

**MOTOR DE BÚSQUEDA INTELIGENTE DE INFORMES DE INVESTIGACIÓN
BASADO EN RECURSOS SEMÁNTICOS PARA EL SISTEMA DE
INVESTIGACIONES DE LA UNIVERSIDAD DE NARIÑO**

**FELIPE CUJAR ROSERO
DAVID SANTIAGO PINCHAO ORTIZ**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2021**

**MOTOR DE BÚSQUEDA INTELIGENTE DE INFORMES DE INVESTIGACIÓN
BASADO EN RECURSOS SEMÁNTICOS PARA EL SISTEMA DE
INVESTIGACIONES DE LA UNIVERSIDAD DE NARIÑO**

**FELIPE CUJAR ROSERO
DAVID SANTIAGO PINCHAO ORTIZ**

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

**DIRECTOR
Ph.D. SILVIO RICARDO TIMARÁN PEREIRA**

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

2021

NOTA DE RESPONSABILIDAD

“Las ideas y conclusiones aportadas en este Trabajo de Grado son responsabilidad de los autores”.

Artículo 1° del Acuerdo No. 324 de octubre 11 de 1966, emanado del Honorable Consejo Directivo de la Universidad de Nariño.

“La Universidad de Nariño no se hace responsable de las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor”.

Artículo 13° del Acuerdo No. 005 de 2010 emanado del Honorable Consejo Académico.



**ACUERDO No. 069
(25 de octubre de 2021)**

**EL CONSEJO DE FACULTAD DE INGENIERIA
DE LA UNIVERSIDAD DE NARIÑO**

En uso de sus atribuciones reglamentarias, estatutarias y,

CONSIDERANDO

Que mediante Acuerdo No. 077 del 10 de diciembre del 2019, el Consejo Académico estableció y unificó la normatividad de los Trabajos de Grado de la Universidad de Nariño.

Que en el Artículo 9 del Acuerdo 077 del 10 de diciembre de 2019, el Consejo Académico de la Universidad de Nariño faculta a los comités curriculares de cada programa para definir y reglamentar los siguientes puntos relacionados con los Trabajos de Grado: a) Criterios de evaluación; b) Criterios para la selección de diplomados y créditos de postgrado; c) Número de estudiantes por Trabajo de Grado; d) Procedimiento de inscripción; e) Aspectos que debe contener cada modalidad de Trabajo de Grado; f) Jurados de evaluación; g) Presentación, sustentación y prerequisitos; h) Criterios de inter y transdisciplinariedad; i) Consideraciones éticas.

Que el Comité Curricular del Departamento de Sistemas, junto con el estamento docente, realizaron sesiones conjuntas que permitieron definir, reglamentar y hacer operativas las modalidades de trabajos de grado acorde con su fundamento disciplinar, acogiéndose a la reglamentación más reciente que es el Acuerdo 077 de 2019 emitido por el Consejo Académico.

Que el cambio en la reglamentación del trabajo de grado no implica una modificación al Estatuto Estudiantil y la aprobación del presente Acuerdo no genera modificación de la normatividad existente, sino que por el contrario lo reglamenta, estableciendo lineamientos claros y definidos para las modalidades acogidas por el Departamento de Sistemas, en lo concerniente a la normatividad relacionada a la presentación, elaboración, seguimiento y calificación de los Trabajos de Grado.

Que el Comité Curricular del Programa de Ingeniería de Sistemas, mediante el Acuerdo No. 095 del 19 de agosto de 2021, aprueba el reglamento para la presentación, aprobación, desarrollo, sustentación, socialización y evaluación de trabajos de grado en las diferentes modalidades de graduación para los programas de pregrado adscritos al Departamento de Sistemas.

Que el Comité Curricular del Programa de Ingeniería de Sistemas, mediante el Acuerdo No. 095 del 19 de agosto de 2021, en el Capítulo 9 artículo 61 define el reconocimiento de las Distinciones para trabajos de grado cuando se realicen bajo la modalidad de Investigación o

Ciudadela Universitaria Torobajo – Bloque 6

Tel: 724 4309 / 731 1449 Ext. 2000 – 320 929 2424

e-mail: ingenieria@udenar.edu.co - facing@udemar.edu.co

<https://www.udenar.edu.co/facultades/ingenieria/>

Pasto – Nariño – Colombia



**ACUERDO No. 069
(25 de octubre de 2021)**

Interacción Social, las cuales serán de Meritorio cuando obtengan una calificación entre 90 a 99 puntos y Laureado para 100 puntos; estos serán asignados si son dignos de presentarse a la comunidad académica en nombre de la Universidad de Nariño, por su aporte original en el campo de las ciencias, la tecnología, las humanidades, las artes o la pedagogía.

- El Consejo de Facultad será el responsable de decidir sobre dichas distinciones y podrá otorgarlas previa presentación de la proposición correspondiente por parte del Comité Curricular, en la cual se adjunte un informe elaborado por los jurados evaluadores que justifique dicho merecimiento.
- Para el otorgamiento de la distinción de Laureado, el Consejo de Facultad podrá solicitar nuevos conceptos de profesionales distintos a los jurados evaluadores.

Que el Comité Curricular del Programa de Ingeniería de Sistemas, mediante Proposición No. 035 de 14 de octubre de 2021, propone al Consejo de Facultad: Otorgar la mención de LAUREADO, al Trabajo de Grado: “Motor de búsqueda inteligente de informes de investigación basado en recursos semánticos para el Sistema de Investigaciones de la Universidad De Nariño”, Modalidad Investigación, presentado por los estudiantes Felipe Cujar Rosero y David Santiago Pinchao Ortiz, bajo la dirección del DhP Silvio Ricardo Timaran

Que en Proposición No. 035 comunica que los ingenieros Manuel Ernesto Bolaños González y Giovanni Albeiro Hernández, recomiendan otorgar Mención de LAUREADO, al Trabajo de Grado presentado por los estudiantes Felipe Cujar Rosero y David Santiago Pinchao Ortiz, teniendo en cuenta lo siguiente:

- Este proyecto presenta un camino teórico para combinar estas dos áreas de conocimiento y se aplican en el dominio de los proyectos de investigación que se desarrollan en la Universidad de Nariño, sin embargo, es trasferible a cualquier otro tipo de ámbito.
- El motor de búsqueda semántico que utiliza una Ontología y un modelo entrenado con Machine Learning, se convierte en una propuesta interesante para que los repositorios con información relacionada con la divulgación de conocimiento científico lo puedan utilizar dentro de sus motores de búsqueda.
- El camino sistemático recorrido para la construcción del motor de búsqueda semántico que utiliza una Ontología y un modelo entrenado con Machine Learning, lo que la convierte en un aporte teórico importante para la comunidad académica y científica que con seguridad será el inicio para nuevos desarrollos tecnológicos.
- Los avances y resultados finales del proyecto han sido socializados a nivel nacional e internacional a través de ponencias en eventos como: a.) Congreso Internacional de Computación (México) - CICOM 2020, b.) Expo póster 2020 Encuentro Académico

Ciudadela Universitaria Torobajo – Bloque 6

Tel: 724 4309 / 731 1449 Ext. 2000 – 320 929 2424

e-mail: ingenieria@udenar.edu.co - facing@udemar.edu.co

<https://www.udenar.edu.co/facultades/ingenieria/>

Pasto – Nariño – Colombia



ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN No.10567 - MAYO 23 DE 2017



GP-CER 112092 SC-CER 110449 CO-SC-CER 110449



**ACUERDO No. 069
(25 de octubre de 2021)**

REDIS Nodo Sur Occidente, c.) 4to Congreso Andino de Computación, Informática y educación – CACIED 2019, d.) Encuentro SOFTHARD 2021, e.) 2nd International Conference on Natural Language Processing and Computational Linguistics (NLPCL 2021), Vancouver, Canada, f.) International Multi-conference for Engineering, Education and Technology 2021 (Argentina).

- La investigación logra publicar los artículos: a.) “K’USIKIYKUY: una ontología para la búsqueda semántica de proyectos de investigación”, b.) “Thaqhaña: Un Motor de Búsqueda Inteligente Basado en Recursos Semánticos”, c.) “Nature: a tool resulting from the union of artificial intelligence and natural language processing for searching research projects in Colombia”, y d.) “Fenix: a semantic search engine based on an ontology and a model trained with machine learning to support research” en revistas de alto reconocimiento e impacto a nivel mundial.

Que los Jurados otorgaron una calificación de (100) puntos al Trabajo de Grado.

Que el Trabajo de Grado cumple con las condiciones necesarias para otorgar la distinción solicitada de Tesis LAUREADO.

Que el Consejo de Facultad mediante sesión de 20 de octubre de 2021 conoció y avaló la Proposición 035 de 14 de octubre de 2021 debido a que las publicaciones y ponencias demuestran la calidad del trabajo siendo evaluado por parte de pares académicos para ser publicado en revistas indexadas internacionales de alto reconocimiento y ha cumplido con la rigurosidad que necesita para ser presentado en eventos internacionales.

En virtud de lo anterior, el Consejo de Facultad,

ACUERDO

ARTÍCULO 1º.

Otorgar la mención de LAUREADO, al Trabajo de Grado: “Motor de búsqueda inteligente de informes de investigación basado en recursos semánticos para el Sistema de Investigaciones de la Universidad De Nariño”, Modalidad Investigación, presentado por los estudiantes Felipe Cujar Rosero y David Santiago Pinchao Ortiz, bajo la dirección del DhP Silvio Ricardo Timaran.

ARTÍCULO 2º.

Notificar del presente Acuerdo a los estudiantes Felipe Cujar Rosero y David Santiago Pinchao Ortiz por medios electrónicos.

ARTICULO 3º

Facultad de Ingeniería, OCARA, Departamento de Sistemas, anotarán lo de su cargo.



**ACUERDO No. 069
(25 de octubre de 2021)**

COMUNÍQUESE Y CÚMPLASE

Dado en San Juan de Pasto, a los veinticinco (25) días del mes de octubre de dos mil veintiuno (2021)


ALEXANDER BARON SALAZAR
Decano

Aprobó: Consejo de Facultad


SANDRA ORDÓÑEZ CERÓN
Secretaria Académica

Nota de Aceptación:

Firma del Presidente del Jurado

References

Firma del Jurado

San Juan de Pasto, octubre de 2021

AGRADECIMIENTOS

Estoy muy alegre por haber culminado exitosamente este proyecto de grado con logros muy destacables en la investigación a niveles nacionales e internacionales.

De esta forma expreso mi agradecimiento primero y antes que nada Al Creador y Dador de la Vida, El Ser Eterno en su máximo esplendor; Dios.

Mi agradecimiento también va dirigido a:

- *Al Alma Máter; la Universidad de Nariño por la calidad y excelencia académica y humana que existe en ella.*
- *A la VIIS (Vicerrectoría de Investigación e Interacción Social) por financiar este proyecto.*
- *Al profesor Ricardo Timarán por su asesoría, buenos consejos, inteligencia y sabiduría.*
- *Al profesor Mateo Guerrero, por su asesoría, su forma de ser juiciosa, sincera y por ser una persona de bien.*
- *A mi compañero de trabajo Santiago Pinchao, por sus constantes deseos de aprender, su mentalidad y su buen carácter para afrontar las situaciones difíciles.*
- *A Constanza de la VIIS por ser un ejemplo de mujer, una persona muy amable, una excelente trabajadora quien nos colaboró con todos los procedimientos que eran sumamente importante realizar en VIIS.*
- *Los seres de mi núcleo familiar que son: mis Padres y Hermanos.*
- *Mis familiares por parte de Padre y Madre.*
- *Amigos y conocidos de la carrera de Ingeniería de Sistemas en la Universidad de Nariño.*

Felipe Cujar Rosero

AGRADECIMIENTOS

Quiero iniciar mis agradecimientos a Dios por la vida y todas las experiencias vividas a lo largo de mi formación profesional.

Agradezco infinitamente a mi siempre Alma Mater la Universidad de Nariño, por brindarme la oportunidad de cumplir ese sueño de cambiar nuestra región y formarme como profesional, con el fin de generar un cambio en nuestro mundo.

Infinitas gracias doy a mis padres JULIO CÉSAR PINCHAO y SANDRA ISABEL ORTIZ, por ser la energía más hermosa que tengo en mi vida al brindarme su total apoyo con el fin de crecer mucho tanto personal como profesionalmente, y por cosechar con éxito el ser humano que soy hoy en día.

No me queda más que agradecerles su amor infinito y espero darles felicidad por siempre, gracias padres.

Gracias a mi otro ser y mi reflejo, mi hermano Julian Pinchao, la persona por quien me he motivado todavía más a que lleguemos a conseguir ese cambio que esperamos aportar en algún momento de nuestras vidas, gracias por ser el cómplice de sabiduría y conocimientos que hemos forjado juntos, que la ciencia nos lleve muy lejos. Te quiero mucho hermano.

Gracias a mi primo y padrino Juan Carlos Ortiz, por guiarme en el camino de quien soy hoy en día, por sus primeras clases de programación, las cuales iniciaron este lindo sueño en mi. Los consejos de vida que guardo y aplico en mi día a día, por muchos años espero cosechar junto a ti logros y alegrías como familia.

Agradezco la total confianza y entrega dada a este proyecto, al asesor Ing. Ricardo Timaran Pereira PhD. gracias por darme la oportunidad de concluir mis estudios en la ciudad de Pasto y permitir cumplir mi sueño de ser investigador, así mismo por sus conocimientos impartidos en todo momento. Igualmente agradezco al Ing. Mateo Guerrero, por su constante apoyo y ayuda en el desarrollo del proyecto, gracias también por permitirme dar mis primeros pasos dentro del mundo de la Ciencia de Datos y la Inteligencia Artificial. A Felipe Cujar, por permitirme trabajar junto con él en esta investigación, que este sea el inicio de muchos grandes proyectos, y gracias por ser parte fundamental en todo momento, por su atención al detalle y finalmente la fuerza necesaria para poder culminar este gran proyecto. Finalmente agradezco a todos los docentes que dejaron sus enseñanzas en mi en el día a día de nuestras clases, gracias a todas esas personas con las que compartí tanto en Pasto como en Ipiales, siempre me motivaron a seguir con este camino y a cumplir esa meta que todos tenemos, gracias por esos maravillosos momentos compartidos que hicieron parte de vida universitaria.

GRACIAS POR SIEMPRE, LOS QUIERO.

David Santiago Pinchao Ortiz

DEDICATORIA

Este trabajo lo dedico a:

- *El ejemplo de lucha, templanza y sabiduría; mi Papá: Álvaro Cujar.*
- *El Amor en persona, la ternura y bondad del alma, la fuerza de los duros momentos que se vivió en el transcurso de la carrera, la sonrisa y ojos más bellos; mi Mamá: Ligia Rosero.*
- *Un cariño sincero lleno de muchos abrazos y expresiones de alegría, una persona inquebrantable y de adaptación a las circunstancias, un papá más; mi Hermano: Pablo Cujar.*
- *Un ejemplo de rectitud, perseverancia, prudencia y razón, un papá más; mi Hermano: Ricardo Cujar.*

Felipe Cujar Rosero

DEDICATORIA

Dedico este logro, a todas las personas que desde un principio estuvieron conmigo. A mis padres por su incondicional apoyo en todo lo que necesitaba, quienes depositaron su confianza en mí para poder cumplir este sueño que es también gran parte de ellos y su formación de vida que me permite cumplir todo lo que anhelo y deseo. Dedico este trabajo a mi hermano y sus futuros logros como científico, que siempre compartas ciencia en toda tu vida.

A toda la familia Pinchao por compartir conmigo todo los momentos alegres y tristes, y saber que siempre podré contar con ellos, que este logro sea un pequeño agradecimiento por todo lo que han hecho por mí. A todos mis primos y primas por ser la motivación de seguir siempre adelante y ser un profesional, a Evita por permitirme saber que en su momento me iba a enfrentar a lo mismo y aconsejarme a nunca rendirme y hacerme saber la persona que soy. Dedico este trabajo a mis tíos Ernesto y Marina, gracias por permitirme saber que les di felicidad en su momento y que estaban muy felices de saber quién puedo llegar a ser. A mi abuelita Chava por todo ese amor compartido todos los días de mi vida, y a mis abuelos Carmen y Milton, por llamarle siempre el niño Santi y tener ese amor infinito hacia mí.

A mis compañeros Luis Carlos Ortega, Daniel Muñoz, Lorena Bravo, Daniela Aza, Jhonantan Chacua, Estefany Díaz, por compartir mucho conmigo en todo momento, por ser los mejores en todo siempre y permitirme llenarlos con mi alegría, durante todos nuestros estudios.

Finalmente dedico esto a la persona que me permitió ser parte de su vida y fue mi cómplice en todo momento, Daniela Vallejo Uscategui, gracias por amarme y permitir amarte en ese hermoso tiempo, por darme las alas que necesitaba para ir más alto de lo que pensé que podía llegar, deseo con muchas ansias que sigas cumpliendo tus sueños, y que todo el universo nos llene de energías hermosas por siempre.

David Santiago Pinchao Ortiz

RESUMEN

En este documento se presenta la construcción de un motor de búsqueda semántico que utilizó una Ontología y un modelo entrenado con Machine Learning para soportar la búsqueda semántica de proyectos de investigación del Sistema de Investigaciones de la Universidad de Nariño. Para la construcción de THAQHAÑA, como se le denomina a este motor, se utilizó la metodología que comprende las etapas de: apropiación del conocimiento; instalación y configuración de las herramientas, bibliotecas y tecnologías; recolección, extracción, preparación, exploración y análisis de proyectos de investigación; diseño, desarrollo, implementación y despliegue del motor de búsqueda semántico; realización de pruebas finales y evaluación del rendimiento del motor de búsqueda semántico.

Los resultados principales del trabajo fueron los siguientes: a) la construcción completa de la Ontología con classes (clases), object properties(predicados), data properties(atributos) e individuals (instancias) en Protegé, consultas SPARQL con Apache Jena Fuseki y la respectiva codificación con Owlready2 usando Jupyter Notebook con Python dentro del entorno virtual de Anaconda; b) el entrenamiento exitoso del modelo para el cual se usaron algoritmos de Machine Learning y específicamente de Procesamiento de Lenguaje Natural como: SpaCy, NLTK, Word2Vec y Doc2Vec, esto también se realizó en Jupyter Notebook con Python dentro del entorno virtual de Anaconda y con Elasticsearch; y c) la creación de THAQHAÑA gestionando y unificando las consultas para la Ontología y para el modelo de Machine Learning. Las pruebas demostraron que THAQHAÑA fue exitoso en todas las búsquedas llevadas a cabo porque sus resultados fueron satisfactorios.

ABSTRACT

In this document, the construction of a semantic search engine was carried out, this semantic search engine used an Ontology and a model trained with Machine Learning to support the semantic search of research projects of the Research System of the University of Nariño. For the construction of THAQHAÑA, as this engine is called, it was used a methodology that comprises the stages of: appropriation of knowledge; installation and configuration of tools, libraries and technologies; collection, extraction, preparation, exploration and analysis of research projects; design, development, implementation and deployment of the semantic search engine; performing of final tests and evaluating the performance of the semantic search engine.

The main results of the work were the following: a) the complete construction of the Ontology with classes, object properties (predicates), data properties (attributes) and individuals (instances) in Protégé, SPARQL queries with Apache Jena Fuseki and the respective coding with Owlready2 using Jupyter Notebook with Python within the Anaconda virtual environment; b) the successful training of the model for which Machine Learning algorithms and specifically Natural Language Processing algorithms were used such as: SpaCy, NLTK, Word2Vec and Doc2Vec, this was also done in Jupyter Notebook with Python within the virtual environment of Anaconda and with Elasticsearch; and c) the creation of THAQHAÑA managing and unifying the queries for the Ontology and for the Machine Learning model. The tests showed that THAQHAÑA was successful in all the searches carried out because its results were satisfactory.

CONTENIDO

	pág.
INTRODUCCIÓN	19
1. GENERALIDADES DEL PROYECTO	21
1.1 MODALIDAD	21
1.2 LÍNEA DE INVESTIGACIÓN	21
1.3 ALCANCE Y DELIMITACIÓN	21
1.4 PLANTEAMIENTO DEL PROBLEMA	21
1.5 FORMULACIÓN DEL PROBLEMA	25
1.6 OBJETIVOS	25
1.6.1 Objetivo general.....	25
1.6.2 Objetivos específicos.....	25
1.7 JUSTIFICACIÓN	26
1.8 ORGANIZACIÓN DEL DOCUMENTO.....	27
2. ESTADO DEL ARTE	27
3. MARCO TEÓRICO	32
3.1 WEB SEMÁNTICA	32
3.2 ONTOLOGÍA.....	36
3.3 LENGUAJE RDF (RESOURCE DESCRIPTION FRAMEWORK)	38
3.4 LENGUAJE XML	40
3.5 LENGUAJE OWL	42
3.6 MOTORES DE BÚSQUEDA O BUSCADORES	42
3.7 SPARQL	44
3.8 SCRUM	44
3.9 APRENDIZAJE DE MÁQUINA O APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING).....	45

3.10 PROCESAMIENTO DE LENGUAJE NATURAL (NATURAL LANGUAGE PROCESSING).....	48
3.11 JUPYTER NOTEBOOK	52
3.12 PROTÉGÉ	52
3.13 OWLREADY2	53
3.14 APACHE JENA FUSEKI (AJF)	54
3.15 ELASTICSEARCH	54
3.16 DOCKER	56
3.17 FLASK	57
3.18 PYTHON	58
3.19 METHONTOLOGY	60
3.20 POSTGRESQL	63
 4. METODOLOGÍA	65
4.1 APROPIACIÓN DEL CONOCIMIENTO	68
4.2 INSTALACIÓN Y CONFIGURACIÓN DE LAS HERRAMIENTAS, BIBLIOTECAS Y TECNOLOGÍAS.....	68
4.3 RECOLECCIÓN Y EXTRACCIÓN DE PROYECTOS DE INVESTIGACIÓN.....	68
4.4 PREPARACIÓN DE PROYECTOS DE INVESTIGACIÓN	68
4.5 EXPLORACIÓN Y ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN.....	69
4.6 DISEÑO DEL MOTOR DE BÚSQUEDA SEMÁNTICO.....	69
4.7 DESARROLLO DEL MOTOR DE BÚSQUEDA SEMÁNTICO	70
4.8 IMPLEMENTACIÓN Y DESPLIEGUE DEL MOTOR DE BÚSQUEDA SEMÁNTICO	71
4.9 REALIZACIÓN DE PRUEBAS FINALES Y EVALUACIÓN DEL RENDIMIENTO DEL MOTOR DE BÚSQUEDA SEMÁNTICO	71
 5. ARQUITECTURA	72
5.1 CAPAS DE LA ARQUITECTURA DEL MOTOR DE BÚSQUEDA	72
5.2 MÓDULOS DE LA ARQUITECTURA DEL MOTOR DE BÚSQUEDA	73

6. RESULTADOS	76
6.1 APROPIACIÓN DEL CONOCIMIENTO	76
6.2 INSTALACIÓN Y CONFIGURACIÓN DE LAS HERRAMIENTAS, BIBLIOTECAS Y TECNOLOGÍAS.....	76
6.3 RECOLECCIÓN Y EXTRACCIÓN DE PROYECTOS DE INVESTIGACIÓN.....	76
6.4 PREPARACIÓN DE PROYECTOS DE INVESTIGACIÓN	78
6.4.1 Organización de datos	78
6.4.2 Creación de corpus	78
6.4.3 Preprocesamiento de datos	78
6.4.3.1 Tokenización de datos	79
6.4.3.2 Normalización de datos	79
6.4.3.3 Limpieza de datos	80
6.4.3.3 Lematización de datos	80
6.5 EXPLORACIÓN Y ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN.....	81
6.6 DISEÑO DEL MOTOR DE BÚSQUEDA SEMÁNTICO.....	85
6.6.1 Fase de especificación	85
6.6.2 Fase de conceptualización	87
6.7 DESARROLLO DEL MOTOR DE BÚSQUEDA SEMÁNTICO.....	96
6.7.1 Desarrollo con Methontology	96
6.7.1.1 Fase de formalización	96
6.7.1.2 Fase de implementación	99
6.7.1.3 Fase de evaluación	101
6.7.2 Desarrollo con Machine Learning	102
6.7.3 Desarrollo unificado de Ontología y Machine Learning.....	115
6.8 IMPLEMENTACIÓN Y DESPLIEGUE DEL MOTOR DE BÚSQUEDA SEMÁNTICO.....	116
6.9 REALIZACIÓN DE PRUEBAS FINALES Y EVALUACIÓN DEL RENDIMIENTO DEL MOTOR DE BÚSQUEDA SEMÁNTICO.....	117
6.9.1 Casos de prueba	117
6.9.2 Pruebas unitarias	119

6.9.3 Pruebas de aceptación	120
6.9.4 Pruebas de comparación	125
7. DISCUSIÓN	129
8. CONCLUSIONES	130
9. RECOMENDACIONES	132
BIBLIOGRAFÍA	134

