

POLARIS VERSIÓN 2.0

Una Herramienta para la Minería Web de Uso y Estructura

JHON CRISTIAN ILES LÓPEZ
CARLOS ANDRÉS RUBIO CORAL

UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2010

POLARIS VERSIÓN 2.0

Una Herramienta para la Minería Web de Uso y Estructura

JHON CRISTIAN ILES LÓPEZ
CARLOS ANDRÉS RUBIO CORAL

Trabajo de grado presentado como requisito parcial para optar al título de
Ingenieros de Sistemas

Director del Proyecto
RICARDO TIMARÁN PEREIRA, Ph. D.

UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2010

“Las ideas y conclusiones aportadas en el trabajo de grado son responsabilidad exclusiva de sus autores”.

Artículo 1º. Del acuerdo No. 324 del 11 de Octubre de 1966 emanado del honorable Consejo Directivo de la Universidad de Nariño.

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

San Juan de Pasto, Octubre 11 de 2010

DEDICATORIA

A Dios, porque en su infinita sabiduría me doto de herramientas maravillosas; puso un sueño, una ilusión y un soplo de esperanza. Me dio unas inmensas alas para ir en busca de mi meta y guió mi camino.

A mis padres que cultivaron en forma perseverante en mí el espíritu de lucha y superación, me infundieron la confianza y me han fortalecido constantemente dándome una palabra de aliento, una voz de consuelo en los momentos difíciles y enseñándome que no hay limitación más grande que la que se ponga uno mismo. Alcanzar este peldaño en mi vida es mérito también de ellos.

A mis hermanas que han creído en mí y me han apoyado incondicionalmente en el sendero hacia este logro.

A mis docentes, por su sabiduría, conocimiento y ejemplo, porque me motivaron a continuar.

A todos quienes estuvieron cerca y me brindaron toda la colaboración que necesité.

Hoy les dedico este trabajo de grado, porque en cada página hay algo de cada uno de ustedes.

Carlos Andrés Rubio Coral

DEDICATORIA

He culminado una importante meta en mi vida, he hecho de un maravilloso sueño una realidad, con esfuerzos, dedicación y trabajo constante he conseguido culminar mis estudios universitarios y este trabajo de grado como muestra de ello, lo dedico a Dios por bendecirme cada día, por haber puesto en mi camino a personas que han sido un soporte y ayuda a lo largo de esta etapa.

A mis padres por creer en mí, por enseñarme a luchar por los sueños, a vencer adversidades, a crecer en los momentos difíciles y que han forjado en mi la perseverancia, el empeño e innumerables valores que hoy hacen de mi una gran persona, a mi Madre especialmente por compartir mis sueños, por su amor, apoyo y esfuerzo incondicional, por encomendarme en cada una de sus oraciones elevadas al cielo, a mi Padre porque sé que desde el cielo me cuidas y oras por mí y sé que ahora estas sonriendo.

Les dedico este trabajo de grado a todas aquellas personas con las que no pude compartir tiempo y sacrifique todos esos pequeños y valiosos momentos de la vida por culminar esta meta especialmente a mis hermanos por sus esfuerzos y buenos deseos a mis sobrinos, a mis familiares más cercanos, a los amigos y conocidos que siempre estuvieron ahí proporcionándome una palabra de aliento y aportándome fuerzas para continuar sin desmayar, a todos los que hicieron posible la realización de esta meta, a ustedes dedico este triunfo, comparto mi felicidad y con ustedes estoy eternamente agradecido.

Jhon Cristian Iles López

AGRADECIMIENTOS

Agradecemos a Dios, por ser quien ha estado a nuestro lado en todo momento, brindándonos energía y fortaleza constantes en cada paso durante nuestra carrera.

A nuestros familiares por su apoyo, comprensión y colaboración incondicional y permanente.

Al Ingeniero Ricardo Timaran Pereira Ph. D., Director de nuestro proyecto, por su tiempo, paciencia y asistencia continúa durante el desarrollo del proyecto.

A nuestros amigos y compañeros, por ser partícipes en el proceso de esta etapa de nuestra vida.

A nuestros profesores por compartir con nosotros sus conocimientos y experiencias.

Y a todas aquellas personas que contribuyeron en la realización de este proyecto de investigación.

GLOSARIO

Autoridad (Authority): Valora la página como recurso de información; para su cálculo se usa una suma ponderada de valores hub de los enlaces que apuntan hacia esta página. Una Web es una autoridad si contiene muchos enlaces entrantes.

Calificación (Score): Es el valor numérico asignado por cada uno de los algoritmos de minería de estructura Web, a este también se le conoce como peso o importancia del sitio Web.

Centro de Actividad (Hub): Valora la información que se consigue siguiendo los enlaces que tiene a otras páginas; se calcula como una suma ponderada de valores authority de las páginas a las que apunta ésta. Una Web es un hub si contiene muchos enlaces salientes.

Dominio: Un dominio de Internet es una red de identificación asociada a un grupo de dispositivos o equipos conectados a la red Internet. El propósito principal de los nombres de dominio en Internet y del sistema de nombres de dominio (DNS), es traducir las direcciones IP de cada nodo activo en la red, a términos memorizables y fáciles de encontrar. Esta abstracción hace posible que cualquier servicio (de red) pueda moverse de un lugar geográfico a otro en la red Internet, aun cuando el cambio implique que tendrá una dirección IP diferente.

Hiperenlace (Link): Link, hipervínculo, vínculo. Conexión entre dos equipos o nodos. Conexión de una página Web con otra mediante una palabra que representa una dirección de Internet (Url). Generalmente un enlace está subrayado y es azul. También sirve para descarga de ficheros, abrir ventanas, etc.

Enlace Dinámico: Los enlaces dinámicos cambian de forma aleatoria. En este tipo de enlace se establece una biblioteca de código, que se entrelaza cuando se ejecuta una aplicación o un programa.

Enlaces Entrante: Es un enlace o hipervínculo hacia una determinada página Web que proviene desde otra página Web.

Enlace Estático: Son los enlaces o direcciones que no cambian y se producen en el tiempo de compilación.

Enlaces Externos: Son enlaces o vínculos que no pertenecen a un mismo dominio de un sitio Web.

Enlaces Internos: Son los enlaces o vínculos que pertenecen a un mismo dominio de un sitio Web.

Enlaces Salientes: Enlace (o hipervínculo) que apunta a una página Web que puede pertenecer a otro sitio Web o al propio. Un enlace saliente es considerado un enlace entrante para el sitio Web que es apuntado.

Factor de Amortiguación (Dampening): Representa la probabilidad de que un navegante continúe pulsando links al navegar por Internet en vez de escribir una URL directamente en la barra de direcciones o pulsar uno de sus marcadores. Por lo tanto, la probabilidad de que el usuario deje de pulsar links y navegue directamente a otra Web aleatoria es 1-d. La introducción del factor de amortiguación en la fórmula resta algo de peso a todas las páginas de Internet y consigue que las páginas que no tienen enlaces a ninguna otra página no salgan especialmente beneficiadas. Si un usuario aterriza en una página sin enlaces, lo que hará será navegar a cualquier otra página aleatoriamente, lo que equivale a suponer que una página sin enlaces salientes tiene enlaces a todas las páginas de Internet.

Grafo: Es un conjunto de nodos con enlaces entre ellos, denominados aristas o arcos.

Indizable (Indexable): Nombre que se le da al proceso de recolectar y almacenar páginas Web por parte de un buscador de internet. Luego de este proceso, pueden estar disponibles en las páginas con resultados de búsquedas que se realicen en ese buscador.

Página semilla: Es una Url inicial con la cual empieza a trabajar el crawler, esta se utiliza para identificar los enlaces que salen a otras páginas Web.

Recuperación de la Información: La recuperación de información, llamada en inglés information retrieval (IR), es la ciencia de la búsqueda de información en documentos, búsqueda dentro de los mismos, búsqueda de metadatos que describan documentos, o también la búsqueda en bases de datos relacionales, ya sea a través de internet, intranet, para textos, imágenes, sonido o datos de otras características, de manera pertinente y relevante.

URL: Un localizador uniforme de recursos, más comúnmente denominado URL (sigla en inglés de uniform resource locator), es una secuencia de caracteres, de acuerdo a un formato modélico y estándar, que se usa para nombrar recursos en Internet para su localización o identificación, como por ejemplo documentos textuales, imágenes, videos, presentaciones digitales, etc.

RESUMEN

Este documento contiene el análisis y diseño del trabajo de grado: **POLARIS V2.0**
– Una Herramienta de Minería de Uso y Estructura Web.

“**POLARIS V2.0**” es un software que se desarrolló pensando en la necesidad de crear una Suite de Minería Web, básicamente busca descubrir patrones de uso y mejorar la estructura de un portal Web en particular, dicha herramienta se encuentra orientada hacia los administradores Web.

Esta herramienta se encuentra dividida en dos grandes módulos: El módulo de Minería de Uso Web, implementado en la primera versión de Polaris y el módulo de Minería de Estructura Web implementado en su segunda versión, y sobre la cual trata el presente documento.

La principal característica de **POLARIS V2.0** es que tiene la posibilidad de trabajar en tiempo real, es decir recuperando datos de la Web teniendo acceso a internet, o se lo puede hacer con datos guardados con anterioridad en la base de datos de la herramienta o si se lo desea en un archivo plano, el cual debe cumplir con el formato que solicita el software para poder ser cargado.

Además, se tiene la posibilidad de utilizar diferentes clases de filtros con el objetivo de mirar la información que en verdad necesita el usuario y no redundar con información que puede ser no necesaria y de poco interés. Estos filtros se presentan en lo concerniente a un nivel de profundidad, enlaces, protocolos y dominios de la URL semilla o página a ser analizada.

Por otra parte, en la categoría de Minería de Estructura Web, se cuenta con cuatro algoritmos implementados, tres de los cuales se encargan de calcular los diferentes pesos de las páginas Web y uno más que se encarga de descubrir la ruta más corta y optima que existe de una página a otra para desplazarnos en la búsqueda de la información que deseamos.

Finalmente cuenta con las diferentes vistas que presenta la herramienta, las cuales aparte de ser muy llamativas son de gran utilidad para tener una imagen de la información recolectada de la Web. Se presentan cuatro tipos de vistas gráficas, las cuales son grafos, radiales, dominios y árboles y por ultimo pero no menos importante se presenta una visualización en tablas de la información procesada para ser interpretada por el usuario.

ABSTRACT

This document contains the analysis and design of the thesis: POLARIS V2.0 - A Tool for Mining Web Usage and Structure.

"POLARIS V2.0" is a software that was developed with the need to create a suite of Web Mining, basically seeks to discover usage patterns and improve the structure of a particular Web site, this tool is oriented to Web managers.

This tool is divided into two main modules: Web Usage Mining, implemented in the first version of Polaris and the module Web Structure Mining implemented in its second version, and which is herein.

The main feature of POLARIS V2.0 is that it has the ability to work in real time, retrieving data from the Web to have access to the internet or it can be done with data previously stored in the database or tool if desired in a flat file, which must comply with the format requested by the software to be loaded.

In addition, there is the possibility of using different kinds of filters in order to look at the information you really need the user and not redundant with information that may be unnecessary and of little interest. These filters are presented in relation to links, protocols and domains of the seed URL or page to be analyzed.

Moreover, in the category of Web Structure Mining, there are four algorithms implemented, three of which are responsible for calculating the different weights of Web pages and one in charge of finding the shortest path and optimal to There is one page to another to move in search of information we want.

Finally, with the different views presented by the tool, which apart from being very showy are very useful to have an image of information dug give the Web. Four types of graphical views, which are graphs, radio, domains, and trees and one more which is entered the information into tables to be performed by the user.

CONTENIDO

	Pág.
INTRODUCCIÓN	27
1. LA TRIPLE WORLD WIDE WEB	32
1.1 DEFINICIÓN Y PRINCIPALES CARACTERÍSTICAS	32
1.2 PARTES DEL WORLD WIDE WEB	33
1.3 TAMAÑO DEL WORLD WIDE WEB	34
1.4 INFOMETRÍA DE LA WEB	35
1.5 ESTRUCTURA DE LA WEB	36
2. MINERÍA WEB	39
2.1 TIPOS DE MINERÍA WEB	39
2.1.1 Minería del contenido de la web	40
2.1.2 Minería de uso de la web	41
2.1.3 Minería de la estructura de la web	41
2.2 ETAPAS DE LA MINERÍA WEB	42
2.2.1 Descubrimiento de las fuentes	42
2.2.2 Selección/extracción y preprocesamiento	43
2.2.3 Generalización	44
2.2.4 Análisis	45
2.3 MINERÍA DE DATOS	45
2.3.1 Fases de la minería de datos	46
2.4 RECUPERACIÓN DE INFORMACIÓN Y BÚSQUEDA EN LA WEB	48
2.4.1 La búsqueda web	48
2.4.2 Arquitectura de los sistemas de recuperación de información	50
2.4.3 Recursos críticos en las búsquedas	51
2.4.4 Motores de búsqueda o buscadores	51
2.4.5 Directorios o índices	52
2.4.6 Los metabuscadores	53

2.4.7	Agentes inteligentes.....	54
2.5	WEB CRAWLER	55
2.5.1	Políticas de rastreo.	56
2.5.2	Arquitectura del crawler.....	57
2.5.3	Ejemplos de rastreadores web.....	58
2.5.4	Funcionamiento del web crawler implementado en Polaris v2.0.....	59
3.	MINERÍA DE LA ESTRUCTURA DE LA WEB	60
3.1	PATRONES DE ENLACE	60
3.2	ALGORITMOS DE ORDENACIÓN POR RELEVANCIA BASADOS EN LA CONECTIVIDAD	63
3.2.1	Algoritmos dependientes de la consulta del usuario	64
3.2.2	Algoritmos cuasi-dependientes de la consulta del usuario.....	64
3.2.3	Algoritmos independientes de la consulta del usuario	65
3.3	ALGORITMOS DE MINERÍA DE ESTRUCTURA WEB IMPLEMENTADOS EN POLARIS VERSIÓN 2.0.....	65
3.3.1	Page Rank	66
3.3.2	HITS.....	77
3.3.3	Web Linker.....	87
4.	ANÁLISIS DE HERRAMIENTAS.....	93
4.1	POLARIS VERSIÓN 1.0.....	93
4.1.1	Descripción preliminar.....	93
4.1.2	Requisitos del sistema.	93
4.1.3	Instalación.....	93
4.1.4	Manejo de la herramienta.....	94
4.1.5	Ventajas.	97
4.1.6	Desventajas.	97
4.2	TOUCHGRAPH NAVIGATOR VERSIÓN 1.3.1.....	97
4.2.1	Descripción Preliminar.	97
4.2.2	Requisitos del Sistema.....	98
4.2.3	Instalación.....	98
4.2.4	Manejo de la Herramienta.....	98

4.2.5	Ventajas.	107
4.2.6	Desventajas.	108
4.3	SURF PATTERN VISUAL ANALYZER.	108
4.3.1	Descripción preliminar.....	108
4.3.2	Requisitos del sistema.	108
4.3.3	Instalación.	109
4.3.4	Manejo de la herramienta.....	109
4.3.5	Desventajas.	112
4.4	SEO ADMINISTRATOR VERSIÓN 3.11	113
4.4.1	Descripción preliminar.....	113
4.4.2	Requisitos del sistema.	114
4.4.3	Instalación.	114
4.4.4	Manejo de la herramienta.....	114
4.4.5	Ventajas.	128
4.4.6	Desventajas.	128
4.5	A1 WEBSITE ANALYZER VERSIÓN 1.4.1	128
4.5.1	Descripción preliminar.....	128
4.5.2	Requisitos del sistema.	129
4.5.3	Instalación.	129
4.5.4	Manejo de la herramienta.....	129
4.5.5	Ventajas.	134
4.5.6	Desventajas.	135
4.6	OTRAS HERRAMIENTAS.....	135
4.6.1	Page Rank Checker Versión 3.0	135
4.6.2	Surf 3d Beta.	136
4.6.3	NetTools Spider.	137
5.	METODOLOGÍA DE DESARROLLO	138
5.1	PROGRAMACIÓN EXTREMA (XP)	138
5.1.1	Fase de planificación.....	139
5.1.2	Fase de diseño.....	140
5.1.3	Fase de desarrollo.	142
5.1.4	Fase de pruebas.	143

6.	RESULTADOS DE LA INVESTIGACIÓN	145
6.1	FASE DE PLANEACIÓN	145
6.1.1	Historias de usuario	145
6.2	PLAN DE ENTREGAS	164
6.2.1	Cronograma de actividades	164
6.2.2	Iteraciones.....	168
6.2.3	Velocidad del proyecto.....	168
6.3	FASE DE DISEÑO	169
6.3.1	Metáfora del sistema.....	169
6.4	TARJETAS CRC	169
6.4.1	Tarjeta CRC – Ventana.....	170
6.4.2	Tarjeta CRC – crawlerRun.....	170
6.4.3	Tarjeta CRC – crawlerSettings.....	171
6.4.4	Tarjeta CRC – txtSettings.....	171
6.4.5	Tarjeta CRC – Conexión.....	172
6.4.6	Tarjeta CRC – Nodo.....	172
6.4.7	Tarjeta CRC – Área de trabajo.....	173
6.4.8	Tarjeta CRC – Configuración.....	173
6.4.9	Tarjeta CRC – domainViewFrame.....	174
6.4.10	Tarjeta CRC – graphViewFrame.....	174
6.4.11	Tarjeta CRC – radialViewFrame.....	175
6.4.12	Tarjeta CRC – Enlace	176
6.4.13	Tarjeta CRC – Host.....	176
6.4.14	Tarjeta CRC – Página.....	177
6.4.15	Tarjeta CRC – Sitio.....	177
6.5	FASE DE DESARROLLO.....	177
6.5.1	Unidades de Prueba	178
6.6	FASE DE PRUEBAS	186
6.6.1	Implementación.....	186
6.6.2	Pruebas de aceptación.....	187
7.	IMPLEMENTACIÓN DE LA HERRAMIENTA POLARIS VERSIÓN 2.0	212

7.1	ARQUITECTURA	212
7.1.1	Módulo de utilidades	212
7.1.2	Módulo del Kernel <i>POLARIS V2.0</i>	212
7.1.3	Módulo de la Interfaz Gráfica.	213
7.2	ESTRUCTURA DE PAQUETES DE APLICACIÓN	215
7.2.1	Paquete Utils.....	217
7.2.2	Paquete polaris2	217
7.2.3	Polaris2.dataSource.....	218
7.2.4	Paquete polaris2.graph	220
7.2.5	Paquete polaris2.graphview	224
7.2.6	Polaris2.mining.....	226
7.2.7	Polaris2.preprocess	229
7.2.8	Paquete polaris2.site.....	230
7.2.9	Paquete polaris2.view	232
7.3	LENGUAJES Y HERRAMIENTAS PARA EL DESARROLLO DE POLARIS VERSIÓN 2.0	232
7.3.1	Lenguaje de Programación Java.....	232
7.3.2	Entorno de Desarrollo Netbeans.	235
7.3.3	PostgreSQL.....	236
7.3.4	Controlador JDBC.	237
7.3.5	WebLA	239
7.3.6	Herramienta de Visualización Prefuse.	239
	CONCLUSIONES.....	240
	RECOMENDACIONES	241
	ANEXOS	2412
	REFERENCIA BIBLIOGRÁFICA.....	243

ÍNDICE DE TABLAS

	Pág.
Tabla 1. Historia de usuario No. 1 - Aplicación de Escritorio.....	146
Tabla 2. Historia de usuario No. 2 - Creación del módulo de minería de estructura.	146
Tabla 3. Historia de usuario No. 3 - Agregar tablas a la base de datos de Polaris.	147
Tabla 4. Historia de usuario No. 4 - Creación de un Web Crawler.....	147
Tabla 5. Historia de usuario No. 5 - Nivel de profundidad del Web Crawler.....	148
Tabla 6. Historia de usuario No. 6 - URL a analizar.	148
Tabla 7. Historia de usuario No. 7 - Filtros para la búsqueda del Crawler.	149
Tabla 8. Historia de usuario No. 8 - Guardar datos del Web Crawler.....	149
Tabla 9. Historia de usuario No. 9 - Implementación de los algoritmos de Minería Web de Estructura	150
Tabla 10. Historia de usuario No. 10 - Visualización de grafos.	150
Tabla 11. Historia de usuario No. 11 - Mejorar la interfaz de Polaris V 1.0.	151
Tabla 12. Historia de usuario No. 12 - Integración de la interfaz.....	151
Tabla 13. Historia de usuario No. 13. - Integración de los algoritmos.	152
Tabla 14. Historia de usuario No. 14 - Preprocesamiento de la información .. recolectada por el Web Crawler.	152
Tabla 15. Historia de usuario No. 15 - Enlaces internos y externos.	153
Tabla 16. Historia de usuario No. 16 - Enlaces entrantes y salientes.	153
Tabla 17. Historia de usuario No. 17 - Cargar información de base de datos	154
Tabla 18. Historia de usuario No. 18 - Visualización de grafos con sus filtros. ..	154

Tabla 19. Historia de usuario No. 19 - Visualización de grafos de los enlaces internos	155
Tabla 20. Historia de usuario No. 20 - Visualización de grafos de los enlaces externos	155
Tabla 21. Historia de usuario No. 21- Configuración de los algoritmos.....	156
Tabla 22. Historia de usuario No. 22 - Parámetros Page Rank.....	156
Tabla 23. Historia de usuario No. 23 - Parámetros HITS.	157
Tabla 24. Historia de usuario No. 24 - Parámetros algoritmo creado.....	157
Tabla 25. Historia de usuario No. 25 - Visualización de información de cada nodo.	158
Tabla 26. Historia de usuario No. 26 - Cálculo de los algoritmos.	158
Tabla 27. Historia de usuario No. 27 - Visualización del grafo final.....	159
Tabla 28. Historia de usuario No. 28 - Algoritmo de la ruta más corta entre dosnodos seleccionados	159
Tabla 29. Historia de usuario No. 29 - Creación de filtros según el protocolo de las páginas Web.	160
Tabla 30. Historia de usuario No. 30 - Reporte en tablas de los cálculos.	160
Tabla 31. Historia de usuario No. 31 - Exportación de datos a otros formatos. .	161
Tabla 32. Historia de usuario No. 32 - Resultados en la mejora de la estructura.	161
Tabla 33. Historia de usuario No. 33. Eliminar datos de la base de datos.	162
Tabla 34. Historia de usuario No. 34. Creación de un nuevo proyecto.	162
Tabla 35. Historia de usuario No. 35 - Cargar un proyecto existente.....	163
Tabla 36. Historia de usuario No. 36 - Guardar proyecto.	163
Tabla 37. Historia de usuario No. 37 - Guardar proyecto como.	164
Tabla 38. Cronograma de Actividades de las Historias de Usuario.....	164

Tabla 39. Versiones de POLARIS V2.0.....	167
Tabla 40. Tarjeta CRC – Ventana.	170
Tabla 41. Tarjeta CRC – crawlerRun.	170
Tabla 42. Tarjeta CRC – crawlerSettings.	171
Tabla 43. Tarjeta CRC – txtSettings.....	171
Tabla 44. Tarjeta CRC – Conexión.	172
Tabla 45. Tarjeta CRC – Nodo.....	172
Tabla 46. Tabla CRC – Área de trabajo	173
Tabla 47. Tarjeta CRC – Configuración.	173
Tabla 48. Tarjeta CRC – domainViewFrame.....	174
Tabla 49. Tarjeta CRC – graphViewFrame.	174
Tabla 50. Tarjeta CRC – radialViewFrame.....	175
Tabla 51. Tabla CRC – Enlace.....	176
Tabla 52. Tabla CRC – Host.	176
Tabla 53. Tarjeta CRC – Página.	177
Tabla 54. Tarjeta CRC – Sitio.....	177
Tabla 55. Test aplicado a la clase: Ventana.....	178
Tabla 56. Test aplicado a la clase: CrawlerRun	179
Tabla 57. Test aplicado a la clase: CrawlerSettings.....	179
Tabla 58. Test aplicado a la clase: TxtSettings	180
Tabla 59. Test aplicado a la clase: Conexión.....	180
Tabla 60. Test aplicado a la clase: Nodo	180
Tabla 61. Test aplicado a la clase: Área de Trabajo	181

Tabla 62. Test aplicado a la clase: Configuración.....	182
Tabla 63. Test aplicado a la clase: domainVlewFrame.....	183
Tabla 64. Test aplicado a la clase: graphViewFrame.....	183
Tabla 65. Test aplicado a la clase: radialViewFrame.....	184
Tabla 66. Test aplicado a la clase: Enlace.....	184
Tabla 67. Test aplicado a la clase: Host.....	185
Tabla 68. Test aplicado a la clase: Página.....	185
Tabla 69. Test aplicado a la clase: Sitio.....	186
Tabla 70. Prueba de aceptación – Iniciar uso de la aplicación.....	188
Tabla 71. Prueba de aceptación – Ingreso al módulo de Minería de EstructuraWeb	189
Tabla 72. Prueba de aceptación – Creación De la base de datos para POLARIS VERSIÓN 2.0.	190
Tabla 73. Prueba de aceptación – Ingreso de la URL a analizar.....	192
Tabla 74. Prueba de aceptación – Nivel de profundidad de la búsqueda.....	193
Tabla 75. Prueba de aceptación – Filtros para el Web Crawler.....	194
Tabla 76. Prueba de aceptación – Recuperación de la información por medio delCrawler	196
Tabla 77. Prueba de aceptación – Recuperar los datos recuperados por elCrawler.	197
Tabla 78. Prueba de aceptación – Visualización de los diferentes tipos de grafos.	198
Tabla 79. Prueba de aceptación – Visualización de los enlaces internos y externos.	199
Tabla 80. Prueba de aceptación – Visualización de los enlaces entrantes y salientes.	200

Tabla 81. Prueba de aceptación – Visualización del grafo teniendo en cuenta losfiltros	202
Tabla 82. Prueba de aceptación – Configuración del nodo Page Rank.	203
Tabla 83. Prueba de aceptación – Configuración del nodo HITS.....	204
Tabla 84. Prueba de aceptación – Configuración del nodo Web Linker.....	204
Tabla 85. Prueba de aceptación – Visualización de la información por cada nodo.	205
Tabla 86. Prueba de aceptación – La ruta más corta.....	206
Tabla 87. Prueba de aceptación – Reportes en tablas.....	207
Tabla 88. Prueba de aceptación – Exportar datos.	207
Tabla 89. Prueba de aceptación – Sugerencias para mejorar la estructura Web.	208
Tabla 90. Prueba de aceptación – Crear proyecto.....	209
Tabla 91. Prueba de aceptación – Cargar proyecto.....	209
Tabla 92. Prueba de aceptación – Guardar proyecto.....	210
Tabla 93. Prueba de aceptación – Guardar proyecto por primera vez.	211

ÍNDICE DE FIGURAS

	Pág.
Figura 1. Partes de la World Wide Web	33
Figura 2. Grafo dirigido que representa la estructura de la Web.....	37
Figura 3. Mapa conceptual de la Minería Web.....	40
Figura 4. Fases de la Minería Web.	42
Figura 5. Fases del proceso KDD	46
Figura 6. Fases dentro de un proceso de Minería de Datos	46
Figura 7. Técnicas de recuperación de información en la Web.....	48
Figura 8. Arquitectura de un sistema de recuperación de información.....	50
Figura 9. Acciones realizadas por un metabuscador al procesar una búsqueda. .	54
Figura 10. Arquitectura de alto nivel de rastreador Web normal.	57
Figura 11. La Web representada a través de un dígrafo.....	61
Figura 12. Matriz de adyacencia del grafo de la Figura 11.....	61
Figura 13. Patrones básicos formados por dos enlaces.....	62
Figura 14. Patrones complejos formados por numerosos nodos y vínculos.	62
Figura 15. Grafo No. 1 para el cálculo del Page Rank (Ejemplo 1)	71
Figura 16. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 PageRank)	72
Figura 17. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 1 PageRank)	73
Figura 18. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 1 PageRank)	73

Figura 19. Grafos de la primera iteración para el nodo No. 4 (Ejemplo 1 PageRank)	74
Figura 20. Grafos de la primera iteración para el nodo No. 5 (Ejemplo 1 PageRank)	74
Figura 21. Grafo No. 2 para el cálculo del Page Rank (Ejemplo 2).....	75
Figura 22. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 2 PageRank)	76
Figura 23. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 2 PageRank)	76
Figura 24. Esquemas de una autoridad (authority) y un centro (hub).	78
Figura 25. Ejemplo de un conjunto base S. El conjunto S contiene al Conjunto W (Root Set) y a las páginas añadidas en la expansión	79
Figura 26. La finalidad de HITS es encontrar un conjunto de páginas relacionadas con el tema de búsqueda que tengan alta puntuación como autoridad y centro.	80
Figura 27. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 HITS) ...	81
Figura 28. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 1 HITS) ...	82
Figura 29. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 1 HITS) ...	82
Figura 30. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 HITS) ...	82
Figura 31. Grafo No. 2 para el cálculo de Hits (Ejemplo 2)	84
Figura 32. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 2 HITS) ...	84
Figura 33. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 2 HITS) ...	84
Figura 34. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 2 HITS) ...	85
Figura 35. Grafos de la primera iteración para el nodo No. 4 (Ejemplo 2 HITS) ...	85
Figura 36. Grafos de la primera iteración para el nodo No. 5 (Ejemplo 2 HITS) ...	86
Figura 37. Grafos de la primera iteración para el nodo No. 6 (Ejemplo 2 HITS) ...	86

Figura 38. Grafo No. 1 para el cálculo del Web Linker (Ejemplo 1).....	87
Figura 39. Mensaje de advertencia para configuración de la base de datos.....	94
Figura 40. Formulario de configuración de la base de datos.....	94
Figura 41. Configuración del nodo Web Server.	95
Figura 42. Formulario de extensiones validadas para realizar el análisis.	96
Figura 43. Área de trabajo de POLARIS.	97
Figura 44. Logo TouchGraph NAVIGATOR	98
Figura 45. Área de trabajo TouchGraph NAVIGATOR.....	100
Figura 46. Archivo de Excel, ejemplo TouchGraph NAVIGATOR	101
Figura 47. Crear o abrir un archivo de Excel.....	102
Figura 48. Ficha “ítem”.....	103
Figura 49. Cuadro de campo.....	104
Figura 50. Ejemplo de la interfaz, con el programa TouchGraph Navigator, para el fichero .xml	106
Figura 51. Ejemplo del gráfico del grafo que se forma con la información suministrada.	107
Figura 52. Formulario de creación de perfiles.	109
Figura 53. Alimentación de archivos para Surf Pattern Visual Analyzer versión 2.5.	110
Figura 54. Ventana para el análisis visual de Surf Pattern Visual Analyzer versión 2.5	110
Figura 55. Detalles del nodo de Surf Pattern Visual Analyzer versión 2.5	111
Figura 56. Opciones del nodo para Surf Pattern Visual Analyzer versión 2.5	112
Figura 57. Ventana de presentación SUITE de SEO Administrator	115
Figura 58. Área de trabajo de Ranking Monitor.....	116

Figura 59. Informe 1 de Posicionamiento de Ranking Monitor	117
Figura 60. Informe 2 de posicionamiento de Ranking Monitor	118
Figura 61. “URL list” de Ranking Monitor	119
Figura 62. “History Table”, de Ranking Monitor.....	120
Figura 63. Área de trabajo de Link Popularity Checker	121
Figura 64. Reporte que generado la herramienta software Link Popularity Checker	122
Figura 65. Reporte generado por un periodo de tiempo, software Link Popularity Checker	123
Figura 66. Área de trabajo Page Rank Analyzer	125
Figura 67. Área de trabajo Log Analyzer	126
Figura 68. Reporte creado por Log Analyzer	127
Figura 69. Área de trabajo A1 Website Analyzer.....	130
Figura 70. Estructura de enlace del sitio Web analizado por A1 Website Analyzer	131
Figura 71. Reporte generado por la herramienta A1 Website Analyzer	132
Figura 72. Exportar reportes de la herramienta A1 Web Analyzer	134
Figura 73. Fases de la metodología XP	138
Figura 74. Ejemplo de una Tarjeta CRC	141
Figura 75. Arquitectura general para POLARIS	214
Figura 76. Arquitectura POLARIS V2.0	215
Figura 77. Ventana que representa la interfaz gráfica de la herramienta Polaris	218
.....versión 2.0.	
Figura 78. Ventana de ejecución para el Web Crawler.	219

Figura 79. Ventana de configuración para el Web Crawler.	219
Figura 80. Ventana que representa la selección de un archivo de texto plano. ..	220
Figura 81. Ventana para la representación del gráfico Domain.	225
Figura 82. Ventana para la representación del gráfico GraphView.	225
Figura 83. Ventana para la representación del gráfico Radial.....	226
Figura 84. Ventana de ejecución algoritmo HITS.	227
Figura 85. Ventana configuración del algoritmo HITS.....	227
Figura 86. Ventana de ejecución algoritmo Page Rank.	227
Figura 87. Ventana de configuración algoritmo Page Rank.	228
Figura 88. Ventana de ejecución algoritmo Web Linker.	228
Figura 89. Ventana de ejecución del nodo enlaces (Links).	229
Figura 90. Ventana que representa la configuración del nodo enlaces (Links). ..	230
Figura 91. Controlador JDBC tipo 4	238

INTRODUCCIÓN

El gran volumen de información contenido en la Web hace cada vez más complejo y caótico el proceso de búsqueda de la misma, sobre todo cuando no se cuentan con los recursos necesarios para llegar a ella o se desconoce una forma de acceso rápido que garantice la obtención del recurso deseado en medio de un mar de información.

En los últimos años se ha planteado el desarrollo de nuevas tecnologías que buscan ofrecer un mejor servicio de la internet, haciendo uso de tecnologías como la minería de datos orientada a la Web o Web Mining, cuyo objetivo es aprovechar la información que circula en Internet para descubrir los diferentes patrones de los usuarios que navegan en la red.

Existen tres tipos de Minería Web: Minería de Uso Web, Minería de Contenido Web y Minería de Estructura Web.

Minería de Uso de la Web (Web Usage Mining). Es la aplicación de las técnicas de data mining para descubrir pautas de conducta a la hora de utilizar la Web por parte de los usuarios.

Esta extracción se refiere a patrones de navegación que se puede descubrir en los usuarios y pueden servir para mejorar la misma, por ejemplo si el 80% de los usuarios recurren al campo de búsqueda cuando entran a un sitio, se debe poner énfasis en la mejora de esa interfaz y que el motor que se encuentre detrás devuelva la información deseada. Este proceso se basa en el uso de los archivos logs de los servidores Web, en donde son almacenados en forma de bitácora, las diferentes peticiones hechas a los servidores.

En conclusión, se tratan seguir una serie de pautas sobre el acceso que utilizan los clientes cuando consultan el sitio Web de una empresa, los usuarios que interrogan a una aplicación que precede a una base de datos y los individuos que navegan por páginas determinadas.

"A partir de datos secundarios derivados de interacciones automáticas de los usuarios mientras navegan por la Web se pueden cubrir mejor las necesidades que se solicitan a través de aplicaciones basadas en protocolos Web." [45]

Minería del Contenido de la Web (Web Content Mining). Busca la regularidad y dinámica de los contenidos en la Web. Los documentos Web pueden ser datos sin estructurar, archivos html parcialmente estructurados, o información procedente de bases de datos generadas en páginas con formato html. Estos documentos de hipertexto incluyen texto, imágenes, audio, vídeo, metadatos e hiperenlaces.

La metodología utilizada en este tipo de minería, va desde las tradicionales relaciones entre términos hasta la tecnología que se utiliza en la minería textual, esta última consiste en analizar elementos textuales con el fin de identificar, deducir y ampliar conocimiento a partir de cualquier organización de documentos.

La extracción de información, intenta inferir la estructura del sitio Web para transformarla y convertirla en una base de datos a nivel lógico.

Además, la localización de patrones en el texto de los documentos, el descubrimiento del recurso basado en conceptos de indexación o la tecnología basada en agentes también puede formar parte de esta categoría. [45]

Minería de la Estructura Web (Web Structure Mining). Intenta descubrir la organización de los enlaces del conjunto de hiperenlaces dentro del documento para generar un informe estructural sobre la página y el sitio Web. Se obtiene información acerca de si los usuarios encuentran la información, si la estructura de sitio es demasiado ancha o demasiado profunda, si los elementos están colocados en los lugares adecuados dentro de la página, si la navegación se entiende, cuáles son las secciones menos visitadas y su relación con el lugar que ocupan en la página central.

Según el objetivo a estudiar, se pueden dar tres tipos de informes: Basándose en los hiperenlaces, clasifica las páginas Web y genera el informe. Revelando la estructura del documento Web en sí y descubriendo la naturaleza de la jerarquía o de la red de hiperenlaces del sitio Web de un dominio particular.

Suele dar como resultado representaciones gráficas para una mejor visión del conocimiento obtenido y pueden utilizarse como guía para el usuario en busca de información. [45]

En la Universidad de Nariño se desarrolló una Herramienta de Minería Web de Uso, denominada **POLARIS** [2], dicha herramienta fue creada en con el lenguaje de programación Java, bajo la licencia de software libre. En este proyecto se logra la construcción de una nueva versión de la herramienta **POLARIS V2.0**, a la cual se la complementó con el módulo de Minería Web de Estructura, transformándola así en una herramienta más competitiva y ampliando su utilidad a miras de convertirla en una Suite de Minería Web.

ALCANCE Y DELIMITACIÓN

El desarrollo de la herramienta se inició a partir de la primera versión de **POLARIS**. Donde se desarrolló e implementó el Módulo de Minería de Estructura Web para **POLARIS VERSIÓN 2.0**. En lo concerniente a la visualización se desarrolló una Interfaz Gráfica que le permite al usuario una sencilla interacción con la herramienta.

Las pruebas de rendimiento para los algoritmos implementados se realizaron utilizando conjuntos de datos reales; del portal de la Universidad de Nariño y otros archivos públicos obtenidos en la Web.

PROBLEMA OBJETO DE ESTUDIO

El Grupo de Investigación GRIAS del programa de Ingeniería de Sistemas de la Universidad de Nariño, con la dirección de RICARDO TIMARAN PEREIRA Ph. D., y la participación de las ingenieras DIANA PATRICIA ANGULO URBANO, JENNY JOHANA DAZA BURBANO y ALEJANDRA ZULETA MEDINA, en la línea KDD, desarrollaron la Herramienta **POLARIS**, la cual en la actualidad se limita únicamente al dominio de la minería Web de uso.

POLARIS en la versión actual, se ve limitada al momento de solucionar problemas que se encuentren relacionados con el Contenido o con la Estructura de la Web, además, se presenta un desconocimiento por parte de las empresas de la herramienta **POLARIS** y de los servicios que esta presta.

Continuando con el proyecto del Grupo de Investigación GRIAS en su línea KDD, se llevará a cabo la construcción de una herramienta que sea una Suite de Minería Web, se propone en este documento complementar la herramienta **POLARIS**, con el Módulo de Minería Web de Estructura, cumpliendo así con la segunda fase del proyecto, ampliando la cobertura y el campo de acción de ésta herramienta.

OBJETIVOS

Objetivo General.

Potenciar la Herramienta **POLARIS** a través del desarrollo del Módulo de Minería Web de Estructura, con el fin de que pueda soportar decisiones que tengan que ver con la Minería Web de Uso y Estructura.

Objetivos Específicos.

Estudiar y analizar la herramienta **POLARIS** en su primera fase.

Analizar las diferentes herramientas de Minería Web de Estructura existentes.

Estudiar, analizar e implementar los algoritmos de Minería de Estructura Web, los cuales se encuentran en investigación.

Analizar, diseñar e implementar programas que permitan la visualización gráfica de los resultados obtenidos en lo concerniente a la Minería de Estructura Web.

Integrar todos los programas en una sola herramienta software **POLARIS VERSIÓN 2.0.**

Obtener depósitos de datos reales del portal de la Universidad de Nariño y otros públicos obtenidos de la Web.

Realizar las pruebas de rendimiento con los depósitos de datos reales en la herramienta **POLARIS VERSIÓN 2.0**.

Analizar los resultados de las pruebas realizadas.

Publicar un artículo que dé a conocer los resultados de la presente investigación.

JUSTIFICACIÓN

“La Minería Web o Web Mining trata de descubrir patrones interesantes en la estructura, el contenido y la utilización de los sitios Web...”. [5]

Se puede definir Web Mining como el proceso de extracción de patrones potencialmente útiles e interesantes y de información implícita (*Data Mining*) de datos provenientes de la Web o, de manera más sencilla, como “la aplicación de técnicas de *Data Mining* a grandes depósitos de datos Web”. [4]

Antes de comenzar a hablar de un proceso de minería, se debe conocer de antemano el tipo de datos al que se le va a aplicar dicha técnica. Hay tres tipos de datos principales en la Web. El más importante y difícil de procesar es el contenido, que es multimedia, en el cual el texto juega un rol dominante. El segundo proviene de la estructura no lineal de la Web; es decir, sus hiperenlaces. Finalmente, el último procede del uso reflejado a través de los logs o bitácoras de los servidores Web.

En la actualidad los tres tipos de Minería Web, no se encuentran integrados en una Suite de Minería Web, sino que por el contrario se han desarrollado diferentes herramientas aisladas entre sí, las cuales resuelven problemas particulares, siendo las más difundidas las aplicaciones de Minería Web de Uso.

Con base en lo anterior, nace la necesidad de actualizar **POLARIS** en **POLARIS VERSIÓN 2.0**, una herramienta de Minería Web de Uso y Estructura, con el grupo GRIAS del programa de Ingeniería de Sistemas de la Universidad de Nariño y bajo la dirección de RICARDO TIMARAN PEREIRA Ph. D.

El desarrollo de **POLARIS VERSIÓN 2.0** como una herramienta software bajo licencia pública GPL para el análisis del tráfico Web, permitirá a todo tipo de empresas, entidades y personas del común, utilicen tecnologías Web Mining como un mecanismo de apoyo en la toma acertada de decisiones.

POLARIS VERSIÓN 2.0, se convierte en un aporte significativo en cuanto a investigación científica de tecnología Web Mining, un reconocimiento más para el

área de Descubrimiento de Conocimiento en Base de Datos de la Universidad de Nariño que a través del programa de Ingeniería de Sistemas contribuye al desarrollo de la región y del país.

ORGANIZACIÓN DEL DOCUMENTO.

El documento inicia con una breve descripción de lo que se va a tratar y la propuesta del presente trabajo de investigación, la cual le brinda al lector una visión general del documento completo. Posteriormente se definen conceptos de gran importancia en la Minería Web, con la finalidad de unificar información, una vez se encuentre los conceptos claros se entra en el tema de la Minería de Estructura Web, el cual es el objeto del presente documento, analizando los algoritmos que van a ser implementados y observando diferentes herramientas que se dedican a trabajar este tipo de minería, analizándolas desde diferentes puntos de vista destacando sus ventaja y debilidades. Siguiendo con el desarrollo del documento encontramos la metodología que se aplicó en el desarrollo del proyecto mostrando sus diferentes fases y pruebas que se llevaron a cabo. Para finalizar están las conclusiones y recomendaciones.

1. LA TRIPLE WORLD WIDE WEB

1.1 DEFINICIÓN Y PRINCIPALES CARACTERÍSTICAS

“Un sistema de documentos de hipertexto y/o hipermedios enlazados y accesibles a través de Internet. Con un navegador Web, un usuario visualiza sitios Web, forjados de páginas Web que pueden contener texto, imágenes u otros contenidos multimedia, y navega a través de ellas usando hiperenlaces”. [58]

Dado que los estudios realizados se van a centrar en la búsqueda y recuperación de información, se va a tratar una abstracción de la Web. Dicho modus operandi no alterará los estudios realizados ya que únicamente se va a “suprimir” aquellos usos de la red que no incumben en tal actividad y que lo único que pueden producir es “ruido” en las argumentaciones.

La Web es el fenómeno más importante que se da en Internet, por ello, ha tomado tal protagonismo que habitualmente se confunden los términos y se designa uno u otro para denominarse entre sí. La búsqueda y recuperación de la información pueden tratarse ambos términos como sinónimos.

La Web tiene muchos usos y aplicaciones. Para el presente estudio de investigación se hará referencia a la Web como base documental, es decir, la Web es un conjunto de información con características particulares cuyo fin es ser consultada por diversos usuarios. Rubén Fuentes en [19] expone que dicha información posee las determinadas características que hacen compleja la labor de recuperación. Dichas características son las siguientes:

- Está distribuida en diversos servidores.
- Es dinámica, cambia constantemente tanto de forma como de contenido.
- No tiene estructura común en todos los documentos. Los documentos carecen de modelo conceptual que estructure semánticamente el contenido.
- Es heterogénea, es decir, está compuesta por partes de diversa naturaleza, ya que soporta grandes cambios de formato y lenguaje.
- Es altamente volátil, aparece información nueva y desaparece la información desactualizada a gran velocidad.
- Es redundante. En numerosas ocasiones se encuentra repetida la información en distintas páginas.

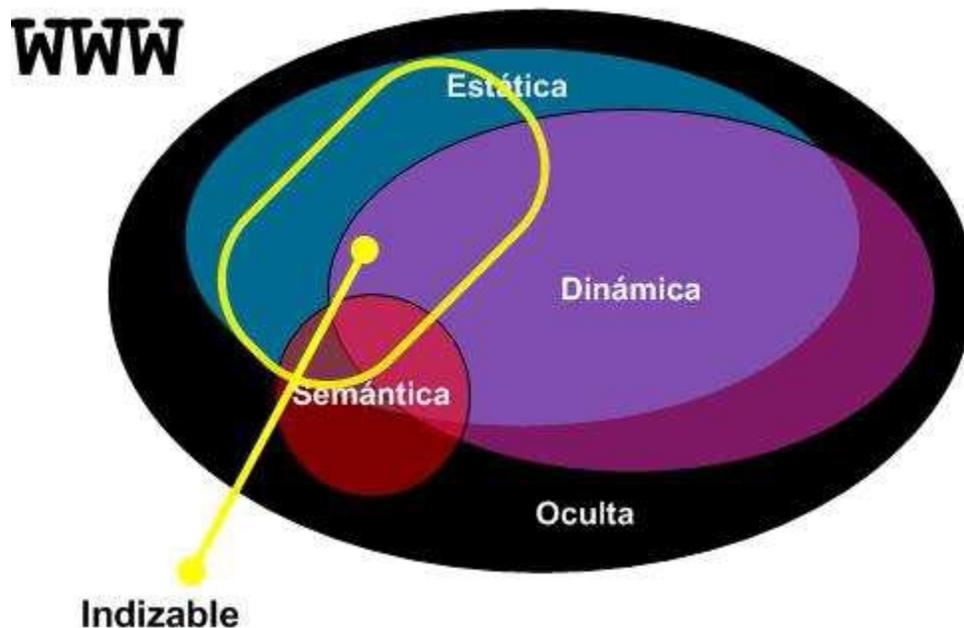
- La información no es únicamente textual.

Las soluciones propuestas a los problemas de la búsqueda y recuperación de información disponibles en la actualidad aunque son muy potentes en tratamiento de texto, en otros formatos pierden bastante efectividad.

1.2 PARTES DEL WORLD WIDE WEB

La estructura del World Wide Web puede estudiarse desde infinidad de perspectivas. Por ejemplo, se puede caracterizar por su contenido, estructura, usuarios que la frecuentan, etc. Como primera aproximación, se logra clasificar el contenido de la World Wide Web por la naturaleza de su contenido, tal y como aparece en la siguiente figura:

Figura 1. Partes de la World Wide Web



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

Por tanto, la Web, desde este punto de vista, estaría dividida en cuatro partes principales: Web Estática, Web Dinámica, Web Semántica y Web Oculta.

- **La web estática:** Corresponde al conjunto de páginas estáticas, es decir, documentos HTML que se encuentran alojados en la Web. Estos documentos cambian con muy poca frecuencia. Un ejemplo de páginas Web estáticas sería el sitio Web creado para la asignatura Seguridad Informática en el 2006 situado en: <http://usuarios.lycos.es/joh/SEGURIDAD/index.html>.

- **La web dinámica:** Contiene aquellas páginas que sólo se crean en respuesta al usuario, al realizar éste una consulta en la Web. Un ejemplo de página dinámica puede ser la página mostrada por Google al buscar el término “página dinámica”, <http://www.google.com/search?q=página+dinamica>.
- **La web semántica:** Contiene páginas que contienen metadatos los cuales dotan de significado al contenido. Este significado es utilizado por los mecanismos automáticos existentes en la Web. A modo de ejemplo, podemos citar las Imágenes que se están tratando Google Labeler, cuya URL es la siguiente: <http://images.google.com/imagelabeler/>
- **La web oculta:** Es el conjunto de páginas que no pueden ser recuperadas por los sistemas de búsqueda tradicionales.

En la **Figura 1**, también puede observarse una zona denominada la Web Indizable, la cual corresponde a la zona que realmente cubren los motores de búsqueda.

Por último, es necesario decir que la figura no se encuentra a escala dado que determinar el tamaño de la Web y sus partes es una tarea bastante compleja.

1.3 TAMAÑO DEL WORLD WIDE WEB

Cuando se realiza un estudio, en el cual es necesario un trabajar con la Web, la persona que lo realiza suele tender a realizarse la siguiente pregunta: ¿Cuál es el tamaño en páginas Web del World Wide Web? Muchos autores se aventuran a dar una respuesta:

- R. Baeza-Yates en [4] asegura que “La Web al menos tiene unas cuatro mil millones de páginas Web estáticas, un número cientos de veces mayor de dinámicas. La Web semántica correspondería a algo menos del 5% de la Web y la Web oculta sería del orden de miles de veces más grande que la pública”.
- P.M. Moreno Jiménez en [41] cita a numerosos autores a la hora de calcular el tamaño de la Web oculta, dando datos como siguientes: o Entre dos tercios y un medio del World Wide Web. o La Web invisible es 550 veces mayor que la Web pública y crece a mucha mayor velocidad. o La Web oculta es 275 veces mayor que la pública. o La Web invisible es entre 2 y 50 veces la Web pública.
- Al introducir en Google “tamaño del World Wide Web”, la página con más relevancia (<http://www.microsiervos.com/archivo/curiosidades/tamanoworld-wide-Web.html>) dice lo siguiente: “El World Wide Web tiene 155.583.825 sitios Web”.
- Maurice de Kunder, en su Web [14], realiza una estimación diaria del tamaño de la Web indizable. A día 26 de mayo de 2008 asegura que el tamaño de la Web

indizable es 47.530 millones de páginas Web.

Debido al alto dinamismo de la Web, el cual produce que ésta cambie constantemente a gran velocidad, los estudios realizados sobre su tamaño no resultan de gran utilidad. Éstos se quedan obsoletos a las pocas horas de ser realizados, o durante el transcurso de la investigación, ya que los datos recogidos han perdido su validez. Estas conclusiones hacen que exista un grupo de ingenieros que creen que el tamaño de la Web pertenece al conjunto de preguntas filosóficas imponderables.

1.4 INFOMETRÍA DE LA WEB

La mejor definición de Infometría es la propuesta en Yahoo! Answers, en la cual se define como “la ciencia que aplica métodos matemáticos a los hechos y situaciones que se producen en el campo de la información, para describir y analizar sus fenómenos, descubrir sus leyes y servir de soporte a sus decisiones.

Se aplica a los datos de cualquier rama, desde las Matemáticas hasta las Ciencias de la Información”. En algunos artículos, a este apartado se le denomina Distribuciones de leyes potenciales (Power law distributions), ya que es la principal medida que se utiliza para caracterizar al WWW.

George Kingsley Zipf, lingüista y filólogo estadounidense de la primera mitad del siglo XX, aplicó el análisis estadístico al estudio de las diferentes lenguas. Tras numerosos estudios, enunció la conocida **Ley de Zipf** [62] que afirma que un pequeño número de palabras son utilizadas con mucha frecuencia, mientras que frecuentemente ocurre que un gran número de palabras son poco empleadas.

Esta afirmación, expresada matemáticamente, quedaría de la siguiente forma:

$$P_n \sim \frac{1}{n^a}$$

Donde P_n representa la frecuencia de una palabra ordenada, ordenadas de mayor a menor por frecuencia de uso n -ésima y, a es una constante que ronda en las cercanías del 1. Esto significa que el segundo elemento en la lista se repetirá aproximadamente en una frecuencia de $\frac{1}{2}$ de la del primero, y el tercer elemento con una frecuencia de $\frac{1}{3}$ y así sucesivamente. Aunque esta ley es empírica es bastante precisa. Sus trabajos son utilizados para estudiar las propiedades estadísticas de grandes conjuntos de datos, y se han aplicado también a la Web. Se estima que aplicando la Ley de Zipf al contenido de las páginas Web, la constante a toma un valor medio que ronda entre 1.8. [4]

Pero no solamente el contenido de las páginas Web sigue un comportamiento acorde con la Ley de Zipf, existen muchas medidas en la World Wide Web que siguen dicha ley, como por ejemplo las siguientes:

- **El tamaño de las páginas.** Para páginas Web de tamaño pequeño no se ajusta muy bien por la presión de la mirada crítica de los internautas, pero a medida que crece el tamaño la ley se va haciendo visible.
- **Número de hiperenlaces entrantes y salientes.** La mayoría de las páginas están apuntadas únicamente por un hiperenlace.
- **Número de componentes conexos en la Web.** Existen pocos conjuntos de páginas conexos, es decir, que existe un camino formado por hiperenlaces que conectan un página con cualquier otra dentro del conjunto.
- **Fecha de actualización.** Existen más páginas nuevas o actualizadas que antiguas.
- **Uso de palabras clave a la hora de realizar búsquedas.** Se realizan gran cantidad de preguntas simples y muy pocas complejas.
- **Número de accesos a una Web por parte del usuario.** Un usuario suele acceder con frecuencia al mismo conjunto de páginas.

1.5 ESTRUCTURA DE LA WEB

La estructura de la Web es compleja y evoluciona en el tiempo. Sin embargo, se han realizado numerosos estudios sobre su estructura, teniendo en cuenta o no estimaciones sobre su tamaño. La mayoría de dichos estudios consideran la Web como un grafo y utilizan algoritmos definidos en teoría de grafos para intentar solucionar los problemas derivados de la búsqueda, recuperación de información y minería de datos presentes en el WWW. En [10] se justifica su necesidad dando las siguientes áreas de aplicación:

- Diseño de Web crawlers, se denominan Agentes Inteligentes de Búsqueda.
- Entender la sociología en la creación de contenido Web.
- Utilizar la información implícita en el uso de hiperenlaces en la Web, por ejemplo algoritmo de Page Rank.
- Realizar predicciones en la evolución de las estructuras Web y desarrollar algoritmos adecuados para encontrar y organizar dichas estructuras.

La ley de Zipf también se presenta en la estructura del World Wide Web, la fracción de páginas que son apuntadas por m páginas es proporcional a $1/m^{2.1}$ [10], mientras que la fracción de páginas que contienen n enlaces es proporcional a $1/n^{2.7}$ [4]. En lenguaje natural esto significa que pocas páginas son muy

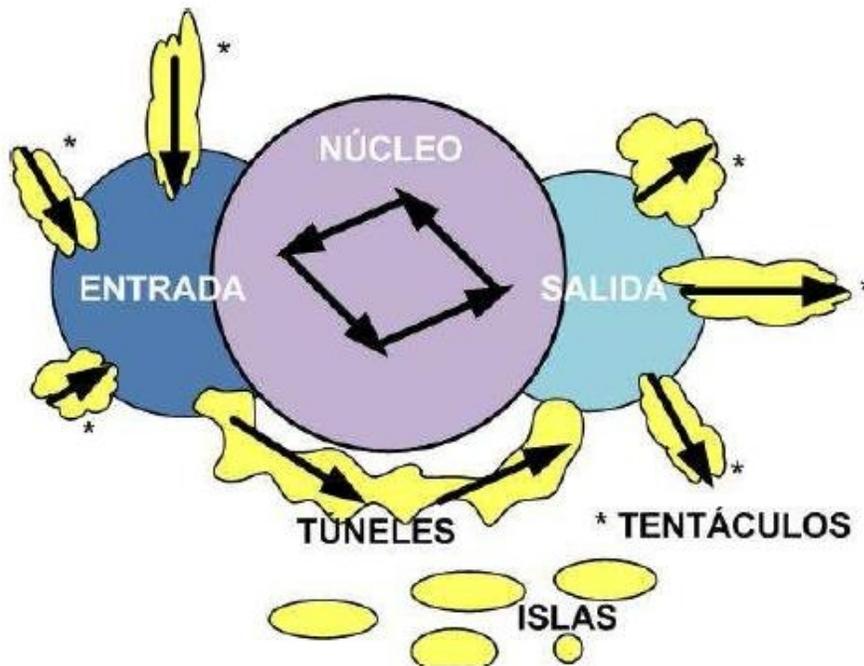
apuntadas y contienen muchos enlaces y que muchas páginas son poco apuntadas y contienen pocos enlaces.

Diversos estudios modelan el World Wide Web como un grafo dirigido, donde los nodos son las páginas Web existentes y las aristas dirigidas de un nodo a otro representan el enlace de una página a otra. Matemáticamente, se expresaría de la siguiente manera:

Sea el World Wide Web un grafo dirigido $G = (V, E)$, donde V es el conjunto de nodos (el conjunto de páginas) y E el conjunto de aristas dirigidas en el grafo (los hiperenlaces), siendo el número total de páginas en la Web $n = |V|$.

En [10], unos de los trabajos pioneros en este campo, tras diversos experimentos, descubrieron que la Web está formada por varios tipos de subgrafos interconectados entre sí. De acuerdo con su trabajo, si fuese posible reducir el zoom del grafo que representa la Web hasta que sea perceptible todo el conjunto, nos encontraríamos un grafo parecido a la figura siguiente:

Figura 2. Grafo dirigido que representa la estructura de la Web.



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

Como se puede observar en la **Figura 2**, la Web está formada por seis tipos

principales de subgrafos: Entrada, Núcleo, Salida, Túneles, Islas y Tentáculos.

- **El Núcleo:** El Núcleo (denominado SCC en [10]), es el conjunto de páginas más estable de todos. Este subgrafo corresponde al conjunto de páginas conexas, es decir, que existe un camino desde cualquier página del conjunto a la otra del conjunto.
- **En el subgrafo Entrada (IN):** Existe un camino desde cualquier página contenida al subgrafo Núcleo, pero no existen caminos entre dichas páginas contenidas. Generalmente está compuesto por las páginas más nuevas que aún no son demasiado conocidas. Estas páginas pasarán al centro si verdaderamente tienen éxito.
- **En el conjunto Salida (OUT):** Existe un camino para llegar hasta sus páginas desde el Núcleo, pero no entre ellas. Suele estar compuesto por las páginas más antiguas, mal hechas, mal mantenidas o simplemente que por su naturaleza no tienen enlaces.
- **Túnel (Tubes):** Es el conjunto de páginas que conectan entrada con salida sin pasar por el núcleo.
- **Las Islas (Disconnected components):** Son conjuntos de páginas aisladas de la Web, no son enlazadas por nadie y no enlazan a nadie, únicamente tienen enlaces con las páginas dentro de su isla. Son los conjuntos menos estables y el menos conocidos.
- **Los Tentáculos (Tendrils):** Son las variaciones debidas a la complejidad de la estructura del World Wide Web, es decir, el conjunto de páginas Web que no se pueden clasificar en los subgrafos anteriores.

2. MINERÍA WEB

La Minería Web o Web Mining consiste en aplicar las técnicas de minería de datos para descubrir y extraer automáticamente información de los documentos y servicios de la Web. [16] En particular, la creación, extracción y mantenimiento de los modelos de usuarios en Sistemas de Recomendación en Internet mejora la experiencia del usuario en relación con la información que le es relevante reduciendo el problema conocido como sobrecarga de la información. [22]

Sin embargo, las técnicas de minería de datos no son fácilmente aplicables a datos de la Web debido a problemas relacionados tanto con la tecnología subyacente como con la ausencia de estándares en el diseño e implementación de páginas Web. La información contenida en archivos log y otras fuentes de información debe ser procesada previamente a la obtención de los modelos. [36]

Etzioni señala que la minería Web está compuesta por tres tareas: descubrimiento de las fuentes, selección y preprocesamiento de la información y descubrimiento de patrones generales desde los sitios Web, en esta última es donde se realiza el proceso de minería en sí. En investigaciones posteriores se consideró una cuarta etapa dirigida al análisis para la validación y/o interpretación de los patrones minados. [48]

La minería Web puede dividirse en tres áreas o categorías principales: minería de contenido, minería de estructura y minería de uso, en función de los datos utilizados para inducir los modelos.

2.1 TIPOS DE MINERÍA WEB

En el caso de la minería Web los datos pueden ser obtenidos desde el lado del servidor, del cliente, de los servidores proxy o de la base de datos corporativa de la entidad a la cual pertenece el sitio. Desde este punto de vista, los datos encontrados en un sitio Web en particular, pueden ser clasificados en tres tipos predominantes, Como se indica en la **Figura 3**.

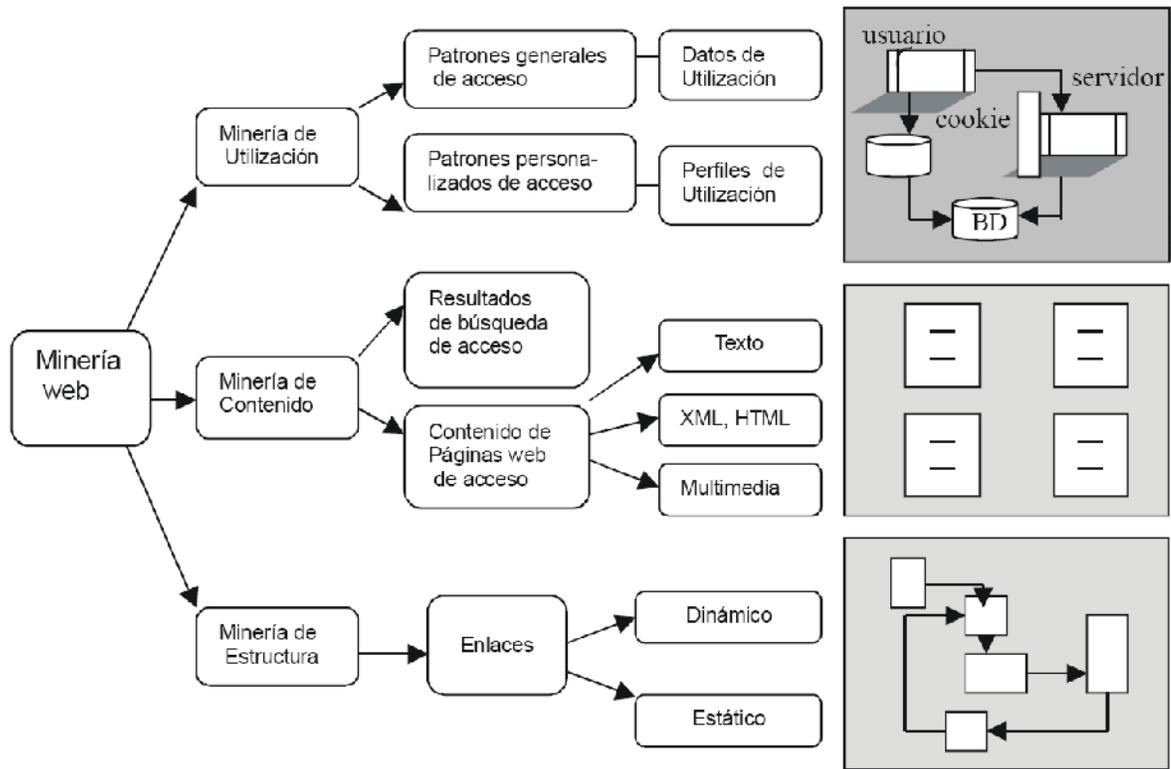
Minería del Contenido de la Web: Son los datos reales que se entregan a los usuarios. Es decir, los datos que almacenan los sitios Web, los cuales consisten generalmente en textos e imágenes u otros medios. Este tipo de dato es el más importante y difícil de procesar, por ser multimedia.

Minería de Uso de la Web: Son aquellos datos que describen el uso al cual se ve sometido un sitio, registrado en los logs de acceso de los servidores Web.

Minería de la Estructura de la Web: Son los datos que describen la organización del contenido en el interior de un sitio. Esto incluye, la organización dentro de una

página, la distribución de los enlaces tanto internos al sitio como externos, y la jerarquía de todo el sitio.