LISTA DE TABLAS

	pág.
Tabla 1. Equivalencias entre componentes lingüísticos y lógicos en RDF.....	40
Tabla 2. Adquisición del conocimiento para la Ontología.....	85
Tabla 3. Glosario de términos de la Ontología	88
Tabla 4. Diccionario de concepto Investigador	91
Tabla 5. Diccionario de concepto Proyecto de investigación.....	91
Tabla 6. Diccionario de concepto Palabra	92
Tabla 7. Relación binaria (pi_tiene_palabra) en detalle	92
Tabla 8. Relación binaria (palabra_sinonimo_palabra) en detalle	93
Tabla 9. Relación binaria (palabra_conecta_palabra) en detalle	93
Tabla 10. Atributos de instancia de clase Investigador.....	94
Tabla 11. Atributos de instancia de clase Proyecto de investigación	94
Tabla 12. Atributos de instancia de clase Palabra	94
Tabla 13. Atributos de clases Investigador, Docente y Proyecto de investigación.....	95
Tabla 14. Hiperparámetros para los algoritmos Word2Vec y Doc2Vec	103

Tabla 15. Caso de prueba 1	118
Tabla 16. Caso de prueba 2	118
Tabla 17. Caso de prueba 3.....	119
Tabla 18. Pruebas Unitarias.....	119
Tabla 19. Pruebas de aceptación para PU-001 Búsqueda de proyectos docentes.....	121
Tabla 20. Pruebas de aceptación para PU-002 Búsqueda de proyectos estudiantiles.....	121
Tabla 21. Pruebas de aceptación para PU-003 Búsqueda de trabajos de grado.....	122
Tabla 22. Pruebas de aceptación para PU-004 Descarga de proyectos docentes.....	123
Tabla 23. Pruebas de aceptación para PU-005 Visualización en línea de proyectos docentes.....	123
Tabla 24. Pruebas de aceptación para PU-006 Descarga de proyectos estudiantiles.....	124
Tabla 25. Pruebas de aceptación para PU-007 Visualización en línea de proyectos estudiantiles.....	124
Tabla 26. Pruebas de aceptación para PU-008 Descarga de trabajos de grado.....	124

Tabla 27. Pruebas de aceptación para PU-009 Visualización en línea de trabajos de grado.....	125
Tabla 28. Pruebas de aceptación para PU-010 Usuario administrador	125
Tabla 29. Pruebas de comparación entre motor de búsqueda semántico y buscador manual	127

LISTA DE FIGURAS

	pág.
Figura 1. Niveles de la Web semántica	35
Figura 2. Arquitectura del motor de búsqueda	72
Figura 3. Ejemplo de texto al cual se le aplicó preprocesamiento de datos	79
Figura 4. Texto tokenizado	79
Figura 5. Texto normalizado	80
Figura 6. Texto limpio	80
Figura 7. Texto lematizado	81
Figura 8. Nube de palabras con BOW	81
Figura 9. Nube de palabras con TF-IDF	82
Figura 10. Histograma con BOW	83
Figura 11. Histograma con TF-IDF	84
Figura 12. Tareas de la metodología Methontology	87
Figura 13. Taxonomía de conceptos de la Ontología	89
Figura 14. Diagrama de relaciones binarias de la clase Proyecto de Investigación.....	90
Figura 15. Clases de la Ontología.....	96
Figura 16. Atributos de la Ontología.....	97
Figura 17. Relaciones de la Ontología.....	98
Figura 18. Instancia de un proyecto de investigación con todos sus datos.....	99

Figura 19. Resultado consulta SPARQL para la relación docente_es_autor_pi.....	101
Figura 20. Resultado consulta SPARQL para la relación estudiante_es_autor_pi.....	102
Figura 21. Resultado consulta SPARQL para la relación convocatoria_tiene_pi.....	102
Figura 22. Resultado de método con Word2Vec para palabra “neuronales”.....	105
Figura 23. Resultado de método con Word2Vec para palabra “carnaval”.....	105
Figura 24. Resultado de método con Word2Vec para palabra “cultivos”.....	106
Figura 25. Resultado de método con Word2Vec para palabra “fresa”	106
Figura 26. Resultado de método con Word2Vec para palabra “psicologia”	107
Figura 27. Resultado de método con Word2Vec para palabra “historia”	107
Figura 28. Resultado de método con Word2Vec para palabra “matematicas”	108
Figura 29. Resultado de método con Word2Vec para palabra “medicos”.....	108
Figura 30. Resultado de método con Word2Vec para palabra “redes”.....	109
Figura 31. Resultado de método con Word2Vec para palabra “nariño”.....	109
Figura 32. Resultado de método con Word2Vec para palabra “pasto”	110
Figura 33. Primer cuadrante palabras semánticamente relacionadas	112
Figura 34. Segundo cuadrante palabras semánticamente relacionadas	113
Figura 35. Tercer cuadrante palabras semánticamente relacionadas	114
Figura 36. Cuarto cuadrante palabras semánticamente relacionadas	115
Figura 37. Implementación de Thaqhaña: motor de búsqueda semántico	116
Figura 38. Resultados para la búsqueda “fisica” en el buscador manual	126

Figura 39. Resultados para la búsqueda “amenaza vocanica” en el buscador manual.....126

Figura 40. Resultados para la búsqueda “algebra” en el buscador manual126

LISTA DE ANEXOS

	pág.
ANEXO A. ANÁLISIS SCRUM	139
ANEXO B. CLASES, ATRIBUTOS Y MÉTODOS	152
ANEXO C. DICCIONARIO DE DATOS.....	162
ANEXO D. LOGROS - PUBLICACIONES - PONENCIAS - ARTÍCULOS - PRESENTACIONES – VIIS	167
ANEXO E. PRUEBAS.....	181

GLOSARIO

ALGORITMO: es un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo otras tareas o actividades.

APRENDIZAJE DE MÁQUINA O APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING): es la ciencia de hacer que las computadoras, las máquinas o equipos de cómputo realicen una tarea y descubran conocimiento sin estar programados explícitamente.

BASE DE DATOS: es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

CLASES DE ONTOLOGÍA: son la base de la descripción del conocimiento en las Ontologías ya que describen los conceptos del dominio. Una clase puede ser dividida en subclases, las cuales representan conceptos más específicos que la clase a la que pertenecen. Una clase cuyos componentes son clases, se denomina superclase o metaclase.

CORPUS: es un conjunto cerrado de textos o de datos destinados a la investigación científica.

INSTANCIAS DE ONTOLOGÍA: son objetos, miembros de una clase, que no pueden ser divididos sin perder su estructura y características funcionales. Pueden ser agrupados en clases.

INTELIGENCIA ARTIFICIAL: es, en informática, la inteligencia expresada por máquinas, sus procesadores y sus softwares, que serían los análogos al cuerpo, el cerebro y la mente, respectivamente, a diferencia de la inteligencia natural demostrada por humanos y ciertos animales con cerebros complejos.

ONTOLOGÍA: es una especificación explícita y formal de una conceptualización compartida.

PROPIEDADES DE ONTOLOGÍA: las instancias se describen por medio de un conjunto de características o atributos que son almacenados en las propiedades. Estas almacenan diferentes clases y tipos de valores. Dichas propiedades tienen especificaciones, rangos, dominios y restricciones. Para una clase dada, las propiedades y las restricciones sobre estas son heredadas para las subclases y las instancias de la clase.

PROCESAMIENTO DE LENGUAJE NATURAL (NATURAL LANGUAGE PROCESSING): es el campo del Machine Learning que permite que una máquina pueda entender, analizar, manipular y potencialmente generar lenguaje humano.

RELACIONES DE ONTOLOGÍA: se establecen entre conceptos de una Ontología para representar las interacciones entre éstos. Definidas por lo general como el producto cartesiano de n conjuntos.

SEMÁNTICA: es la parte de la lingüística que estudia diversos aspectos del significado, sentido o interpretación de signos lingüísticos como símbolos, palabras, expresiones o representaciones formales.

SINTAXIS: es la parte de la lingüística que estudia el orden y la relación de las palabras o sintagmas en la oración, así como las funciones que cumplen.

INTRODUCCIÓN

De acuerdo con Berners Lee y Hendler, en la Web Semántica la información se ofrece con un significado bien definido, permitiendo a ordenadores y personas trabajar de forma cooperativa. La idea que existe detrás de la Web Semántica es tener datos en la Web definidos y enlazados de manera que puedan ser usados de forma más efectiva para descubrimiento, automatización, integración y reutilización entre diferentes aplicaciones. El reto de la Web Semántica es ofrecer el lenguaje que exprese tanto datos como reglas para razonar sobre diversos datos y además permita que las reglas sobre cualquier sistema de representación del conocimiento sean exportadas a la Web, aportando un importante grado de “flexibilidad y frescura” a los sistemas de representación de conocimiento centralizados tradicionales, que se vuelven sumamente agobiantes, crecen rápidamente de tamaño y se vuelven inmanejables. Diferentes sistemas web pueden utilizar diversos identificadores para un mismo concepto; así, un programa que quiera comparar o combinar información entre dichos sistemas tiene que conocer cuáles términos tienen el mismo significado; idealmente, el programa debería tener una forma de descubrir los significados comunes de cualquier base de datos que encuentre. Una solución a este problema es incluir un nuevo elemento a la Web Semántica; colecciones de información denominadas Ontologías¹.

De igual forma, se sabe que la gran cantidad de información textual disponible en la Red junto con el aumento de la demanda por parte de los usuarios hacen necesaria la existencia de sistemas que permitan un acceso a aquella información de interés de una forma eficiente y efectiva, ahorrando así tiempo en su búsqueda y consulta. Entre las técnicas existentes para lograr esa eficiencia y efectividad, y a su vez para proporcionar acceso o facilitar la gestión de información de documentos de texto se encuentran técnicas de Machine Learning, usarlas es altamente conveniente, esto se puede evidenciar en un elevado número de aplicaciones en diferentes ámbitos².

Actualmente en el Sistema de Investigaciones de la Universidad de Nariño en la VIIS, el ingeniero encargado, al momento de realizar búsquedas para estas investigaciones de tipo: trabajo de grado, proyectos estudiantiles y proyectos docentes, comenta que existe demora en los procesos, él narra que dichos procesos no son óptimos, que a veces es difícil encontrar lo que se quiere, en muchas ocasiones no ha logrado encontrar lo requerido. Esto indica que no existe ni eficiencia ni efectividad garantizada al momento de realizar búsquedas sobre las investigaciones, puesto que la búsqueda se está realizando de forma manual. Es decir que al tener un sistema de búsqueda manual de la

¹ GARCÍA, Francisco. Web Semántica y Ontologías. [En línea]. Disponible en: https://www.researchgate.net/publication/267222548_Web_Semantica_y_Ontologias

² MOURIÑO, M. Clasificación multilingüe de documentos utilizando machine learning y la wikipedia. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/tesis?codigo=150295>

información que no tiene el calificativo de motor de búsqueda sintáctico, se deduce que la información no tiene una estructura clara para ser presentada y que los procesos son ineficientes en las búsquedas. Esto lleva a la conclusión de que es totalmente necesario construir el motor de búsqueda semántico inteligente, dado que si el problema persiste, a medida que la información vaya aumentando, las búsquedas serán más tediosas y dispendiosas, adicionalmente con el motor de búsqueda semántico inteligente la cantidad de población de usuarios que realicen búsquedas sobre determinado tema será mayor y estará satisfecha por encontrar los resultados deseados.

Así las cosas este proyecto le brinda una herramienta que le permita a los docentes, estudiantes y demás investigadores realizar búsquedas y consultas sobre las investigaciones que se han realizado en la Universidad de Nariño. En el presente proyecto de modalidad investigación se ha construido un motor de búsqueda semántico haciendo uso de la semántica mediante el lenguaje de consultas SPARQL y del lenguaje RDF con el manejo de Ontologías, todo esto integrado con un modelo de Machine Learning que utiliza algoritmos y bibliotecas como: NLTK, SpaCy, Word2Vec, Doc2Vec, TF-IDF (Term Frequency- Inverse Document Frequency), BOW (Bag Of Words), entre otros. De esta forma se puede facilitar el trabajo y permitir a los investigadores y a la comunidad en general recuperar y encontrar la información solicitada eficientemente de los proyectos de investigación que se encuentran digitalizados en el Sistema de Investigaciones de la Universidad de Nariño. El proyecto fue desarrollado por estudiantes investigadores del grupo de investigación GRIAS del programa de Ingeniería de Sistemas, de la facultad de Ingeniería de la Universidad de Nariño.

1. GENERALIDADES DEL PROYECTO

1.1 MODALIDAD

Este proyecto corresponde a la modalidad de trabajo de grado de tipo investigación, fue desarrollado por estudiantes investigadores del Grupo de Investigación Aplicado en Sistemas (GRIAS) del Departamento de Sistemas de la Universidad de Nariño.

1.2 LÍNEA DE INVESTIGACIÓN

Este proyecto corresponde a la línea de investigación titulada: "Herramientas y sistemas de gestión de conocimiento y recuperación de la información". Dicha línea se sustenta según proposición No. 010 de septiembre de 2002 del Comité Curricular y de Investigaciones del Departamento de Sistemas, está ratificada por el Consejo de Facultad mediante Acuerdo No 045 de octubre de 2002.

1.3 ALCANCE Y DELIMITACIÓN

Desde un comienzo se planteó que el proyecto estaría delimitado en la etapa de diseño y construcción de la herramienta con la respectiva realización de pruebas con una muestra representativa de proyectos de investigación de docentes, estudiantes y trabajos de grado, además se delimitó en no abarcar toda la información de la VIIS sino cierto porcentaje.

1.4 PLANTEAMIENTO DEL PROBLEMA

Internet fue concebida por Tim Berners-Lee como un proyecto para gestionar y compartir conocimiento e información entre un selecto grupo de científicos. Con el pasar del tiempo y a la par de los avances en la elaboración del hardware que hacía posible la comunicación alrededor del mundo, se fueron desarrollando los aplicativos necesarios para satisfacer las necesidades de usuarios cada vez más elevadas en número y en importancia. El gran volumen de contenidos disponibles en línea dificulta su búsqueda y procesamiento, ha sido vital pues la necesidad de ingeniar nuevas formas de optimizar el tratamiento que se da a dichos contenidos; para que la información disponible en la Web sea interpretada por los computadores sin necesidad de intervención humana, se requiere de la Web Semántica. Se habla de una Internet en la que los computadores no

sólo son capaces de presentar la información contenida en las páginas Web, sino que además pueden “entender” dicha información³.

Según Kappel se ve pertinente hacer uso de la semántica, la cual se refleja en las respuestas que recibe un usuario a sus peticiones en los buscadores, ya que estas van más allá del estado en el que los usuarios simplemente realizaban una pregunta y recibían un conjunto ordenado por prioridad de páginas web. Los usuarios desean respuestas dirigidas a sus preguntas sin información superflua. Las respuestas deben contener información de fuentes autorizadas, términos con el mismo significado a los usados en la pregunta, enlaces relevantes, etc. Entonces, la Web Semántica trata de dotar de una estructura semántica a los contenidos significativos de la Web, creando un entorno en el que agentes de software naveguen por las páginas realizando complejas tareas para los usuarios⁴.

Los factores que han generado el éxito de Internet, también han originado problemas tales como: sobrecarga de información, heterogeneidad de fuentes y problemas consiguientes de interoperabilidad. La Web Semántica ayuda a resolver estos problemas, al permitir a los usuarios delegar tareas en herramientas de software. Incorporando semántica en la Web, el software es capaz de procesar contenidos, razonar con ellos, combinarlos y realizar deducciones lógicas para resolver problemas automáticamente. La habilidad automática es el resultado de la aplicación de técnicas de inteligencia artificial, las cuales exigen la participación de agentes inteligentes que perfeccionen las búsquedas, agregando valores de razonamiento y toma de decisión a los servicios web que almacenan altos contenidos⁵.

Se parte del supuesto que esta Web tiene la capacidad de construir una base de conocimiento sobre las preferencias de los usuarios y que, a través de una combinación entre su capacidad de comprensión y la información disponible en Internet, es capaz de atender de forma exacta las demandas de información por parte de los usuarios, por ejemplo: la reserva de restaurante, programación de vuelos, consultas médicas, compra de libros, etc. Así, el usuario obtendría resultados exactos sobre su búsqueda, sin mayores complicaciones porque la Web Semántica aporta un camino para razonar en la Web al ser una infraestructura basada en metadatos (datos altamente estructurados que describen información), extendiendo así sus capacidades. Es decir, no se trata de una inteligencia artificial mágica que permita a los servidores Web entender las palabras de

³ VELÁSQUEZ, Torcoroma, PUENTES, Andrés & GUZMÁN, Jaime. Ontologías: una técnica de representación de conocimiento. En: Avances en Sistemas e Informática. Vol. 8. No. 2. (Julio, 2011), p. 211-216. [En línea]. Disponible en: <https://revistas.unal.edu.co/index.php/avances/article/view/26750>

⁴ Ibíd.

⁵ EFIGENIA, Ana & CANTOR, Sandoval. USO DE ONTOLOGÍAS Y WEB SEMÁNTICA PARA APOYAR LA GESTIÓN DEL CONOCIMIENTO. En: Ciencia e Ingeniería Neogranadina. Vol. 17 No. 2. (Diciembre, 2007), p.111-129. [En línea]. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/2512191.pdf>

los usuarios, es por el contrario la construcción de una habilidad dispuesta en una máquina, con el fin de resolver problemas bien definidos, a través de operaciones igualmente bien definidas que se llevarán a cabo sobre datos existentes⁶.

Un motor de búsqueda es una aplicación y/o recurso informático que permite localizar información en los servidores de determinada Red, dando como resultado una lista que va acorde con los archivos o materiales almacenados en los servidores correspondientes y que da respuesta a las necesidades del usuario. Los motores de búsqueda facilitan la tarea de localizar la información que está dispersa por la Red, pero es crucial conocer la forma en la que se está realizando la búsqueda⁷.

Los motores de búsqueda sintácticos hacen uso de palabras clave, donde el resultado de la búsqueda depende de un proceso de indexación, que es aquel que va a permitir organizar las búsquedas con estas palabras clave o mediante el uso de árboles jerárquicos categorizados por determinado tema. A pesar de la potencia que demuestran los motores de búsqueda sintácticos, aún quedan lejos de poder proporcionar al usuario los resultados adecuados a las consultas realizadas, ya que la cantidad de resultados pueden ser demasiados y por ende será bastante dispendioso encontrar el resultado deseado o bien no obtener ningún resultado, con el agregado de que gran parte de la responsabilidad de la búsqueda puede caer en manos del usuario, quien tendría que filtrar y categorizar su búsqueda para llegar a obtener una respuesta clara y concisa.

En síntesis, como se puede observar, estos problemas pueden ser solucionados con el uso de motores de búsqueda semánticos los que, en cambio, le facilitan el trabajo al usuario, son eficientes en la búsqueda ya que encuentran resultados en función del contexto, proporcionando así información más exacta acerca de lo que se busca, ofreciendo una cantidad de resultados más sesgada, facilitando la labor de filtrar los resultados por parte del usuario. Es de esta forma como estos motores de búsqueda interpretan las búsquedas de los usuarios haciendo uso de algoritmos que simbolizan comprensión o entendimiento, ofreciendo resultados precisos de manera rápida y así reconocen el contexto correcto para las palabras o sentencias de búsqueda. No es más que un motor de búsqueda semántico, es decir, aquel que realiza la búsqueda fijándose en el significado del grupo de palabras que están escritas⁸.

Actualmente en el Sistema de Investigaciones de la Universidad de Nariño en la VIIS, el ingeniero encargado, al momento de realizar búsquedas para estas investigaciones de

⁶ Ibíd.

⁷ GALLO, Manuel, FABRE, Ernesto & GALLO, Manuel. ¿Qué es un buscador? [En línea]. Disponible en: http://media.axon.es/pdf/98234_1.pdf

⁸ DE PEDRO, A. Buscadores Semánticos, para qué sirven. Usos en la AAPP. [En línea]. Disponible en: <http://www.alejandrolepedro.es/buscadores-semanticos-el-paso-al-30>

tipo: trabajo de grado, proyectos estudiantiles y proyectos docentes, comenta que existe demora en los procesos, él narra que dichos procesos no son óptimos, que a veces es difícil encontrar lo que se quiere, en muchas ocasiones no ha logrado encontrar lo requerido. Esto indica que no existe ni eficiencia ni efectividad garantizada al momento de realizar búsquedas sobre las investigaciones, puesto que la búsqueda se está realizando de forma manual. Es decir que al tener un sistema de búsqueda manual de la información que no tiene el calificativo de motor de búsqueda sintáctico, se deduce que la información no tiene una estructura clara para ser presentada y que los procesos son ineficientes en las búsquedas. Esto lleva a la conclusión de que es totalmente necesario construir el motor de búsqueda semántico inteligente, dado que si el problema persiste, a medida que la información vaya aumentando, las búsquedas serán más tediosas y dispendiosas, adicionalmente con el motor de búsqueda semántico inteligente la cantidad de población de usuarios que realicen búsquedas sobre determinado tema será mayor y estará satisfecha por encontrar los resultados deseados.

Así las cosas este proyecto le brinda una herramienta que le permita a los docentes, estudiantes y demás investigadores realizar búsquedas y consultas sobre las investigaciones que se han realizado en la Universidad de Nariño. En el presente proyecto de modalidad investigación se ha construido un motor de búsqueda semántico haciendo uso de la semántica mediante el lenguaje de consultas SPARQL y del lenguaje RDF con el manejo de Ontologías, todo esto integrado con un modelo de Machine Learning que utiliza algoritmos y bibliotecas como: NLTK, SpaCy, Word2Vec, Doc2Vec, TF-IDF (Term Frequency- Inverse Document Frequency), BOW (Bag Of Words), entre otros. De esta forma se puede facilitar el trabajo y permitir a los investigadores y a la comunidad en general recuperar y encontrar la información solicitada eficientemente de los proyectos de investigación que se encuentran digitalizados en el Sistema de Investigaciones de la Universidad de Nariño. El proyecto fue desarrollado por estudiantes investigadores del grupo de investigación GRIAS del programa de Ingeniería de Sistemas, de la facultad de Ingeniería de la Universidad de Nariño.

1.5 FORMULACIÓN DEL PROBLEMA

¿Cómo soportar la búsqueda y recuperación inteligente de los proyectos de investigación que se encuentran en formato digital en el Sistema de Investigaciones de la Universidad de Nariño?

1.6 OBJETIVOS

1.6.1 Objetivo general. Soportar la búsqueda y recuperación inteligente de los proyectos de investigación que se encuentran en formato digital en el Sistema de Investigaciones de la Universidad de Nariño a través de la construcción de un motor de búsqueda semántico con el fin de facilitar las consultas sobre proyectos de investigación realizados en la Universidad de Nariño y las posibles relaciones con otros proyectos de investigación.

1.6.2 Objetivos específicos.

- ✓ Definir el ciclo de vida del proyecto de investigación.
- ✓ Apropiar el conocimiento de todas las bases teóricas que componen la armonía de la investigación basado en la abstracción de la literatura y el estado del arte.
- ✓ Identificar los requerimientos del motor de búsqueda semántico, la base de datos y el dominio de la Ontología teniendo en cuenta los proyectos de investigación de la Universidad de Nariño.
- ✓ Diseñar el modelo del motor de búsqueda semántico, la estructura de la Ontología y la base de datos.
- ✓ Desarrollar e implementar el motor de búsqueda semántico, acoplando la Ontología a las consultas con SPARQL y RDF en el Sistema de Investigaciones de la Universidad de Nariño.
- ✓ Realizar pruebas de funcionalidad de la herramienta y evaluar los resultados.
- ✓ Elaborar un artículo o ponencia para ser publicado en una revista indexada o en las memorias de un evento especializado de índole regional, nacional y/o internacional.
- ✓ Elaborar el documento final del proyecto de grado y sustentarlo.

1.7 JUSTIFICACIÓN

Las diferentes investigaciones y trabajos de grado de diversos niveles se apoyan de otras investigaciones que se encuentran en el Sistema de Investigaciones de la Universidad de Nariño. En este proyecto, aparte de la meta de obtener un motor de búsqueda semántico con innovación y eficiencia, se quiso también fomentar la participación

colaborativa y la unión en el ámbito de la investigación en la Universidad de Nariño; herramientas como ésta propenden al crecimiento de la investigación, ayudan y aportan a todo aquel que sea y se considere un investigador de la Universidad de Nariño.

Ésta herramienta orientada a la comunidad investigadora adquiere un componente de innovación fundamental porque incorpora tecnologías, conocimientos y herramientas innovadoras como son: la Web Semántica, las Ontologías, los motores de búsqueda, SPARQL, el lenguaje RDF y Machine Learning fundamentándose en el esquema de las bases de datos y la IA (Inteligencia Artificial); dos potentes ámbitos que en la actualidad son la raíz de muchos frutos obtenidos en la tecnología, que lo que hacen es facilitar la vida a las personas, fomentar la competitividad, motivar al cambio y generar conocimiento que enriquezca la vida informática y computacional.

Así las cosas, en este proyecto se ha construido un motor de búsqueda semántico que permite a los investigadores recuperar y acceder de manera inteligente y fácil a los proyectos de investigación que se encuentran digitalizados en el Sistema de investigaciones de la Universidad de Nariño y se enfoca en ofrecer una herramienta que ayude en el crecimiento de la calidad de las investigaciones que se realizan en la Universidad de Nariño, desarrollado por estudiantes investigadores del grupo de investigación GRIAS del programa de Ingeniería de Sistemas, de la facultad de Ingeniería de la Universidad de Nariño.

De esta manera la Universidad de Nariño a través del grupo de investigación GRIAS, emprende iniciativas para reforzar el campo investigativo de la Universidad de Nariño.