Figura 3. Mapa conceptual de la Minería Web.



Fuente: Baeza Yates, Ricardo y Poblete, B. Una herramienta de minería de consultas para el diseño de contenido y la estructura de un sitio Web. Granada: Thomson, 2005. págs. 39-48. ISBN: 84-9732-449-8. [6]

2.1.1 Minería del contenido de la web: Web Content Mining (Minería de Contenido Web) se centra en el contenido, por lo que se pueden obtener datos acerca de la forma de escribir que sea más atractiva para el usuario, de si la catalogación que se usa sirve para mejorar la relevancia del sitio, si los temas que se tratan interesan o no.

Esta área de la minería Web tiene dos vertientes: recuperación de la información y base de datos.

Como se conoce, los sitios de la Web están compuestos de colecciones de documentos de hipertexto. La recuperación de la información se realiza a través de la exploración semántica de los documentos, mediante dos enfoques: la minería de textos y el análisis semántico de los textos.

Considerando que los sitios de la Web también son colecciones de documentos semiestructurados, se pueden descubrir y extraer esquemas para formularios que capturen información semántica relevante de fuentes de datos heterogéneas. Los enfoques están basados en lenguajes de consultas para Web (XML, WebSQL, WebML), base de datos múltiples y descubrimiento de jerarquías.

2.1.2 Minería de uso de la web: Web Usage Mining (Minería de uso Web) se refiere a la técnicas de data mining para descubrir pautas de conducta a la hora de utilizar la Web por parte de los usuarios.

Este tipo de minería utiliza los archivos logs convenientemente preprocesados para desarrollar la minería de datos sobre ellos y así descubrir reglas que permitan identificar las experiencias de los usuarios cuando navegan por un sitio Web.

En definitiva, se trata de seguir una serie de patrones sobre:

- El acceso que utilizan los clientes cuando consultan el sitio Web de una empresa
- Los usuarios que interrogan a una aplicación que precede a una base de datos
- Los individuos que navegan por páginas determinadas.

2.1.3 Minería de la estructura de la web: Web Structure Mining (Minería de Estructura Web) se refiere al grado de dificultad que tienen los usuarios para encontrar la información, si la estructura del sitio es simple o muy profunda, si los elementos están colocados en los lugares adecuados dentro de la página, si la navegación es comprensible, cuáles son las secciones menos visitadas y su relación con el lugar que ocupan en la página central.

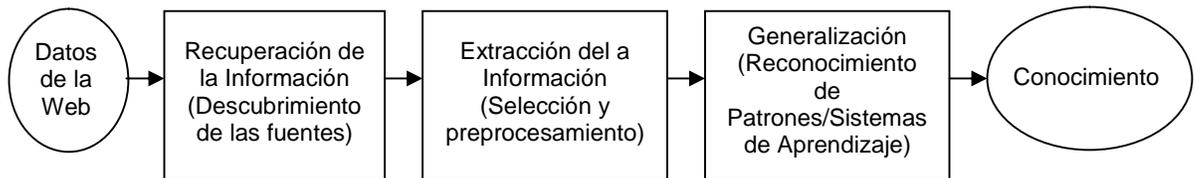
La minería de estructura Web revela más información que simplemente la información contenida en los documentos. Por ejemplo, enlaces o eslabones que apuntan a un documento indican su nivel de popularidad, mientras que los enlaces o eslabones que salen de un documento indican la riqueza o quizás la variedad de temas que se abarcan en el documento. Esto fue resaltado por Kleinberg en el algoritmo HITS, (del inglés Hypertext Induced Topic Selection), es un algoritmo diseñado para valorar y de paso clasificar la importancia de una página Web. [20]

Este tipo de minería Web está muy relacionada con la temática de los grafos tanto en la parte visual como los diferentes procesos que se llevan a cabo en cada uno de ellos, la teoría de grafos y la visualización de los mismos son campos muy amplios que en muchos estudios e investigaciones se trabajan totalmente independientes.

2.2 ETAPAS DE LA MINERÍA WEB

A continuación se describen las cuatro tareas que forman parte de la minería Web, como se visualizan en la **Figura 4**.

Figura 4. Fases de la Minería Web.



Fuente: Pal, S.K., Talwar, V. y Mitra, P. Web Mining in Soft Computing Framework: Relevance, state of the art and future directions. *IEEE Transactions on Neural Networks*. 2002, Vol. 13, págs. 1163-1177. [48]

2.2.1 Descubrimiento de las fuentes: El descubrimiento de las fuentes o la recuperación de la información (RI) consisten en la recuperación automática de los documentos relevantes, mientras que se asegura al mismo tiempo, tanto como sea posible, que los no relevantes no sean considerados. El proceso de RI principalmente incluye la representación del documento, uso de índices y búsqueda de documentos.

El gran número de páginas en la Web, su dinamismo, y la actualización frecuente hace a las técnicas de uso de índices aparentemente imposible. En la actualidad, existen cuatro enfoques para poner índices a los documentos en la Web: índice humano o manual, índice automático, índice inteligente o basado en agentes e índices basados en metadatos.

Los futuros sistemas de descubrimiento de las fuentes harán uso de la tecnología de categorización de texto automática para clasificar los documentos de la Web en categorías. Esta tecnología podría facilitar la construcción automática de directorios de la Web como Yahoo, que localiza documentos y los presenta en categorías. Alternativamente, esta tecnología podría ser usada para filtrar los resultados de consultas a los índices de búsqueda.

Los estudios en la recuperación de información incluyen la modelación, desarrollo de interfaces con el usuario, visualización de los datos, y filtros. [7]

2.2.2 Selección/extracción y preprocesamiento: Una vez los documentos se han recuperado, el desafío es extraer conocimiento automáticamente y otras informaciones requeridas sin la interacción humana. La extracción de la información (EI) es la tarea de identificar fragmentos específicos de un documento que constituyen su contenido semántico fundamental. Hasta ahora, los principales métodos de extracción de información involucran wrappers de escritura (codificación de la escritura) que asigna los documentos a algún modelo de datos.

Los wrappers actúan como interfaces de cada fuente de datos, proporcionando una semiestructura a aquellas fuentes no estructuradas o bien mapean la estructura de datos original en la búsqueda de un patrón común [1]. Los métodos wrapper, aunque son eficaces para eliminar atributos irrelevantes y redundantes, son muy lentos, variando en cada ejecución el número de atributos, siguiendo algún criterio de búsqueda y de parada. [30]

La extracción de información tiene como objetivo extraer el nuevo conocimiento de los documentos recuperados en la estructura y representación del documento mediante la conversión en mayúsculas, teniendo en cuenta que los expertos de la recuperación de información consideran que el texto del documento es una bolsa de palabras y no prestan atención a la estructura del documento. La escalabilidad es el mayor desafío más para los expertos de extracción de información; no es factible construir sistemas de extracción de información que sean escalables al tamaño y dinamismo de la Web. Por tanto, la mayoría de los sistemas de Extracción de información extraen información de sitios específicos y se enfocan en áreas definidas.

Algunos de los agentes inteligentes de la Web han sido desarrollados para buscar información relevante usando características de un dominio particular (y posiblemente un perfil de usuario) para organizar e interpretar la información descubierta.

Es necesario un sistema de procesamiento robusto para extraer cualquier tipo de conocimiento, incluso a partir bases de datos medianas. Cuando un usuario solicita una página Web se accede a una variedad de archivos como imágenes, sonido, video, ejecutables y HTML. Como resultado, el log del servidor contiene muchas entradas que son redundantes o irrelevantes para las tareas de minería. Esto significa, que estas entradas deben eliminarse a través del preprocesamiento. Una de las técnicas de preprocesamiento usadas para la extracción de información es el índice semántico latente, del inglés latent semantic index, que busca transformar los vectores del documento original a un espacio dimensional más bajo mediante el análisis de la estructura correlacional en esa colección del documento de modo que documentos similares que no comparten los mismos términos se colocan en la misma categoría (tema).

2.2.3 Generalización: Una vez que se ha automatizado el descubrimiento y la extracción de la información procedente de los sitios Web, el siguiente paso es tratar de generalizar a partir de la experiencia acumulada. Para ello, la Minería Web ha adaptado técnicas de minería de datos (reglas de asociación, clustering, entre otras), de la recolección de información y ha desarrollado algunas técnicas propias, como por ejemplo el análisis de caminos, que ha sido usado para extraer secuencias de patrones de navegación desde archivos log. [25]

Actualmente, la mayoría de los sistemas de aprendizaje desplegados en la Web se dedican más a aprender acerca de los intereses de sus usuarios, en lugar de aprender sobre el propio contenido y organización de la Web. El problema del etiquetado es un obstáculo importante cuando se aprende sobre la Web, ya que los datos son abundantes pero no están etiquetados. [16]

Algunas técnicas, como las pruebas de incertidumbre, reducen la cantidad de datos no etiquetados necesarios, pero no eliminan el problema del etiquetado. Un enfoque para resolver este problema se basa en el hecho de que la Web es mucho más que una colección enlazada de documentos, es un medio interactivo. Por ejemplo, Ahoy [17] toma como entrada el nombre de una persona y su afiliación y se intenta localizar el homepage de esa persona, de esta forma se les pregunta a los usuarios acerca de las páginas recuperadas, para etiquetar sus respuestas como correctas o incorrectas. [48]

Las técnicas de clustering no requieren las entradas etiquetadas y se han aplicado con éxito a las grandes colecciones de documentos. De hecho, la Web ofrece un terreno fértil para investigaciones de clustering y clasificación de documentos.

Las reglas de asociación también son una parte íntegra de esta fase. Básicamente, las reglas de asociación son expresiones del tipo $X \Rightarrow Y$, donde X e Y son conjuntos de elementos.

$X \Rightarrow Y$ expresa que siempre que una transacción T contenga a X entonces probablemente T también contiene a Y . La probabilidad o confianza de la regla se define como el porcentaje de transacciones que contienen a Y y además a X en relación con el total de transacciones que contienen a X .

2.2.4 Análisis: El análisis es un problema de manipulación de datos que requiere que existan datos suficientes disponibles para que la información potencialmente útil se pueda extraer y analizar. Los humanos juegan un papel importante en el proceso de descubrimiento del conocimiento en la Web, considerando que la Web es un medio interactivo. Esto es especialmente importante para la validación y/o interpretación de los patrones minados que tienen lugar en esta fase. Una vez que los patrones se han descubierto, los analistas necesitan herramientas apropiadas para entender, visualizar e interpretar estos patrones. Algunos usan el Procesamiento Analítico en Línea, del inglés Online Analytical Processing (OLAP) con el propósito de simplificar el análisis de las estadísticas a partir de los logs de los servidores, y otros mecanismos SQL, como el sistema WEBMINER [40] que propone un lenguaje de consultas, similar a SQL, que posee un mecanismo de consultas para preguntar acerca del conocimiento descubierto (en forma de reglas de la asociación y modelos secuenciales).

2.3 MINERÍA DE DATOS

La Minería de Datos (DM) por las siglas en inglés Data Mining es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos [26]. Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios la toma de decisiones.

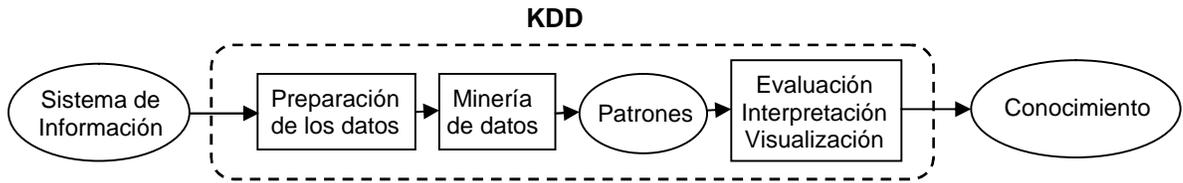
Existen términos que se utilizan frecuentemente como sinónimos de la minería de datos. Uno de ellos se conoce como "análisis (inteligente) de datos" [8], que suele hacer un mayor hincapié en las técnicas de análisis estadístico. Otro término muy utilizado, y el más relacionado con la minería de datos, es la extracción o "descubrimiento de conocimiento en bases de datos" (Knowledge Discovery in Databases o KDD, según sus siglas en inglés). [25]

Aunque algunos autores usan los términos Minería de Datos y KDD indistintamente, como sinónimos, existen claras diferencias entre los dos. Así la mayoría de los autores coinciden en referirse al KDD como un proceso que consta de un conjunto de fases, una de las cuales es la minería de datos. De acuerdo con esto, el proceso de minería de datos consiste únicamente en la aplicación de un algoritmo para extraer patrones de datos y se llamará KDD al proceso completo que incluye pre-procesamiento, minería y post-procesamiento de los datos.

El KDD según [18] es la extracción automatizada de conocimiento o patrones interesantes, no triviales, implícitos, previamente desconocidos, potencialmente útiles y predictivos de la información de grandes Bases de Datos.

La **Figura 5**, muestra las fases del proceso de KDD, una de las cuales es la Minería de Datos

Figura 5. Fases del proceso KDD

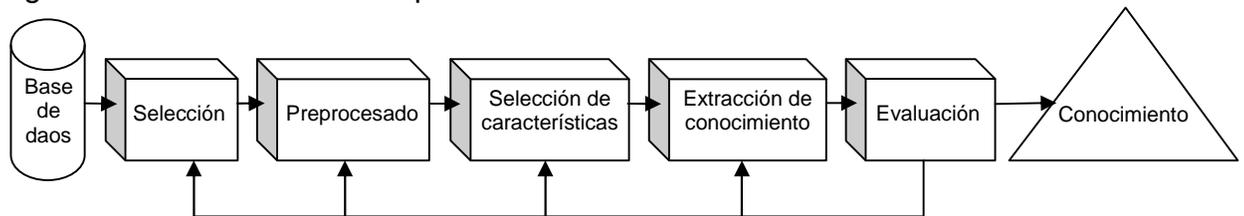


Fuente: González Díaz, Ernesto; Pérez Hernández, Zady y Espinosa Conde, Ivet. monografias.com. *Técnicas de minería de datos*. [En línea] [Citado el: 8 de Mayo de 2009.] <http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml>. [21]

Las investigaciones en temas de KDD incluyen análisis estadístico, técnicas de representación del conocimiento y visualización de datos, entre otras. Algunas de las tareas más frecuentes en procesos de KDD son la clasificación y clustering, el reconocimiento de patrones, las predicciones y la detección de dependencias o relaciones entre los datos.

2.3.1 Fases de la minería de datos: Los pasos a seguir para la realización de un proyecto de minería de datos son siempre los mismos, independientemente de la técnica específica de extracción de conocimiento usada.

Figura 6. Fases dentro de un proceso de Minería de Datos



Fuente: González Díaz, Ernesto; Pérez Hernández, Zady y Espinosa Conde, Ivet. monografias.com. *Técnicas de minería de datos*. [En línea] [Citado el: 8 de Mayo de 2009.] <http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml>. [21]

El proceso de minería de datos pasa por las siguientes fases:

- **Comprensión del problema:** Se debe comprender plenamente el problema al cual se le quiere dar solución.
- **Filtrado de datos:** El formato de los datos contenidos en la fuente de datos nunca es el correcto, y la mayoría de las veces no es posible ni siquiera utilizar algún algoritmo de minería sobre los datos iniciales sin que requieran alguna transformación. En este paso se filtran los datos con el objetivo de eliminar valores incorrectos, no válidos o desconocidos; según las necesidades del algoritmo a utilizar. Además se obtienen muestras de los datos en busca de mayor velocidad y eficiencia de los algoritmos, o se reducen el número de valores posibles para los atributos de análisis.

➤ **Selección de variables:** Después de realizar la limpieza de los datos, en la mayoría de los casos se tiene una gran cantidad de variables o atributos. La selección de características reduce el tamaño de los datos, sin apenas sacrificar la calidad del modelo de conocimiento obtenido del proceso de minería; seleccionando las variables más influyentes en el problema.

Los métodos para la selección de los atributos que más influencia tienen en el problema son básicamente dos:

- Aquellos basados en la elección de los mejores atributos del problema.
- Aquellos que buscan variables independientes mediante test de sensibilidad, algoritmos de distancia o heurísticos.

➤ **Extracción de conocimiento:** La extracción del conocimiento es la esencia de la Minería de Datos donde mediante una técnica, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. Los modelos que se generan son expresados de diversas formas:

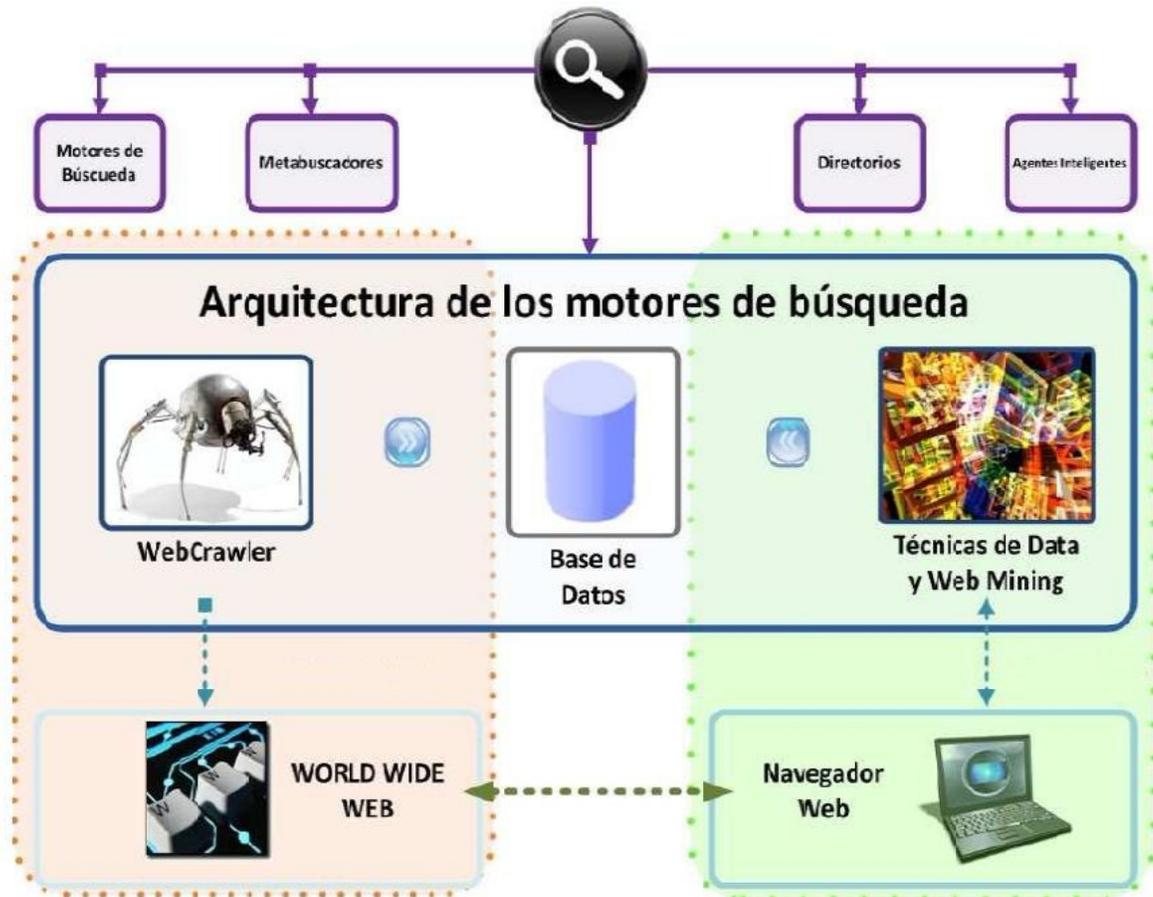
- Reglas
- Árboles
- Redes neuronales

También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre-procesado diferente de los datos.

➤ **Interpretación y evaluación:** Una vez obtenido el modelo, se procede a su validación; donde se comprueba que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos para buscar el que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

2.4 RECUPERACIÓN DE INFORMACIÓN Y BÚSQUEDA EN LA WEB

Figura 7. Técnicas de recuperación de información en la Web



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

2.4.1 La búsqueda web: “La búsqueda en la Web no necesita introducción. Debido al uso de internet y la riqueza de la información contenida en la misma, la búsqueda Web se ha convertido en el método de obtención de información. Mientras los viajes a las bibliotecas disminuyen, las búsquedas en Internet crecen.” [34]

La búsqueda en la Web se clasifica dentro del área de estudio denominada Recuperación de Información (Information Retrieval o IR), en la cual se estudian las diversas formas de ayudar al usuario a encontrar la información que éste necesita en un conjunto grande de documentos de texto. Un ejemplo de un sistema IR tradicional sería la clasificación por fichas que se realizaba antiguamente en las bibliotecas. En la Web, estos documentos se corresponden

con las páginas Web.

La recuperación de información es simplemente encontrar un conjunto de documentos que sean relevantes en relación a una consulta realizada, la cual suele tener el formato de una lista de palabras clave. A dichas palabras clave se les conoce como términos de la búsqueda. La solución propuesta por los sistemas de recuperación de información suele ser una lista de documentos ordenados por relevancia en relación a la consulta realizada.

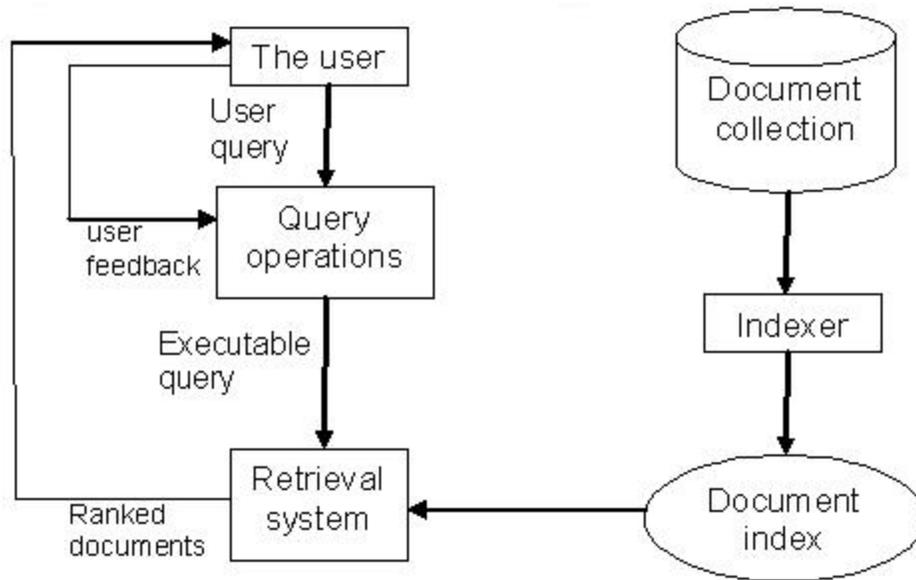
En el documento [34] se expone que, a pesar de que la búsqueda en la Web forma parte del campo de estudio del área IR, ésta presenta nuevos problemas que deben de ser solucionados. Dichos problemas son los siguientes:

- En la búsqueda Web la eficiencia trata un papel muy importante, siendo un problema secundario en los sistemas IR tradicionales. Esto se debe principalmente a que los usuarios generalmente exigen respuestas rápidas, además de que la cantidad de información en la Web es muchísimo mayor que en los sistemas tradicionales. No importa lo efectivo que sea un algoritmo, si éste no es capaz de trabajar con eficiencia, muy pocos usuarios lo utilizarán. De esto podemos deducir que el intervalo de tiempo que el usuario está dispuesto a emplear es bastante reducido, ello implica que el tiempo de respuesta de los sistemas también lo debe ser.
- Las características del conjunto de documentos en los que se van a realizar la búsqueda Web difieren de los de las búsquedas tradicionales. En las búsquedas tradicionales los documentos son independientes unos de otros, en la Web, en cambio, las páginas Web contienen hiperenlaces que unen unas con otras. Otra diferencia es que las páginas Web son documentos semiestructurados mientras que los documentos tradicionales son un conjunto de párrafos relacionados sobre un tema concreto. Las páginas Web contienen diversos campos, siendo unos más importantes que otros, por ejemplo, el contenido de un artículo es más importante que la publicidad. Por tanto, los sistemas de búsqueda en la Web deberían diferenciar estos campos.
- El término spamming, es uno de los mayores problemas en la Web. Se considera spam a todo documento que satisface la consulta realizada por el usuario pero que no satisface la necesidad de información del mismo. En la IR tradicional no existe este problema.

En la búsqueda Web, éste es un verdadero problema debido a que en ocasiones aparece spam como documentos relevantes, lo que produce una disminución de la calidad de la herramienta de búsqueda.

2.4.2 Arquitectura de los sistemas de recuperación de información: La Recuperación de Información estudia conceptos de adquisición, organización, almacenamiento, recuperación y distribución de información [34]. En la Figura 8, se puede observar la arquitectura que tiene un sistema de recuperación de información:

Figura 8. Arquitectura de un sistema de recuperación de información.



Fuente: Liu, B. *Information Retrieval and Web Search*. En Liu B., *Web Data Mining*. 2006. págs. 183-236. [34]

En la **Figura 8**, un usuario (the user) con una necesidad de información formula una consulta (user query) a un sistema de recuperación de información a través de un módulo denominado operaciones de consulta (query operations). Este módulo puede traspasar la consulta tal y como está o puede enriquecerla.

El sistema de recuperación (Retrieval System) utiliza el índice de documentos (document index) para recuperar los documentos que satisfacen la consulta del usuario, además éste computa la relevancia de los mismos y los ordena por relevancia. Tras ello, los documentos ordenados (ranked documents) son remitidos al usuario. La colección de documentos (document collection) también se suele denominar base de datos textual, en la cual los documentos están indizados o indexados por el indizador (indexer).

La consulta realizada por el usuario puede corresponder con alguna de las siguientes formas:

- **Consulta de palabras clave:** Ésta consiste en una lista de términos relevantes.
- **Consultas booleanas:** Conjunto de palabras clave acompañadas de operadores booleanos.
- **Consultas de frases:** Una oración que debe aparecer en el documento tal y como se especifica en la consulta.
- **Consultas de documentos completos:** Se corresponde con la búsqueda de documentos similares a uno propuesto.
- **Consultas en lenguaje natural:** El usuario expresa su consulta en el lenguaje que utiliza para comunicarse con otros seres humanos. Actualmente esta tecnología se encuentra bastante limitada y en proceso de investigación.

2.4.3 Recursos críticos en las búsquedas: En la actualidad, y con mayor énfasis en el mundo que rodea a las Ciencias de la Computación, se suele utilizar el término recurso crítico. Recurso se define como el medio o elemento que se emplea para conseguir un fin pretendido, siendo el recurso crítico el medio o elemento de un conjunto de recursos que más escasea.

Se acostumbra tratar como recurso crítico al dinero, al personal o a la tecnología. El tiempo es el recurso crítico por excelencia para el ser humano por sus características particulares. El tiempo no puede ser acumulado, ni almacenado, ni reemplazado y se está obligado a consumirlo a una tasa fija, se quiera o no. En muchas ocasiones el tiempo es quién decide si una tarea es factible o no, y más en este contexto debido al colosal tamaño de la Web.

2.4.4 Motores de búsqueda o buscadores: Un buscador (Search Engine) es un sistema informático que indexa por palabras clave diversos archivos almacenados en servidores Web. Se les suele denominar también Motores de Búsqueda o Buscadores jerárquicos. Como ejemplos de buscadores tenemos Google, MSN Live Search o AltaVista.

Los buscadores rastrean e indexan de forma automática páginas Web, así como todos los documentos referenciados en ellas. Esto lo realizan mediante el uso de arañas (Web Crawlers). También son añadidas a los servidores de los buscadores las páginas, las cuales son informadas directamente por el autor. Su interfaz suele ser un cuadro de texto donde realizar búsquedas de forma sencilla. Tras introducir las palabras clave en dicho cuadro, se nos presenta una página Web en la cual aparecen diversos resultados ordenados por relevancia.

Los buscadores presentan resultados realizados directamente sobre sus servidores y no sobre la Web, y dado que las arañas realizan el rastreo con

periodicidad variable y no con mucha frecuencia, puede existir diferencia entre lo encontrado y lo que verdaderamente se encuentre en la Web. Se creó que la frecuencia de rastreo ha aumentado en los últimos años y que ésta se hace de forma diferenciada para algunos recursos. Por ejemplo, las páginas de noticias se rastrearían con mayor frecuencia que las páginas personales.

Las principales diferencias entre los buscadores presentes en la Web son la forma de calcular la relevancia de la información obtenida, Page Rank y demás, y el poco solapamiento que existe entre sus resultados, es decir, que cada buscador tiende a indizar una parte diferente de la Web. Según [31] existe una diferencia significativa al realizar una misma búsqueda utilizando buscadores distintos.

Los puntos fuertes de usar Motores de búsqueda son que éstos realizan búsquedas exhaustivas y que posibilitan encontrar recursos muy específicos acotando los términos de la búsqueda según [38]. En dicho documento, también se expone que la principal desventaja es el ruido producido, es decir, los buscadores presentan un gran número de resultados y en bastantes ocasiones éstos están duplicados. Según [41] la principal limitación sería que, en la actualidad, los buscadores sólo pueden recuperar información que se encuentra disponible en la Web, no aquella que se ofrece a través de la Web.

A pesar de las limitaciones, en los últimos años han mejorado considerablemente en la precisión en las búsquedas y en la forma de ofrecer los resultados. Además, se pueden realizar búsquedas de diversos archivos, además de páginas Web, como documentos de MS Word, documentos Pdf, etc.

2.4.5 Directorios o índices: Un directorio (directory) es un sistema informático que indexa por temas o categorías jerarquizados los archivos almacenados en servidores Web.

También se les suele denominar Índices o Índices temáticos. Como principal ejemplo está Yahoo!, aunque en la actualidad también usa búsqueda por palabras clave. Los directorios se organizan manualmente a partir del registro de páginas por parte de sus editores e implican una selección y un proceso de categorización jerarquizada.

Según [41] existen dos tipos de directorios:

- **Directorios de modelo cerrado (Closed model directories):** Los editores de estos directorios son empleados de la empresa que los realiza, esto conlleva a una mayor calidad de la información y a un mayor costo.
- **Directorios de modelo abierto (Open model directories):** Los editores de los directorios son voluntarios, lo que reduce el coste y la calidad de los mismos.

En [38] se expone que el punto fuerte de usar Directorios, es que existe poco riesgo de duplicación en los resultados al tratarse de recursos seleccionados. El punto débil sería que son menos exhaustivos que los Buscadores.

Por último, es importante comentar que existen directorios contenedores de buscadores, directorios con anotaciones de los usuarios, directorios con BD, es decir, que un directorio puede indizar cualquier tipo de servicio o formato. Algunos autores, véase el artículo [41], tratan a las guías como una herramienta más de búsqueda.

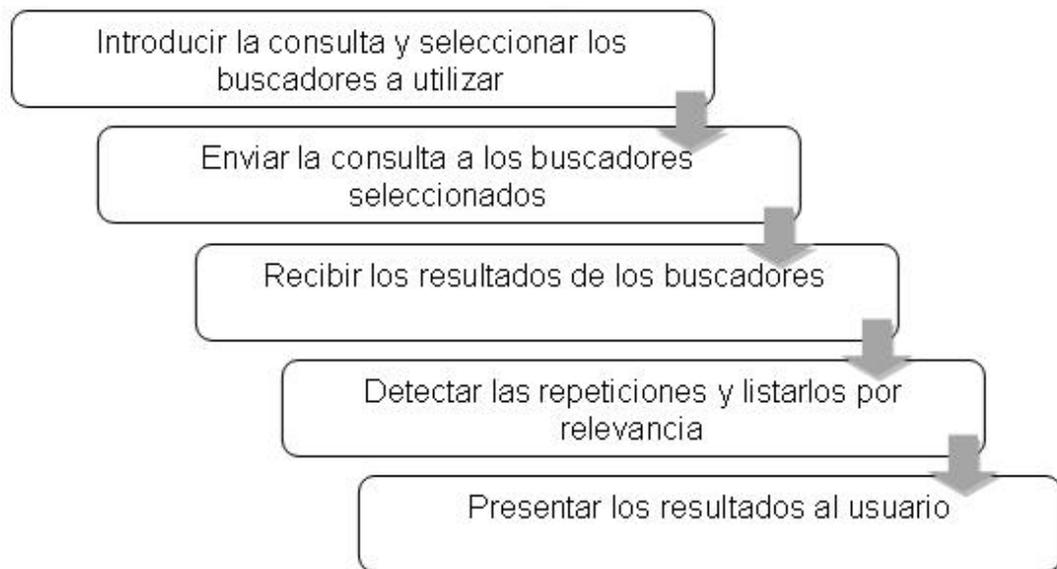
Creadas por especialistas y entidades académicas, las guías son un contenedor de recursos Web seleccionados por área de especialidad, los cuales pueden no ser recuperables por los buscadores. Además éstas pueden incluir algún mecanismo de búsqueda, tecnología basada en buscadores, en sus páginas o en la Web en general.

Las guías son reconocidas por su alto grado de calidad, ya que en la mayoría de los casos comprometen el prestigio de sus autores o las instituciones a las que pertenezcan. Por ello, la selección de recursos suele ser muy cuidadosa y con una constante actualización.

En ocasiones las entidades encargadas se asocian formando circuitos (Web rings) para la elaboración cooperativa de estas guías.

2.4.6 Los metabuscadores: Los metabuscadores (Metasearch Engines) son aplicaciones que realizan búsquedas simultáneas en diversos buscadores y/o directorios presentando a posteriori un resumen de los resultados obtenidos y eliminando los duplicados. Un ejemplo de metabuscadores sería Metacrawler[®]. En la Figura 9, podemos observar cómo se realiza una búsqueda en un metabuscador.

Figura 9. Acciones realizadas por un metabuscador al procesar una búsqueda.



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

Los metabuscadores son utilizados principalmente para búsquedas muy complejas. Dado que cada buscador se dedica casi en totalidad a una parte de la zona indizable, al utilizar metabuscadores se multiplica el área de búsqueda y por tanto, la probabilidad de éxito. Como limitaciones decir que los metabuscadores no suelen presentar opciones de búsqueda avanzada, además éstos presentan las limitaciones de cada buscador utilizado. Otra limitación sería que al depender de otros sistemas, en ocasiones fallen.

Existe relativa controversia entre los diferentes expertos del tema sobre el uso de los metabuscadores, unos recomiendan utilizarlos fervientemente ya que para ellos son el siguiente eslabón de la cadena alimenticia de la búsqueda en la Web mientras que otros advierten constantemente de sus limitaciones. [41]

2.4.7 Agentes inteligentes: Tal y como comenta [31] un agente inteligente es un programa que basándose en su propio conocimiento, realiza un conjunto de operaciones para satisfacer las necesidades de un usuario o de otro programa, bien por iniciativa propia o porque alguno de éstos se lo requiere.

Youngblood [30] comenta que el corazón de todo motor de búsqueda se encuentra en el agente de búsqueda, cuya principal función es recorrer la Web en busca de información que posteriormente será guardada en la base de datos del sistema. Por ello, antes de diseñar un agente de búsqueda, nos deberíamos preguntar ¿Qué es un Agente Inteligente? Aunque existe controversia en el ámbito científico

sobre dicha definición debido su aplicación en múltiples disciplinas, la definición propuesta por [59] es la más aceptada: “Un agente inteligente es un sistema informático, hardware o software, caracterizado por ser autónomo, social, reactivo y proactivo.”

Otros autores, como por ejemplo Moreno Jiménez, denominan a este conjunto “Software especializado” y lo utilizan para clasificar todas aquellas herramientas que no se corresponden a ninguno de los tipos propuestos anteriormente y que generalmente utilizan técnicas basadas en la inteligencia artificial.

Este es el grupo más heterogéneo de todos, ya que existen multitud de diversas aplicaciones; existen agentes que filtran los resultados de los buscadores, otros que buscan y recuperan información, otros que “espían” Webs avisando al usuario cuando se produce un cambio en éstas, etc.

En [19] se caracteriza a este grupo por usar búsqueda dinámica en vez de la búsqueda estática utilizada por los sistemas de búsqueda clásicos. La búsqueda dinámica obtiene los datos de la Web en el momento en el que se está realizando la búsqueda, estos datos no son obtenidos de ningún índice pre construido.

2.5 WEB CRAWLER

Un Crawler o rastreador Web, es un programa de computadora que navega por la World Wide Web, es una herramienta automatizada en forma metódica y de manera ordenada. Otros términos para los rastreadores Web son: hormigas, indizadores automáticos, robots, arañas Web, orugas o robots Web. [29]

Este proceso es llamado Web gatear o rastreo. Muchos sitios, en particular, los motores de búsqueda, usan crawlers con la finalidad de mirar la estructura de la Web y encontrar los enlaces rotos. Su uso principal es crear una copia de todas las páginas visitadas para su posterior procesamiento por un motor de búsqueda que indexe las páginas descargadas para proporcionar búsquedas rápidas.

Los crawlers también se pueden utilizar para automatizar las tareas de mantenimiento en un sitio Web, como la comprobación de enlaces o validación de código HTML. Además, los rastreadores pueden ser usados para obtener información específica de las páginas Web, tales como la recolección de direcciones de correo electrónico.

Un rastreador Web es un tipo de robot, o un agente de software. En general, comienza con un URL a visitar, llamada semilla. A medida que el rastreador visita esta URL, identifica todos los hipervínculos en la página y los agrega a la lista de URLs a visitar, llamada la frontera de rastreo. Los URL de la frontera son visitados de forma recursiva de acuerdo con un conjunto de políticas.

2.5.1 Políticas de rastreo: Hay características importantes de la Web que hacen muy difícil el rastreo:

- Su gran volumen.
- Su rápido ritmo de cambio.
- La generación de páginas dinámicas.

El gran volumen implica que el rastreador sólo puede descargar una fracción de las páginas Web en un tiempo determinado, por lo que debe dar prioridad a sus descargas. La alta tasa de cambio implica que para el momento en que el rastreador está descargando las últimas páginas de un sitio es muy probable que las primeras hayan cambiado o que páginas nuevas se hayan agregado al sitio, incluso que las páginas hayan sido suprimidas.

El comportamiento de un rastreador Web es el resultado de una combinación de políticas: [12]

- Una política de selección que indica las páginas para descargar,
- Una política de reinserción para verificar si hay cambios en las páginas,
- Una política de cortesía que indica la manera de evitar la sobrecarga de los sitios Web,
- Una política de paralelización que indica la forma de coordinar rastreadores Web distribuidos.

Teniendo en cuenta el tamaño actual de la Web, incluso grandes motores de búsqueda sólo cubren una porción de la parte a disposición del público. Un estudio del 2005 demostró que los grandes motores de búsqueda únicamente pueden llegar a indexar del 40% al 70% de la Web indexable, [23] un estudio previo realizado por el Dr. Steve Lawrence y Giles Lee demostró que ningún motor de búsqueda indexó más del 16% de la Web en 1999. [32]

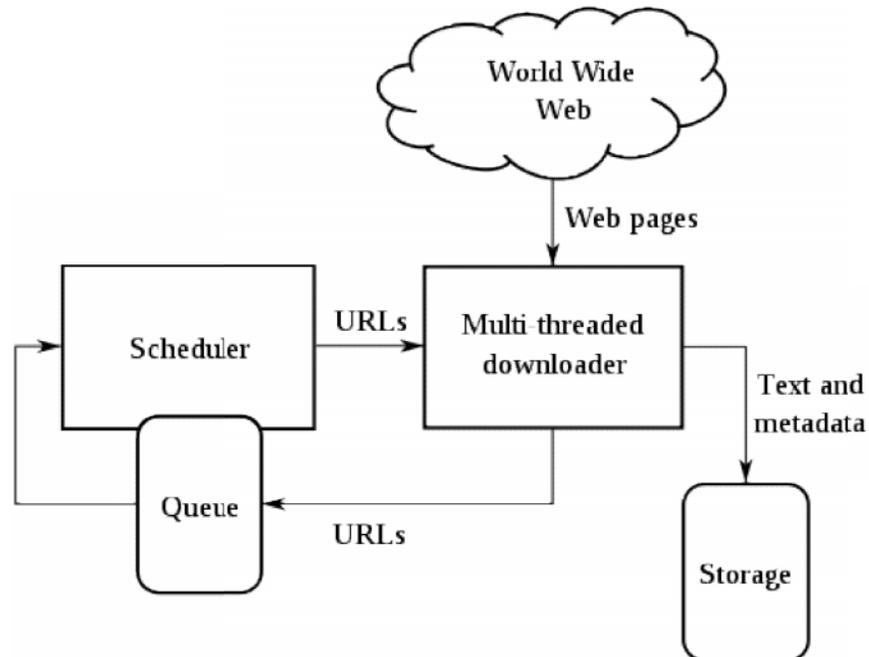
Como un rastreador siempre descarga sólo una fracción de las páginas Web, es muy conveniente que la fracción descargada contenga las páginas más relevantes y no sólo una muestra aleatoria de la Web.

Esto requiere una métrica de importancia para priorizar las páginas Web. La importancia de una página es una función de su calidad intrínseca, su popularidad en términos de vínculos o visitas, e incluso de su URL (este último es el caso de los motores de búsqueda vertical limitada a un nivel superior de dominio único, o motores de búsqueda restringidos a un sitio Web). El diseño de una política de buena selección tiene una dificultad añadida: debe trabajar con información

parcial, ya que el conjunto completo de páginas Web no se conoce durante el rastreo.

2.5.2 Arquitectura del crawler: Un rastreador no sólo debe tener una buena estrategia de rastreo, sino que también debe tener una arquitectura altamente optimizada.

Figura 10. Arquitectura de alto nivel de rastreador Web normal.



FUENTE: Castillo, Carlos; Effective Web Crawling .Rastreo Web eficaces (Ph.D. thesis). (Tesis doctoral). [12]

Shkapenyuk y Suel [54] señaló que "si bien es bastante fácil de construir un rastreador lento que descargue algunas páginas por segundo durante un corto período de tiempo, la creación de un sistema de alto rendimiento que se pueden descargar millones de páginas en varios semana presenta una serie de retos en el diseño del sistema, entrada y salida, eficiencia de la red, la robustez y capacidad de gestión".

Los rastreadores Web son una parte central de motores de búsqueda, y detalles de sus algoritmos y arquitectura se conservan como secretos comerciales. Cuando los diseños de orugas se publican, a menudo existe una importante falta de detalle que los demás no pueden reproducir la obra.

2.5.3 Ejemplos de rastreadores web: La siguiente es una lista de Crawlers de propósito general creados por diferentes compañías y motores de búsqueda.

- **Yahoo! Slurp:** Es el nombre del rastreador de búsqueda de Yahoo.
- **Msnbot:** Es el nombre de Microsoft, Bing Web Crawler.
- **Fast Crawler:** Es un rastreador distribuido, utilizado por Fast Search & Transfer.
- **Googlebot:** Se describe con cierto detalle, pero la referencia es sólo una primera versión de su arquitectura, que se basa en C++ y Python. El rastreador se integró con el proceso de indexación, ya que analizar el texto se hizo para la indexación de texto completo y también para la extracción de URL. Hay una dirección URL del servidor que envía listas de URLs a ser leída por varios procesos de rastreo. [9]

Durante el análisis, las URL que se encontraron se pasan a una URL del servidor que comprueba si la dirección se ha visto anteriormente. En caso contrario, la dirección ha sido añadida a la cola de la URL del servidor.

- **Methabot:** Es un rastreador Web scripts escritos en C, publicado bajo la licencia de ISC.
- **Arachnode.NET:** Es un código abierto. NET el rastreador Web se encuentra escrito en C # con SQL Server 2005 / SQL Server 2008 y Lucene.
- **Polybot:** Es un rastreador distribuido escrito en C++ y Python, que se compone de un gestor de rastreo. Las URLs recopilada se añaden a una cola en el disco, y el posterior tratamiento es buscar las direcciones URL que fueron previamente almacenadas. [54]
- **Webrace:** Es un módulo de rastreo y almacenamiento en caché implementado en Java, y se utiliza como parte de un sistema más genérico llamado erace. El sistema recibe las solicitudes de los usuarios para la descarga de páginas Web, de modo que el rastreador de actos, en parte, como un servidor proxy inteligente. El sistema también controla las solicitudes de "suscripciones" a las páginas Web que deben ser controlados. La característica más sobresaliente de WebRACE es que, mientras la mayoría de los rastreadores de comenzar con un conjunto de URLs, WebRACE está continuamente recibiendo nuevas URL que se generan. [61]

2.5.4 Funcionamiento del web crawler implementado en Polaris v2.0: La implementación del Crawler en Polaris versión 2.0 es parte fundamental del módulo de Minería de Estructura Web, ya que por medio del Crawler se procede a realizar la recolección de información Web que se necesita para calcular los pesos de los nodos mediante los algoritmos de minería de estructura Web.

El crawler tiene diferentes filtros para trabajar acorde a las necesidades del usuario, entre los cuales se encuentra el tipo de protocolo, los dominios de los sitios Web, el nivel que desea trabajar, el número de enlaces por nivel y el tipo de enlace que se desee trabajar: internos o externos (los cuales a su vez se clasifican en entrantes y salientes), esta última clasificación se hace internamente dentro del Web crawler como también diferentes filtros para eliminar los enlaces a documentos con los cuales no se puede trabajar, como los archivos pdf, imágenes, etc. También existe diferentes filtros para eliminar carpetas repetidas dentro de un enlaces del sitio Web, los formatos más comunes de páginas Web y validaciones de parámetros dentro de la misma.

La forma como se realiza la recuperación de la información es mediante una URL principal, la cual se ha denominado página semilla, a partir de esta se inicia a cargar el código HTML, ya que es el código Web que se ejecuta en las terminales clientes, y se procede a identificar dentro del código donde se encuentran los diferentes tipos de enlaces, se almacenan estos en una lista de Urls a visitar, frontera de rastreo, y posteriormente se da inicio a cada uno de estos enlaces para aplicar nuevamente el procedimiento de cargar su código HTML y buscar los vínculos que tiene. Este procedimiento se hace de manera continua, dinámica y recursiva hasta un nivel de profundidad escogido por el usuario de la herramienta.

Toda la información recuperada por el crawler es guardada en la base de datos de la herramienta o en un archivo de texto plano con la finalidad de hacer una herramienta que trabaja tanto on-line como off-line.

3. MINERÍA DE LA ESTRUCTURA DE LA WEB

Los motores de búsqueda primitivos (implementados antes de 1998) solían utilizar análisis de similitud entre los términos de búsqueda propuestos por el usuario y las páginas Web indexadas. En 1996, se dieron cuenta de que únicamente el análisis de similitud en el contenido no era suficiente por las siguientes razones:

- El número de páginas Web creció rápidamente durante la segunda mitad de los años 90 y dada una búsqueda y utilizando dicho método, el número de páginas relevantes podía ser enorme.
- Los métodos de análisis de similitud podían ser fácilmente manipulados (spamming). El dueño de una página podía repetir palabras importantes o introducir palabras relacionadas para aparecer en las primeras posiciones de los buscadores.

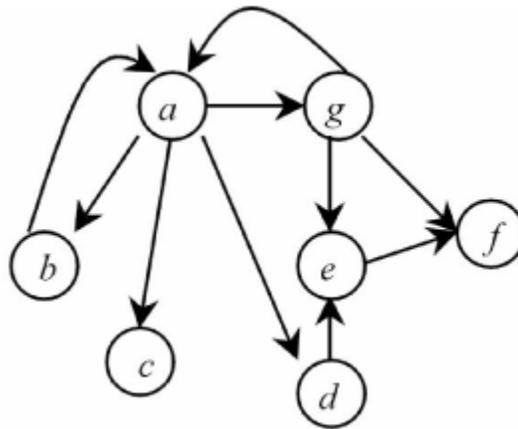
Por ello, los investigadores comenzaron a trabajar en la búsqueda de una solución al problema. Finalmente recurrieron a los hiperenlaces. A pesar de que los documentos tradicionales se consideraban independientes unos de otros, las páginas Web están conectadas por hiperenlaces, lo que proporciona gran información sobre la relevancia de las mismas. Los hiperenlaces son utilizados para organizar grandes cantidades de información en un sitio Web o para unir páginas con contenido relacionado.

En 1998, se diseñaron los dos algoritmos basados en hiperenlaces que más han influido en la Web: Page Rank (usado por Google) y HITS. Hoy en día, los buscadores consideran multitud de factores al producir los resultados presentados al usuario además de los métodos de evaluación de páginas basados en hiperenlaces.

3.1 PATRONES DE ENLACE

La World Wide Web puede ser vista estructuralmente como un grafo dirigido o dígrafo, donde cada página Web es un nodo o vértice y cada enlace es un arco o arista (**Figura 11**). Aunque los documentos Web ofrecen información textual, la conectividad del dígrafo, además de permitir la navegación, puede ayudar a caracterizarlos. Esa caracterización se logra mediante el procesamiento de su conectividad en matrices de adyacencia, en las que se indica con un valor uno (1) la existencia de un arco y con un cero (0) la inexistencia del mismo, como se muestra en la **Figura 12**. [39]

Figura 11. La Web representada a través de un dígrafo.



Fuente: Merlino-Santesteban, Cristian. *Análisis de conectividad en la recuperación de*. Brasilia : Ci. Inf., 2003. págs. 113-119. Vol. 32. [39]

Figura 12. Matriz de adyacencia del grafo de la Figura 11.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
<i>a</i>	-	1	1	1	0	0	1
<i>b</i>	1	-	0	0	0	0	0
<i>c</i>	0	0	-	0	0	0	0
<i>d</i>	0	0	0	-	1	0	0
<i>e</i>	0	0	0	0	-	1	0
<i>f</i>	0	0	0	0	0	-	0
<i>g</i>	1	0	0	0	1	1	-

Arcos salientes

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
<i>a</i>	-	1	0	0	0	0	1
<i>b</i>	1	-	0	0	0	0	0
<i>c</i>	1	0	-	0	0	0	0
<i>d</i>	1	0	0	-	0	0	0
<i>e</i>	0	0	0	1	-	0	1
<i>f</i>	0	0	0	0	1	-	1
<i>g</i>	1	0	0	0	0	0	-

Arcos entrantes

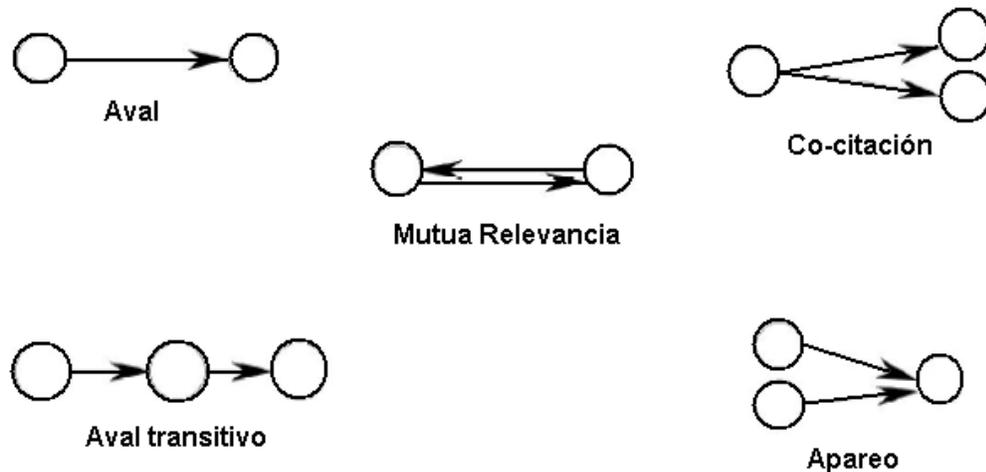
Fuente: Merlino-Santesteban, Cristian. *Análisis de conectividad en la recuperación de*. Brasilia : Ci. Inf., 2003. págs. 113-119. Vol. 32. [39]

Un hipervínculo en una página Web que conecta un documento con otro y representa un aval implícito con la página destino. Si un vínculo de la página p_a enlaza a la página p_b , el autor de p_a (por suposición) está recomendando a p_b . Si este enlace no estuviera presente la probabilidad que ambas páginas trataran el mismo o parecido tema sería mucho menor. Esta presunción deriva principalmente de la teoría hipertextual donde el contenido externo introducido por los hiperenlaces tiende a ser de alta calidad y utilidad.

Cuando consideramos dos vínculos, obtenemos un variado número de patrones básicos, como lo indica la **Figura 13**. Dos páginas apuntándose entre sí refuerzan la intuición acerca de su mutua relevancia. Una página referenciando dos páginas

distintas (co-citación) sugiere la probabilidad de que estén relacionadas en contenido. Dos documentos vinculando a un tercero producen un apareo (coupling), es decir, dos páginas están relacionadas mutuamente sin estar enlazadas entre sí.

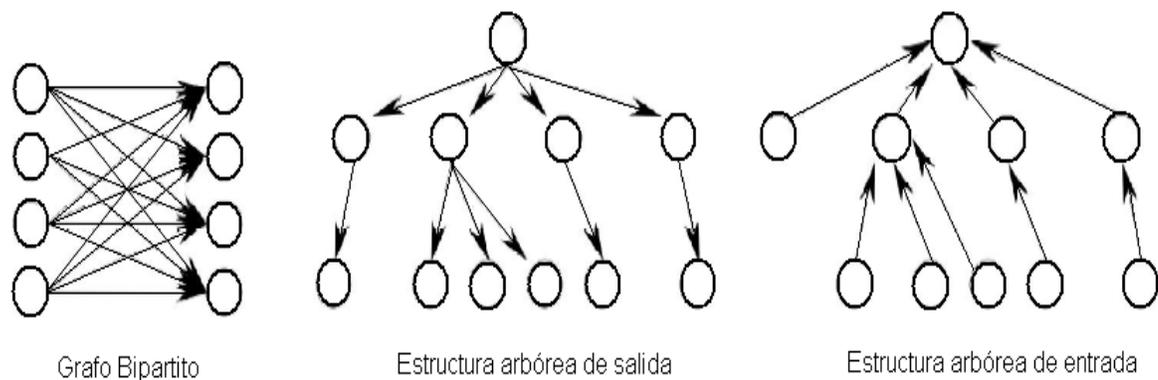
Figura 13. Patrones básicos formados por dos enlaces.



Fuente: Merlino-Santesteban, Cristian. *Análisis de conectividad en la recuperación de.* Brasilia : Ci. Inf., 2003. págs. 113-119. Vol. 32. [39]

Un aval transitivo ocurre cuando la página p_g vincula a p_u , la cual a su vez enlaza a p_v . Transitivamente, p_g es considerada como aval de p_v aunque este sea un aval débil. Y por consiguiente estas estructuras pueden combinarse unas con otras formando patrones muy complejos que fortalecen la relación entre las páginas (**Figura 14**).

Figura 14. Patrones complejos formados por numerosos nodos y vínculos.



Fuente: Merlino-Santesteban, Cristian. *Análisis de conectividad en la recuperación de.* Brasilia : Ci. Inf., 2003. págs. 113-119. Vol. 32. [39]

De las múltiples combinaciones posibles, la generalización de la co-citación con la estructura arbórea de salida (Out Degree) y la generalización del apareo con la estructura arbórea de entrada (In Degree) han sido de particular interés para el campo de la recuperación de información, especialmente aquellas estructuras con grandes grados de conectividad a un nodo raíz.

La estructura arbórea de entrada indica que si muchas páginas diferentes apuntan directamente o transitivamente a un nodo, es probable que éste sea un recurso valioso (autoridad) sobre algún tema o de interés compartido por las otras páginas. Esto es análogo a la medición del impacto de un artículo científico por el número de citas que recibe. El interés en la estructura arbórea de salida está dado por la suposición de que si un nodo vincula a muchas páginas Web valiosas en un tópico, denominado hub, luego podemos considerarlo como una buena fuente para buscar información. En este sentido, un hub, es análogo a un artículo de revisión bibliográfica. Partiendo de estas interpretaciones podemos esperar que los documentos con alto valor de autoridad tengan contenido relevante mientras que los documentos con alto valor de hub tengan hiperenlaces a contenidos relevantes.

En este acercamiento al estudio de patrones, los algoritmos de conectividad, también incluyen en su análisis el criterio de cantidad e importancia de los enlaces entrantes y salientes dado que no todos los vínculos tienen el mismo peso, y la simple cuenta de citas a un documento no siempre implica importancia. En otras palabras, un documento Web con un alto grado de citación y/o luminosidad podría verse como un documento muy importante, pero la calidad de los enlaces que apunta a otras páginas o recibe no es siempre la misma. En muchos casos, una página con unos pocos enlaces entrantes de calidad se juzga más visible que una página con centenares de vínculos de sitios menos valiosos.

3.2 ALGORITMOS DE ORDENACIÓN POR RELEVANCIA BASADOS EN LA CONECTIVIDAD

Básicamente, los algoritmos de ordenación por relevancia basados en la conectividad pueden dividirse en tres tipos:

- Dependientes de la consulta del usuario, los cuales asignan una puntuación (valor de relevancia) a una página en el contexto de un proceso de equiparación.
- Cuasi dependientes de la consulta del usuario, los cuales otorgan una puntuación a una página en el entorno de un corpus documental (URL) dado.
- Independientes de la consulta del usuario, los cuales asignan una puntuación a una página independientemente de la búsqueda proporcionada. A diferencia de lo expuesto por Henzinger [24], se incorporó el tipo cuasi dependiente porque un

URL (Uniform Resource Locator) también constituye un punto de partida y brinda un contexto de búsqueda.

3.2.1 Algoritmos dependientes de la consulta del usuario: Este esquema de ordenamiento tiene su implementación más popular en el algoritmo HITS (Hyperlink Induced Topic Search) desarrollado por Kleinberg [28] como mejora al trabajo realizado por Carriere y Kazman [11]. La idea básica es coleccionar un conjunto de páginas iniciales sobre un tópico a partir de los resultados obtenidos de la búsqueda de una o más palabras en un sistema de recuperación de información, ampliar el mismo con la adición de las páginas que le enlazan y las páginas que son referenciadas por él, construir una matriz de adyacencia W y calcular el vector representativo de $A = W^T W$ y $H = W W^T$. Estos vectores corresponden a los pesos de los nodos autoridad y hub respectivamente.

Chakrabarti et al [13]. Extendieron HITS en el ARC (Automatic Resource Compiler) incorporando el texto próximo a los anclajes en la computación de los pesos de los nodos autoridad y hub. La utilidad del texto alrededor de los anclajes había sido mostrada años antes por McBryan, pero fusionarlo con HITS fue un paso innovador.

La creación del ARC fue ideada como un mecanismo para compilar y mantener automáticamente listas de recursos de alta "calidad". Chakrabarti continuó mejorando el ARC en el proyecto CLEVER [27], el cual conjuga varios algoritmos de conectividad para potenciar su funcionalidad.

Un algoritmo alternativo en esta línea de investigación, SALSA (Stochastic Approach for Link-Structure Analysis), fue propuesto por Lempel y Moran [33]. SALSA emplea el mismo meta-algoritmo que HITS, pero asigna a cada nodo un peso autoridad proporcional a la suma de los pesos de sus vínculos entrantes y un peso hub proporcional a la suma de los pesos de sus vínculos salientes, es decir, SALSA calcula, para todas las páginas, la suma de los pesos de sus enlaces salientes y entrantes, y normaliza estos dos vectores. Este algoritmo resalta básicamente la "calidad" de una página a través de la ponderación de su conectividad dentro del subgrafo generado.

3.2.2 Algoritmos cuasi-dependientes de la consulta del usuario: En estos algoritmos la entrada del proceso de búsqueda no son los términos de una consulta, sino el URL de una página (o un sitio Web) y la salida es un conjunto ordenado de páginas Web relacionadas. Una página relacionada es aquella que contiene el mismo tópico que la página original, pero no son necesariamente idénticas semánticamente.

El algoritmo más implementado es el de co-citación. Dean y Henzinger [15] implementaron esta técnica con buenos resultados. El método usado se basó en encontrar páginas que vinculasen al URL modelo y después determinar las

páginas enlazadas por él, obteniendo como respuesta una lista ordenada por autoridad.

Rafiei y Mendelzon [51] construyeron un prototipo híbrido entre Page Rank y SALSA, denominado TOPIC (Toronto Page Influence Computation), que dado un URL inicial identifica los temas donde la página tiene una buena reputación y genera una lista ordenada por autoridad de estos tópicos (asociados a URLs).

3.2.3 Algoritmos independientes de la consulta del usuario: Los algoritmos independientes de la consulta del usuario producen un ranking independientemente de la similitud consulta-documento puesto que su objetivo principal es medir la "calidad" intrínseca de una página. El mejor ejemplo de esta implementación es el algoritmo Page Rank (PR) usado por el motor de búsqueda Google como uno de sus componentes para ayudar a ordenar los resultados retornados por una búsqueda. Partiendo de un grafo (off-line) construido a priori, Page Rank calcula la importancia de una página otorgando a cada hipervínculo que le apunta un peso proporcional a la autoridad de la página que lo contiene. Para determinar la autoridad de la página citante, el Page Rank es utilizado recursivamente unas 100 veces (hasta que los valores converjan). El Page Rank de una página Web no es influenciado por la página en sí misma o alguna consulta potencial, pero alcanza a una media objetiva universal basándose solamente en determinaciones subjetivas de importancia aportadas por los hiperenlaces.

3.3 ALGORITMOS DE MINERÍA DE ESTRUCTURA WEB IMPLEMENTADOS EN POLARIS VERSIÓN 2.0

En internet, cada enlace que entra a una página es un punto para ella. Por ejemplo, si se escribe algo, y ese algo es enlazado por el *New York Times*, la página ha ganado una importancia por el solo hecho de ser enlazada desde el *New York Times*. Así es como se clasifica la importancia de las páginas para Google. El Page Rank fue creado por Larry Page y Sergey Brin, fundadores de Google, en el año 1998.

Pero este método de clasificación no es nuevo. Observando la historia, encontramos a Jon Kleinberg, quien creó HITS (Hypertext Induced Topic Search) unos años antes que Page y Brin. De hecho, los fundadores de Google le citan en la creación del Page Rank. Este algoritmo tuvo mucha importancia antes del furor de las *puntocom*, antes de que Google se convirtiera en un éxito.

Pero aun más atrás, encontramos el trabajo de Gabriel Pinski y Francis Narin, quien, al igual que el algoritmo HITS o Page Rank, desarrollaron un modo de clasificar las publicaciones según las veces que era citada en otras publicaciones. Era la teoría iterativa de clasificación. Ya en el año 1965, Charles H Hubblell había desarrollado algoritmos de este tipo.

Ahora se ha descubierto un algoritmo de este tipo que data de 1941, creado por Wassily Leontief, economista de Harvard. Dividió un país en sectores que daban y recibían recursos de otros, pero no en la misma medida. Este trabajo le mereció el Premio Nobel de Economía en el año 1973.

Para el caso en particular de la herramienta **POLARIS VERSIÓN 2.0**, se estudió e implementó dos algoritmos de Minería de Estructura Web y se desarrolló e implementó un algoritmo, teniendo como base el estudio realizado a la estructura Web. Los algoritmos en mención son: Page Rank, HITS y Web Linker, respectivamente.

3.3.1 Page Rank: Page Rank es un algoritmo de análisis de hipervínculos para medir la popularidad de las páginas Web. Su objetivo es medir de forma relativa la importancia de cada página Web dentro del total. El resultado es un valor numérico que representa la relevancia de dicha página y es, entre otros factores, uno de los métodos empleados para posicionar las páginas Web en los resultados de búsquedas de Google.

Page Rank fue desarrollado por Larry Page (de allí su nombre Page-Rank) y Sergey Brin como parte de un proyecto de investigación acerca de nuevas formas de búsquedas. El proyecto comenzó en el año 1995.

Page Rank es una marca registrada y patentada por Google el 9 de enero de 1999 que ampara una familia de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos (o páginas Web) indexados por un motor de búsqueda. Sus propiedades son muy discutidas por expertos en optimización de motores de búsqueda.

No son conocidos públicamente los métodos exactos que se emplean para determinar el Page Rank de una página Web, pero existen estimaciones bastante acertadas al respecto.

Page Rank confía en la naturaleza democrática de la Web utilizando su vasta estructura de enlaces como un indicador del valor de una página en concreto. Google interpreta un enlace de una página A a una página B como un voto, de la página A, para la página B. Pero Google mira más allá del volumen de votos, o enlaces que una página recibe; también analiza la página que emite el voto. Los votos emitidos por las páginas consideradas "importantes", es decir con un Page Rank elevado, valen más, y ayudan a hacer a otras páginas "importantes". Por lo tanto, el Page Rank de una página refleja la importancia de la misma en Internet.