1.8 ORGANIZACIÓN DEL DOCUMENTO

Este documento se ha organizado de la siguiente manera: primero: se encuentran las generalidades del proyecto ya mencionadas; segundo: está el estado del arte en el cual se describen proyectos similares o relacionados, que se han realizado en la historia en los ámbitos: internacional, nacional y regional; tercero: se encuentra una breve descripción de los componentes teóricos utilizados, tanto herramientas, lenguajes y conceptos en general, cuarto: está la metodología utilizada para el proyecto, donde se describe cada una de las diversas etapas y pasos para lograr el objetivo general; quinto: se encuentra la arquitectura en la cual se pueden evidenciar las capas y módulos utilizados; sexto: están los múltiples resultados obtenidos en el proyecto de investigación; séptimo: se encuentra la sección de discusión con un análisis objetivo y crítico del proyecto; octavo: están las conclusiones finales del proyecto; noveno las recomendaciones y sugerencias; finalmente décimo: se encuentra la bibliografía con la cual fue orientado y guiado este proyecto. Se culmina presentando los anexos más importantes del proyecto de investigación.

2. ESTADO DEL ARTE

Los motores de búsqueda semánticos encuentran la información de forma más precisa, aunque en la actualidad no se ha masificado el uso de estos, ya se tienen algunos funcionando y sirven de referencia para el futuro de la búsqueda de información⁹.

Algunos motores de búsqueda semánticos, que utilizan la semántica para encontrar resultados en función del contexto, información más exacta acerca de lo que se busca, son: biblioteca digital de ERCIM¹⁰, NDLTD¹¹, Wolfram Alpha¹². Este último es capaz de responder directamente a las preguntas que hace el usuario en lugar de proporcionar una lista de los documentos o páginas web que podrían contener la respuesta, tal y como lo hace Google. Una vez formulada la pregunta, la herramienta calcula diferentes respuestas eligiendo de forma selectiva la información de la Red para acabar dando una respuesta precisa. Swotti es otro buscador que utiliza tecnologías de la Web Semántica para extraer las opiniones que realizan los usuarios en blogs y foros sobre empresas o productos. Es capaz de identificar los adjetivos y verbos que definen aquello que se está buscando, y que por tanto permiten deducir si el comentario es positivo o negativo. Cuando se hace una búsqueda en Swotti se obtiene no sólo resultados, sino sobre todo una valoración cualitativa¹³. Swoogle es buscador de documentos para la Web Semántica, un Google para la Web Semántica aunque todavía no está dirigido al usuario final, se ha creado en la Universidad de Maryland, no está destinado al usuario común, sino para el rastreo de documentos web semánticos cuyos formatos son OWL, RDF y DAML. Swoogle es un buscador que detecta, analiza e indexa el conocimiento codificado como documentos de Web Semántica. Swoogle entiende por documentos de Web Semántica aquéllos que están escritos con algunos de los lenguajes orientados a la construcción de Ontologías (RDF, OWL, DAML, N3, etc). Recupera tanto documentos escritos completamente en estos lenguajes (que para Swoogle son documentos de Web Semántica estrictos) como documentos escritos parcialmente con alguno de ellos. Asimismo, proporciona un algoritmo inspirado también en el algoritmo Page Rank de Google, que para Swoogle se ha denominado Ontology Rank. El Ontology Rank se ha adaptado a la semántica y a los modelos y patrones de uso que se encuentran en los documentos de Web Semántica. Actualmente, Swoogle tiene indexados alrededor de 1.5

⁹ FAZZINGA, Bettina, GIANFORME, Giorgio, GOTTLÖB, Georg & LUKASIEWICZ, Thomas. Semantic Web Search Based On Ontological Conjunctive Queries. En: SSRN Electronic Journal. [En línea]. Disponible en: https://www.researchgate.net/publication/326473981_Semantic_Web_Search_Based_on_Ontological_Conjunctive_Queries

¹⁰ ANDREONI, Antonella, BALDACCI Maria, BIAGONI, Stefania, CARLESI, Carlo, CASTELLI, Donatella, PAGANO, Pasquale, PETERS, Carol & PISANI, Serena. The ERCIM Technical Reference Digital Library. En: D-Lib Magazine. Vol. 5. No. 12. (Diciembre, 1999). [En línea]. Disponible en: <http://www.dlib.org/dlib/december99/peters/12peters.html>

¹¹ NDLTD. Networked Digital Library of Theses and Dissertations. [En línea]. Disponible en: <http://www.ndltd.org>

¹² WolframAlpha Computational Intelligence. [En línea]. Disponible en: <https://www.wolframalpha.com>

¹³ MARTÍN, Javier. Swotti buscador de opiniones. [En línea]. Disponible en: <https://loogic.com/swotti-buscador-de-opiniones>

M de documentos de Web Semántica. Dicha información está disponible mediante un enlace interno a los datos estadísticos relativos a su estado¹⁴.

Entre algunos trabajos estudiantiles a nivel internacional, relacionados con el tema de esta investigación están: el de Camacho Rodríguez en su trabajo de grado para optar el título de Ingeniería de Telemática propone incorporar un buscador semántico en la plataforma LdShake para la selección de patrones educativos. Este trabajo fue desarrollado en la Universidad Pompeu Fabra- UPF de Barcelona, España en el año 2013. En este trabajo se analiza la eficiencia de utilizar Ontologías para mejorar considerablemente los resultados y a la vez ganar velocidad en la búsqueda¹⁵. Amaral presenta un buscador semántico para el idioma portugués donde se hace uso de herramientas de Procesamiento de Lenguaje Natural y un corpus léxico multilingüe donde se evalúan las consultas del usuario, para la desambiguación de palabras polisémicas se hace uso de pivotes mostrados en la pantalla con los diferentes significados de la palabra donde el usuario escoge el sentido con el que desea realizar la consulta¹⁶. Aucapiña y Plaza en su tesis para la obtención del Título de Ingeniero en Sistemas proponen un Buscador semántico universitario para la Universidad de Cuenca en Cuenca, Ecuador en el año 2018, donde describen de forma detallada el uso de SPARQL como lenguaje de consulta y las diversas etapas llevadas a cabo para la consecución del prototipo del buscador semántico siguiendo metodologías probadas y en ciertos casos son soportados por procesos automatizados¹⁷. Umpiérrez Rodríguez en su trabajo final de grado en Ingeniería Informática denominado “SPARQL Interpreter” en la Universidad de Las Palmas de Gran Canaria, desarrollado en el año de 2014, donde expone como SPARQL Interpreter afronta el problema de la comunicación entre un lenguaje de consulta y una base de datos específica¹⁸. Baculima y Cajamarca en su tesis de grado en Ingeniería de Sistemas denominado “Diseño e Implementación de un Repositorio Ecuatoriano de Datos Enlazados Geoespaciales”, desarrollada en la Universidad de Cuenca Ecuador, en el año 2014, trabajan la solución de generación, publicación y visualización datos Enlazados Geoespaciales, para lo cual se apoyan de los buscadores web, esto ya que la Web se enfoca en la publicación de este tipo de datos,

¹⁴ BARBERÁ, Consuelo, MILLET, Mercé & TORRES, Emiliano. Estudio del buscador semántico Swoogle. [En línea]. Disponible en: <https://www.uv.es/etomar/trabajos/swoogle/swoogle.pdf>

¹⁵ CAMACHO, María. Incorporación de un buscador semántico en la plataforma LdShake para la selección de patrones educativos. Barcelona, 2013, 76p. Trabajo de grado. Universidad Pompeu Fabra. Escuela Superior Politécnica UPF. Ingeniería de Telemática. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/22172>

¹⁶ AMARAL, Carlos, LAURENT, Dominique, MARTINS, André, MENDES, Alfonso & PINTO, Cláudia. Design and Implementation of a Semantic Search Engine for Portuguese. [En línea]. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.4090&rep=rep1&type=pdf>

¹⁷ AUCAPIÑA, Yolanda & PLAZA, C. Buscador semántico universitario: Caso de estudio Universidad de Cuenca. Cuenca, 2018, 200p. Trabajo de grado (Tesis previa a la obtención del Título de Ingeniero en Sistemas). Universidad de Cuenca. Facultad de Ingeniería. Ingeniería de Sistemas. [En línea]. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/30291>

¹⁸ UMPIÉRREZ, Francisco. SPARQL Interpreter. Las Palmas de Gran Canaria, 2014, 65p. Trabajo de grado (Trabajo Final de Grado en Ingeniería Informática). Universidad de Las Palmas de Gran Canaria. Escuela Ingeniería Informática. Ingeniería Informática. [En línea]. Disponible en: https://nanopdf.com/download/0701044000000000pdf_pdf

permitiendo que estén estructurados de tal manera que se puedan interconectar entre diferentes fuentes. En este trabajo se apoyan de SPARQL Y GEOSPARQL para poder realizar consultas, inserción modificación y eliminación de los datos¹⁹.

Entre algunos proyectos estudiantiles sobre motores de búsqueda semánticos desarrollados a nivel nacional están: el de Iglesias, desarrollado en la Universidad Simón Bolívar de Barranquilla y cuyo objetivo fue construir un motor de búsqueda ontológico que permita realizar búsquedas semánticas en línea de trabajos formativos de maestrías y doctorados, donde se puedan encontrar esta clase de trabajos o temas que puedan servir como una guía para que surja una nueva investigación, de esta forma mejorar las búsquedas a la hora de seleccionar los temas de investigación para trabajos de grado²⁰. Bustos Quiroga en su tesis de grado en la Maestría en Ingeniería de Sistemas y Computación desarrolla un “Prototipo de un sistema de integración de recursos científicos, diseñado para su funcionamiento en el espacio de los datos abiertos enlazados para mejorar la colaboración, la eficiencia y promover la innovación en Colombia” en 2015 en la Universidad Nacional de Colombia. En este trabajo se utiliza Web Semántica en datos enlazados para mejorar la integración en intemporalidad entre aplicaciones y facilitar el acceso a la información a través de modelos unificados y formatos de datos compartidos²¹. Moreno y Sánchez en su trabajo de grado para optar el título de Ingeniero de Sistemas y Computación proponen un prototipo de motores de búsqueda semántico aplicado a la búsqueda de libros de Ingeniería de Sistemas y Computación en la biblioteca Jorge Roa Martínez de la Universidad Tecnológica de Pereira. Este trabajo fue desarrollado en el año 2012. Este prototipo se desarrolló partiendo de los fundamentos teóricos existentes y, del análisis que se llevó a cabo acerca de las tecnologías que se involucran, como son los agentes inteligentes de software, las Ontologías implementadas en lenguajes como RDF y XML, y demás herramientas de desarrollo²².

¹⁹ BACULIMA, Jhon & CAJAMARCA, Marcelo. Diseño e Implementación de un Repositorio Ecuatoriano de Datos Enlazados Geoespaciales. Cuenca, 2014, 131p. Trabajo de grado (Tesis de Grado previa a la obtención del Título: Ingeniero de Sistemas). Universidad de Cuenca. Facultad de Ingeniería. Ingeniería de sistemas. [En línea]. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/19876>

²⁰ IGLESIAS, Daniela, MEJÍA, Omar, NIETO, Julio, SÁNCHEZ, Steven & MORENO, Silvia. Construcción de un buscador ontológico para búsquedas semánticas de proyectos de maestría y doctorado. En: Investigación y Desarrollo en TIC. Vol. 7. No. 1. (Mayo, 2017), p. 7-13. [En línea]. Disponible en: <https://revistas.unisimon.edu.co/index.php/identic/article/view/2501>

²¹ BUSTOS, Gabriel. Prototipo de un sistema de integración de recursos científicos, diseñado para su funcionamiento en el espacio de los datos abiertos enlazados para mejorar la colaboración, la eficiencia y promover la innovación en Colombia. Bogotá, 2018. Tesis de Maestría. Universidad Nacional de Colombia. Facultad de Ingeniería. Ingeniería de Sistemas e Industrial. [En línea]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/55245>

²² MORENO, Carlos & SÁNCHEZ, Yakeline. Prototipo de buscador semántico aplicado a la búsqueda de libros de Ingeniería de Sistemas y Computación en la biblioteca Jorge Roa Martínez de la Universidad Tecnológica de Pereira. Pereira, 2012, 66p. Trabajo de grado. Universidad Tecnológica de Pereira. Facultad de Ingenierías: Eléctrica, Electrónica, Física y Ciencias de la Computación. Ingeniería De Sistemas y Computación. [En línea]. Disponible en: <http://repositorio.utp.edu.co/dspace/bitstream/11059/2671/1/0057565M843.pdf>

A nivel regional, en la Universidad de Nariño, Benavides y Guerrero desarrollaron el proyecto de trabajo de grado para optar el título de Ingeniero de Sistemas, en el año 2013, denominado “UMAYUX: un modelo de software de gestión de conocimiento soportado en una Ontología dinámica débilmente acoplada con un gestor de base de datos para la Universidad de Nariño” cuyo objetivo fue definir un modelo de software de gestión de conocimiento soportado en una Ontología dinámica débilmente acoplado con un gestor de bases de datos que convierta el conocimiento que era tácito, dentro de los procesos académicos y administrativos de la Universidad de Nariño, en conocimiento explícito y que permita recopilar, estructurar, almacenar información y transformarla mediante el uso de Ontologías de dominio específico, de tal forma que cada unidad académica o dependencia administrativa puede construir y acoplar al modelo. El modelo UMAYUX se implementó mediante la construcción de MASKANA, una herramienta de gestión de conocimiento soportada en una Ontología dinámica sobre trabajos de grado de los estudiantes de pregrado del programa de Ingeniería de Sistemas del departamento de Sistemas de la Facultad de Ingeniería, débilmente acoplada con el SGBD PostgreSQL²³.

²³ BENAVIDES, Mauricio & GUERRERO, Jimmy. UMAYUX: un modelo de software de gestión de conocimiento soportado en una ontología dinámica débilmente acoplado con un gestor de base de datos. San Juan de Pasto, 2014, 145p. Trabajo de grado (Trabajo de grado presentado como requisito parcial para optar al título de Ingeniero de Sistemas). Universidad de Nariño. Facultad de Ingeniería. Ingeniería de Sistemas. [En línea]. Disponible en: <http://sired.udenar.edu.co/2030>

3. MARCO TEÓRICO

3.1 WEB SEMÁNTICA

Internet fue concebida por Tim Berners-Lee como un proyecto para gestionar y compartir conocimiento e información entre un selecto grupo de científicos. Con el pasar del tiempo y a la par de los avances en la elaboración del hardware que hacía posible la comunicación alrededor del mundo, se fueron desarrollando los aplicativos necesarios para satisfacer las necesidades de usuarios cada vez más elevadas en número y en importancia. El gran volumen de contenidos disponibles en línea dificulta su búsqueda y procesamiento, ha sido vital pues la necesidad de ingeniar nuevas formas de optimizar el tratamiento que se da a dichos contenidos; para que la información disponible en la Red sea interpretada por los computadores sin necesidad de intervención humana, se requiere de la Web Semántica. Se habla de una Internet en la que los computadores no sólo son capaces de presentar la información contenida en las páginas Web, sino que además pueden “entender” dicha información²⁴.

De acuerdo con Berners Lee y Hendler en la Web Semántica la información se ofrece con un significado bien definido, permitiendo a ordenadores y personas trabajar de forma cooperativa. La idea que existe detrás de la Web Semántica es tener datos en la Web definidos y enlazados de manera que puedan ser usados de forma más efectiva para un descubrimiento, una automatización, una integración y una reutilización entre diferentes aplicaciones. El reto de la Web Semántica es ofrecer el lenguaje que exprese tanto datos como reglas para razonar sobre los datos y además permita que las reglas sobre cualquier sistema de representación del conocimiento sean exportadas a la Web, aportando un importante grado de flexibilidad y “frescura” a los sistemas de representación de conocimiento centralizados tradicionales, que se vuelven sumamente agobiantes, crecen rápidamente de tamaño y se vuelven inmanejables. Diferentes sistemas web pueden utilizar diferentes identificadores para un mismo concepto; así, un programa que quiera comparar o combinar información entre dichos sistemas tiene que conocer qué términos significan lo mismo; idealmente, el programa debería tener una forma de descubrir los significados comunes de cualquier base de datos que encuentre. Una solución a este problema es incluir un nuevo elemento a la Web Semántica; colecciones de información denominadas Ontologías²⁵.

Según Kappel se ve pertinente hacer uso de la semántica, la cual se refleja en las respuestas que recibe un usuario a sus peticiones en los buscadores, ya que éstas van más allá del estado en el que los usuarios simplemente realizaban una pregunta y

²⁴ VELÁSQUEZ, Torcoroma, PUENTES, Andrés & GUZMÁN, Jaime. Ontologías: una técnica de representación de conocimiento. En: Avances en Sistemas e Informática. Vol. 8. No. 2. (Julio, 2011), p. 211-216. [En línea]. Disponible en: <https://revistas.unal.edu.co/index.php/avances/article/view/26750>

²⁵ GARCÍA, Francisco. Web Semántica y Ontologías. [En línea]. Disponible en: https://www.researchgate.net/publication/267222548_Web_Semantica_y_Ontologias

recibían un conjunto ordenado por prioridad de páginas web. Los usuarios desean respuestas dirigidas a sus preguntas sin información superflua. Las respuestas deben contener información de fuentes autorizadas, términos con el mismo significado a los usados en la pregunta, enlaces relevantes, etc. (McGuinness, 2004). Entonces, la Web Semántica trata de dotar de una estructura semántica a los contenidos significativos de la Web, creando un entorno en el que agentes de software naveguen por las páginas realizando complejas tareas para los usuarios²⁶.

Los factores que han generado el éxito de Internet, también han originado problemas tales como: sobrecarga de información, heterogeneidad de fuentes y problemas consiguientes de interoperabilidad. La Web Semántica ayuda a resolver estos problemas, al permitir a los usuarios delegar tareas en herramientas de software. Incorporando semántica en la Web, el software es capaz de procesar contenidos, razonar con ellos, combinarlos y realizar deducciones lógicas para resolver problemas automáticamente. La habilidad automática es el resultado de la aplicación de técnicas de inteligencia artificial, las cuales exigen la participación de agentes inteligentes que perfeccionen las búsquedas, agregando valores de razonamiento y toma de decisión a los servicios web que almacenan altos contenidos²⁷.

Se parte del supuesto que esta Web tiene la capacidad de construir una base de conocimiento sobre las preferencias de los usuarios y que, a través de una combinación entre su capacidad de comprensión y la información disponible en Internet, es capaz de atender de forma exacta las demandas de información por parte de los usuarios, por ejemplo: la reserva de restaurante, programación de vuelos, consultas médicas, compra de libros, etc. Si esto ocurriese de tal forma en la vida real, el usuario, obtendría resultados exactos sobre su búsqueda, sin mayores complicaciones, pero aunque la realidad es otra, la Web Semántica aporta un camino para razonar en la Web al ser una infraestructura basada en metadatos (datos altamente estructurados que describen información), extendiendo así sus capacidades. Es decir, no se trata de una inteligencia artificial mágica que permita a los servidores Web entender las palabras de los usuarios, es sólo la construcción de una habilidad dispuesta en una máquina, con el fin de resolver problemas bien definidos, a través de operaciones igualmente bien definidas que se llevarán a cabo sobre datos existentes²⁸.

La Web semántica proporciona un marco común que permite que los datos sean compartidos y reutilizados a través de aplicaciones, empresas y fronteras comunitarias. Es un esfuerzo colaborativo liderado por el W3C con la participación de un gran número

²⁶ Ibíd.

²⁷ EFIGENIA, Ana & CANTOR, Sandoval. USO DE ONTOLOGÍAS Y WEB SEMÁNTICA PARA APOYAR LA GESTIÓN DEL CONOCIMIENTO. En: Ciencia e Ingeniería Neogranadina. Vol. 17 No. 2. (Diciembre, 2007), p.111-129. [En línea]. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/2512191.pdf>

²⁸ Ibíd.

de investigadores y socios industriales. Está basado en Resource Description Framework (RDF) e integra una variedad de aplicaciones utilizando XML para la sintaxis²⁹.

Según Pedraza la carencia de un modelo bien definido de representación de la información en la Web ha traído consigo problemas de cara a diversos aspectos relacionados con su procesamiento. Para intentar solucionarlos, el W3C, organismo encargado de guiar la evolución de la Web, ha propuesto su transformación hacia una nueva Web denominada Web semántica³⁰.

La idea es que los datos puedan ser utilizados y “comprendidos” por los ordenadores sin necesidad de supervisión humana, de forma que los agentes web puedan ser diseñados para tratar la información situada en las páginas web de manera semiautomática. Se trata de convertir la información en conocimiento referenciando datos dentro de las páginas web como metadatos con un esquema común consensuado sobre algún dominio. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que además podrán tener información adicional de cómo hacer deducciones con ellos, es decir, axiomas que podrán aplicarse en los diferentes dominios que trate el conocimiento almacenado. Con ello, se mejorará la búsqueda de información y se potenciará el desarrollo de diversas aplicaciones, ya que las anotaciones de información seguirán un esquema común y los buscadores web compartirán con las anotaciones web los mismos esquemas. Los agentes web no sólo encontrarán la información de forma precisa, si no que podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas, y con los requerimientos de la consulta indicada por el usuario³¹.

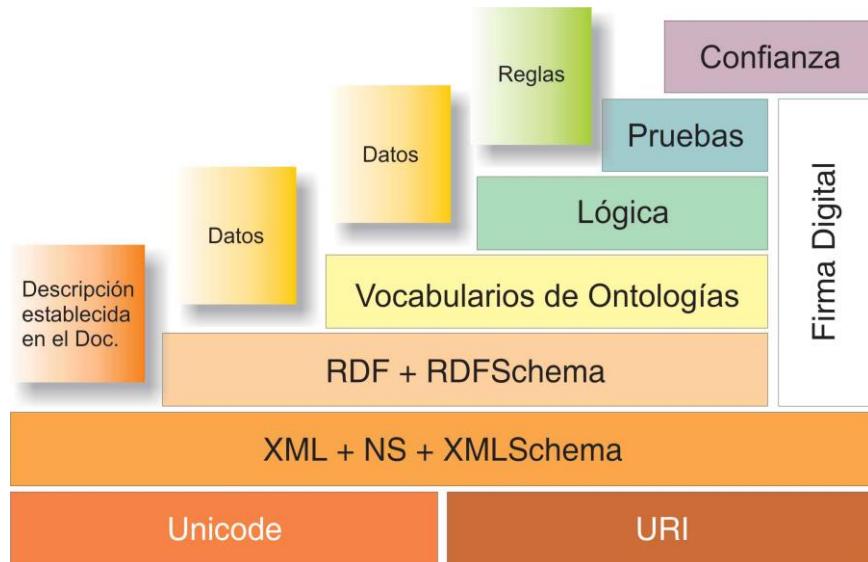
En la figura 1 se puede evidenciar los niveles de la Web semántica:

Figura 1. Niveles de la Web semántica

²⁹ CODINA, Lluís & CRISTÒFOL, Rovira. La Web Semántica. En: Jesús Tramullas (coord.). Tendencias en documentación digital. Guijón: Trea, 2006. p. 9-54. [En línea]. Disponible en: <http://eprints.rclis.org/8899>

³⁰ PEDRAZA, Rafael, CODINA, Lluís & CRISTÒFOL, Rovira. Web semántica y ontologías en el procesamiento de la información documental. En: El profesional de la información. Vol. 16. No. 6. (Noviembre, 2007), p. 569-579. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/13141>

³¹ LOZANO, Adolfo. Ontologías en la Web Semántica. [En línea]. Disponible en: <http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Ontologias%20en%20la%20Web%20Semantica.pdf>



Fuente: Codina y Rovira³²

- ✓ Nivel de recursos (Unicode + URI): en el primer nivel se incluye la identificación de recursos web, estableciendo así la gran importancia que tiene definir el conjunto de recursos distribuidos por la Red.
- ✓ Nivel sintáctico (XML + NS + XMLSchema): en este nivel se soluciona el problema de cómo definir distintos lenguajes de etiquetado para añadir contenido semántico a las páginas web (XML + NS + XMLSchema). Esta hace referencia a la capa más técnica de la Web Semántica.
- ✓ Nivel de descripción de recursos (RDF + RDFSchema): basado y apoyado en el nivel anterior, este nivel define el lenguaje universal con el cual podemos expresar diferentes ideas en la Web Semántica. RDF es un lenguaje simple mediante el cual definimos sentencias en el formato de una triple (sujeto, predicado y objeto). Tanto este nivel como el anterior corresponden a las anotaciones de la información (metadatos).
- ✓ Nivel de vocabulario de Ontologías: este nivel permite extender la funcionalidad de la Web Semántica, agregando nuevas clases y propiedades para describir los recursos, siendo las Ontologías la piedra angular de la propuesta de Tim Berners-Lee, ya que ofrecen un criterio para catalogar y clasificar la información.

³² CODINA, Lluís & CRISTÒFOL, Rovira. La Web Semántica. En: Jesús Tramullas (coord.). Tendencias en documentación digital. Guijón: Trea, 2006. p. 9-54. [En línea]. Disponible en: <http://eprints.rclis.org/8899>

- ✓ Nivel de lógica: además de las Ontologías se precisan reglas de inferencia; eso es de lo que trata este nivel. El computador no “entiende” nada de lo que está procesando, pero puede manipular los términos de modo mucho más eficiente beneficiando la inteligibilidad humana. El nivel de la lógica, pretende así, dar flexibilidad a la arquitectura para realizar consultas e inferir conocimiento a partir de las Ontologías del nivel anterior.
- ✓ Nivel de pruebas: para este nivel, será necesario el intercambio de pruebas escritas en el lenguaje unificador de la Web Semántica (se trata del lenguaje que hace posible las inferencias lógicas a través del uso de reglas de inferencia tal como es especificado por las Ontologías).
- ✓ Nivel de confianza: es importante que los agentes sean muy escépticos acerca de lo que leen en la Web Semántica hasta que hayan podido comprobar de forma exhaustiva las fuentes de información que le suministran la información requerida; así se tiene este nivel dentro de la Web Semántica.
- ✓ Firma digital: es un bloque encriptado de datos que será utilizado por los computadores y los agentes para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable³³.

3.2 ONTOLOGÍA

Una de las definiciones más completas y utilizadas en el mundo investigativo para explicar el concepto de Ontología es la de Borst que completa la versión original de Gruber, Borst dice que una Ontología es “una especificación explícita y formal de una conceptualización compartida”. Studer explica que una “conceptualización” es un modelo abstracto de un fenómeno del mundo construido mediante la observación, análisis e identificación de los conceptos más importantes dentro del ámbito de ese fenómeno, esto es, el dominio del conocimiento. “Explícito” significa que los conceptos utilizados en la Ontología y las restricciones para su uso están claramente definidos y son entendibles. “Formal” se refiere al hecho de que debe estar en los términos de un lenguaje comprensible para las máquinas, es decir estar expresada mediante una sintaxis (por ejemplo el lenguaje OWL) que permita a un ordenador operar sobre ella. Por último, “compartida” refiere a la noción de que tiene conocimiento consensuado

³³ VELÁSQUEZ, Torcoroma, PUENTES, Andrés & GUZMÁN, Jaime. Ontologías: una técnica de representación de conocimiento. En: Avances en Sistemas e Informática. Vol. 8. No. 2. (Julio, 2011), p. 211-216. [En línea]. Disponible en: <https://revistas.unal.edu.co/index.php/avances/article/view/26750>

y/o aprobado en algún grado por expertos en un determinado dominio³⁴. Entre las principales razones para crear Ontologías se puede citar la de Noy & McGuinness quienes afirman que la Ontología sirve para “compartir un entendimiento común de una estructura de información entre personas o agentes de software, posibilitar la reutilización de conocimiento de un dominio, hacer explícitas suposiciones de un dominio, separar el conocimiento de un dominio del conocimiento operacional, o analizar el conocimiento de un dominio”³⁵.

Las Ontologías tienen estos componentes clave, para el desarrollo exitoso de la semántica:

- ✓ Clase: son la base de la descripción del conocimiento en las Ontologías ya que describen los conceptos del dominio. Una clase puede ser dividida en subclases, las cuales representan conceptos más específicos que la clase a la que pertenecen. Una clase cuyos componentes son clases, se denomina superclase o metaclass.
- ✓ Instancias o individuos: son objetos, miembros de una clase, que no pueden ser divididos sin perder su estructura y características funcionales. Pueden ser agrupados en clases.
- ✓ Relaciones: se establecen entre conceptos de una Ontología para representar las interacciones entre éstos. Definidas por lo general como el producto cartesiano de n conjuntos. Algunas de las relaciones más utilizadas son:
 - Instancia de: asocian objetos a clases.
 - Relaciones temporales: implican precedencia en el tiempo.
 - Relaciones topológicas: establecen conexiones espaciales entre conceptos.
- ✓ Propiedades o slots: las instancias se describen por medio de un conjunto de características o atributos que son almacenados en las propiedades. Estas almacenan diferentes clases y tipos de valores. Dichas propiedades tienen especificaciones, rangos, dominios y restricciones. Para una clase dada, las propiedades y las restricciones sobre estas son heredadas para las subclases y las instancias de la clase.

³⁴ PEDRAZA, Rafael, CODINA, Lluís & CRISTÒFOL, Rovira. Web semántica y ontologías en el procesamiento de la información documental. En: El profesional de la información. Vol. 16. No. 6. (Noviembre, 2007), p. 569-579. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/13141>

³⁵ GARCÍA, Francisco. Web Semántica y Ontologías. [En línea]. Disponible en: https://www.researchgate.net/publication/267222548_Web_Semantica_y_Ontologias

- ✓ Axioma: elementos que permiten el modelado de verdades que se cumplen siempre en la realidad. Los axiomas pueden ser estructurales y/o no estructurales. Un axioma estructural establece condiciones relacionadas con la jerarquía de la Ontología, conceptos y atributos definidos. Ejemplo: el concepto A no es una clase de A. Un axioma no estructural establece relaciones entre atributos de un concepto y son específicos de un dominio. Ejemplo: la relación $f = m \cdot a$, que debe cumplirse siempre entre los atributos f (fuerza), m (masa) y a (aceleración) de un determinado concepto.
- ✓ Conceptualización: conjunto de conceptos, relaciones, objetos y restricciones que caracterizan un dominio.
- ✓ Taxonomía: conjunto de conceptos organizados jerárquicamente. Las taxonomías definen las relaciones entre los conceptos, pero no los atributos de estos.
- ✓ Vocabulario: conjunto de palabras con una explicación y documentación que persigue la universalidad y el formalismo en el contexto de un dominio³⁶.

3.3 LENGUAJE RDF (RESOURCE DESCRIPTION FRAMEWORK)

El lenguaje RDF o Infraestructura para la Descripción de Recursos es muy útil en situaciones en las que la información necesita ser procesada por aplicaciones que intercambian información legible por máquina, más que por humanos. RDF provee un marco común de trabajo para expresar esta información y para intercambiárla entre aplicaciones distintas mediante una serie de “parsers” o analizadores RDF y otras herramientas de procesamiento automatizado. RDF puede utilizarse en diferentes áreas como en la recuperación de recursos para los motores de búsqueda, robots, agentes inteligentes y en general IA (Inteligencia Artificial), junto con otras áreas como catalogación para describir el contenido y las relaciones de contenido disponibles en un sitio web o en una colección de documentos para describir los derechos de propiedad intelectual o las políticas de privacidad de un sitio web, entre otros³⁷.

RDF es el sistema que permite utilizar metadatos para describir recursos (normalmente sitios web) en la Web Semántica. El objetivo de RDF es habilitar la extracción del significado de la estructura de un documento, descrito en XML, con el fin de garantizar la interoperabilidad entre aplicaciones sin necesidad de intervención humana. Para que

³⁶ Ibíd.

³⁷ LINCOLN, Matthew. Uso de SPARQL para acceder a datos abiertos enlazados. [En línea]. Disponible en: <https://programminghistorian.org/es/lecciones/sparql-datos-abiertos-enlazados>

un ordenador pueda entender este tipo de estructuras, denominadas triples, será necesario representar dicha información mediante RDF/XML³⁸.

Como en otros elementos de la Web semántica, la base conceptual de RDF proviene de la lógica, pero en este caso con componentes de la lingüística. Todo el sistema RDF parte de tres entidades lógicas que son las siguientes:

- ✓ Recursos: pueden ser sitios o páginas web, pero también pueden ser cosas que no están en la Web, como personas o cualquier objeto del mundo real o conceptual.
- ✓ Propiedades: son las características relevantes de los recursos. Pueden ser propiedades de dato (atributos de objeto) o propiedades de objeto (métodos de objeto). Un ejemplo de propiedad de dato es el color de piel de una persona y un ejemplo de propiedad de objeto es la acción que realiza la persona; caminar.
- ✓ Valores: los valores son los datos en los que se concreta un atributo de un recurso determinado.

Estas tres entidades lógicas de acuerdo con Codina y Cristòfol³⁹ se corresponden respectivamente a los siguientes elementos de la lingüística:

- ✓ Sujeto: es el elemento del que se dice algo concreto.
- ✓ Predicado: es lo que se dice del sujeto.
- ✓ Objeto: es el elemento que el sujeto predica.

La tabla 1 expresa las equivalencias entre los componentes lingüísticos y lógicos basados en los componentes básicos de RDF:

³⁸ PEDRAZA, Rafael, CODINA, Lluís & CRISTÒFOL, Rovira. Web semántica y ontologías en el procesamiento de la información documental. En: El profesional de la información. Vol. 16. No. 6. (Noviembre, 2007), p. 569-579. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/13141>

³⁹ CODINA, Lluís & CRISTÒFOL, Rovira. La Web Semántica. En: Jesús Tramullas (coord.). Tendencias en documentación digital. Guijón: Trea, 2006. p. 9-54. [En línea]. Disponible en: <http://eprints.rclis.org/8899>

Tabla 1. Equivalencias entre componentes lingüísticos y lógicos en RDF

Declaración			
Término lingüístico	Sujeto	Predicado	Objeto
Término lógico	Recurso	Propiedad	Valor
Ejemplo generalizado	Jurado	Evalúa	Trabajo de grado

Fuente: Codina y Rovira⁴⁰

3.4 LENGUAJE XML

XML es un estándar que, junto con su norma asociada XML SCHEMA, permite definir tipos de documentos y los conjuntos de etiquetas necesarias para codificar tales tipos de documentos. Una vez los documentos están marcados con un conjunto de etiquetas XML es posible procesarlos y explotarlos de forma automática con diversos propósitos, de la misma manera que un conjunto de registros de una base de datos se puede explotar de formas diversas e incluso exportarse a diversos sistemas de gestión de bases de datos si la estructura de registros sigue algún tipo de estándar.

Miller y Clarke examinan las características de XML y las generalizan en tres ejes principales:

- ✓ No propiedad, interoperabilidad y neutralidad de plataforma: se refiere al hecho de que XML es un sistema abierto, no propietario e independiente de la plataforma y del sistema operativo.
- ✓ Longevidad, persistencia y reutilización: esta faceta de XML deriva directamente de la primera puesto que la información codificada en XML es independiente de la plataforma y el código fuente puede interpretarse directamente por el ojo humano, entonces su persistencia, reutilización y conservación están garantizadas para eventos futuros.
- ✓ Separación de contenido, visualización y funcionalidad: incide en uno de los aspectos más interesantes e inteligentes de XML al insistir en separar el marcado del contenido de su presentación, XML es un instrumento de primer orden para procesar información. En efecto, la información puede almacenarse en un único formato y

⁴⁰ Ibíd.

reutilizarse en tantos formatos como sea necesario, ejemplo: HTML, PDF, página web, documento impreso, etc⁴¹.

Ejemplo:

```
<Trabajo de grado>
  <Autor1>
    Felipe Cujar Rosero
  </Autor1>
  <Autor2>
    David Santiago Pinchao Ortiz
  </Autor2>
  <Asesor>
    Silvio Ricardo Timarán Pereira
  </Asesor>
<Trabajo de grado>
```

3.5 LENGUAJE OWL

De acuerdo con Pedraza, Codina y Cristòfol “Con el objeto de que las máquinas puedan realizar tareas de razonamiento útil sobre los recursos de la Web Semántica, es necesario definir un lenguaje o herramienta de descripción que permita comparar y combinar documentos o recursos con distinta estructura; es decir, que sea capaz de reconocer, por ejemplo, el elemento A y B como iguales, permitiendo la interoperabilidad entre ambos”⁴².

OWL es el lenguaje estándar de la Web Semántica para expresar y codificar Ontologías, por tanto puede ser utilizado para representar explícitamente el significado de términos en vocabularios y las relaciones semánticas entre ellos. OWL utiliza RDF para representar y codificar las Ontologías. Esta recomendación sigue la tendencia de W3C de proceder mediante “extensiones⁴³.

⁴¹ Ibíd.

⁴² PEDRAZA, Rafael, CODINA, Lluís & CRISTÒFOL, Rovira. Web semántica y ontologías en el procesamiento de la información documental. En: El profesional de la información. Vol. 16. No. 6. (Noviembre, 2007), p. 569-579. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/13141>

⁴³ Ibíd.

De acuerdo con W3C: OWL está concebido para ser utilizado cuando la información contenida en los documentos necesita ser procesada por aplicaciones informáticas, en oposición a las situaciones donde el contenido solamente debe ser presentado a seres humanos⁴⁴.

3.6 MOTORES DE BÚSQUEDA O BUSCADORES

Un motor de búsqueda es un tipo de software que crea índices de bases de datos o de sitios web en función de los títulos de los archivos, de palabras clave, o del texto completo de dichos archivos. El usuario conecta con un motor de búsqueda y especifica la palabra o las palabras clave del tema que desea buscar. El motor de búsqueda devuelve una lista de resultados presentados en hipertexto, es decir que se pueden pulsar y acceder directamente al fichero correspondiente. Un motor de búsqueda de Internet es una dirección más de la Web que ofrece al usuario las direcciones URL de otras páginas o servicios u otro tipo de información en el caso de motores de búsqueda especializados atendiendo al criterio de búsqueda que se haya seleccionado. Además, permite en la mayoría de los casos acceder a los recursos localizados mediante enlaces, facilitando así la navegación, ya que sólo es necesario conocer la dirección de los motores de búsqueda para moverse dentro de Internet⁴⁵.

Los motores de búsqueda facilitan la tarea de localizar la información que está dispersa por la Red, razón por la que son las páginas más visitadas como punto de partida para obtener cualquier información que se encuentre en Internet sobre un tema concreto. Encontrar la información deseada en la Red no es fácil, porque localizar un tema concreto puede ser muy difícil ya que existen millones de páginas web y cualquier búsqueda que se realice sin criterio será una pérdida de tiempo sin encontrar una respuesta concreta⁴⁶.

Algunos tipos de motores de búsqueda que actualmente se dispone son:

- ✓ Motores de búsqueda por palabras clave: son los más comunes, se puede introducir una palabra clave y el buscador examina su base de datos para mostrar los resultados que coincidan.

⁴⁴ CODINA, Lluís & CRISTÒFOL, Rovira. La Web Semántica. En: Jesús Tramullas (coord.). Tendencias en documentación digital. Guijón: Trea, 2006. p. 9-54. [En línea]. Disponible en: <http://eprints.rclis.org/8899>

⁴⁵ GALLO, Manuel, FABRE, Ernesto & GALLO, Manuel. ¿Qué es un buscador? [En línea]. Disponible en: http://media.axon.es/pdf/98234_1.pdf

⁴⁶ Ibíd.

- ✓ Motores de búsqueda por categorías: estos están organizados por temas, de forma que, al realizar una búsqueda determinada, muestran opciones con más respuestas, así sucesivamente hasta llegar a los contenidos que existen de un tema específico.
- ✓ Meta buscadores: también llamados motores de búsqueda múltiples, ya que realizan varias búsquedas simultáneas con otros motores de búsqueda y muestran los resultados obtenidos realizando una ponderación entre dichos motores a fin de retornar la mejor respuesta.
- ✓ Motores de búsqueda específicos: son motores de búsqueda que sólo contienen información sobre un tema concreto y se utilizan a nivel local únicamente.
- ✓ Motores de búsqueda semánticos: son motores de búsqueda que a través de Ontologías u otros procesos semánticos hacen inferencias para realizar las búsquedas por medio de agentes inteligentes de software.

3.7 SPARQL

Las bases de datos RDF son muy apropiadas para expresar relaciones complejas entre múltiples entidades como: personas, lugares, eventos y conceptos ligados a objetos individuales. Estas bases de datos se denominan habitualmente bases de datos orientadas a grafos (graph databases) porque estructuran la información como un grafo o red, donde un conjunto de recursos o nodos están conectados entre sí mediante aristas (o enlaces) que describen las relaciones establecidas entre dichos recursos y/o nodos⁴⁷.

SPARQL es el lenguaje utilizado para interrogar este tipo de bases de datos. Este lenguaje es particularmente potente por su eficiencia al momento de retornar una respuesta. Así, SPARQL es la mejor alternativa para realizar consultas a la Ontología y así poder obtener las mejores respuestas a dichas consultas⁴⁸.

3.8 SCRUM

Scrum es una metodología ágil y flexible creada para gestionar el desarrollo de software, es de recalcar que ágil se entiende como la optimización del uso de recursos y la búsqueda de la mejor forma de llevar a cabo un proyecto. Scrum es en sí un proceso en

⁴⁷ LINCOLN, Matthew. Uso de SPARQL para acceder a datos abiertos enlazados. [En línea]. Disponible en: <https://programminghistorian.org/es/lecciones/sparql-datos-abiertos-enlazados>

⁴⁸ Ibíd.

el que se aplican buenas prácticas asociadas al software que permiten trabajar colaborativamente, es decir, en equipo, para así obtener el mejor resultado posible de un proyecto.

De acuerdo con lo mencionado en el caso de estudio de Flores y Portillo⁴⁹ en Scrum se definen roles y prácticas que se ejecutarán acorde la duración del proyecto. Los roles principales son: ScrumMaster quien mantiene los procesos y trabaja como director de proyecto; ProductOwner que representa a los stakeholders (clientes externos o internos) y el Team que incluye a los desarrolladores.

Durante cada sprint, un periodo de determinados días que es la magnitud definida por el equipo, el equipo crea un incremento de software potencialmente entregable, es decir utilizable. En cada sprint se crea el Product Backlog, que es un conjunto de requisitos de alto nivel que tienen prioridad y que definen el trabajo a realizar por el Team. Luego, los elementos del Product Backlog que forman parte del sprint se determinan durante la reunión de Sprint Planning. Durante esta reunión, el Product Owner identifica los elementos del Product Backlog que quiere ver completados y los hace conocer para el Team. Durante el sprint, nadie puede cambiar el Product Backlog, lo que significa que los requisitos están congelados durante el sprint.

Scrum permite la creación de equipos auto-organizados impulsando la comunicación verbal y fluida entre todos los miembros y disciplinas involucrados en el proyecto. Un principio clave de Scrum es el reconocimiento de que durante un proyecto los clientes pueden cambiar de idea sobre lo que quieren y necesitan, o sea en cualquier momento los requisitos de alto nivel mencionados en el anterior párrafo cambian o es probable que aumenten o disminuyan, es entonces cuando los desafíos impredecibles no pueden ser fácilmente enfrentados de una forma predictiva y planificada. Por este hecho, Scrum adopta una aproximación pragmática, aceptando que el problema no puede ser completamente entendido o definido, y centrándose en maximizar la capacidad del equipo de entregar rápidamente y responder a requisitos emergentes o que surgen en momentos inesperados.

⁴⁹ FLORES, Pedro & PORTILLO, Julio. ELABORACIÓN DE PROPUESTA DE GUÍA DE IMPLEMENTACIÓN DE SCRUM PARA EMPRESA SALVADOREÑA, UN CASO DE ESTUDIO. Antiguo Cuscatlán, 2017, 117p. Trabajo de grado (MAESTRO EN ARQUITECTURA DE SOFTWARE). Universidad Don Bosco. Arquitectura de Software. [En línea]. Disponible en: <http://rd.udb.edu.sv:8080/jspui/bitstream/11715/1264/1/documento.pdf>

3.9 APRENDIZAJE DE MÁQUINA O APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

Según el científico de datos Joseph Rocca, el Machine Learning es la ciencia de hacer que las computadoras, las máquinas o equipos de cómputo realicen una tarea y descubran conocimiento sin estar programados explícitamente.

Adicionalmente, el ingeniero e investigador en IA Francois Chollet en 2017 dijo que la ingeniería tradicional se combina con las reglas creadas por los humanos y con los datos para crear respuestas a un problema, es decir soluciones. Pero el Machine Learning usa datos y respuestas para descubrir las reglas detrás de un problema⁵⁰.

El proceso de Machine Learning se puede simplificar en las siguientes etapas:

- ✓ Recolección de datos: se refiere a extraer todos los datos de los cuales el algoritmo va a aprender.
- ✓ Preparación de datos: aquí se aplica ingeniería de datos obteniendo las características más importantes del dataset, se quita lo innecesario, es decir se deja lo que sirve.
- ✓ Entrenamiento: en esta etapa se aplica uno o diversos algoritmos de Machine Learning para entrenar con los datos definidos en las etapas previas.
- ✓ Evaluación del modelo: en esta etapa se realizan las pruebas respectivas para medir la eficiencia y rendimiento del modelo.
- ✓ Retroalimentación: es la etapa que sirve para dar mantenimiento al modelo, donde se retroalimenta el mismo hasta que el resultado satisface las necesidades de negocio.

El aprendizaje que usa el Machine Learning se clasifica así:

⁵⁰ ROCCA, Joseph. A simple introduction to Machine Learning. [En línea]. Disponible en: <https://towardsdatascience.com/introduction-to-machine-learning-f41aabc55264>

- ✓ Aprendizaje supervisado: se tiene una entrada y una salida (como la mayoría de procesos en ingeniería), sólo que aquí se analiza un gran conjunto de datos que le van a servir a los algoritmos para su aprendizaje, esto es análogo a los sistemas de razonamiento basado en casos (en donde ya se tiene experiencia de soluciones de problemas previos y para los nuevos problemas se basa en esas experiencias). En este orden de ideas en el aprendizaje supervisado lo que hace es aprender con base en la experiencia (datos) que se tiene y así se realizan predicciones.

Ejemplo: se tiene una gran cantidad de datos sobre el clima y se ha aplicado todo el proceso y la ingeniería detrás de Machine Learning con aprendizaje supervisado, entonces se realizó la predicción, basado en estadística y probabilidades, de que el clima para el día 10 (diez) de octubre va a ser de 40 grados, es decir cálido. Después se aplica otro algoritmo que indica que los días cálidos en una pequeña playa de Tanzania suelen haber alrededor de 100 (cien) turistas, esto es beneficioso para los comerciantes y les lleva a fabricar 100 (cien) productos que saben con un alto porcentaje de certeza les generará grandes ganancias. Esto es a grosor modo el aprendizaje supervisado con toda la importancia y peso que conlleva para las empresas y el análisis de datos.

En este aprendizaje se encuentran algoritmos de clasificación, regresión, etc.

- ✓ Aprendizaje no supervisado: aquí pasa lo contrario al anterior aprendizaje (supervisado), porque no se tienen los datos para basarse y hacer predicciones, sino que hay que hacer el proceso inverso; se pasa al algoritmo unos datos sin relación entre sí y a partir de ese proceso el algoritmo empieza a entrenar para hallar la relación entre los mismos.

Ejemplo: un supermercado Z tiene datos aleatorios de todos los clientes que han entrado al supermercado y aplica Machine Learning con aprendizaje no supervisado para analizar a sus clientes y sus preferencias de compra, aplica técnicas de clustering para agrupar a cierto tipo de clientes y se da cuenta que a muchos clientes que les gusta la comida picante, también les gusta las hamburguesas. Entonces si el supermercado llega a confirmar con un alto porcentaje de certeza este patrón o tendencia, podrá colocar en su estante la comida picante con las hamburguesas y generará grandes ganancias.

Aquí en este aprendizaje se encuentran algoritmos de clustering, asociación, detección de anomalías, reducción dimensional, etc.

- ✓ Aprendizaje semi-supervisado: en este aprendizaje se combinan los dos anteriores tipos de aprendizaje (supervisado y no supervisado).

Se encuentran las redes adversarias generativas; dos redes neuronales, las cuales compiten entre sí, luego la que gane será la que “tenga la razón”, porque la salida de sus capas es más convincente de acuerdo al contexto y así aprende.

Se tiene mucha aplicabilidad en el área de reconocimiento, procesamiento y edición de imágenes, pues se manejan muchos vectores y fórmulas.

- ✓ Aprendizaje reforzado: este tipo de aprendizaje va muy de la mano con conceptos del ambiente de la psicología, los cuales dictan que cuando una persona realiza una actividad bien hecha, en su cerebro se liberan sustancias “de aprobación” de dichos actos como: la serotonina, dopamina, entre otras. En cambio cuando la persona no realiza bien su trabajo, o comete una acción ilegal o inmoral, entonces normalmente se liberan sustancias de “desaprobación” que causan la huida como: la adrenalina, noradrenalina, cortisol, entre otros. La relación aquí con el aprendizaje es que se tiene algoritmos que cuando se realiza un trabajo bien hecho (una predicción) se empieza a premiar, sino se castiga. Esto deriva en una situación en la que el algoritmo será como un niño que se va a ir por los premios y evitará los castigos.

Entre los ejemplos a destacar, se encuentran muchos video-juegos. Google DeepMind usó el aprendizaje reforzado para que en juegos como Go y Atari la máquina adquiera “poderes superhumanos”, es decir, que gane siempre⁵¹.

3.10 PROCESAMIENTO DE LENGUAJE NATURAL (NATURAL LANGUAGE PROCESSING)

El PLN, según el científico de datos Badreesh Shetty, es el campo del Machine Learning que permite que una máquina pueda entender, analizar, manipular y potencialmente generar lenguaje humano⁵².

Las principales etapas y técnicas de PLN son:

⁵¹ EDWARDS, Gavin. Machine Learning An Introduction. [En línea]. Disponible en: <https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0>

⁵² SHETTY, Badreesh. Natural Language Processing (NLP) for Machine Learning. [En línea]. Disponible en: <https://towardsdatascience.com/natural-language-processing-nlp-for-machine-learning-d44498845d5b>

- ✓ Eliminación de signos de puntuación: se eliminan los signos de puntuación y otras expresiones que no significan nada para el dataset (conjunto de datos).
- ✓ Eliminación de stopwords: son palabras sin significado que en una oración tienen sentido, al igual que los signos de puntuación, pero para el dataset preposiciones, artículos y otro tipo de conectores no sirven y por eso se eliminan.
- ✓ Stemming: se trata de la eliminación de sufijos de la palabra. Es una forma de limpieza de la palabra con el fin de trabajar de manera estándar, ordenada y ahorrando almacenamiento en memoria.
- ✓ Lematización: es la alternativa a stemming, en donde también existe una limpieza de la palabra, pero la lematización es más exacta al momento de obtener la raíz de la palabra. Un ejemplo es: la raíz de las palabras “aprendiendo, aprenderás, aprenderán” es “aprender”. La lematización usa diccionarios morfológicos y es mucho más exacta, eficiente y coherente que el stemming, aunque el proceso demora más.
- ✓ N-gramas: es básicamente una combinación de palabras, de n cantidad; unigrama sería una palabra, bigrama serían dos palabras y así sucesivamente. Esto se realiza para contemplar la posibilidad de trabajar con más de una palabra (sin necesariamente llegar a formar oraciones pero sí que tengan sentido) y ver qué tan eficiente resulta.
- ✓ Tokenización: se trata de separar las palabras dándole así una estructura al dataset. Esto servirá de insumo para una fase posterior que se llama vectorización.
- ✓ Vectorización: aquí se convierten las palabras tokenizadas a datos numéricos y es este tipo de dato que los algoritmos de Machine Learning pueden entenderlo dada su matemática y estadística.