➤ **Algoritmo Page Rank:** El algoritmo inicial del Page Rank se lo puede encontrar en el documento original donde sus creadores presentaron el prototipo de Google: "The Anatomy of a Large-Scale Hypertextual Web Search Engine". [9]

$$r(i) = d * \sum_{j \in \overline{B(i)}} \frac{r(j)}{n(j)} + \frac{(1-d)}{m}$$

Donde:

r(i) = Page Rank de la página i.

N(i) = Número de enlaces salientes de la página i.

B(i) = Número de páginas que apuntan a la página i.

m = Número total de nodos del gráfico.

d = Factor de decaimiento entre **0** y **1**. [35]

Algunos expertos aseguran que el valor de la variable *d* suele ser 0,85. Representa la probabilidad de que un navegante continúe pulsando links al navegar por Internet en vez de escribir una URL directamente en la barra de direcciones o pulsar uno de sus marcadores y es un valor establecido por Google. Por lo tanto, la probabilidad de que el usuario deje de pulsar links y navegue directamente a otra Web aleatoria es *1-d*. La introducción del factor de decaimiento en la fórmula resta algo de peso a todas las páginas de Internet y consigue que las páginas que no tienen enlaces a ninguna otra página no salgan especialmente beneficiadas. Si un usuario aterriza en una página sin enlaces, lo que hará será navegar a cualquier otra página aleatoriamente, lo que equivale a suponer que una página sin enlaces salientes tiene enlaces a todas las páginas de Internet.

El peso o importancia de una página es el resultado de una "votación" entre todas las demás páginas de la World Wide Web acerca del nivel de importancia que tiene esa página. Un hiperenlace a una página cuenta como un voto de apoyo. El Page Rank de una página se define recursivamente y depende del número y Page Rank de todas las páginas que la enlazan. Una página que está enlazada por muchas páginas con un Page Rank alto consigue también un Page Rank alto. Si no hay enlaces a una página Web, no hay apoyo a esa página específica. El Page Rank de la barra de Google va de 0 a 10. Diez es el máximo Page Rank posible y son muy pocos los sitios que gozan de esta calificación, 1 es la calificación mínima que recibe un sitio normal, y cero significa que el sitio ha sido penalizado o aún no ha recibido una calificación de Page Rank. Parece ser una escala logarítmica. Los detalles exactos de esta escala son desconocidos.

➤ **Pseudocódigo:** El pseudocódigo del algoritmo Page Rank es el siguiente:

- (1) $S \leftarrow \text{cualquier vector}$
- (2) $r \leftarrow A^T \times S$
- (3) Si $|r - s| < \epsilon$ fin. *r* es el vector Page Rank
- (4) $S \leftarrow r$, ir a (2).

Donde r es el vector $m \times 1$ $r = [r(1), r(2), \dots, r(m)]$

Y los elementos de A son $a_{ij} = \begin{cases} \frac{1}{N(i)}, & \text{si } i \rightarrow j \\ 0, & \text{en otro caso} \end{cases}$ [35]

➤ **Cálculo del Page Rank:** Para el cálculo de una página, su Page Rank existente (si tiene alguno) es abandonado completamente y un nuevo cálculo es realizado donde la página se basa solamente en el Page Rank "votado" para ello por sus enlaces entrantes actuales, los cuales pueden haber cambiado desde la última vez que se calculó el Page Rank.

La ecuación muestra claramente cómo llega el Page Rank de una página. Pero lo que no es inmediatamente obvio es que no puede funcionar si el cálculo se hace una sola vez. Supongamos que tenemos 2 páginas, A y B, cada una enlazada a la otra, y ninguna de las dos tienen enlaces de otro tipo. Esto es lo que ocurre:

Paso 1: Calcular el Page Rank de la página A con el valor de sus enlaces entrantes.

La página A ahora tiene un nuevo valor de Page Rank. En el cálculo se usó el valor del enlace entrante de la página B. Pero la página B tiene un enlace entrante (de la página A) y su nuevo valor de Page Rank no ha sido trabajado todavía, por eso el nuevo valor de Page Rank de la página A está basado en datos imprecisos y por lo tanto no puede ser preciso.

Paso2: Calcular el Page Rank de la página B con el valor de sus enlaces entrantes.

La página B tiene ahora un nuevo valor de Page Rank, pero no puede ser preciso ya que el cálculo usó el nuevo valor de Page Rank del enlace entrante desde la página A, el cual es impreciso.

No se puede concluir con el Page Rank de A hasta que se conozca el Page Rank de B, y tampoco se puede concluir con el Page Rank de B hasta que se conozca el Page Rank de A.

No se puede rehacer los cálculos usando los nuevos valores y conseguir resultados más precisos, ya que siempre se está usando valores imprecisos para los cálculos, por lo que los resultados serán siempre imprecisos.

El problema se soluciona repitiendo los cálculos muchas veces. En cada repetición se consiguen resultados ligeramente más precisos. De hecho, es imposible conseguir una precisión total ya que los cálculos siempre estarán basados en valores imprecisos. Cuarenta o cincuenta iteraciones son suficientes para alcanzar un punto donde más iteraciones no alterarían de manera notable el resultado. Esto

es precisamente lo que hace Google en cada actualización y esa es la razón por la que tarda tanto.

Una cosa a tener en cuenta es que los resultados que se obtienen de los cálculos son proporcionales. Los números obtenidos deben ser puestos en una escala (conocida solo por Google) para conocer el Page Rank real de cada página. Aún así, no se puede usar los cálculos para canalizar el Page Rank dentro de un sitio alrededor de sus páginas ya que algunas páginas reciben una proporción mayor de Page Rank que otras.

➤ **Enlaces colgantes:** Los enlaces colgantes son simplemente enlaces que apuntan a alguna página que no tiene enlaces salientes. Afectan al modelo porque no está claro donde debe ser distribuido su peso, y hay un gran número de ellas. A menudo estos enlaces colgantes son páginas que no se han descargado todavía. Porque los enlaces colgantes no afectan el ranking de ninguna otra página directamente, se plantea simplemente borrarlos del sistema hasta que el Page Rank de todas las páginas es calculado. Después de que todos los Page Rank hayan sido calculados, las páginas pueden ser añadidas otra vez sin afectar significativamente las cosas. [9]

Un enlace colgante es un enlace hacia una página que no tiene enlaces salientes, o un link hacia una página que Google no ha indexado. En ambos casos Google elimina los enlaces justo al empezar los cálculos y los reinstaura justo después de acabar los cálculos. De esa manera, los enlaces colgantes afectan mínimamente al Page Rank de otras páginas.

➤ **Enlaces entrantes:** Los enlaces entrantes (enlaces hacia el sitio Web desde el exterior) son una manera de incrementar el Page Rank total de un sitio Web. La otra es añadir más páginas. No hay que preocuparse de donde provengan los enlaces. Google reconoce que un Webmaster no tiene control sobre que sitios Web le enlazan, así que los sitios Web no son penalizados por enlaces entrantes.

El Page Rank de la página que nos enlaza es importante, pero también el número de enlaces salientes de esa página.

Una vez que el Page Rank es inyectado en su sitio Web, los cálculos se rehacen y el Page Rank de cada página cambia. Dependiendo de la estructura de enlaces interna, el Page Rank de algunas páginas se incrementa y en otras permanece inalterado pero en ninguna se pierde Page Rank.

Es beneficioso tener enlaces entrantes de las páginas a las que se les está canalizando Page Rank. Una inyección de Page Rank a cualquier página será repartida por todo el sitio Web a través de enlaces internos. Las páginas importantes recibirán un incremento, pero no tanto como si estuvieran enlazadas directamente. La página que recibe el enlace entrante es la que más ganancia

obtiene.

Se puede considerar una Web como una pequeña red autónoma de páginas. Cuando se hacen cálculos de Page Rank, se está atendiendo esta pequeña red. Si hace un enlace a otro sitio, se pierde algo de Page Rank de red, y si se recibe un enlace, el Page Rank de la red es incrementado. Pero esto no es así. Para el cálculo de Page Rank, solo hay una red, cualquier página que Google tenga en su índice. Cualquier iteración del cálculo se hace en la red entera y no en sitios individuales.

La red entera está interconectada, y cada enlace de cada página juega su papel en cada iteración de los cálculos, es imposible calcular el efecto de los enlaces entrantes a un sitio Web con una precisión realista.

➤ **Enlaces salientes:** Los enlaces salientes son el desagüe del Page Rank total de un sitio Web. Hacen "gotear" el Page Rank. Para contrarrestar el efecto drenado, hay que asegurarse de que esos enlaces son recíprocos. Por el Page Rank de las páginas de cada extremo de cada link externo, y por el número de enlaces salientes de esas páginas, los enlaces recíprocos pueden hacer ganar o perder Page Rank. Se necesita tener cuidado cuando se escoge donde intercambiar enlaces.

Cuando el Page Rank se pierde en un sitio Web vía un enlace a otro sitio Web, todas las páginas en la estructura de enlaces internos se ven afectadas. (Esto no se muestra siempre después de 1 iteración). La página a la que enlaza hace una diferencia entre las páginas que sufren mayor pérdida. Sin un programa para realizar los cálculos en una estructura de enlaces específica, es muy difícil decidir la página correcta para enlazar, pero generalmente es la página con menos Page Rank.

Muchos sitios Web necesitan contener algunos enlaces salientes que no tienen nada que ver con Page Rank. Desafortunadamente, todos los enlaces salientes normales, pierden Page Rank. Pero hay maneras "anormales" de enlazar a otros sitios Web que no resulten en pérdidas de Page Rank. Page Rank tiene pérdidas cuando Google reconoce un enlace a otro sitio.

➤ **Manipulación:** Debido a la importancia comercial que tiene aparecer entre los primeros resultados del buscador, se han diseñado métodos para manipular artificialmente el Page Rank de una página. Entre estos métodos hay que destacar el spam, consistente en añadir enlaces a una cierta página Web en lugares como blogs, libros de visitas, foros de Internet, etc. con la intención de incrementar el número de enlaces que apuntan a la página.

A principios del 2005 Google implementó un nuevo atributo para hiperenlaces llamado "nofollow", el cual es un atributo en código HTML que se coloca a un

enlace cuando se quiere que este no sea tenido en cuenta por los buscadores. De esta manera, Google y otros buscadores cuando calculan el peso de una página, no tienen en cuenta los links que tengan este atributo.

➤ **Ejemplos Page Rank**

Ejemplo 1 Page Rank.

Dampening (d) = 0.8

Iteraciones = 1

Nodos (m) = 5

Scores = (1/5, 1/5, 1/5, 1/5, 1/5) = (0.2, 0.2, 0.2, 0.2, 0.2)

$$r(i) = d \cdot \sum_{j \in \overline{B(i)}} \frac{r(j)}{n(j)} + \frac{(1-d)}{m}$$

r(i) = Page Rank de la página i.

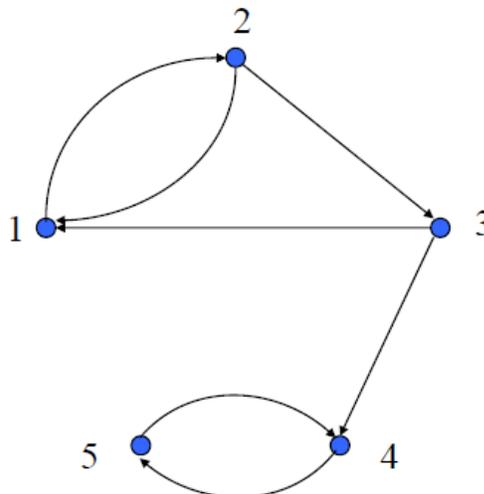
N(i) = Número de enlaces salientes de la página i.

B(i) = Número de páginas que apuntan a la página i.

m = Número total de nodos del gráfico.

d = Factor de decaimiento entre 0 y 1.

Figura 15. Grafo No. 1 para el cálculo del Page Rank (Ejemplo 1)



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

$$Aux = \frac{(1 - d)}{m} = \frac{1 - 0.8}{5} = 0.04$$

Acumulador = 0

Primera iteración.

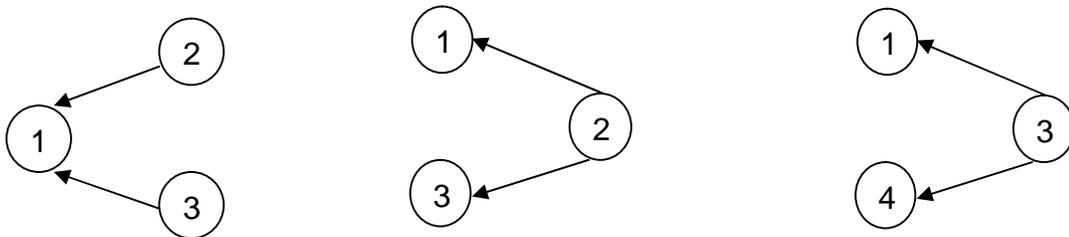
Para el nodo número 1.

In Links de 1 (2, 3)

Out Links de 2 (1, 3)

Out Links 3 (1, 4)

Figura 16. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 Page Rank)



Los links de salida tanto para el nodo dos como el nodo tres son de 2 para cada uno.

$$Acumulador = Acumulador + \frac{Score(2)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.2}{2}$$

$$Acumulador = 0.1$$

$$Acumulador = Acumulador + \frac{Score(3)}{NumLinks}$$

$$Acumulador = 0.1 + \frac{0.2}{2}$$

$$Acumulador = 0.2$$

$$r(1) = (Acumulador \cdot d) + Aux$$

$$r(1) = (0.2 \cdot 0.8) + 0.04$$

$$r(1) = 0.2$$

Para el nodo número 2.

In Links de 2 (1)

Out Links de 1 (2)

Figura 17. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 1 Page Rank)



Los links de salida del nodo uno es uno.

$$Acumulador = Acumulador + \frac{Score(1)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.2}{1}$$

$$Acumulador = 0.2$$

$$r(2) = (Acumulador \cdot d) + Aux$$

$$r(2) = (0.2 \cdot 0.8) + 0.04$$

$$r(2) = 0.2$$

Para el nodo número 3.

In Links de 3 (2)

Out Links de 2 (1, 3)

Figura 18. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 1 Page Rank)



Los links de salida para el nodo dos son 2.

$$Acumulador = Acumulador + \frac{Score(2)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.2}{2}$$

$$Acumulador = 0.1$$

$$r(3) = (Acumulador \cdot d) + Aux$$

$$r(3) = (0.1 \cdot 0.8) + 0.04$$

$$r(3) = 0.12$$

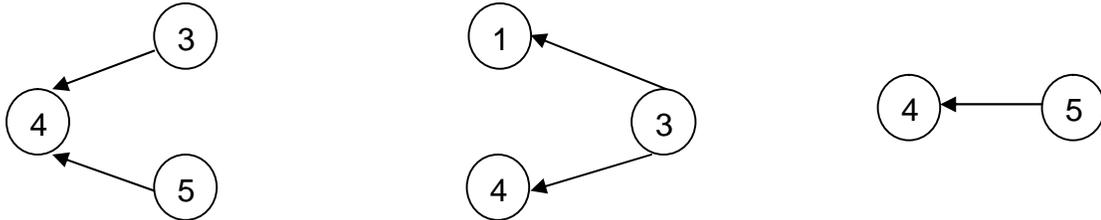
Para el nodo número 4.

In Links de 4 (3, 5)

Out Links de 3 (1, 4)

Out Links 5 (4)

Figura 19. Grafos de la primera iteración para el nodo No. 4 (Ejemplo 1 Page Rank)



Los links de salida para el nodo tres son 1 y 3 y para el nodo cinco es el 4.

$$Acumulador = Acumulador + \frac{Score(3)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.2}{2}$$

$$Acumulador = 0.1$$

$$Acumulador = Acumulador + \frac{Score(5)}{NumLinks}$$

$$Acumulador = 0.1 + \frac{0.2}{1}$$

$$Acumulador = 0.2$$

$$r(4) = (Acumulador \ d) + Aux$$

$$r(4) = (0.3 \ 0.8) + 0.04$$

$$r(4) = 0.28$$

Para el nodo número 5.

In Links de 5 (4)

Out Links de 4 (5)

Figura 20. Grafos de la primera iteración para el nodo No. 5 (Ejemplo 1 Page Rank)



Los links de salida del nodo cuatro es el 5.

$$Acumulador = Acumulador + \frac{Score(4)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.2}{1}$$

$$Acumulador = 0.2$$

$$r(5) = (Acumulador \cdot d) + Aux$$

$$r(5) = (0.2 \cdot 0.8) + 0.04$$

$$r(5) = 0.2$$

Las iteraciones pueden continuar sucesivamente, al terminar la iteración los scores de cada nodo con actualizados para el cálculo siguiente.

Ejemplo 2 Page Rank.

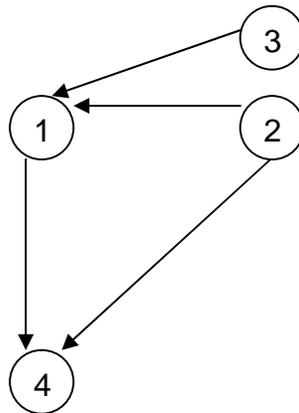
Dampening (d) = 0.85

Iteraciones = 1

Nodos (m) = 4

Scores = (1/4, 1/4, 1/4, 1/4) = (0.25, 0.25, 0.25, 0.25)

Figura 21. Grafo No. 2 para el cálculo del Page Rank (Ejemplo 2)



$$Aux = \frac{(1 - d)}{m} = \frac{1 - 0.8}{4} = 0.0375$$

Acumulador = 0

Primera iteración.

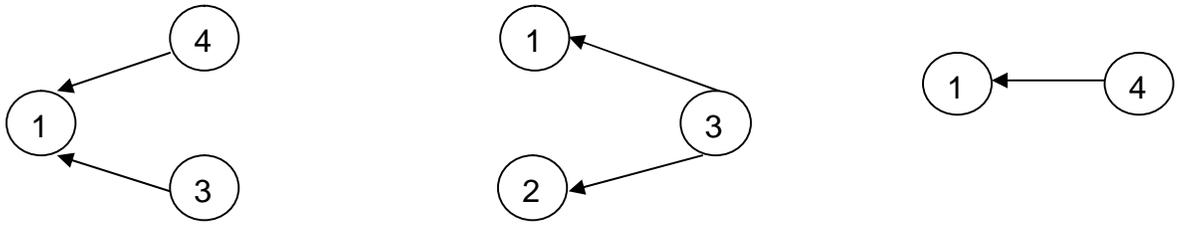
Para el nodo número 1.

In Links de 1 (3, 4)

Out Links de 3 (1, 2)

Out Links 4 (1)

Figura 22. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 2 Page Rank)



Los links de salida para el nodo tres son el 1 y 2 y para el nodo cuatro es el 1.

$$Acumulador = Acumulador + \frac{Score(3)}{NumLinks}$$

$$Acumulador = 0 + \frac{0.25}{2}$$

$$Acumulador = 0.125$$

$$Acumulador = Acumulador + \frac{Score(4)}{NumLinks}$$

$$Acumulador = 0.125 + \frac{0.25}{1}$$

$$Acumulador = 0.375$$

$$r(1) = (Acumulador \cdot d) + Aux$$

$$r(1) = (0.375 \cdot 0.85) + 0.0375$$

$$r(1) = 0.3562$$

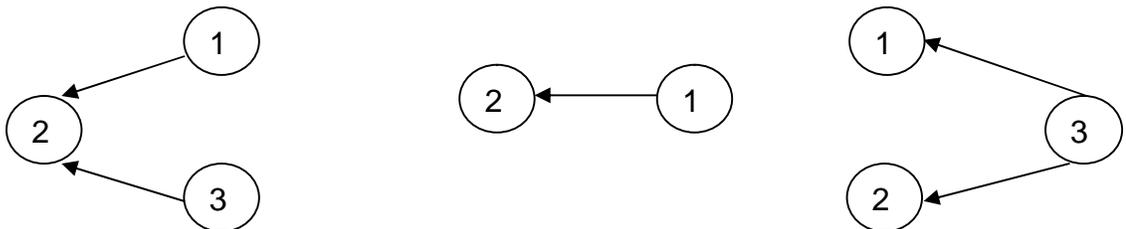
Para el nodo número 2.

In Links de 2 (1, 3)

Out Links de 1 (2)

Out Links 3 (1, 2)

Figura 23. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 2 Page Rank)



Los links de salida para el nodo uno es el 2 y para el nodo tres son el 1 y 2.

$$Acumulado = Acumulado + \frac{Score(1)}{NumLinks}$$

$$Acumulado = 0 + \frac{0.25}{1}$$

$$Acumulado = 0.25$$

$$Acumulado = Acumulado + \frac{Score(3)}{NumLinks}$$

$$Acumulado = 0.25 + \frac{0.25}{2}$$

$$Acumulado = 0.375$$

$$r(2) = (Acumulado \cdot d) + Aux$$

$$r(2) = (0.375 \cdot 0.85) + 0.0375$$

$$r(2) = 0.3562$$

Para el nodo número 3.

In Links de 3 (vacío)

$$r(3) = (Acumulado \cdot d) + Aux$$

$$r(3) = (0 \cdot 0.85) + 0.0375$$

$$r(3) = 0.0375$$

Para el nodo número 4.

In Links de 4 (vacío)

$$r(4) = (Acumulado \cdot d) + Aux$$

$$r(4) = (0 \cdot 0.85) + 0.0375$$

$$r(4) = 0.0375$$

Traducción:

Id 1 = nodo número 1;

Id 2 = nodo número 4;

Id 3 = nodo número 2;

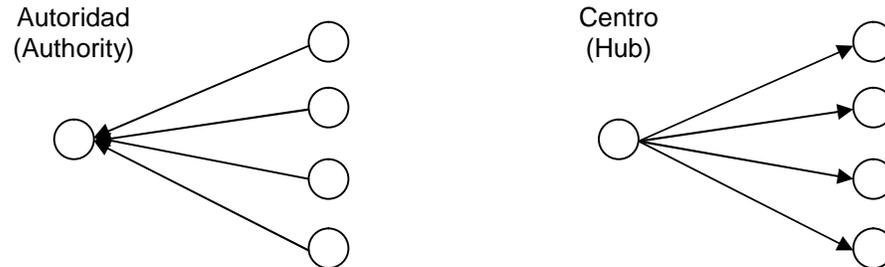
Id 4 = nodo número 3

Esto se debe a que en el momento de colocar los datos en el grafo se hace de la manera como van apareciendo.

3.3.2 HITS: En Enero de 1998, HITS (El acrónimo de Hypertext Induced Topic Search) fue propuesto por Jon Kleinberg. Este algoritmo, a diferencia de Page Rank, es dependiente de las consultas del usuario.

HITS se comporta de la siguiente manera: primero, expande el conjunto de páginas relevantes devueltas por un buscador para, a posteriori, listar los resultados de acuerdo a dos medidas: lista de autoridades y lista de centros.

Figura 24. Esquemas de una autoridad (authority) y un centro (hub).



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

Una autoridad (authority) es una página con muchos enlaces entrantes. La idea es que una página tiene autoridad si mucha gente confía en ella y la enlaza con sus páginas Webs.

Un centro (hub) es una página que contiene muchos enlaces salientes. Una página centro sirve como organizador de muchas páginas con autoridad.

La idea clave de HITS es que un buen centro apunta a muchas buenas autoridades, y que una buena autoridad es apuntada por muchos buenos centros.

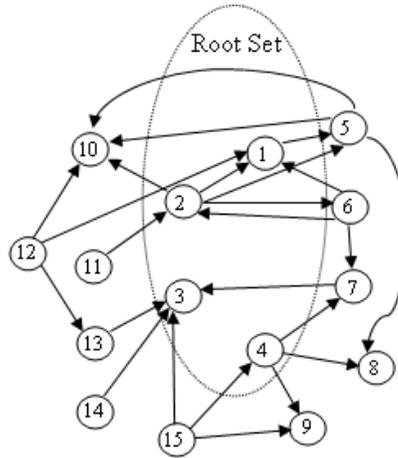
Esto significa que autoridades y centros deben tener una relación de refuerzo mutuo (mutual reinforcement relationship).

➤ **Algoritmo HITS:** HITS funciona de la siguiente manera.

1. Se envía una consulta q a un buscador. El buscador le devuelve t resultados más relevantes. Este conjunto de resultados lo denominamos W o conjunto raíz (root set).

2. Posteriormente, se crea el conjunto S , que es el resultado de engordar el conjunto W con todas las páginas que apuntan a o son apuntadas por cada página contenida en W . Como restricción tenemos que cada página de W puede introducir k páginas en S . El conjunto S se llama conjunto base (base set).

Figura 25. Ejemplo de un conjunto base S. El conjunto S contiene al Conjunto W (Root Set) y a las páginas añadidas en la expansión



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

3. Después, HITS trabaja con las páginas de **S**, dándoles a cada página en **S** una puntuación de autoridad y otra de centro. Sea la puntuación de autoridad de una página **i** **a(i)**, y su puntuación de centro **h(i)**. La relación de refuerzo mutuo se expresa como:

$$a(i) = \sum_{(j,i) \in E} h(j)$$

$$h(i) = \sum_{(i,j) \in E} a(j)$$

4. Finalmente, HITS selecciona las páginas con mayor puntuación en autoridad y centro, y las devuelve al usuario.

➤ **Pseudocódigo:** El pseudocódigo del algoritmo HITS es el siguiente:

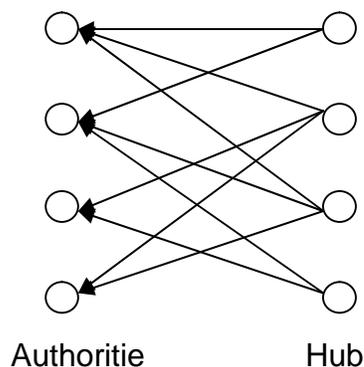
- (1) $G :=$ set of pages
- (2) for each page p in G do
- (3) $p.auth = 1$ // $p.auth$ is the authority score of the page p
- (4) $p.hub = 1$ // $p.hub$ is the hub score of the page p
- (5) function HubsAndAuthorities(G)
- (6) for step from 1 to k do // Run the algorithm for k steps
- (7) for each page p in G do // update all authority values first

- ```

(8) for each page q in $p.incomingNeighbors$ do // $p.incomingNeighbors$
 // is the set of pages that link to p
(9) $p.auth += q.hub$
(10) for each page p in G do // Then update all hub values
(11) for each page r in $p.outgoingNeighbors$ do // $p.outgoingNeighbors$
 // is the set of pages that p links to
(12) $p.hub += r.auth$

```

Figura 26. La finalidad de HITS es encontrar un conjunto de páginas relacionadas con el tema de búsqueda que tengan alta puntuación como autoridad y centro.



**Fuente:** Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

➤ **Ventajas y desventajas de HITS:** La ventaja más importante de HITS es la habilidad de listar los resultados de acuerdo a la consulta realizada, lo cual debería proporcionar páginas más importantes.

Sin embargo, HITS tiene grandes desventajas, como:

HITS no posee la capacidad de combatir la manipulación que posee Page Rank. Es relativamente fácil engañar a HITS añadiendo enlaces en nuestra página a páginas que son buenas autoridades.

HITS puede polucionar el tema de la consulta. Al expandir el conjunto proporcionado por el buscador, se pueden añadir fácilmente páginas que no tienen nada que ver el tema de consulta. A este problema se le denomina topic drift. El tiempo empleado en la consulta es bastante superior al empleado con otros algoritmos como Page Rank.

➤ **Ejemplos HITS**

**Ejemplo 1 HITS.**

Nodos = 4

Iteración = 1

VAscore = (1, 1, 1, 1)

VHscore = (1, 1, 1, 1)

Hacum = 1 + 0 + 4 + 2 = 7

Aacum = 2 + 2 + 0 + 0 = 4

Cálculo realizado para el grafo de la **Figura 21**.

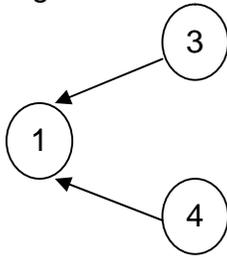
**Primera iteración.**

**Para el nodo número 1.**

In Links de 1 (3, 4)

Out Links de 1 (2)

Figura 27. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 HITS)



$$Ascore(1) = h(3) + h(4)$$

$$Ascore(1) = 1 + 1$$

$$Ascore(1) = 2$$

$$VAscore = (2, 1, 1, 1)$$

$$VHscore = (1, 1, 1, 1)$$

$$Hscore(1) = a(2)$$

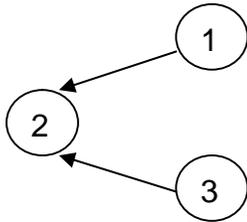
$$Hscore(1) = 1$$

**Para el nodo número 2.**

In Links de 2 (1, 3)

Out Links de 2 (vacío)

Figura 28. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 1 HITS)



$$\begin{aligned} \text{Ascore}(2) &= h(1) + h(3) \\ \text{Ascore}(2) &= 1 + 1 \\ \text{Ascore}(2) &= 2 \\ \text{VAscore} &= (2, 2, 1, 1) \\ \text{VHscore} &= (1, 0, 1, 1) \end{aligned}$$

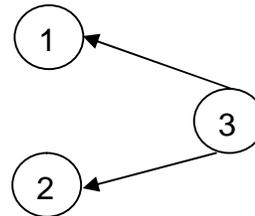
$$\text{Hscore}(2) = 0$$

**Para el nodo número 3.**

In Links de 3 (Vacío)

Out Links de 3 (1, 2)

Figura 29. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 1 HITS)



$$\text{Ascore}(3) = 0$$

$$\begin{aligned} \text{Hscore}(3) &= a(2) + a(1) \\ \text{Hscore}(3) &= 2 + 2 \\ \text{Hscore}(3) &= 4 \end{aligned}$$

$$\begin{aligned} \text{VAscore} &= (2, 2, 0, 1) \\ \text{VHscore} &= (1, 0, 4, 1) \end{aligned}$$

**Para el nodo número 4.**

In Links de 4 (Vacío)

Out Links de 4 (1)

Figura 30. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 1 HITS)



$$\text{Ascore}(4) = 0$$

$$\begin{aligned} \text{Hscore}(4) &= a(1) \\ \text{Hscore}(4) &= 2 \end{aligned}$$

VAscore = (2, 2, 0, 0)  
 VHscore = (1, 0, 4, 2)

$$Authority = \frac{VAsore}{Aacum}$$

$$Authority = \frac{2}{4}, \frac{2}{4}, \frac{0}{4}, \frac{0}{4}$$

Authority = (0.5, 0.5, 0, 0)

$$Hub = \frac{VHscore}{Hacum}$$

$$Authority = \frac{1}{7}, \frac{0}{7}, \frac{4}{7}, \frac{2}{7}$$

Authority = (0.1428, 0, 0.57, 0.2857)

Las iteraciones pueden continuar, generalmente este algoritmo en menos de 25 iteraciones converge.

**Traducción:**

- Id 1 = nodo número 1;
- Id 2 = nodo número 4;
- Id 3 = nodo número 2;
- Id 4 = nodo número 3.

Esto se debe a que en el momento de colocar los datos en el grafo se hace de la manera como van apareciendo.

**Ejemplo 2 HITS.**

Nodos = 6

Iteración = 1

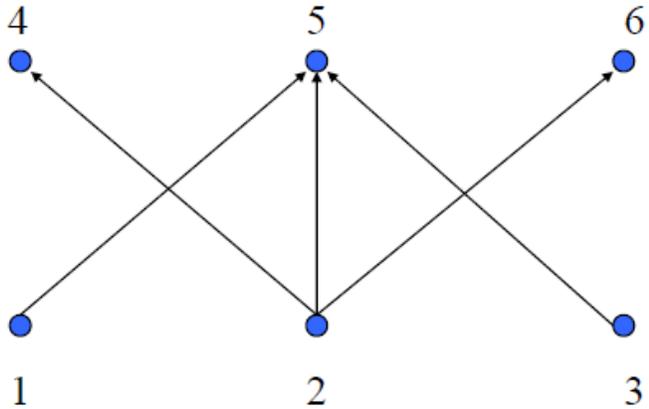
VAscore = (1, 1, 1, 1, 1, 1)

VHscore = (1, 1, 1, 1, 1, 1)

Hacum = 1 + 0 + 5 + 0 + 0 + 3 = 9

Aacum = 0 + 3 + 0 + 5 + 5 + 0 = 13

Figura 31. Grafo No. 2 para el cálculo de Hits (Ejemplo 2)



Fuente: Ortigosa Hernández, Jonathan. Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. 2008. pág. 10. [46]

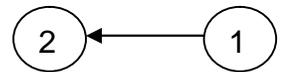
**Primera iteración.**

**Para el nodo número 1.**

In Links de 1 (vacío)

Out Links de 1 (2)

Figura 32. Grafos de la primera iteración para el nodo No. 1 (Ejemplo 2 HITS)



$Ascore(1) = 0$

$Hscore(1) = a(2)$   
 $Hscore(1) = 1$

$VAscore = (0, 1, 1, 1, 1, 1)$

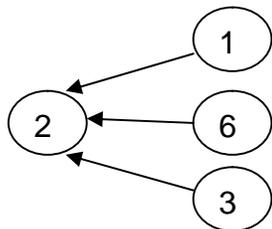
$VHscore = (1, 1, 1, 1, 1, 1)$

**Para el nodo número 2.**

In Links de 2 (1, 3, 6)

Out Links de 2 (vacío)

Figura 33. Grafos de la primera iteración para el nodo No. 2 (Ejemplo 2 HITS)



$$\begin{aligned} \text{Ascore}(2) &= h(1) + h(3) + h(6) \\ \text{Ascore}(2) &= 1 + 1 + 1 \\ \text{Ascore}(2) &= 3 \end{aligned}$$

$$\text{Hscore}(2) = 0$$

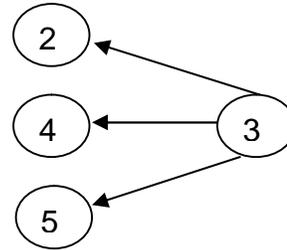
$$\begin{aligned} \text{VAscore} &= (0, 3, 1, 1, 1, 1) \\ \text{VHscore} &= (1, 0, 1, 1, 1, 1) \end{aligned}$$

**Para el nodo número 3.**

In Links de 3 (vacío)

Out Links de 3 (2, 4, 5)

Figura 34. Grafos de la primera iteración para el nodo No. 3 (Ejemplo 2 HITS)



$$\text{Ascore}(3) = 0$$

$$\begin{aligned} \text{Hscore}(2) &= a(2) + a(4) + a(5) \\ \text{Hscore}(3) &= 3 + 1 + 1 \\ \text{Hscore}(3) &= 5 \end{aligned}$$

$$\begin{aligned} \text{VAscore} &= (0, 3, 0, 1, 1, 1) \\ \text{VHscore} &= (1, 0, 5, 1, 1, 1) \end{aligned}$$

**Para el nodo número 4.**

In Links de 4 (3)

Out Links de 4 (vacío)

Figura 35. Grafos de la primera iteración para el nodo No. 4 (Ejemplo 2 HITS)



$$\begin{aligned} \text{Ascore}(4) &= h(3) \\ \text{Ascore}(4) &= 5 \end{aligned}$$

$$\text{Hscore}(2) = 0$$

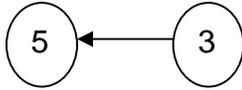
$$\begin{aligned} \text{VAscore} &= (0, 3, 0, 5, 1, 1) \\ \text{VHscore} &= (1, 0, 5, 0, 1, 1) \end{aligned}$$

**Para el nodo número 5.**

In Links de 5 (3)

Out Links de 5 (vacío)

Figura 36. Grafos de la primera iteración para el nodo No. 5 (Ejemplo 2 HITS)



$$Ascore(5) = h(3)$$

$$Hscore(5) = 0$$

$$Ascore(5) = 5$$

$$VAscore = (0, 3, 0, 5, 5, 1)$$

$$VHscore = (1, 0, 5, 0, 0, 1)$$

**Para el nodo número 6.**

In Links de 6 (vacío)

Out Links de 6 (2)

Figura 37. Grafos de la primera iteración para el nodo No. 6 (Ejemplo 2 HITS)



$$Ascore(6) = 0$$

$$Hscore(6) = a(2)$$

$$Hscore(6) = 3$$

$$VAscore = (0, 3, 0, 5, 5, 0)$$

$$VHscore = (1, 0, 5, 0, 0, 3)$$

$$Authority = \frac{VAsore}{Aacum}$$

$$Authority = \frac{0}{13}, \frac{3}{13}, \frac{0}{13}, \frac{5}{13}, \frac{5}{13}, \frac{0}{13}$$

$$Authority = (0, 0.2307, 0, 0.3846, 0.3846, 0)$$

$$Hub = \frac{VHsore}{Hacum}$$

$$Authority = \frac{1}{9}, \frac{0}{9}, \frac{5}{9}, \frac{0}{9}, \frac{0}{9}, \frac{3}{9}$$

$$Authority = (0.111, 0, 0.555, 0, 0, 0.333)$$

Las iteraciones pueden continuar, generalmente este algoritmo en menos de 25 iteraciones converge.

**Traducción:**

- ld 1 = nodo número 1;
- ld 2 = nodo número 5;
- ld 3 = nodo número 2;
- ld 4 = nodo número 4;
- ld 5 = nodo número 6;
- ld 6 = nodo número 3.

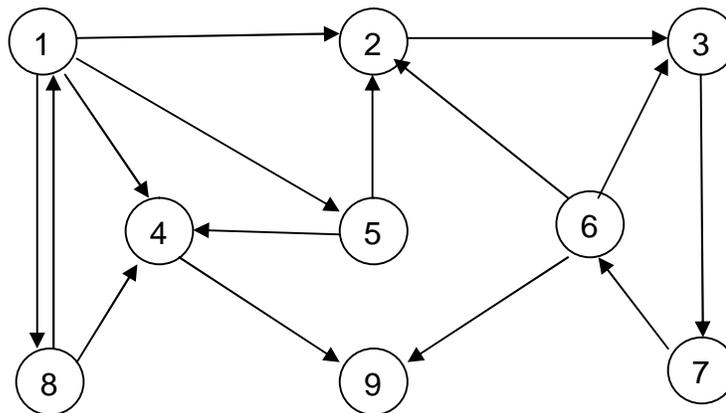
Esto se debe a que en el momento de colocar los datos en el grafo se hace de la manera como van apareciendo.

**3.3.3 Web Linker:** Este algoritmo busca dar un valor a cada nodo teniendo en cuenta la cantidad de enlaces de cada uno, independientemente de que sean entrantes o salientes. En pocas palabras busca un porcentaje de cada uno de los nodos encontrados a partir de los enlaces entrantes y salientes del grafo.

➤ **Ejemplos Web Linker.**

**Ejemplo 1 Web Linker.**

Figura 38. Grafo No. 1 para el cálculo del Web Linker (Ejemplo 1)



**Cantidad de Nodos = 9**  
**Cantidad de Enlaces = 15**

Para lo cual se genera una matriz donde se determinen las entradas y salidas de los nodos, identificando con un 1 cuando se presenta la existencia de un link, independientemente de que este será de entrada o de salida.

Las filas representan las entradas de cada nodo y las columnas las salidas.

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

**Para el nodo número. 1**

Nodo 1 = In Links + Out Links  
 Nodo 1 = 1 + 4  
 Nodo 1 = 5

**Para el nodo número. 2**

Nodo 2 = In Links + Out Links  
 Nodo 2 = 3 + 1  
 Nodo 2 = 4

**Para el nodo número. 3**

Nodo 3 = In Links + Out Links  
 Nodo 3 = 2 + 1  
 Nodo 3 = 3

**Para el nodo número. 4**

Nodo 4 = In Links + Out Links  
 Nodo 4 = 3 + 1  
 Nodo 4 = 4

**Para el nodo número. 5**

Nodo 5 = In Links + Out Links  
 Nodo 5 = 1 + 2  
 Nodo 5 = 3

**Para el nodo número. 6**

Nodo 6 = In Links + Out Links  
 Nodo 6 = 1 + 3  
 Nodo 6 = 4

**Para el nodo número. 7**

Nodo 7 = In Links + Out Links  
 Nodo 7 = 1 + 1

Nodo 7 = 2

**Para el nodo número. 8**

Nodo 8 = In Links + Out Links

Nodo 8 = 1 + 2

Nodo 8 = 3

**Para el nodo número. 9**

Nodo 9 = In Links + Out Links

Nodo 9 = 2 + 0

Nodo 9 = 2

Se busca el nodo que tenga la mayor cantidad de links, para este caso el nodo número 1, al cual se le coloca una calificación de 10, teniendo en cuenta que la calificación más alta es de 10 y la más baja es de 0.

Mayor = Nodo 1

**Para el nodo número. 1**

*Score 1 = Mayor*

*Score 1 = 10*

**Para el nodo número. 2**

$$Score 2 = \frac{Nodo 2 \times 10}{Nodo 1}$$

$$Score 2 = \frac{4 \times 10}{5}$$

*Score 2 = 8*

**Para el nodo número. 3**

$$Score 3 = \frac{Nodo 3 \times 10}{Nodo 1}$$

$$Score 3 = \frac{3 \times 10}{5}$$

*Score 3 = 6*

**Para el nodo número. 4**

$$\text{Score 4} = \frac{\text{Nodo 4} \times 10}{\text{Nodo 1}}$$

$$\text{Score 4} = \frac{4 \times 10}{5}$$

$$\text{Score 4} = 8$$

**Para el nodo número. 5**

$$\text{Score 5} = \frac{\text{Nodo 5} \times 10}{\text{Nodo 1}}$$

$$\text{Score 5} = \frac{3 \times 10}{5}$$

$$\text{Score 5} = 6$$

**Para el nodo número. 6**

$$\text{Score 6} = \frac{\text{Nodo 6} \times 10}{\text{Nodo 1}}$$

$$\text{Score 6} = \frac{4 \times 10}{5}$$

$$\text{Score 6} = 8$$

**Para el nodo número. 7**

$$\text{Score 7} = \frac{\text{Nodo 7} \times 10}{\text{Nodo 1}}$$

$$\text{Score 7} = \frac{2 \times 10}{5}$$

$$\text{Score 7} = 4$$

**Para el nodo número. 8**

$$\text{Score 8} = \frac{\text{Nodo 8} \times 10}{\text{Nodo 1}}$$

$$\text{Score } 8 = \frac{3 \times 10}{5}$$

$$\text{Score } 8 = 6$$

**Para el nodo número. 9**

$$\text{Score } 9 = \frac{\text{Nodo } 9 \times 10}{\text{Nodo } 1}$$

$$\text{Score } 9 = \frac{2 \times 10}{5}$$

$$\text{Score } 9 = 4$$

**Ejemplo 2 Web Linker.**

Cálculo realizado para el grafo de la **Figura 21**.

**Cantidad de Nodos = 4**

**Cantidad de Enlaces = 4**

$$M = \begin{vmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{vmatrix}$$

**Para el nodo número. 1**

Nodo 1 = In Links + Out Links

Nodo 1 = 2 + 1

Nodo 1 = 3

**Para el nodo número. 2**

Nodo 2 = In Links + Out Links

Nodo 2 = 0 + 1

Nodo 2 = 1

**Para el nodo número. 3**

Nodo 3 = In Links + Out Links

Nodo 3 = 0 + 1

Nodo 3 = 1

**Para el nodo número. 4**

Nodo 4 = In Links + Out Links

Nodo 4 = 2 + 0

Nodo 4 = 2

Mayor = Nodo 1

**Para el nodo número. 1**

*Score 1 = Mayor*

*Score 1 = 10*

**Para el nodo número. 2**

$$Score 2 = \frac{Nodo 2 \times 10}{Nodo 1}$$

$$Score 2 = \frac{2 \times 10}{3}$$

$$Score 2 = 6.66 \quad 7$$

**Para el nodo número. 3**

$$Score 3 = \frac{Nodo 3 \times 10}{Nodo 1}$$

$$Score 3 = \frac{1 \times 10}{3}$$

$$Score 3 = 3.33 \quad 3$$

**Para el nodo número. 4**

$$Score 4 = \frac{Nodo 4 \times 10}{Nodo 1}$$

$$Score 4 = \frac{2 \times 10}{3}$$

$$Score 4 = 6.66 \quad 7$$

## 4. ANÁLISIS DE HERRAMIENTAS

### 4.1 POLARIS VERSIÓN 1.0

**4.1.1 Descripción preliminar:** Polaris es una herramienta de minería Web de uso desarrollada por estudiantes pertenecientes al Grupo de Investigación GRIAS (Grupo de Investigación Aplicado a Sistemas) dirigido por Ricardo Timaran Pereira, Ph.D. del Programa de Ingeniería de Sistemas de la Universidad de Nariño de la ciudad de Pasto.

Polaris es un software que realiza minería Web de uso a partir de un log de acceso de servidores Web: Apache y IIS Server.

Polaris tiene una completa variedad de algoritmos de asociación y clasificación, y un algoritmo de minería de uso llamado HPG (Gramática Probabilística de Hipertexto), también tiene siete formatos de visualización que ayuda a entender todo el proceso de minería.

**4.1.2 Requisitos del sistema:** Para el correcto funcionamiento de la herramienta se necesita tener instalado el siguiente software:

- Postgres 8.2 o superior.
- La Máquina Virtual de Java jre o superior.
- Componente de Java3D 1.5.0 o superior.
- Usuario del SGBD Postgres que tenga permiso de creación de base de datos.
- Instalador de Polaris v1.0

Para poder obtener el software necesario para el funcionamiento de Polaris Versión 1.0 se lo puede descargar de las siguientes URL.

- <http://www.postgresql.org/download>
- <http://www.java.com/en/download/index.jsp>

**4.1.3 Instalación:** Para esto es necesario el archivo de instalación Setup Polaris.exe y seguir los siguientes pasos:

- Hacer doble clic en el archivo.
- Seguir las instrucciones del instalador.
- Aceptar la licencia GPL (General Public License).

- Hacer clic en install para comenzar la instalación, esperar a que instale todos los archivos necesarios.
- Ejecutar la aplicación.

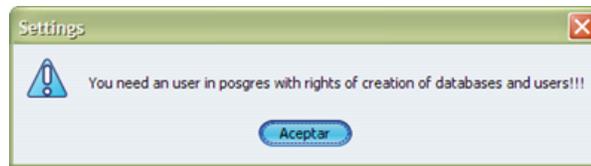
El instalador de Polaris Versión 1.0 puede ser descargado desde la siguiente dirección Web:

<http://polaris.berlios.de>

**4.1.4 Manejo de la herramienta:** Antes de iniciar a trabajar con le herramienta software Polaris Versión 1.0 es necesario realizar su respectiva configuración. Para esto se debe seguir los siguientes pasos:

Se presiona el botón de la barra de herramientas que indica la configuración y aparece un mensaje de advertencia:

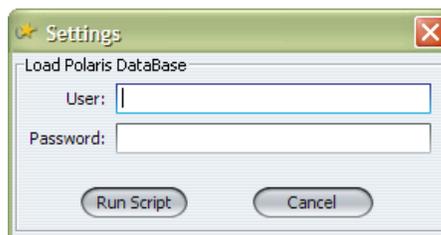
Figura 39. Mensaje de advertencia para configuración de la base de datos.



**Fuente: Angulo Diana, Daza Johana y Zuleta Alejandra** Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007. [2]

Una vez se lee el mensaje y se hace clic en el botón aceptar, aparece la ventana:

Figura 40. Formulario de configuración de la base de datos.



**Fuente: Angulo Diana, Daza Johana y Zuleta Alejandra** Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007. [2]

Se escribe el usuario con derechos de creación de bases de datos y usuario, y su contraseña, inmediatamente se hace clic en **Run Script**, si no hay ningún error aparece un mensaje de éxito, en caso contrario un mensaje de error.

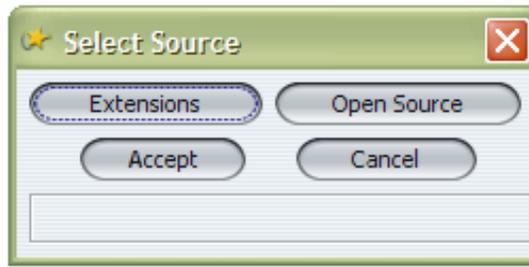
Si todo salió bien, se puede utilizar el software perfectamente.

Una vez abierta la aplicación, un Nuevo proyecto es abierto automáticamente, por tanto podemos realizar un algoritmo de minería de uso.

Para armar el algoritmo se debe tener en cuenta que este siempre empieza por un WEB DATA SOURCE, por tanto, ninguna otra herramienta se habilita hasta que un elemento de este sea arrastrado al área de trabajo. Posteriormente se debe configurar el Web server.

Lo primero que se debe hacer es dar clic derecho en el nodo Web Server, y seleccionar la opción Settings entonces aparecerá una pequeña ventana:

Figura 41. Configuración del nodo Web Server.

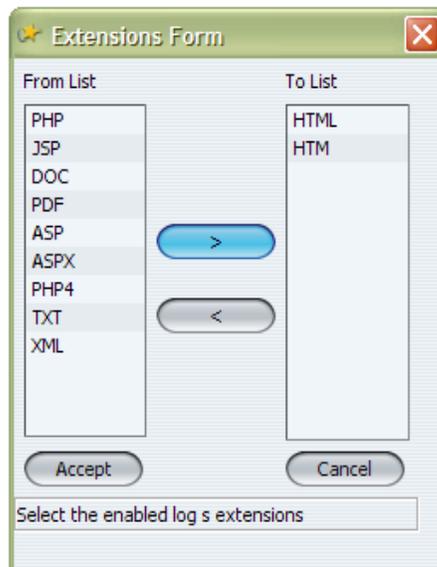


**Fuente:** Angulo Diana, Daza Johana y Zuleta Alejandra Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007. [2]

Hay dos opciones de configuración:

**Extensions:** Se establece las extensiones validas para realizar el análisis.

Figura 42. Formulario de extensiones validadas para realizar el análisis.



**Fuente:** Angulo Diana, Daza Johana y Zuleta Alejandra Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007. [2]

**Open Source:** Se escoge el archivo Web server log que se va a analizar. Estos archivos tienen en su mayoría extensión .log,

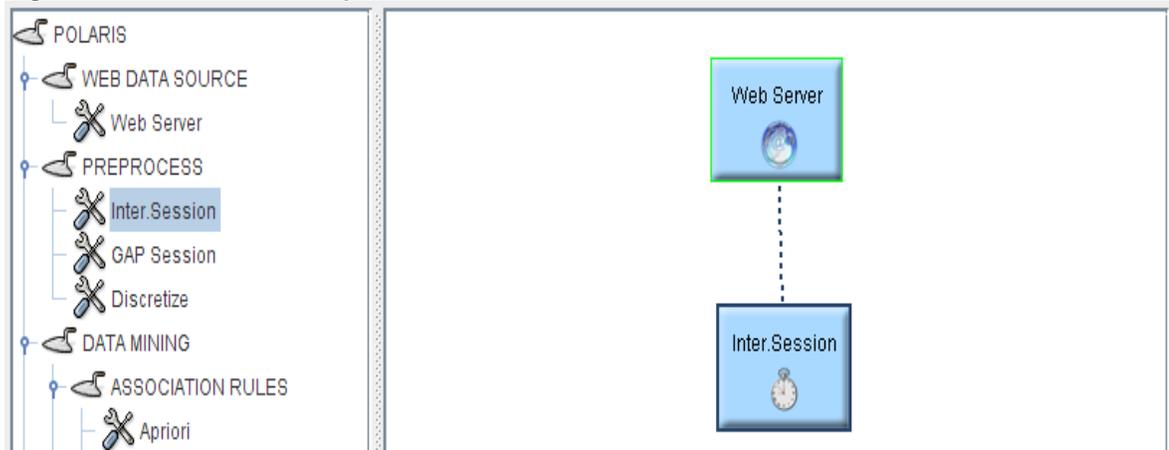
Una vez configurado el nodo se hace clic derecho en el nodo y se selecciona la opción **Run** Al terminar la ejecución el nodo cambia de estado y aparece un mensaje en la barra de estado con el resultado de la ejecución.

Posteriormente arrastramos el elemento **Inter.Session**

Una vez arrastrado el Inter.Session, existen dos opciones para hacer en el momento:

- Se puede configurar y ejecutar el nodo
- Se puede hacer la conexión con el Web Server

Figura 43. Área de trabajo de POLARIS.



**Fuente:** Angulo Diana, Daza Johana y Zuleta Alejandra Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007. [2]

Posteriormente se debe realizarla respectiva configuración del nodo y así poder ejecutarlo.

#### 4.1.5 Ventajas:

- Es una herramienta libre.
- Requerimientos mínimos de software y hardware.
- Genera reportes gráficos.
- Trabaja sobre sistemas operativos Windows y Linux.
- Posee una interfaz gráfica amigable.

#### 4.1.6 Desventajas:

- Únicamente soporta Minería Web de Uso.

## 4.2 TOUCHGRAPH NAVIGATOR VERSIÓN 1.3.1.

**4.2.1 Descripción Preliminar:** TouchGraph es una herramienta que muestra la conectividad entre sitios Web de una forma visual bastante llamativa. Lo que hace esta herramienta es tomar la información de la base de datos de Google de sitios relacionados, y mostrar un mapa interactivo de sitios Web interconectados o términos de búsqueda interconectados.

El TouchGraph Navigator permite cargar, visualizar y analizar datos sobre las entidades, sus relaciones y atributos asociados. También ofrece navegadores para Amazon y Facebook, además de ofrecer trabajos personalizados a compañías por encargo. TouchGraph representa gráficamente la conectividad entre sitios Web a partir de los “related sites” detectados por Google.

**4.2.2 Requisitos del Sistema:** La versión de escritorio de la herramienta software TouchGraph Navigator puede ejecutarse en cualquier máquina que soporte Java como Windows, Mac o Linux.

La versión basada en Web se puede ejecutar en cualquier navegador, incluido Internet Explorer, Firefox y Safari.

TouchGraph en las versiones de escritorio y Web son basadas en Java 1.5, por lo cual se requiere que esta versión o versiones superiores se encuentren instaladas en el ordenador cliente. La última versión de Java se puede descargar desde aquí:

<http://www.java.com/en/download/index.jsp>

**4.2.3 Instalación:** Para descargar la copia de evaluación de la versión TouchGraph Navigator Versión 1.3.1 debe ir a la página:

<http://www.touchgraph.com/navigator.html>

El proceso de instalación es sencillo, basta con hacer clic en el botón Next o Siguiente, únicamente se debe tener precaución al momento de ingresar el nombre, la organización y el serial, los cuales son proporcionados cuando se realiza la descarga del software; dicha información es enviada al correo y tendrá una vigencia de 30 días como prueba.

Al final de este proceso aparece la ventana de instalación exitosa, la cual nos indica que toda la configuración de instalación se ha realizado satisfactoriamente y el TouchGraph Navigator está listo para ser utilizado.

**4.2.4 Manejo de la Herramienta:** El trabajo con la herramienta es sencillo, sin embargo el usuario puede ir a la opción Help, Ayuda, del menú principal, donde se desplegará el manual de uso de TouchGraph Navigator, con toda la información que se requiere para el adecuado uso del producto.

Figura 44. Logo TouchGraph NAVIGATOR



**Fuente:** TouchGraph, LLC. TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

Antes de iniciar a trabajar con TouchGraph debe tener en claro lo siguientes conceptos, los cuales son los tipos de datos que maneja la herramienta:

**Entidad:** Una entidad es un objeto con un número de identificación exclusivo. Las entidades pueden ser cualquier tipo de objetos, incluidas las personas, documentos, temas, o productos.

**Relación:** Una relación es una relación entre dos entidades. Las relaciones pueden ser la amistad entre dos personas, hipervínculos entre un par de páginas Web, o la conexión entre un documento y un tema.

**Atributos:** Un atributo es una propiedad de una entidad o una relación. Por ejemplo, una persona puede tener una edad, altura, peso, y la dirección. Una relación de amistad puede tener una fecha que corresponde a la primera reunión, y una fuerza.

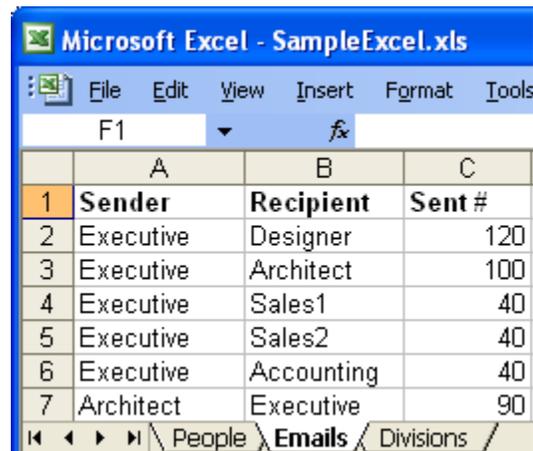
**Tipo de entidad:** Un tipo de entidad es la clase de entidades. Por ejemplo, Bob, Bill y Ben son tres entidades, con "Persona" como su tipo de entidad. IBM, Microsoft, Apple son tres entidades, con 'Empresa' como su tipo de entidad.

**Tipo de Relación:** Una relación de tipo es una clase de relaciones. No puede haber más de una relación de diferentes tipos entre las mismas dos Entidades. Por ejemplo, dos autores pueden tener una relación de amistad, y que también podría haber co-autor de un artículo en conjunto. "Amistad" y "coautoría" son dos tipos de relación.

TouchGraph en su versión libre, únicamente permite visualizar los ejemplos que incluye la herramienta, lastimosamente no permite crear una nueva aplicación para poder clarificar su funcionamiento. En la versión demo on-line nos permite ingresar nuevas páginas, palabras o frases de la cual se desee obtener la estructura Web y su respectiva visualización.



Figura 46. Archivo de Excel, ejemplo TouchGraph NAVIGATOR



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - SampleExcel.xls". The menu bar includes File, Edit, View, Insert, Format, and Tools. The active cell is F1. The spreadsheet contains a table with the following data:

|   | A             | B                | C             |
|---|---------------|------------------|---------------|
| 1 | <b>Sender</b> | <b>Recipient</b> | <b>Sent #</b> |
| 2 | Executive     | Designer         | 120           |
| 3 | Executive     | Architect        | 100           |
| 4 | Executive     | Sales1           | 40            |
| 5 | Executive     | Sales2           | 40            |
| 6 | Executive     | Accounting       | 40            |
| 7 | Architect     | Executive        | 90            |

At the bottom of the window, there are navigation arrows and a tab labeled "Emails" which is currently selected, with other tabs labeled "People" and "Divisions" visible.

**Fuente: TouchGraph, LLC.** TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

Los datos pueden ser almacenados en múltiples hojas de Excel. El Asistente para Excel permite la elección de las hojas a usar e ignora las que no se debe tener en cuenta, ya que no todas las hojas contienen datos utilizables.

Un formato adecuado para una hoja Excel debe contener únicamente una tabla de datos para una entidad o una relación.

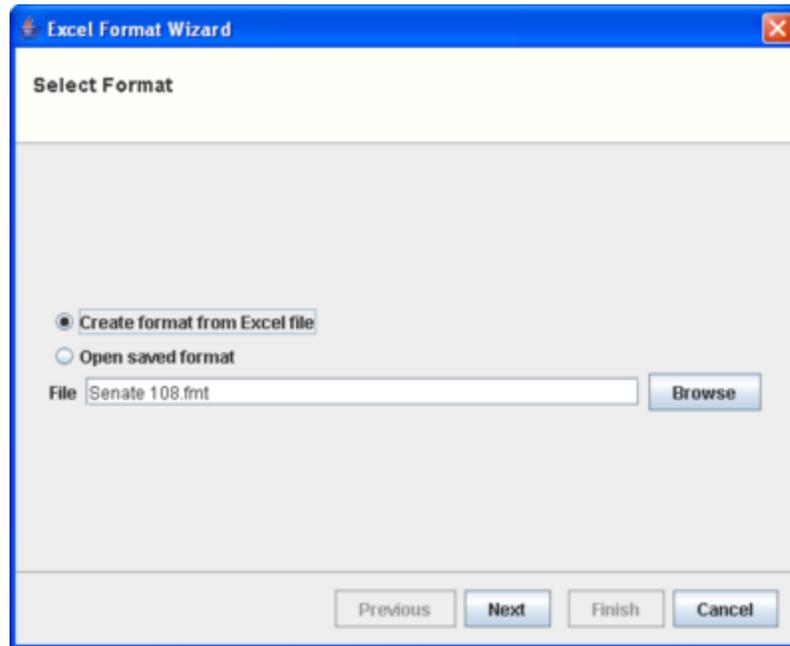
La fila superior de una hoja de Excel debe contener los nombres de columnas.

Las columnas deben estar alineadas a la izquierda, si hay 4 columnas, las columnas A, B, C, D deben tener los datos, sin columnas en blanco en el medio.

Múltiples valores pueden ser colocados en una sola celda usando un delimitador. Por ejemplo, si una columna contiene los autores, una célula puede tener el valor "Vonnegut, Dickens, Twain".

Seleccionar el formato de página, es el primer paso del asistente de Excel. Este permite usar un archivo ya guardado o crear uno nuevo si se desea.

Figura 47. Crear o abrir un archivo de Excel.



**Fuente: TouchGraph, LLC.** TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

Presionamos en Next para continuar con el asistente de carga de Excel.

La siguiente página del asistente nos muestra la ficha “ítem”, la cual nos permite definir los atributos asociados, el tema y categorías, de la hoja de Excel que se está cargando.

Figura 48. Ficha "Item".

**Item Sheet**

Item Name:

Select Sheet: **Skip**, 1 (senators), 2 (bills), 3 (info)

Category Delim:

Use the Item Sheet to load a list of Items and associated values. One column must be selected as the Item Id (Id), and multiple columns can be selected as Categories (Cat) and Attributes (Attr). Categories are used for grouping items and will have nodes created in the graph. Attributes contain Item data like text for the Label, number for Size/Filtering, and URLs for links and Item Images.

| Excel Column | Alias | Id                               | Cat                              | Attr                             | Type  | X                     |
|--------------|-------|----------------------------------|----------------------------------|----------------------------------|-------|-----------------------|
| ID #         | ID    | <input checked="" type="radio"/> | <input type="radio"/>            | <input type="radio"/>            |       | <input type="radio"/> |
| Name         |       | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | Text  | <input type="radio"/> |
| Link         |       | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | URI   | <input type="radio"/> |
| Image        |       | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | Image | <input type="radio"/> |
| State Name   | State | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/>            |       | <input type="radio"/> |
| Party        |       | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/>            |       | <input type="radio"/> |

Previous Next Finish Cancel

**Fuente: TouchGraph, LLC.** TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

El Nombre del artículo de un campo permite elegir el nombre del tipo de artículo. Configurar el Nombre del artículo personaliza el aspecto de la partida el cuadro y el gráfico. Se ajusta los componentes que luego mostrar el nombre del tema, tales como "senador" en lugar de "tema".

Seleccione la ficha de lista muestra los nombres de las hojas de Excel en el archivo de Excel, y permite recoger los datos de una hoja.

Use la tabla para especificar el tipo de los datos en el archivo de Excel, los cuales se convertirán en entidades. Una vez terminado se puede presionar Next para avanzar.

El Cuadro de campo es la siguiente ventana del asistente, donde se muestra una lista de las columnas de Excel.

Figura 49. Cuadro de campo



**Fuente:** TouchGraph, LLC. TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

Los campos de selección de la tabla tienen las siguientes columnas:

**Columna de Excel:** Muestra el nombre de la columna de Excel de la hoja Excel.

Si se repiten los nombres de columna en la hoja de Excel se mostrará como una sola columna.

**Alias:** Hace referencia al nombre que debe aparecer dentro de los cuadros y ajustes del Navegador.

**Id:** Una y sólo una columna debe ser seleccionada como el tema Id. El tema de identificación se utiliza para conectar el tema a las categorías, y para crear las relaciones.

**Cat:** Múltiples columnas de Excel pueden ser seleccionados como categorías.

Para un determinado tipo de categoría, la categoría de múltiples valores pueden estar asociados con el mismo tema de identificación. Por ejemplo, un único "autor" Tipo Categoría puede contener múltiples nombres de los autores para el mismo libro.

**Attr:** Múltiples columnas de Excel se pueden seleccionar como atributos.

Para un determinado atributo, un valor sólo puede ser asociado con el mismo tema ID.