Entre los algoritmos de PLN, se destacan los siguientes:

- ✓ BOW (Bag Of Words): bolsa de palabras en español, es uno de los importantes algoritmos usados para PLN, porque cuenta las veces que está el término en un documento, esto da una idea acerca de los términos, qué tanta importancia tienen, cual es la frecuencia de cada uno, qué tan grande será el dataset, etc.

- ✓ TF-IDF (Term Frequency – Inverse Document Frequency): es de los algoritmos más usados por su eficiencia, encuentra el peso del término en todo el corpus; determina qué tan importante es el término dentro de todo el conjunto de documentos. El peso de un término t en un documento d_1 es más grande, es decir más importante si aparece muchas veces (si son todas mejor para el peso) en el menor número de documentos posible (si es uno solo, mejor).
- ✓ Word2Vec: el algoritmo Word2Vec utiliza una red neuronal para aprender asociaciones de palabras de un gran corpus de texto. Una vez entrenado, dicho modelo puede detectar sinónimos entre las palabras o sugerir palabras adicionales que tengan relación semántica. Como su nombre lo indica, Word2Vec representa cada palabra distinta con una lista particular de números llamada vector. Los vectores se eligen cuidadosamente de modo que una función matemática (similitud coseno entre los vectores) indica el nivel de similitud semántica entre las palabras representadas por esos vectores.

Word2Vec puede utilizar cualquiera de las dos arquitecturas modelo para producir una representación distribuida de palabras: bolsa de palabras continua (CBOW) y skip-gram. En la arquitectura de CBOW, el modelo predice la palabra actual desde un conjunto de palabras con contexto. El orden de las palabras con contexto no influye en la predicción. En la arquitectura de skip-gram, el modelo utiliza la palabra actual para predecir el conjunto de palabras con contexto y pesan más las palabras con contexto cercanas que las palabras con contexto más distantes. CBOW es más rápido, mientras que skip-gram es más lento, pero funciona mejor con palabras poco frecuentes⁵³.

- ✓ Doc2Vec: el algoritmo Doc2Vec está basado en el algoritmo Word2Vec (previamente mencionado) y su objetivo es crear una representación numérica de un documento, independientemente de su longitud. Pero a diferencia de las palabras, los documentos no vienen en estructuras lógicas, por lo que en su momento se pensó que lo mejor era usar el algoritmo Word2Vec y agregar otro vector denominado Paragraph ID, el cual es usado para identificar de manera única a los documentos o secciones del mismo.

Un esquema usado en Doc2Vec se denomina Paragraph Vector - Distributed Memory (PV-DM), el cual es análogo al algoritmo de CBOW en Word2Vec con la adición del Paragraph ID.

⁵³ SHPERBER, Gidi. A gentle introduction to Doc2Vec. [En línea]. Disponible en: <https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e>

El otro esquema se denomina Paragraph Vector - Distributed Bag of Words (PV-DBOW), el cual es análogo al algoritmo de skip-gram de Word2Vec pero con la diferencia de que incuye el Paragraph ID.

Así las cosas, mientras que los vectores de palabras (en Word2Vec) representan el concepto de una palabra, los vectores de documentos (en Doc2Vec) representan el concepto de un documento⁵⁴.

- ✓ NLTK: es una plataforma para crear programas Python que funcionen con datos de lenguaje humano. Proporciona interfaces fáciles de usar para bibliotecas de procesamiento de texto para clasificación, tokenización, derivación, etiquetado, análisis y razonamiento semántico, envoltorios para bibliotecas de PNL de nivel industrial, y un activo foro de discusión.

NLTK es conocido como “una herramienta maravillosa para enseñar y trabajar en lingüística computacional usando Python” y “una biblioteca increíble para jugar con el lenguaje natural”⁵⁵.

- ✓ SpaCy: es una biblioteca gratuita de código abierto para avanzado Procesamiento de Lenguaje Natural en Python. Está diseñado específicamente para uso en producción y ayuda a construir aplicaciones que procesan y “comprenden” grandes volúmenes de texto. Puede ser usado para construir la extracción de información o la comprensión del lenguaje natural sistemas o para preprocesar texto para el aprendizaje profundo⁵⁶.

La aplicación de PLN en la vida real se relaciona con:

- ✓ Extracción de la información: se trata de obtener la información o los datasets (conjuntos de datos) por cualquier medio sea físico o digital.
- ✓ Recuperación de la información: esta aplicación se da al momento de buscar determinada información obteniendo una respuesta coherente y concreta a dicha búsqueda.

⁵⁴ BUDHIRAJA, Amar. A simple explanation of document embeddings generated using Doc2Vec. [En línea]. Disponible en: <https://medium.com/@amarbudhiraja/understanding-document-embeddings-of-doc2vec-bfe7237a26da>

⁵⁵ Kit de herramientas de lenguaje natural. [En línea]. Disponible en: <https://www.nltk.org>

⁵⁶ SpaCy 101: todo lo que necesita saber. [En línea]. Disponible en: <https://spacy.io/usage/spacy-101>

- ✓ Traducción: es una de las funciones más usadas en el PLN, que ocurre al momento de usar traductores disponibles en cualquier medio con el fin de traducir una o múltiples palabras.
- ✓ Síntesis de texto: se trata de una reducción significativa del texto, basado en un análisis previo.
- ✓ Análisis de sentimiento: observando el caso de Twitter cuando se analizaba qué usuarios estaban tristes o felices dada su escritura (cabe destacar lo difícil que fue y sigue siendo el análisis de la ironía o los sarcasmos en los escritos de los usuarios).
- ✓ Filtro de spam: empresas de correos electrónicos usan el PLN para dar manejo al spam, que no es más que un tipo de correo con fines maliciosos.
- ✓ Predicción: búsquedas con algoritmos de predicción, tal y como lo hace Google, cuando se realiza una búsqueda usualmente se predice lo que el usuario va a buscar.
- ✓ Auto-correctores: a modo de ejemplo está Google Keyboard, que básicamente corrige la gramática y sintaxis del texto del usuario.
- ✓ Respuesta a preguntas: IBM Watson es un claro ejemplo de respuestas a determinadas preguntas y el uso de “diálogo” humano-máquina.
- ✓ Expresión de lenguaje natural: a partir del texto de una imagen o los datos de un video, se puede generar el lenguaje natural contenido⁵⁷.

3.11 JUPYTER NOTEBOOK

Jupyter Notebook es una aplicación web de código abierto que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Los usos incluyen: limpieza y transformación de datos, simulación numérica, modelado estadístico, visualización de datos, aprendizaje automático, etc⁵⁸.

⁵⁷ SHETTY, Badreesh. Natural Language Processing (NLP) for Machine Learning. [En línea]. Disponible en: <https://towardsdatascience.com/natural-language-processing-nlp-for-machine-learning-d44498845d5b>

⁵⁸ PEREZ, Fernando & GRANGER, Brian. Project Jupyter. [En línea]. Disponible en: <https://jupyter.org>.

3.12 PROTÉGÉ

Protégé es un software libre de código abierto implementado en Java, desarrollado en la Universidad de Stanford, que permite la construcción de Ontologías de dominio. Es capaz de operar como una plataforma para acceder a otros sistemas basados en conocimiento o aplicaciones integradas, o como una librería que puede ser usada por otras aplicaciones para acceder y visualizar bases de conocimiento. La herramienta ofrece una interfaz gráfica que permite al desarrollador de Ontologías enfocarse en la modelación conceptual sin que requiera de conocimientos de la sintaxis de los lenguajes de salida⁵⁹.

El modelo de Protégé está basado en un metamodelo de sistemas orientados a objetos y sistemas basados en marcos. Este metamodelo básicamente puede representar Ontologías que consisten en clases, propiedades e instancias. La estructura del metamodelo de Protégé habilita fácilmente la extensión y adaptación a otras representaciones.

El plugin OWL es una extensión compleja que puede ser usada para editar archivos OWL y bases de datos. Este y otros plugins incluyen una colección de características para los componentes denominados widgets que proveen acceso a servicios basados en clasificación, chequeo consistente y pruebas.

Protégé es una herramienta que hace posible la descripción de conceptos y provee otras facilidades, por ejemplo: tiene un enriquecido conjunto de operadores (and, or y la negación); está basado sobre un modelo lógico que garantiza que los conceptos estén bien descritos; se pueden describir conceptos complejos y el modelo lógico permite usar un razonador que chequea la consistencia descrita en la Ontología garantizando que la jerarquía esté correcta⁶⁰.

3.13 OWLREADY2

Owlready2 es un módulo para programación orientada a Ontologías en Python 3, que incluye un modelo basado en triples (quadstore) RDF optimizado. Owlready2 fue creado

⁵⁹ TABARES, John & JIMÉNEZ, Jovani. Ontología para el proceso evaluativo en la educación superior. En: Revista Virtual Universidad Católica del Norte. Vol. 1. No. 42. (Agosto, 2014), p. 68-79. [En línea]. Disponible en: <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/495>

⁶⁰ FLÓREZ, Héctor. Construcción de ontologías OWL. En: VÍNCULOS. Vol. 4. No. 1. (Diciembre, 2007), p. 19-34. [En línea]. Disponible en: <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/4112>

por Jean-Baptiste Lamy en el laboratorio de investigación LIMICS. Está disponible bajo la licencia GNU GPL v3⁶¹.

Owlready2 tiene las siguientes características:

- ✓ Importa Ontologías OWL 2.0 en formato NTriples, RDF / XML u OWL / XML.
- ✓ Exporta Ontologías OWL 2.0 a NTriples o RDF / XML.
- ✓ Manipula clases, instancias y propiedades de la Ontología de forma transparente como si fueran objetos normales de Python.
- ✓ Agrega métodos de Python a las clases de Ontología.
- ✓ Realiza una clasificación automática de clases e instancias utilizando el razonador HermiT o Pellet.
- ✓ Soporta hasta mil millones de triples de RDF (pero potencialmente puede soportar más).
- ✓ Además, el modelo basado en triples (quadstore) es compatible con el módulo RDFLib Python, que se puede utilizar para realizar consultas SPARQL.
- ✓ Owlready2 también se puede utilizar como ORM (Mapeador Relacional de Objetos), teniendo en cuenta que como base de datos gráficos / objetos supera a Neo4j, MongoDB, SQLObject y SQLAlchemy en términos de rendimiento.

3.14 APACHE JENA FUSEKI (AJF)

AJF puede funcionar como sistema operativo, como una aplicación web Java (archivo WAR) o como un servidor independiente. AJF está disponible en dos formas: la primera

⁶¹ LAMY, Jean. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. En: Artificial Intelligence in Medicine. Vol. 80. (Agosto, 2017), p. 11-28. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0933365717300271>

es un único sistema que funciona tipo aplicación web, combinado con una interfaz de usuario para administradores y consultas; la segunda es un servidor principal que está adecuado para la fase de despliegue, incluso se lo puede trabajar con Docker o en ejecución integrada. Ambas formas utilizan el mismo protocolo central y la misma configuración de formato de archivo⁶².

AJF proporciona el servicio para consulta y actualización SPARQL, así como el protocolo de almacenamiento de grafos SPARQL. Está estrechamente integrado con TDB (el cual es un componente de Jena) para proporcionar una capa de almacenamiento robusta, transaccional y persistente que a su vez incorpora la herramienta de consulta de texto de Jena.

3.15 ELASTICSEARCH

Elasticsearch es un motor de analítica y análisis distribuido, y de código abierto para todos los tipos de datos, incluidos textuales, numéricos, geoespaciales, estructurados y desestructurados. Elasticsearch está desarrollado en Apache Lucene y fue presentado por primera vez en 2010 por Elasticsearch N.V. y ahora es conocido como Elastic.

La velocidad y escalabilidad de Elasticsearch y su capacidad de indexar muchos tipos de contenido significan que puede usarse para una variedad de casos de uso como los siguientes:

- ✓ Búsqueda de aplicaciones
- ✓ Búsqueda de sitio web
- ✓ Búsqueda Empresarial
- ✓ Logging y analíticas de log
- ✓ Métricas de infraestructura y monitoreo de contenedores

⁶² Apache Software Foundation. Apache Jena Fuseki. [En línea]. Disponible en: <https://jena.apache.org/documentation/fuseki2>

- ✓ Monitoreo de rendimiento de aplicaciones
- ✓ Análisis y visualización de datos geoespaciales
- ✓ Analítica de Seguridad
- ✓ Analítica de Negocios

Como Elasticsearch está desarrollado sobre Lucene, es excelente en la búsqueda de texto completo. Elasticsearch también es una plataforma de búsqueda en casi tiempo real, lo que implica que la latencia entre el momento en que se indexa un documento hasta el momento en que se puede buscar en él es muy breve, normalmente puede ser un segundo. Como resultado, Elasticsearch está bien preparado para casos de uso con restricciones de tiempo como analítica de seguridad y monitoreo de infraestructura.

Elasticsearch es distribuido por naturaleza. Los documentos almacenados en Elasticsearch se distribuyen en distintos contenedores conocidos como shards, que están duplicados para brindar copias redundantes de los datos en caso de que falle el hardware. La naturaleza distribuida de Elasticsearch le permite escalar horizontalmente a cientos (o incluso miles) de servidores y gestionar petabytes de datos.

Además de su velocidad, la escalabilidad y la resistencia, Elasticsearch tiene una cantidad de características integradas muy útiles que contribuyen a que el almacenamiento y la búsqueda de datos sean incluso más eficientes, como data rollup y gestión de ciclo de vida del índice⁶³.

3.16 DOCKER

Docker es una plataforma abierta para que desarrolladores y administradores de sistemas desarrollen, envíen y ejecuten aplicaciones distribuidas, ya sea en computadoras portátiles, máquinas virtuales de centros de datos o en la nube⁶⁴.

⁶³ Elasticsearch B.V. Elasticsearch. [En línea]. Disponible en: <https://www.elastic.co/es/what-is/elasticsearch>

⁶⁴ ARAUJO, Joaquín. ¿Qué es Docker? ¿Qué son los contenedores? y ¿Por qué no usar VMs? [En línea]. Disponible en: <https://platzi.com/tutoriales/1432-docker/1484-guia-del-curso-de-docker>

Docker empaqueta software en “contenedores” que incluyen en ellos todo lo necesario para que dicho software se ejecute, incluidas librerías. Con Docker se puede implementar y ajustar la escala de aplicaciones de una forma rápida en cualquier entorno con la garantía de que el código se ejecutará.

A primera vista se piensa en Docker como una especie de máquina virtual liviana, pero no es así. En Docker no existe un hypervisor que virtualice hardware sobre el cual corra un sistema operativo completo. En Docker lo que se hace es usar las funcionalidades del kernel para encapsular un sistema, de esta forma el proyecto que corre dentro de él no tendrá conocimiento que está en un contenedor. Los contenedores se encuentran aislados entre sí y se comportan como máquinas independientes. Gracias al uso de contenedores la demanda de recursos baja limitándose sólo al consumo de la aplicación que contenga. Un contenedor inicia en milisegundos.

Docker trabaja con algo que se llama “contenedores de Linux”, estos son un conjunto de tecnologías que juntas forman un contenedor (de Docker) y se clasifican así:

- ✓ Namespaces: permite a la aplicación que corre en un contenedor de Docker tener una vista de los recursos del sistema operativo.
- ✓ Cgroups: permite limitar y medir los recursos que se encuentran disponibles en el sistema operativo.
- ✓ Chroot: permite tener en el contenedor una vista de un sistema “falso” para el mismo, es decir, crea su propio entorno de ejecución con su propio root y home.

Algunas de las características más notables de un contenedor son:

- ✓ Los contenedores son más livianos (ya que trabajan directamente sobre el kernel) que las máquinas virtuales.
- ✓ No es necesario instalar un sistema operativo por contenedor.
- ✓ Menor uso de los recursos de la máquina.
- ✓ Mayor cantidad de contenedores por equipo físico.

- ✓ Mejor portabilidad.

3.17 FLASK

Flask es un micro framework escrito en Python y concebido para facilitar el desarrollo de aplicaciones⁶⁵.

Entre las razones para usar flask en el presente proyecto se encuentra:

- ✓ Al ser un micro framework se enfoca en proporcionar lo necesario para permitir el funcionamiento de cualquier tipo de aplicación en muy poco tiempo.
- ✓ Incluye un servidor web de desarrollo para probar las aplicaciones sin tener que instalar un servidor web por aparte.
- ✓ Trae un depurador y soporte integrado para pruebas unitarias.
- ✓ Es compatible con Python 3, por lo tanto se puede usar la codificación de caracteres Unicode, además es 100% compatible con el estándar WSGI.
- ✓ Tiene un buen manejo de rutas porque con el uso de un decorador de Python se puede simplificar muchos procesos en este manejo de rutas en la aplicación.
- ✓ Flask soporta el uso de cookies seguras y el uso de sesiones.
- ✓ Flask se apoya en el motor de plantillas Jinja2 que permite de forma sencilla renderizar vistas y respuestas.
- ✓ Este framework resulta ideal para construir servicios web (como APIs REST) o aplicaciones de contenido estático.

⁶⁵ MUÑOZ, José. Introducción a flask. [En línea]. Disponible en: <https://plataforma.josedomingo.org/pledin/cursos/flask/curso/u05/>

- ✓ Flask es Open Source y está amparado bajo una licencia BSD.

3.18 PYTHON

Es un lenguaje de programación interpretado de alto nivel que contiene implícitas algunas estructuras de datos como listas, diccionarios, conjuntos y tuplas, que permiten realizar algunas tareas complejas en pocas líneas de código y de manera legible⁶⁶.

Python cuenta con decenas de módulos que cubre la mayoría de las necesidades básicas de un programador. En esta se le da cobertura de forma muy intuitiva a tópicos como:

- ✓ Cadenas.
- ✓ Estructura de datos.
- ✓ Funciones numéricas y matemáticas.
- ✓ Compresión de datos.
- ✓ Formatos de archivo.
- ✓ Criptografía.
- ✓ Servicios de los sistemas operativos.
- ✓ Comunicación entre procesos.
- ✓ Manejo de datos de Internet.
- ✓ Servicios multimedia.

⁶⁶ CHALLENGER, Ivet, DÍAZ, Yanet & BECERRA, Roberto. El lenguaje de programación Python. En: Ciencias Holguín. Vol. XX. No. 2. (Junio, 2014), p. 1-13. [En línea]. Disponible en: www.redalyc.org/articulo.oa?id=181531232001

- ✓ Manejo de excepciones.

Existe una lista de principios que se deberían seguir cuando se escribe código en Python:

- ✓ Hermoso es mejor que feo.
- ✓ Explícito es mejor que implícito.
- ✓ Simple es mejor que complejo.
- ✓ Plano es mejor que anidado.
- ✓ Disperso es mejor que denso.
- ✓ El código legible cuenta.
- ✓ Casos especiales no son lo suficientemente especiales para romper las reglas.
- ✓ Casi siempre lo práctico vence a lo formal.
- ✓ Los errores no deben pasar nunca desapercibidos, a menos que se especifique este comportamiento.
- ✓ Ante una ambigüedad, es preferible descartar la tentación a adivinar.
- ✓ Debe haber una, y preferentemente una sola, manera obvia de lograr algo, aunque esta generalmente no siempre está clara a primera vista a menos de ser un genio.
- ✓ Ahora es mejor que nunca, aunque en muchas ocasiones nunca es mejor que ahora mismo.

- ✓ Los espacios de nombre son una buena idea.
- ✓ Si la implementación es difícil de explicar, entonces es una mala idea.
- ✓ Si la implementación es fácil de explicar, entonces puede ser una buena idea.

3.19 METHONTOLOGY

De acuerdo con García⁶⁷, Methontology es una metodología desarrollada en el laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid. Permite la construcción de Ontologías a nivel de conocimiento e incluye la identificación del proceso de desarrollo de la Ontología, el cual es un ciclo de vida basado en la evolución de prototipos y técnicas particulares a usar en cada paso.

Las bases de dicha metodología son actividades del desarrollo de software propuesto por la organización IEEE y algunas otras metodologías de conocimientos. El objetivo de la metodología es crear una Ontología desde unos pocos requisitos iniciales, partiendo desde lo más general hasta llegar a lo más específico⁶⁸.

En Methontology la transición por cada fase en que se puede encontrar una Ontología durante su proceso de construcción se denomina “ciclo de vida de la Ontología”.

A continuación se describen de manera general cada una de las fases del ciclo de vida:

- ✓ Especificación: fase en la que se determina la necesidad, el tipo de formalidad, los usuarios finales y la forma de representación en lenguaje formal o informal.

⁶⁷ GARCÍA, Francisco. Web Semántica y Ontologías. [En línea]. Disponible en: https://www.researchgate.net/publication/267222548_Web_Semantica_y_Ontologias

⁶⁸ CHECA, Diego & ROJAS, Oscar. ONTOLOGÍA PARA LOS SISTEMAS HOLÓNICOS DE MANUFACTURA BASADOS EN LA UNIDAD DE PRODUCCIÓN. En: Revista Colombiana de Tecnologías de Avanzada. Vol. 1. No. 23. (Noviembre, 2013), p. 134-141. [En línea]. Disponible en: http://revistas.unipamplona.edu.co/ojs_viceinves/index.php/RCTA/article/view/2334

- ✓ Conceptualización: durante esta fase se realizan descripciones, tabulaciones y gráficas (si es necesario) para representar de manera conceptual todo el conocimiento del dominio.
- ✓ Formalización: fase en la cual se usa una herramienta de descripción o un framework para la especificación de la Ontología en términos semi-computables.
- ✓ Implementación: uso de diferentes lenguajes formales de implementación de Ontologías para crear una versión computable de la misma.
- ✓ Evaluación: comprobación de que los conceptos, sus relaciones, instancias y demás elementos son coherentes en toda la Ontología.

Cabe destacar que dentro de la fase de conceptualización en Methontology, se desarrollan once tareas puntuales que son⁶⁹:

- ✓ Tarea 1. Construir el glosario de términos: se construye una lista o un glosario con todos los términos que puedan considerarse importantes para el dominio de la Ontología.
- ✓ Tarea 2. Construir una taxonomía de conceptos: se realiza la taxonomía de conceptos para la Ontología, en la cual se define una jerarquía de dichos conceptos; dicha taxonomía se elabora a partir del glosario de términos construido en la tarea 1.
- ✓ Tarea 3. Construir un diagrama de relaciones binarias: se elabora un diagrama donde se establecen las diferentes relaciones entre los conceptos especificados en la taxonomía.
- ✓ Tarea 4. Construir el diccionario de conceptos: se define un diccionario de conceptos que contiene a los conceptos o clases más importantes del dominio, así como sus atributos, relaciones e instancias.
- ✓ Tarea 5. Definir las relaciones binarias en detalle: se observan las relaciones de manera detallada mediante una tabla donde se listan todas las relaciones binarias

⁶⁹ TABARES, John & JIMÉNEZ, Jovani. Ontología para el proceso evaluativo en la educación superior. En: Revista Virtual Universidad Católica del Norte. Vol. 1. No. 42. (Agosto, 2014), p. 68-79. [En línea]. Disponible en: <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/495>

existentes en la Ontología. Para cada relación binaria se especifica el nombre, conceptos origen y destino (que refieren al dominio y rango de la clase respectivamente), relación inversa y cardinalidad.

- ✓ Tarea 6. Definir los atributos de instancias en detalle: dentro de la Ontología se definen atributos (también denominados data properties) que describen a cada objeto de determinada clase; estos atributos son los atributos de instancia. Se representan en una tabla con el nombre de atributo, la clase a la que pertenecen, el tipo de dato y la cardinalidad.
- ✓ Tarea 7. Definir los atributos de clases: los atributos de las clases son las características que van a tener cada una de las clases de la Ontología. Denotan restricciones para elementos de algunas clases.
- ✓ Tarea 8. Definir en detalle las constantes: las constantes se definen utilizando una tabla donde para cada constante se especifica el nombre, tipo de valor, valor y unidad de medida (para constantes numéricas).
- ✓ Tarea 9. Definir los axiomas formales: los axiomas necesarios en la Ontología se describen con precisión en una tabla. Para cada definición de un axioma formal se especifica el nombre, descripción, expresión lógica que lo describe, los conceptos, atributos y relaciones binarias a los cuales el axioma hace referencia para el dominio determinado.
- ✓ Tarea 10. Definir las reglas: se definen en una tabla las reglas. Para cada regla se especifica el nombre, descripción, conceptos a los que hace referencia.
- ✓ Tarea 11. Definir las instancias: finalmente en esta tarea se definen las instancias que han sido identificadas para cada clase o concepto en la Ontología.

3.20 POSTGRESQL

Es un sistema de gestión de bases de datos relacionales orientado a objetos y de código abierto.