**Tipo:** Para los atributos, es necesario especificar el tipo de valor que el atributo contiene. Entre estos tipos tenemos:

- **Texto:** Esto puede ser una cadena corta o una larga. Los atributos de texto se ordenan alfabéticamente en los cuadros.
- **Entero, decimal:** Para definir número enteros o decimillas. Estos atributos se ordenan numéricamente en las tablas.
- **Boolean:** Muestra un valor de Verdadero o Falso en los cuadros.
- **Fecha:** Formatea el valor como una fecha, y permite la clasificación cronológica de las tablas.
- **URI:** Permite crear un enlace a una página Web remota en la ventana de información, o la vinculación a un archivo local.
- **Imagen:** Un URI apunta a una imagen en un servidor remoto o un archivo local.

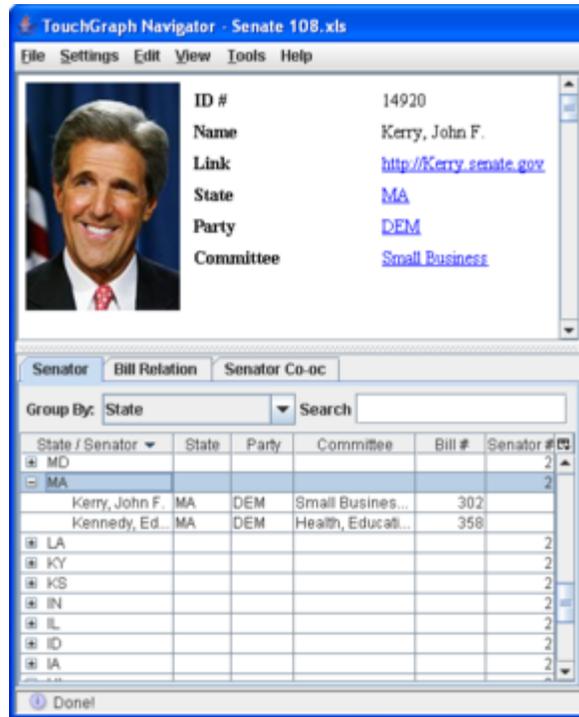
**X:** La selección de 'X' ignorará la columna de la hoja de Excel.

Una vez definidas las relaciones el asistente termina y realiza la carga de la información que se encuentra en la base de datos creada en la hoja de Excel.

- **Transformar datos a formato .xml:** Se aplica un filtro **awk** al fichero de texto separado por tabulaciones y se genera un fichero en formato **xml**.
- **Generar el gráfico:** Se procesa el fichero **.xml** con el programa **TouchGraph Navigator**.

En la interfaz nos indicará al lado izquierdo los cuadros que muestran la entidad y relación de datos.

Figura 50. Ejemplo de la interfaz, con el programa TouchGraph Navigator, para el fichero .xml.

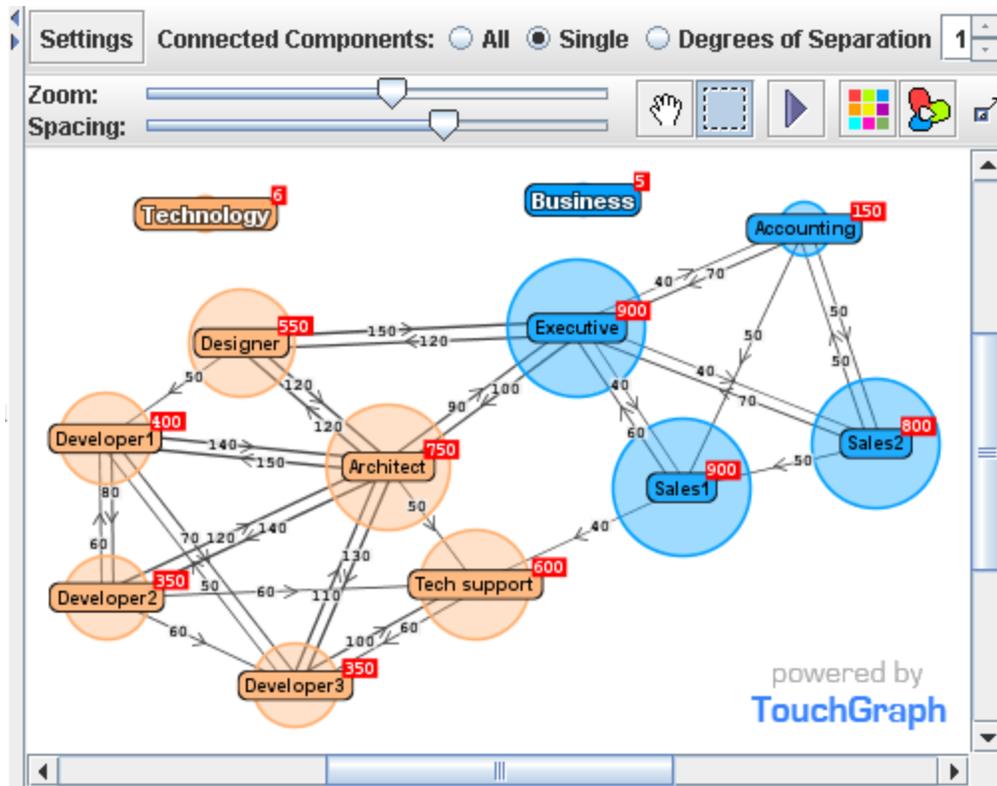


**Fuente: TouchGraph, LLC.** TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

El panel de información se encuentra en la parte superior izquierda de la ventana del Navegador. Cuando un elemento es seleccionado, el panel de información muestra imágenes asociadas, los hipervínculos, las características y las descripciones.

El gráfico se ubica en el lado derecho del navegador, así como otras barras de herramientas y controles para interactuar con el gráfico.

Figura 51. Ejemplo del gráfico del grafo que se forma con la información suministrada.



Fuente: TouchGraph, LLC. TouchGraph. [En línea] 2007. [Citado el: 10 de Diciembre de 2008.] <http://www.touchgraph.com/navigator.html>. [55]

El gráfico ofrece una gama de Interacción Gráfica como hacer clic, arrastrar, y la selección.

El gráfico se compone de nodos y aristas, que tienen un número de componentes visuales para mostrar los datos y la representación de la interacción.

#### 4.2.5 Ventajas:

- En poco tiempo se puede **obtener una información global, actual y resumida** de la que se podrá seleccionar rápidamente lo que realmente sea interesante para profundizar.
- Permite capturar los datos desde una hoja de cálculo **Excel** además de la opción del fichero **XML**. Esto facilita la captura de datos desde herramientas ofimáticas y no precisa una transformación de los datos a un formato **XML**.
- Amplía las posibilidades de visualización y mejora la estética.

- TouchGraph es una herramienta software que puede ejecutarse bajo el sistema operativo Windows o Linux.

#### **4.2.6 Desventajas:**

- Una aplicación que se preocupa de buscar datos y a la vez generar una gráfica de ellos, es bastante más lenta que una interfaz de solo texto.
- La versión liberada de TouchGraph no permite la creación de nuevas redes, únicamente permite trabajar con los ejemplos que ya vienen precargados.
- Obtener la versión licenciada de TouchGraph es muy costoso.

### **4.3 SURF PATTERN VISUAL ANALYZER.**

**4.3.1 Descripción preliminar:** La herramienta Surf Pattern Visual Analyzer (SPVA) es usada para el análisis y visualización de búsquedas y navegación en la Web, permite observar los patrones de visualización de enlaces que los surfistas de las páginas Web usan más comúnmente.

Surf Pattern Visual Analyzer es útil porque los internautas pueden localizar una página de un sitio Web escribiendo la URL, seleccionando un enlace, con la búsqueda de frases o de palabras claves de los motores de búsqueda (como Google, MSN, Yahoo, AskJeeves, etc.).

La comprensión de este patrón de navegación por el sitio (o comportamiento del surfista) es muy importante en la optimización de contenido Web. El objetivo del Surf Pattern Visual Analyzer es contribuir a mejorar los contenidos y enlaces entre las páginas, ofreciendo información acerca de cómo era una página y donde se encuentra. [52]

**4.3.2 Requisitos del sistema:** Surf Pattern Visual Analyzer es Multi-plataforma: Windows, MAC OS X, Solaris, Linux, HPUX, Irix, etc.

La versión Surf Pattern Visual Analyzer 2.5, se descargo gratuitamente para la plataforma Mac OS X 10.1 o superiores, ya que en las otras plataformas no se puede descargar el software de manera libre.

Se debe tener en cuenta antes de instalar el software de tener la versión Java que contenga el paquete AWT / GUI. Este trabajará con el actual sistema de la interfaz de Mac, pero con limitada widgets (una pequeña aplicación a programa). Versión alternativa basada en Eclipse SWT / GUI.

**4.3.3 Instalación:** Para descargar la copia gratuita de Surf Pattern Visual Analyzer versión 2.5 debe ir a la página:

<http://www.macupdate.com/info.php/id/30143/surf-pattern-analyzer>

Para descargar la copia comercial de Surf Pattern Visual Analyzer versión 2.5, y cualquier otra de sus versiones debe ir a la página:

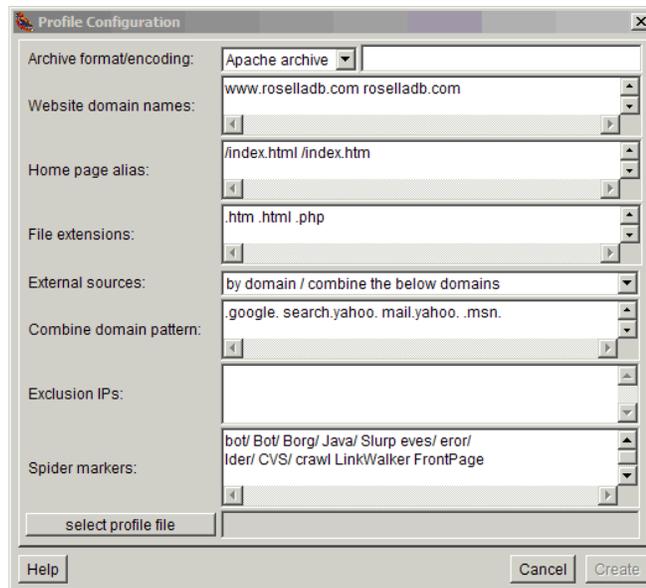
<http://mac.softpedia.com/get/HTML-Tools/Surf-Pattern-Analyzer.shtml>

El proceso de instalación para plataforma MAC es sencillo, basta con descomprimir el archivo SurfAnalyzer2p5\_MAC.zip en el cual se encuentra el archivo ejecutable SurfPatternAnalyzer.jar, el cual al ser ejecutado coloca en marcha el programa.

**4.3.4 Manejo de la herramienta:** Surf Pattern Visual Analyzer se basa en "Profile" (perfiles).

➤ **Creación de perfiles:** El primer paso es crear un perfil vacío. Desde el " Profile" del menú, seleccione "Open Profile" para obtener el siguiente cuadro de diálogo:

Figura 52. Formulario de creación de perfiles.



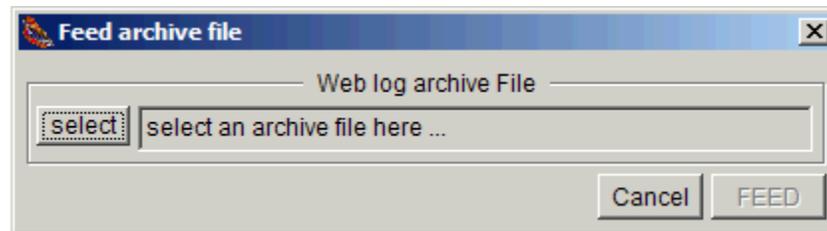
**Fuente: Rosella.** Predictive Knowledge & Data Mining. *Data Mining & Database Analytics*. [En línea] 2005. [Citado el: 20 de Enero de 2009.] [52]

Posteriormente se debe especificar las configuraciones, seleccione un archivo para almacenar el perfil y pulse el botón "**Create**". Detalles de la especificación se

puede encontrar en la ayuda pulsando el botón "Help".

➤ **Archivos de alimentación:** Una vez que se crea un perfil, el siguiente paso es la alimentación de los archivos de registro Web para el perfil. SPVA recoge información de los archivos y el espacio de almacenamiento en los perfiles.

Figura 53. Alimentación de archivos para Surf Pattern Visual Analyzer versión 2.5.



**Fuente: Rosella.** Predictive Knowledge & Data Mining. *Data Mining & Database Analytics*. [En línea] 2005. [Citado el: 20 de Enero de 2009.] [52]

➤ **Análisis visual:** Utilice las barras de desplazamiento para ampliar o para mostrar más nodos de la siguiente manera:

- **Top scrollbars:** Para controlar la distancia entre dos nodos, en las columnas.
- **Left scrollbars:** Para observar todos los nodos existentes.

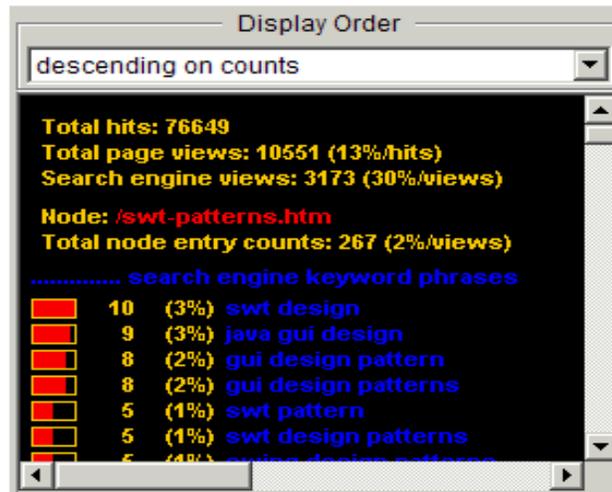
Figura 54. Ventana para el análisis visual de Surf Pattern Visual Analyzer versión 2.5



**Fuente: Rosella.** Predictive Knowledge & Data Mining. *Data Mining & Database Analytics*. [En línea] 2005. [Citado el: 20 de Enero de 2009.] [52]

**Detalles de nodo:** El panel de la derecha muestra los detalles del nodo seleccionado. Cuando se selecciona un nodo de la columna izquierda el panel muestra páginas de navegación salientes (**outgoing navigation**). Asimismo, cuando se selecciona un nodo de la columna, en el panel se muestran las páginas de navegación entrantes (**incoming navigation**).

Figura 55. Detalles del nodo de Surf Pattern Visual Analyzer versión 2.5



**Fuente: Rosella.** Predictive Knowledge & Data Mining. *Data Mining & Database Analytics*. [En línea] 2005. [Citado el: 20 de Enero de 2009.] [52]

Otros controles que pueden ser utilizados son los siguientes:

**Nodo Opciones (Node Options):** Selecciona las fuentes de navegación entre páginas Web, motor de búsqueda de palabras y frases claves.

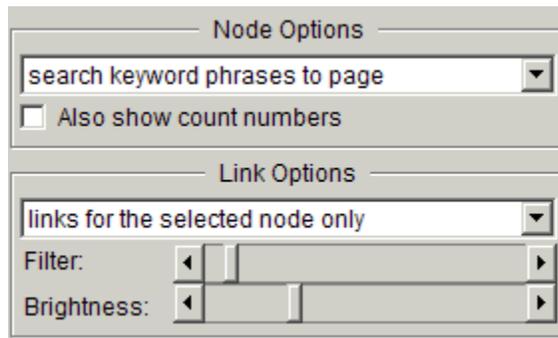
También se muestran la cantidad en números y porcentaje de las conexiones entre las páginas.

**Enlace Opciones (Link Options):** Indica los enlaces de todas las páginas o desde el nodo seleccionado solamente.

**Filtro (Filter):** Para filtrar los vínculos de navegación visualizada.

**Brillo (Brightness):** Se muestra los Enlaces con menor incidencia en colores más oscuros. Este filtro hace más brillante los arcos entre los nodos.

Figura 56. Opciones del nodo para Surf Pattern Visual Analyzer versión 2.5



**Fuente: Rosella.** Predictive Knowledge & Data Mining. *Data Mining & Database Analytics*. [En línea] 2005. [Citado el: 20 de Enero de 2009.] [52]

Con este tipo de visualización que nos brinda el software nos permite tomar decisiones sobre la creación de sitios Web, en los que nos indicaría que enlaces utilizan más los usuarios y para mejorar un sitio Web ya implementado.

- Analizar los vínculos entre las palabras claves, frases claves y páginas Web puede dar lugar a diseños de páginas que lleve los surfistas al contenido y páginas que se necesitan.
- Surf Pattern Analyzer (SPVA) puede utilizarse para identificar patrones de navegación fundamentales de los sitios Web.
- Identificación de la mayoría de las grandes rutas de navegación.
- Motor de búsqueda de palabras claves y frases de palabras clave, análisis de enlaces (Google, MSN, Yahoo).
- Complementa las funciones de su sitio de análisis, tales como las estadísticas de AWStats y Urchin
- Surf Pattern Analyzer es Multi-plataforma: Windows, MAC OS X, Solaris, Linux, HP-UX, Irix, etc.

#### 4.3.5 Desventajas:

- El software Surf Pattern Visual Analyzer maneja archivos Apache Log Format.
- La herramienta es comercial, por lo cual no se encuentra una versión libre.

## 4.4 SEO ADMINISTRATOR VERSIÓN 3.11

**4.4.1 Descripción preliminar:** SEO Administrator es una suite integrada que cubren todos los aspectos de optimización de sitio Web. Profesionales y los recién llegados por igual han utilizado con éxito SEO Administrator para mejorar sus sitios Web y han encontrado que es una herramienta simple e intuitiva de manejar.

A continuación se presenta un panorama general de la herramienta SEO Administrator y sus respectivos módulos.

**"Ranking Monitor"**. Esta herramienta utiliza palabras claves de un sitio Web para obtener la posición de la clasificación del sitio Web a partir de una amplia gama de motores de búsqueda. La lista de más de treinta motores de búsqueda internacionales incluye: Google, Yahoo, MSN, AltaVista, Lycos, HotBot, etc.

**"Link Popularity Checker"**. El número de enlaces entrantes a su sitio Web es un factor importante en el éxito del motor de búsqueda y promoción, esta herramienta analiza estos vínculos. Este paquete le permite crear una lista de enlaces a su Web y seguir su dinámica mediante la revelación de nuevos y obsoletos enlaces. El módulo también puede analizar los sitios Web de su competidor y la búsqueda de potenciales recursos de intercambio de enlaces.

**"Site Indexation Tool"**. Esta herramienta comprueba el sitio Web para la indexación de la página. Antes de obtener los resultados de la búsqueda a base de palabras clave, la página Web debe ser indexada por los motores de búsqueda. Seo Administrator revela las páginas de un sitio Web que han sido indexadas por los distintos motores de búsqueda.

**"Log Analyzer"**. Está diseñado para el análisis de un sitio Web de registro de archivos. Se proporciona estimaciones de la utilización de recursos, el número de visitantes, las consultas de búsqueda que se utilizaron para encontrar su sitio, sitio remitentes, los visitantes de los países y mucho más.

**"Page Rank Analyzer"**. Obtiene automáticamente el valor de Page Rank de Google para su lista de URL. El número de enlaces entrantes también se obtiene.

**"Keywords Suggestion Tool"**. Escoger el derecho de palabras clave y frases clave para orientar en sus páginas es muy importante para ayudar a la gente a encontrar su sitio Web. Esta herramienta utiliza una variedad de fuentes para sugerir palabras clave pertinentes para cada página.

**"HTML Analyzing Tool"**. Es un programa que analiza el contenido HTML de la página. El programa proporciona un informe completo sobre el peso y la densidad de palabras clave y frases clave en sus páginas Web. Así como un análisis preliminar de su propio sitio.

**"Google Data Centers Monitor"**. Este programa es una herramienta que le permite la consulta de todos los centros de datos de Google desde una interfaz amigable.

**"Snippets Viewer"**. Las descripciones automatizadas (fragmentos) de sus recursos mostrados por los motores de búsqueda también son muy importantes. Una atractiva descripción hará que los visitantes elijan el vínculo entre todas las demás propuestas por el motor de búsqueda. Este programa compila fragmentos y una base de datos de ellos. [53]

**4.4.2 Requisitos del sistema:** Una de las características de SEO Administrator es que a pesar de ser una suite de Administración Web, no necesita gran cantidad de recursos de memoria y procesamiento al igual que almacenamiento en disco, por lo cual un computador con los siguientes requisitos puede trabajar perfectamente con la herramienta software:

Memoria de 128 Mb o superior.  
Espacio libre en disco de 30 Mb o superior.  
Procesador 800 MHz o Superior.

Se debe tener en cuenta que es una herramienta cuyo funcionamiento lo hace on-line, por lo cual es necesario tener conexión a Internet para poder obtener los resultados deseados.

**4.4.3 Instalación:** Se puede descargar la versión Trial de SEO Administrator de la siguiente página Web:

<http://www.seoadministrator.com/>

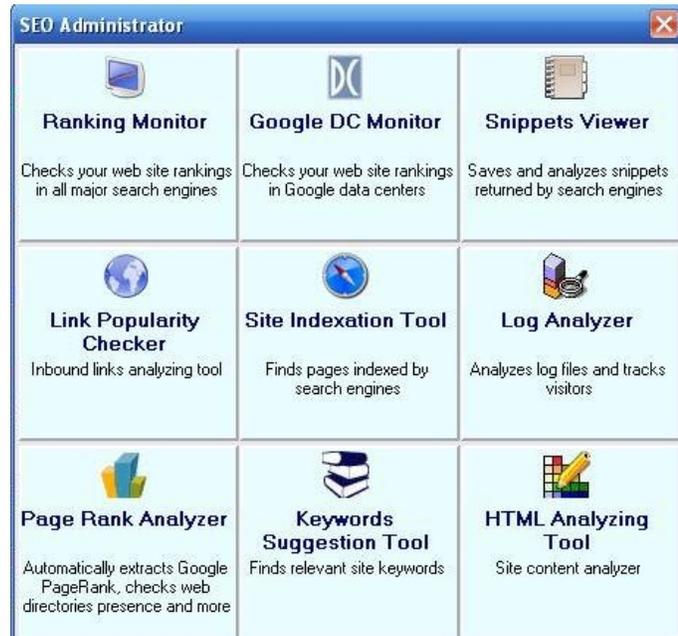
Además si se está interesado en conseguir la versión completa, igualmente desde la anterior Web se lo puede hacer, el valor se encuentra sujeto a cambios dependiendo la cantidad de módulo que se necesite para el funcionamiento de la herramienta.

El proceso de instalación es sencillo, basta con hacer clic en el botón Next o Siguiente, hasta concluir con la instalación.

Al finalizar este proceso aparece la ventana de instalación exitosa, y la herramienta se encuentra lista para ser utilizada. La versión Trial de SEO Administrator se encuentra disponible por 60 días a partir de la instalación.

**4.4.4 Manejo de la herramienta:** Una vez se ejecute la herramienta, presentara la siguiente ventana donde se encuentran los diferentes módulos que conforman la Suite de SEO Administrator.

Figura 57. Ventana de presentación SUITE de SEO Administrator

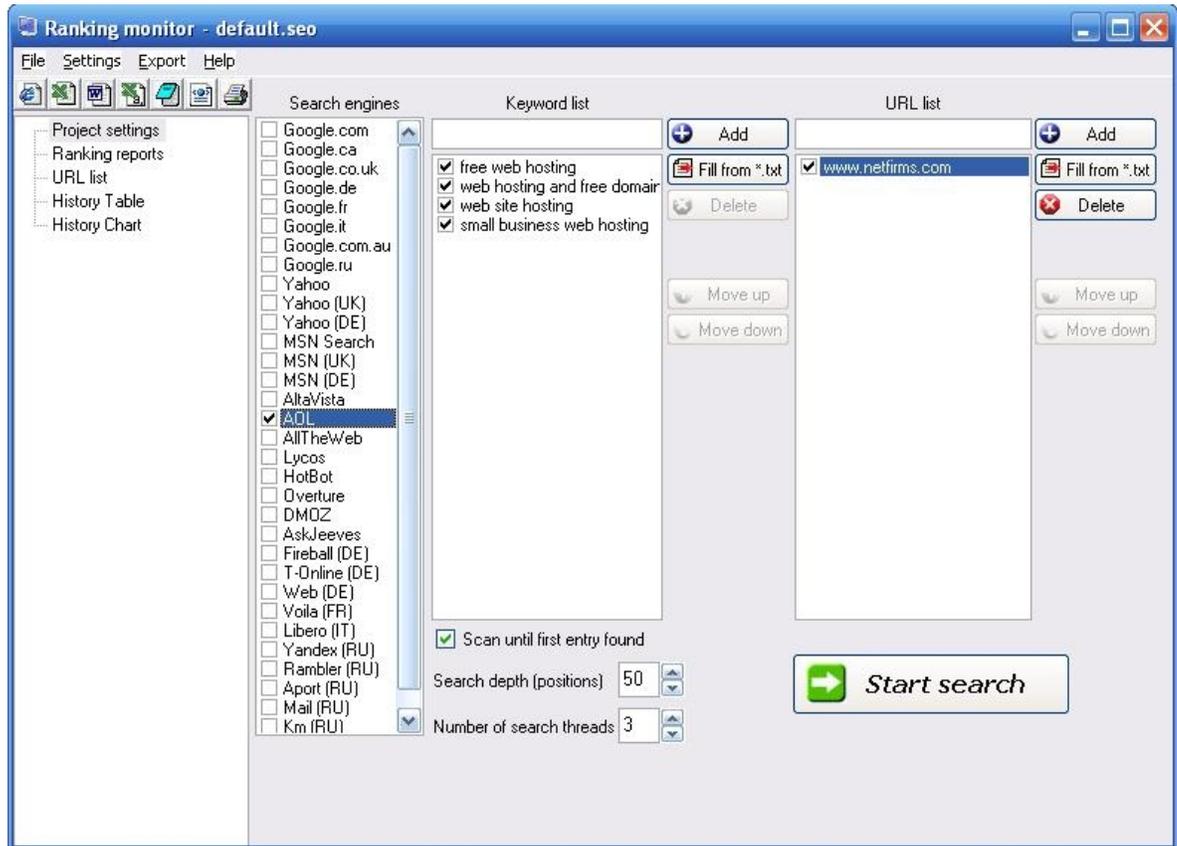


**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

➤ **Ranking Monitor:** Cada especialista debe vigilar periódicamente el posicionamiento del sitio Web en los principales motores de búsqueda. Manualmente es posible, pero esto puede ocupar una cantidad significativa de tiempo. El software de SEO Administrator automatiza completamente esta tarea y le permite crear fácilmente los informes de posición en cuestión de minutos.

Actualmente, el motor de búsqueda más popular con el que cuenta la Web es Google, seguido de Yahoo, MSN y AOL. SEO Administrator apoya todos estos motores de búsqueda importantes, así como más de otros treinta. El funcionamiento del módulo "Ranking Monitor" es de uso fácil e intuitivo. Sólo tiene que especificar las palabras clave, direcciones de su sitio Web que se deben controlar, y continuación, seleccione sus motores de búsqueda. Ranking Monitor inicia a recopila la información necesaria. En las informaciones recolectadas, se mostrará de manera fácil de interpretar los informes de posicionamiento dentro de los motores de búsqueda del sitio Web.

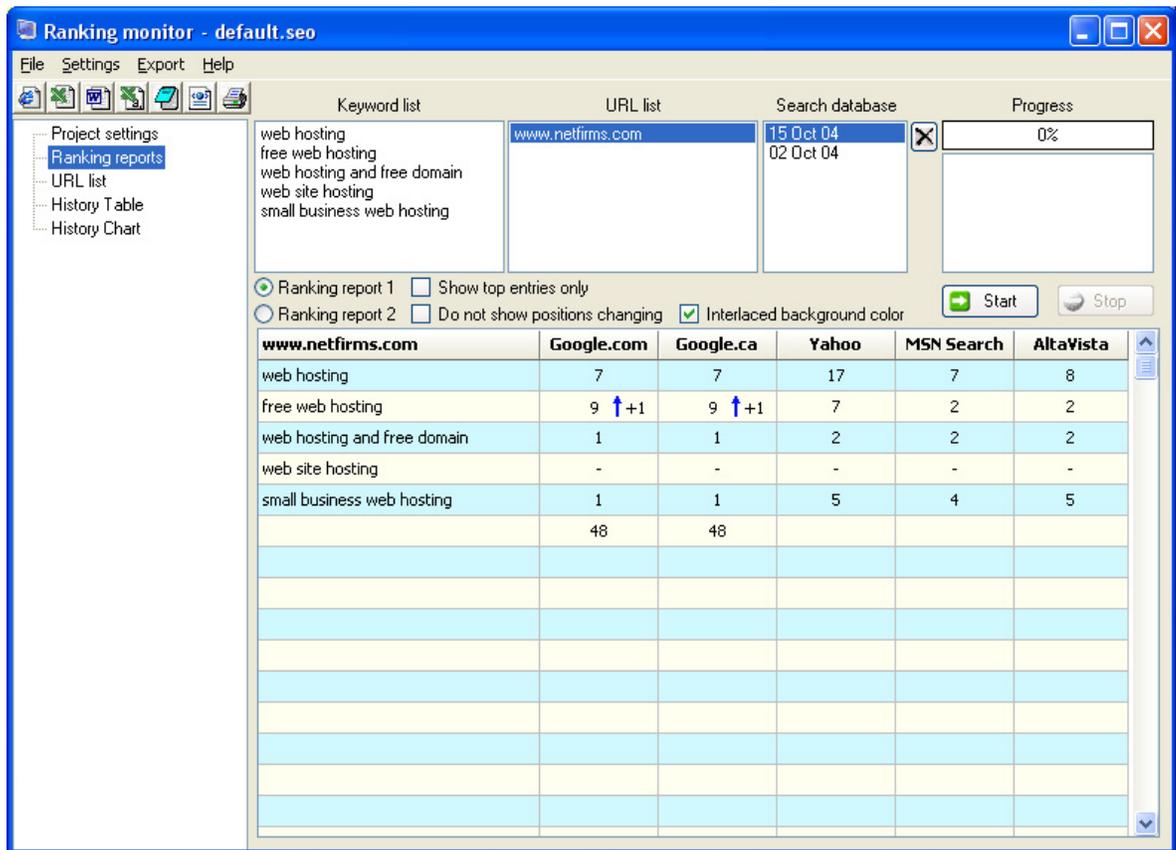
Figura 58. Área de trabajo de Ranking Monitor



**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

**Informe 1 de Posicionamiento**, muestra los datos en una forma breve pero informativa. Así como la posición actual, el informe muestra la posición en que haya cambios con respecto al informe anterior.

Figura 59. Informe 1 de Posicionamiento de Ranking Monitor



Fuente: **SEO ADMINISTRATOR**, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

**Informe 2 de posicionamiento**, es más detallado. La página Web que ocupa la posición se especifica (El símbolo "/" indica la raíz del sitio). Se debe tener en cuenta que si varias páginas del mismo sitio Web ocupan diferentes posiciones en el motor de búsqueda (por ejemplo, en las posiciones 5ª, 6ª y 45ª) esta información se mostrará.

Figura 60. Informe 2 de posicionamiento de Ranking Monitor

The screenshot shows the 'Ranking monitor - default.seo' application window. The interface is divided into several sections:

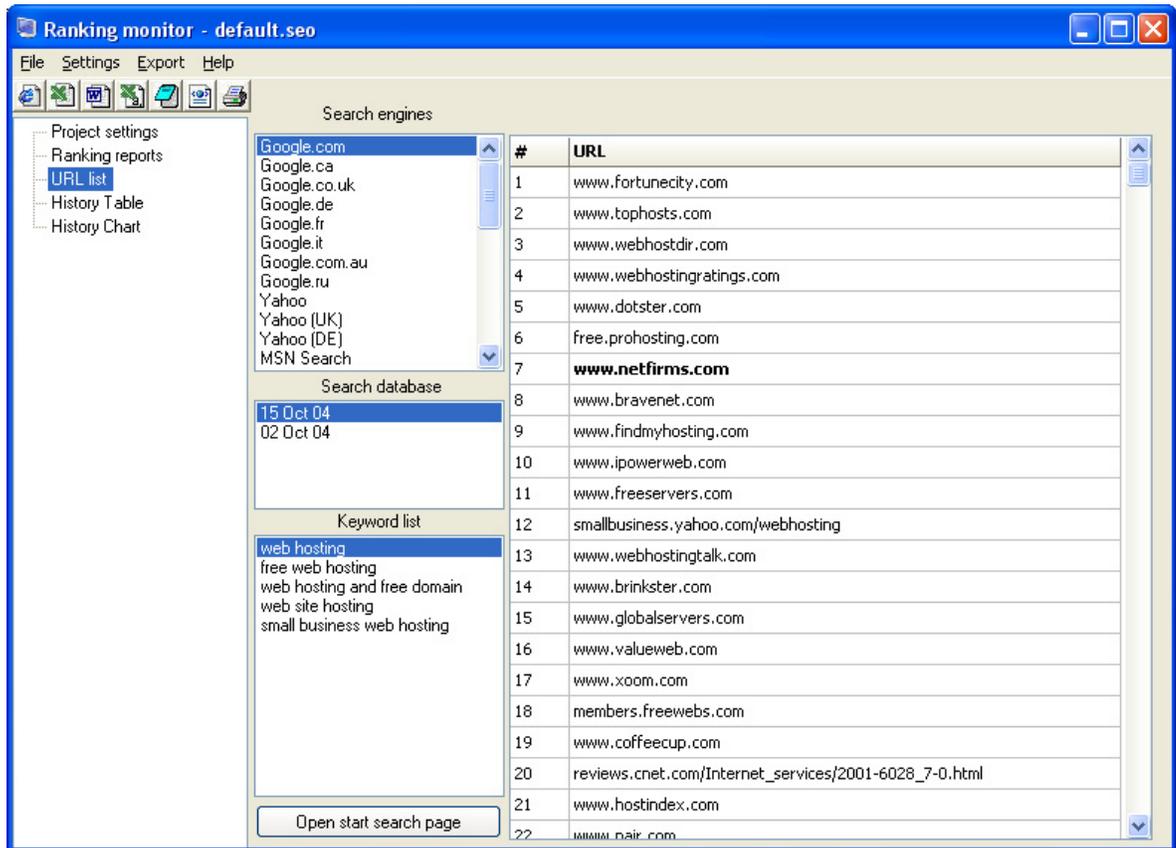
- Keyword list:** Contains the keywords: web hosting, free web hosting, web hosting and free domain, web site hosting, and small business web hosting.
- URL list:** Contains the URL: www.netfirms.com.
- Search database:** Shows dates: 15 Oct 04 and 02 Oct 04.
- Progress:** Shows 0% completion.
- Options:** Includes radio buttons for 'Ranking report 1' and 'Ranking report 2', checkboxes for 'Show top entries only', 'Do not show positions changing', and 'Interlaced background color'. There are 'Start' and 'Stop' buttons.
- Table:** A table with columns: Keywords, Web page, Position, and Previous/Change. It shows results for Google.com, Google.ca, and Yahoo.

| Keywords                    | Web page                               | Position | Previous/Change |
|-----------------------------|----------------------------------------|----------|-----------------|
| <b>Google.com</b>           |                                        |          |                 |
| web hosting                 | <a href="#">/</a>                      | 7        | 7               |
| free web hosting            | <a href="#">/</a>                      | 9        | 10 ↑ +1         |
| web hosting and free domain | <a href="#">/</a>                      | 1        | 1               |
| web site hosting            |                                        |          |                 |
| small business web hosting  | <a href="#">/</a>                      | 1        | 1               |
|                             | <a href="#">/web-hosting-advantage</a> | 48       |                 |
| <b>Google.ca</b>            |                                        |          |                 |
| web hosting                 | <a href="#">/</a>                      | 7        | 7               |
| free web hosting            | <a href="#">/</a>                      | 9        | 10 ↑ +1         |
| web hosting and free domain | <a href="#">/</a>                      | 1        | 1               |
| web site hosting            |                                        |          |                 |
| small business web hosting  | <a href="#">/</a>                      | 1        | 1               |
|                             | <a href="#">/web-hosting-advantage</a> | 48       |                 |
| <b>Yahoo</b>                |                                        |          |                 |

Fuente: **SEO ADMINISTRATOR**, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

Este informe "URL list" muestra todos los eslabones de entrega de un motor de búsqueda en una lista única.

Figura 61. "URL list" de Ranking Monitor



**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

SEO Administrator crea la base de datos "History Table", donde se registra el posicionamiento del sitio Web de manera cronológica, detallando la posición y el respectivo motor de búsqueda.

Figura 62. "History Table", de Ranking Monitor

| Keyword                     | 02 Oct 04 | 15 Oct 04 | 12 Feb 05 |  |  |  |
|-----------------------------|-----------|-----------|-----------|--|--|--|
| <b>Google.com</b>           |           |           |           |  |  |  |
| web hosting                 | 7         | 7         | 3         |  |  |  |
| free web hosting            | 10        | 9         | 7         |  |  |  |
| web hosting and free domain | 1         | 1         | 1         |  |  |  |
| small business web hosting  | 1         | 1         | 1         |  |  |  |
| <b>Yahoo</b>                |           |           |           |  |  |  |
| web hosting                 | 17        | 17        | 23        |  |  |  |
| free web hosting            | 7         | 7         | 10        |  |  |  |
| web hosting and free domain | 2         | 2         | 2         |  |  |  |
| small business web hosting  | 5         | 5         | 9         |  |  |  |
| <b>AltaVista</b>            |           |           |           |  |  |  |
| web hosting                 | 8         | 8         | 20        |  |  |  |
| free web hosting            | 2         | 2         | 5         |  |  |  |
| web hosting and free domain | 2         | 2         | 3         |  |  |  |
| small business web hosting  | 5         | 5         | 9         |  |  |  |

Fuente: **SEO ADMINISTRATOR**, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

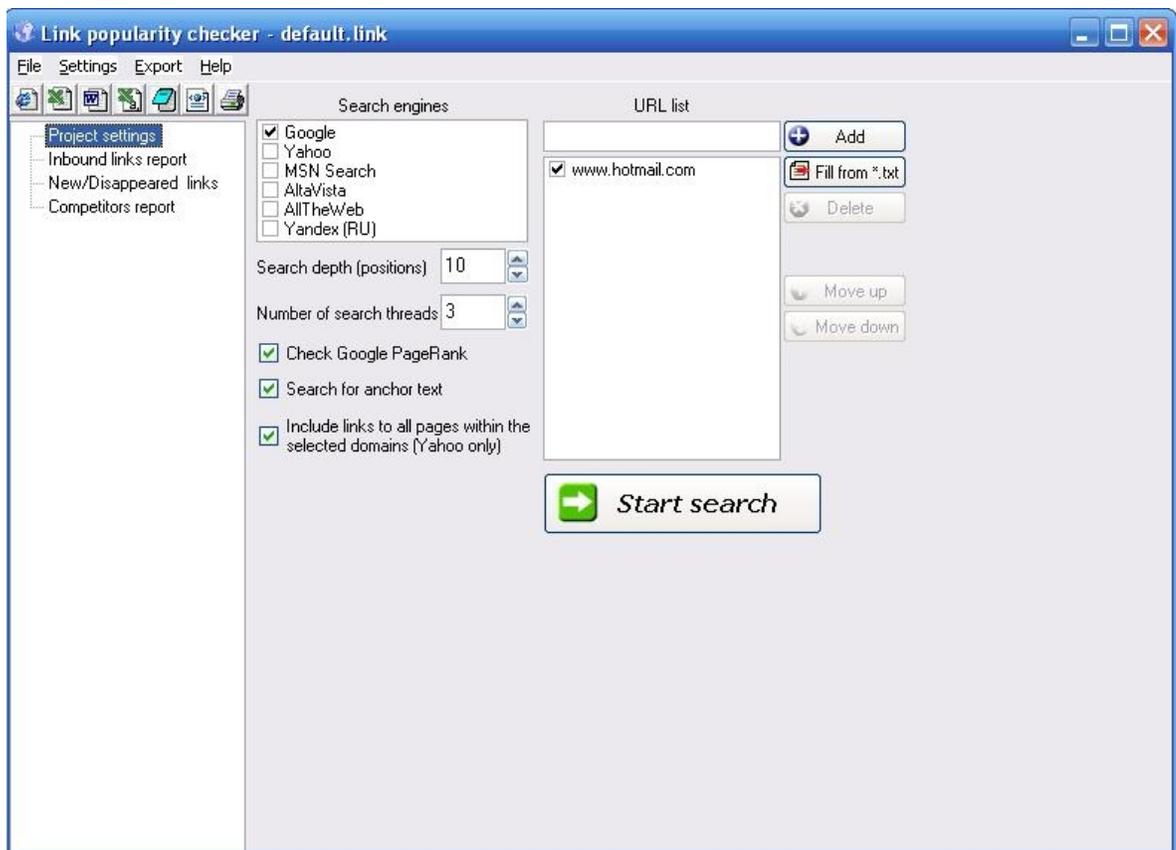
➤ **Link Popularity Checker:** Un gran número de sitios con enlaces en referencia a un determinado sitio Web (enlaces entrantes) es una garantía de altos cargos en los motores de búsqueda. Este número de vínculos es para un sitio la popularidad y es el más importante parámetro utilizado por los motores de búsqueda para evaluar la importancia y el valor de su sitio Web.

La mayoría de los motores de búsqueda tienen una base de datos de enlaces de entrada (acoplamiento de bases de datos) y la mayoría de ellos le permite ver su lista de sitios de enlaces entrantes con la ayuda de los agentes especiales de idiomas. Las bases de datos de diversos motores de búsqueda son fundamentalmente similares entre sí, pero cada uno tendrá únicas de información en el enlace de popularidad que no esté presente en los demás. SEO Administrator comprueba los enlaces entrantes a través de una serie de motores de búsqueda.

Un segundo parámetro importante que afecta a la relación valor de popularidad es el peso y la importancia del sitio de referencia. Un enlace de microsoft.com es mucho más importante que una página de algunos anónimos. Una de las maneras más convenientes para estimar la importancia y popularidad de un enlace Web de referencia es obtener su Google Page Rank.

Por último, el tercer parámetro que afecta el vínculo de popularidad es el ancla de texto. Este es el texto en la página de referencia, que está asociado con el enlace a su sitio. Las palabras clave dentro de este texto mejorarán significativamente sus posiciones en los motores de búsqueda.

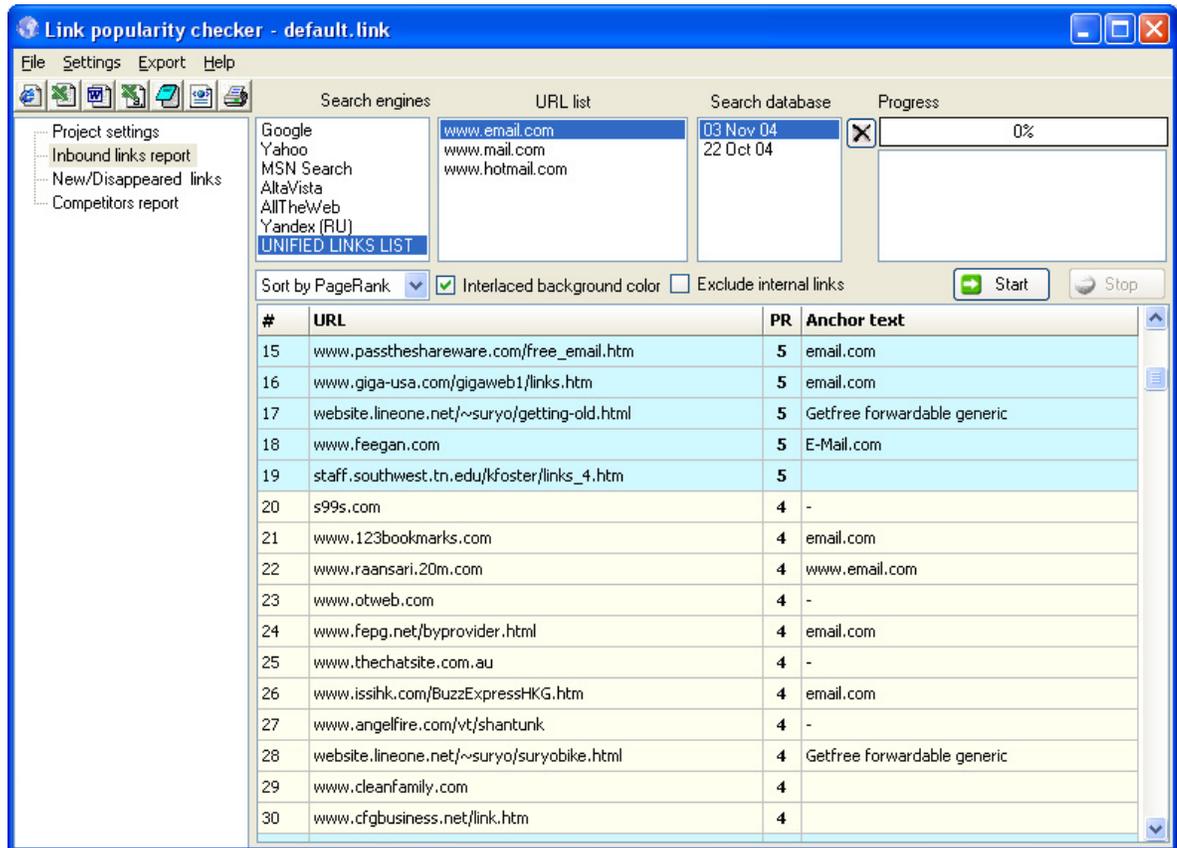
Figura 63. Área de trabajo de Link Popularity Checker



**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

Así como la recogida de la información entrante vínculo, esta herramienta de software también define Page Rank y ancla de texto para cada página de referencia. A continuación se presenta una imagen del primer reporte que genera la herramienta software.

Figura 64. Reporte que generado la herramienta software Link Popularity Checker



**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

En muchos casos, se estará interesado en la dinámica siguiente enlace. En otras palabras, tendrá que seguir los cambios en los vínculos señalando los vínculos nuevos y los vínculos obsoletos. Link Popularity Checker permite ver y seguir la pista de tales vínculos en un período de tiempo determinado.

Figura 65. Reporte generado por un periodo de tiempo, software Link Popularity Checker

| #  | Links appeared                               | PR | #  | Links disappeared                                 | PR |
|----|----------------------------------------------|----|----|---------------------------------------------------|----|
| 1  | www.geocities.com/w_ijen                     | 0  | 1  | www.echoecho.com/email01.htm                      | 4  |
| 2  | www.careervoyager.com                        | 6  | 2  | temp.starklawlibrary.org/blog/archive/2004_06.htm | 4  |
| 3  | www.angelfire.com/vt/shantunk                | 4  | 3  | www.exploora.com.br/ebusca.php?busca=email        | 2  |
| 4  | www.raansari.20m.com                         | 4  | 4  | www.auctiontamer.com/auction/getstarted4.htm      | 3  |
| 5  | www.fayez.net                                | 3  | 5  | www.amherstguide.com/articles/farming/foodproc    | 2  |
| 6  | www.creativelistings.com/bridal/florists.htm | 5  | 6  | www.linux-mips.org/archives/linux-mips/2002-06/n  | 2  |
| 7  | website.lineone.net/~suryo/getting-old.html  | 5  | 7  | www.maedde.de/gbook.php                           | 3  |
| 8  | www.ethailand.com/express                    | 5  | 8  | www.i-dns.net/support_download/downloads/iClie    | 2  |
| 9  | www.connectonline.com                        | 5  | 9  | www.mail-archive.com/netdynamics.public.support   | 2  |
| 10 | www.abab.8m.com                              | 1  | 10 | www.casi.org.uk/discuss/2000/msg00750.html        | 1  |
| 11 | www.geocities.com/eu84                       | 0  | 11 | groups.yahoo.com/group/info-cvs/message/13812     | 1  |
| 12 | www.gomonster.com                            | 5  | 12 | www.mail-archive.com/tomcat-user@jakarta.apache   | 2  |
| 13 | www.alabamaed.com                            | 5  | 13 | lists.contesting.com/archives/html/WriteLog/2001- | 2  |
| 14 | www.gofallon.com                             | 6  | 14 | www.startselect.com/                              | 1  |
| 15 | www.al-kinani.8m.com                         | 3  | 15 | www.beefheart.com/fireparty/archive/2000/03/m     | 1  |
| 16 | www.feegan.com                               | 5  | 16 | www.burmalibrary.org/reg.burma/archives/19991:    | 1  |

Fuente: **SEO ADMINISTRATOR**, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

Casi todos los sitios Web, especialmente si es nuevo, tendrá la competencia que tienen los sitios con mejor posicionamiento en buscadores. Seo Administrator permite comparar la base de datos de enlaces de un sitio Web con los sitios competidores. Estos datos se pueden precisar dentro del reporte de competidores.

➤ **Page Rank Analyzer:** El módulo Page Rank Analyzer está diseñado para analizar automáticamente un gran número de sitios Web y los siguientes parámetros se obtienen para cada sitio en la lista:

- Google Page Rank;
- Alexa ranking de tráfico
- El número de enlaces entrantes, de acuerdo con diversos buscadores;

La tarea básica de los motores de búsqueda es pertinente volver a las páginas Web en respuesta a una consulta de búsqueda. Para facilitar esto, el texto de la página es analizado. Peso y densidad de palabras clave, etiquetas y otros muchos utilizan parámetros. Obviamente, la mayoría de estos parámetros se controlan por parte del autor de la página Web. Un Check con el motor de búsqueda muestra resultados de muchas páginas Web (los llamados "portales"), que están optimizadas para la calidad de las consultas populares, pero que obviamente no contienen información útil para los humanos. Así, los motores de búsqueda no pueden clasificar los sitios basados sólo en el texto de la página.

Por esta razón, los motores de búsqueda utiliza el parámetro de **acoplamiento**, que toma el número y la calidad de los enlaces entrantes en cuenta a la hora de evaluar la pertinencia y el interés de páginas Web.

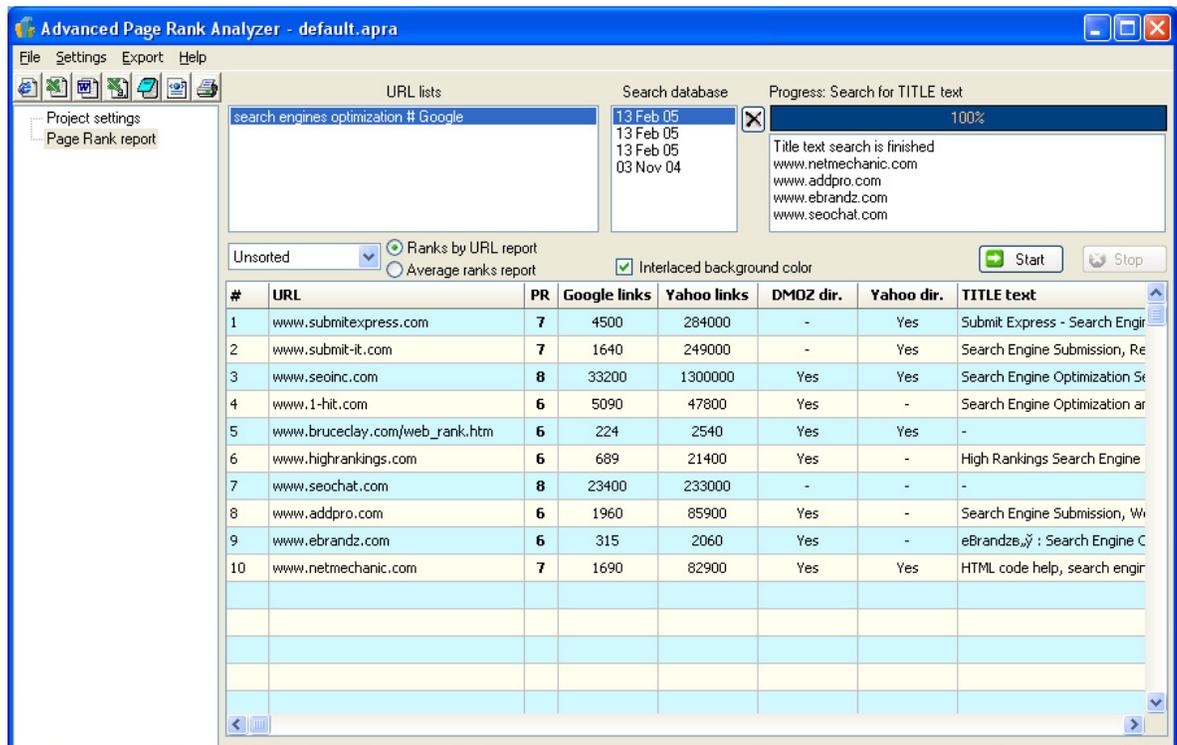
Un gran número de enlaces entrantes por sí sola no es una garantía de la calidad del recurso Web. Además, la calidad de cada vínculo de referencia se toma en cuenta.

**¿Qué es el Page Rank, vínculo popularidad, índice de citas de parámetro?** La cita es un índice numérico de la estimación de la popularidad de un sitio de recursos. Es una medida de absoluta importancia de la página. Cada motor de búsqueda da algoritmos para calcular los índices de citas. Estos valores no se publican normalmente en cualquier lugar. Existe una escala de índice para realizar las respectivas comparaciones, esta comparativa muestra el valor de popularidad (ranking) de una página determinada, frente a la popularidad de otras páginas de Internet.

**Page Rank (PR)** es la escala del índice de citas utilizadas por el motor de búsqueda de Google. Este valor caracteriza la "importancia" o popularidad de cada página. Page Rank no tiene en cuenta la relación de una página a sus recursos, sino que depende del número y la calidad de los enlaces entrantes. Es posible visualizar un valor de Page Rank a la vez con la ayuda de la barra Google gratuita. SEO Administrator es capaz de mostrar el Page Rank de un gran número de sitios.

El siguiente parámetro, es el número de enlaces entrantes de acuerdo con Google, Yahoo, MSN. El número de enlaces entrantes no dan una estimación real de la importancia de la página porque ignora el vínculo de importancia. Sin embargo, también se obtiene una estimación de popularidad de la página.

Figura 66. Área de trabajo Page Rank Analyzer



**Fuente:** **SEO ADMINISTRATOR**, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

➤ **Log Analyzer:** Cada solicitud de su sitio es manejado por el servidor y un adecuado registro se guarda en el archivo de registro (Archivos Log). Una gran cantidad de información sobre los visitantes de un sitio Web se puede capturar de los registros de los Archivos Log del servidor cuando se muestran en un formato comprensible. El archivo de registro del servidor es un archivo de texto que contienen los registros, en un formato especial, de cada solicitud en el sitio.

SEO Administrator se encarga de analizar estos registros en estos dos formatos:

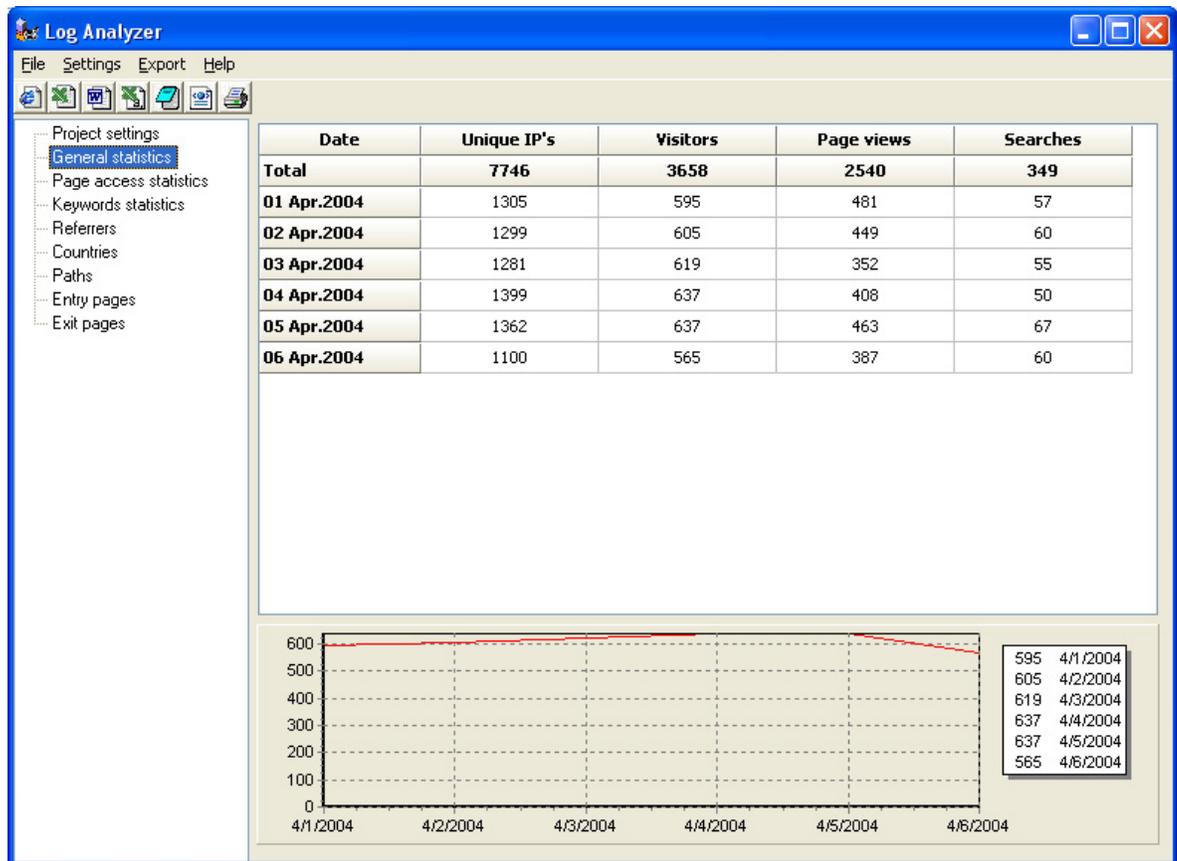
**IIS:** Internet Information Server, generados por los servidores Windows.

**Apache:** Como el formato del archivo de registro generado por el sistema operativo Linux.

Un análisis del sitio y el número de visitantes es la primera fase, seguida por la extracción de parámetros tales como el número de direcciones IP únicas, el número de vistas de página y el número de entradas procedentes de los motores de búsqueda. Log Analyzer automatiza todas estas tareas.

**Número de direcciones IP únicas:** Muestra el número de solicitudes únicas en el servidor. Se debe tener en cuenta que una única dirección IP no significa necesariamente un único visitante. Los archivos de imagen y otros recursos a veces se pidió a un determinado sitio por medio de otros sitios Web y un registro de esta solicitud es presentada.

Figura 67. Área de trabajo Log Analyzer



**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

**Número de visitantes:** El visitante es una persona que ha recuperado la información de los archivos de un sitio Web. Los archivos de información son páginas html (con las extensiones htm, html, asp, etc.) y archivos de datos (con extensiones exe, zip, etc.). Las Imágenes y otros tipos de archivo no son archivos de información.

**Páginas vistas:** Estos son a veces llamadas "impresiones" y muestra el número total de descargas desde la página html del sitio. Por regla general, el número de vistas de la página es significativamente más que el número de visitantes porque

cada visitante examina más de una página. Sin embargo, la situación inversa a veces puede surgir. Por ejemplo, en sitios de descarga de software puede haber muchas solicitudes de descarga de archivos, pero unas cuantas visitas a otras páginas.

**Las entradas de los motores de búsqueda:** Una persona que entre en un sitio Web de un motor de búsqueda es tratada como cualquier otro visitante. Usted tendrá un interés especial en la optimización del motor de búsqueda, por lo que el número de estos visitantes será especialmente interesante. Esta es la razón por la cual el actual el Log Analyzer muestra los visitantes que llegan a través de motores de búsqueda en un informe separado por categoría.

Figura 68. Reporte creado por Log Analyzer

| Page                  | Visits      | Searches   | Bandwidth (Mb)  | Google     |
|-----------------------|-------------|------------|-----------------|------------|
| <b>Total</b>          | <b>8622</b> | <b>346</b> | <b>10769.34</b> | <b>272</b> |
| download/coach-e.zip  | 3236 (37%)  | 0 (0%)     | 5898.99 (54%)   | -          |
| index.htm             | 1508 (17%)  | 302 (87%)  | 10.74 (0%)      | 244        |
| download/coach-eg.zip | 1336 (15%)  | 0 (0%)     | 2057.63 (19%)   | -          |
| download/coach-ei.zip | 1301 (15%)  | 0 (0%)     | 2163.64 (20%)   | -          |
| screenshots.html      | 230 (2%)    | 1 (0%)     | 0.83 (0%)       | 1          |
| download/serie-a.exe  | 206 (2%)    | 0 (0%)     | 633.42 (5%)     | -          |
| registration.html     | 173 (2%)    | 14 (4%)    | 1.26 (0%)       | 14         |
| features.htm          | 139 (1%)    | 5 (1%)     | 0.70 (0%)       | 5          |
| news.htm              | 133 (1%)    | 1 (0%)     | 0.58 (0%)       | 1          |
| RUS/index.html        | 110 (1%)    | 15 (4%)    | 0.28 (0%)       | 5          |
| rus/descr.html        | 45 (0%)     | 2 (0%)     | 0.21 (0%)       | 1          |
| awards.html           | 43 (0%)     | 0 (0%)     | 0.25 (0%)       | -          |
| contacts.html         | 43 (0%)     | 2 (0%)     | 0.19 (0%)       | 1          |
| RUS/register.html     | 30 (0%)     | 0 (0%)     | 0.22 (0%)       | -          |
| rus/screen.html       | 26 (0%)     | 0 (0%)     | 0.04 (0%)       | -          |
| index.html            | 21 (0%)     | 1 (0%)     | 0.03 (0%)       | -          |
| coach/descr.html      | 17 (0%)     | 2 (0%)     | 0.09 (0%)       | -          |
| MSOffice/cltreq.asp   | 8 (0%)      | 0 (0%)     | 0.01 (0%)       | -          |
| coach/screen.html     | 3 (0%)      | 0 (0%)     | 0.0 (0%)        | -          |
| coach/register.html   | 2 (0%)      | 1 (0%)     | 0.0 (0%)        | -          |

**Fuente:** SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software. SEO ADMINISTRATOR. *Seo tools, site promotion software program*. [En línea] 2002-2010. [Citado el: 15 de Enero de 2009.] <http://www.seoadministrator.com/>. [53]

#### **4.4.5 Ventajas:**

- SEO Administrator ofrece un conjunto completo de funciones y los informes necesarios para verificar un sitio Web y realizar un seguimiento de posicionamiento basado en las palabras clave especificadas.
- SEO Administrator es muy fácil de usar.
- SEO Administrator ofrece pleno apoyo de los principales motores de búsqueda, significa que esta herramienta de posicionamiento de sitio Web es un componente esencial de todos los Web master.

#### **4.4.6 Desventajas:**

- La versión demo en algunos de su módulos únicamente permite trabajar con tres motores de búsqueda de manera simultánea y con cinco palabras y frases clave al igual que con las direcciones Web.
- La suite de SEO Administrator es comercial, por lo cual la descarga de la misma tiene un costo de 499 dólares con una licencia para diez computadores.

### **4.5 A1 WEBSITE ANALYZER VERSIÓN 1.4.1**

**4.5.1 Descripción preliminar:** A1 Website Analyzer es una completa solución para examinar enlaces (Link Checker) porque presenta diferentes opciones para la observación y manipulación de los documentos que conforman un sitio Web. A diferencia de las herramientas en línea Link Checker, la herramienta permite inspeccionar los sitios Web de pruebas situado en cualquier lugar, incluyendo los locales, Internet, sistema de archivos local, y la red de área local. Con este software no se ha fijado el límite superior de las páginas y los enlaces que se pueden comprobar.

**A1 Website Analyzer** permite determinar el contenido del sitio Web y funciona como una herramienta de análisis de la estructura del sitio Web.

**A1 Website Analyzer** es un software que puede ser utilizado por un programador, administrador de sitios Web, gerente de sistemas, etc., y además ofrece una rentable alternativa a la hora de analizar la creación, actualización de los sitios que se desee. Su manera de funcionamiento, es básicamente realizar la extracción de toda la estructura de datos de un sitio Web, proporcionando de esta manera el almacenamiento de aquellos datos que se necesiten, tales como documentos, imágenes, páginas Web, y manejar los redireccionamientos de vínculos, ofreciendo los tiempos de respuesta, tamaño de los archivos, tiempo de descarga. Asimismo Puede integrarse con herramientas de análisis de sitio Web en línea.

Toda la información de los sitios Web e incluso datos de exploración del sitio Web

puede ser exportada o importada como XML o CSV (o archivos de texto).

**A1 Website Analyzer** cuenta con un poderoso motor de búsqueda, con el cual puede examinar las palabras clave y obtener la estructura de datos en formatos CSV o XML. [42]

**4.5.2 Requisitos del sistema:** La versión de escritorio de la herramienta puede ejecutarse en los siguientes **Sistemas Operativos:** Win 98, 2000, Me, Windows Xp, y Windows Vista.

**4.5.3 Instalación:** Para descargar la copia de evaluación de **A1 Website Analyzer** Versión 1.4.1 debe ir a la siguiente página:

<http://www.microstools.com/products/Website-analyzer/>

El proceso de instalación es el siguiente: basta con hacer clic en el archivo ejecutable del programa, escoger el idioma en el cual se va hacer la instalación, luego hacer clic en el botón Next o siguiente, una vez instalado se puede trabajar con la herramienta de prueba por un tiempo de de 30 días.

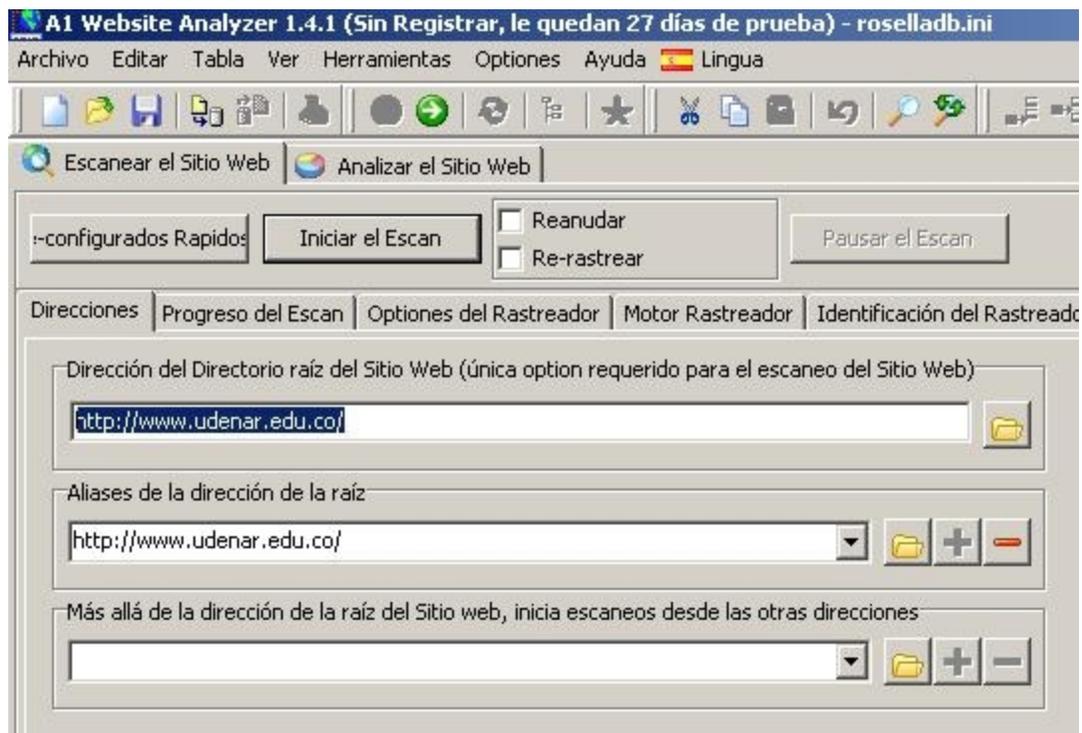
Al final de este proceso aparece la ventana de instalación exitosa, la cual nos indica que toda la configuración de instalación se ha realizado satisfactoriamente y el **A1 Website Analyzer** está listo para ser utilizado.

**4.5.4 Manejo de la herramienta:** El trabajo con la herramienta es sencillo, sin embargo el usuario puede ir a la opción Help (Ayuda), del menú principal, donde se desplegará el manual de uso de **A1 Website Analyzer**, con toda la información que se requiere para el adecuado uso del producto.

Para empezar a trabajar con **A1 Website Analyzer** se debe estar conectado a Internet, tener un proyecto de un sitio Web, o un sitio Web de prueba guardado en el disco local, en la LAN, etc., en el caso de tener un proyecto este se carga seleccionando la pestaña “Archivo” y luego la opción “Abrir Proyecto” de estas forma se puede examinar la información del sitio Web guardada en el proyecto.

Otra opción es la de cargar un sitio Web local o de prueba, para lo cual se debe seleccionar la pestaña “Escanear el Sitio Web”, y dentro de ella se selecciona la pestaña “Direcciones”, en la cual se coloca la dirección del sitio Web con el cual se va a trabajar, o la ruta donde se encuentra el sitio Web de prueba, en este caso se utilizo la URL <http://www.udenar.edu.co/>, para realizar el estudio de la herramienta.

Figura 69. Área de trabajo A1 Website Analyzer



Después de haber ingresado la dirección del sitio Web, se puede ingresar un alias del mismo y nos presenta otra opción en la cual permite incluir desde que dirección URL puede empezar a trabajar la herramienta.

La herramienta brinda varias opciones para filtrar páginas y documentos, como también la utilización de rastreadores (crawlers), por lo tanto para que **A1 Website Analyzer** empiece a funcionar se presiona el botón Iniciar “Escan” (en el caso de que esté conectado a Internet), en este momento el software se demora un tiempo para extraer la información del sitio Web, si es un proyecto el que va a cargar nos muestra la información que se ha guardado en él. Más sin embargo se debe tener en cuenta que es necesario estar conectado a internet para utilizar algunas de las opciones de la herramienta.

Al instante de que el programa termine de encontrar la información del sitio Web se podrá observarla, gracias a las opciones que nos brinda **A1 Website Analyzer**, en los siguientes párrafos se describen algunas opciones para el análisis de la estructura de los sitios Web.

Al seleccionar la pestaña “Analizar el Sitio Web”, se presenta una ventana la cual se divide en dos columnas, la primera en el lado izquierdo muestra la información de la estructura del enlace del sitio Web.

Como se puede observar en la siguiente gráficas, se indica la jerarquía del sitio Web, las páginas, documentos, imágenes, etc., con la respectiva información de cada uno de sus elementos. Para observar esta información se debe hacer clic en el elemento al cual se quiera referenciar.

Figura 70. Estructura de enlace del sitio Web analizado por A1 Website Analyzer

The screenshot shows the 'Estructura de enlace del Sitio web' window in A1 Website Analyzer. The window has two tabs: 'Escanear el Sitio Web' and 'Analizar el Sitio Web'. The main area is titled 'Mapa del Sitio' and contains a table with the following columns: 'Dirección', 'Objetos', 'R.Cód...', and 'R.Tie...'. The table lists various website elements, including folders and files, with their respective counts and response codes. Below the table, there are sections for 'Active address' and 'Active user agent'.

| Dirección                 | Objetos | R.Cód... | R.Tie... |
|---------------------------|---------|----------|----------|
| http://www.udenar.edu.co/ | 5869    | 200      | 1        |
| Default.aspx              | 1       | 200      | 1        |
| GIA/                      | 1       | 404      | 1        |
| actualidad/               | 45      | 0        |          |
| acuicola                  | 1       | 301      | 2        |
| acuicola/                 | 123     | 200      | 1        |
| admisiones.aspx           | 1       | 200      | 1        |
| archivocentral/           | 1       | 404      | 1        |
| artefacto/                | 3       | 0        |          |
| banda/                    | 31      | 200      | 1        |
| banner/                   | 2       | 0        |          |
| bg.gif                    | 1       | 404      | 1        |
| becas.aspx                | 1       | 200      | 1        |
| botones/                  | 37      | 0        |          |
| bproyectos/               | 8       | 200      | 1        |
| ceilat/                   | 3       | 0        |          |
| centros.aspx              | 1       | 200      | 1        |
| centros/                  | 22      | 0        |          |
| cesun/                    | 34      | 200      | 2        |
| cesun1/                   | 14      | 0        |          |
| cidomas/                  | 32      | 0        |          |
| ciesju.aspx               | 1       | 200      | 1        |
| ciesju/                   | 7       | 0        |          |

Active address  
http://www.udenar.edu.co/

Active user agent

Por otra parte en la siguiente gráfica, en el lado derecho de la pantalla, se observa una columna que nos presenta mucha más información acerca del sitio Web, la primera pestaña "Datos Recogidos", está constituida por diferentes opciones:

Figura 71. Reporte generado por la herramienta A1 Website Analyzer

| Enlaces a 48 páginas                                                             | Count 52 |
|----------------------------------------------------------------------------------|----------|
| http://www.udenar.edu.co/directorio/                                             | 2        |
| http://www.udenar.edu.co/eventos.aspx                                            | 2        |
| http://www.udenar.edu.co/noticias.aspx                                           | 2        |
| http://www.udenar.edu.co/sistmatri.aspx                                          | 2        |
| http://www.udenar.edu.co/                                                        | 1        |
| http://www.udenar.edu.co/actualidad/BOLETIN_118._PLANETARIO_DE_PASTO.pdf         | 1        |
| http://www.udenar.edu.co/actualidad/BOLETIN_120._PLAN_DE_DESARROLLO_RENDICI%C... | 1        |
| http://www.udenar.edu.co/actualidad/DIPLOMADOcooperacion.pdf                     | 1        |
| http://www.udenar.edu.co/admisiones.aspx                                         | 1        |
| http://www.udenar.edu.co/archivocentral/                                         | 1        |
| http://www.udenar.edu.co/becas.aspx                                              | 1        |
| http://www.udenar.edu.co/centros.aspx                                            | 1        |
| http://www.udenar.edu.co/centros/cjuridicos.aspx                                 | 1        |
| http://www.udenar.edu.co/comunicados.aspx                                        | 1        |
| http://www.udenar.edu.co/comunicados/estampillapro.html                          | 1        |
| http://www.udenar.edu.co/contacto/                                               | 1        |
| http://www.udenar.edu.co/correo.aspx                                             | 1        |
| http://www.udenar.edu.co/Default.aspx                                            | 1        |
| http://www.udenar.edu.co/destudiantes/index.html                                 | 1        |
| http://www.udenar.edu.co/editorial/                                              | 1        |
| http://www.udenar.edu.co/educacion.aspx                                          | 1        |
| http://www.udenar.edu.co/especializaciones/creatividad.aspx                      | 1        |
| http://www.udenar.edu.co/especializaciones/disenio.aspx                          | 1        |
| http://www.udenar.edu.co/eventos/horarios.htm                                    | 1        |
| http://www.udenar.edu.co/extension.aspx                                          | 1        |
| http://www.udenar.edu.co/gestion.aspx                                            | 1        |
| http://www.udenar.edu.co/gestionBBVA.aspx                                        | 1        |
| http://www.udenar.edu.co/historia.aspx                                           | 1        |
| http://www.udenar.edu.co/intranet.aspx                                           | 1        |
| http://www.udenar.edu.co/investigacin.aspx                                       | 1        |
| http://www.udenar.edu.co/mapadelsitio.aspx                                       | 1        |
| http://www.udenar.edu.co/mecij/                                                  | 1        |
| http://www.udenar.edu.co/mision.aspx                                             | 1        |
| http://www.udenar.edu.co/moni.aspx                                               | 1        |
| http://www.udenar.edu.co/planeacion.aspx                                         | 1        |
| http://www.udenar.edu.co/portafolio                                              | 1        |
| http://www.udenar.edu.co/postgrado.aspx                                          | 1        |

Página de Datos Enlaces [Interno] Enlaces [externo] Enlazado por Usa [interno] Usa [externo] Usado [Interno]

**Página de Datos:** Información relacionada con el sitio Web en general, como cada uno de sus componentes (páginas Web, documentos, imágenes, etc.). En nuestro caso de estudio una información relevante es la que se presenta en el campo importancia del resultado escalado, el cual es un número que se da por el ponderado de enlaces entrantes y transformado a un algoritmo (Page Rank) basado en escala de 0 – 10.

**Enlaces [Interno]:** Muestra las direcciones URLs internas del sitio Web, lo cual brinda la información de los enlaces internos a los que se dirige el sitio Web o uno de sus componentes.

**Enlaces [Externo]:** Muestra las direcciones URLs externas a las que el sitio Web se conecta, lo cual brinda la información de los enlaces externos a los que se dirige el sitio Web o uno de sus componentes.

**Enlazado por:** Muestra las direcciones URLs que se enlazan al sitio Web o uno de sus componentes.

Las demás opciones indican las direcciones de las cuales el sitio Web extrae alguna información como documentos, imágenes, etc. Y otra información no relevante en nuestro caso de estudio.

La segunda pestaña "Ver Archivo": muestra el archivo como tal del sitio Web o del componente que se escoja.

La tercera pestaña "Ver Fuente": muestra la fuente del sitio Web o del componente que se escoja.

La cuarta pestaña "W3C validar HTML" validar las páginas Web: **A1 Website Analyzer** usa la versión en línea del W3C de validación de HTML incrustado en el interior de la herramienta de control del navegador. [57]

HTML es el acrónimo de Hyper Text Markup Language. La extensión de los archivos es a menudo ".Html":

Puede validar los archivos HTML de la LAN o del disco local.  
Puede validar los archivos HTML de la Internet o del localhost.

La última pestaña "W3C validar CSS" para validar la página Web: **A1 Website Analyzer** usa la versión en línea de la herramienta de validación W3C CSS. [56]

CSS es el acrónimo de Cascading Style Sheets. La extensión de archivo es "Css":

Puede validar los archivos CSS de la LAN o disco local.  
Puede validar los archivos CSS de Internet o del localhost.  
Puede validar archivos de la página Web de Internet.

La herramienta A1 Web Analyzer El sitio Web Se Puede Exporta A Xml Y Csv

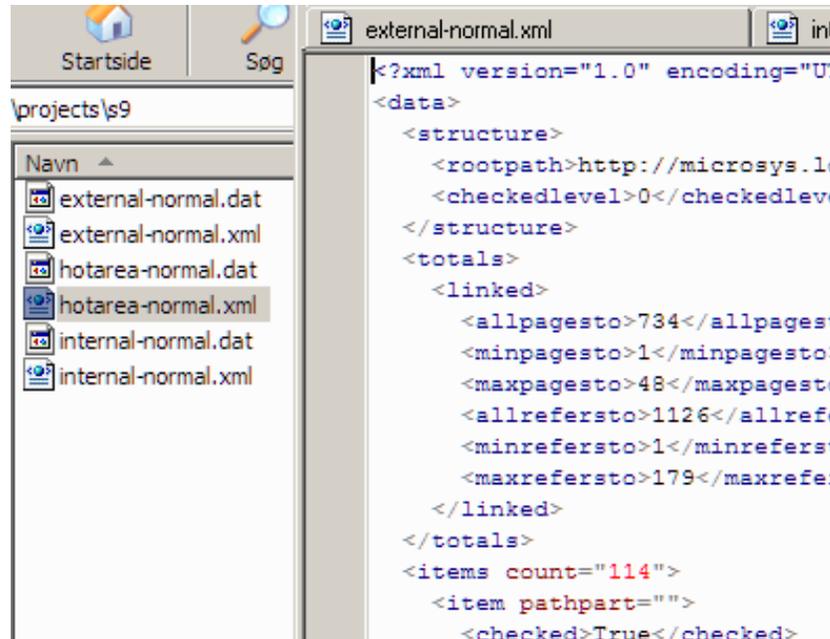
**Sitio Web exportador de datos a CSV**, todas las listas, cuadros de texto, de red y puntos de vista de árbol, etc. se pueden exportar a archivos de texto o CSV, la herramienta brinda la opción de ocultar/mostrar columnas antes de la exportación.

**Proyectos de sitio Web en XML.** Estructura de datos extraídos de un recurso que a menudo se denomina metadatos "datos sobre datos". Al guardar proyectos en A1 Web Analyzer, una gran cantidad de esos datos se guardan en archivos XML.

XML ofrece una gran alternativa ya que a través de él, se puede realizar análisis de datos y extracción de datos. Existen contenedores para ello en casi todos los

lenguajes de programación, por ejemplo, Java, PHP, C #, Visual Basic, Delphi, etc.

Figura 72. Exportar reportes de la herramienta A1 Web Analyzer



Las páginas Web del proyecto se almacenan en metadatos en documentos XML perfecto para minería de datos, ya que ofrece información que se puede analizar o minar.

#### 4.5.5 Ventajas:

- Genera informe de análisis del sitio Web.
- La exploración del sitio Web recoge gran cantidad de datos, Esto significa que se puede analizar el sitio Web teniendo en cuenta importantes parámetros.
- Detecta Enlaces rotos, permitiendo la actualización del sitio Web.
- Permite determinar toda la estructura de datos y páginas Web, que contenga el Sitio Web.
- Muestra el panorama de los archivos utilizados por el sitio Web, más específicamente su Ruta: El camino asumido por o relacionado con algún archivo.
- Detecta rápidamente los enlaces internos y externos existentes, dentro de un sitio Web.

- Todos los enlaces se ponderan para calcular mejor e rango de la página.

#### **4.5.6 Desventajas:**

- Es totalmente comercial, y la herramienta para evaluación tan solo dura por 30 días de prueba.
- La herramienta fue creada exclusivamente para la plataforma Windows en cualquiera de sus versiones, por lo tanto no permite ser ejecutada en otros sistemas operativos.
- El software maneja un algoritmo para el análisis de la estructura Web como lo es Page Rank, pero no presenta información relevante para un tipo de minería Web de estructura.
- La visualización de los enlaces es muy general, y no permite la identificación de patrones en el sitio Web.
- Brinda información para poder hacer minería de datos y no minería Web.

#### **4.6 OTRAS HERRAMIENTAS**

Como nos hemos podido dar cuenta la mayoría de herramientas existentes permiten analizar la cantidad de enlaces entrantes y salientes de las páginas Web y en algunos de los casos permite visualizar el grafo que genera una respectiva búsqueda de palabras clave, frases clave o direcciones URL.

A continuación se mencionan algunas herramientas que no tienen relevancia debido a su limitada capacidad de trabajo y de análisis de la información.

**4.6.1 Page Rank Checker Versión 3.0:** Page Rank Checker es una herramienta gratuita para comprobar el ranking de las páginas de cualquier sitio Web de manera fácil, además nos muestra el valor del Page Rank en las páginas.

Las páginas son cargadas a la herramienta software por medio de archivos planos con extensión .txt, en donde se deben encontrar las direcciones Web y separadas unas de otras por medio de un salto de línea (enter).

La calificación del Page Rank es de 0 a 10, considerándose a 10 como el número más alto otorgado. Este valor numérico indica la popularidad de un Sitio Web. La calificación que es otorgada a los Sitios Web es definida por el motor de búsqueda google.

Page Rank Checker además de comprobar el ranking de los Sitios Web, genera reportes sencillos de entender, los cuales puedes ser de gran importancia para el Webmaster, una vez sean analizados e interpretados de manera correcta. [47]

**Ventajas:**

- Es una herramienta que no requiere alta capacidad de procesamiento.
- Crea reportes HTML y se pueden guardar los resultados obtenidos.
- Fácil de utilizar.

**Desventajas:**

- El pago por concepto de licencias es muy elevado.
- Los reportes que realizan sólo indican los datos obtenidos sin ningún tipo análisis que orienten al Webmaster.
- Se debe crear un archivo plano de las direcciones Web para poder analizar los respectivos sitios.
- Únicamente trabaja bajo la plataforma de Windows en sus diferentes versiones.

**4.6.2 Surf 3d Beta:** Surf3D Beta permite a los usuarios personalizar el descubrimiento y recolección de información en la Web, ofrece potentes búsquedas y visualización de datos.

Surf3d Beta ofrece soluciones profesionales para el ahorro de tiempo en la navegación Web, esta herramienta software reduce el tiempo de búsqueda en más de un 80%, en comparación a lo que normalmente lleva hacerlo a través del motor de búsqueda.

Este software permite que cualquier motor de búsqueda en la red determine el ranking de los resultados obtenido, además evita las rutas irrelevantes, obtiene panoramas visuales de muchas páginas Web a la vez, es fácil de integrar con aplicaciones de software existentes o protegidas por contraseñas del navegador o de bases de datos accesibles. <http://www.navagent.com/> [44]

**Ventajas:**

- Recuperar cualquier archivo de la Web y obtiene una vista previa.
- Obtiene panoramas visuales en 3 dimensiones de las páginas Web.
- Evita las rutas irrelevantes con lo cual se optimiza la búsqueda de información.

**Desventajas:**

- La versión demo únicamente muestra los gráficos en tres dimensiones de páginas previamente cargadas, por lo cual no ofrece libertad al momento de analizar la página que se desea.
- El pago por concepto de licencias es muy elevado.
- Solo se basa en visualización de relevancias de las páginas Web, más sin embargo no Realiza Minería Web de Estructura.

**4.6.3 NetTools Spider:** NetTools Spider, es una herramienta software multifuncional de gran utilidad para internet, el cual permite descargar, navegar, buscar datos y sitios Web.

NetTools Spider, es lo suficientemente flexible como para los usuarios más exigentes, pero bastante simple que se puede descargar un sitio Web con sólo 2 clics del ratón. Literalmente, se puede empezar a descargar un sitio Web en menos de 30 segundos.

NetTools Spider, convertirá un sitio localizado a una copia, lo que hace posible copiar el sitio Web en un CD y verlo con cualquier navegador. Se puede compartir las descargas de los sitios Web con amigos, compañeros o clientes.

NetTools Spider, realiza búsqueda para los archivos que contienen palabras clave. Puede buscar en miles de sitios Web y descargar sólo los archivos que contienen las palabras que se está buscando.

NetTools Spider, es capaz de controlar todos los eslabones de un sitio Web incluidos los enlaces generados en contenido dinámico y enlaces a sitios Web externos.

NetTools Spider, posee como característica más relevante que puede ser usado como una herramienta de minería Web en tiempo real. Con el uso de scripts simples, NetTools Spider puede extraer elementos de información desde un sitio Web y almacenar esa información en una base de datos o archivo de texto. Estos scripts pueden ser escritos en cualquiera de los dos lenguajes de programación Java Script y VBScript.

NetTools Spider, permite exportar fácilmente información de la página Web y enlaces en diferentes formatos, incluyendo Excel y csv. [50]

**Ventajas:**

- Logra recuperar los enlaces rotos encontrados en los motores de búsqueda.
- Exporta fácilmente la información que se encuentran en las páginas Web.
- Trabaja en tiempo real.

**Desventajas:**

- La herramienta software completa tiene un costo de 69.95 dólares.
- Únicamente tiene en cuenta los enlaces externos.
- Se enfoca más en la duplicación y copia de portales Web.

## 5. METODOLOGÍA DE DESARROLLO

### 5.1 PROGRAMACIÓN EXTREMA (XP)

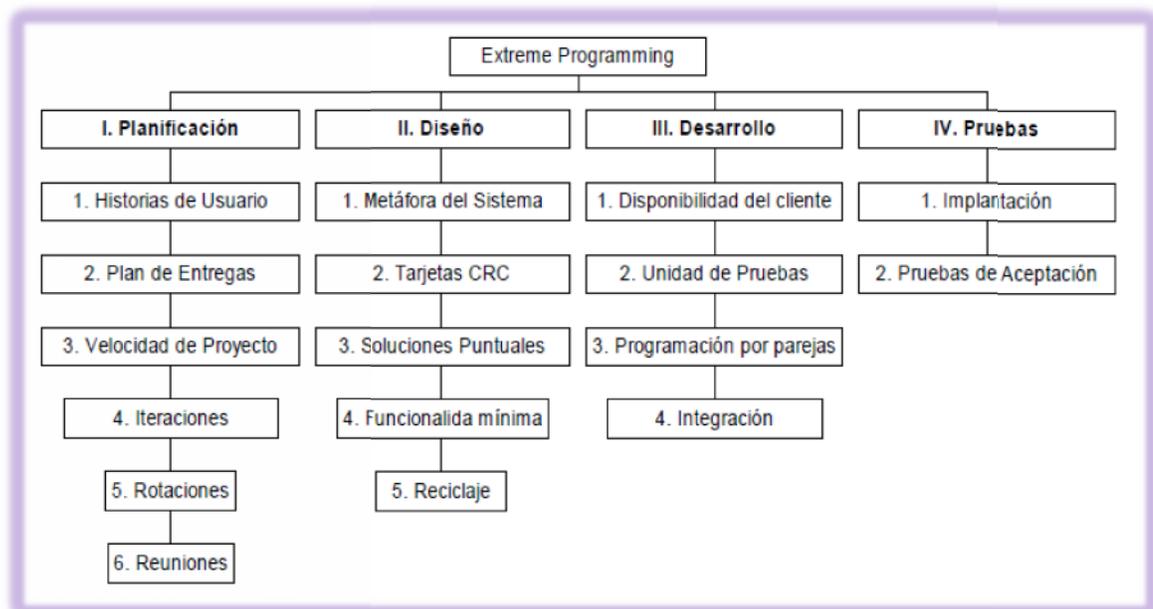
Esta es una de las metodologías de desarrollo de software más utilizada en la actualidad, para proyectos de corto plazo y con equipo de bajos requerimientos tecnológicos. Consiste en una programación rápida o extrema, cuya particularidad es tener como parte del equipo, al usuario final como uno de los requisitos para llegar al éxito del proyecto.

En esta metodología, el cliente se convierte en un miembro más del equipo de trabajo y es el encargado de decidir que se implementa, puede añadir, cambiar o quitar requerimientos en cualquier momento para lo cual debe estar enterado constantemente del estado real y el progreso del proyecto obteniendo lo máximo de cada semana de trabajo.

XP trabaja cuatro fases principales: Planificación, Diseño, Desarrollo y Pruebas.

Estas fases se dividen, a su vez, en subfases de desarrollo que poseen una serie de pasos que permiten realizar un adecuado desarrollo del proyecto.

Figura 73. Fases de la metodología XP



**Fuente:** B I POWER BLOGGERS. Metodología Extreme Programming(XP) parte I. [En línea] 28 de Abril de 2009. [Citado el: 18 de Enero de 2010.] [http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte\\_28.html](http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte_28.html). [3]

Algo que se debe tener en cuenta en esta metodología, es que muchas de los procesos de cada una de las fases consisten en actividades que se relacionan entre sí. Por ello, en lugar de verlas como pasos estrictos y rígidos, deben verse como recomendaciones que se deben tener en cuenta al momento de realizar las etapas principales, en cuanto al establecimiento de los requerimientos del sistema, la asignación de actividades, las reuniones con el equipo de trabajo, la programación y las pruebas.

**5.1.1 Fase de planificación:** En esta primera fase, se realiza toda la planeación del proyecto, los clientes relatan a grandes rasgos las historias de usuario o requerimientos, al tiempo que los desarrolladores se familiarizan con las herramientas tecnológicas y prácticas que se implementan en el proceso de desarrollo. Una vez seleccionada la tecnología, se exploran las posibilidades de arquitectura del sistema y se hace un diseño previo.

En general esta fase contiene los siguientes procesos que se desarrollan para determinar la planeación: Historias de Usuario, Plan de entregas, Iteraciones y Reuniones. A continuación se realiza una descripción breve de cada uno de los procesos de esta fase:

➤ **Historias de usuario:** El primer paso de cualquier proyecto que siga la metodología X.P es definir las historias de usuario con el cliente. Las historias de usuario tienen la misma finalidad que los casos de uso pero con algunas diferencias: Constan de unas cuantas ideas escritas por el cliente en un lenguaje no técnico sin hacer mucho hincapié en los detalles; no se debe hablar ni de posibles algoritmos para su implementación ni de diseños de base de datos adecuados, etc. Son usadas para estimar tiempos de desarrollo de la parte de la aplicación que describen.

También se utilizan en la fase de pruebas, para verificar si el programa cumple con lo que especifica la historia de usuario. Cuando llega la hora de implementar una historia de usuario, el cliente y los desarrolladores se reúnen para concretar y detallar lo que tiene que hacer dicha historia. El tiempo de desarrollo ideal para una historia de usuario es entre 1 y 3 semanas.

➤ **Plan de entregas:** Después de tener ya definidas las historias de usuario es necesario crear un plan de publicaciones, en inglés "Release plan" (Planificación de publicaciones), donde se indiquen las historias de usuario que se crearán para cada versión del programa y las fechas en las que se publicarán estas versiones. Un "Release plan" es una planificación donde los desarrolladores y clientes establecen los tiempos de implementación ideales de las historias de usuario, la prioridad con la que serán implementadas y las historias que serán implementadas en cada versión del programa.

Después de un "Release plan" tienen que estar claros estos cuatro factores: los objetivos que se deben cumplir (que son principalmente las historias que se deben desarrollar en cada versión), el tiempo que tardarán en desarrollarse y publicarse las versiones del programa, el número de personas que trabajarán en el desarrollo y cómo se evaluará la calidad del trabajo realizado.

➤ **Iteraciones:** Todo proyecto que siga la metodología XP se ha de dividir en iteraciones de aproximadamente 3 semanas de duración. Al comienzo de cada iteración los clientes deben seleccionar las historias de usuario definidas en el "Release planning" que serán implementadas.

En esta fase, también se seleccionan las historias de usuario que no pasaron el test de aceptación que se realizó al terminar la iteración anterior. Estas historias de usuario son divididas en tareas de entre 1 y 3 días de duración que se asignarán a los programadores.

➤ **Velocidad del proyecto:** La velocidad del proyecto es una medida que representa la rapidez con la que se desarrolla el proyecto; estimarla es muy sencillo, basta con contar el número de historias de usuario que se pueden implementar en una iteración; de esta forma, se sabrá el cupo de historias que se pueden desarrollar en las distintas iteraciones.

Usando la velocidad del proyecto controlaremos que todas las tareas se puedan desarrollar en el tiempo del que dispone la iteración. Es conveniente reevaluar esta medida cada 3 ó 4 iteraciones y si se aprecia que no es adecuada hay que negociar con el cliente un nuevo "Release Plan".

➤ **Programación en parejas:** La metodología X.P. aconseja la programación en parejas pues incrementa la productividad y la calidad del software desarrollado. El trabajo en pareja involucra a dos programadores trabajando en el mismo equipo; mientras uno codifica haciendo hincapié en la calidad de la función o método que está implementando, el otro analiza si ese método o función es adecuado y está bien diseñado. De esta forma se consigue un código y diseño con gran calidad.

➤ **Reuniones diarias:** Es necesario que los desarrolladores se reúnan diariamente y expongan sus problemas, soluciones e ideas de forma conjunta. Las reuniones tienen que ser fluidas y todo el mundo tiene que tener voz y voto.

**5.1.2 Fase de diseño:** En la fase de Diseño, se establecen unas recomendaciones o premisas para lograr un diseño simple pero funcional, estas recomendaciones son: elegir una metáfora para el sistema, usar tarjetas CRC, crear soluciones puntuales para reducir riesgos, trabajar con una funcionalidad mínima y hacer un buen reciclaje.

➤ **Metáfora del Sistema:** Es una historia en la que se relata al usuario el

funcionamiento básico del sistema; esta debe ser lo más clara posible para que le sea de fácil entendimiento al usuario.

En la metáfora se deben utilizar nombres que orienten la implantación del proyecto para no perder la esencia del mismo y reflejar lo que se quiere representar para el mundo real.

➤ **Tarjetas CRC:** Las tarjetas C.R.C (Clases, Responsabilidades y Colaboraciones) representan objetos. El uso de las tarjetas C.R.C permite al programador centrarse y apreciar el desarrollo orientado a objetos olvidándose de la clásica programación procedural.

La clase a la que pertenece el objeto se puede escribir en la parte de arriba de la tarjeta, en una columna a la izquierda se pueden escribir las responsabilidades u objetivos que debe cumplir el objeto y a la derecha, las clases que colaboran con cada responsabilidad.

En esta etapa hay que usar un glosario de términos y una correcta especificación de los nombres de métodos y clases que ayudará a comprender el diseño y facilitará sus posteriores ampliaciones y la reutilización del código.

Figura 74. Ejemplo de una Tarjeta CRC

|                                           |          |                                          |
|-------------------------------------------|----------|------------------------------------------|
| Nombre del Proyecto:                      |          |                                          |
| Nombre de la Clase: _____                 |          |                                          |
| <b>Responsabilidades (qué debe hacer)</b> | <b>R</b> | <b>Clases Asistentes (que colaboran)</b> |
|                                           |          |                                          |
|                                           |          |                                          |
|                                           |          |                                          |
|                                           |          |                                          |
| <b>Desarrollador</b>                      | _____    |                                          |
| <b>Observaciones:</b>                     |          |                                          |
|                                           |          |                                          |

**Fuente:** B I POWER BLOGGER. Metodología Extreme Programming(XP) parte I. [En línea] 28 de Abril de 2009. [Citado el: 18 de Enero de 2010.] [http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte\\_28.html](http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte_28.html). [3]

➤ **Soluciones Puntuales:** En la metodología XP sugiere que hay que conseguir diseños simples y sencillos. Hay que procurar hacerlo todo lo menos complicado posible para conseguir un diseño de fácil entendimiento e implementación, que a la larga costará menos tiempo y esfuerzo desarrollar.

Si surgen problemas potenciales durante el diseño, X.P sugiere utilizar una pareja de desarrolladores para que investiguen y reduzcan al máximo el riesgo que supone ese problema.

➤ **Funcionalidad Mínima:** Nunca se debe añadir funcionalidad extra al programa aunque se piense que en un futuro será utilizada. Sólo el 10% de la misma es utilizada, lo que implica que el desarrollo de funcionalidad extra es un desperdicio de tiempo y recursos.

➤ **Reciclaje:** Al momento de reutilizar código de otros proyectos (o del mismo) que haya sido creado para otro propósito, es mejorar y modificar la estructura y codificación (sacando un nuevo código) sin alterar su funcionalidad. En XP, a este proceso se le conoce como refactorizar.

Refactorizar supone revisar de nuevo estos códigos para procurar optimizar su funcionamiento. Es muy común rehusar códigos ya creados que contienen funcionalidades que no serán usadas y diseños obsoletos. Esto supone un error porque puede generar código completamente inestable y muy mal diseñado; por este motivo, es necesario refactorizar cuando se va a utilizar código ya creado.

**5.1.3 Fase de desarrollo:** La fase de Desarrollo reúne las características o cualidades en las que se basa todo el desarrollo general del proyecto. Está la disponibilidad del cliente, la unidad de pruebas, la programación en parejas y la integración de los programadores.

➤ **Disponibilidad del Cliente:** Como ya se dijo en la introducción, el cliente es una parte más del equipo de desarrollo; su presencia es indispensable en las distintas fases de X.P. A la hora de codificar una historia de usuario su presencia es aún más necesaria. No olvidemos que los clientes son los que crean las historias de usuario y negocian los tiempos en los que serán implementadas.

Antes del desarrollo de cada historia de usuario el cliente debe especificar detalladamente lo que ésta hará y también tendrá que estar presente cuando se realicen los test que verifiquen que la historia implementada cumple la funcionalidad especificada.

La codificación debe hacerse ateniendo a estándares de codificación ya creados. Programar bajo estándares mantiene el código consistente y facilita su comprensión y escalabilidad.

➤ **Unidad de Pruebas:** Crear test que prueben el funcionamiento de los distintos códigos implementados nos ayudará a desarrollar dicho código. Crear estos test antes nos ayuda a saber qué es exactamente lo que tiene que hacer el código a implementar y sabremos que una vez implementado pasará dichos test sin problemas ya que dicho código ha sido diseñado para ese fin.

Se puede dividir la funcionalidad que debe cumplir una tarea a programar en pequeñas unidades, de esta forma se crearán primero los test para cada unidad y a continuación se desarrollará dicha unidad, así poco a poco conseguiremos un desarrollo que cumpla todos los requisitos especificados.

➤ **Programación en Parejas:** Como ya se comentó anteriormente, X.P opta por la programación en pareja ya que permite un código más eficiente y con una gran calidad. X.P sugiere un modelo de trabajo usando repositorios de código dónde las parejas de programadores publican cada pocas horas sus códigos implementados y corregidos junto a los test que deben pasar. De esta forma el resto de programadores que necesiten códigos ajenos trabajarán siempre con las últimas versiones. Para mantener un código consistente, publicar un código en un repositorio es una acción exclusiva para cada pareja de programadores.

La optimización del código siempre se debe dejar para el final. Hay que hacer que funcione y que sea correcto, más tarde se puede optimizar.

➤ **Integración:** X.P también propone un modelo de desarrollo colectivo en el que todos los programadores están implicados en todas las tareas; cualquiera puede modificar o ampliar una clase o método de otro programador si es necesario y subirla al repositorio de código.

El permitir al resto de los programadores modificar códigos que no son suyos no supone ningún riesgo ya que para que un código pueda ser publicado en el repositorio tiene que pasar los test de funcionamiento definidos para el mismo.

**5.1.4 Fase de pruebas:** Las unidades o test de pruebas son parte indispensable de la Programación Extrema. Estas se convierten en una herramienta de desarrollo y material de apoyo, más no en un paso de verificación que puede despreciarse por más que parezca que el código esté funcionando correctamente.

Es por esto que la metodología XP exige constantes pruebas en el desarrollo, ya que descubrir todos los errores que puedan presentarse lleva tiempo y más si se dejan para el final del proyecto. En esta fase es importante tener en cuenta las siguientes características: implantación y pruebas de aceptación.

➤ **Implementación:** El proceso de implantación se realiza luego de garantizar que el código se encuentra completo. Lo que quiere decir que se han realizado las unidades de pruebas y se ha hecho una revisión de los requerimientos planteados por el cliente, que aseguran que el código está de acuerdo con la solicitud. El código será implantado cuando supere sus correspondientes test de unidades.

Un punto importante es usar test de unidad (que llegan a ser test de implementación y no de aceptación) que no tengan ninguna dependencia del

código que en un futuro evaluará. Hay que crear los test abstrayéndose del futuro código, de esta forma aseguraremos la independencia del test respecto al código que evalúa.

Los distintos test se deben subir al repositorio de código acompañados del código que verifican. Ningún código puede ser publicado en el repositorio sin que haya pasado su test de funcionamiento, de esta forma, aseguramos el uso colectivo del código.

➤ **Pruebas de aceptación:** Los test mencionados anteriormente sirven para evaluar las distintas tareas en las que ha sido dividida una historia de usuario. Para asegurar el funcionamiento final de una determinada historia de usuario se deben crear "Pruebas de aceptación".

Estas pruebas son creadas y usadas por los clientes para comprobar que las distintas historias de usuario cumplen su cometido.

Las pruebas de aceptación están basadas en las historias de usuario. Se hace un planteamiento de los puntos del desarrollo que deben probarse para corroborar que funcionen correctamente. Es para esto que se crean las pruebas de aceptación, determinando cuales son los aspectos que el cliente desea revisar y así garantizar su correcto funcionamiento. Por ello la importancia de crear cuantas pruebas de aceptación sean necesarias para que todo el trabajo realizado cuente con la calidad que exige.

Es fundamental desarrollar las pruebas de aceptación, porque es de esta forma que se demuestra el progreso y el proceso que se sigue en la ejecución del plan de iteraciones para las historias de usuario, ya que estas no se consideran completas hasta que no superan sus pruebas de aceptación.

## 6. RESULTADOS DE LA INVESTIGACIÓN

En el presente proyecto se aplicó la metodología XP teniendo en cuenta que sus características para el desarrollo de software son compatibles con las especificaciones propias de este. **POLARIS VERSIÓN 2.0** es un proyecto a corto plazo y cuenta con un grupo de trabajo en el que participan dos desarrolladores, además, se trabaja en equipo con el cliente como elemento esencial para el planteamiento de los requerimientos y el éxito del proyecto.

En esta sección se detallan todos los resultados obtenidos al aplicar las fases de la metodología, brindándole al lector una visión general del trabajo realizado y del proceso de construcción que se utilizó para el desarrollo de **POLARIS VERSIÓN 2.0**.

Es necesario recordar que, los procesos de cada una de las fases de la metodología, no son pasos estrictos que deben darse uno tras otro. Más bien, se tratan de recomendaciones que se relacionan entre sí para la comprensión total de la fase en cuestión.

Debido a eso, en lugar de exponer en detalle los procesos de cada fase, se mencionarán los resultados de los procesos más significativos de éstas, pero haciendo alusión a los demás sub puntos inmersos en la etapa en cuestión. Lo que permitirá ser más específicos y poco redundantes, en la redacción de los resultados de la investigación.

### 6.1 FASE DE PLANEACIÓN

En esta fase se realizó toda la planeación general de proyecto, el Ingeniero Ricardo Timaran Pereira, Ph. D. hizo las veces de usuario final y, junto al grupo de trabajo, relato los requerimientos del proyecto. Con base en estas historias de usuario, se realizó la distribución de tareas, el cronograma de entregas y la programación de las reuniones que se muestran a continuación.

**6.1.1 Historias de usuario:** De acuerdo a las especificaciones del Ingeniero Ricardo Timaran Pereira, Ph. D. director del proyecto, se realizaron 37 historias de usuario basadas en los requerimientos principales del proyecto. Se asignó un número y nombre a cada historia, así como el desarrollador responsable y la fecha de entrega prevista.

Las historias de usuario planteadas se describen a continuación.

➤ **Historia de usuario No. 1 - Aplicación de Escritorio.**

Tabla 1. Historia de usuario No. 1 - Aplicación de Escritorio.

| <b>HISTORIA DE USUARIO No. 1 – APLICACIÓN DE ESCRITORIO</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 24 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                          |
| <p>El programa que se desea desarrollar es la segunda versión de una herramienta de Minería Web, en la cual se pretende implementar el módulo de Minería Web de Estructura, logrando con esto ampliar el campo de acción de dicha herramienta que su versión inicial únicamente cuenta con Minería Web de Uso. Esto se hace con la finalidad de encaminarla a una suite de Minería Web, a la cual le haría falta únicamente la implementación de la Minería Web de Contenido.</p> <p>Esta herramienta debe ser una aplicación de escritorio, la cual podrá ser instalada sobre cualquier sistema operativo y en cualquier computador que cumpla unos requisitos mínimos de sistema.</p> <p>Para el funcionamiento del módulo de Minería Web de Estructura es necesario que se tenga un acceso a Internet.</p> |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                          |
| <p>Esta aplicación va dirigida principalmente a los Web Master o Administradores de portales Web, que necesiten estudiar y mejorar la estructura de enlaces de sus páginas y la navegabilidad de las mismas. Además esta aplicación puede ser de gran uso para cualquier empresa que maneje un sitio Web.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Es conforme              |

➤ **Historia de usuario No. 2 - Creación del módulo de minería de estructura.**

Tabla 2. Historia de usuario No. 2 - Creación del módulo de minería de estructura.

| <b>HISTORIA DE USUARIO No. 2 – CREACIÓN DEL MÓDULO DE MINERÍA DE ESTRUCTURA WEB</b>                                                                                                                                                                  |                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                     | Cristian Iles       |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                     | 25 de enero de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                       |                     |
| <p>Investigación, Análisis e implementación del módulo de Minería de Estructura Web, para la herramienta POLARIS.</p>                                                                                                                                |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                             |                     |
| <p>Este módulo va dirigido a los Web Master a la hora de implementar o mejorar los enlaces dentro de un sitio Web, ya que permite tomar decisiones teniendo en cuenta los cálculos del peso de cada enlace. El diseño y la explicación del nuevo</p> |                     |

módulo se encuentran en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 3 - Agregar tablas a la base de datos de Polaris.**

Tabla 3. Historia de usuario No. 3 - Agregar tablas a la base de datos de Polaris.

**HISTORIA DE USUARIO No. 3 – AGREGAR TABLAS A LA BASE DE DATOS DE POLARIS**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 10 de febrero de 2010

**Descripción de la Tarea**

Aumentar tablas a la estructura de la base de datos de la herramienta POLARIS en su primera versión, las cuales serán necesarias para el almacenamiento de la información recolectada por el nuevo módulo de Minería de Estructura Web.

**Notas Adicionales**

El diseño y la explicación de las nuevas tablas a aumentar tablas de la base de datos se encuentran en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 4 - Creación de un Web Crawler.**

Tabla 4. Historia de usuario No. 4 - Creación de un Web Crawler.

**HISTORIA DE USUARIO NO. 4 – CREACIÓN DE UN WEB CRAWLER**

**Desarrollador responsable** Carlos Rubio

**Fecha de entrega prevista** 20 junio de 2010

**Descripción de la Tarea**

Investigación, Análisis e implementación de un Web Crawler, el cual es un mecanismo que permite la Recuperación de Información (RI) para el módulo de Minería de Estructura Web, con el cual se obtiene los datos necesarios para que el módulo a implementar trabaje de manera adecuada.

Para el funcionamiento del Web Crawler se debe tener una conexión a Internet ya que desde la Web se extrae la información necesaria para lograr su funcionamiento inicial, para esto se debe poseer ciertos filtros y parámetros para la obtención de la información.

**Notas Adicionales**

El Web Crawler es indispensable para el funcionamiento de la herramienta POLARIS V2.0. El diseño y la explicación del Web Crawler se encuentran en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 5 - Nivel de profundidad del Web Crawler.**

Tabla 5. Historia de usuario No. 5 - Nivel de profundidad del Web Crawler.

| <b>HISTORIA DE USUARIO NO. 5 – NIVEL DE PROFUNDIDAD DEL WEB CRAWLER</b>                                                                                                                  |                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                         | Carlos Rubio        |
| <b>Fecha de entrega prevista</b>                                                                                                                                                         | 25 de junio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                           |                     |
| El Web Crawler necesita recibir como parámetro un número, el cual indica el nivel de profundidad que se debe tener al momento de recolectar la información para la generación del grafo. |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                 |                     |
| El nivel de profundidad debe ser mayor o igual a 1 y se da en números enteros.                                                                                                           |                     |
| <b>Estado de implementación</b>                                                                                                                                                          | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                           | Es conforme         |

➤ **Historia de usuario No. 6 - URL a analizar.**

Tabla 6. Historia de usuario No. 6 - URL a analizar.

| <b>HISTORIA DE USUARIO NO. 6 – URL A ANALIZAR</b>                                                                                                                                                                                                                                     |                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                      | Carlos Rubio        |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                      | 28 de junio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                        |                     |
| El Web Crawler necesita recibir como parámetro una dirección Web, la cual va a ser objeto de estudio, esta será tomada como la página central o de inicio de la cual se iniciara a desprender el grafo en su totalidad dependiendo del nivel de profundidad que se haya especificado. |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                              |                     |
| La URL que se digite debe ser válida para iniciar el análisis, es indispensable encontrarse conectado a internet.                                                                                                                                                                     |                     |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                       | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                        | Es conforme         |

➤ **Historia de usuario No. 7 - Filtros para la búsqueda del Crawler.**

Tabla 7. Historia de usuario No. 7 - Filtros para la búsqueda del Crawler.

| <b>HISTORIA DE USUARIO NO. 7 – FILTROS PARA LA BÚSQUEDA DEL CRAWLER</b>                                                                                                                                        |                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                               | Carlos Rubio        |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                               | 15 de julio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                 |                     |
| En el momento que el Web Crawler de inicio a la recolección de información desde internet, se debe establecer unos filtros, los cuales permitan escoger el tipo de protocolo con los cuales se desea trabajar. |                     |
| Algunos ejemplos de estos protocolos son HTTP, HTTPS, FTP. Se debe brindar la oportunidad al usuario de trabajar únicamente con uno de ellos, con varios o todos si es el caso.                                |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                       |                     |
| El grafo se genera con base a las páginas que le Web Crawler haya recolectado y estas dependen a su vez del filtro establecido, esto se hace para limitar la búsqueda en la Web.                               |                     |
| <b>Estado de implementación</b>                                                                                                                                                                                | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                 | Es conforme         |

➤ **Historia de usuario No. 8 - Guardar datos del Web Crawler.**

Tabla 8. Historia de usuario No. 8 - Guardar datos del Web Crawler.

| <b>HISTORIA DE USUARIO NO. 8 – GUARDAR DATOS DEL WEB CRAWLER</b>                                                                                                                                             |                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                             | Carlos Rubio        |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                             | 22 de julio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                               |                     |
| Cuando el usuario mire conveniente, debe tener la posibilidad de guardar la información recolectada por el Web Crawler, la cual se almacenara en la nueva estructura de la base de datos previamente creada. |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                     |                     |
| Esta información quedara almacenada hasta que el usuario necesite nuevamente de ella.                                                                                                                        |                     |
| <b>Estado de implementación</b>                                                                                                                                                                              | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                                               | Es conforme         |

➤ **Historia de usuario No. 9 - Implementación de los algoritmos de Minería Web de Estructura.**

Tabla 9. Historia de usuario No. 9 - Implementación de los algoritmos de Minería Web de Estructura.

| <b>HISTORIA DE USUARIO NO. 9 – IMPLEMENTACION DE LOS ALGORITMOS DE MINERÍA WEB DE ESTRUCTURA</b>                                                                                                                                   |                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                   | Cristian Iles      |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                   | 30 de mayo de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                     |                    |
| Una vez se haya realizado el respectivo análisis y desarrollo de los algoritmos de Minería de Estructura Web, deben ser implementados dentro del nuevo módulo que se ha desarrollado.                                              |                    |
| Estos algoritmos deben ser acoplados de tal manera que no interfieran en la Minería de Uso implementada en la primera versión de la herramienta POLARIS.                                                                           |                    |
| Los algoritmos implementados son los encargados de realizar los cálculos y de dar la calificación de la relevancia de cada página, dependiendo de la cantidad de enlaces entrantes y salientes y si estos son internos o externos. |                    |
| <b>Notas Adicionales</b>                                                                                                                                                                                                           |                    |
| Ninguna                                                                                                                                                                                                                            |                    |
| <b>Estado de implementación</b>                                                                                                                                                                                                    | Implementada       |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                     | Es conforme        |

➤ **Historia de usuario No. 10 - Visualización de grafos.**

Tabla 10. Historia de usuario No. 10 - Visualización de grafos.

| <b>HISTORIA DE USUARIO No. 10 – VISUALIZACIÓN DE GRAFOS</b>                                                                                                                                                                       |                      |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                  | Carlos Rubio         |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                  | 12 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                    |                      |
| La visualización del grafo se podrá realizar en cualquier momento, por ejemplo se podrá realizar después de la recolección del Web Crawler, o también de aplicar el algoritmo escogido para observar los datos que son arrojados. |                      |
| El grafo debe ser el reflejo de la red, partiendo de la URL digitada y del nivel de profundidad que se le haya establecido al Web Crawler en el momento de realizar la estructura.                                                |                      |
| La visualización puede darse de diferentes maneras, a conveniencia del usuario, se debe poder filtrar por los tipos de enlaces, si son internos, externos, entrantes o                                                            |                      |

salientes, dependiendo de las necesidades del usuario y de lo que se vaya a analizar.

#### **Notas Adicionales**

El grafo es de gran utilidad para el usuario de la herramienta, ya que en este se refleja todos los datos que se hayan obtenido al aplicar los algoritmos de Minería de Estructura Web y además le da una idea visual de la red partiendo de la URL que ingresó.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

#### ➤ **Historia de usuario No. 11 - Mejorar la interfaz de Polaris V 1.0.**

Tabla 11. Historia de usuario No. 11 - Mejorar la interfaz de Polaris V 1.0.

### **HISTORIA DE USUARIO No. 11 – MEJORAR LA INTERFAZ DE POLARIS V 1.0**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 5 de marzo de 2010

#### **Descripción de la Tarea**

El programa que se desea desarrollar es la segunda versión de una herramienta de Minería Web de Uso, en la cual se pretende implementar el módulo de Minería Web de Estructura, por tal motivo es prescindible modificar la interfaz de la primera versión de POLARIS, haciéndola más dinámica e interactiva con el usuario.

#### **Notas Adicionales**

Aunque la aplicación a desarrollar va dirigida a personas que tengan conocimiento sobre la estructura de un sitio Web, debe tener una interfaz amigable para cualquier tipo de usuario. El diseño y la explicación de la modificación de la interfaz se encuentran en la sección Metáfora del Sistema de la Fase de Diseño.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

#### ➤ **Historia de usuario No. 12 - Integración de la interfaz.**

Tabla 12. Historia de usuario No. 12 - Integración de la interfaz.

### **HISTORIA DE USUARIO No. 12 – INTEGRACIÓN DE LA INTERFAZ**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 1 de septiembre de 2010

#### **Descripción de la Tarea**

La interfaz debe ser integrada de tal manera que tengan diseño igual o similar los dos módulos (módulo de minería de Uso Web y módulo de minería de Estructura Web) de la herramienta POLARIS.

#### **Notas Adicionales**

Permite que la interfaz sea una sola para manejar los dos módulos dentro de la

---

aplicación POLARIS

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

---

➤ **Historia de usuario No. 13. - Integración de los algoritmos.**

Tabla 13. Historia de usuario No. 13. - Integración de los algoritmos.

**HISTORIA DE USUARIO No. 13 – INTEGRACIÓN DE LOS ALGORITMOS**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 31 de agosto de 2010

**Descripción de la Tarea**

Debe integrarse los algoritmos de minería de estructura Web al módulo que se quiere crear, estos algoritmos son el mecanismo para que el proceso de minería de estructura se realice de forma correcta.

**Notas Adicionales**

La implementación e integración es fundamental de los algoritmos de minería de estructura es fundamental para el módulo q se desea implementar.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

---

➤ **Historia de usuario No. 14 - Preprocesamiento de la información recolectada por el Web Crawler.**

Tabla 14. Historia de usuario No. 14 - Preprocesamiento de la información recolectada por el Web Crawler.

**HISTORIA DE USUARIO No. 14 – PREPROCESAMIENTO DE LA INFORMACIÓN RECOLECTADA POR EL WEB CRAWLER**

**Desarrollador responsable** Carlos Rubio

**Fecha de entrega prevista** 2 de julio de 2010

**Descripción de la Tarea**

Antes de empezar a trabajar con los datos recolectados por el Web Crawler es necesario realizar un preprocesamiento a toda la información, en pocas palabras dividirla para hacerla más eficiente.

**Notas Adicionales**

El preprocesamiento de la información dará información importante para trabajar con los datos.

En este momento la información puede ser filtrada para trabajar con los datos que se requiera.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

---

➤ **Historia de usuario No. 15 - Enlaces internos y externos.**

Tabla 15. Historia de usuario No. 15 - Enlaces internos y externos.

| <b>HISTORIA DE USUARIO No. 15 – ENLACES INTERNOS Y EXTERNOS</b>                                                                                                                                                                                                           |                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                          | Carlos Rubio       |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                          | 8 de julio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                            |                    |
| Se debe reconocer los enlaces internos y externos de un sitio Web, de este modo a la hora de la visualización se podrá observar independientemente como se conforma un sitio Web (enlaces internos) y con quien se conecta o a que enlaces se conecta (enlaces externos). |                    |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                  |                    |
| Permite observar gráficamente el grafo de la estructura interna del sitio Web y la externa a ella, aunque si el usuario desea puede observar todo el grafo.                                                                                                               |                    |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                           | Implementada       |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                            | Es conforme        |

➤ **Historia de usuario No. 16 - Enlaces entrantes y salientes.**

Tabla 16. Historia de usuario No. 16 - Enlaces entrantes y salientes.

| <b>HISTORIA DE USUARIO No. 16 – ENLACES ENTRANTES Y SALIENTES</b>                                                                                                                                                                                   |                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                    | Carlos Rubio        |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                    | 10 de julio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                      |                     |
| Se debe reconocer los enlaces salientes y entrantes de cada uno de los nodos (páginas Web). Hacer una división de los enlaces salientes y entrantes de los Internos y externos del sitio Web, para poder visualizar todo los grafos independientes. |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                            |                     |
| Este preprocesamiento es muy importante para el buen funcionamiento de los algoritmos, ya que ellos trabajan con los enlaces salientes y entrantes de cada uno de los nodos, independientemente si son Externos o Internos del sitio Web.           |                     |
| <b>Estado de implementación</b>                                                                                                                                                                                                                     | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                      | Es conforme         |

➤ **Historia de usuario No. 17 - Cargar información de base de datos**

Tabla 17. Historia de usuario No. 17 - Cargar información de base de datos

| <b>HISTORIA DE USUARIO No. 17 – CARGAR INFORMACIÓN DE BASE DE DATOS</b>                                      |                     |
|--------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                             | Cristian Iles       |
| <b>Fecha de entrega prevista</b>                                                                             | 1 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                               |                     |
| La información guardada en la base de datos debe recuperarse en cada uno de los procesos que se la necesite. |                     |
| <b>Notas Adicionales</b>                                                                                     |                     |
| El preprocesamiento de la información dará información importante para trabajar con los datos.               |                     |
| <b>Estado de implementación</b>                                                                              | Implementada        |
| <b>Conformidad del Cliente</b>                                                                               | Es conforme         |

➤ **Historia de usuario No. 18 - Visualización de grafos con sus filtros.**

Tabla 18. Historia de usuario No. 18 - Visualización de grafos con sus filtros.

| <b>HISTORIA DE USUARIO No. 18 – VISUALIZACIÓN DE GRAFOS CON SUS FILTROS</b>                                                                                                                              |                      |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                         | Carlos Rubio         |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                         | 15 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                           |                      |
| La visualización de la información deberá realizarse de una forma dinámica y que permita observar el grafo (verlo gráficamente sus nodos y aristas (enlaces)) de los datos obtenidos por el Web Crawler. |                      |
| <b>Notas Adicionales</b>                                                                                                                                                                                 |                      |
| El diseño para la visualización se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                     |                      |
| <b>Estado de implementación</b>                                                                                                                                                                          | Implementada         |
| <b>Conformidad del Cliente</b>                                                                                                                                                                           | Es conforme          |

➤ **Historia de usuario No. 19 - Visualización de grafos de los enlaces internos.**

Tabla 19. Historia de usuario No. 19 - Visualización de grafos de los enlaces internos.

| <b>HISTORIA DE USUARIO No. 19 – VISUALIZACIÓN DE GRAFOS DE LOS ENLACES INTERNOS</b>                                                                                                                                                                                                                                                                                                                                                       |                      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                          | Carlos Rubio         |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                          | 10 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                            |                      |
| La visualización de la información deberá realizarse de una forma dinámica y que permita observar el grafo (verlo gráficamente sus nodos y aristas (enlaces)), en esta parte la visualización solo mostrara los enlaces internos de la URL estudiada, se podrá definir los datos que realmente sean de interés para el usuario como los enlaces salientes y entrantes, la visualización se hará totalmente independiente o conjuntamente. |                      |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                  |                      |
| Permite observar independientemente en forma gráficas el grafo interno del sitio Web.                                                                                                                                                                                                                                                                                                                                                     |                      |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                           | Implementada         |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                            | Es conforme          |

➤ **Historia de usuario No. 20 - Visualización de grafos de los enlaces externos.**

Tabla 20. Historia de usuario No. 20 - Visualización de grafos de los enlaces externos.

| <b>HISTORIA DE USUARIO No. 20 – VISUALIZACIÓN DE GRAFOS DE LOS ENLACES EXTERNOS</b>                                                                                                                                                                                                                                                                                                                                                                                                                                 |                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Carlos Rubio         |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 13 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                      |
| La visualización de la información deberá realizarse de una forma dinámica y que permita observar el grafo (verlo gráficamente sus nodos y aristas (enlaces)), en esta parte la visualización solo mostrara los enlaces externos de la Url estudiada, se podrá definir los datos que realmente sean de interés para el usuario como los enlaces salientes y entrantes, la visualización se hará totalmente independiente (esto quiere decir que se podrán observar los enlaces salientes ó los entrantes, o ambos). |                      |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                      |
| Permite observar independientemente en forma gráficas el grafo únicamente con                                                                                                                                                                                                                                                                                                                                                                                                                                       |                      |

---

los enlaces y los nodos externos del sitio Web.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

---

➤ **Historia de usuario No. 21- Configuración de los algoritmos.**

Tabla 21. Historia de usuario No. 21- Configuración de los algoritmos.

| <b>HISTORIA DE USUARIO No. 21 – CONFIGURACIÓN DE LOS ALGORITMOS</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Cristian Iles       |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 10 de abril de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                     |
| Cada algoritmo implementado tienes sus propias configuraciones y parámetros para empezar a trabajar con la información recolectada, esta configuración puede depender del número de iteraciones que se desean realizar o la certeza de confianza que se desea en la información entre más baja, el cálculo de los algoritmos se hará menos confiable pero mucho más rápido, que cuando se maneja una alta confianza y mayor iteraciones porque la confiabilidad es mucho más alta pero se excede en el tiempo del proceso por parte de la aplicación. |                     |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                     |
| El diseño para la configuración de cada uno de los algoritmos implementados se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                                                                                                                                                                                                                                                                                                                      |                     |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Implementada        |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Es conforme         |

---

➤ **Historia de usuario No. 22 - Parámetros Page Rank**

Tabla 22. Historia de usuario No. 22 - Parámetros Page Rank

| <b>HISTORIA DE USUARIO No. 22 – PARÁMETROS PAGE RANK</b>                                                                                                                                                                                                                                                                            |                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                    | Cristian Iles       |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                    | 22 de abril de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                      |                     |
| Este algoritmo recibe dos parámetros para empezar a trabajar uno de ellos es la cantidad de iteraciones que va a realizar este valor es un número entero que está entre 70 y 100, entre más alto este factor solo nos indica que los resultados que se observarán serán mucho más exactos que con una menor cantidad de iteraciones |                     |
| El otro parámetro es el factor de decaimiento (Dampening), el cual es el valor de confianza de los resultados que se van a obtener, este valor debe de ser mayor a cero y menor a uno. Se puede recomendar trabajar con un Dampening de 0.85 y un número real con dos decimales y dos cifras significativas.                        |                     |

---

#### Notas Adicionales

El diseño para la configuración del algoritmo Page Rank se encuentra en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

#### ➤ Historia de usuario No. 23 - Parámetros HITS.

Tabla 23. Historia de usuario No. 23 - Parámetros HITS.

#### HISTORIA DE USUARIO No. 23 – PARÁMETROS HITS

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 8 de mayo de 2010

#### Descripción de la Tarea

Este algoritmo recibe como parámetro el número de iteraciones con los cuales va a trabajar. Se puede recomendar que el número de iteraciones mínimas sea de 25 esto para que los valores se han aprobado y el número de iteraciones máxima sea de 50 para que el procesamiento por parte de la aplicación no lleve demasiado tiempo.

#### Notas Adicionales

El diseño para la configuración del algoritmo HITS se encuentra en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

#### ➤ Historia de usuario No. 24 - Parámetros algoritmo creado.

Tabla 24. Historia de usuario No. 24 - Parámetros algoritmo creado.

#### HISTORIA DE USUARIO No. 24 – PARÁMETROS ALGORITMO LINKER WEB

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 20 de mayo de 2010

#### Descripción de la Tarea

Es tiempo.