Entre sus principales características para ser usado en el proyecto se encuentran⁷⁰:

- ✓ Tipos de datos
 - Primitivos: entero, numérico, cadena, booleano
 - Estructurado: fecha / hora, matriz, rango, UUID
 - Documento: JSON / JSONB, XML
 - Geometría: punto, línea, círculo, polígono
 - Personalizaciones: compuestos, tipos de datos personalizados
- ✓ Integridad de los datos
 - Llaves primarias
 - Llaves foráneas
 - Restricciones múltiples
- ✓ Simultaneidad, rendimiento
 - Indexación: árbol B, multicolumna, expresiones, parcial
 - Transacciones, transacciones anidadas (a través de savepoints)
 - Paralelización de consultas de lectura y construcción de índices de árbol B

⁷⁰ PostgreSQL. [En línea]. Disponible en <https://www.postgresql.org/about/>

✓ Fiabilidad y tolerancia a fallos

- Registro de escritura anticipada
- Replicación: asíncrona, síncrona y lógica
- Recuperación en un punto de tiempo, espera activa

✓ Seguridad

- Autenticación: GSSAPI, SSPI, LDAP, SCRAM-SHA-256
- Sistema de control de acceso robusto
- Seguridad a nivel de columna y fila
- Autenticación multifactor con certificados

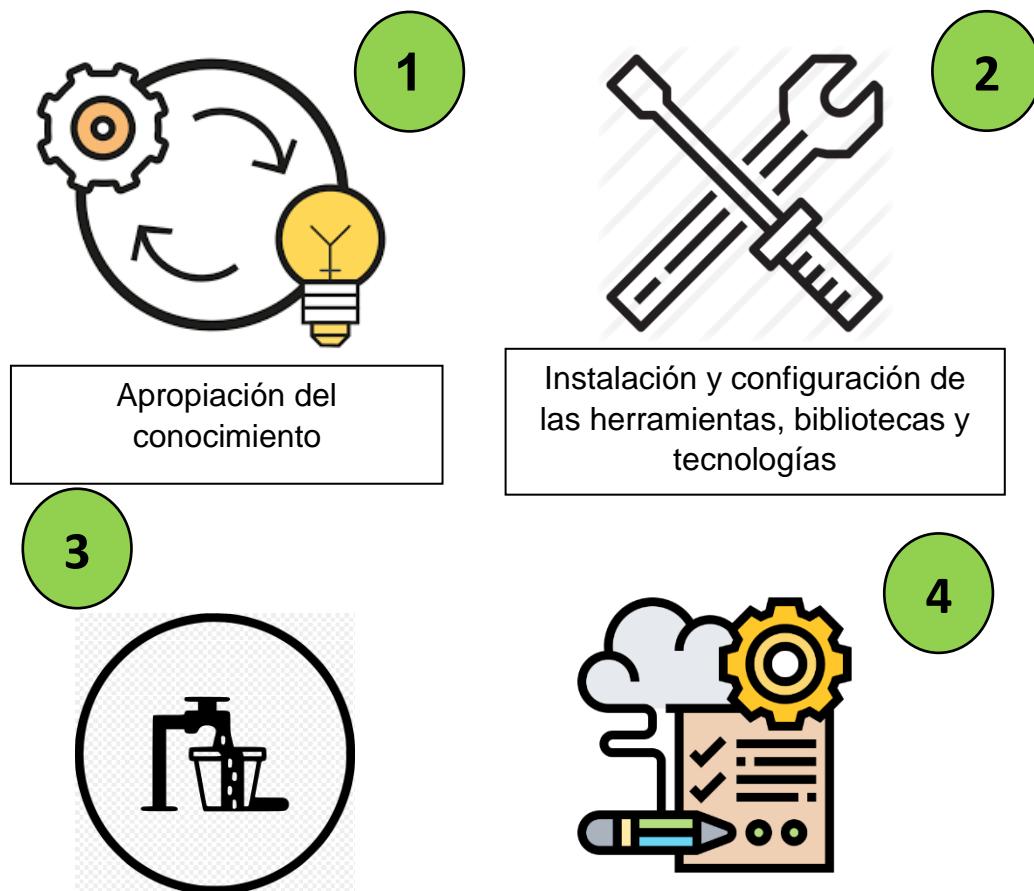
✓ Extensibilidad

- Funciones y procedimientos almacenados
- Lenguajes de procedimiento: PL / PGSQL, Perl, Python y otros
- Expresiones de ruta SQL / JSON
- Interfaz de almacenamiento personalizable para tablas
- Muchas extensiones que brindan funcionalidades adicionales como PostGIS

4. METODOLOGÍA

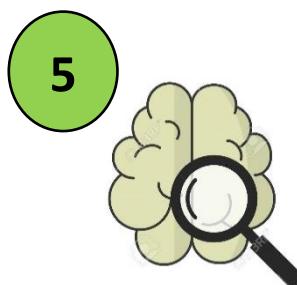
La metodología que se empleó para el desarrollo del proyecto comprendió las siguientes etapas: apropiación del conocimiento; instalación y configuración de las herramientas, bibliotecas y tecnologías; recolección y extracción de proyectos de investigación; preparación de proyectos de investigación; exploración y análisis de proyectos de investigación; diseño del motor de búsqueda semántico; desarrollo del motor de búsqueda semántico, implementación y despliegue del motor de búsqueda semántico; realización de pruebas finales y evaluación del rendimiento del motor de búsqueda semántico.

El modelo visual obtenido fue el siguiente:



Recolección y extracción
de proyectos de
investigación

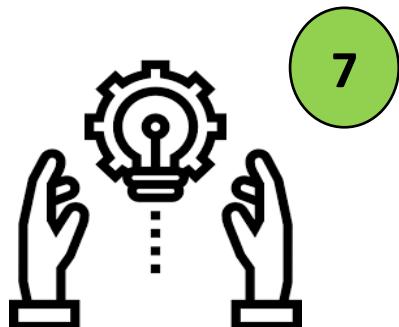
Preparación de proyectos
de investigación



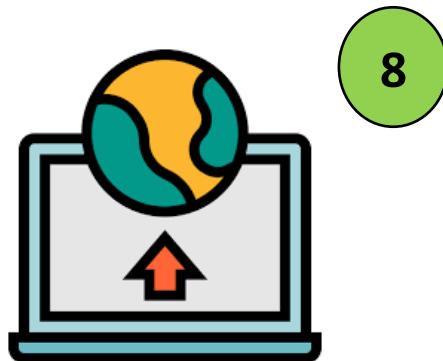
Exploración y análisis de
proyectos de investigación



Diseño del motor de
búsqueda semántico



Desarrollo del motor de
búsqueda semántico



Implementación y
despliegue del motor de
búsqueda semántico



9

Realización de pruebas finales y evaluación del rendimiento del motor de búsqueda semántico

4.1 APROPIACIÓN DEL CONOCIMIENTO

En esta etapa se entiende y comprende el problema para poder definirlo y delimitarlo, establecer unos objetivos acompañados de unos pasos claros y concretos siguiendo una metodología en específico, buscando siempre soluciones para tener un desarrollo exitoso del motor de búsqueda semántico. Así mismo se adquiere conocimiento de todas las temáticas que abarcan el proyecto, así como también se aprende a manejar las diversas herramientas y lenguajes usados.

Para todo esto se realiza una revisión exhaustiva de la literatura, el estado del arte y todas las bases sólidas teóricas y prácticas, sujetas a la experiencia, que componen la armonía de la investigación.

4.2 INSTALACIÓN Y CONFIGURACIÓN DE LAS HERRAMIENTAS, BIBLIOTECAS Y TECNOLOGÍAS

En esta etapa se realiza la instalación y configuración de las herramientas, bibliotecas y tecnologías mencionadas en el marco teórico, las cuales se han visto pertinentes y factibles de usar por razones como: rendimiento, facilidad de manejo, eficacia y eficiencia en las tareas o procesos correspondientes. Se destaca que estas herramientas, bibliotecas y tecnologías son de uso abierto para el público, fueron descargadas de páginas oficiales y su configuración se ciñe a los principios que dictan su documentación.

4.3 RECOLECCIÓN Y EXTRACCIÓN DE PROYECTOS DE INVESTIGACIÓN

En esta etapa se recolecta y se extrae toda la información de los proyectos de investigación que incluyen: proyectos docentes, proyectos estudiantiles y trabajos de grado que se encuentran almacenados en el sistema de investigaciones de la Universidad de Nariño.

Toda esta información se recolecta y extrae mediante la comunicación con el ingeniero encargado del sistema de investigaciones de la Universidad de Nariño y de otras fuentes web que se encuentran dentro del dominio de la Universidad de Nariño.

4.4 PREPARACIÓN DE PROYECTOS DE INVESTIGACIÓN

Esta etapa de preparación de proyectos de investigación es fundamental porque es aquí donde se hace un amplio recorrido de los proyectos de investigación, se avanza por cada uno cumpliendo una serie de fases que van a permitir obtener el núcleo de la semántica. En este orden de ideas, se va a quitar lo innecesario, se va a dejar lo que sirve, lo que es útil y de vital importancia para la siguiente etapa.

En esta etapa se completan las fases siguientes: organización de datos, creación de corpus y preprocesamiento de datos, haciendo uso de las bibliotecas de NLTK y SpaCy; esta fase de preprocesamiento se divide en las subfases de: tokenización, normalización, limpieza (en esta subfase se eliminan datos nulos, signos de puntuación, caracteres no ascii y stopwords) y lematización de los datos.

En esta etapa se hace uso de Jupyter notebook con scripts de Python y se utiliza Pandas para facilitar el manejo de los datos en las series y data frames.

4.5 EXPLORACIÓN Y ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN

En esta etapa se realiza la exploración y el análisis de todos los proyectos de investigación, con el fin de tener una idea general sobre lo que tratan estos, realizar breves estadísticas, de este modo analizar los datos generando ciertas afirmaciones para poder continuar con la siguiente fase en donde se realiza el diseño del motor de búsqueda semántico.

Para cumplir con el acometido se usan los algoritmos de BOW (Bag Of Words) y TF-IDF (Term Frequency – Inverse Document Frequency) y se realizan diversos histogramas y nubes de palabras.

Esta etapa se realiza dentro del contexto de la aplicación de Jupyter notebook con técnicas de manejo de data frame con la ayuda de bibliotecas como Pandas y Numpy.

4.6 DISEÑO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

En esta etapa se diseña el funcionamiento del motor de búsqueda semántico haciendo uso de la metodología Methontology con sus fases de especificación y conceptualización. Se aclara que las otras fases de Methontology se ven reflejadas en la siguiente etapa que corresponde al desarrollo del motor de búsqueda semántico.

Se construye un glosario de términos, se construye la taxonomía de conceptos, se construye un diagrama de relaciones binarias, se construye un diccionario de conceptos, se definen las relaciones binarias, se definen los atributos de instancia, se definen los atributos de clase, se observa si es o no necesario usar constantes, axiomas y reglas, y se definen finalmente las instancias.

De este modo se definen los principales componentes ontológicos como son: clases, relaciones, atributos e instancias.

4.7 DESARROLLO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

En esta etapa se desarrolla el motor de búsqueda semántico, siguiendo una adaptación de la metodología ágil Scrum basado en tres fases:

En la primera se realiza un desarrollo con Methontology donde se continúa con las subfases de formalización, implementación y evaluación donde se formaliza la etapa de diseño del motor de búsqueda semántico porque se realiza la creación y carga de la Ontología, creación de clases, atributos, relaciones e instancias, todo esto con la herramienta Protégé de la mano del módulo de Owlready2 y Python, también se realizan diversas pruebas con el lenguaje de consultas SPARQL de la mano del servidor Apache Jena Fuseki. En esta fase de desarrollo se utilizan las Ontologías como componente y técnica principal para el motor de búsqueda semántico.

Es importante aclarar que las subfases de implementación y evaluación, que en este apartado se mencionan, no hacen referencia a las etapas del proyecto que implican la implementación y evaluación del motor de búsqueda en el contexto de la Universidad de Nariño sino a las subfases previamente explicadas de Methontology para el manejo de Ontologías.

Para esta fase se usan los datos recolectados en las etapas anteriores para la construcción e instanciación del corpus que compone el tesauro de la Ontología.

La segunda fase es el desarrollo con Machine Learning que se basa en el entrenamiento de algoritmos de Machine Learning con procesamiento de lenguaje natural como son: Word2Vec y Doc2Vec. Se crean modelos con los datos preprocesados en la etapa de preparación de proyectos de investigación, se entrena dichos modelos, se evalúan los resultados y se retroalimenta hasta obtener resultados satisfactorios. En esta fase de desarrollo se utiliza Machine Learning como componente y técnica secundaria para el motor de búsqueda semántico.

En esta tercera fase de desarrollo se integran las anteriores fases y se unifica la Ontología con el Machine Learning a fin de tener un orden, claridad e integridad en el desarrollo exitoso del motor de búsqueda semántico. Además se construye el gestor de búsquedas y el gestor de resultados, que como se verá en el apartado de de arquitectura (Ver 5. ARQUITECTURA), estos son integrados y están constantemente interactuando en la capa de lógica con las capas de presentación y de datos.

4.8 IMPLEMENTACIÓN Y DESPLIEGUE DEL MOTOR DE BÚSQUEDA SEMÁNTICO

El motor de búsqueda semántico se implementa y despliega, para ello se utiliza Flask, manejando el patrón de arquitectura de software: MVC (Modelo Vista Controlador) en donde se realizan los modelos soportados con PostgreSQL, los controladores en lenguaje de Flask para finalmente realizar las vistas con HTML y CSS de la mano de Boostrap Studio.

4.9 REALIZACIÓN DE PRUEBAS FINALES Y EVALUACIÓN DEL RENDIMIENTO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

En esta etapa se realizan casos de prueba, pruebas unitarias y pruebas de aceptación. Así, se describen las pruebas aplicadas al motor de búsqueda para medir si este encuentra resultados satisfactorios a diversas búsquedas y si lo hace de manera efectiva

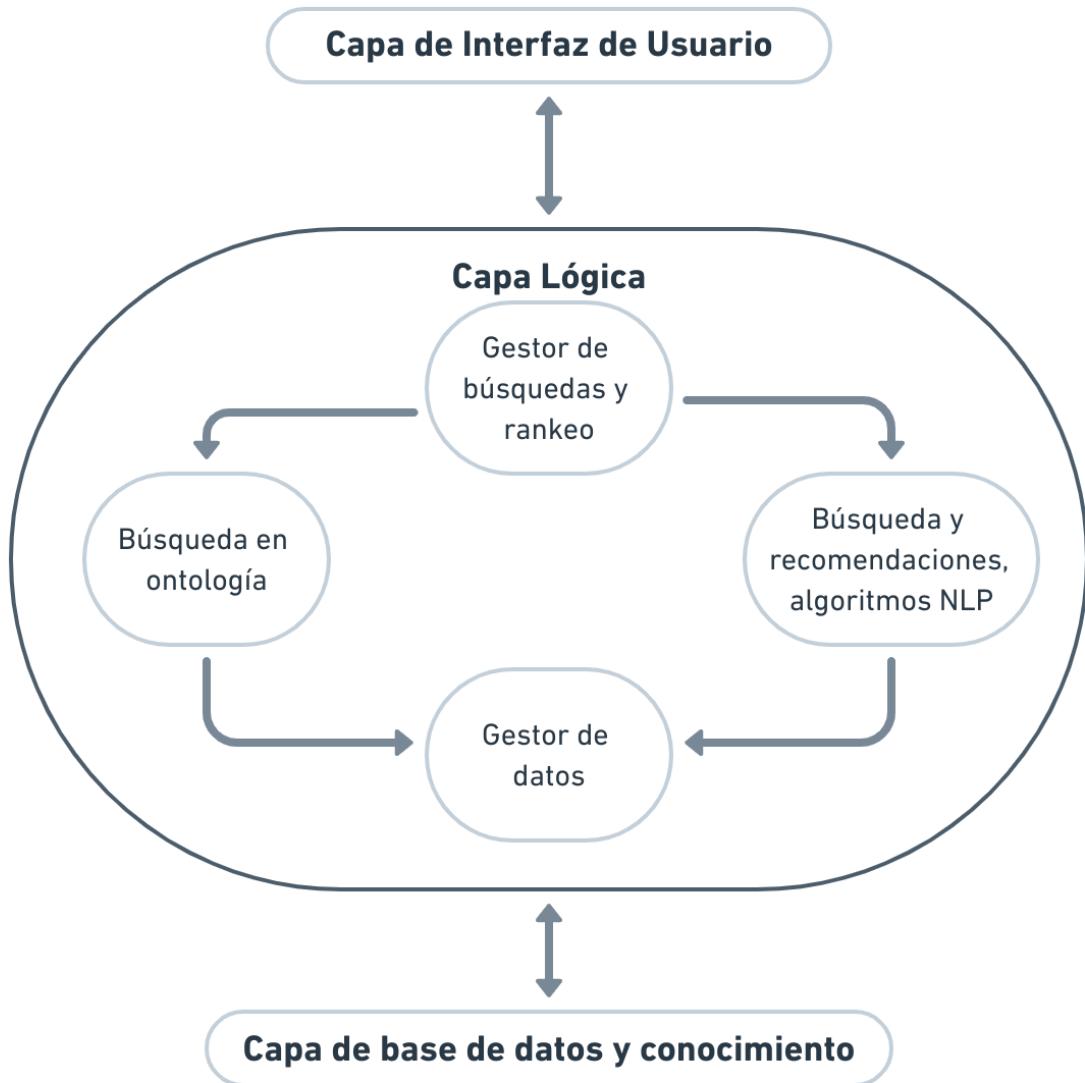
siguiendo determinados lineamientos. De esta forma se mide y se evalúa el rendimiento del motor de búsqueda semántico y del proyecto en general.

5. ARQUITECTURA

La arquitectura del motor de búsqueda semántico se puede evidenciar en la figura 2.

Obs8. Ajustar todas las figuras según las indicaciones de la guía, ver indicaciones en la página 45

Figura 2. Arquitectura del motor de búsqueda



Fuente: Creación propia

5.1 CAPAS DE LA ARQUITECTURA DEL MOTOR DE BÚSQUEDA

✓ **Capa de Presentación**

En esta capa se encuentra el módulo IGU (Interfaz Gráfica de Usuario), el cual presenta la interfaz del motor de búsqueda semántico e interactúa con el usuario final para realizar consultas y mostrar resultados. Se conecta con la capa de lógica.

✓ **Capa de Lógica**

En esta capa se encuentran los módulos: Gestor de Búsquedas, Ontología, Algoritmos de Machine Learning y Gestor de Datos. Estos módulos se encargan de procesar todos los datos de los proyectos de investigación para el motor de búsqueda semántico. Esta capa se conecta con las capas de presentación y de datos.

✓ **Capa de Datos**

En esta capa se encuentra el módulo de Base de Datos, el cual almacena todos los datos de los proyectos de investigación y permite la conexión con la capa de lógica.

5.2 MÓDULOS DE LA ARQUITECTURA DEL MOTOR DE BÚSQUEDA

✓ **Módulo IGU**

Este módulo presenta la Interfaz Gráfica de Usuario del motor de búsqueda semántico, este módulo interactúa con el usuario final presentándole los elementos necesarios para realizar la búsqueda y satisfacer sus necesidades y experiencia de usuario. Este módulo se conecta con el módulo de Gestor de Búsquedas de manera bidireccional para el evento en el que se realice la búsqueda o en el que se muestre el resultado.

✓ **Módulo Gestor de Búsquedas**

Este módulo permite realizar consultas a los módulos de Ontología y Algoritmos de Machine Learning así como retornar el resultado de esas búsquedas al módulo de IGU, para esta labor se instruye de SPARQL, Elasticsearch y scripts que gestionan las búsquedas y las retornan exitosamente.

✓ **Módulo Ontología**

Este módulo almacena todo el conocimiento semántico de la Ontología recorriendo clases, atributos, relaciones e instancias. Se conecta con el módulo Gestor de Búsquedas en el evento en el que se realiza una búsqueda o en el momento en el que se retorna el resultado de esa búsqueda, es por eso que esta conexión se realiza de manera bidireccional. Se conecta con el módulo Gestor de Datos en las etapas iniciales o cuando se alimenta la base de datos, esto es porque este Gestor de Datos realiza las fases de preparación, exploración y análisis de datos y estos datos son insumo para el módulo de Ontología.

✓ **Módulo Algoritmos de Machine Learning**

Este módulo almacena todo el conocimiento semántico de los algoritmos de Machine Learning como: Word2Vec y Doc2Vec. Se conecta con el módulo Gestor de Búsquedas en el evento en el que se realiza una búsqueda o en el momento en el que se retorna el resultado de esa búsqueda, es por eso que esta conexión se realiza de manera bidireccional. Se conecta con el módulo Gestor de Datos en las etapas iniciales o cuando se alimenta la base de datos, esto es porque este Gestor de Datos realiza las fases de preparación, exploración y análisis de datos y estos datos son insumo para el módulo de Algoritmos de Machine Learning.

✓ **Módulo Gestor de Datos**

Este módulo se encarga de gestionar todos los datos que se encuentran en el módulo de Base de Datos con el cual está conectado, esto se logra con la preparación, exploración y análisis de los proyectos de investigación, en donde se destacan los algoritmos de Procesamiento de Lenguaje Natural como: SpaCy, NLTK, BOW y TF-IDF. A su vez se conecta con los módulos de Ontología y Algoritmos de Machine Learning por motivo de suministrar a modo de insumo todos los datos que fueron transformados en este módulo y que posteriormente estos puedan ser fuente para la semántica.

✓ **Módulo Base de Datos**

En este módulo se almacenan los datos de los proyectos de investigación que incluyen: proyectos docentes, proyectos estudiantiles y trabajos de grado. Esto se logra mediante el sistema de gestión de bases de datos PostgreSQL. Este módulo se conecta con el módulo de Gestor de Datos para enviarle datos y que este último pueda procesar los datos y enviarle a los demás módulos respectivos.

6. RESULTADOS

Se presentan los resultados obtenidos para cada etapa de la metodología aplicada:

6.1 APROPIACIÓN DEL CONOCIMIENTO

Partiendo de una revisión exhaustiva de la literatura, el estado del arte y todas las bases sólidas teóricas que componen la armonía de la investigación, se apropió el conocimiento evidenciado mediante: la definición de la modalidad, la línea de investigación, el alcance y delimitación del proyecto, en el planteamiento y formulación del problema, en la redacción de objetivos, justificación, antecedentes, marco teórico, metodología, resultados, análisis y discusión de resultados, conclusiones, recomendaciones, bibliografía y anexos presentes en este proyecto.

Se destaca el resultado del conocimiento adquirido de todas las temáticas que abarcan el proyecto, así como también de las diversas herramientas y lenguajes usados. Se obtuvo el aprendizaje de temáticas tales como: semántica, web semántica, Ontologías, motores de búsqueda, aprendizaje de máquina, procesamiento de lenguaje natural, Scrum y Methontology. De igual forma se conoció y reforzó aprendizaje en lenguajes como: python, XML, RDF, OWL y SPARQL.

6.2 INSTALACIÓN Y CONFIGURACIÓN DE LAS HERRAMIENTAS, BIBLIOTECAS Y TECNOLOGÍAS

Se destaca el resultado de la instalación y configuración de: Jupyter notebook, Protégé, Owlready2, Apache Jena Fuseki, Elasticsearch, Visual Studio Code, Anaconda, Gensim con Word2Vec y Doc2Vec, Pandas, Numpy, TF-IDF, BOW, NLTK, SpaCy, etc.

6.3 RECOLECCIÓN Y EXTRACCIÓN DE PROYECTOS DE INVESTIGACIÓN

Se destaca el resultado de recolectar y extraer información de los proyectos de investigación de proyectos docentes, proyectos estudiantiles y trabajos de grado que se encuentran almacenados en el sistema de investigaciones de la Universidad de Nariño.

Se aclara que actualmente la diferencia entre proyectos estudiantiles y trabajos de grado radica en que los proyectos estudiantiles se inscriben desde los primeros semestres de la carrera universitaria (desde primero hasta octavo) mientras que los trabajos de grado se inscriben desde los últimos semestres de la carrera universitaria (séptimo en adelante) hasta el momento de figurar como egresado (si es el caso).

La información que se recolectó y extrajo incluye:

- ✓ El código del proyecto.
- ✓ El título del proyecto.
- ✓ El resumen del proyecto.
- ✓ El estado del proyecto.
- ✓ Palabras clave del proyecto.
- ✓ Código de (los) autor(es).
- ✓ Nombres de (los) autor(es).
- ✓ Facultad de (los) autor(es).
- ✓ Departamento de (los) autor(es).
- ✓ Programa de (los) autor(es).
- ✓ Nombres de (los) asesor(es).
- ✓ El tipo de convocatoria con su respectivo año.
- ✓ El grupo de investigación de (los) autor(es).
- ✓ La línea de investigación de (los) autor(es).

6.4 PREPARACIÓN DE PROYECTOS DE INVESTIGACIÓN

Se destaca el resultado de preparar los proyectos de investigación, de modo tal que esto permitió navegar por las siguientes etapas, anticipando y evitando inconvenientes, errores o problemas con respecto a la calidad de los datos.

En este orden de ideas, se resalta la realización de las siguientes fases:

6.4.1 Organización de datos. En esta fase se aplicaron algoritmos de ordenamiento (creados por los autores de este proyecto) para los proyectos de investigación, esto dado que los proyectos en la fase de recolección y extracción se encontraban desordenados y en condiciones no aptas para ser tratados, manejados y trabajados. Todos los datos se encuentran en el Anexo C (Ver Anexo C).

6.4.2 Creación de corpus. En esta fase se creó el corpus para los proyectos de investigación, el cual fue el insumo más potente de la semántica como se podrá observar en las etapas posteriores. Dicho corpus resultó de unificar todos los datos de los proyectos de investigación (ya organizados en la fase previa) que fueron: título y resumen de la investigación; palabra clave 1, palabra clave 2, palabra clave 3, palabra clave 4, palabra clave 5; nombres, apellidos, programa, facultad, departamento, grupo de investigación y línea de investigación para cada uno de los autores y asesores.

6.4.3 Preprocesamiento de datos. En esta fase se utilizó las bibliotecas de NLTK y SpaCy para realizar un preprocesamiento de los datos obtenidos en la anterior fase.

En la figura 3 se puede visualizar un fragmento de un proyecto de investigación, al cual como se observará, a modo de ejemplo, se le aplicarán las técnicas de preprocesamiento de datos. Para ello se trabajó con las siguientes subfases:

Figura 3. Ejemplo de texto al cual se le aplicó preprocesamiento de datos

En este trabajo se aplica una técnica novedosa conocida como Deep Learning que forma parte de las herramientas de inteligencia computacional, particularmente del área de las Redes Neuronales Artificiales para clasificar dos tipos de sismos volcánicos: volcán-tectónicos o VT y largo periodo o LP. Se introduce al sistema ejemplos que contienen las series de tiempo correspondientes a los registros de sismos de los tipos LP y VT para realizar el 'entrenamiento' del sistema, éste extrae características generales que diferencian un tipo del otro; una vez el sistema ha 'aprendido', es capaz de reconocer, en un registro que nunca ha conocido, las características que distinguen un tipo de otro y catalogarlo de manera correcta.

Fuente: Creación propia

6.4.3.1 Tokenización de datos. En esta subfase se ejecutaron algoritmos de la biblioteca NLTK para separar todas las palabras y poder trabajar con ellas de manera individual. En la figura 4 se observa que el texto observado en la figura 3 fue tokenizado.

Figura 4. Texto tokenizado

```
['En', 'este', 'trabajo', 'se', 'aplica', 'una', 'técnica', 'novedosa', 'conocida', 'como', 'Deep', 'Learning', 'que', 'forma', 'parte', 'de', 'las', 'herramientas', 'de', 'inteligencia', 'computacional', ' ', 'particularmente', 'del', 'área', 'de', 'las', 'Redes', 'Neuronales', 'Artificiales', 'para', 'clasificar', 'dos', 'tipos', 'de', 'sismos', 'volcánicos', ':', 'volcano', 'tectónicos', 'o', 'VT', 'y', 'largo', 'periodo', 'o', 'LP', '!', 'Se', 'introduce', 'al', 'sistema', 'ejemplos', 'que', 'contienen', 'las', 'series', 'de', 'tiempo', 'correspondientes', 'a', 'los', 'registros', 'de', 'sismos', 'de', 'los', 'tipos', 'LP', 'y', 'VT', 'para', 'realizar', 'el', '"entrenamiento", "", "del", "sistema", " ", "éste", "extrae", "características", "generales", "que", "diferencian", "un", "tipo", "del", "otro", ":", "una", "vez", "el", "sistema", "ha", '"aprendido", "", "", "es", "capaz", "de", "reconocer", ":", "en", "un", "registro", "que", "nunca", "ha", "conocido", ":", "las", "características", "que", "distinguen", "un", "tipo", "de", "otro", "y", "catalogarlo", "de", "manera", "correcta", ".']
```

Fuente: Creación propia

6.4.3.2 Normalización de datos. Para esta subfase se aplicaron algoritmos para que todos los datos queden bajo un mismo estándar. Esto se puede observar en la figura 5, en la cual se normalizaron los tokens de la anterior subfase.

Figura 5. Texto normalizado

en este trabajo se aplica una técnica novedosa conocida como deep learning que forma parte de las herramientas de inteligencia computacional , particularmente del área de las redes neuronales artificiales para clasificar dos tipos de sismos volcánicos : volcano tectónicos o vt y largo periodo o lp . se introduce al sistema ejemplos que contienen las series de tiempo correspondientes a los registros de sismos de los tipos lp y vt para realizar el 'entrenamiento ' del sistema , éste extrae características generales que diferencian un tipo del otro ; una vez el sistema ha 'aprendido ' , es capaz de reconocer , en un registro que nunca ha conocido , las características que distinguen un tipo de otro y catalogarlo de manera correcta .

Fuente: Creación propia

6.4.3.4 Limpieza de datos

En esta subfase se aplicaron algoritmos de NLTK y SpaCy junto con expresiones regulares para que los datos queden totalmente limpios, esto con la eliminación de datos nulos, signos de puntuación, caracteres “no ascii” y stopwords. Se puede evidenciar esta fase en la figura 6, en donde se realizó la limpieza de los datos previamente normalizados.

Figura 6. Texto limpio

trabajo aplica tecnica novedosa conocida deep learning forma parte herramientas inteligencia computacional particularmente area redes neuronales artificiales clasificar dos tipos sismos volcanicos volcano tectonicos largo periodo introduce sistema ejemplos contienen series tiempo correspondientes registros sismos tipos realizar entrenamiento sistema extrae caracteristicas generales diferencian tipo vez sistema aprendido capaz reconocer registro nunca conocido caracteristicas distinguen tipo catalogarlo manera correcta

Fuente: Creación propia

6.4.3.3 Lematización de datos

Finalmente en esta subfase se lematizaron los datos a los cuales ya se les realizó la etapa de limpieza, tal y como lo indica la figura 7.

Figura 7. Texto lematizado

trabajar aplicar tecnica novedoso conocido deep learning formar partir herramienta inteligencia computacional particularmente area radar neuronal artificial clasificar do tipo sismo volcanicos volcano tectonicos largar periodo introducir sistema ejemplo contener serie tiempo correspondiente registro sismo tipo realizar entrenamiento sistema extraer caracteristicas general diferenciar tipo vez sistema aprender capaz reconocer registrar nunca conocer caracteristicas distinguir tipo catalogarlo manera correcto

Fuente: Creación propia

6.5 EXPLORACIÓN Y ANÁLISIS DE PROYECTOS DE INVESTIGACIÓN

Se destaca el resultado de obtener nubes de palabras e histogramas a partir de los algoritmos de procesamiento de lenguaje natural como son: BOW (Bag Of Words) y TF-IDF (Term Frequency – Inverse Document Frequency).

En la figura 8 se pueden contemplar las palabras que más se repiten en los proyectos de investigación de la Universidad de Nariño dentro de una nube de palabras, esto se realizó con ayuda del algoritmo de procesamiento de lenguaje natural BOW (Bag Of Words).

Figura 8. Nube de palabras con BOW



Fuente: Creación propia

En la figura 9 se puede observar la nube de palabras de los proyectos de investigación, análoga a la nube de palabras de la figura 8 con la diferencia que el algoritmo usado fue el de procesamiento de lenguaje natural TF-IDF.

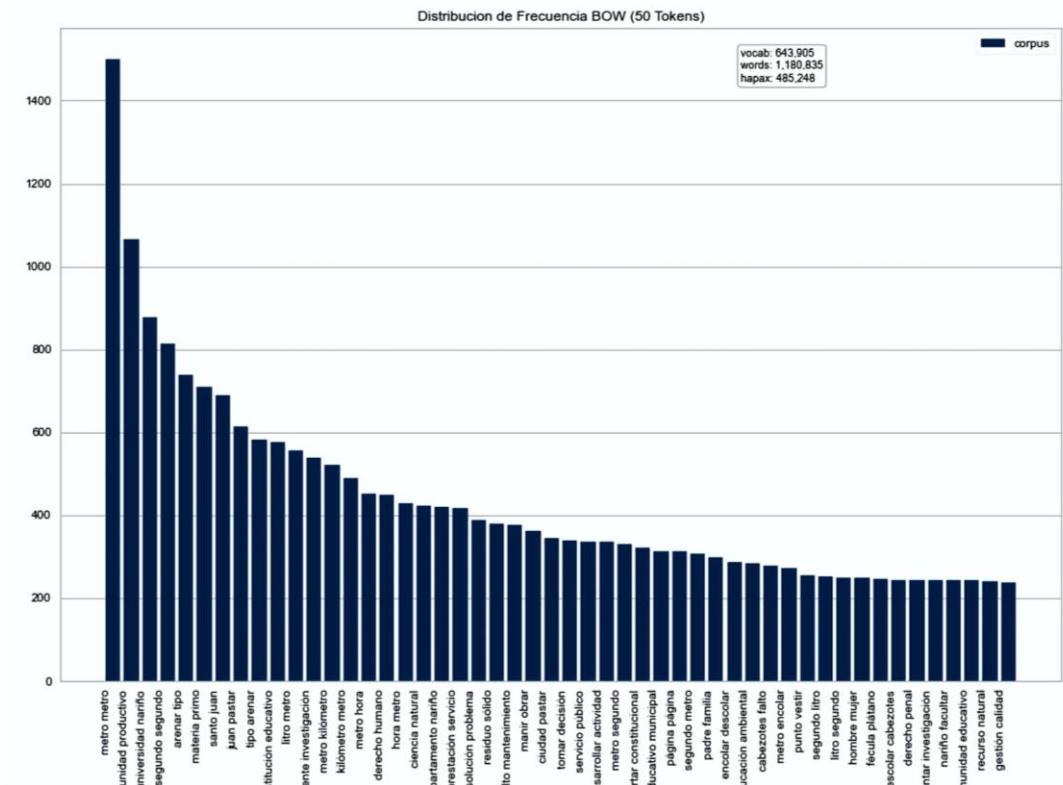
Figura 9. Nube de palabras con TF-IDF



Fuente: Creación propia

En la figura 10 se puede observar el histograma que se realizó con el algoritmo de BOW, para esto se usaron los n-gramas que consisten en combinar palabras de manera aleatoria (específicamente se usaron bigramas que son conjuntos de dos palabras) donde se observó si estas palabras se repetían o tenían alguna influencia en el corpus de proyectos de investigación. En el eje horizontal del histograma se reflejan las palabras del corpus y en el eje vertical está el número de veces que se repiten.

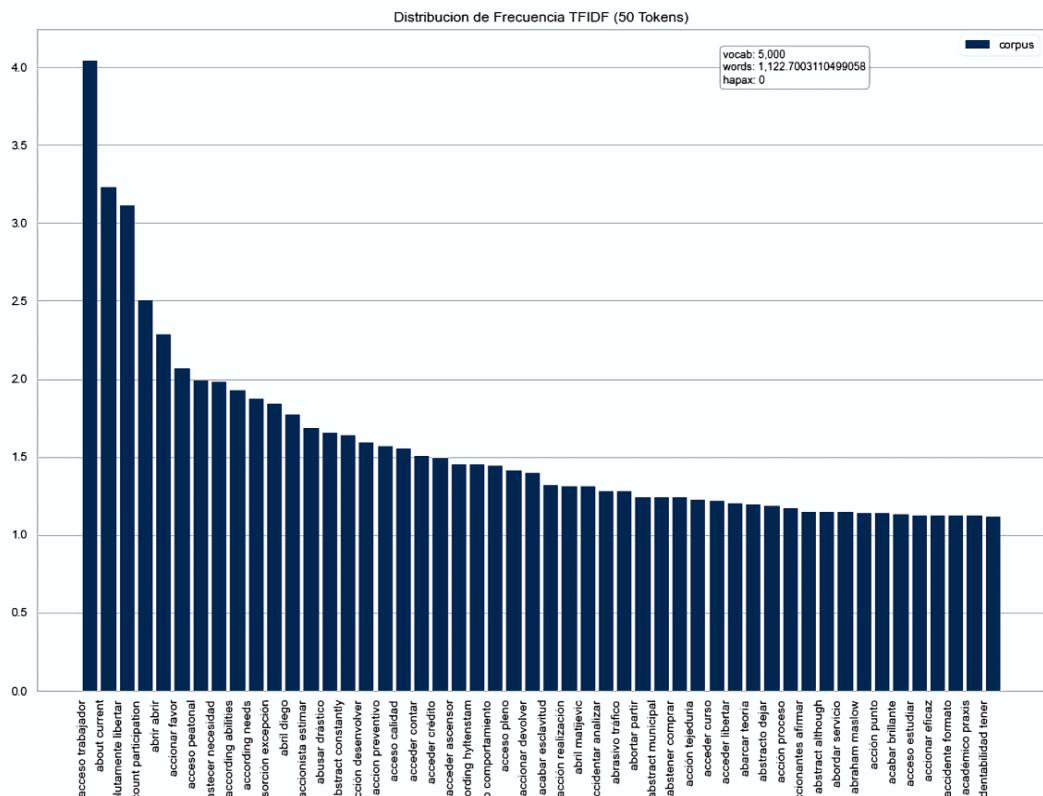
Figura 10. Histograma con BOW



Fuente: Creación propia

En la figura 11 se puede observar el histograma que se realizó con el algoritmo de TF-IDF, al igual que con BOW también se usaron los bigramas observando si estas palabras se repetían o tenían alguna influencia en el corpus de proyectos de investigación. En el eje horizontal del histograma se reflejan las palabras del corpus y en el eje vertical está el número de veces que se repiten.

Figura 11. Histograma con TF-IDF



Fuente: Creación propia

De esta forma, con los resultados observados en las figuras 8, 9, 10 y 11 se determinaron cuales palabras estaban más presentes en el proyecto, qué tipos de palabras fueron o a que categoría gramatical pertenecían (sustantivos, verbos, adjetivos, etc.) y de igual forma se obtuvo una idea y un posible acercamiento a los tópicos y contextos de los proyectos de investigación. A su vez fue fundamental porque ayudó para encontrar nuevas posibles palabras en el corpus y así armar el vocabulario como se contemplará en las etapas posteriores.

De esta forma se logró el análisis y la exploración de los datos donde se usaron ambos algoritmos BOW y TF-IDF con el fin de abarcar un mayor número de posibilidades frente a las palabras del corpus.

6.6 DISEÑO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

Una vez culminada la etapa anterior de exploración y análisis de los datos, se procedió a diseñar el motor de búsqueda semántico. Este diseño se realizó teniendo en cuenta las fases de especificación y conceptualización de la metodología Methontology, en donde se destacan los siguientes resultados:

6.6.1 Fase de especificación. Dentro de esta fase se identificaron las razones que llevaron a realizar la Ontología, también se describió quienes se pretende sean los usuarios finales que hagan uso del motor de búsqueda semántico y a su vez se realizó un proceso de adquisición de conocimiento; este proceso de adquisición de conocimiento difiere de la apropiación del conocimiento general del proyecto, dado que esta adquisición se enfoca fundamentalmente en el diseño del motor de búsqueda.

Así las cosas, se resalta que la Ontología se creó para darle un componente robusto de semántica al motor de búsqueda interactuando con los proyectos de investigación y sus componentes. Los usuarios finales serán todos aquellos investigadores que deseen acceder mediante diversas búsquedas a los recursos de los proyectos de investigación.

La adquisición del conocimiento, que se puede entender como el principal insumo para la Ontología, se basó en recolectar información de fuentes web particularmente del Sistema de Investigaciones de VIIS⁷¹. La tabla 2 muestra la categoría y el contenido relacionado acerca del conocimiento sobre los proyectos de investigación junto con todos los componentes en los que se ven inmersos para poder llevar el conocimiento a la Ontología.

Este es un proceso inicial y en la fase de conceptualización se muestra qué componentes se han elegido para la Ontología y de manera más profunda a qué refiere cada uno:

Tabla 2. Adquisición del conocimiento para la Ontología

Categoría	Contenido relacionado
VIIS (Vicerrectoría de Investigaciones e Interacción Social)	Proyectos de Investigación Convocatorias
Proyectos de investigación	Modalidades Convocatorias Facultades Departamento Grupos de investigación Líneas de investigación Autores

⁷¹ Sistema de Información de Investigaciones. [En línea]. Disponible en: <http://sisinfoviis.udnar.edu.co>

Categoría	Contenido relacionado
Convocatorias	Asesores Contenidos de proyecto de investigación Estado de la investigación Trabajos de Grado Proyectos estudiantiles Proyectos docentes Año Acuerdo
Modalidades	Investigación Científica Desarrollo Tecnológico Innovación Creación Artística y Cultural
Facultades	Departamentos
Departamentos	Programas Grupos de investigación
Programas	Docentes Estudiantes
Centros de investigación	Grupo de investigación
Grupos de investigación	Código Nombre Grupo Departamentos Área de Investigación Líneas de investigación Clasificación Colciencias Director Personas
Líneas de investigación	Director Grupos de Investigación
Personas	Nombres, apellidos Cedula Correo Investigadores Asesores Autores
Investigadores	Código Categoría Estudiantes Docentes
Asesores	Docentes
Contenidos de proyecto de investigación	Titulo Resumen Palabras clave

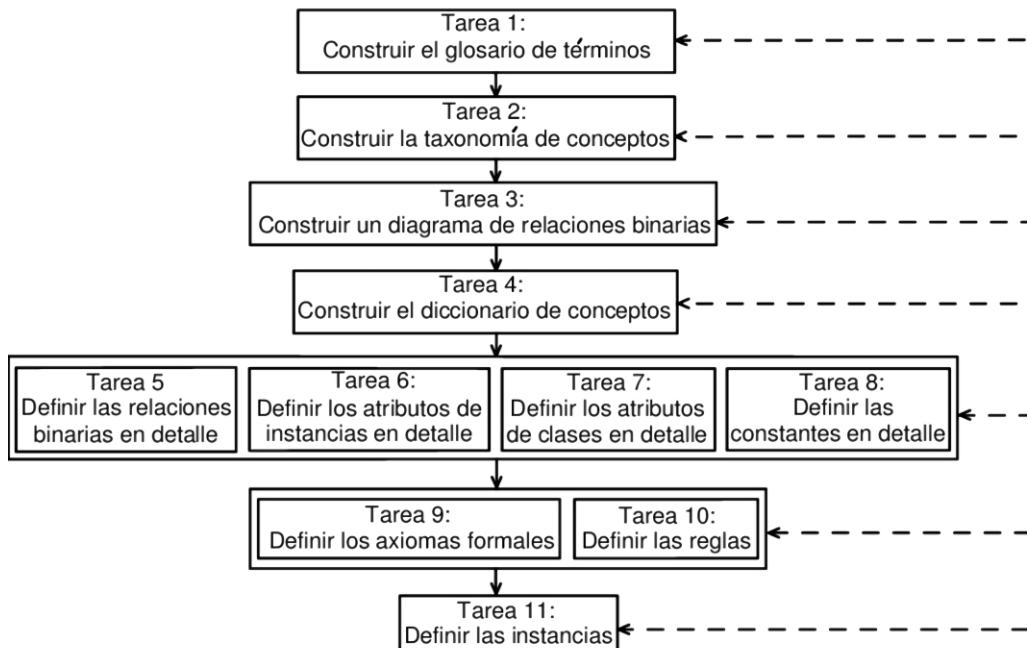
Categoría	Contenido relacionado
Tema/categoría	Tema/categoría
Vocabulario	Vocabulario Palabras Relaciones
Palabras	Tipo Sinónimo Significado

Fuente: Creación propia

6.6.2 Fase de conceptualización. Dentro de la fase de conceptualización se desarrollaron once tareas puntuales que permitieron conceptualizar exitosamente: clases, atributos, relaciones e instancias de la Ontología. La mayoría de estos y de manera completa se pueden visualizar en el Anexo B (Ver Anexo B).

Las tareas que se realizaron utilizando esta metodología se muestran en la figura 12:

Figura 12. Tareas de la metodología Methontology



Fuente: Tabares y Jiménez⁷²

Tarea 1. Construir el glosario de términos:

⁷² TABARES, John & JIMÉNEZ, Jovani. Ontología para el proceso evaluativo en la educación superior. En: Revista Virtual Universidad Católica del Norte. Vol. 1. No. 42. (Agosto, 2014), p. 68-79. [En línea]. Disponible en: <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/495>

En esta tarea se listaron todos los términos importantes seleccionados después de analizar la anterior fase de especificación con su proceso de adquisición de conocimiento, así mismo se presenta una breve descripción de cada uno como se indica en la tabla 3.

Tabla 3. Glosario de términos de la Ontología

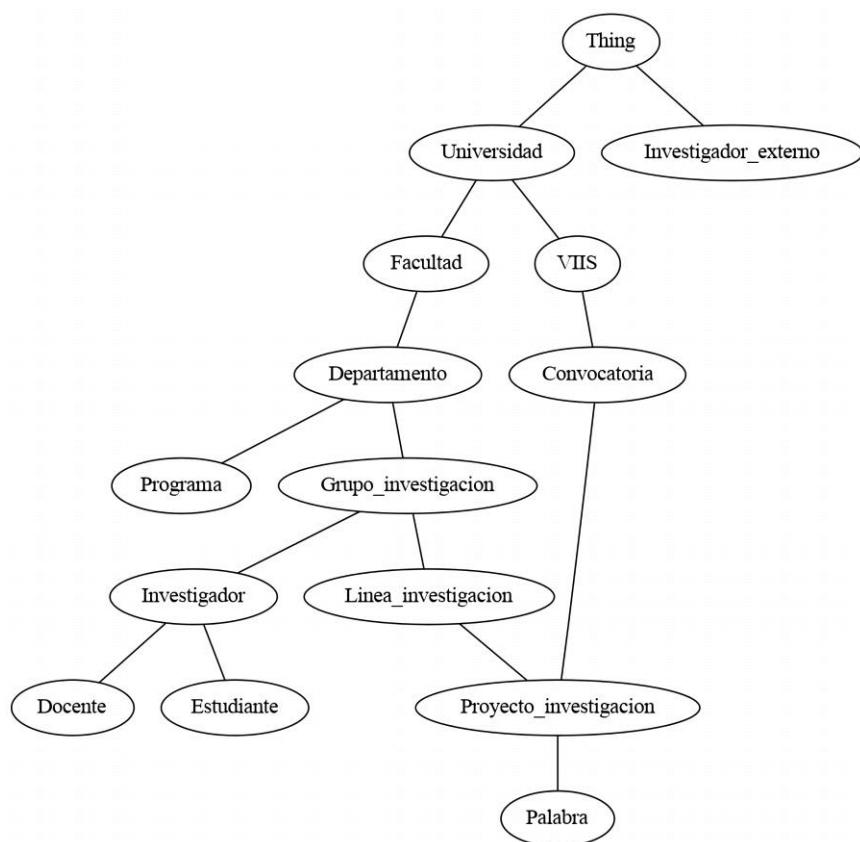
Término	Descripción
Universidad	Es la entidad orientada a la educación que contiene facultades.
Facultad	Es la entidad que contiene departamentos académicos.
VIIS	Vicerrectoría de Investigaciones e Interacción Social, es la entidad encargada del aspecto investigativo a lo largo de la Universidad, es quien gestiona los recursos económicos para proyectos de investigación.
Departamento	Es la entidad que contiene programas académicos.
Convocatoria	Este término refiere al llamado por parte de la VIIS para que los investigadores acudan a esta convocatoria y presenten proyectos con el fin de que estos sean financiados.
Programa	Es el programa académico dentro del cual se encuentran docentes y estudiantes.
Grupo de investigación	Es el grupo conformado por docentes y/o estudiantes investigadores para presentar proyectos a la convocatoria de la VIIS.
Docente	Es un investigador que pertenece a la Universidad, que realiza proyectos de tipo docente.
Estudiante	Es un investigador que pertenece a la Universidad que realiza proyectos estudiantiles y/o trabajos de grado.
Investigador externo	Es un investigador que es externo a la Universidad pero que se presenta a la convocatoria de VIIS.
Línea de investigación	Es una rama que el grupo de investigación maneja, enfocada a un ámbito específico del conocimiento.
Investigador	Es quien elabora proyectos de investigación y los presenta a la convocatoria de VIIS. Este investigador, puede ser docente, estudiante o investigador externo.
Proyecto de investigación	Es la entidad que contiene todo lo referente a un proyecto de investigación.
Palabra	Es quizá la entidad más importante dentro del dominio investigativo, porque hace referencia a cada una de las palabras que conforma el proyecto de investigación con las cuales se generó gran parte de la semántica y se armó el tesauro.

Fuente: Creación propia

Tarea 2. Construir taxonomía de conceptos:

En esta tarea se definió la taxonomía o jerarquía de los conceptos o clases de la Ontología que se obtuvo a partir del glosario de términos de la tarea 1, esta taxonomía se muestra en la figura 13.