#### Notas Adicionales

El diseño para la configuración del algoritmo HITS se encuentra en la sección *Metáfora del Sistema* de la *Fase de Diseño*.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 25 - Visualización de información de cada nodo.**

Tabla 25. Historia de usuario No. 25 - Visualización de información de cada nodo.

| <b>HISTORIA DE USUARIO No. 25 – VISUALIZACIÓN DE INFORMACIÓN DE CADA NODO</b>                                                                                                                                                                                                                                                                                                                                                   |                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                | Carlos Rubio            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                | 5 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                  |                         |
| Al ser visualizado el grafo en cualquier momento también se puede seleccionar uno de cualquiera de los nodos en su representación gráfica para observar la información correspondiente al nodo seleccionado, esta información será la recolectada por el Web Crawler y en ella se podrán ver el logo o nombre del sitio al que pertenece el logo, el nombre de su enlace o link, entre otra información relevante de cada nodo. |                         |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                        |                         |
| El diseño para la visualización de información de cada nodo se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                                                                                                                                                                                                                |                         |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                 | Implementada            |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                  | Es conforme             |

➤ **Historia de usuario No. 26 - Cálculo de los algoritmos.**

Tabla 26. Historia de usuario No. 26 - Cálculo de los algoritmos.

| <b>HISTORIA DE USUARIO No. 26 – CÁLCULO DE LOS ALGORITMOS</b>                                                                                                                                                                   |                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                | Cristian Iles      |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                | 1 de junio de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                  |                    |
| Cada uno de los algoritmos implementados tiene sus cálculos para poder dar el ponderado (el peso) a cada nodo, el cómputo de los algoritmos se hace a través de varias iteraciones y de una fórmula matemática antes planteada. |                    |
| <b>Notas Adicionales</b>                                                                                                                                                                                                        |                    |
| La implementación de los algoritmos y su respectivo diseño se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                 |                    |
| <b>Estado de implementación</b>                                                                                                                                                                                                 | Implementada       |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                  | Es conforme        |

➤ **Historia de usuario No. 27 - Visualización del grafo final.**

Tabla 27. Historia de usuario No. 27 - Visualización del grafo final.

| <b>HISTORIA DE USUARIO No. 27 – VISUALIZACIÓN DEL GRAFO FINAL</b>                                                                                                                                                                 |                          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                  | Carlos Rubio             |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                  | 15 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                    |                          |
| En el momento de que uno de los algoritmos sea seleccionado para calcular el peso, de cada nodo del grafo a estudiar, se debe visualizar gráficamente el grafo con los pesos correspondientes al valor agregado por el algoritmo. |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                          |                          |
| El diseño para la visualización del grafo final se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                              |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                   | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                    | Es conforme              |

➤ **Historia de usuario No. 28 - Algoritmo de la ruta más corta entre dos nodos seleccionados.**

Tabla 28. Historia de usuario No. 28 - Algoritmo de la ruta más corta entre dos nodos seleccionados.

| <b>HISTORIA DE USUARIO No. 28 – ALGORITMO DE LA RUTA MÁS CORTA PARA DOS NODOS SELECCIONADOS</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Cristian Iles      |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 30 de mayo de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                    |
| Implementar un algoritmo que nos permita realizar el cálculo de la ruta más corta entre dos nodos seleccionados, teniendo en cuenta los pesos calculados por los algoritmos implementados y así poder dar soporte a la toma de decisiones del sitio Web que se encuentre en análisis. Al disponer de una visualización para el grafo final, se permite una interacción en tiempo real con el usuario final, de esta forma puede seleccionar dos nodos diferentes a los cuales dese aplicarles el algoritmo de la ruta más corta, con la finalidad de terminar con el proceso de minería de estructura Web. |                    |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                    |
| El diseño del algoritmo para la ruta más corta se encuentra en la sección <i>Metáfora del Sistema</i> de la <i>Fase de Diseño</i> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                    |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Implementada       |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Es conforme        |

➤ **Historia de usuario No. 29 - Creación de filtros según el protocolo de las páginas Web.**

Tabla 29. Historia de usuario No. 29 - Creación de filtros según el protocolo de las páginas Web.

| <b>HISTORIA DE USUARIO No. 29 – CREACIÓN DE FILTROS SEGÚN EL PROTOCOLO DE LAS PÁGINAS WEB</b>                                                                                                                                                                                                                 |                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                              | Carlos Rubio         |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                              | 10 de agosto de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                |                      |
| <p>Crear un filtro que permita escoger con que páginas se desea trabajar siendo estas clasificadas por su protocolo. Este tipo de filtro no será único, ya que nos brinda la facilidad de trabajar con uno o más tipos de protocolos y si el usuario final lo desea obviar lógicamente se lo puede hacer.</p> |                      |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                      |                      |
| Ninguna.                                                                                                                                                                                                                                                                                                      |                      |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                               | Implementada         |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                | Es conforme          |

➤ **Historia de usuario No. 30 - Reporte en tablas de los cálculos.**

Tabla 30. Historia de usuario No. 30 - Reporte en tablas de los cálculos.

| <b>HISTORIA DE USUARIO No. 30 – REPORTE EN TABLAS DE LOS CÁLCULOS</b>                                                                                                                                                                                                                                                                                                                                              |                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                   | Cristian Iles           |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                   | 1 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                     |                         |
| <p>Una vez se hayan realizado todos los cálculos pertinentes, estos serán exportados a tablas en las cuales se reflejan cifras de los respectivos resultados obtenidos después de haber sido procesados los diferentes algoritmos implementados. Estas tablas irán acompañadas de gráficas estadísticas donde se mirara reflejada la importancia de las diferentes páginas y la concurrencia que estas tienen.</p> |                         |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                           |                         |
| <p>Estas tablas son fáciles de entender e interpretar, serán únicamente de lectura con la finalidad de no ser manipuladas por parte del usuario.</p>                                                                                                                                                                                                                                                               |                         |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                    | Implementada            |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                     | Es conforme             |

➤ **Historia de usuario No. 31 - Exportación de datos a otros formatos.**

Tabla 31. Historia de usuario No. 31 - Exportación de datos a otros formatos.

| <b>HISTORIA DE USUARIO No. 31 – EXPORTACIÓN DE DATOS A OTROS FORMATOS</b>                                                                                                                                                                                                                                                                                                                                                                                                           |                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 13 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                          |
| <p>La herramienta está dispuesta a exportar los datos en diferentes tipos de formatos de archivo, esto con la finalidad de poder llevar la información de un lugar a otro de manera fácil y servir como soporte o anexos en la presentación de un documento para un Web Máster.</p> <p>Además de esta manera se facilita la realización de análisis posteriores, sin necesidad de estar ejecutando el software varias veces para obtener resultados que ya se habían calculado.</p> |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                          |
| <p>Se debe tener en cuenta que la información en la Web se mantiene en constante actualización y modificación por lo cual estos datos que se exportan pueden variar de un día para otro.</p>                                                                                                                                                                                                                                                                                        |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Es conforme              |

➤ **Historia de usuario No. 32 - Resultados en la mejora de la estructura.**

Tabla 32. Historia de usuario No. 32 - Resultados en la mejora de la estructura.

| <b>HISTORIA DE USUARIO No. 32 – RESULTADOS EN LA MEJORA DE LA ESTRUCTURA</b>                                                                                                                                                                                                                                                                                                                                                                                                             |                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 20 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                          |
| <p>Una vez se haya realizado la respectiva aplicación de los diferentes algoritmos de minería de estructura Web o de la aplicación de uno en particular, se procede a obtener los resultados que la herramienta indica acerca de la mejora en la estructura del sitio Web que se está analizando.</p> <p>Estos resultados son notificaciones que conllevan a mejorar la navegabilidad del usuario final de la página Web, haciéndole más fácil la búsqueda de información relevante.</p> |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                          |
| <p>La herramienta únicamente notifica al usuario no realiza modificación alguna sobre</p>                                                                                                                                                                                                                                                                                                                                                                                                |                          |

la estructura de la página, las modificaciones sugeridas deben ser tenidas en cuenta por el Web Máster y puestas en marcha.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 33. Eliminar datos de la base de datos.**

Tabla 33. Historia de usuario No. 33. Eliminar datos de la base de datos.

**HISTORIA DE USUARIO No. 33 – ELIMINAR DATOS DE LA BASE DE DATOS**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 25 de agosto de 2010

**Descripción de la Tarea**

Cuando los datos que se encuentran dentro de la base de datos de la herramienta ya no es necesaria, se puede eliminar de manera permanente, con la finalidad de dar cabida a nueva información que sea útil para el usuario.

**Notas Adicionales**

Una vez se eliminen datos el proceso no podrá ser revertido, por lo cual el usuario debe estar seguro de lo que va a borrar.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 34. Creación de un nuevo proyecto.**

Tabla 34. Historia de usuario No. 34. Creación de un nuevo proyecto.

**HISTORIA DE USUARIO No. 34 – CREACIÓN DE UN NUEVO PROYECTO**

**Desarrollador responsable** Cristian Iles

**Fecha de entrega prevista** 22 de septiembre de 2010

**Descripción de la Tarea**

Cada vez que se realice en análisis a una URL diferente, se debe crear un nuevo proyecto, esto con la finalidad de diferenciar la información y mantenerla de forma clara para ser analizada de la manera más adecuada.

La creación de un nuevo proyecto se debe realizar por medio del menú de la herramienta que ofrece diferentes opciones dependiendo de las necesidades del usuario.

**Notas Adicionales**

Ninguna.

**Estado de implementación** Implementada

**Conformidad del Cliente** Es conforme

➤ **Historia de usuario No. 35 - Cargar un proyecto existente.**

Tabla 35. Historia de usuario No. 35 - Cargar un proyecto existente.

| <b>HISTORIA DE USUARIO No. 35 – CARGAR UN PROYECTO EXISTENTE</b>                                                                                                                                                                                                                                                                                                                                                               |                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                                                               | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                                                                                               | 22 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                                                                                                 |                          |
| <p>Todo análisis que se realice en la herramienta y haya sido guardado previamente se podrá cargar nuevamente, el cual se presentará al usuario justo de la misma manera como lo dejo en el momento que guardo modificaciones por última vez.</p> <p>Para cargar un proyecto existente se lo debe realizar por medio del menú de la herramienta que ofrece diferentes opciones dependiendo de las necesidades del usuario.</p> |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                                                                                                       |                          |
| Ninguna.                                                                                                                                                                                                                                                                                                                                                                                                                       |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                                                                                                | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                                                                                                 | Es conforme              |

➤ **Historia de usuario No. 36 - Guardar proyecto.**

Tabla 36. Historia de usuario No. 36 - Guardar proyecto.

| <b>HISTORIA DE USUARIO No. 36 – GUARDAR PROYECTO</b>                                                                                                                                                                                                                                                                                                   |                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                       | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                                                                                                                       | 23 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                                                                                                                         |                          |
| <p>En el momento que el usuario lo desee tiene la posibilidad de guardar el proyecto no importa que este se encuentre inconcluso o se haya terminado el análisis por completo.</p> <p>Para guardar un proyecto se lo debe realizar por medio del menú de la herramienta que ofrece diferentes opciones dependiendo de las necesidades del usuario.</p> |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                                                                                                                               |                          |
| La opción guardar sobrescribe el proyecto actual no saca una copia diferentes.                                                                                                                                                                                                                                                                         |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                                                                                                                                        | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                                                                                                                         | Es conforme              |

➤ **Historia de usuario No. 37 - Guardar proyecto como.**

Tabla 37. Historia de usuario No. 37 - Guardar proyecto como.

| <b>HISTORIA DE USUARIO No. 37 – GUARDAR PROYECTO COMO</b>                                                                                                                                                                                       |                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                | Cristian Iles            |
| <b>Fecha de entrega prevista</b>                                                                                                                                                                                                                | 23 de septiembre de 2010 |
| <b>Descripción de la Tarea</b>                                                                                                                                                                                                                  |                          |
| Esta opción le permitirá al usuario guardar el proyecto como uno nuevo, es decir obtener una copia. Esta opción será útil para hacer modificaciones a diferentes proyectos de los cuales no se desea perder la información actual que se tiene. |                          |
| <b>Notas Adicionales</b>                                                                                                                                                                                                                        |                          |
| Esta opción no sobrescribe el proyecto actual sino que por el contrario saca una copia en el estado actual del mismo.                                                                                                                           |                          |
| <b>Estado de implementación</b>                                                                                                                                                                                                                 | Implementada             |
| <b>Conformidad del Cliente</b>                                                                                                                                                                                                                  | Es conforme              |

## 6.2 PLAN DE ENTREGAS

En la reunión de planificación se realizó el cronograma de actividades o plan de entregas del proyecto, estas entregas se programaron de acuerdo a las historias de usuario. Se estableció el tiempo de inicio y finalización de cada una de las historias. Además, se plantearon cuáles serían las versiones del proyecto, en qué fechas se publicarían y qué historias tendrían implementadas.

**6.2.1 Cronograma de actividades:** A continuación se indica el cronograma de actividades que se tuvo en cuenta para la implementación de las historias de usuario.

Aunque el tiempo ideal para el desarrollo de una Historia de Usuario es de 1 a 3 semanas, en este proyecto se plantearon algunas historias con tiempos de duración que ascendían a varios meses. Esto se debe a que aquellas historias presentaban tareas algo largas que sólo se podían ir implementando en el transcurso del tiempo que, para su implementación, dependían de la implementación de otras historias de usuario.

Tabla 38. Cronograma de Actividades de las Historias de Usuario

| No | Nombre de la Historia de Usuario              | Fecha de Inicio Estimada | Fecha de Entrega Prevista | Duración en Semanas |
|----|-----------------------------------------------|--------------------------|---------------------------|---------------------|
| 1  | Aplicación de escritorio                      | 10 de enero de 2010      | 24 de septiembre de 2010  | 36                  |
| 2  | Creación del módulo de minería de estructura. | 10 de enero de 2010      | 25 de enero de 2010       | 2                   |

|           |                                                                    |                       |                         |    |
|-----------|--------------------------------------------------------------------|-----------------------|-------------------------|----|
| <b>3</b>  | Agregar tablas a la base de datos de Polaris.                      | 28 de enero de 2010   | 10 de febrero del 2010  | 2  |
| <b>4</b>  | Creación de un Web Crawler.                                        | 15 de febrero de 2010 | 20 de junio de 2010     | 18 |
| <b>5</b>  | Nivel de profundidad del Web Crawler.                              | 20 de junio de 2010   | 25 de junio de 2010     | 1  |
| <b>6</b>  | URL a analizar.                                                    | 25 de junio de 2010   | 28 de junio de 2010     | 1  |
| <b>7</b>  | Filtros para la búsqueda del Crawler.                              | 1 de julio de 2010    | 15 de julio de 2010     | 2  |
| <b>8</b>  | Guardar datos del Web Crawler.                                     | 17 de julio de 2010   | 22 de julio de 2010     | 1  |
| <b>9</b>  | Implementación de los algoritmos de Minería Web de Estructura.     | 1 de marzo de 2010    | 30 de mayo de 2010      | 13 |
| <b>10</b> | Visualización de grafos.                                           | 20 de julio de 2010   | 12 de agosto de 2010    | 3  |
| <b>11</b> | Mejorar la interfaz de Polaris V 1.0.                              | 15 de abril de 2010   | 5 de marzo de 2010      | 3  |
| <b>12</b> | Integración de la interfaz.                                        | 28 de agosto de 2010  | 1 de septiembre de 2010 | 2  |
| <b>13</b> | Integración de los algoritmos.                                     | 22 de agosto de 2010  | 31 de agosto de 2010    | 1  |
| <b>14</b> | Preprocesamiento de la información recolectada por el Web Crawler. | 25 de junio de 2010   | 2 de julio de 2010      | 1  |
| <b>15</b> | Enlaces internos y externos.                                       | 3 de julio de 2010    | 8 de julio de 2010      | 1  |
| <b>16</b> | Enlaces entrantes y salientes.                                     | 5 de julio de 2010    | 10 de julio de 2010     | 1  |
| <b>17</b> | Cargar información de base de datos                                | 23 de julio de 2010   | 1 de agosto de 2010     | 1  |
| <b>18</b> | Visualización de grafos con sus filtros.                           | 2 de agosto de 2010   | 15 de agosto de 2010    | 2  |
| <b>19</b> | Visualización de grafos de los enlaces internos.                   | 5 de agosto de 2010   | 10 de agosto de 2010    | 1  |
| <b>20</b> | Visualización de grafos de los enlaces externos.                   | 8 de agosto de 2010   | 13 de agosto de 2010    | 1  |
| <b>21</b> | Configuración de los algoritmos.                                   | 1 de abril de 2010    | 10 de abril de 2010     | 2  |
| <b>22</b> | Parámetros Page Rank                                               | 10 de abril de 2010   | 22 de abril de 2010     | 2  |
| <b>23</b> | Parámetros HITS.                                                   | 2 de mayo de 2010     | 8 de mayo de 2010       | 1  |
| <b>24</b> | Parámetros algoritmo Web Linker.                                   | 16 de mayo de 2010    | 20 de mayo de 2010      | 1  |
| <b>25</b> | Visualización de información de cada nodo.                         | 20 de agosto de 2010  | 5 de septiembre de 2010 | 2  |

|    |                                                               |                          |                          |   |
|----|---------------------------------------------------------------|--------------------------|--------------------------|---|
| 26 | Cálculo de los algoritmos.                                    | 22 de mayo de 2010       | 1 de junio de 2010       | 1 |
| 27 | Visualización del grafo final.                                | 2 de septiembre de 2010  | 15 de septiembre de 2010 | 2 |
| 28 | Algoritmo de la ruta más corta entre dos nodos seleccionados. | 22 de mayo de 2010       | 30 de mayo de 2010       | 1 |
| 29 | Creación de filtros según el protocolo de las páginas Web.    | 3 de agosto de 2010      | 10 de agosto de 2010     | 1 |
| 30 | Reporte en tablas de los cálculos.                            | 18 de agosto de 2010     | 1 de septiembre de 2010  | 2 |
| 31 | Exportación de datos a otros formatos.                        | 9 de septiembre de 2010  | 13 de septiembre de 2010 | 1 |
| 32 | Resultados en la mejora de la estructura.                     | 13 de septiembre de 2010 | 20 de septiembre de 2010 | 1 |
| 33 | Eliminar datos de la base de datos.                           | 20 de agosto de 2010     | 25 de agosto de 2010     | 1 |
| 34 | Creación de un nuevo proyecto.                                | 18 de septiembre de 2010 | 22 de septiembre de 2010 | 1 |
| 35 | Cargar un proyecto existente.                                 | 18 de septiembre de 2010 | 22 de septiembre de 2010 | 1 |
| 36 | Guardar proyecto.                                             | 19 de septiembre de 2010 | 23 de septiembre de 2010 | 1 |
| 37 | Guardar proyecto como.                                        | 19 de septiembre de 2010 | 23 de septiembre de 2010 | 1 |

En la sección anterior, en cada historia de usuario, se indicó el nombre del Desarrollador responsable. Puesto que en este proyecto solo intervienen dos programadores, no es posible establecer más personas para la implementación de determinada historia de usuario.

Sin embargo, se realizó la asignación de la responsabilidad del desarrollo de las historias a cada uno. Mientras uno se concentra en una historia, el otro interviene en el proceso de desarrollo de otra, pero ambos participan en la programación, de manera que se cumple la programación en parejas.

**Versiones del Proyecto:** De este proyecto se han creado nueve versiones, cada una de las cuales contiene la implementación de las historias de usuario relacionadas con el módulo que identifica su nombre.

En los casos de las historias que tienen una duración de más de 2 semanas, y superan la fecha de publicación de la versión, solo se midió el porcentaje correspondiente que debió estar implementado hasta la fecha. En las versiones siguientes, se realizaron las implementaciones en relación con las respectivas fechas de publicación.

Tabla 39. Versiones de POLARIS V2.0

| No | Nombre de la Versión del Programa   | Historias de Usuario que fueron implementadas                                                                                                                                                                                                                                                                                                                                                                                                                                      | Fecha de Publicación |
|----|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| 1  | Módulo de minería de estructura Web | <ol style="list-style-type: none"> <li>1. Aplicación de escritorio</li> <li>2. Creación del módulo de minería de estructura.</li> <li>3. Agregar tablas a la base de datos de Polaris.</li> </ol>                                                                                                                                                                                                                                                                                  | Febrero de 2010      |
| 2  | Web Crawler                         | <ol style="list-style-type: none"> <li>4. Creación de un Web Crawler.</li> <li>5. Nivel de profundidad del Web Crawler.</li> <li>6. URL a analizar.</li> <li>7. Filtros para la búsqueda del Crawler.</li> <li>8. Guardar datos del Web Crawler.</li> <li>9. Implementación de los algoritmos de Minería Web de Estructura.</li> </ol>                                                                                                                                             | Mayo de 2010         |
| 3  | Interfaz gráfica                    | <ol style="list-style-type: none"> <li>10. Visualización de grafos.</li> <li>11. Mejorar la interfaz de Polaris Versión 1.0.</li> <li>12. Integración de la interfaz.</li> <li>13. Integración de los algoritmos.</li> </ol>                                                                                                                                                                                                                                                       | Junio de 2010        |
| 4  | Preprocesamiento                    | <ol style="list-style-type: none"> <li>14. Preprocesamiento de la información recolectada por el Web Crawler.</li> <li>15. Enlaces internos y externos.</li> <li>16. Enlaces entrantes y salientes.</li> <li>17. Cargar información de base de datos.</li> </ol>                                                                                                                                                                                                                   | Julio de 2010        |
| 5  | Visualización de grafos             | <ol style="list-style-type: none"> <li>18. Visualización de grafos con sus filtros.</li> <li>19. Visualización de grafos de los enlaces internos.</li> <li>20. Visualización de grafos de los enlaces externos.</li> <li>21. Configuración de los algoritmos.</li> <li>22. Parámetros Page Rank</li> <li>23. Parámetros HITS.</li> <li>24. Parámetros algoritmo creado.</li> <li>25. Visualización de información de cada nodo.</li> <li>26. Cálculo de los algoritmos.</li> </ol> | Agosto de 2010       |

|   |                      |                                                                                                                                                                                  |                        |
|---|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| 6 | Grafo final          | 27. Visualización del grafo final.<br>28. Algoritmo de la ruta más corta entre dos nodos seleccionados.<br>29. Creación de filtros según el protocolo de las páginas Web.        | Agosto de 2010         |
| 7 | Reportes             | 30. Reporte en tablas de los cálculos.<br>31. Exportación de datos a otros formatos.<br>32. Resultados en la mejora de la estructura.<br>33. Eliminar datos de la base de datos. | Septiembre de 2010     |
| 8 | Proyectos            | 34. Creación de un nuevo proyecto.<br>35. Cargar un proyecto existente.<br>36. Guardar proyecto.<br>37. Guardar proyecto como.                                                   | Septiembre 2010        |
| 9 | <b>Versión Final</b> | <b>Todas las Historias de Usuario</b>                                                                                                                                            | <b>Octubre de 2010</b> |

**6.2.2 Iteraciones:** En las tablas de los ítems anteriores se establecieron, tanto la duración de cada una de las historias de usuario, como las historias que fueron implementadas para cada una de las versiones del proyecto.

Casi todas las historias (salvo algunas excepciones) se desarrollan dentro del tiempo normal de implementación de una Historia de Usuario, entre 1 y 3 semanas. En el caso de las versiones del proyecto, su tiempo de desarrollo no excede al mes (con la excepción de la segunda versión, que consume un tiempo de alrededor de un mes y medio).

El desarrollo de cada una de las versiones del proyecto constituye una iteración. De manera, que se cuentan con 9 iteraciones, en las cuales se implementan un promedio de 5 historias de usuario por iteración (no se tienen en cuenta el tiempo de implementación de las historias de usuario).

**6.2.3 Velocidad del proyecto:** Teniendo en cuenta que algunas de las historias se las desarrolló simultáneamente, se deben realizar las respectivas operaciones para calcular la velocidad del proyecto. Esta velocidad se calcula hallando el cociente entre la sumatoria de los puntos de las historias (semanas de duración de implementación) y el número de iteraciones.

Las sumatoria total de puntos es de 60 semanas (a las historias con una duración superior a 4 semanas, se les asignó como valor, 4, para no desproporcionar

demasiado la velocidad del proyecto). Este valor dividido entre 9 iteraciones, da como resultado 6,67 historias por iteración, un ritmo bastante rápido teniendo en cuenta que solo se cuenta con dos programadores.

Este valor es superior al mencionado al final del ítem anterior, acerca del promedio de 5 historias de usuario por iteración, debido a que en ese cálculo no se tuvo en cuenta el tiempo que toma codificar cada una de las historias. Además, tampoco se consideró que algunas historias se realizaron en el transcurso del proyecto, a lo largo de los meses.

Esto hace evidente el hecho de que, además de las 4 historias promedio, se implementaron las partes de las historias correspondientes a la versión en curso, en el caso de aquellas que tenían una duración superior a 4 semanas.

### **6.3 FASE DE DISEÑO**

Para esta fase de la metodología se realizó el análisis de las historias de usuario provistas por el cliente en la Fase de Planeación. Con el análisis de estos requerimientos se procedió a realizar un adecuado diseño del software.

Este diseño tiene que ver con la creación de una *Metáfora del Sistema*, en la cual se relata el funcionamiento del sistema en términos entendibles para el usuario. Otro elemento utilizado en la Fase de Diseño, son las *Tarjetas CRC*, cuya función consistió en realizar la distribución de las tareas específicas que se realizaron en la creación del sistema; en este caso en términos de programación. Además, es importante contar con un plan de posibles errores que puedan presentarse al tratar implementar la solución.

**6.3.1 Metáfora del sistema:** Después de realizar un análisis de las *Historias de Usuario* descritas en la *Fase de Planeación*, se establece que POLARIS VERSIÓN 2.0 es una aplicación de escritorio.

La metáfora del sistema se amplía en el capítulo **No. 7 IMPLEMENTACIÓN DE LA HERRAMIENTA POLARIS VERSIÓN 2.0.**

### **6.4 TARJETAS CRC**

Se realizó la distribución de las principales tareas, registrando en las tarjetas CRC las responsabilidades básicas del sistema y especificando las acciones a realizar por cada una.

Puesto que el lenguaje de programación, utilizado para la construcción de **POLARIS VERSIÓN 2.0**, es *Java* y toda su tecnología, resultó muy apropiado recurrir a esta herramienta del diseño de sistemas de información. Las Tarjetas CRC que se detallan en los siguientes ítems representan las principales clases de construidas para el código Java para **POLARIS VERSIÓN 2.0.**

Es necesario resaltar que las clases bosquejadas en las Tarjetas CRC son aquellas que resumen las principales funcionalidades del sistema en relación a la implementación de los algoritmos encargados de calcular el peso de las páginas Web y la creación del grafo. No se han esquematizado todas las clases que utilizadas para realizar el resto de tareas y procedimientos a fin de proveer simplicidad a la apropiación del análisis y diseño del proyecto.

#### 6.4.1 Tarjeta CRC – Ventana.

Tabla 40. Tarjeta CRC – Ventana.

| <b>TARJETA CRC – VENTANA</b>                                                                                                                                                                                                                                                                          |                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                            | Polaris versión 2.0                                                          |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                             | Ventana                                                                      |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                | <b>CLASES ASISTENTES</b>                                                     |
| <ul style="list-style-type: none"> <li>Es la representación gráfica de la herramienta Polaris versión 2.0, en la cual se acoplo los dos tipos de minería implementados hasta el momento.</li> </ul>                                                                                                   | <ul style="list-style-type: none"> <li>Nodo</li> <li>Area_Trabajo</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                  |                                                                              |
| <ul style="list-style-type: none"> <li>Presenta el área de trabajo y las herramientas de los dos tipos de minería.</li> <li>tanto la minería de uso como la minería de estructura trabajan totalmente independiente compartiendo dentro de la herramientas aspectos como el Drag and Drop.</li> </ul> |                                                                              |
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                      | Cristian Iles                                                                |

#### 6.4.2 Tarjeta CRC – crawlerRun.

Tabla 41. Tarjeta CRC – crawlerRun.

| <b>TARJETA CRC – CRAWLERRUN</b>                                                                                                                                                                             |                                                                                                                |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                  | Polaris versión 2.0                                                                                            |
| <b>Nombre de la clase</b>                                                                                                                                                                                   | crawlerRun                                                                                                     |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                      | <b>CLASES ASISTENTES</b>                                                                                       |
| <ul style="list-style-type: none"> <li>Permite ejecutar el proceso del Web Crawler.</li> </ul>                                                                                                              | <ul style="list-style-type: none"> <li>Area_Trabajo</li> <li>Configuración</li> <li>crawlerSettings</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                        |                                                                                                                |
| <ul style="list-style-type: none"> <li>Presenta un formulario en el que se muestra el progreso del proceso del Crawler.</li> <li>Muestra información relevante acerca de las páginas procesadas.</li> </ul> |                                                                                                                |
| <b>Desarrollador responsable</b>                                                                                                                                                                            | Carlos Andrés Rubio Coral                                                                                      |

### 6.4.3 Tarjeta CRC – crawlerSettings.

Tabla 42. Tarjeta CRC – crawlerSettings.

| <b>TARJETA CRC – CRAWLERSETTINGS</b>                                                                                                                                                                                                                                                                                                                   |                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                                                                             | Polaris versión 2.0                                                                                       |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                                                                              | crawlerSettings                                                                                           |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                                                                 | <b>CLASES ASISTENTES</b>                                                                                  |
| <ul style="list-style-type: none"> <li>Permite la configuración del Web Crawler.</li> </ul>                                                                                                                                                                                                                                                            | <ul style="list-style-type: none"> <li>Area_Trabajo</li> <li>Configuración</li> <li>crawlerRun</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                                                                   |                                                                                                           |
| <ul style="list-style-type: none"> <li>Presenta un formulario para la configuración Web Crawler.</li> <li>Selecciona el nivel de búsqueda del Web Crawler.</li> <li>El número de enlaces por página.</li> <li>Filtros por protocolos.</li> <li>Filtros por nombres de dominios, como los dominios por organización que son los más comunes.</li> </ul> |                                                                                                           |
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                       | Carlos Andrés Rubio Coral                                                                                 |

### 6.4.4 Tarjeta CRC – txtSettings.

Tabla 43. Tarjeta CRC – txtSettings.

| <b>TARJETA CRC – TXTSETTINGS</b>                                                                                                                                                |                                                                                                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                      | Polaris versión 2.0                                                                                       |
| <b>Nombre de la clase</b>                                                                                                                                                       | txtSettings                                                                                               |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                          | <b>CLASES ASISTENTES</b>                                                                                  |
| <ul style="list-style-type: none"> <li>Permite seleccionar un archivo que contiene la información del sitio analizado por el Web Crawler.</li> </ul>                            | <ul style="list-style-type: none"> <li>Area_Trabajo</li> <li>Configuración</li> <li>crawlerRun</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                            |                                                                                                           |
| <ul style="list-style-type: none"> <li>Presenta un formulario para Cargar un archivo de texto plano.</li> <li>Muestra la barra de progreso del archivo seleccionado.</li> </ul> |                                                                                                           |
| <b>Desarrollador responsable</b>                                                                                                                                                | Cristian Iles                                                                                             |

#### 6.4.5 Tarjeta CRC – Conexión.

Tabla 44. Tarjeta CRC – Conexión.

| <b>TARJETA CRC – CONEXIÓN</b>                                                                                                                                                                                                              |                                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                 | Polaris versión 2.0                                                           |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                  | Conexión                                                                      |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                     | <b>CLASES ASISTENTES</b>                                                      |
| <ul style="list-style-type: none"><li>• Se encarga de especificar el Origen y el Destino de cada uno de los nodos.</li><li>• Administrar una lista de apuntadores que permite reconocer que tipo de conexión se está realizando.</li></ul> | <ul style="list-style-type: none"><li>• Nodo</li><li>• Area_Trabajo</li></ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                       |                                                                               |
| <ul style="list-style-type: none"><li>• Para especificar el origen y destino de cada uno de los nodos se lo hace por medio de dos apuntadores.</li></ul>                                                                                   |                                                                               |
| <b>Desarrollador responsable</b>                                                                                                                                                                                                           | Cristian Iles                                                                 |

#### 6.4.6 Tarjeta CRC – Nodo

Tabla 45. Tarjeta CRC – Nodo

| <b>TARJETA CRC – NODO</b>                                                                                                                                                                                                                                                                                                                                                                  |                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                                                                                                                 | Polaris versión 2.0                                        |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                                                                                                                  | Nodo                                                       |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                                                                                                     | <b>CLASES ASISTENTES</b>                                   |
| <ul style="list-style-type: none"><li>• Es la encargada de llevar la información completa de cada ítem del árbol.</li></ul>                                                                                                                                                                                                                                                                | <ul style="list-style-type: none"><li>• Conexión</li></ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                                                                                                       |                                                            |
| <ul style="list-style-type: none"><li>• Maneja las coordenadas X y Y para la ubicación de los nodos.</li><li>• En esta clase se guarda la información de que tipo de nodo es el que se encuentra trabajando.</li><li>• Presenta tres apuntadores: el primero que apunta al siguiente nodo, el segundo que apunta al anterior nodo y el tercero que apunta a la primera conexión.</li></ul> |                                                            |
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                                                           | Cristian Iles                                              |

#### 6.4.7 Tarjeta CRC – Área de trabajo

Tabla 46. Tabla CRC – Área de trabajo

| <b>TABLA CRC – ÁREA DE TRABAJO</b>                                                                                                                                    |                                                                          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                            | Polaris versión 2.0                                                      |
| <b>Nombre de la clase</b>                                                                                                                                             | Area_Trabajo                                                             |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                | <b>CLASES ASISTENTES</b>                                                 |
| <ul style="list-style-type: none"> <li>Se encarga de eliminar, modificar, mover y presentar los nodos en el área donde el usuario se encuentra trabajando.</li> </ul> | <ul style="list-style-type: none"> <li>Nodo</li> <li>Conexión</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                  |                                                                          |
| <ul style="list-style-type: none"> <li>Esta clase es la encargada de apuntar siempre al primer nodo para iniciar a realizar las conexiones.</li> </ul>                |                                                                          |
| <b>Desarrollador responsable</b>                                                                                                                                      | Cristian Iles                                                            |

#### 6.4.8 Tarjeta CRC – Configuración.

Tabla 47. Tarjeta CRC – Configuración.

| <b>TARJETA CRC – CONFIGURACIÓN</b>                                                                                                                                                                                      |                                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                              | Polaris versión 2.0                                                                                                                                                                                          |
| <b>Nombre de la clase</b>                                                                                                                                                                                               | Configuración                                                                                                                                                                                                |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                  | <b>CLASES ASISTENTES</b>                                                                                                                                                                                     |
| <ul style="list-style-type: none"> <li>Permite la configuración de los nodos dependiendo de qué tipo de nodo ha sido arrastrado al área de trabajo, ya que los nodos tienen diferente tipo de configuración.</li> </ul> | <ul style="list-style-type: none"> <li>LinksSettings</li> <li>pageRankSettings</li> <li>hitsSettings</li> <li>WebGraph</li> <li>Page Rank</li> <li>HITS</li> <li>crawlerSettings</li> <li>ventana</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                    |                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>La configuración de un elemento del árbol de herramientas debe realizarse antes de ejecutarse los nodos.</li> </ul>                                                              |                                                                                                                                                                                                              |
| <b>Desarrollador responsable</b>                                                                                                                                                                                        | Cristian Iles                                                                                                                                                                                                |

#### 6.4.9 Tarjeta CRC – domainViewFrame.

Tabla 48. Tarjeta CRC – domainViewFrame.

| <b>TARJETA CRC – DOMAINVIEWFRAME</b>                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                                                                                     | Polaris versión 2.0                                                                                                                                                    |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                                                                                      | domainViewFrame                                                                                                                                                        |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                                                                         | <b>CLASES ASISTENTES</b>                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• Representa la ventana de visualización del grafo, teniendo en cuenta el conjunto de dominios al que pertenece cada uno los enlaces de la página Web semilla.</li> <li>• Recursivamente realiza la misma actividad para cada uno de los enlaces encontrados convirtiéndolos en páginas semillas.</li> </ul>            | <ul style="list-style-type: none"> <li>• LinksRun</li> <li>• pageRankRun</li> <li>• hitsRun</li> <li>• visitRun</li> <li>• Dijkstra</li> <li>• area_trabajo</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>• Esta clase permite observar el grafo de la información recolectada luego de aplicar las configuraciones de los enlaces.</li> <li>• También permite presentar el grafo después de ejecutado cualquier algoritmo, de esta forma al seleccionar los nodos se puede ver el valor calculado por los algoritmos.</li> </ul> |                                                                                                                                                                        |
| <b>Desarrollador responsable</b>                                                                                                                                                                                                                                                                                                                               | Cristian Iles                                                                                                                                                          |

#### 6.4.10 Tarjeta CRC – graphViewFrame.

Tabla 49. Tarjeta CRC – graphViewFrame.

| <b>TARJETA CRC – GRAPHVIEWFRAME</b>                                                                                                                                                                                                                                                                                        |                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                                                 | Polaris versión 2.0                                                                                                                                                    |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                                                  | graphViewFrame                                                                                                                                                         |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                                     | <b>CLASES ASISTENTES</b>                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• Representa la ventana de visualización del grafo de la información recolectado.</li> <li>• Recursivamente realiza la misma actividad para cada uno de los enlaces encontrados convirtiéndolos en páginas semillas y representándolos en forma de nuevos nodos y arcos.</li> </ul> | <ul style="list-style-type: none"> <li>• LinksRun</li> <li>• pageRankRun</li> <li>• hitsRun</li> <li>• visitRun</li> <li>• Dijkstra</li> <li>• area_trabajo</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>• Esta clase permite observar el grafo de la información recolectada luego de aplicar las configuraciones de los enlaces.</li> <li>• También permite presentar el grafo después de ejecutado cualquier algoritmo,</li> </ul>                                                        |                                                                                                                                                                        |

de esta forma al seleccionar los nodos se puede ver el valor calculado por los algoritmos.

- Se puede observar los diferentes niveles y el número de enlaces definidos por el usuario.

**Desarrollador responsable** Carlos Andrés Rubio Coral

#### 6.4.11 Tarjeta CRC – radialViewFrame.

Tabla 50. Tarjeta CRC – radialViewFrame.

| <b>TARJETA CRC – RADIALVIEWFRAME</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Polaris versión 2.0                                                                                                                                                    |
| <b>Nombre de la clase</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | radialViewFrame                                                                                                                                                        |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <b>CLASES ASISTENTES</b>                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• Representa la ventana de visualización del grafo, en forma radial, la información recolectado.</li> <li>• Recursivamente realiza la misma actividad para cada uno de los enlaces encontrados convirtiéndolos en páginas semillas y representándolos en forma de nuevos nodos y arcos.</li> </ul>                                                                                                                                                                                                                                                                                        | <ul style="list-style-type: none"> <li>• LinksRun</li> <li>• pageRankRun</li> <li>• hitsRun</li> <li>• visitRun</li> <li>• Dijkstra</li> <li>• area_trabajo</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>• Esta clase permite observar el grafo, en forma radial, de la información recolectada luego de aplicar las configuraciones de los enlaces.</li> <li>• También permite presentar el grafo después de ejecutado cualquier algoritmo, de esta forma al seleccionar los nodos se puede ver el valor calculado por los algoritmos.</li> <li>• Permite interactuar al usuario con el grafo radial de una forma mucho más fácil que con los otros dos tipos de visualizaciones.</li> <li>• Se puede observar los diferentes niveles y el número de enlaces definidos por el usuario.</li> </ul> |                                                                                                                                                                        |
| <b>Desarrollador responsable</b> Carlos Andrés Rubio Coral                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                        |

#### 6.4.12 Tarjeta CRC – Enlace

Tabla 51. Tabla CRC – Enlace.

| <b>TABLA CRC – ENLACE</b>                                                                                                                 |                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                | Polaris versión 2.0                                      |
| <b>Nombre de la clase</b>                                                                                                                 | Página                                                   |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                    | <b>CLASES ASISTENTES</b>                                 |
| <ul style="list-style-type: none"> <li>Se encarga de especificar la página de origen y de destino de cada uno de los enlaces.</li> </ul>  | <ul style="list-style-type: none"> <li>Página</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                      |                                                          |
| <ul style="list-style-type: none"> <li>Crea los respectivos vínculos para la generación del grafo con sus concernientes links.</li> </ul> |                                                          |
| <b>Desarrollador responsable</b>                                                                                                          | Cristian Iles                                            |

#### 6.4.13 Tarjeta CRC – Host.

Tabla 52. Tabla CRC – Host.

| <b>TABLA CRC – HOST</b>                                                                                                                                                  |                                                                    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| <b>Nombre del proyecto</b>                                                                                                                                               | Polaris versión 2.0                                                |
| <b>Nombre de la clase</b>                                                                                                                                                | Host                                                               |
| <b>RESPONSABILIDADES DE LAS CLASES</b>                                                                                                                                   | <b>CLASES ASISTENTES</b>                                           |
| <ul style="list-style-type: none"> <li>Genera la estructura para el grafo Domain para agrupar las páginas según su host.</li> </ul>                                      | <ul style="list-style-type: none"> <li>domainViewFrame.</li> </ul> |
| <b>OBSERVACIONES</b>                                                                                                                                                     |                                                                    |
| <ul style="list-style-type: none"> <li>Crea los respectivos vínculos para la generación del grafo con sus concernientes links que pertenecen a un mismo host.</li> </ul> |                                                                    |
| <b>Desarrollador responsable</b>                                                                                                                                         | Cristian Iles                                                      |

#### 6.4.14 Tarjeta CRC – Página.

Tabla 53. Tarjeta CRC – Página.

| TARJETA CRC – PÁGINA                                                                                                                                                                                                                                                                         |                                                        |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| Nombre del proyecto                                                                                                                                                                                                                                                                          | Polaris versión 2.0                                    |
| Nombre de la clase                                                                                                                                                                                                                                                                           | Página                                                 |
| RESPONSABILIDADES DE LAS CLASES                                                                                                                                                                                                                                                              | CLASES ASISTENTES                                      |
| <ul style="list-style-type: none"><li>Manejar toda la información perteneciente a una determinada página (Numeración, Nivel, URL, Nombre e Id de página).</li></ul>                                                                                                                          | <ul style="list-style-type: none"><li>Enlace</li></ul> |
| OBSERVACIONES                                                                                                                                                                                                                                                                                |                                                        |
| <ul style="list-style-type: none"><li>Cada página maneja cuatro tipos de apuntadores, en los cuales se reflejan los enlaces entrantes internos, entrantes externos, salientes internos y salientes externos.</li><li>Inserta los enlaces.</li><li>Crea un nuevo listado de enlaces</li></ul> |                                                        |
| Desarrollador responsable                                                                                                                                                                                                                                                                    | Cristian Iles                                          |

#### 6.4.15 Tarjeta CRC – Sitio.

Tabla 54. Tarjeta CRC – Sitio.

| TARJETA CRC – SITIO                                                                                                                                                                                                                                                                                           |                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Nombre del proyecto                                                                                                                                                                                                                                                                                           | Polaris versión 2.0                                                                |
| Nombre de la clase                                                                                                                                                                                                                                                                                            | Sitio                                                                              |
| RESPONSABILIDADES DE LAS CLASES                                                                                                                                                                                                                                                                               | CLASES ASISTENTES                                                                  |
| <ul style="list-style-type: none"><li>Administrar una lista de páginas o de objetos de tipo página en una lista doblemente enlazada, con la finalidad de crear todos los nodos nuevos de la Web.</li></ul>                                                                                                    | <ul style="list-style-type: none"><li>Página</li><li>Enlace</li><li>Host</li></ul> |
| OBSERVACIONES                                                                                                                                                                                                                                                                                                 |                                                                                    |
| <ul style="list-style-type: none"><li>Permite la inserción de una página nueva en la lista.</li><li>Permite buscar páginas por el identificador de página.</li><li>Permite buscar páginas por numeración.</li><li>Permite buscar páginas por URL.</li><li>Indica cual es la página inicial y final.</li></ul> |                                                                                    |
| Desarrollador responsable                                                                                                                                                                                                                                                                                     | Carlos Andrés Rubio Coral                                                          |

### 6.5 FASE DE DESARROLLO.

En la fase de desarrollo se concentra prácticamente todo el trabajo pesado: la codificación de todo el sistema. La metodología XP plantea que en esta fase se

cumplan algunas recomendaciones que, además de ajustarse a la esencia de la metodología de crear software de manera rápida y extrema, permiten satisfacer las necesidades del cliente en forma amplia, de los programadores y producir código que cumpla las especificaciones y que quede bien probado.

Lo que se persigue es la participación activa del cliente y que los programadores trabajen en pareja e intercambien o integren el código que producen para que pueda ser reutilizado y no se haga trabajo de más. Otro requisito de esta fase es la elaboración de unidades de prueba que permitan evaluar el funcionamiento de los distintos códigos que hayan de ser implementados.

**6.5.1 Unidades de Prueba:** Las unidades de prueba son pequeños test que permiten conocer exactamente qué es lo que deben hacer cada una de las clases que se piensan implementar y, una vez implementadas, determinar si se satisfacen los requisitos establecidos por el usuario final.

En los ítems siguientes se presentarán las unidades de prueba que fueron utilizadas para evaluar las clases planteadas en la Fase de Diseño, en la sección Tablas CRC.

➤ **Test aplicado a la clase Ventana**

Tabla 55. Test aplicado a la clase: Ventana

| <b>TEST APLICADO A LA CLASE VENTANA</b>                                                                                        |                     |
|--------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                                                                     | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                                                                      | Ventana             |
| <b>TEST DE PRUEBA</b>                                                                                                          | <b>ACEPTACIÓN</b>   |
| • ¿Presenta la interfaz gráfica de la herramienta Polaris versión 2.0?                                                         |                     |
| • ¿La clase contiene el módulo de minería de uso y estructura Web?                                                             |                     |
| • ¿Permite eliminar, modificar, arrastrar y mover los diferentes objetos con los cuales se trabaja dentro del área de trabajo? |                     |
| • ¿Permite configurar y determinar las diferentes restricciones para los objetos presentes dentro del área de trabajo?         |                     |
| • ¿Permite realizar las conexiones entra cada uno de los nodos?                                                                |                     |
| • ¿Permite el manejo del Drag and Drop?                                                                                        |                     |
| • ¿Reconoce los iconos que se están seleccionando?                                                                             |                     |
| <b>Desarrollador responsable:</b> Cristian Iles                                                                                |                     |

➤ **Test aplicado a la clase CrawlerRun**

Tabla 56. Test aplicado a la clase: CrawlerRun

| <b>TEST APLICADO A LA CLASE CRAWLERRUN</b>                               |                     |
|--------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                               | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                | Ventana             |
| <b>TEST DE PRUEBA</b>                                                    | <b>ACEPTACIÓN</b>   |
| • ¿Ejecuta el proceso del Web Crawler?                                   |                     |
| • ¿Despliega una nueva ventana donde se observa el progreso del Crawler? |                     |
| • ¿Accede a internet y realiza las solicitudes de las páginas?           |                     |
| • ¿Determina cuales son los enlaces validos y no validos?                |                     |
| • ¿Adiciona las páginas al sitio Web?                                    |                     |
| <b>Desarrollador responsable:</b> Carlos Andrés Rubio Coral              |                     |

➤ **Test aplicado a la clase CrawlerSettings**

Tabla 57. Test aplicado a la clase: CrawlerSettings

| <b>TEST APLICADO A LA CLASE CRAWLERSETTINGS</b>                                  |                     |
|----------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                       | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                        | Ventana             |
| <b>TEST DE PRUEBA</b>                                                            | <b>ACEPTACIÓN</b>   |
| • ¿Despliega una nueva ventana donde se observa la configuración del Crawler?    |                     |
| • ¿Permite escoger el nivel de búsqueda para el Crawler?                         |                     |
| • ¿Permite seleccionar la cantidad máxima de Links por página?                   |                     |
| • ¿Permite seleccionar uno o varios dominios con los cuales se desee trabajar?   |                     |
| • ¿Permite seleccionar uno o varios protocolos con los cuales se desea trabajar? |                     |
| • ¿Permite agregar más protocolos a la lista de los predeterminados?             |                     |
| • ¿Permite agregar más dominios a la lista de los predeterminados?               |                     |
| <b>Desarrollador responsable:</b> Carlos Andrés Rubio Coral                      |                     |

➤ **Test aplicado a la clase TxtSettings**

Tabla 58. Test aplicado a la clase: TxtSettings

| <b>TEST APLICADO A LA CLASE TXTSETTINGS</b>                                                |                     |
|--------------------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                                 | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                                  | Ventana             |
| <b>TEST DE PRUEBA</b>                                                                      | <b>ACEPTACIÓN</b>   |
| • ¿Genera una ventana que permite la selección de un archivo?                              |                     |
| • ¿Valida que el archivo sea aceptado por la herramienta?                                  |                     |
| • ¿Carga el archivo seleccionado?                                                          |                     |
| • ¿Genera una barra de progreso donde se indica el estado de la carga que está realizando? |                     |
| <b>Desarrollador responsable:</b> Cristian Iles                                            |                     |

➤ **Test aplicado a la clase Conexión**

Tabla 59. Test aplicado a la clase: Conexión

| <b>TEST APLICADO A LA CLASE CONEXIÓN</b>                         |                     |
|------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                       | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                        | Ventana             |
| <b>TEST DE PRUEBA</b>                                            | <b>ACEPTACIÓN</b>   |
| • ¿Especifica el origen de los nodos?                            |                     |
| • ¿Especifica el destino de los nodos?                           |                     |
| • ¿Permite reconocer el tipo de conexión que se está realizando? |                     |
| • ¿Administra una lista de apuntadores?                          |                     |
| <b>Desarrollador responsable:</b> Cristian Iles                  |                     |

➤ **Test aplicado a la clase Nodo**

Tabla 60. Test aplicado a la clase: Nodo

| <b>TEST APLICADO A LA CLASE NODO</b>                            |                     |
|-----------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                      | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                       | Ventana             |
| <b>TEST DE PRUEBA</b>                                           | <b>ACEPTACION</b>   |
| • ¿Maneja las coordenadas (x,y) para la ubicación de los nodos? |                     |
| • ¿Guarda la información del tipo de nodo sobre el cual se está |                     |

|                                                                                |  |
|--------------------------------------------------------------------------------|--|
| trabajando por medio de una lista?                                             |  |
| • ¿El apuntador id_nodo se encarga de identificar al nodo actual de los demás? |  |
| • ¿Maneja los apuntadores al siguiente nodo y al anterior?                     |  |
| • ¿Maneja los nodos de la primera y la ultima conexión?                        |  |
| • ¿Informa si un nodo ya ha sido configurado?                                  |  |
| • ¿Informa si un nodo ya ha sido ejecutado?                                    |  |
| • ¿Informa si el nodo ya se encuentra conectado a otro?                        |  |
| <b>Desarrollador responsable:</b> Cristian Iles                                |  |

➤ **Test aplicado a la clase Área de Trabajo**

Tabla 61. Test aplicado a la clase: Área de Trabajo

| <b>TEST APLICADO A LA CLASE ÁREA DE TRABAJO</b>                                                                                      |                     |
|--------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                                                                           | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                                                                            | Ventana             |
| <b>TEST DE PRUEBA</b>                                                                                                                | <b>ACEPTACIÓN</b>   |
| • ¿Presenta los nodos en el área de trabajo?                                                                                         |                     |
| • ¿Elimina los nodos que se encuentren en el área de trabajo?                                                                        |                     |
| • ¿Modifica los nodos que se encuentren en el área de trabajo?                                                                       |                     |
| • ¿Mueve los nodos que se encuentren en el área de trabajo?                                                                          |                     |
| • ¿Contiene una variable nodo que se encarga de apuntar al primero y al último de los nodos que se encuentren en el área de trabajo? |                     |
| • ¿Informa si un nodo ya ha sido configurado?                                                                                        |                     |
| • ¿Informa si un nodo ya ha sido ejecutado?                                                                                          |                     |
| • ¿Informa si el nodo ya se encuentra conectado a otro?                                                                              |                     |
| <b>Desarrollador responsable:</b> Cristian Iles                                                                                      |                     |

➤ **Test aplicado a la clase Configuración.**

Tabla 62. Test aplicado a la clase: Configuración.

| <b>TEST APLICADO A LA CLASE – CONFIGURACIÓN</b>                                           |                     |
|-------------------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                                | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                                 | Configuración       |
| <b>TEST DE PRUEBA</b>                                                                     | <b>ACEPTACIÓN</b>   |
| • Si el nodo es de tipo LINKS, ¿informa si este nodo debe tomar links internos entrantes? |                     |
| • Si el nodo es de tipo LINKS, ¿informa si este nodo debe tomar links internos salientes? |                     |
| • Si el nodo es de tipo LINKS, ¿informa si este nodo debe tomar links externos entrantes? |                     |
| • Si el nodo es de tipo LINKS, ¿informa si este nodo debe tomar links externos salientes? |                     |
| • ¿Muestra la cantidad de datos procesados por cada algoritmo seleccionado?               |                     |
| • ¿Anuncia cual es la URL de inicio (página semilla)?                                     |                     |
| • ¿Advierte si la URL de inicio (página semilla), se ha ingresado de forma incorrecta?    |                     |
| • ¿Muestra el nivel de búsqueda seleccionado por el usuario?                              |                     |
| • ¿Muestra el total de enlaces por página?                                                |                     |
| • ¿Permite la selección de los protocolos, para definir su filtro?                        |                     |
| • ¿Permite la selección de los nombres de dominios, para definir su filtro?               |                     |
| • ¿Permite ingresar el valor dampening?                                                   |                     |
| • ¿Permite ingresar el valor de iteraciones, para el algoritmo Page Rank?                 |                     |
| • ¿Permite ingresar el valor de iteraciones, para el algoritmo HITS?                      |                     |
| • ¿Guarda el valor calculado del algoritmo Page Rank?                                     |                     |
| • ¿Guarda el valor calculado del algoritmo HITS?                                          |                     |
| • ¿Guarda la ruta del archivo de texto plano?                                             |                     |
| • ¿Muestra la información del archivo de texto plano?                                     |                     |
| • ¿Advierte si uno de los nodos ha sido configurado?                                      |                     |
| <b>Desarrollador responsable</b>                                                          | Cristian Iles       |

➤ **Test aplicado a la clase: domainViewFrame.**

Tabla 63. Test aplicado a la clase: domainViewFrame.

| <b>TEST APLICADO A LA CLASE – DOMAINVIEWFRAME</b>                                                                                      |                     |
|----------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                                                                             | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                                                                              | domanViewFrame      |
| <b>TEST DE PRUEBA</b>                                                                                                                  | <b>ACEPTACIÓN</b>   |
| <ul style="list-style-type: none"> <li>• ¿Contiene la instancia de la clase Domain, que pertenece a la herramienta Prefuse?</li> </ul> |                     |
| <ul style="list-style-type: none"> <li>• ¿Presenta una ventana de visualización?</li> </ul>                                            |                     |
| <ul style="list-style-type: none"> <li>• ¿Presenta la información correspondiente a cada uno de los nodos?</li> </ul>                  |                     |
| <ul style="list-style-type: none"> <li>• ¿Advierte el tipo de conexión que se debe realizar para su funcionamiento?</li> </ul>         |                     |
| <ul style="list-style-type: none"> <li>• ¿Obtiene la información recolectada por la instancia de la clase host?</li> </ul>             |                     |
| <b>Desarrollador responsable</b>                                                                                                       | Cristian Iles       |

➤ **Test aplicado a la clase: graphViewFrame.**

Tabla 64. Test aplicado a la clase: graphViewFrame.

| <b>TEST APLICADO A LA CLASE – GRAPHVIEWFRAME</b>                                                                                          |                           |
|-------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| <b>Nombre del proyecto</b>                                                                                                                | Polaris versión 2.0       |
| <b>Nombre de la clase</b>                                                                                                                 | graphViewFrame            |
| <b>TEST DE PRUEBA</b>                                                                                                                     | <b>ACEPTACIÓN</b>         |
| <ul style="list-style-type: none"> <li>• ¿Contiene la instancia de la clase GraphView, que pertenece a la herramienta Prefuse?</li> </ul> |                           |
| <ul style="list-style-type: none"> <li>• ¿Presenta una ventana de visualización?</li> </ul>                                               |                           |
| <ul style="list-style-type: none"> <li>• ¿Presenta la información correspondiente a cada uno de los nodos?</li> </ul>                     |                           |
| <ul style="list-style-type: none"> <li>• ¿Advierte el tipo de conexión que se debe realizar para su funcionamiento?.</li> </ul>           |                           |
| <b>Desarrollador responsable</b>                                                                                                          | Carlos Andrés Rubio Coral |

➤ **Test aplicado a la clase: radialViewFrame.**

Tabla 65. Test aplicado a la clase: radialViewFrame.

| <b>TEST APLICADO A LA CLASE – RADIALVIEWFRAME</b>                                        |                           |
|------------------------------------------------------------------------------------------|---------------------------|
| <b>Nombre del proyecto</b>                                                               | Polaris versión 2.0       |
| <b>Nombre de la clase</b>                                                                | radialViewFrame           |
| <b>TEST DE PRUEBA</b>                                                                    | <b>ACEPTACIÓN</b>         |
| • ¿Contiene la instancia de la clase RadialView, que pertenece a la herramienta Prefuse? |                           |
| • ¿Presenta una ventana de visualización?                                                |                           |
| • ¿Presenta la información correspondiente a cada uno de los nodos?                      |                           |
| • ¿Advierte el tipo de conexión que se debe realizar para su funcionamiento?             |                           |
| <b>Desarrollador responsable</b>                                                         | Carlos Andrés Rubio Coral |

➤ **Test aplicado a la clase Enlace.**

Tabla 66. Test aplicado a la clase: Enlace.

| <b>TEST APLICADO A LA CLASE – ENLACE</b>                   |                     |
|------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                 | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                  | Enlace              |
| <b>TEST DE PRUEBA</b>                                      | <b>ACEPTACIÓN</b>   |
| • ¿Determina las páginas de origen de los enlaces?.        |                     |
| • ¿Determina las páginas de destino de los enlaces?.       |                     |
| • ¿Genera una lista de enlaces, para las páginas origen?.  |                     |
| • ¿Genera una lista de enlaces, para las páginas destino?. |                     |
| • ¿Es utilizada por la clase página?.                      |                     |
| <b>Desarrollador responsable</b>                           | Cristian Iles       |

➤ **Test aplicado a la clase Host.**

Tabla 67. Test aplicado a la clase: Host.

| <b>TEST APLICADO A LA CLASE – HOST</b>        |                     |
|-----------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                    | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                     | Host                |
| <b>TEST DE PRUEBA</b>                         | <b>ACEPTACION</b>   |
| • ¿Almacena el nombre del host?               |                     |
| • ¿Les asigna un identificador?               |                     |
| • ¿Crea un apuntador para el siguiente host?  |                     |
| • ¿Crea un apuntador para el anterior host?   |                     |
| • ¿Es utilizada por la clase domainViewFrame? |                     |
| <b>Desarrollador responsable</b>              | Cristian Iles       |

➤ **Test aplicado a la clase: Página.**

Tabla 68. Test aplicado a la clase: Página.

| <b>TEST APLICADO A LA CLASE – PÁGINA</b>                                       |                     |
|--------------------------------------------------------------------------------|---------------------|
| <b>Nombre del proyecto</b>                                                     | Polaris versión 2.0 |
| <b>Nombre de la clase</b>                                                      | Página              |
| <b>TEST DE PRUEBA</b>                                                          | <b>ACEPTACIÓN</b>   |
| • ¿Almacena en una lista la numeración?                                        |                     |
| • ¿Almacena en una lista el nivel?                                             |                     |
| • ¿Almacena en una lista la URL?                                               |                     |
| • ¿Almacena en una lista el nombre de una página Web?                          |                     |
| • ¿Almacena en una lista el identificador?                                     |                     |
| • ¿Se adiciona una conexión o enlaces entre las páginas o nodos?               |                     |
| • ¿Se identifican las conexiones como enlaces internos, entrantes y salientes? |                     |
| • ¿Se identifican las conexiones como enlaces externos, entrantes y salientes? |                     |
| <b>Desarrollador responsable</b>                                               | Cristian Iles       |

➤ **Test aplicado a la clase Sitio.**

Tabla 69. Test aplicado a la clase: Sitio.

| <b>TEST APLICADO A LA CLASE – SITIO</b>            |                           |
|----------------------------------------------------|---------------------------|
| <b>Nombre del proyecto</b>                         | Polaris versión 2.0       |
| <b>Nombre de la clase</b>                          | Sitio                     |
| <b>TEST DE PRUEBA</b>                              | <b>ACEPTACIÓN</b>         |
| • ¿Administra una lista de objetos de tipo página? |                           |
| • ¿Crea un listado de todos los host?              |                           |
| • ¿Reconoce la cantidad de host?                   |                           |
| • ¿Agrupa los host encontrados en un solo grupo?   |                           |
| • ¿Crea una lista de host concatenados?            |                           |
| • ¿Busca un host dentro de la lista?               |                           |
| • ¿Muestra la información de una página?           |                           |
| • ¿Inserta nuevas páginas?                         |                           |
| • ¿Elimina páginas sin conexión?                   |                           |
| • ¿Busca una página predeterminada?                |                           |
| <b>Desarrollador responsable</b>                   | Carlos Andrés Rubio Coral |

## 6.6 FASE DE PRUEBAS.

Uno de los pilares de la programación extrema es la recurrencia constante a las pruebas que certifiquen el cumplimiento de los requisitos planteados por el usuario final y promuevan su satisfacción. En esta fase, se realizan los procesos de implementación del código y se aplican las pruebas de aceptación.

En el proceso de implementación se deben aplicar las pruebas de unidad de la fase de desarrollo y comprobar que el código satisface los requerimientos planteados por el usuario. Además de estas pruebas, es necesario aplicar las pruebas de aceptación que se basan en las historias de usuario y están relacionadas con la funcionalidad del sistema y no con su implementación en sí.

**6.6.1 Implementación:** La implementación consiste en añadir al sistema completo las nuevas funcionalidades que sean programadas a medida que se van codificando las clases propuestas en las Tarjetas CRC, luego de cada iteración. Sin embargo, para realizar la implementación de cada una de las clases programadas, éstas deben satisfacer los requisitos propios de cada clase, que se han plasmado en las Pruebas de Unidad.

Estos test, que fueron diseñados libres de la dependencia de código, permiten evaluar de forma específica la programación realizada por las parejas en cada sección del sistema y facilitan la reutilización del código. En los casos en que un programador añada alguna funcionalidad se debe percatar de que la clase sigan

satisfaciendo los requerimientos del test correspondiente.

En la implementación de **POLARIS V 2.0** se ha asegurado que todas las clases pasen sus respectivos test y así se lleve a término todas las exigencias del cliente.

**6.6.2 Pruebas de aceptación:** Las pruebas de aceptación son otro tipo de pruebas que deben realizarse al sistema antes de ser usado por el usuario final. Las pruebas mencionadas en los ítems anteriores permiten evaluar el código fuente, mientras que las pruebas que se analizarán en esta sección permiten evaluar el funcionamiento del sistema y comprobar que se han implementado todas las historias de usuario propuestas.

Por otro lado, este tipo de pruebas permite establecer funcionalidades que no se detallan en las Historias de Usuario. Con las pruebas de aceptación se determinarán algunos tipos de requisitos no funcionales.

Para que las pruebas de aceptación sean redactadas de manera completa, sin omitir detalles importantes sobre lo que debe hacer el sistema, es necesaria la presencia del Cliente. Esta es otra de las secciones en que la intervención del cliente se hace indispensable, debido a que sus conocimientos del problema que el sistema solucionará lo convierten en la persona idónea para bosquejar la respuesta que debe dar el sistema ante cada eventualidad.

Todas las pruebas de aceptación presentan cuatro secciones en la que se describen: las condiciones de ejecución (lo que ya debe estar en funcionamiento para que se pueda realizar la prueba), los pasos de ejecución (lo que debe hacer el usuario para obtener los resultados esperados), los resultados esperados (lo que el sistema debe hacer en respuesta a los pasos de ejecución) y la evaluación de la prueba (lo que el sistema hizo en respuesta a los pasos de ejecución).

Existen casos en los que la ejecución de una prueba de aceptación (condiciones de ejecución), dependen de haberse ejecutado otras pruebas de aceptación.

A continuación se presentan las pruebas de aceptación con las que ha sido evaluado el desempeño y funcionamiento de **POLARIS VERSIÓN 2.0**.

➤ **Prueba de aceptación – Iniciar uso de la aplicación.**

Tabla 70. Prueba de aceptación – Iniciar uso de la aplicación.

| <b>PRUEBA DE ACEPTACIÓN – INICIAR USO DE LA APLICACIÓN</b>                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                           | HU. No. 1 – Aplicación de escritorio                                                                                                                                                                              |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                  | Iniciar el uso de la herramienta POLARIS VERSIÓN 2.0 y mostrar la <i>Página de Bienvenida</i> .                                                                                                                   |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                      | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• La aplicación debe estar debidamente instalada y configurada.</li> <li>• La base de datos deben estar en ejecución.</li> <li>• El equipo donde se iniciará el aplicativo debe tener acceso a internet para realizar nuevos proyectos.</li> <li>• En el equipo donde se ejecuta la aplicación puede tener como sistema operativo Windows o Linux.</li> </ul> | <ul style="list-style-type: none"> <li>• Presentación de la interfaz de usuario.</li> <li>• Se escoge el sobre que módulo se desea trabajar (Minería de Uso Web o Minería de Estructura Web).</li> </ul>          |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                          | <b>Evaluación de la Prueba</b>                                                                                                                                                                                    |
| <ul style="list-style-type: none"> <li>• La aplicación debe iniciarse de manera satisfactoria y debe presentar al usuario la opción de escoger el tipo de minería que desea trabajar.</li> </ul>                                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>• La aplicación inició el proceso de cargue en el computador.</li> <li>• La aplicación se cargó de manera satisfactoria y se visualizó la Página de Bienvenida.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                       | Prueba Aprobada                                                                                                                                                                                                   |

➤ **Prueba de aceptación – Ingreso al módulo de Minería de Estructura Web.**

Tabla 71. Prueba de aceptación – Ingreso al módulo de Minería de Estructura Web.

| <b>PRUEBA DE ACEPTACIÓN – INGRESO AL MÓDULO DE MINERÍA DE ESTRUCTURA WEB</b>                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | HU. No. 2 – Creación del módulo de Minería de Estructura Web                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                           | Ingreso al módulo de Minería de Estructura Web y muestra la interfaz inicial del módulo.                                                                                                                                                                                                                                                                                                                                              |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                               | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                   |
| <ul style="list-style-type: none"> <li>La aplicación debió haber sido inicializada manera adecuada.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                | <p><b>Caso 1:</b> <u>Ingreso módulo Minería de Estructura Web.</u></p> <ul style="list-style-type: none"> <li>Se escoge la pestaña en la cual se especifica el ingreso al módulo de Minería de Estructura Web.</li> </ul> <p><b>Caso 2:</b> <u>Ingreso módulo Minería de Uso Web.</u></p> <ul style="list-style-type: none"> <li>Se escoge la pestaña en la cual se especifica el ingreso al módulo de Minería de Uso Web.</li> </ul> |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                        |
| <p><b>Caso 1:</b> <u>Ingreso módulo Minería de Estructura Web.</u></p> <ul style="list-style-type: none"> <li>Carga la presentación del módulo de Minería de Estructura Web e indica su interfaz o panel inicial para comenzar a trabajar.</li> </ul> <p><b>Caso 2:</b> <u>Ingreso módulo Minería de Uso Web.</u></p> <ul style="list-style-type: none"> <li>Carga la presentación del módulo de Minería de Uso Web e indica su interfaz o panel inicial para comenzar a trabajar.</li> </ul> | <p><b>Caso 1:</b> <u>Ingreso módulo Minería de Estructura Web.</u></p> <ul style="list-style-type: none"> <li>La aplicación cargó de manera satisfactoria el módulo de Minería de Estructura Web.</li> </ul> <p><b>Caso 2:</b> <u>Ingreso módulo Minería de Uso Web.</u></p> <ul style="list-style-type: none"> <li>La aplicación cargó de manera satisfactoria el módulo de Minería de Uso Web.</li> </ul>                           |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                       |

➤ Prueba de aceptación – Creación de la base de datos para POLARIS VERSIÓN 2.0.

Tabla 72. Prueba de aceptación – Creación De la base de datos para POLARIS VERSIÓN 2.0.

| PRUEBA DE ACEPTACIÓN – CREACIÓN DE LA BASE DE DATOS PARA POLARIS VERSIÓN 2.0                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                    | HU. No. 3 – Agregar tablas a la base de datos de Polaris.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                           | Crear la estructura de la base de datos para trabajar los dos tipos de Minería Web implementados en <b>POLARIS VERSIÓN 2.0</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Condiciones de Ejecución                                                                                                                                                                                                      | Entrada / Pasos de Ejecución                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <ul style="list-style-type: none"> <li>Haber ingresado a alguno de los dos módulos disponibles en la herramienta <b>POLARIS VERSIÓN 2.0</b>.</li> </ul>                                                                       | <p><b>Caso 1:</b> <u>Usuario correcto y Password correcta:</u></p> <ul style="list-style-type: none"> <li>Se ingresa un Usuario correcto.</li> <li>Se ingresa una Password correcta.</li> <li>Se pulsa la opción “Aceptar”.</li> </ul> <p><b>Caso 2:</b> <u>Usuario correcto y Password incorrecta:</u></p> <ul style="list-style-type: none"> <li>Se ingresa un Usuario correcto.</li> <li>Se ingresa una Password incorrecta.</li> <li>Se pulsa la opción “Aceptar”.</li> </ul> <p><b>Caso 3:</b> <u>Usuario incorrecto y Password correcta:</u></p> <ul style="list-style-type: none"> <li>Se ingresa un Usuario incorrecto.</li> <li>Se ingresa una Password correcta.</li> <li>Se pulsa la opción “Aceptar”.</li> </ul> <p><b>Caso 4:</b> <u>Usuario incorrecto y Password incorrecta:</u></p> <ul style="list-style-type: none"> <li>Se ingresa un Usuario incorrecto.</li> <li>Se ingresa una Password incorrecta.</li> <li>Se pulsa la opción “Aceptar”.</li> </ul> |
| Resultados Esperados                                                                                                                                                                                                          | Evaluación de la Prueba                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <p><b>Caso 1:</b> <u>Usuario correcto y Password correcta:</u></p> <ul style="list-style-type: none"> <li>El sistema debe realizar la validación de los datos ingresados.</li> <li>El sistema debe informar que se</li> </ul> | <p><b>Caso 1:</b> <u>Usuario correcto y Password correcta:</u></p> <ul style="list-style-type: none"> <li>El sistema inició la validación de los datos ingresados.</li> <li>El sistema informa que la conexión a</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>creó la estructura de la base de datos de manera satisfactoria.</p> <ul style="list-style-type: none"> <li>• Se debe cargar la interfaz del aplicativo para iniciar a trabajar.</li> </ul> <p><b>Caso 2: <u>Usuario correcto y Password incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• El sistema debe realizar la validación de los datos ingresados.</li> <li>• El sistema informa que la autenticación fue incorrecta.</li> <li>• Se presenta la ventana de conexión a la base de datos para que el usuario lo intente nuevamente.</li> </ul> <p><b>Caso 3: <u>Usuario incorrecto y Password correcta:</u></b></p> <ul style="list-style-type: none"> <li>• El sistema debe realizar la validación de los datos ingresados.</li> <li>• El sistema informa que la autenticación fue incorrecta.</li> <li>• Se presenta la ventana de conexión a la base de datos para que el usuario lo intente nuevamente.</li> </ul> <p><b>Caso 4: <u>Usuario incorrecto y Password incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• El sistema debe realizar la validación de los datos ingresados.</li> <li>• El sistema informa que la autenticación fue incorrecta.</li> <li>• Se presenta la ventana de conexión a la base de datos para que el usuario lo intente nuevamente.</li> </ul> | <p>la base de datos se realizó de manera correcta.</p> <ul style="list-style-type: none"> <li>• Inició el proceso de cargue de la venta principal del módulo.</li> <li>• Se cargó la ventana principal del módulo a trabajar.</li> </ul> <p><b>Caso 2: <u>Usuario correcto y Password incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• Se inició la validación de los datos.</li> <li>• El sistema muestro un Mensaje de Error.</li> <li>• Carga nuevamente la ventana de conexión a la base de datos.</li> </ul> <p><b>Caso 3: <u>Usuario incorrecto y Password correcta:</u></b></p> <ul style="list-style-type: none"> <li>• Se inició la validación de los datos.</li> <li>• El sistema muestro un Mensaje de Error.</li> <li>• Carga nuevamente la ventana de conexión a la base de datos.</li> </ul> <p><b>Caso 4: <u>Usuario incorrecto y Password incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• Se inició la validación de los datos.</li> <li>• El sistema muestro un Mensaje de Error.</li> <li>• Carga nuevamente la ventana de conexión a la base de datos.</li> </ul> |
| <p><b>Calificación de Aprobación de la Prueba</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <p>Prueba Aprobada</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

➤ **Prueba de aceptación – Ingreso de la URL a analizar**

Tabla 73. Prueba de aceptación – Ingreso de la URL a analizar.

| <b>PRUEBA DE ACEPTACIÓN – INGRESO DE LA URL A ANALIZAR</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | HU. No. 6 – URL a analizar                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Ingresar la URL a analizar y verificar la valides de la misma.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <ul style="list-style-type: none"> <li>• Haberse realizado la conexión a la base de datos de manera satisfactoria.</li> <li>• Haber ingresado al Módulo de Minería de Estructura Web.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <p><b>Caso 1: <u>URL correcta:</u></b></p> <ul style="list-style-type: none"> <li>• Se ingresa la URL a analizar.</li> <li>• Se pulsa la opción “Analizar”.</li> </ul> <p><b>Caso 2: <u>URL incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• Se ingresa la URL a analizar.</li> <li>• Se pulsa la opción “Analizar”.</li> </ul>                                                                                                                                                                                                                                                                                                                                 |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <p><b>Caso 1: <u>URL correcta:</u></b></p> <ul style="list-style-type: none"> <li>• La herramienta realiza la petición de la URL ingresada.</li> <li>• Se encuentra de manera satisfactoria el portal Web solicitado.</li> <li>• Se activan los parámetros de configuración para iniciar a trabajar con el Crawler</li> </ul> <p><b>Caso 2: <u>URL incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• La herramienta realiza la petición de la URL ingresada.</li> <li>• No se encuentra la URL especificada.</li> <li>• Se notifica que la URL no existe o se ingresó de manera incorrecta.</li> <li>• Se solicita una URL nueva.</li> </ul> | <p><b>Caso 1: <u>URL correcta:</u></b></p> <ul style="list-style-type: none"> <li>• Se inició la validación de la URL.</li> <li>• Se encontró de manera satisfactoria la URL ingresada.</li> <li>• Se activaron las herramientas para la configuración de los diferentes parámetros del Web Crawler.</li> </ul> <p><b>Caso 2: <u>URL incorrecta:</u></b></p> <ul style="list-style-type: none"> <li>• Se inició la validación de la URL</li> <li>• No se encuentra la página Web ingresada.</li> <li>• Se informa al usuario que el formato o la dirección no existe o el formato ingresado no es válido.</li> <li>• Se solicitó una URL nueva y de manera adecuada.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

➤ **Prueba de aceptación – Nivel de profundidad de la búsqueda**

Tabla 74. Prueba de aceptación – Nivel de profundidad de la búsqueda.

| <b>PRUEBA DE ACEPTACIÓN – NIVEL DE PROFUNDIDAD DE LA BÚSQUEDA</b>                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | HU. No. 5 – Nivel de profundidad del Web Crawler                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Escoger el nivel de búsqueda en la Web para generar el grafo.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <ul style="list-style-type: none"> <li>Se debe haber ingresado una URL correcta.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                 | <p><b>Caso 1: <u>Nivel de profundidad correcto:</u></b></p> <ul style="list-style-type: none"> <li>Se ingresa el nivel de profundidad.</li> <li>Se valida el valor del nivel de profundidad.</li> </ul> <p><b>Caso 2: <u>Nivel de profundidad incorrecto:</u></b></p> <ul style="list-style-type: none"> <li>Se ingresa el nivel de profundidad.</li> <li>Se valida el valor del nivel de profundidad.</li> </ul>                                                                                                                                                 |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <p><b>Caso 1: <u>Nivel de profundidad correcto:</u></b></p> <ul style="list-style-type: none"> <li>Se valida el valor ingresado en el nivel de profundidad.</li> <li>Se continua con la configuración del Crawler</li> </ul> <p><b>Caso 2: <u>Nivel de profundidad incorrecto:</u></b></p> <ul style="list-style-type: none"> <li>Se valida el valor ingresado en el nivel de profundidad.</li> <li>Se notifica que el valor no es válido.</li> <li>Solicita nuevamente el nivel de profundidad.</li> </ul> | <p><b>Caso 1: <u>Nivel de profundidad correcto:</u></b></p> <ul style="list-style-type: none"> <li>Se validó el nivel de profundidad.</li> <li>Se detectó que el valor ingresado es correcto.</li> <li>Se procedió con el siguiente paso de configuración del Web Crawler.</li> </ul> <p><b>Caso 2: <u>Nivel de profundidad incorrecto:</u></b></p> <ul style="list-style-type: none"> <li>Se validó el nivel de profundidad.</li> <li>Se detectó que el valor ingresado no es válido.</li> <li>Se solicitó nuevamente otro valor para este parámetro.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

➤ Prueba de aceptación – Filtros para el Web Crawler

Tabla 75. Prueba de aceptación – Filtros para el Web Crawler

| PRUEBA DE ACEPTACIÓN – FILTROS PARA EL WEB CRAWLER                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                      | HU. No. 7 – Filtros para la búsqueda del Crawler                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                             | Escoger los filtros por medio de casillas de verificación para la generación del grafo, los filtros que se pueden usar si se desean son de protocolo Web y de dominio Web.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Condiciones de Ejecución                                                                                                                                                                        | Entrada / Pasos de Ejecución                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <ul style="list-style-type: none"> <li>• Se debe haber ingresado una URL correcta.</li> <li>• Se debe haber ingresado un valor valido en el nivel de profundidad.</li> </ul>                    | <p><b>Caso 1:</b> <u>Filtrado por protocolo y dominio:</u></p> <ul style="list-style-type: none"> <li>• Seleccionar las casillas de verificación para el filtro de dominio Web.</li> <li>• Escoger otro tipo de dominio en caso de no encontrarse el deseado.</li> <li>• Ingresar el nuevo filtro de dominio.</li> <li>• Seleccionar las casillas de verificación para el filtro de protocolo Web.</li> </ul> <p><b>Caso 2:</b> <u>Filtrado por protocolo:</u></p> <ul style="list-style-type: none"> <li>• Seleccionar las casillas de verificación para el filtro de protocolo Web.</li> <li>• Dejar en blanco los filtros de dominio.</li> </ul> <p><b>Caso 3:</b> <u>Filtrado por dominio:</u></p> <ul style="list-style-type: none"> <li>• Seleccionar las casillas de verificación para el filtro de dominio Web.</li> <li>• Escoger otro tipo de dominio en caso de no encontrarse el deseado.</li> <li>• Ingresar el nuevo filtro de dominio.</li> <li>• Dejar en blanco los filtros de protocolo.</li> </ul> <p><b>Caso 4:</b> <u>Sin filtros:</u></p> <ul style="list-style-type: none"> <li>• Dejar en blanco los filtros de dominio y protocolo Web.</li> </ul> |
| Resultados Esperados                                                                                                                                                                            | Evaluación de la Prueba                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <p><b>Caso 1:</b> <u>Filtrado por protocolo y dominio:</u></p> <ul style="list-style-type: none"> <li>• Las casillas de verificación del filtro por protocolo escogidas deben quedar</li> </ul> | <p><b>Caso 1:</b> <u>Filtrado por protocolo y dominio:</u></p> <ul style="list-style-type: none"> <li>• Se activaron las casillas de verificación escogidas para el filtrado</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>marcadas.</p> <ul style="list-style-type: none"> <li>• Las casillas de verificación del filtro por dominio escogidas deben quedar marcadas.</li> <li>• En caso de no encontrar un filtro por dominio adecuado el usuario ingresa el deseado.</li> <li>• Se verifica el la validez del nuevo dominio.</li> <li>• Se termina la configuración del Crawler.</li> </ul> <p><b>Caso 2: <u>Filtrado por protocolo:</u></b></p> <ul style="list-style-type: none"> <li>• Las casillas de verificación del filtro por protocolo escogidas deben quedar marcadas.</li> <li>• Se termina la configuración del Crawler.</li> </ul> <p><b>Caso 3: <u>Filtrado por dominio:</u></b></p> <ul style="list-style-type: none"> <li>• Las casillas de verificación del filtro por dominio escogidas deben quedar marcadas.</li> <li>• En caso de no encontrar un filtro por dominio adecuado el usuario ingresa el deseado.</li> <li>• Se verifica el la validez del nuevo dominio.</li> <li>• Se termina la configuración del Crawler.</li> </ul> <p><b>Caso 4: <u>Sin filtros:</u></b></p> <ul style="list-style-type: none"> <li>• Las casillas de verificación de los filtros de dominio y protocolo Web se deben encontrar en blando (Al no existir filtros se trabaja con todos los protocolos y dominios Web)</li> <li>• Se termina la configuración del Crawler.</li> </ul> | <p>por protocolo.</p> <ul style="list-style-type: none"> <li>• Se activaron las casillas de verificación escogidas para el filtrado por dominio.</li> <li>• Se ingresó un dominio nuevo.</li> <li>• Se validó el dominio nuevo.</li> <li>• Terminó la configuración del Crawler.</li> </ul> <p><b>Caso 2: <u>Filtrado por protocolo:</u></b></p> <ul style="list-style-type: none"> <li>• Se activaron las casillas de verificación escogidas para el filtrado por protocolo.</li> <li>• Terminó la configuración del Crawler.</li> </ul> <p><b>Caso 3: <u>Filtrado por dominio:</u></b></p> <ul style="list-style-type: none"> <li>• Se activaron las casillas de verificación escogidas para el filtrado por dominio.</li> <li>• Se ingresó un dominio nuevo.</li> <li>• Se validó el dominio nuevo.</li> <li>• Terminó la configuración del Crawler.</li> </ul> <p><b>Caso 4: <u>Sin filtros:</u></b></p> <ul style="list-style-type: none"> <li>• Las casillas de verificación de los filtros tanto por protocolo como por dominio quedaron sin activar.</li> <li>• Terminó la configuración del Crawler.</li> </ul> |
| <p><b>Calificación de Aprobación de la Prueba</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p>Prueba Aprobada</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

➤ **Prueba de aceptación – Recuperación de la información por medio del Crawler**

Tabla 76. Prueba de aceptación – Recuperación de la información por medio del Crawler.

| <b>PRUEBA DE ACEPTACIÓN – RECUPERACIÓN DE LA INFORMACIÓN POR MEDIO DEL CRAWLER</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | HU. No. 4 – Generación de un Web Crawler                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Realiza el recorrido por la Web, tomando como inicio la URL ingresado por el usuario, el grafo que toma lo hace de manera real, los filtros y cantidad de enlaces por nodo se limitan únicamente para generar el gráfico.                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <ul style="list-style-type: none"> <li>• Se debe haber ingresado una URL correcta.</li> <li>• Se debe haber ingresado un valor valido en el nivel de profundidad.</li> <li>• Se deben haber escogidos los filtros que el usuario desee.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>• Pulsar la opción “Iniciar”.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <ul style="list-style-type: none"> <li>• Se realiza la solitud de la URL ingresada por el usuario.</li> <li>• Se solicita el código fuente HTML de la URL ingresada por el usuario.</li> <li>• Se recorre el código fuente HTML y se seleccionan todos los links de esta página.</li> <li>• Se añaden los link encontrados a una lista doblemente enlazada, siendo un nodo cada página nueva que se encuentre en el recorrido.</li> <li>• Posteriormente se ingresa a la primera página que se haya encontrado el código HTML de la página semilla.</li> <li>• El procedimiento se sigue sucesivamente dependiendo del nivel de profundidad que se haya establecido.</li> </ul> | <ul style="list-style-type: none"> <li>• Se solicitó de manera satisfactoria la URL ingresada por el usuario.</li> <li>• Se capturó el código HTML de la página semilla o inicial.</li> <li>• Se eliminaron los links repetidos que se hayan encontrado en la página semilla.</li> <li>• Se realizó la recuperación de la información a las páginas nuevas que han sido encontradas.</li> <li>• El procedimiento se repitió hasta el nivel de profundidad escogido por parte del usuario.</li> <li>• La barra de progreso demostró que se terminó de realizar la recuperación de la información.</li> <li>• Se activaron los botones para mirar el grafo generado.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

➤ **Prueba de aceptación – Guardar los datos recuperados por el Crawler**

Tabla 77. Prueba de aceptación – Recuperar los datos recuperados por el Crawler.

| <b>PRUEBA DE ACEPTACIÓN – GUARDAR LOS DATOS RECUPERADOS POR EL CRAWLER</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | HU. No. 8 – Guardar datos del Web Crawler                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Los datos obtenidos por el Crawler deben ser guardados de manera permanente en la base de datos de la herramienta o en un archivo plano.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | <p><b>Caso 1: <u>Guardar:</u></b></p> <ul style="list-style-type: none"> <li>• Presionar sobre el icono “Configurar Crawler”.</li> <li>• Escoger la opción “guardar”.</li> </ul> <p><b>Caso 2: <u>Exportar:</u></b></p> <ul style="list-style-type: none"> <li>• Presionar sobre el icono “Configurar Crawler”.</li> <li>• Escoger la opción “exportar datos”.</li> <li>• Seleccionar la ruta en la cual se desea guardar el archivo.</li> <li>• Presionar en “guardar”</li> </ul>                                                                                                                                                                                                                |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <p><b>Caso 1: <u>Guardar:</u></b></p> <ul style="list-style-type: none"> <li>• Se abre un menú emergente del nodo “Crawler”.</li> <li>• Al tomar la opción “Guardar” se baja la información de memoria a la estructura de la base de datos de la herramienta.</li> <li>• Se informa que la información fue guardada de manera exitosa.</li> </ul> <p><b>Caso 2: <u>Exportar:</u></b></p> <ul style="list-style-type: none"> <li>• Se abre un menú emergente del nodo “Crawler”.</li> <li>• Al tomar la opción “Exportar” se debe abrir una nueva ventana en la cual se debe seleccionar la ruta de destino para guardar el archivo.</li> <li>• Se informa que el archivo se creó de</li> </ul> | <p><b>Caso 1: <u>Guardar:</u></b></p> <ul style="list-style-type: none"> <li>• Se desplegó el menú emergente del nodo Crawler.</li> <li>• Se seleccionó la opción “Guardar”.</li> <li>• Una nueva ventana apareció informando que se guardaron los datos de manera satisfactoria.</li> </ul> <p><b>Caso 2: <u>Exportar:</u></b></p> <ul style="list-style-type: none"> <li>• Se desplegó el menú emergente del nodo Crawler.</li> <li>• Se seleccionó la opción “Exportar”.</li> <li>• Apareció una nueva venta para seleccionar la ruta donde se va a guardar el archivo.</li> <li>• Se escribió el nombre con el cual se desea guardar el archivo.</li> <li>• Se presionó “Guardar”.</li> </ul> |

|                                                |                                                                                                                                         |
|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| manera satisfactoria.                          | <ul style="list-style-type: none"> <li>• Una nueva ventana apareció confirmando que el archivo se guardó satisfactoriamente.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b> | Prueba Aprobada                                                                                                                         |

➤ **Prueba de aceptación – Visualización de los diferentes tipos de grafos.**

Tabla 78. Prueba de aceptación – Visualización de los diferentes tipos de grafos.

| <b>PRUEBA DE ACEPTACIÓN – VISUALIZACIÓN DE LOS DIFERENTES TIPOS DE GRAFOS</b>                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                 | HU. No. 10 – Visualización de grafos                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                        | Debe presentar los diferentes tipos grafos, en el cual se representa la Web teniendo como origen la URL ingresada por el usuario.                                                                                                                                                                                                                                                                                                                            |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                            | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Presionar en “Generar gráfico”</li> <li>• Escoger el tipo de gráfico que se desea mirar.</li> <li>• Ampliar el gráfico si se lo desea.</li> </ul>                                                                                                                                                                                                                                                                   |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <ul style="list-style-type: none"> <li>• Indicar el tipo de gráficos que se pueden generar.</li> <li>• Después de escoger uno de los tipos de gráfico, se deben generar en el área de trabajo.</li> <li>• Al presionar en la opción ampliar se debe cargar el gráfico en una nueva ventana para ser manipulado de manera independiente.</li> <li>• Cerrar la ventana al momento de terminar de usar el gráfico.</li> </ul> | <ul style="list-style-type: none"> <li>• Del menú o árbol se tomó uno de los tipos de gráficos que se pueden visualizar.</li> <li>• Se desplegó el gráfico en el área de trabajo, con la opción ampliar.</li> <li>• Al presionar la opción ampliar, el grafo se presentó en una ventana aparte.</li> <li>• Cada nodo del grafo tenía su información correspondiente, información que ha sido obtenida hasta el momento únicamente por el Crawler.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                             | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                              |

➤ **Prueba de aceptación – Visualización de los enlaces internos y externos.**

Tabla 79. Prueba de aceptación – Visualización de los enlaces internos y externos.

| <b>PRUEBA DE ACEPTACIÓN – VISUALIZACIÓN DE LOS ENLACES INTERNOS Y EXTERNOS</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | HU. No. 15 – Enlaces internos y externos                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Se debe visualizar el grafo únicamente con los enlaces internos o externos de la URL ingresada por el usuario.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <p><b>Caso 1: <u>Enlaces internos:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces internos.</li> <li>• Generar el grafo.</li> </ul> <p><b>Caso 2: <u>Enlaces externos:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces Externos.</li> <li>• Generar el grafo.</li> </ul> <p><b>Caso 3: <u>Enlaces internos y externos:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces internos.</li> <li>• Marcar la casilla de verificación de los enlaces externos.</li> <li>• Generar el grafo.</li> </ul> |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <p><b>Caso 1: <u>Enlaces internos:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces internos.</li> <li>• Se genera los diferentes gráficos únicamente con los enlaces internos de la URL ingresada por el usuario.</li> </ul> <p><b>Caso 2: <u>Enlaces externos:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces externos.</li> </ul> | <p><b>Caso 1: <u>Enlaces internos:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó únicamente la casilla de enlaces internos.</li> <li>• Se presionó el botón para generar el grafo.</li> <li>• Se obtuvo el grafo únicamente con enlaces internos.</li> </ul> <p><b>Caso 2: <u>Enlaces externos:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó únicamente la casilla de</li> </ul>                                                                                                               |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Se genera los diferentes gráficos únicamente con los enlaces externos de la URL ingresada por el usuario.</li> </ul> <p><b>Caso 3: <u>Enlaces internos y externos:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces internos.</li> <li>• Se activa la casilla de verificación de los enlaces externos.</li> <li>• Se genera los diferentes gráficos con los enlaces internos como externos de la URL ingresada por el usuario.</li> </ul> | <ul style="list-style-type: none"> <li>enlaces externos.</li> <li>• Se presionó el botón para generar el grafo.</li> <li>• Se obtuvo el grafo únicamente con enlaces externos.</li> </ul> <p><b>Caso 3: <u>Enlaces internos y externos:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó las casillas de enlaces internos y externos.</li> <li>• Se presionó el botón para generar grafo.</li> <li>• Se obtuvo el grafo completo tanto con enlaces internos como externos.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

➤ **Prueba de aceptación – Visualización de los enlaces entrantes y salientes.**

Tabla 80. Prueba de aceptación – Visualización de los enlaces entrantes y salientes.

| <b>PRUEBA DE ACEPTACIÓN – VISUALIZACIÓN DE LOS ENLACES ENTRANTES Y SALIENTES</b>                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                       | HU. No. 16 – Enlaces entrantes y salientes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                              | Se debe visualizar el grafo únicamente con los enlaces entrantes o salientes de la URL ingresada por el usuario.                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Condiciones de Ejecución</b>                                                                                                                  | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul> | <p><b>Caso 1: <u>Enlaces entrantes:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces entrantes.</li> <li>• Generar el grafo.</li> </ul> <p><b>Caso 2: <u>Enlaces salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces Salientes.</li> <li>• Generar el grafo.</li> </ul> <p><b>Caso 3: <u>Enlaces entrantes y salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de enlaces entrantes.</li> </ul> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <ul style="list-style-type: none"> <li>• Marcar la casilla de verificación de los enlaces salientes.</li> <li>• Generar el grafo.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <p><b>Caso 1: <u>Enlaces entrantes:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces entrantes.</li> <li>• Se genera los diferentes gráficos únicamente con los enlaces entrantes de la URL ingresada por el usuario.</li> </ul> <p><b>Caso 2: <u>Enlaces salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces salientes.</li> <li>• Se genera los diferentes gráficos únicamente con los enlaces salientes de la URL ingresada por el usuario.</li> </ul> <p><b>Caso 3: <u>Enlaces entrantes y salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Se desprende el menú emergente de la configuración del Crawler.</li> <li>• Se activa la casilla de verificación de los enlaces entrantes.</li> <li>• Se activa la casilla de verificación de los enlaces salientes.</li> <li>• Se genera los diferentes gráficos con los enlaces entrantes como salientes de la URL ingresada por el usuario.</li> </ul> | <p><b>Caso 1: <u>Enlaces entrantes:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó únicamente la casilla de enlaces entrantes.</li> <li>• Se presionó el botón para generar el grafo.</li> <li>• Se obtuvo el grafo únicamente con enlaces entrantes.</li> </ul> <p><b>Caso 2: <u>Enlaces salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó únicamente la casilla de enlaces salientes.</li> <li>• Se presionó el botón para generar el grafo.</li> <li>• Se obtuvo el grafo únicamente con enlaces salientes.</li> </ul> <p><b>Caso 3: <u>Enlaces entrantes y salientes:</u></b></p> <ul style="list-style-type: none"> <li>• Se activó el menú de configuración del Crawler.</li> <li>• Se marcó las casillas de enlaces entrantes y salientes.</li> <li>• Se presionó el botón para generar grafo.</li> <li>• Se obtuvo el grafo completo tanto con enlaces entrantes como salientes.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

➤ **Prueba de aceptación – Visualización del grafo teniendo en cuenta los filtros.**

Tabla 81. Prueba de aceptación – Visualización del grafo teniendo en cuenta los filtros

| <b>PRUEBA DE ACEPTACIÓN – VISUALIZACIÓN DEL GRAFO TENIENDO EN CUENTA LOS FILTROS</b>                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                                               | HU. No. 18 – Visualización de grafos teniendo en cuenta filtros y parámetros.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                                      | Se debe visualizar el grafo teniendo en cuenta los filtros establecidos por el usuario (protocolo, dominio y enlaces)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                          | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>• Configurar el Crawler</li> <li>• Escoger el tipo de protocolo que se desea trabajar (Uno o varios).</li> <li>• Escoger el tipo de enlaces con los cuales se desea trabajar (internos, externos, entrantes, salientes).</li> <li>• Visualizar el grafo.</li> </ul>                                                                                                                                                                                                                                                                                                  |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                                              | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>• Se debe desplegar el menú de configuración del Crawler.</li> <li>• Escoger el tipo de filtro en cuanto a protocolo o dominio Web se refiere.</li> <li>• Configurar el nodo de enlaces.</li> <li>• Activar las casillas de verificación de el tipo de enlaces que se desea trabajar (cualquier combinación es posible).</li> <li>• Visualizar el grafo únicamente con la información requerida por parte del usuario.</li> </ul> | <ul style="list-style-type: none"> <li>• Se desplegó el menú de configuración del nodo Crawler.</li> <li>• Se desplegó la ventana emergente de configuración.</li> <li>• Se escogió el tipo de protocolo y el tipo de dominio Web con los cuales se desea trabajar (pueden ser uno o varios).</li> <li>• Se desplegó el menú de configuración de enlaces.</li> <li>• Se activó las casillas de verificación de los tipos de enlaces con los cuales se quiere trabajar.</li> <li>• Se visualiza el grafo de manera satisfactoria, únicamente con la información solicitada por parte del usuario.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                           | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

➤ **Prueba de aceptación – Configuración del nodo Page Rank**

Tabla 82. Prueba de aceptación – Configuración del nodo Page Rank.

| <b>PRUEBA DE ACEPTACIÓN – CONFIGURACIÓN DEL NODO PAGE RANK</b>                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                                                                                            | HU. No. 22 – Parámetros Page Rank.                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                                                                                   | Se debe configurar el algoritmo Page Rank para visualizar el valor de las páginas tanto en el grafo como en las tablas.                                                                                                                                                                                                                                                                                                                          |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                       | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Desplegar el menú de configuración del nodo Page Rank.</li> <li>• Ingresar un número de iteraciones por medio del Spinner o manualmente.</li> <li>• Ingresar el Dampening por medio del Spinner o manualmente.</li> <li>• Escoger la opción correr.</li> </ul>                                                                                                                                          |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                                                           | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <ul style="list-style-type: none"> <li>• Se debe desplegar el menú de configuración del nodo Page Rank.</li> <li>• Se ingresa el número de iteraciones el cual se encuentra validado para que únicamente sean valores posibles.</li> <li>• Se ingresa el número de Dampening el cual se encuentra validado para que únicamente sean valores posibles.</li> <li>• Se presiona la opción “Run” para que se realice el cálculo del algoritmo.</li> </ul> | <ul style="list-style-type: none"> <li>• Se desplegó el menú de configuración del nodo Page Rank.</li> <li>• Por medio del Spinner se tomo un valor para la cantidad de iteraciones. (70)</li> <li>• Por medio del Spinner se toma un valor para la cantidad de Dampening (0.85)</li> <li>• Se presionó “Run” para ejecutar el algoritmo.</li> <li>• Los nodos del grafo ahora cuentan con el peso de sus páginas de manera adecuada.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                        | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                  |

➤ **Prueba de aceptación – Configuración del nodo HITS**

Tabla 83. Prueba de aceptación – Configuración del nodo HITS.

| <b>PRUEBA DE ACEPTACIÓN – CONFIGURACIÓN DEL NODO HITS</b>                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                 | HU. No. 23 – Parámetros HITS                                                                                                                                                                                                                                                                                                                         |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                        | Se debe configurar el algoritmo HITS para visualizar el valor de las páginas tanto en el grafo como en las tablas.                                                                                                                                                                                                                                   |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                            | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                  |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul>                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Desplegar el menú de configuración del nodo HITS.</li> <li>• Ingresar un número de iteraciones por medio del Spinner o manualmente.</li> <li>• Escoger la opción “Run”.</li> </ul>                                                                                                                          |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                       |
| <ul style="list-style-type: none"> <li>• Se debe desplegar el menú de configuración del nodo HITS.</li> <li>• Se ingresa el número de iteraciones el cual se encuentra validado para que únicamente sean valores posibles.</li> <li>• Se presiona la opción “Run” para que se realice el cálculo del algoritmo.</li> </ul> | <ul style="list-style-type: none"> <li>• Se desplegó el menú de configuración del nodo HITS.</li> <li>• Por medio del Spinner se tomo un valor para la cantidad de iteraciones. (35)</li> <li>• Se presionó “Run” para ejecutar el algoritmo.</li> <li>• Los nodos del grafo ahora cuentan con el peso de sus páginas de manera adecuada.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                             | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                      |

➤ **Prueba de aceptación – Configuración del nodo Web Linker.**

Tabla 84. Prueba de aceptación – Configuración del nodo Web Linker.

| <b>PRUEBA DE ACEPTACIÓN – CONFIGURACIÓN DEL NODO LINKER WEB</b>                                                                                  |                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                       | HU. No. 24 – Parámetros Algoritmo creado                                                                                                               |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                              | Se debe configurar el algoritmo Creado para visualizar el valor de las páginas tanto en el grafo como en las tablas.                                   |
| <b>Condiciones de Ejecución</b>                                                                                                                  | <b>Entrada / Pasos de Ejecución</b>                                                                                                                    |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> </ul> | <ul style="list-style-type: none"> <li>• Desplegar el menú de configuración del nodo Algoritmo creado.</li> <li>• Escoger la opción correr.</li> </ul> |

| Resultados Esperados                                                                                                                                                                                           | Evaluación de la Prueba                                                                                                                                                                                                                                                 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Se debe desplegar el menú de configuración del nodo Algoritmo Creado.</li> <li>• Se presiona la opción “Run” para que se realice el cálculo del algoritmo.</li> </ul> | <ul style="list-style-type: none"> <li>• Se desplegó el menú de configuración del nodo Algoritmo Creado</li> <li>• Se presionó “Run” para ejecutar el algoritmo.</li> <li>• Los nodos del grafo ahora cuentan con el peso de sus páginas de manera adecuada.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                 | Prueba Aprobada                                                                                                                                                                                                                                                         |

➤ **Prueba de aceptación – Visualización de la información por cada nodo**

Tabla 85. Prueba de aceptación – Visualización de la información por cada nodo.

| <b>PRUEBA DE ACEPTACIÓN – VISUALIZACIÓN DE LA INFORMACIÓN POR CADA NODO</b>                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                     | HU. No. 25 – Visualización de información de cada nodo.                                                                                                                                                                                                                                                                                                                                                         |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                            | Al presionar sobre cualquier nodo del grafo, debe presentar la información de cada uno de estos (calificación generada por los diferentes algoritmos, URL)                                                                                                                                                                                                                                                      |
| Condiciones de Ejecución                                                                                                                                                                                                                                                                       | Entrada / Pasos de Ejecución                                                                                                                                                                                                                                                                                                                                                                                    |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> <li>• Si se desea mirar el peso de los nodos, se debe haber corrido previamente alguno de los algoritmos de Minería de Estructura Web.</li> </ul>   | <ul style="list-style-type: none"> <li>• Generar el grafo en cualquier momento, indiferentes de los filtros que tenga o no aplicado.</li> <li>• Dar doble clic sobre el nodo al cual queremos acceder.</li> </ul>                                                                                                                                                                                               |
| Resultados Esperados                                                                                                                                                                                                                                                                           | Evaluación de la Prueba                                                                                                                                                                                                                                                                                                                                                                                         |
| <ul style="list-style-type: none"> <li>• Se mira el grafo.</li> <li>• Al dar doble clic sobre cualquier nodo del grafo se debe desplegar una nueva ventana.</li> <li>• En la ventana se muestra los valores de los algoritmos aplicados, la URL y un hipervínculo a esta dirección.</li> </ul> | <ul style="list-style-type: none"> <li>• Se observo el grafo en el área de trabajo.</li> <li>• Al dar doble clic se desplegó una nueva ventana.</li> <li>• En la ventana aparecen los valores del nodo identificándose cada uno de los algoritmos implementados.</li> <li>• Aparece la URL a la cual hace parte.</li> <li>• La URL genera un hipervínculo al sitio Web para acceder a esta dirección</li> </ul> |

|                                                |                      |
|------------------------------------------------|----------------------|
|                                                | desde la aplicación. |
| <b>Calificación de Aprobación de la Prueba</b> | Prueba Aprobada      |

➤ **Prueba de aceptación – La ruta más corta.**

Tabla 86. Prueba de aceptación – La ruta más corta.

| <b>PRUEBA DE ACEPTACIÓN – LA RUTA MAS CORTA</b>                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                                                                         | HU. No. 28 – Algoritmo de la ruta más corta entre dos nodos seleccionados.                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                                                                | Se seleccionan dos nodos del grafo y se debe indicar la ruta más corta para desplazarse de un nodo al otro.                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                    | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> <li>• Establecer la conexión entre el algoritmo Dijkstra y el grafo.</li> </ul>                                                                                                                                                         | <ul style="list-style-type: none"> <li>• Se configura el nodo Dijkstra y se ejecuta</li> <li>• Del grafo desplegado se selecciona un nodo de origen y un nodo destino.</li> </ul>                                                                                                                                                                                                                                                                                                                                           |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                        | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>• Se debe desplegar el menú emergente de la configuración del nodo Dijkstra.</li> <li>• Al correr el algoritmo debe cargarse la imagen del grafo.</li> <li>• Al marcar el nodo origen y el nodo destino estos deben quedar en un color diferente al actual.</li> <li>• Se marca la ruta más corta posible entre estos dos nodos.</li> </ul> | <ul style="list-style-type: none"> <li>• El menú emergente del nodo Dijkstra de desplegó sin ningún inconveniente.</li> <li>• Se presionó “Run” para ejecutar el algoritmo.</li> <li>• Se generó el grafo partiendo de la URL digitada por el usuario.</li> <li>• Al marcar el nodo origen y el destino se colocaron de otro color con la finalidad de diferenciarlos.</li> <li>• Se marcó la ruta más corta entre estos dos nodos, marcándose los enlaces por los cuales toca hacer el recorrido en otro color.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                     | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

➤ **Prueba de aceptación – Reportes en tablas.**

Tabla 87. Prueba de aceptación – Reportes en tablas.

| <b>PRUEBA DE ACEPTACIÓN – REPORTES EN TABLAS</b>                                                                                                                                                                                                    |                                                                                                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                          | HU. No. 30 – Reporte en tablas de los cálculos.                                                                                                                                                                                                           |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                 | La información calculada por parte de los algoritmos se presenta por medio de tablas para que el usuario la pueda imprimir o guardar.                                                                                                                     |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                     | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                       |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> <li>• Se debe haber ejecutado como mínimo uno de los algoritmos de minería de estructura Web.</li> </ul> | <ul style="list-style-type: none"> <li>• En el menú de vistas se toma la opción tablas.</li> <li>• Se ejecuta la visualización de la tabla.</li> </ul>                                                                                                    |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                         | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                            |
| <ul style="list-style-type: none"> <li>• En una ventana nueva debe aparecer la información de cada uno de los nodos, agrupados por nivel y con el peso que se ha sido otorgado por los diferentes algoritmos que se hayan ejecutado.</li> </ul>     | <ul style="list-style-type: none"> <li>• Se generó una nueva ventana en la cual se encuentran los nodos agrupados por nivel, además se presenta la información de el puntaje obtenido por parte de los algoritmos de Minería e Estructura Web.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                      | Prueba Aprobada                                                                                                                                                                                                                                           |

➤ **Prueba de aceptación – Exportar datos.**

Tabla 88. Prueba de aceptación – Exportar datos.

| <b>PRUEBA DE ACEPTACIÓN – EXPORTAR DATOS</b>                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                        | HU. No. 31 – Exportación de datos en otros formatos.                                                                                                                                                                                                                                                     |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                               | La información calculada por parte de los algoritmos se presenta por medio de tablas para que el usuario la pueda imprimir o guardar.                                                                                                                                                                    |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                   | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                                                      |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> <li>• Se debe haber ejecutado como mínimo uno de los algoritmos de minería de estructura Web.</li> <li>• Haber escogido una de las tablas o</li> </ul> | <ul style="list-style-type: none"> <li>• En el menú de vistas se tomar la opción tablas.</li> <li>• Se ejecuta la visualización de la tabla.</li> <li>• Escoger la opción exportar y el formato en el cual se desea exportar la información.</li> <li>• Escoger la ruta destino donde se va a</li> </ul> |

|                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| vista la cual se quiera exportar.                                                                                                                                                                                                                                                                                                                                                                                   | guardar el archivo.                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                                                                                                         | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <ul style="list-style-type: none"> <li>• Se debe visualizar la tabla la cual se desea exportar.</li> <li>• Presionar sobre el botón “exportar”.</li> <li>• Se toma el tipo de formato en el cual se desea exportar el archivo.</li> <li>• Se despliega la una ventana para escoger la ruta de almacenamiento del archivo.</li> <li>• Se presiona guardar para almacenar el archivo de manera permanente.</li> </ul> | <ul style="list-style-type: none"> <li>• Una vez se visualizó la tabla de la manera expuesta en la anterior prueba de aceptación, se procedió a presionar la opción “exportar”.</li> <li>• Se seleccionó el tipo de formato que deseamos tenga el archivo.</li> <li>• Se desplegó la opción en la cual se tomó la ruta para almacenar el archivo.</li> <li>• Se presionó “Guardar” y el archivo quedo exportado en el formato deseado de manera satisfactoria.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                                      | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

➤ **Prueba de aceptación – Sugerencias para mejorar la estructura Web.**

Tabla 89. Prueba de aceptación – Sugerencias para mejorar la estructura Web.

| <b>PRUEBA DE ACEPTACIÓN – SUGERENCIAS PARA MEJORAR LA ESTRUCTURA WEB</b>                                                                                                                                                                            |                                                                                                                                                                                                                                                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                          | HU. No. 32 – Resultado en la mejora de escritura.                                                                                                                                                                                              |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                 | Se debe presentar posibles sugerencias donde se recomiende cambios posibles que conlleven a mejorar la navegabilidad de la URL digitada por el usuario.                                                                                        |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                     | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                            |
| <ul style="list-style-type: none"> <li>• Se debe haber corrido el Web Crawler para realizar la recuperación de información de la Web.</li> <li>• Se debe haber ejecutado como mínimo uno de los algoritmos de minería de estructura Web.</li> </ul> | <ul style="list-style-type: none"> <li>• Seleccionar la opción “Recomendaciones”.</li> </ul>                                                                                                                                                   |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                         | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                 |
| <ul style="list-style-type: none"> <li>• Se debe visualizar una tabla con las recomendaciones de mejorar la navegabilidad.</li> <li>• Esta tabla igualmente puede ser guardada o exportada de la misma manera como se explico en las</li> </ul>     | <ul style="list-style-type: none"> <li>• Se presionó en “Recomendaciones”.</li> <li>• Generó la tabla de recomendaciones de estructura Web.</li> <li>• Finalmente se guardó la tabla de recomendaciones de la estructura de la URL.</li> </ul> |

|                                                |                 |
|------------------------------------------------|-----------------|
| pruebas mencionadas con anterioridad.          |                 |
| <b>Calificación de Aprobación de la Prueba</b> | Prueba Aprobada |

### Prueba de aceptación – Crear proyecto

Tabla 90. Prueba de aceptación – Crear proyecto.

| <b>PRUEBA DE ACEPTACIÓN – CREAR PROYECTO</b>                                                                                                                                                                             |                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                               | HU. No. 34 – Crear proyecto nuevo                                                                                                                                                                                                                                          |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                      | Si el usuario lo desea se puede generar un nuevo proyecto.                                                                                                                                                                                                                 |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                          | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>Haber cargado la herramienta <b>POLARIS VERSIÓN 2.0</b></li> </ul>                                                                                                                | <ul style="list-style-type: none"> <li>Seleccionar la opción “nuevo proyecto”</li> </ul>                                                                                                                                                                                   |
| <b>Resultados Esperados</b>                                                                                                                                                                                              | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                             |
| <ul style="list-style-type: none"> <li>Al presionar sobre el icono de nuevo proyecto se preguntará si está seguro de hacerlo.</li> <li>Aparece el área de trabajo en limpio para hacer uso de la herramienta.</li> </ul> | <ul style="list-style-type: none"> <li>Se presionó la opción “nuevo proyecto”.</li> <li>Se desplegó un mensaje de advertencia en el cual el usuario confirma la creación de un nuevo proyecto.</li> <li>El área de trabajo queda en limpio y lista para usarse.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                           | Prueba Aprobada                                                                                                                                                                                                                                                            |

### ➤ Prueba de aceptación – Cargar proyecto

Tabla 91. Prueba de aceptación – Cargar proyecto.

| <b>PRUEBA DE ACEPTACIÓN – CARGAR PROYECTO</b>                                                                                                     |                                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                        | HU. No. 35 – Cargar un proyecto existente.                                                                                                                       |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                               | Todo proyecto que haya sido guardado previamente se puede cargar nuevamente.                                                                                     |
| <b>Condiciones de Ejecución</b>                                                                                                                   | <b>Entrada / Pasos de Ejecución</b>                                                                                                                              |
| <ul style="list-style-type: none"> <li>Haberse guardado previamente un proyecto generado por la herramienta <b>POLARIS VERSIÓN 2.0</b></li> </ul> | <ul style="list-style-type: none"> <li>Seleccionar la opción “Cargar proyecto”.</li> <li>Escoger la ruta de origen del proyecto previamente guardado.</li> </ul> |

|                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                    | <ul style="list-style-type: none"> <li>• Seleccionar la opción abrir.</li> </ul>                                                                                                                                                                                                                                                                                                      |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                        | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>• Al seleccionar la opción “Cargar proyecto” se debe desplegar una ventana para buscar el documento.</li> <li>• Se busca el proyecto que se desea cargar.</li> <li>• Al presionar abrir el proyecto debe cargarse en el área de trabajo y listo para seguir trabajando sobre él.</li> </ul> | <ul style="list-style-type: none"> <li>• Se seleccionó del menú la opción “Cargar proyecto”.</li> <li>• Se desplegó la ventana en la cual ubicamos el proyecto en el lugar que se encontraba guardado.</li> <li>• Se seleccionó el proyecto y se tomó la opción abrir.</li> <li>• El proyecto quedó cargado y listo para trabajar en el área de trabajo de la herramienta.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                     | Prueba Aprobada                                                                                                                                                                                                                                                                                                                                                                       |

➤ **Prueba de aceptación – Guardar proyecto**

Tabla 92. Prueba de aceptación – Guardar proyecto

| <b>PRUEBA DE ACEPTACIÓN – GUARDAR PROYECTO</b>                                                                                                                                |                                                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                    | HU. No. 36 – Guardar proyecto.                                                                                                                                                                                                                                                   |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                           | Todo proyecto puede ser guardado en el momento que se desee                                                                                                                                                                                                                      |
| <b>Condiciones de Ejecución</b>                                                                                                                                               | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                              |
| <ul style="list-style-type: none"> <li>• Encontrarse trabajando sobre cualquier proyecto.</li> <li>• El proyecto debe haberse guardado previamente.</li> </ul>                | <ul style="list-style-type: none"> <li>• Seleccionar del menú la opción guardar.</li> </ul>                                                                                                                                                                                      |
| <b>Resultados Esperados</b>                                                                                                                                                   | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                   |
| <ul style="list-style-type: none"> <li>• Al seleccionar la opción guardar el proyecto debe guardarse automáticamente. (el proceso es transparente para el usuario)</li> </ul> | <ul style="list-style-type: none"> <li>• Se seleccionó del menú la opción “Guardar proyecto”</li> <li>• El proyecto quedó guardado de manera satisfactoria.</li> <li>• La única forma de verificar si el proyecto fue guardado es cerrándolo y cargándolo nuevamente.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                | Prueba Aprobada                                                                                                                                                                                                                                                                  |

➤ **Prueba de aceptación – Guardar proyecto por primera vez**

Tabla 93. Prueba de aceptación – Guardar proyecto por primera vez.

| <b>PRUEBA DE ACEPTACIÓN – GUARDAR PROYECTO POR PRIMERA VEZ</b>                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                               |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Historia de Usuario</b>                                                                                                                                                                                                                                                                                                         | HU. No. 37 – Guardar proyecto como                                                                                                                                                                                                                                            |
| <b>Descripción del Caso de Prueba de Aceptación</b>                                                                                                                                                                                                                                                                                | Todo proyecto puede ser almacenado de manera permanente en el disco duro del computador.                                                                                                                                                                                      |
| <b>Condiciones de Ejecución</b>                                                                                                                                                                                                                                                                                                    | <b>Entrada / Pasos de Ejecución</b>                                                                                                                                                                                                                                           |
| <ul style="list-style-type: none"> <li>• Estar en ejecución la herramienta <b>POLARIS VERSIÓN 2.0</b></li> </ul>                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>• Seleccionar del menú la opción “Guardar como”</li> <li>• Establecer la ruta donde se desea guardar el proyecto.</li> <li>• Presionar la opción guardar.</li> </ul>                                                                   |
| <b>Resultados Esperados</b>                                                                                                                                                                                                                                                                                                        | <b>Evaluación de la Prueba</b>                                                                                                                                                                                                                                                |
| <ul style="list-style-type: none"> <li>• Al seleccionar la opción “Cargar proyecto” se debe desplegar una ventana para buscar el documento.</li> <li>• Se busca el proyecto que se desea cargar.</li> <li>• Al presionar abrir el proyecto debe cargarse en el área de trabajo y listo para seguir trabajando sobre él.</li> </ul> | <ul style="list-style-type: none"> <li>• Se selecciono del menú la opción “Guardar como”.</li> <li>• Se estableció la ruta donde se desea almacenar el proyecto.</li> <li>• Se presiona en guardar.</li> <li>• El proyecto queda guardado de manera satisfactoria.</li> </ul> |
| <b>Calificación de Aprobación de la Prueba</b>                                                                                                                                                                                                                                                                                     | Prueba Aprobada                                                                                                                                                                                                                                                               |

## 7. IMPLEMENTACIÓN DE LA HERRAMIENTA POLARIS VERSIÓN 2.0

La implementación de la herramienta software **POLARIS V2.0** se realizó sobre:

- **Sistema Operativo:** *Windows Seven Ultimate de 64 bits*
- **Lenguaje de programación:** Java 6.0.
- **Plataforma de Desarrollo:** NetBeans 6.9 Milestone, como IDE de desarrollo.

### 7.1 ARQUITECTURA

Los módulos de software de la herramienta **POLARIS V2.0** son:

**7.1.1 Módulo de utilidades:** Este módulo es el encargado de dos tareas principales:

- Realizar la conexión a la base de datos y poder acceder a la fuente de datos que repose en el disco duro.
- Contener la colección de clases principales y librerías que son utilizadas por otras clases en la manipulación y visualización de datos, de esta manera puede hacerse la administración de las mismas y se hace factible la reutilización del código.

### 7.1.2 Módulo del Kernel **POLARIS V2.0**

En este módulo se encuentran los módulos de Preprocesamiento y Algoritmos, necesarios para poder realizar la recuperación de la información desde la Web y el análisis de la importancia o peso de los diferentes nodos Web.

➤ **Módulo Crawler:** Contiene las clases necesarias para realizar la recuperación la información de la estructura Web en tiempo real, teniendo una conexión a internet. Además se encarga de aplicar los filtros necesarios en el momento de extraer la información desde la Web, dichos filtros son seleccionados por el usuario.

➤ **Módulo de Algoritmos:** Este módulo contiene las clases necesarias para la aplicación de las técnicas y algoritmos de minería de estructura Web, necesarios para valorar la importancia de las páginas y de la dirección URL que se está analizando.

➤ **Módulo de Links:** Este módulo contiene las clases que permite escoger únicamente el tipo de enlaces con los cuales se quiere trabajar, los tipos de enlaces dentro de la estructura Web son cuatro (Internos entrantes, internos salientes, externos entrantes y externos salientes), con los cuales se pueden

presentar quince gráficos diferentes dependiendo de las diferentes combinaciones que el usuario escoja.

**7.1.3 Módulo de la Interfaz Gráfica:** Este módulo contiene las diferentes formas de visualizar la información recolectada por el Crawler y los cálculos realizados después de haber aplicados los diferentes algoritmos de minería de estructura Web.

➤ **Módulo Graph Views:** Contiene las clases necesarias para construir y desplegar los diferentes tipos de grafos que permiten la visualización estructurada del sitio Web que se encuentra en análisis.

Los tipos de Graph Views son tres: Grafo, en el cual se observa la red como un grafo dirigido; Radial, en el cual se mira el grafo de una manera más estática formado en su centro con la página semilla, desplegándose de esta sus diferentes enlaces como radios hasta el segundo nivel y así sucesivamente y por último tenemos la visualización por Dominio, en el cual el grafo se agrupa dependiendo del host al cual pertenezca cada una de las páginas Web.

➤ **Módulo Views:** Este módulo contiene las clases necesarias para dar a conocer la información del análisis del sitio Web, en el se muestran tablas en las cuales se encuentra asignada la información obtenida por el Crawler, en una manera ordenada y entendible para el usuario, dando además las recomendaciones del caso para mejorar la navegabilidad en la Web que se esté analizando.

Por otra parte este módulo en su opción Tree, permite mirar el grafo de la red como un árbol.

Figura 75. Arquitectura general para **POLARIS**

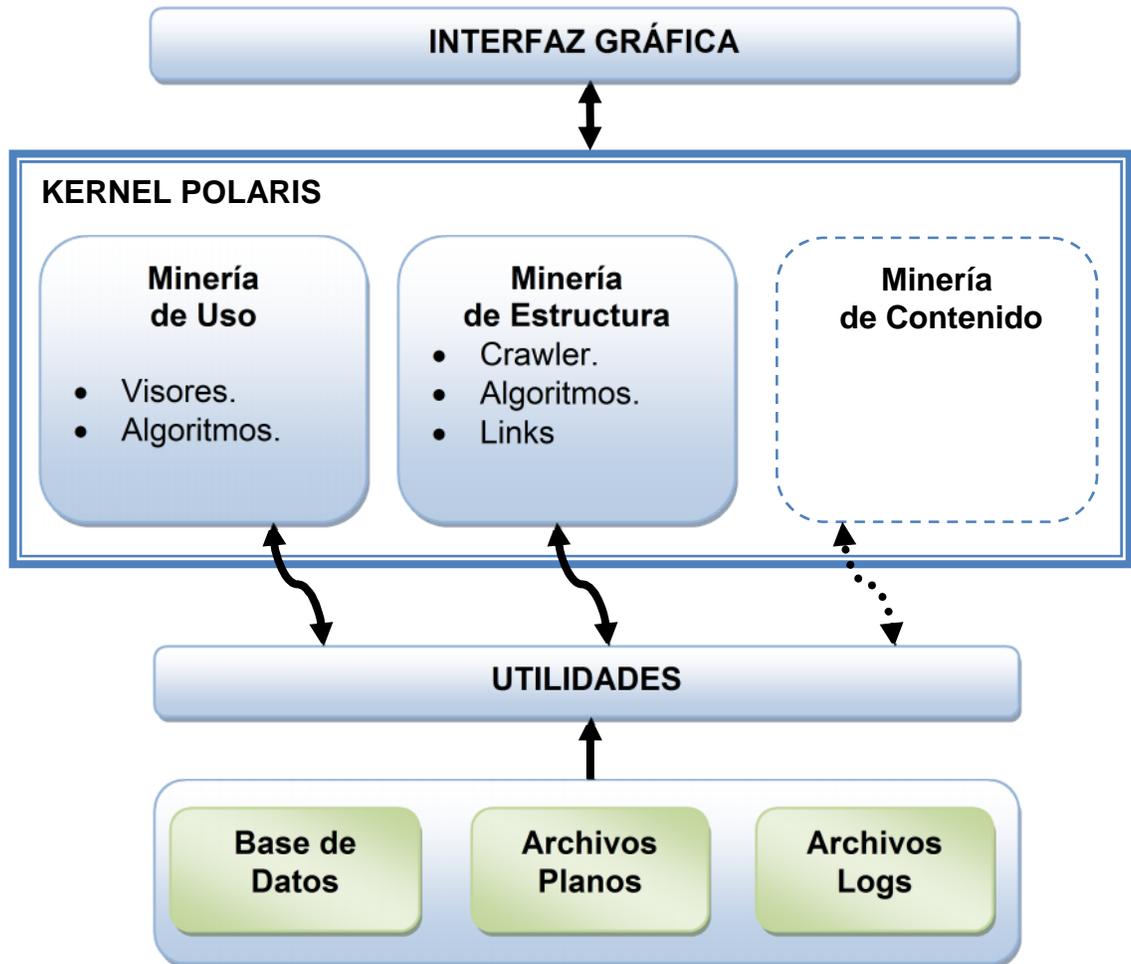
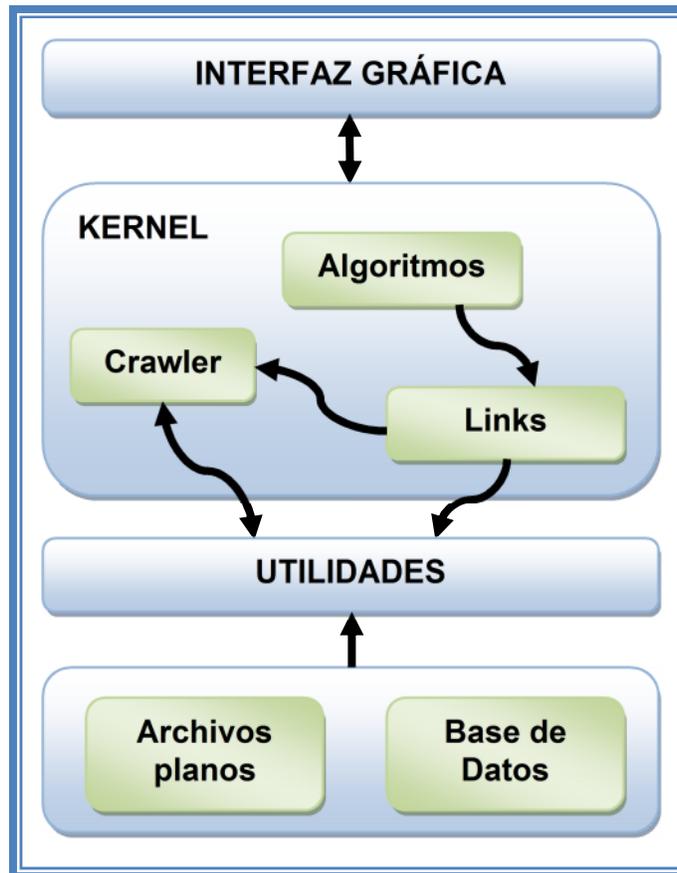


Figura 76. Arquitectura **POLARIS V2.0**



## 7.2 ESTRUCTURA DE PAQUETES DE APLICACIÓN

La estructura general de los paquetes de la herramienta software **Polaris Versión 2.0** para el módulo de minería de estructura Web y el acoplamiento con la versión anterior, es la siguiente:

- **Utils:** De este paquete que pertenece a la versión anterior de Polaris se reutiliza la clase conexión, para permitir el enlace a la base de datos y poder guardar la información recolectado por el Web Crawler.
- **polaris2:** En este paquete se encuentra la clase principal para la visualización de la interfaz gráfica implementado el acople del módulo de minería de uso Web y el de Minería de estructura Web, y otras clases que permiten el buen funcionamiento del acople del producto final. Es el paquete que permite las diferentes conexiones de los nodos en el área de trabajo, como también las restricciones, eliminación, modificación de los nodos.

- **polaris2.dataSource**: Paquete donde se incluyen las clases para la obtención de los datos a ser procesados por la minería de estructura Web, en este paquete se encuentra la clase implementada del Web Crawler, el cual es el mecanismo para la recolección de información para este tipo de minería, también se encuentra implementado una clase para poder obtener la información de archivos de texto plano.
- **polaris2.graph**: Este paquete contiene las clases que se necesitan para unir los diferentes nodos de un gráfico (gráfico es el dibujo que se forma con los diferentes nodos que se pueden arrastrar del árbol de herramientas que está en la parte izquierda de la presentación) y las diferentes conexiones que se realizan entre estos por el usuario.
- **polaris2.graphview**: Representa el paquete donde se encuentran las áreas de trabajo de los diferentes tipos de visualización de grafos implementados en el módulo de minería de estructura Web.
- **polaris2.mining**: En este paquete se encuentran las clases utilizadas para los diferentes algoritmos de minería de estructura Web, como también las clases que representan las ventanas para las configuraciones de los diferentes algoritmos.
- **polaris2.preproces**: Se encuentra las clases relacionadas al preprocesamiento que se le hace a la información recolectada por el Web Crawler, este tipo de preprocesamiento tiene que ver con qué tipo de enlaces se quiere procesar la información si con los internos y/o externos de la página semilla, y con los enlaces entrantes y/o salientes correspondientes a cada uno de ellos, también se encuentra la clase implementada para la configuración de este tipo de preprocesamiento.
- **polaris2.site**: Representa las clases con las cuales se forma el grafo con diferentes tipo listas para la representación del grafo que se forma al reunir la información recolectada y procesada por el Web Crawler.
- **polaris2.view**: Es el paquete en el cual se concentran las clases para la visualización de los datos en formato texto, existen dos tipos uno en formato tabla y el otro en formato árbol.

Cabe destacar en este aspecto que para la implementación y el posterior acople del módulo de herramienta minería de estructura se reutilizaron algunas clases implementadas en la versión anterior de Polaris,

Es por esto que aprovechando las ventajas de la Programación Orientada a Objetos tales como la reusabilidad o la capacidad para que programas que traten las mismas estructuras de información reutilicen las definiciones de objetos y procedimientos empleados en otros programas y de las posibilidades que presenta

un software libre. Sin embargo, debido al tipo de información que es analizada en **Polaris Versión 2.0**, se desarrollaron e implementaron clases propias de la herramienta, y también para la interfaz gráficas se replanteo una nueva clase para poder acoplar el módulo de estructura y mejorar algunos aspectos de diseño como el Drag and Drop para una mejor presentación de la herramienta y su fácil uso para el usuario final.

A continuación se amplia y se detallan conceptos sobre los paquetes más importantes dentro del desarrollo de la herramienta y la manera como fueron implementadas las mismas.

**7.2.1 Paquete Utils:** De este paquete se reutiliza la siguiente clase:

➤ **Clase conexión:** Se encarga de establecer, administrar y cerrar la conexión a una Base de Datos.

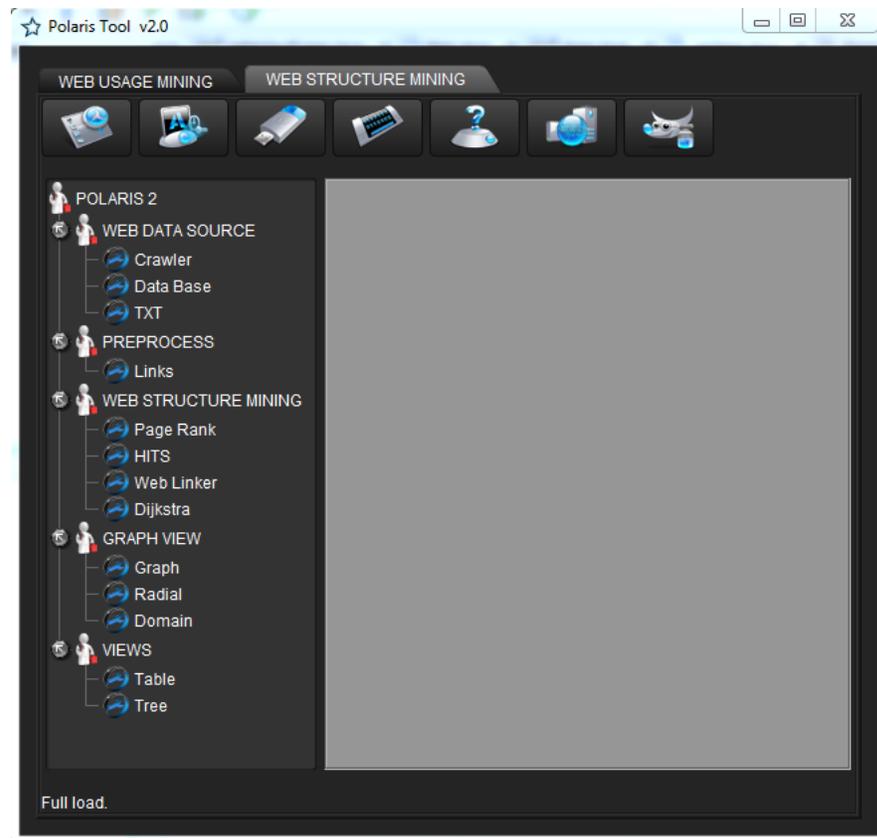
La clase conexión maneja los siguientes métodos:

- conectar(): Establece la conexión con la Base de Datos teniendo en cuenta el nombre de la misma, el usuario y la contraseña.
- cerrarConexion(): Cierra la conexión con la Base de Datos.
- consulta(): Para realizar un query o consulta a la Base de Datos.
- insertar(): Para realizar la inserción de elementos o registros a la Base de Datos.
- eliminar(): Para eliminar un registro de la Base de Datos.
- modificar(): Para modificar los registros de la Base de Datos.

**7.2.2 Paquete polaris2:** La clase que compone este paquete es la siguiente:

➤ **Clase ventana:** Se encarga de presentar la interfaz gráfica de la herramienta Polaris versión 2.0, conteniendo en ella el módulo de minería de uso y estructura Web.

Figura 77. Ventana que representa la interfaz gráfica de la herramienta Polaris versión 2.0.



Los métodos principales manejados por esta clase son:

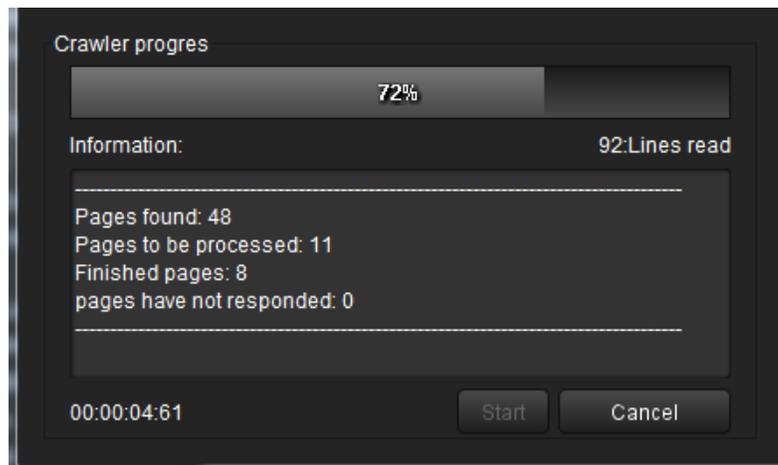
**eventos\_area\_trabajo():** Es el método que permite eliminar, modificar, arrastrar y mover los diferentes objetos con los cuales se trabaja dentro del área de trabajo, permite configurar y determinar las diferentes restricciones para los objetos presentes dentro de una área de trabajo y que cada uno de ellos se pueda conectar al objeto correspondiente. Trabaja con diferentes métodos para lograr este objetivo.

**eventos\_arbol():** Permite reconocer que icono se está seleccionando y arrastrando a la área de trabajo permitiendo el manejo del Drag and Drop en el momento de seleccionar un icono de la árbol de herramientas.

**7.2.3 Polaris2.dataSource:** Las clases que componen este paquete son las siguientes:

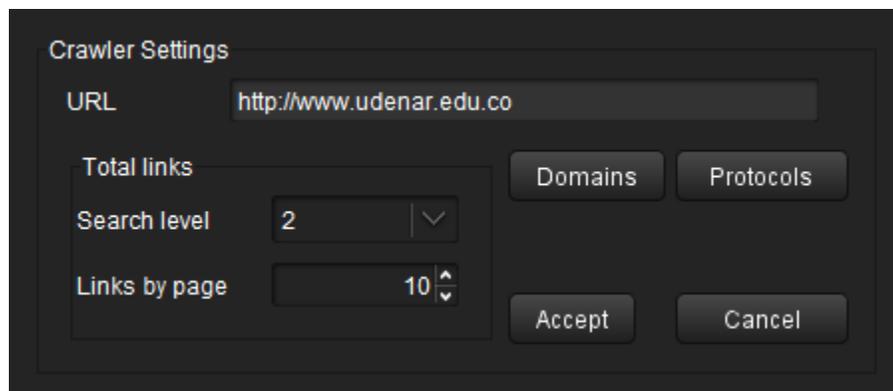
➤ **Clase crawlerRun:** Esta clase permite ejecutar el proceso del Web Crawler. Esta clase presenta una ventana en donde se observa el progreso que tiene el Crawler, al instancias esta clase es la encargada de acceder a internet y realizar las solicitudes a las páginas determinando los enlaces y validando cuales son aceptados y rechazados, estos enlaces automáticamente se los almacena en la variable conf.site del nodo Crawler (Accede a la variable conf luego accede a la variable site, esta variable es de tipo sitio aquí es donde se adicionan las páginas).

Figura 78. Ventana de ejecución para el Web Crawler.



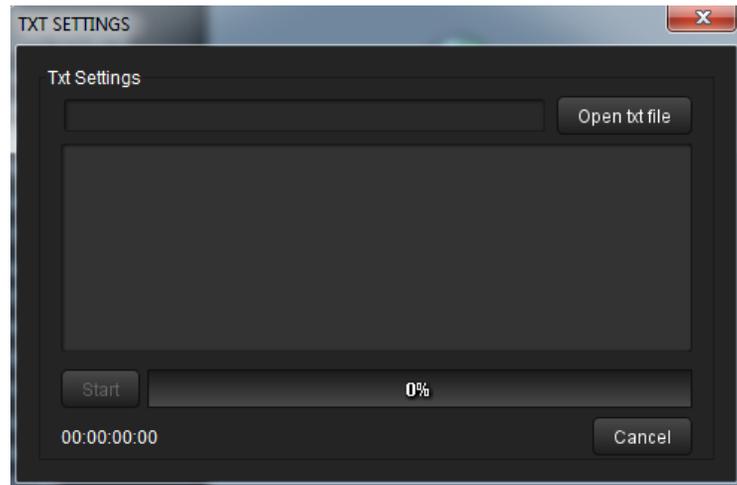
➤ **Clase crawlerSettings:** Es el Ventana de configuración para el Web Crawler, presenta diferentes opciones de filtrado para su ejecución. Es una ventana que permite configurar las características que tiene un Crawler, internamente lo que hace es acceder a la variable conf y de aquí configurar las variables que tengan que ver con el Crawler.

Figura 79. Ventana de configuración para el Web Crawler.



➤ **Clase txtSettings:** Es una ventana que permite seleccionar un archivo que contiene la información de un sitio analizado por el Crawler y muestra el progreso de esta carga.

Figura 80. Ventana que representa la selección de un archivo de texto plano.



**7.2.4 Paquete polaris2.graph:** Las clases que componen este paquete son las siguientes:

➤ **Clase conexion:** Esta clase no tiene métodos la clase conexión se forma de tres apuntadores que indican a que nodo se está entrelazando otro nodo y si tiene más conexiones se van concatenando en los apuntadores (sig\_conexion y ant\_conexion public Nodo destino) nodo con el cual se está conectado (public Conexion sig\_conexion) si hay más conexiones este apuntador nos lleva a esa conexión (public Conexion ant\_conexion) si hay anteriores conexiones este apuntador nos lleva esa conexión.

➤ **Clase nodo:** Esta clase hace referencia a cada uno de los nodos o iconos para la representación de las herramientas que se encuentran definidas en el área de trabajo (contenidos en el árbol a la izquierda de la pantalla), tiene los siguientes variables para su funcionamiento

- id\_nodo: entero que me identifica a este nodo de los demás (identificador).
- int cordenadaX: donde esta dibujado en la pantalla (coordenada x).
- cordenadaY: Donde esta dibujado en la pantalla (coordenada y).

- Tipo: Variable String (cadena) especifica a que tipo pertenece de la siguiente lista,
  - CRAWLER
  - TXT
  - DATA BASE
  - LINKS
  - PAGE RANK
  - HITS
  - WEB LINKER
  - DIJKSTRA
  - GRAPH
  - RADIAL
  - DOMAIN
  - TABLE
  - TREE
  
- sig\_nodo: Es un apuntador a un siguiente nodo (cuando esta dentro de una lista de nodos).
  
- ant\_nodo: Es un apuntador a un nodo anterior (cuando esta dentro de una lista de nodos).
  
- pri\_conexion: Apuntador a una variable tipo conexión que definirá cuando dos nodos esta conectados (cuando dos nodos están conectados aparece una línea blanca en el área de trabajo).
  
- ult\_conexion: Apunta al último nodo al cual se encuentra conectado con un apuntador a una variable de tipo configuración que contiene toda la información que corresponde a este nodo.
  
- Settings: Puede ser true o false y, informa si el nodo está configurado.
  
- Run: Puede ser true o false y, informa si el nodo está funcionando
  
- Connect: Puede ser true o false y, informa si el nodo está conectado con otro nodo.

La clase nodo no tiene métodos, pero depende del tipo de nodo que se cree modifica las variables de configurado y ejecutado de un nodo, por ejemplo:

Si es un nodo de tipo RADIAL este se instancia con las variables, corriendo=true, y configurado=true, porque no tiene configuración y no necesita de un run para

funcionar, pero si es un nodo de tipo Crawler las variables tomaran los siguientes valores corriendo=false y configurado=false, porque se debe configurar y ejecutar (Run) para que se active.

➤ **Clase area\_trabajo:** El área de trabajo es la parte gris de la ventana principal, esta clase es la que administra todos los nodos que se arrastran desde el árbol hasta el área de trabajo donde se ubican y conectan las diferentes partes del sistema.

Se encarga de eliminar, modificar, mover y presentar los nodos en el área donde el usuario se encuentra trabajando.

Contiene variable de tipo nodo que apunta al primer y último nodo que se encuentra en el área de trabajo y tiene funciones que permiten en el área de trabajo.

- public void parar\_nodo(Nodo n): En el área de trabajo cuando un nodo es reconfigurado o eliminado los nodos que estén conectados a este deben volverse a ejecutar, esta función se llama para reconocer los nodos conectados a este nodo.
- public void paint(Graphics g): Función realiza el dibujo de los nodos y conexiones del área de trabajo.
- public void eliminar\_nodo(Nodo nod): Elimina un nodo del área de trabajo.
- public void insertar\_nodo(String tipo,int pos\_x,int pos\_y): Inserta un nodo en le área de trabajo, se le envían como parámetros las coordenadas en donde será dibujado.
- public void conectar(Nodo fuente, Nodo destino): Realiza una conexión entre dos nodos (es llamada por insertar\_conexion, que es el método encargado de validarlo).
- public int revisar(Conexion c, String t): Revisa si en las conexiones de un nodo existe un nodo tipo "t" configurado y arrancado.
- public int insertar\_conexion(Nodo fuente, Nodo destino): Esta función es la que realiza la conexión entre dos nodos, primero verifica de que tipo son los nodos que se van a conectar, luego valida si la conexión entre estos nodos del área de trabajo es permitida, si es permitida los conecta sino devuelve un mensaje indicando por qué no se pudieron conectar.

➤ **Clase configuración:** La clase nodo instancia en un objeto una clase de tipo

configuración, esta es un apuntador a todas las características que puede contener un nodo, dependiendo del tipo que sea.

por ejemplo si el nodo es de tipo PAGE RANK en la variable conf del nodo se configuraran las variables de `page_rank` y la variable `iterationsPageRank` que son las utilizadas en el algoritmo de page Rank las demás no se toman en cuenta.

A continuación se describen las variables que se han definido en esta clase:

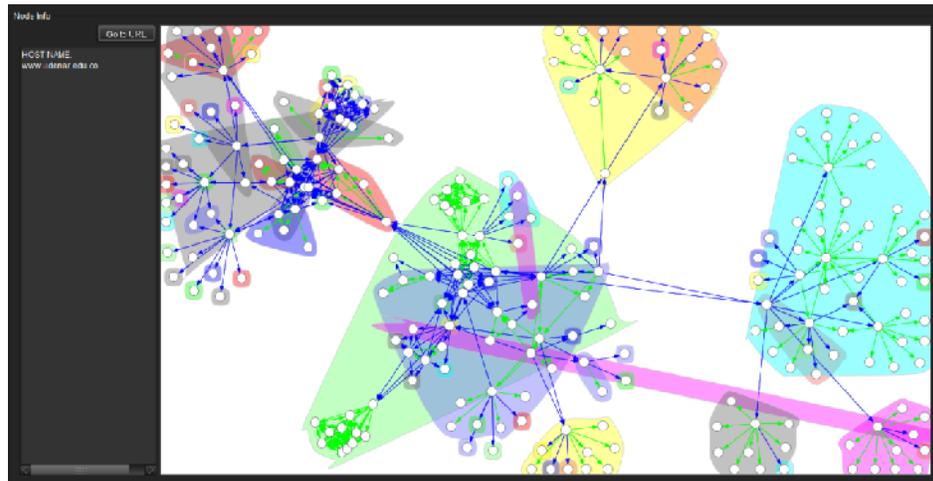
- boolean `InternalInLinks`: Si el nodo es de tipo LINKS, informa si este nodo debe tomar links internos entrantes.
- boolean `InternalOutLinks`: Si el nodo es de tipo LINKS, informa si este nodo debe tomar links internos salientes.
- boolean `ExternalInLinks`: Si el nodo es de tipo LINKS, informa si este nodo debe tomar links externos entrantes.
- boolean `ExternalOutLinks`: Si el nodo es de tipo LINKS, informa si este nodo debe tomar links externos salientes.
- int `totalDataProcces`: cuando se aplica un algoritmo (HITS, Page Rank, WebLinker), informa cuando los datos fueron procesados.
- String `url`: Si el nodo es de tipo CRAWLER, informa la cual es la Url inicial, digitada por el usuario para realizar el estudio de la página, por lo tanto esta se convierte en una página semilla.
- int `search_level`: Si el nodo es de tipo CRAWLER, informa el nivel de búsqueda.
- int `links_by_page`: Si el nodo es de tipo CRAWLER, informa cuantos links por página.
- String `other_potoclos`: Si el nodo es de tipo CRAWLER, informa que otros protocolos debo filtrar.
- boolean `http`: Si el nodo es de tipo CRAWLER, informa si debo aceptar protocolo.
- boolean `https`: Si el nodo es de tipo CRAWLER, informa si debo aceptar protocolo.
- boolean `ftp`: Si el nodo es de tipo CRAWLER, informa si debo aceptar protocolo.

- String other\_domain: Si el nodo es de tipo CRAWLER, informa que otros dominios debo filtrar.
- boolean com: Si el nodo es de tipo CRAWLER que este dominio lo debo aceptar.
- boolean net: Si el nodo es de tipo CRAWLER, notifica que este dominio lo debo aceptar
- boolean gov: Si el nodo es de tipo CRAWLER, notifica que este dominio lo debo aceptar
- boolean edu: Si el nodo es de tipo CRAWLER, notifica que este dominio lo debo aceptar.
- boolean org: Si el nodo es de tipo CRAWLER, notifica que este dominio lo debo aceptar.
- double dampening: Si el nodo es de tipo HITS, indica el valor de dampening.
- int iterationsPageRank: Si el nodo es de tipo Page Rank, indica cuantas iteraciones se deben realizar.
- int iterationsHits: Si el nodo es de tipo HITS indica cuantas iteraciones realizar.
- sitio site: esta variable es un apuntador a un objeto instanciado de la clase **sitio**.
- Page Rank pague\_rank: Contiene la información del Page Rank calculado.
- HITS hits: Contiene la información del HITS calculado.
- String ruta\_txt si el nodo es tipo txt, informa la ruta del archivo.
- String info\_txt: si el nodo es de tipo txt, informa que texto está contenido en el archivo.

**7.2.5 Paquete polaris2.graphview:** Las clases que componen este paquete son las siguientes:

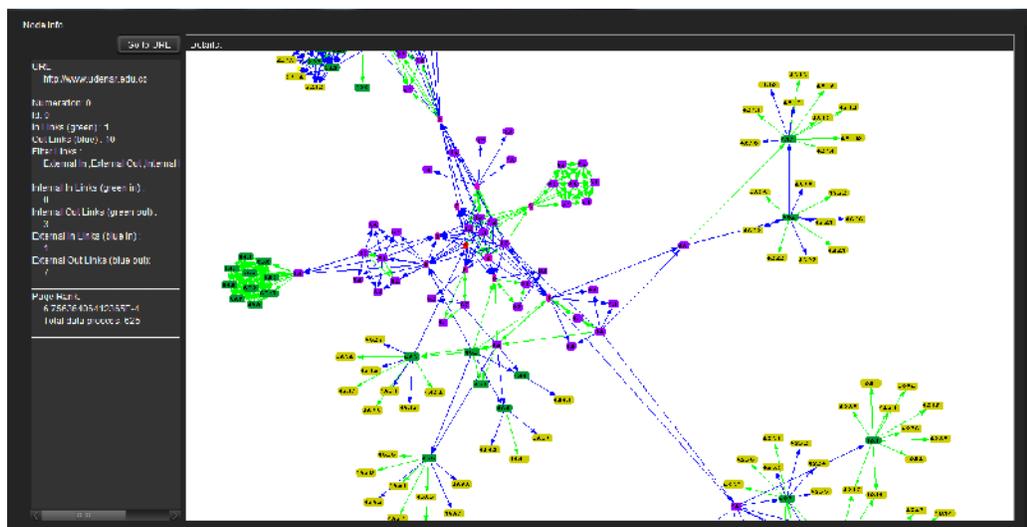
➤ **Clase domainViewFrame:** Ventana donde se presenta un gráfico de tipo Domain haciendo uso de la herramienta Prefuse, se instancia un objeto de la clase Domain para poder utilizarlo dentro del Ventana.

Figura 81. Ventana para la representación del gráfico Domain.



➤ **Clase graphViewFrame:** Ventana donde se presenta un gráfico GraphView. haciendo uso de la herramienta Prefuse, se instancia un objeto de la clase GraphView. para poder utilizarlo dentro del Ventana.

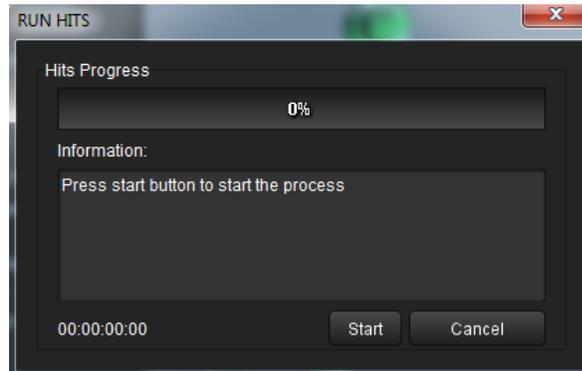
Figura 82. Ventana para la representación del gráfico GraphView.



➤ **Clase radialViewFrame:** Ventana donde se presenta un gráfico Radial. haciendo uso de la herramienta Prefuse, se instancia un objeto de la clase Radial. para poder utilizarlo dentro de la Ventana.

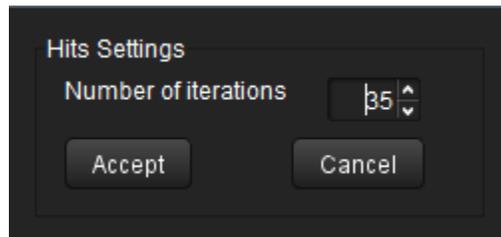


Figura 84. Ventana de ejecución algoritmo HITS.



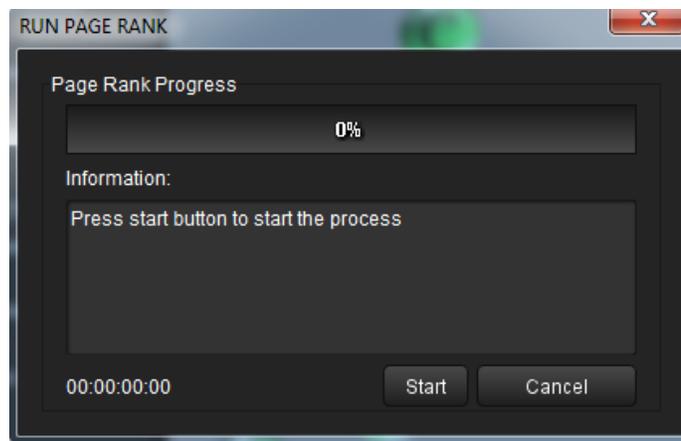
➤ **Clase hitsSettings:** Esta clase presenta el Ventana para la configuración del algoritmo HITS, solo recibe como parámetro el número de iteraciones a realizar por el usuario.

Figura 85. Ventana configuración del algoritmo HITS



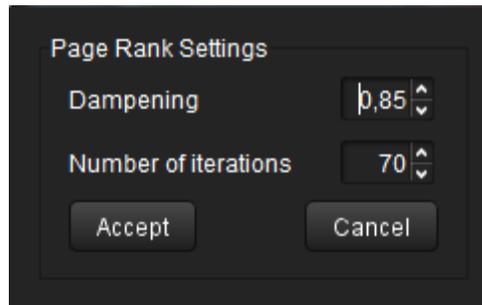
➤ **Clase pageRankRun:** Permite la ejecución del algoritmo Page Rank.

Figura 86. Ventana de ejecución algoritmo Page Rank.



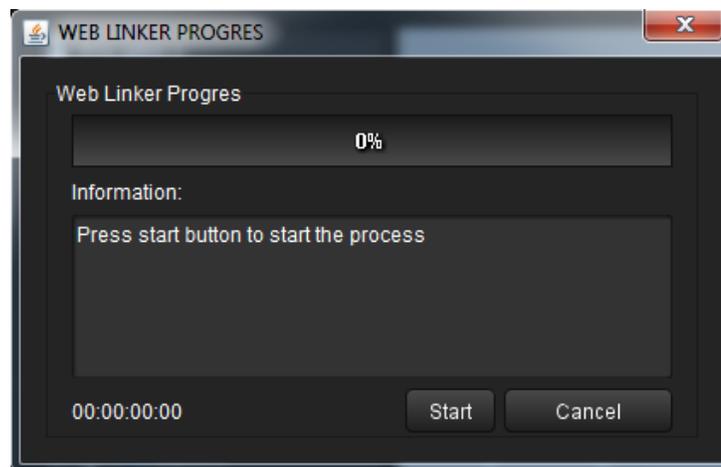
➤ **Clase pageRankSettings:** Esta clase presenta el Ventana para la configuración del algoritmo Page Rank, recibe dos parámetros, uno de ellos es el número de iteraciones a realizar y el otro es el valor Dampening o factor de amortiguamiento que toma un valor de 0 a 1 y por defecto este valor es 0.85.

Figura 87. Ventana de configuración algoritmo Page Rank.



➤ **Clase visitRun:** Permite la ejecución del algoritmo Web Linker.

Figura 88. Ventana de ejecución algoritmo Web Linker.



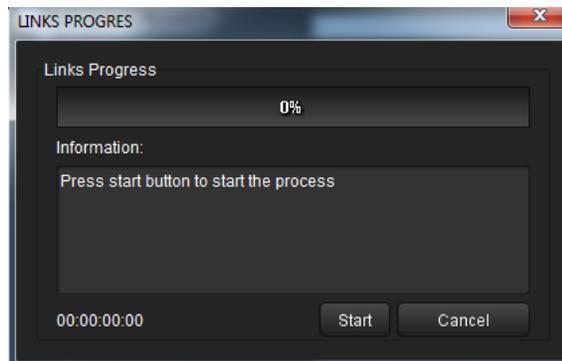
➤ **Clase Dijkstra:** Es la clase encargada de realizar el algoritmo Dijkstra entre dos nodos seleccionados, este genera un gráfico en el cual representa la ruta más corta entre ellos teniendo en cuenta el cálculo de los algoritmos de minería de estructura Web, los enlaces que formen esta ruta más corta se muestran en otro color, permitiendo al usuario la toma de decisiones entre los nodos seleccionado.

**7.2.7 Polaris2.preprocess:** Las clases que componen este paquete son las siguientes:

➤ **Clase linksRun:** Es una ventana la cual se encarga de generar un nuevo sitio un objeto que se encuentre en el árbol de herramientas denominado WEB DATA SOURCE y se le aplica filtros seleccionados por el usuario a través de la clase **linksSettings**.

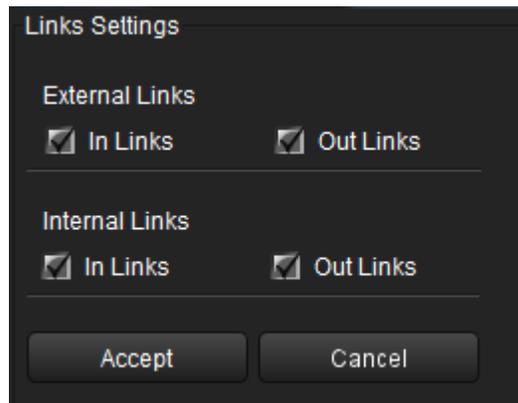
Ejemplo se tiene un Crawler con 100 páginas (el Crawler contiene una variable de tipo sitio que contiene todas las páginas que pertenecen a este Crawler) y se tiene un nodo de tipo Links conectado a el Crawler cuando se ejecuta linksRun esta clase lo que hace es crear un sitio dentro del nodo Links y este sitio tendrá menos páginas si no se han activado todas las casillas del nodo Links, Permite la ejecución del nodo enlaces (Links).

Figura 89. Ventana de ejecución del nodo enlaces (Links).



➤ **Clase linksSettings:** Esta clase representa la configuración del nodo enlaces (Links), en la cual se puede preprocesar la información recolectada por el Web Crawler en diferentes formas, enlaces internos entrantes y/o salientes, enlaces externos entrantes y/o salientes, esta clase me permite por medio de una ventana configurar el un nodo de tipo Links, al dar clic en aceptar se modifica la variable conf de este nodo y se cambian los valores de InternallnLinks, InternalOutLinks, ExternallnLinks, ExternalOutLinks;

Figura 90. Ventana que representa la configuración del nodo enlaces (Links).



**7.2.8 Paquete polaris2.site:** Las clases que componen este paquete son las siguientes:

➤ **Clase enlace:** Se encarga de especificar la página de origen y de destino de cada uno de los enlaces.

Enlace no tiene métodos, esta clase solo contiene la dirección de una página destino. Esta clase esta instanciada dentro de la clase página para saber a cual está conectada.

➤ **Clase host:** Esta clase se utiliza cuando se está generando el gráfico Domain para agrupar las páginas según su host.

Esta clase no tiene métodos solo tres variables:

- public String nom\_host: Hace referencia al nombre del host (servidor que contiene la página, ejemplos de hosts son www.google.com, www.facebook.com y www.udenar.edu.co).
- public int id\_host: Se refiere al número que identifica al host en una lista.
- public host sig\_host: Apuntador a un siguiente host (usada para concatenar los host en una lista), es null si no hay siguiente host en la lista.
- public host ant\_host: Apuntador a un anterior host (usada para concatenar los host en una lista), es null si no hay anterior host en la lista.

➤ **Clase página:** Manejar toda la información perteneciente a una determinada página (Numeración, Nivel, URL, Nombre e Id de página).

`public void insertar_enlace(página des, String tipo):` Sirve para adicionar un enlace (conexión a otra página), por lo tanto se debe enviar como parámetro una variable de tipo página que será el destino (a la cual se va a conectar), y debo enviar una variable de tipo String con cualquiera de los siguientes valores: `entraI`, `entraE`, `saleI`, `saleE`, haciendo referencia al tipo de conexión que estas dos páginas tienen.

- `sitio:` Administrar una lista de páginas o de objetos de tipo página en una lista doblemente enlazada, con la finalidad de crear todos los nodos nuevos de la Web.
- `public void crear_lista_host():` Este método crea un listado de todos los host diferentes que contiene un grafo. Este método se usa cuando se crea un grafo de tipo Domain para saber cuantos host hay y poderlos agrupar en un solo grupo.
- `public void insertar_host(String n):` La clase sitio contiene una lista concatenada de los host existentes esta función se utiliza para concatenar otro host a esta lista, la clase sitio tiene una variable `pri_host` que apunta al primer host a partir de este se siguen concatenando los demás.
- `public int buscar_host(String n):` Sirve para buscar un host en la lista de hosts del sitio, cuando esta función es llamada devuelve un número que indica el id de un host que se está buscando, si es -1 es que no existe el host especificado.
- `public void cargar_página(String info_nodo):` Esta función recibe como parámetro una variable de tipo string que contiene la información de una página a ser adicionada al sitio, este método es utilizado cuando se está leyendo un archivo txt en donde se lee una línea del archivo y se llama a esta función para que ingrese una nueva página al sitio partiendo de esa línea sacada del archivo de texto.
- `public void insertar_página(página nueva_página,int nivel):` Inserta una nueva página en el sitio, el sitio tiene una variable `página_inicial` que es un apuntador a la primer página que tiene el sitio, a partir de esta se siguen concatenando las demás páginas que tenga el sitio creando una lista de páginas.
- `public void borrar_desconectados():` Se llama a esta función para que elimine páginas que no tienen ninguna conexión, estas páginas sin ninguna conexión aparecen si una página no responde a una petición de acceso, esto se realiza al ejecutar el Crawler.
- `public página buscar_página(String url_pag):` Es la función se llama para buscar la página en el sitio y así conocer la Url con la cual se está trabajando, por lo tanto

retorna null si la página no existe y una variable de tipo páginas encuentra el valor.

**7.2.9 Paquete polaris2.view:** Las clases que componen este paquete son las siguientes:

➤ **Clase tableSetings:** Ventana que muestra la configuración de los datos que se desean ser mostrados en la tabla, esto se hace a través de check box (por ejemplo los check box se usarían para indicar que se muestren totales, id, iteración, etc.. todos los datos que contenga la tabla)

➤ **tableRunFrame:** Ventana que me presenta en pantalla la tabla y me permite su exportación a html txt o pdf. Esta clase permite la visualización en formato tabla (table), de la información del nodo perteneciente al árbol de herramientas al que se conecte, permite observar los registros procesados teniendo en cuenta si se ha aplicado los algoritmos o solo se quiere ver la información recolectada por el Crawler.

➤ **treeRunFrame:** Ventana que me muestra en pantalla un árbol que representa las páginas y conexiones del sitio. El tree es otra forma de visualizar un grafo a través de formato texto, esta representación la hace por medio de un árbol sino en el cual se inicia con la Url inicial y cuando se abre se mira las conexiones que tenga, estas conexiones representan las hojas del árbol que se va expandiendo.

## **7.3 LENGUAJES Y HERRAMIENTAS PARA EL DESARROLLO DE POLARIS VERSIÓN 2.0**

Para llevar a cabo el desarrollo del proyecto **POLARIS VERSION 2.0**, fue necesario hacer uso de herramientas que permitan la construcción de este, a continuación se relacionan los diferentes medios que se utilizaron en la consecución de **POLARIS**.

**7.3.1 Lenguaje de Programación Java:** Java [43] es un lenguaje de programación orientado a objetos desarrollado por James Gosling y sus compañeros de Sun Microsystems al inicio de la década de 1990. A diferencia de los lenguajes de programación convencionales, que generalmente están diseñados para ser compilados a código nativo, Java es compilado en un bytecode que es ejecutado (usando normalmente un compilador JIT), por una maquina virtual Java. El lenguaje Java se crea con cinco objetivos principales:

1. Deberá usar la metodología de la programación orientada a objetos.
2. Deberá permitir la ejecución de un mismo programa en múltiples sistemas operativos.

3. Deberá incluir por defecto soporte para trabajo en red.
4. Deberá diseñarse para ejecutar código en sistemas remotos de forma segura.
5. Deberá ser fácil de usar y tomar lo mejor de otros lenguajes orientados a objetos, como C++.

#### ➤ **Características Principales**

**Orientado a Objetos:** La primera característica, orientado a objetos ("OO"), se refiere a un método de programación y al diseño del lenguaje. Aunque hay muchas interpretaciones para OO, una primera idea es diseñar el software de forma que los distintos tipos de datos que use están unidos a sus operaciones. Así los datos y el código (funciones o métodos) se combinan en entidades llamadas objetos. Un objeto puede verse como un paquete que contiene el "comportamiento" (el código) y el "estado" (datos).

El principio es separar aquello que cambia de las cosas que permanecen inalterables. Frecuentemente, cambiar una estructura de datos implica un cambio en el código que opera sobre los mismos, o viceversa. Esta separación en objetos coherentes e independientes ofrece una base más estable para el diseño de un sistema software. El objetivo es hacer que grandes proyectos sean fáciles de gestionar y manejar, mejorando como consecuencia su calidad y reduciendo el número de proyectos fallidos. Otra de las grandes promesas de la programación orientada a objetos es la creación de entidades más Genéricas (objetos) que permitan la reutilización del software entre proyectos, una de las premisas fundamentales de la Ingeniería del Software. Un objeto genérico "cliente", por ejemplo, Deberá en teoría tener el mismo conjunto de comportamiento en diferentes proyectos, sobre todo cuando estos coinciden en cierta medida, algo que suele suceder en las grandes organizaciones. En este sentido, los objetos Podría verse como piezas reutilizables que pueden emplearse en múltiples proyectos distintos, posibilitando así a la industria del software a construir proyectos de envergadura empleando componentes ya existentes y de comprobada calidad; conduciendo esto finalmente a una reducción drástica del tiempo de desarrollo. Podemos usar como ejemplo de objeto el aluminio. Una vez definidos datos (peso, maleabilidad, etc.), y su "comportamiento" (soldar dos piezas, etc.), el objeto "aluminio" puede ser reutilizado en el campo de la construcción, del automóvil, de la aviación, etc.

**Independencia de la plataforma:** La segunda característica, la independencia de la plataforma, significa que programas escritos en el lenguaje Java pueden ejecutarse igualmente en cualquier tipo de hardware. Es lo que significa ser capaz de escribir un programa una vez y que pueda ejecutarse en cualquier dispositivo, tal como reza el axioma de Java, "write once, run everywhere". Para ello, se

compila el código fuente escrito en lenguaje Java, para generar un código conocido como “bytecode” (específicamente Java bytecode) que son instrucciones maquina simplificadas específicas de la plataforma Java. Esta pieza esta “a medio camino” entre el código fuente y el código maquina que entiende el dispositivo destino. El bytecode es ejecutado entonces en la maquina virtual (VM), un programa escrito en código nativo de la plataforma destino (que es el que entiende su hardware), que interpreta y ejecuta el código. Además, se suministran bibliotecas adicionales para acceder a las características de cada dispositivo (como los gráficos, ejecución mediante hebras o threads, la interfaz de red) de forma unificada. Se debe tener presente que, aunque hay una etapa explicita de compilación, el bytecode generado es interpretado o convertido a instrucciones maquina del código nativo por el compilador JIT (Just In Time).

**El recolector de basura:** Un argumento en contra de lenguajes como C++ es que los programadores se encuentran con la carga añadida de tener que administrar la memoria de forma manual. En C++, el desarrollador debe asignar memoria en una zona conocida como heap (monticulo) para crear cualquier objeto, y posteriormente desalojar el espacio asignado cuando desea borrarlo. Un olvido a la hora de desalojar memoria previamente solicitada, o si no lo hace en el instante oportuno, puede llevar a una fuga de memoria, ya que el sistema operativo piensa que esa zona de memoria está siendo usada por una aplicación cuando en realidad no es así un programa mal diseñado Podría consumir una cantidad desproporcionada de memoria. Además, si una misma Región de memoria es desalojada dos veces el programa puede volverse inestable y llevar a un eventual cuelgue.

En Java, este problema potencial es evitado en gran medida por el recolector automático de basura (o automatic garbage collector). El programador determina cuando se crean los objetos y el entorno en tiempo de ejecución de Java (Java runtime) es el responsable de gestionar el ciclo de vida de los objetos. El programa, u otros objetos pueden tener localizado un objeto mediante una referencia a este (que, desde un punto de vista de bajo nivel es una dirección de memoria). Cuando no quedan referencias a un objeto, el recolector de basura de Java borra el objeto, liberando así la memoria que ocupaba previniendo posibles fugas (ejemplo: un objeto creado y únicamente usado dentro de un método solo tiene entidad dentro de este; al salir del método el objeto es eliminado), aun así es posible que se produzcan fugas de memoria si el código almacena referencias a objetos que ya no son necesarios es decir, pueden aun ocurrir, pero en un nivel conceptual superior. En definitiva, el recolector de basura de Java permite una fácil creación y eliminación de objetos, mayor seguridad y frecuentemente más rápida que en C++.

La recolección de basura de Java es un proceso prácticamente invisible al desarrollador. Es decir, el programador no tiene conciencia de cuando la recolección de basura tendrá lugar, ya que esta no tiene necesariamente que

guardar relación con las acciones que realiza el código fuente. Debe tenerse en cuenta que la memoria es solo uno de los muchos recursos que deben ser gestionados.

**7.3.2 Entorno de Desarrollo Netbeans:** El entorno en el que se desarrollo POLARIS Versión 2.0 fue Neatbeans 6.9 Milestone. NetBeans es un proyecto exitoso de código abierto con una gran base de usuarios, una comunidad en constante crecimiento. Sun Microsystems fundó el proyecto de código abierto NetBeans en junio 2000 y continúa siendo el patrocinador principal de los proyectos.

Al día de hoy hay disponibles dos productos: el NetBeans IDE y NetBeans Platform.

NetBeans IDE es un entorno de desarrollo integrado, una herramienta para que los programadores puedan escribir, compilar, depurar y ejecutar programas. Está escrito en Java, pero puede servir para cualquier otro lenguaje de programación. Existe además un número importante de módulos para extender el NetBeans IDE. NetBeans IDE es un producto libre y gratuito sin restricciones de uso.

También está disponible NetBeans Platform; una base modular y extensible usada como estructura de integración para crear grandes aplicaciones de escritorio. Empresas independientes asociadas, especializadas en desarrollo de software, proporcionan extensiones adicionales que se integran fácilmente en la plataforma y que pueden también utilizarse para desarrollar sus propias herramientas y soluciones.

La plataforma ofrece servicios comunes a las aplicaciones de escritorio, permitiéndole al desarrollador enfocarse en la lógica específica de su aplicación. Entre las características de la plataforma están:

- Administración de las interfaces de usuario (ej. menús y barras de herramientas).
- Administración de las configuraciones del usuario.
- Administración del almacenamiento (guardando y cargando cualquier tipo de dato)
- Administración de ventanas.
- Framework basado en asistentes (diálogos paso a paso).

Ambos productos son de código abierto y gratuitos para uso tanto comercial como no comercial. El código fuente está disponible para su reutilización de acuerdo con la Common Development and Distribution License ( CDDL) v1.0 and the GNU General Public License (GPL) v2.

**7.3.3 PostgreSQL:** PostgreSQL es un sistema de gestión de base de datos relacional orientada a objetos y libre, publicado bajo la licencia BSD.

Como muchos otros proyectos de código abierto, el desarrollo de PostgreSQL no es manejado por una sola empresa sino que es dirigido por una comunidad de desarrolladores y organizaciones comerciales las cuales trabajan en su desarrollo. Dicha comunidad es denominada el PGDG (*PostgreSQL Global Development Group*).

➤ **Características:** Algunas de sus principales características son, entre otras:

**Alta concurrencia:** Mediante un sistema denominado MVCC (Acceso concurrente multiversión, por sus siglas en inglés) PostgreSQL permite que mientras un proceso escribe en una tabla, otros accedan a la misma tabla sin necesidad de bloqueos. Cada usuario obtiene una visión consistente de lo último a lo que se le hizo *commit*. Esta estrategia es superior al uso de bloqueos por tabla o por filas común en otras bases, eliminando la necesidad del uso de bloqueos explícitos.

#### **Amplia variedad de tipos nativos**

- PostgreSQL provee nativamente soporte para:
- Números de precisión arbitraria.
- Texto de largo ilimitado.
- Figuras geométricas (con una variedad de funciones asociadas)
- Direcciones IP (IPv4 e IPv6).
- Bloques de direcciones estilo CIDR.
- Direcciones MAC.
- Arrays.

Adicionalmente los usuarios pueden crear sus propios tipos de datos, los que pueden ser por completo indexables gracias a la infraestructura GiST de PostgreSQL. Algunos ejemplos son los tipos de datos GIS creados por el proyecto PostGIS.

**Foreign Keys:** Claves ajenas también denominadas Llaves ajenas o Claves Foráneas

**Disparadores (*triggers*):** Un disparador o *trigger* se define en una acción específica basada en algo ocurrente dentro de la base de datos. En PostgreSQL esto significa la ejecución de un procedimiento almacenado basado en una

determinada acción sobre una tabla específica. Ahora todos los disparadores se definen por seis características:

- El nombre del disparador o *trigger*
- El momento en que el disparador debe arrancar
- El evento del disparador deberá activarse sobre...
- La tabla donde el disparador se activará
- La frecuencia de la ejecución
- La función que podría ser llamada

Entonces combinando estas seis características, PostgreSQL le permitirá crear una amplia funcionalidad a través de su sistema de activación de disparadores (*triggers*).

**Soporte para transacciones distribuidas:** Permite a PostgreSQL integrarse en un sistema distribuido formado por varios recursos (ejemplo una base de datos PostgreSQL, otra Oracle, una cola de mensajes IBM MQ JMS y un ERP SAP) gestionado por un servidor de aplicaciones donde el éxito ("commit") de la transacción global es el resultado del éxito de las transacciones locales.

**Funciones:** Bloques de código que se ejecutan en el servidor. Pueden ser escritos en varios lenguajes, con la potencia que cada uno de ellos da, desde las operaciones básicas de programación, tales como bifurcaciones y bucles, hasta las complejidades de la programación orientada a objetos o la programación funcional.

Los disparadores (*triggers* en inglés) son funciones enlazadas a operaciones sobre los datos.

PostgreSQL soporta funciones que retornan "filas", donde la salida puede tratarse como un conjunto de valores que pueden ser tratados igual a una fila retornada por una consulta (query en inglés).

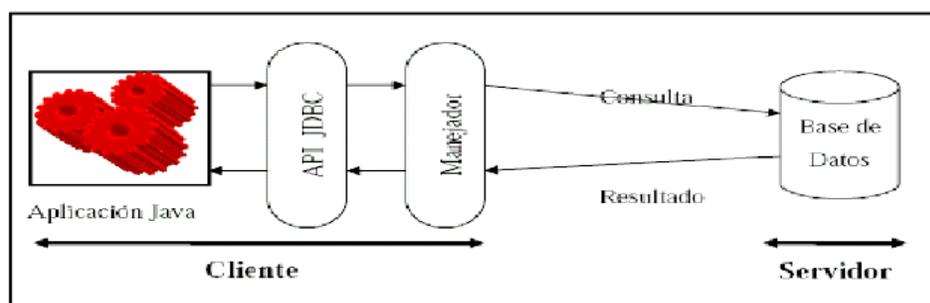
Las funciones pueden ser definidas para ejecutarse con los derechos del usuario ejecutor o con los derechos de un usuario previamente definido. El concepto de funciones, en otros DBMS, son muchas veces referidas como "procedimientos almacenados" (stored procedures en inglés).

**7.3.4 Controlador JDBC:** JDBC [43] es el acrónimo de Java Database Connectivity, un API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java independientemente del sistema de operación donde se ejecute o de la base de datos a la cual se accede utilizando el dialecto SQL del modelo de base de datos que se utilice.

El API JDBC se presenta como una colección de interfaces Java y métodos de gestión de manejadores de conexión hacia cada modelo específico de base de datos. Un manejador de conexiones hacia un modelo de base de datos en particular es un conjunto de clases que implementan las interfaces Java y que utilizan los métodos de registro para declarar los tipos de localizadores a base de datos (URL) que pueden manejar. Para utilizar una base de datos particular, el usuario ejecuta su programa junto con la biblioteca de conexión apropiada al modelo de su base de datos, y accede a ella estableciendo una conexión, para ello provee un localizador a la base de datos y los parámetros de conexión específicos. A partir de allí puede realizar cualquier tipo de tareas con la base de datos a las que tenga permiso: consultas, actualizaciones, creado modificado y borrado de tablas, ejecución de procedimientos almacenados en la base de datos, etc. Cada base de datos emplea un protocolo diferente de comunicación, protocolos que normalmente son propietarios. El uso de un manejador, una capa intermedia entre el código del desarrollador y la base de datos, permite independizar el código Java que accede a la BD del sistema de BD concreto a la que estamos accediendo, ya que en nuestro código Java emplearemos comandos estándar, y estos comandos serían traducidos por el manejador a comandos propietarios de cada sistema de BD concreto. Si queremos cambiar el sistema de BD que empleamos lo único que deberemos hacer es reemplazar el antiguo manejador por el nuevo, y seremos capaces de conectarnos a la nueva BD.

**Manejador de Protocolo Nativo (tipo 4):** El manejador traduce directamente las llamadas al API JDBC al protocolo nativo de la base de datos. Es el manejador que tiene mejor rendimiento, pero está más ligado a la base de datos que empleamos que el manejador tipo JDBCNet, donde el uso del servidor intermedio nos da una gran flexibilidad a la hora de cambiar de base de datos. Este tipo de manejadores también emplea Tecnología 100 % Java.

Figura 91. Controlador JDBC tipo 4



**Fuente:** Narvaez Teran, H R. RASEMUS: UNA HERRAMIENTA PARA EL DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS CON TÉCNICAS DE CLUSTERING, DÉBILMENTE ACOPLADO CON EL SGBD POSTGRESQL. San Juan de Pasto, Colombia, 2009.

**7.3.5 WebLA:** WebLA [37] es un paquete Java para manipulación de grafos en la Web, la aplicación de algoritmos difundidos, tales como Page Rank, HITS, CoCitation Similarity y SimRank. Es de particular interés para la investigación en la recuperación de información, ya que proporciona un conjunto de APIs (Application Programming Interfaces) que permiten experimentar con este tipo de algoritmos.

Se implementan métodos eficientes para el manejo en memoria de grafos Web relativamente grandes, y el paquete también incluye la utilidad de muchos métodos, como la detección de enlaces no-nepotista, la extracción de información de enlaces de una base de datos relacional, etc.

El software trabaja en un motor de búsqueda Web portugués denominado tumba!, donde se utiliza como base del algoritmo de páginas relacionadas, así como en otras herramientas de minería Web.

WebLA fue desarrollado por el grupo XLDB del Departamento de Informática de la Facultad de Ciencias de la Universidad de Lisboa en Portugal.

WebLA fue escrito por Bruno Martins .

WebLA está publicado bajo la licencia BSD es la licencia de software otorgada principalmente para los sistemas BSD (*Berkeley Software Distribution*). Es una licencia de software libre permisiva como la licencia de OpenSSL o la MIT License. Esta licencia tiene menos restricciones en comparación con otras como la GPL estando muy cercana al dominio público. La licencia BSD al contrario que la GPL permite el uso del código fuente en software no libre.

**7.3.6 Herramienta de Visualización Prefuse:** Prefuse [49] es un conjunto de herramientas de software para visualización de datos interactivos. El conjunto de herramientas **Prefuse** original proporciona un marco de visualización para el lenguaje de programación Java. Prefuse proporciona herramientas de visualización y animación para ActionScript y el Flash Player de Adobe.

Prefuse soporta un amplio conjunto de características de modelado de datos, visualización e interacción. Proporciona datos para optimizar las estructuras de tablas, grafos y árboles, una serie de diseño y técnicas de codificación visual, apoyo a la animación, consultas dinámicas, búsqueda integrada y conectividad de base de datos. Prefuse está escrito en Java, usando la biblioteca de gráficos Java 2D y es fácil su integración en aplicaciones Java Swing y applets Web. Prefuse está bajo los términos de una licencia BSD , y puede ser libremente utilizado con fines comerciales y no comerciales. La galería de visualización y vídeo de demostración proporcionan numerosos ejemplos de los tipos de aplicaciones que se pueden construir con las herramientas Prefuse.

## CONCLUSIONES

Se presenta la segunda versión de la herramienta “POLARIS”, *lográndola convertir en una “Herramienta de minería de Uso y Estructura Web”*.

La arquitectura de POLARIS V2.0 es modular, compuesta por los módulos de interfaz gráficas, kernel y utilidades, lo que permite la reutilización de sus componentes para incluirlos en otras herramientas de este tipo y facilita su mantenimiento.

Al estar esta herramienta desarrollada bajo un lenguaje multiplataforma, como lo es JAVA<sup>TM</sup>, la convierte en una herramienta portable a cualquier sistema operativo.

POLARIS V2.0 está elaborada bajo licencia GPL, aunque algunas Herramientas de Visualización utilizadas se encuentran bajo licenciamiento BSD, como lo es Prefuse.

Las pruebas de funcionalidad realizadas con POLARIS V2.0 confirman que esta herramienta soporta las etapas de minería de datos y minería Web, y presentando la visualización de resultados de una manera correcta y fiable.

La interfaz gráfica de POLARIS V2.0 es amigable, presentado un diseño atractivo y de fácil manejo, que permite exportar gráficas y reportes de los resultados presentados por los algoritmos, en formatos pdf, txt y html, que facilitan el análisis de informes.

En el Kernel de POLARIS V2.0 se encuentran implementados tres tipos de algoritmos de minería de estructura Web como lo son: Page Rank, HITS y Web Linker, y un algoritmo más que busca rutas cortas y optimas dentro de la Web, que permiten al usuario escoger el algoritmo y los parámetros de acuerdo a sus necesidades.

## RECOMENDACIONES

Disponer de una conexión a internet para el correcto funcionamiento de la herramienta, que permita la recuperación de la información en un lapso de tiempo relativamente corto.

Realizar otros tipos de pruebas, como pueden ser ingresando nuevas páginas semillas y niveles de profundidad más elevados, con la finalidad de evaluar el rendimiento de **POLARIS V2.0** con respecto a otras herramientas que involucren la tarea de Minería de Uso y Estructura Web.

Utilizar **POLARIS V2.0** en proyectos de investigación que apliquen la tarea de Minería Web con conjunto de datos reales para soportar la toma de decisiones en la organización estructural de un portal Web.

Implementar el módulo de Minería de Contenido Web para lograr que **POLARIS** se convierta en una suite de minería de Web.

Utilizar **POLARIS** en las electivas de Base de Datos del programa de Ingeniería de Sistemas de la Universidad de Nariño, como herramienta de apoyo, que permita realizar futuros estudios a la estructura y organización del portal Web de la Universidad de Nariño.

Crear un módulo para la automatización de las estadísticas de los datos recolectados por el Crawler y procesados por los algoritmos.

## **ANEXOS**

Los siguientes anexos se encuentran en la carpeta denominada ANEXOS

### **ANEXO 1**

#### **DOCUMENTOS EN FORMATO WORD 2007**

**A.** PRUEBAS PARA LOS ALGORITMOS.docx

**B.** MANUAL DE USUARIO.docx

**C.** ANALISIS DE FUNCIONALIDAD.docx

### **ANEXO 2**

#### **DOCUMENTOS EN FORMATO TXT (ARCHIVO PLANO)**

**A.** En la siguiente carpeta se encuentran los archivos planos para las pruebas de los algoritmos.

TXT PRUEBA ALGORITMOS

**B.** y en la siguiente carpeta se encuentran los archivos en formato .txt, de donde se sacaron las pruebas para el documento de ANALISIS DE FUNCIONALIDAD.docx.

TXT PRUEBAS DE FUNCIONALIDAD

## REFERENCIA BIBLIOGRÁFICA

- [1] **Aldana J.F. [y otros]** Metadata functionality for semantic Web integration. Proceedings of the Seventh International Society of Knowledge Organization (ISKO'02) conference. - Granada, 2002. - págs. 10-13.
- [2] **Angulo Diana, Daza Johana y Zuleta Alejandra** Polaris: Herramienta de minería de uso para la Web. - San Juan de Pasto, 2007.
- [3] **B I POWER BLOGGER** Metodología Extreme Programming(XP) parte I [En línea]. - 28 de Abril de 2009. - 18 de Enero de 2010. - [http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte\\_28.html](http://allxd24.blogspot.com/2009/04/metodologia-extreme-programmingxp-parte_28.html).
- [4] **Baeza Yates Ricardo** Escavando la Web. - Santiago de Chile, 2004. - Vol. 13 : 1 : págs. 4-10.
- [5] **Baeza Yates Ricardo y Garrido Palma Marcelo** Taller: "Desde HTML a la amabilidad: Un Modelo Integral para el Diseño Web". - Santiago de Chile, 2004.
- [6] **Baeza Yates Ricardo y Poblete B.** Una herramienta de minería de consultas para el diseño de contenido y la estructura de un sitio web. - Granada : Thomson, 2005. - págs. 39-48. - ISBN: 84-9732-449-8.
- [7] **Baeza Yates Ricardo y Ribierto Neto B.** Modern information retrieval. - [s.l.] : Addison Wesley Longman, 1999.
- [8] **Berthold M. y Hand D.J.** Intelligent Data Analysis. An Intoduction. - [s.l.] : Springer, 2003. - 2ndEdition.
- [9] **Brin S. y Page L.** The anatomy of a large-scale hypertextual Web search engine . Computer Networks and ISDN Systems. - 1998. - Vols. 30(1-7) : págs. 107–117.
- [10] **Broder A. [y otros]** Graph structure in the web. Computer Networks. - 2000. - págs. 309-320.
- [11] **Carriere J. y Kazman R.** Web Query: Searching and visualizing the Web through structure and associated text. In: International World Wide Web Conference. - 1997.
- [12] **Castillo Carlos** Effective Web Crawling .Rastreo Web eficaces (Ph.D. thesis). (Tesis doctoral).. - 2004.
- [13] **Chakrabarti S. et al** Automatic resource list compilation by analyzing

hyperlink structure and associated text. In: International World Wide Web Conference. - 1998.

[14] **de Kunder M.** WorldWideWebSize.com [En línea] // Daily estimated size of the World Wide Web Tilburg University, Holanda. - 2008. - 25 de Septiembre de 2009. - <http://www.worldwidewebsize.com/>.

[15] **Dean J. y Henzinger M. R** Finding related pages on the web. [En línea] // In: International World Wide Web Conference. - 1999. - 10 de Mayo de 2009. - <http://decweb.ethz.ch/www8/data/2148/html/index.htm>.

[16] **Etzioni O** The World Wide Web: Quagmire or gold mine? Communications of the ACM. - 1996. - Vol. 39 : 11 : págs. 65 - 68.

[17] **Etzioni O., Shakes J. y Langheinrich M.** Ahoy! The homepage finder. Proceedings of the 6th WWW conference. - Santa Clara, 1997.

[18] **Fayyad U.M. [y otros]** Advances in Knowledge Discovery and Data-Mining, AAai Press / The MIT Press. - 1996.

[19] **Fuentes Fernández R. y Pavón J.** Agentes para la recuperación de información especializada en Internet. Facultad de Informática de la Universidad Complutense de Madrid. - Madrid, 2003.

[20] **Gibson D., Kleinberg Jon y Raghavan P.** Inferring Web Communities from Link Topologies. Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, objects, time and space structure in hypermedia system. - 1998. - págs. 225-234.

[21] **González Díaz Ernesto, Pérez Hernández Zady y Espinosa Conde Ivet** monografias.com [En línea] // Técnicas de minería de datos. - 8 de Mayo de 2009. - <http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml>.

[22] **González G, Delfín S y Lluís J** Preprocesamiento de base de datos masivas y multidimensionales en minería de uso web para modelar usuarios: Comparacion de herramientas y técnicas con un caso de estudio.. - Granada : Thomson, 2005. - págs. 193 - 202. - 84-9732-449-8.

[23] **Gulli A. y Signorini A.** "La web es indexable más de 11,5 millones de páginas" . Especial pistas de interés y los carteles de la conferencia internacional sobre el 14<sup>o</sup> en la World Wide Web. [En línea]. - ACM Press., 2005. - 05 de 09 de 2010. - <http://doi.acm.org/10.1145/1062745.1062789>.

[24] **Henzinger R.** Hyperlink analysis for the Web. IEEE Internet Computing. -

2001. - Vol. 5 : 1 : págs. 45-50.

[25] **Hernández Orallo J., Ramírez M.J. y Ferri C.** Introducción a la minería de datos. Pearson Educación. - Madrid : Prentice Hall, 2004.

[26] **Ian H. Witten y Eibe Frank** Data Mining: Practical Machine Learning Tools and techniques with Java Implementations. - [s.l.] : Morgan Kaufmann, 2000.

[27] **IBM** human computer interaction [En línea] // CLEVER Search Technology. - 15 de Marzo de 2009. - <http://www.research.ibm.com/topics/popups/innovate/hci/html/clever.html>.

[28] **Kleinberg Jon M.** Authoritative sources in hyperlinked environment. ACM-SIAM Symposium on Discrete Algorithms. - 1998.

[29] **Kobayashi, M; Takeda, K** "Recuperación de información en la web" . [En línea] // ACM Computing Surveys (ACM Press). - 2000. - 05 de 09 de 2010. - <http://doi.acm.org/10.1145/358923.358934..>

[30] **Kohaavi R. y John G.** Wrappers for the features subset selection. Artificial Intelligence. - 1997. - Vol. 97 : 1-2 : págs. 273-324.

[31] **Lara Navarra P y Martínez Usero J. A** Agentes Inteligentes para la recuperación de Información. - BARCELONA : Planeta UOC S.L., 2004. - ISBN: 84-9707-571-4.

[32] **Lawrence Steve y Giles C. Lee** "Accessibility of information on the web". Nature 400 (6740): 107. doi : 10.1038/21987 . PMID 10428673 . [En línea]. - 08 de 07 de 1999.

[33] **Lempel R. y Moran S.** The stochastic approach for link-structure analysis (SALSA) and the TKC effect. In: International World Wide Web Conference [En línea]. - 2000. - 21 de Mayo de 2009. - <http://www9.org/w9cdrom/175/175.html>.

[34] **Liu B** Information Retrieval and Web Search. En Liu B., Web Data Mining. - 2006. - págs. 183-236.

[35] **Llamas Nistal Martín** Recuperación de Información en Internet: El valor añadido de los hiperenlaces. - 2007-2008.

[36] **Marbán O y Menasalvas E** Estudio de perfiles de visitantes de un sitio Web a partir de los logs de los servidores Web aplicando técnicas de Data Mining (Web Mining). - Madrid, 2002.

[37] **Martins B.** WebLA [En línea] // was developed at the XLDB group of the

Department of Informatics of the Faculty of Sciences of the University of Lisbon.. - 2004. - 05 de 01 de 2009. - <http://webla.sourceforge.net/>.

[38] **Méndez Rodríguez E.M.** Sistemas de búsqueda y recuperación de Información. - Universidad Carlos III de Madrid, 1999.

[39] **Merlino-Santesteban Cristian** Análisis de conectividad en la recuperación de. - Brasília : Ci. Inf., 2003. - Vol. 32 : 3 : págs. 113-119.

[40] **Mobasher B. [y otros]** Web Mining: Patterns from WWW Transactions. Department of computer Science & Engineering. - Minnesota : University of Minnesota, 1996. - TR96-050.

[41] **Moreno Jiménez P.M.** Estrategias y mecanismos de búsqueda en la web invisible. [En línea]. - 2005. - 12 de Junio de 2009. - [http://biblio.colmex.mx/recelec/web\\_invisible.htm..](http://biblio.colmex.mx/recelec/web_invisible.htm..)

[42] **Mycrosys** - Custom designed software development and consulting [En línea] // A1 Website Analyzer and Link Checker Website Analysis. - 1997-2010. - 23 de Marzo de 2009. - <http://www.microsystools.com/products/website-analyzer/>.

[43] **Narvaez Teran H R.** RASEMUS: UNA HERRAMIENTA PARA EL DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS CON TÉCNICAS DE CLUSTERING, DÉBILMENTE ACOPLADO CON EL SGBD POSTGRESQL.. - San Juan de Pasto, Colombia, 2009.

[44] **Navagent, Inc.** Navagent Surf3D [En línea]. - 2001-2008. - 25 de Febrero de 2009. - <http://www.navagent.com/>.

[45] **Nuñez Herrero Julio Alberto** Minería de Textos Web [En línea] // Recuperación y organización de la información. - 12 de Marzo de 2009. - <http://minería-textos-web.awardspace.com/>

[46] **Ortigosa Hernández Jonathan** Técnicas actuales de búsqueda y recuperación de información en Internet: Dominio y uso de técnicas de Data Mining. - 2008. - pág. 10.

[47] **Page Rank Checker PR Checker .info** [En línea]. - 2004-2009. - 25 de Mayo de 2009. - <http://www.prchecker.info/>.

[48] **Pal S.K., Talwar V. y Mitra P.** Web Mining in Soft Computing Framework: Relevance, state of the art and future directions [Sección del libro] // IEEE Transactions on Neural Networks. - 2002. - Vol. 13 : 5.

[49] **Prefuse.org** PREFUSE [En línea] // information visualization toolkit . - 24 de

01 de 2009. - 20 de 03 de 2009. - <http://prefuse.org/>.

[50] **QuestronixSoftware** NetTools Spider [En línea]. - 25 de Mayo de 2009. - <http://www.questronixsoftware.com/>.

[51] **Rafiei D. y Mendelzon A. O.** What is this page known for?: Computing Web Wage Weputations. [En línea] // In: International World Wide Web Conference. - 2000. - 29 de Abril de 2009. - <http://www9.org/w9cdrom/368/368.html>.

[52] **Rosella**, Predictive Knowledge & Data Mining [En línea] // Data Mining & Database Analytics . - 2005. - 20 de Enero de 2009. - <http://www.roselladb.com/surf-pattern-analyzer.htm>.

[53] **SEO ADMINISTRATOR, FlamingoSoft google seo tools - google seo software** SEO ADMINISTRATOR [En línea] // Seo tools, site promotion software program . - 2002-2010. - 15 de Enero de 2009. - <http://www.seoadministrator.com/>.

[54] **Shkapenyuk V. y Suel T.** Design and implementation of a high performance distributed web crawler . In Proceedings of the 18th International Conference on Data Engineering (ICDE). - San Jose, California : IEEE CS Press., 2002. - págs. 357-368.

[55] **TouchGraph, LLC.** TouchGraph [En línea]. - 2007. - 10 de Diciembre de 2008. - <http://www.touchgraph.com/navigator.html>.

[56] **W3C, CSS Validation Service** Verifica Hojas de Estilo en Cascada (CSS) y documentos (X)HTML con hojas de estilo [En línea]. - 1994 - 2009. - 10 de Marzo de 2009. - <http://jigsaw.w3.org/css-validator/>.

[57] **W3C, Markup Validation Service** Check the markup (HTML, XHTML, ...) of Web documents [En línea]. - 1994-2010. - <http://validator.w3.org/>.

[58] **Wikipedia la enciclopedia libre** World Wide Web [En línea]. - 5 de Mayo de 2009. - [http://es.wikipedia.org/wiki/World\\_Wide\\_Web](http://es.wikipedia.org/wiki/World_Wide_Web).

[59] **Wooldridge M. y N.R. Jennings** Intelligent Agents: Theory and Practice. The Knowledge Engineering Review. - 1995. - 10 : págs. 115-152.

[60] **Youngblood G. M.** Web Hunting: Design of a Simple Intelligent Web Search Agent [En línea] // Association for Computer Machinery. - 1999. - 20 de Enero de 2009. - <http://www.acm.org/crossroads/xrds5-4/webhunting.html>.

[61] **Zeinalipour-Yazti D. y and Dikaiakos MD** Design and implementation of a distributed crawler and filtering processor . In Proceedings of the Fifth Next Generation Information Technologies and Systems (NGITS), of Lecture Notes in

Computer. - 2002. - Vol. 2382 : págs. 58-74.

[62] **Zipf G. K.** Human behavior and the principle of least effort: An introduction to human ecology, Addison-Wesley.. - 1949.