Figura 13. Taxonomía de conceptos de la Ontología

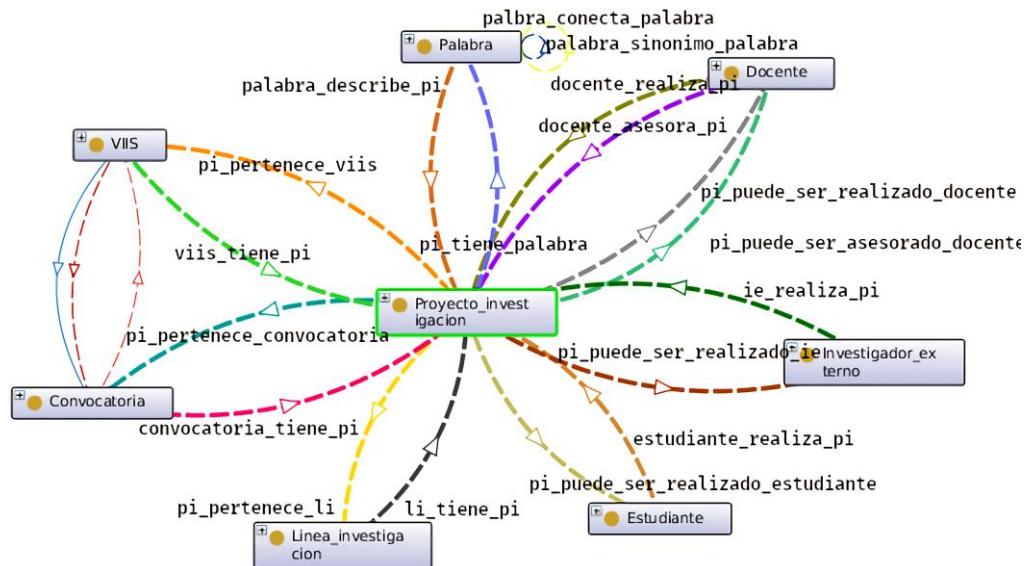


Fuente: Creación propia

Tarea 3. Construir un diagrama de relaciones binarias:

En esta tarea se elaboró el diagrama de relaciones binarias que son los predicados (object properties en Protegé) de la Ontología. Se visualizan en la figura 14 las relaciones de la clase más importante de la Ontología que es: Proyecto de investigación.

Figura 14. Diagrama de relaciones binarias de la clase Proyecto de Investigación



Fuente: Creación propia

Tarea 4. Construir el diccionario de conceptos:

En esta tarea se detallaron los conceptos o clases más importantes o relevantes dentro del dominio investigativo de la Ontología, destacando sus atributos, relaciones e instancias. Se eligieron: Investigador porque es el actor principal que le da vida al proyecto de investigación como se observa en la tabla 4. Proyecto de investigación porque es donde reposa la materia prima de donde se extrajeron todas las palabras para descubrir el poder semántico, como se observa en la tabla 5. Palabra porque son las palabras con las cuales se armó el tesoro, además con estas se encontraron diversas relaciones que son la fuente base para las búsquedas y así obtener el conocimiento, esto se observa en la tabla 6.

Tabla 4. Diccionario de concepto Investigador

Clase Investigador	
Atributos	
	id_investigador
	nombres_investigador
	apellidos_investigador
	codigo_investigador
	cedula_investigador
	correo_investigador

Clase Investigador	
Relaciones	investigador_puede_ser_docente investigador_puede_ser_estudiante investigador_puede_ser_ie investigador_pertenece_gi investigador_pertenece_viiis investigador_se_encuentra_convocatoria
Instancias	Investigador
Fuente:	Creación propia

Tabla 5. Diccionario de concepto Proyecto de investigación

Clase Proyecto de investigación	
Atributos	id_proyecto_investigacion titulo_proyecto_investigacion resumen_proyecto_investigacion palabras_clave estado_proyecto_investigacion tipo_proyecto_investigacion
Relaciones	pi_tiene_palabra pi_pertenece_li pi_pertenece_convocatoria pi_puede_ser_realizado_estudiante pi_puede_ser_realizado_docente pi_puede_ser_realizado_ie pi_puede_ser_asesorado_docente pi_pertenece_viiis
Instancias	Proyecto investigación
Fuente:	Creación propia

Tabla 6. Diccionario de concepto Palabra

Clase Palabra	
Atributos	id_palabra descripcion_palabra lema_palabra tipo_palabra concepto_palabra
Relaciones	palabra_sinonimo_palabra palabra_conecta_palabra palabra_describe_pi
Instancias	Palabra
Fuente:	Creación propia

Tarea 5. Definir las relaciones binarias en detalle:

Se obtuvo un total de 55 relaciones binarias, de las cuales se destacan 3: pi_tiene_palabra porque cada proyecto se relaciona con sus palabras de modo tal que esta relación facilita las búsquedas como se visualiza en la tabla 7, palabra_sinonimo_palabra que es fundamental para construir el tesauro como se visualiza en la tabla 8 y palabra_conecta_palabra dado que de este modo se encontraron diversas relaciones semánticas como se visualiza en la tabla 9.

Para cada relación se obtuvo su clase origen (dominio), clase destino (rango), relación inversa y cardinalidad.

Tabla 7. Relación binaria (pi_tiene_palabra) en detalle

Relación pi_tiene_palabra	
Clase Origen (Dominio)	Proyecto de investigación
Clase Destino (Rango)	Palabra
Relación Inversa	palabra_describe_pi
Cardinalidad	1:N

Fuente: Creación propia

Tabla 8. Relación binaria (palabra_sinonimo_palabra) en detalle

Relación palabra_sinonimo_palabra	
Clase Origen (Dominio)	Palabra
Clase Destino (Rango)	Palabra
Relación Inversa	Tiene relación simétrica, es decir la relación directa es igual a la relación inversa.
Cardinalidad	1:N

Fuente: Creación propia

Tabla 9. Relación binaria (palabra_conecta_palabra) en detalle

Relación palabra_conecta_palabra	
Clase Origen (Dominio)	Palabra
Clase Destino (Rango)	Palabra
Relación Inversa	Tiene relación simétrica, es decir la relación directa es igual a la relación inversa.
Cardinalidad	1:N

Fuente: Creación propia

Tarea 6. Definir los atributos de instancias en detalle:

Se obtuvo un total de 45 atributos de los cuales se muestra los de las clases más representativas que son las que se mencionaron en el diccionario de conceptos en la tarea 4 (Investigador, Proyecto de investigación y Palabra), los atributos de instancias de estas clases se pueden observar en: tabla 10, tabla 11 y tabla 12 respectivamente.

Para cada clase se visualizan: los atributos de instancia, el nombre de la clase (dominio), el tipo de dato (rango) y su cardinalidad.

Tabla 10. Atributos de instancia de clase Investigador

Atributo	Clase (Dominio)	Tipo de Dato (Rango)	Cardinalidad
id_investigador	Investigador	Int	1
nombres_investigador		String	1
apellidos_investigador		String	1
codigo_investigador		String	1
cedula_investigador		String	1
correo_investigador		String	1

Fuente: Creación propia

Tabla 11. Atributos de instancia de clase Proyecto de investigación

Atributo	Clase (Dominio)	Tipo de Dato (Rango)	Cardinalidad
id_proyecto_investigacion	Proyecto_investigacion	Int	1
titulo_proyecto_investigacion		String	1
resumen_proyecto_investigacion		String	1
palabras_clave		String	0:5
estado_proyecto_investigacion		String	1
tipo_proyecto_investigacion		String	1

Fuente: Creación propia

Tabla 12. Atributos de instancia de clase Palabra

Atributo	Clase (Dominio)	Tipo de Dato (Rango)	Cardinalidad
id_palabra	Palabra	String	1
descripcion_palabra		String	1
lema_palabra		String	1
tipo_palabra		String	1
concepto_palabra		String	1

Fuente: Creación propia

Tarea 7. Definir los atributos de clases en detalle:

En esta tarea se definieron atributos de clase que sirven como restricciones cardinales para cada clase. Estos pueden ser observados en la tabla 13 para las clases: Investigador, Docente y Proyecto de investigación.

Tabla 13. Atributos de clases Investigador, Docente y Proyecto de investigación

Clase	Atributo
Investigador	Máximo 4 investigadores por proyecto de investigación.
Docente	Máximo 2 docentes pueden asesorar proyectos de investigación.
Proyecto de investigación	Máximo 2 años y mínimo 6 meses de duración del proyecto de investigación.

Fuente: Creación propia

Tarea 8. Definir las constantes en detalle:

No se contempló la necesidad de usar constantes para esta Ontología.

Tarea 9. Definir los axiomas formales:

No se contempló la necesidad de usar axiomas que son predicados (relaciones) que siempre se cumplen, es decir siempre son afirmativos.

Tarea 10. Definir las reglas:

Debido a que esta Ontología no presentó axiomas formales, no fue necesario aplicar reglas.

Tarea 11. Definir las instancias:

Se obtuvo instancias para cada una de las clases contempladas en la Ontología: Universidad, Facultad, VIIS, Departamento, Convocatoria, Programa, Grupo de investigación, Docente, Estudiante, Investigador externo, Línea de investigación, Investigador, Proyecto de investigación y Palabra.

6.7 DESARROLLO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

Se desarrolló el motor de búsqueda semántico siguiendo la metodología de Scrum basado en tres fases en las cuales se destacan los siguientes resultados:

6.7.1 Desarrollo con Methontology. Para esta fase se aplicaron las tres subfases de Methontology que son: formalización, implementación y evaluación.

6.7.1.1 Fase de formalización.

En esta fase se presentan los resultados obtenidos después usar la herramienta de Protégé para la construcción de la Ontología en términos semi-computables.

Se destaca el resultado de la creación de clases de la Ontología como indica la figura 15.

Figura 15. Clases de la Ontología



Fuente: Creación propia

Se destaca el resultado de la creación de atributos de la Ontología como se indica la figura 16.

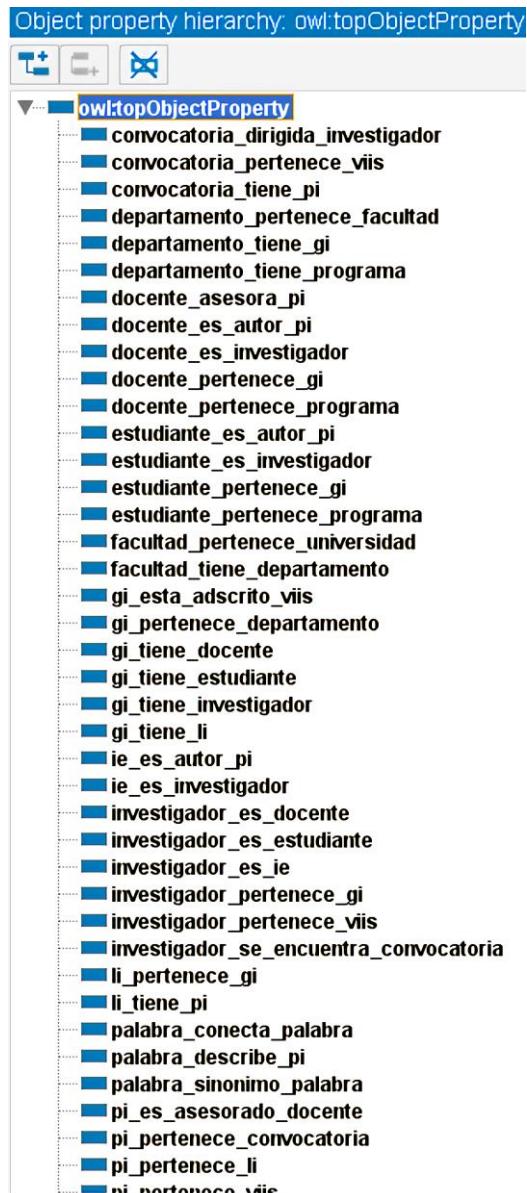
Figura 16. Atributos de la Ontología



Fuente: Creación propia

Se destaca el resultado de la creación de relaciones de la Ontología como se indica la figura 17.

Figura 17. Relaciones de la Ontología



Fuente: Creación propia

Se destaca el resultado de la creación de instancias de la Ontología como se indica la figura 18.

Figura 18. Instancia de un proyecto de investigación con todos sus datos

Fuente: Creación propia

6.7.1.2 Fase de implementación

En esta fase se destacan los resultados de haber usado la biblioteca Owlready2 para codificar una versión computable de la Ontología. Se crearon scripts y se realizó la codificación para el manejo de la Ontología con Python donde se realizó todo un proceso de instanciar objetos de las siguientes clases:

- ✓ **Grupo_investigacion:** esta clase hereda de la clase Thing (es la clase principal y origen de Protégé que viene por defecto). Tiene los métodos de “get” y “set”, también llamados “getters” y “setters”, para obtener y fijar valores sobre cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Linea_investigacion, Docente y Estudiante.
- ✓ **Linea_investigacion:** esta clase hereda de la clase Grupo_investigacion. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene un método para relacionarse de manera directa con instancias de la clase Proyecto_investigacion.
- ✓ **Investigador:** esta clase hereda de la clase Thing. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Docente, Estudiante e Investigador_externo.

- ✓ Docente: esta clase hereda de la clase Investigador. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de la clase Proyecto_investigacion.
- ✓ Estudiante: esta clase hereda de la clase Investigador. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene un método para relacionarse de manera directa con instancias de la clase Proyecto_investigacion.
- ✓ Investigador_externo: esta clase hereda de la clase Investigador. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene un método para relacionarse de manera directa con instancias de la clase Proyecto_investigacion.
- ✓ Palabra: esta clase hereda de la clase Thing. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de la misma clase Palabra (relación simétrica).
- ✓ Proyecto_investigacion: esta clase hereda de la clase Thing. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene un método para relacionarse de manera directa con instancias de la clase Palabra.
- ✓ Universidad: esta clase hereda de la clase Thing. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Facultad y VIIS.
- ✓ Facultad: esta clase hereda de la clase Universidad. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de la clase Departamento.
- ✓ Departamento: esta clase hereda de la clase Facultad. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Programa y Grupo_investigacion.
- ✓ Programa: esta clase hereda de la clase Departamento. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Estudiante y Docente.
- ✓ VIIS: esta clase hereda de la clase Thing. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Convocatoria, Investigador, Grupo_investigacion y Proyecto_investigacion.

- ✓ Convocatoria: esta clase hereda de la clase VIIS. Tiene “getters” y “setters” para cada uno de sus atributos. Tiene métodos para relacionarse de manera directa con instancias de las clases: Proyecto_investigacion e Investigador.

Así mismo dentro de los scripts se codificaron los “DataProperties” de Owlready2 que corresponden a los atributos de la Ontología, junto con los “ObjectProperties” de Owlready2 que corresponden a las relaciones de la Ontología; para cada uno de estos elementos mencionados se determinó el dominio y el rango. Es de resaltar que las relaciones inversas Owlready2 las ejecuta en segundo plano, de modo tal que sólo basta ejecutar la relación directa para que se ejecute la relación: directa e inversa al tiempo.

En síntesis, todas las clases, atributos y relaciones fueron instanciados dentro de la Ontología.

6.7.1.3 Fase de evaluación

En esta fase se destacan resultados después de haber realizado pruebas funcionales de manera local y haber recuperado los proyectos de investigación y demás componentes de la Ontología de manera exitosa con el uso de SPARQL y del servidor de Apache Jena Fuseki mediante el manejo de triples (sujeto predicado objeto).

Para las siguientes figuras se aclara que el prefijo “onto” es un nombre que se le da al recurso donde está ubicada físicamente la Ontología y se usa para llamar a cada uno de sus componentes.

En la figura 19 se observa un fragmento del resultado de la consulta SPARQL para la relación docente_es_autor_pi que se refiere a los docentes que son autores de diversos proyectos de investigación.

Figura 19. Resultado consulta SPARQL para la relación docente_es_autor_pi

Docente	Proyecto_investigacion
1 on:Pablo_Santacruz_Guerrero	on:La_dimensión_socio-política_en_los_procesos_de_creación_artística_en_la_región,_realizados_por_egresados_del_programa_de_Artes_visuales_de_la_Universidad_de_Nariño.
2 on:Jonnathan_Harvey_Narváez_Burbano	on:Canales_de_animación_sociocultural_para_activar_procesos_de_resiliencia_comunitaria_frente_al_fenómeno_de_violencia_barrial_en_la_comuna_10_del_municipio_de_Pasto
3 on:Cristhian_Alexander_Pereira_Otero	on:Justiciabilidad_de_los_Derechos_Sociales_Económicos_y_Culturales_en_el_marco_de_los_Nuevos_Constitucionalismos_Latinoamericanos.
4 on:Gabriel_Vicente_Obando_Guerrero	on:Learners_satisfaction_and_Language_Teacher_Education_programs_at_Universidad_de_Nariño
5 on:Sandra_Calvachi_Arciniegas	on:Memoria_Sonora_del_Centro_Histórico_de_Pasto,_2018-2019

Fuente: Creación propia

En la figura 20 se observa un fragmento del resultado de la consulta SPARQL para la relación estudiante_es_autor_pi que se refiere a los estudiantes que son autores de diversos proyectos de investigación.

Figura 20. Resultado consulta SPARQL para la relación estudiante_es_autor_pi

Estudiante	Proyecto_investigacion
1 on:GUSTAVO_MILCIADES_ALVAREZ_MELO	on:COMPARACIÓN_DE_ANÁLISIS_Y_DISEÑO_ESTRUCTURAL_DE_EDIFICACIONES_DE_CONCRETO_REFORZADO_BAJO LAS_NORMAS_COLOMBIANA_Y_ECUATORIANA_VIGENTES.
2 on:José_Andrés_Córdoba_Mora	on:Ensambaje_de_Abejas_en_Bosque_Seco_Tropical_(Bs-T)_de_los_Municipios_de_Chachagüí,_San_Lorenzo,_El_Tambo_y_Taminango,_Nariño,_Colombia.
3 on:William_Geovanny_Imbacuán_Tutistar	on:MAKIPURAY:_UNA_HERRAMIENTA_WEB_DE_INTELIGENCIA_DE_NEGOCIOS_PARA_EL_DESCUBRIMIENTO_DE_CONOCIMIENTO_ORIENTADA_A LAS_PYME_DEL_DEPARTAMENTO_DE_NARIÑO
4 on:Johnnatan_David_Bastidas_Obando	on:EVALUACIÓN_DE_LA_REACCIÓN_DE_ALFA-ALQUILACIÓN_DE_LA-L-PROLINA_CON_HALOGENUROS_DE_BENCILo_CoMo: BROMURO_DE_BENCILo_CLORURO_DE_4-METOXBENCILo_Y_CLORURO_DE_4-METILBENCILo
5 on:Monica_Paola_De_La_Cruz_Caicedo	on:LA_VACUNACIÓN_CoMo_ESTRATEGIA_PARA_CONTROLAR_LA_INCIDENCIA_DE_LA_INFLUENZA_A_EN_EL_MUNICIPIO_DE_SAN_JUAN_DE_PASTO.

Fuente: Creación propia

En la figura 21 se observa un fragmento del resultado de la consulta SPARQL para la relación convocatoria_tiene_pi que se refiere a que determinada convocatoria tiene un proyecto de investigación financiado.

Figura 21. Resultado consulta SPARQL para la relación convocatoria_tiene_pi

Convocatoria	Proyecto_investigacion
1 on:Docente_2012	on:La_dimensión_socio-política_en_los_procesos_de_creación_artística_en_la_región,_realizados_por_egresados_del_programa_de_Artes_visuales_de_la_Universidad_de_Nariño.
2 on:Docente_2014	on:Canales_de_animación_sociocultural_para_activar_procesos_de_resiliencia_comunitaria_frente_al_fenómeno_de_violencia_barrial_en_la_comuna_10_del_municipio_de_Pasto
3 on:Docente_2017	on:Justiciabilidad_de_los_Derechos_Sociales_Económicos_y_Culturales_en_el_marco_de_los_Nuevos_Constitucionalismos_Latinamericanos.
4 on:Docente_2012	on:Learners_satisfaction_and_Language_Teacher_Education_programs_at_Universidad_de_Nariño
5 on:Trabajos_de_Grado_2017	on:COMPARACIÓN_DE_ANÁLISIS_Y_DISEÑO_ESTRUCTURAL_DE_EDIFICACIONES_DE_CONCRETO_REFORZADO_BAJO LAS_NORMAS_COLOMBIANA_Y_ECUATORIANA_VIGENTES.

Fuente: Creación propia

6.7.2 Desarrollo con Machine Learning. En esta fase se destacan los resultados producto del entrenamiento con el algoritmo de Machine Learning con Procesamiento de Lenguaje Natural como es Word2Vec, el cual ayudó a encontrar el contexto que una palabra tiene, además se entrenó un modelo con el algoritmo de Doc2Vec, el cual se basa en Word2Vec para encontrar documentos que se relacionan entre sí, estos modelos hacen uso de redes neuronales. En este caso se entrenó el modelo con los algoritmos mencionados previamente en base al modelo Skip-Gram, el cual intenta predecir las palabras o documentos en contexto dada una palabra o un conjunto de palabras base a buscar.

Cabe aclarar que la salida que retornó Word2Vec fue la entrada para el proceso realizado con Doc2Vec, esto es posible dado que ambos algoritmos trabajan de la mano para lograr descubrir relaciones semánticas y recuperar semánticamente la información de manera efectiva.

Para realizar la búsqueda de similitud entre palabras o documentos, de un conjunto de palabras dadas, se utilizó la biblioteca Gensim, la cual hace uso de la normalización de los vectores obtenidos a partir de la palabras a buscar y el cálculo del producto punto entre el vector normalizado y cada uno de los vectores correspondiente a cada palabra o documento entrenado.

Se creó el modelo con los datos de la etapa de preparación de proyectos de investigación, se asignaron los respectivos hiperparámetros, se entrenó dicho modelo, se evaluaron los resultados y se retroalimentó ajustando los hiperparámetros hasta obtener resultados satisfactorios, como se puede evidenciar en la tabla 14.

Tabla 14. Hiperparámetros para los algoritmos Word2Vec y Doc2Vec

Nombre	Valor	Descripción
vector_size	300	Dimensión del vector de cada una de las palabras del corpus.
window	5	Refiere al contexto donde se elige la distancia entre palabras predichas.
min_count	1	Mínimo de palabras a buscar.
dm	0	0 indica que se usa PV-DBOW de Doc2Vec que es análogo al modelo de Skip-Gram usado en Word2Vec. 1 indica que se usa PV-DM de Doc2Vec que es análogo al modelo de CBOW usado en Word2Vec.
dbow_word	1	0 indica que se va a entrenar con Doc2Vec. 1 indica que se va a entrenar con Doc2Vec teniendo de insumo a Word2Vec.
s		

Nombre	Valor	Descripción
hs	0	Es el valor con el que se va a castigar a la neurona en caso de que la tarea efectuada no sea la requerida.
negative	20	Número de palabras irrelevantes para el muestreo negativo.
ns_exponent	-0.5	Indica que se van a muestrear frecuencias por igual.
alpha	0.015	Tasa de aprendizaje de la red neuronal.
min_alpha	0.0001	Tasa que se reducirá durante el entrenamiento.
seed	25	Semilla para generar hash para palabras.
sample	5	Número de reducción para palabras con alta frecuencia
epochs	150	Épocas, número de iteraciones para el entrenamiento.

Fuente: Creación propia

En las siguientes figuras se puede observar el resultado de ejecutar la orden de encontrar 10 de las palabras más similares y relacionadas (según la similitud coseno del algoritmo ordenadas porcentualmente de mayor a menor) a otra palabra que se especifica dentro de todo el corpus investigativo con un método del algoritmo Word2Vec. Es importante aclarar que la palabra que se especifica no está con tilde puesto que las búsquedas realizadas fueron dentro del corpus y este como se explicó anteriormente fue preprocesado en su totalidad.

La figura 22 indica las 10 palabras más similares y relacionadas a la palabra “neuronales”.

Figura 22. Resultado de método con Word2Vec para palabra “neuronales”

```
modelo_cargado.wv.most_similar(
    positive=['neuronales'], topn=10)

[('clasificar', 0.8764573931694031),
 ('volcano', 0.7713260650634766),
 ('sismos', 0.7701717615127563),
 ('inalambricas', 0.7511094212532043),
 ('artificiales', 0.7483046650886536),
 ('tectonicos', 0.7314063310623169),
 ('deep', 0.7180562019348145),
 ('milimetrica', 0.6849448680877686),
 ('recursivos', 0.6747711896896362),
 ('pragmatico', 0.6735430955886841)]
```

Fuente: Creación propia

La figura 23 indica las 10 palabras más similares y relacionadas a la palabra “carnaval”.

Figura 23. Resultado de método con Word2Vec para palabra “carnaval”

```
modelo_cargado.wv.most_similar(  
    positive=['carnaval'], topn=10)  
  
[('negros', 0.8737403154373169),  
 ('blancos', 0.8074564337730408),  
 ('promotor', 0.8038805723190308),  
 ('organizador', 0.7975124716758728),  
 ('corpocarnaval', 0.7947887182235718),  
 ('alegoricos', 0.7947543859481812),  
 ('poliesrireno', 0.7759032845497131),  
 ('muñoz', 0.7492727041244507),  
 ('hacedores', 0.7245036363601685),  
 ('gesta', 0.7210983633995056)]
```

Fuente: Creación propia

La figura 24 indica las 10 palabras más similares y relacionadas a la palabra “cultivos”.

Figura 24. Resultado de método con Word2Vec para palabra “cultivos”

```
modelo_cargado.wv.most_similar(  
    positive=['cultivos'], topn=10)  
  
[('andinos', 0.5301966667175293),  
 ('sustituir', 0.5119062662124634),  
 ('agrotecnologias', 0.4994411766529083),  
 ('totipotencia', 0.49643680453300476),  
 ('cebada', 0.49597498774528503),  
 ('invitro', 0.49116745591163635),  
 ('transitorios', 0.4897231459617615),  
 ('hechas', 0.48805299401283264),  
 ('potencializador', 0.4799562096595764),  
 ('recesivo', 0.47675687074661255)]
```

Fuente: Creación propia

La figura 25 indica las 10 palabras más similares y relacionadas a la palabra “fresa”.

Figura 25. Resultado de método con Word2Vec para palabra “fresa”

```
modelo_cargado.wv.most_similar(  
    positive=['fresa'], topn=10)  
  
[('cereza', 0.9345278739929199),  
 ('citricos', 0.9311402440071106),  
 ('ciruela', 0.9139297604560852),  
 ('pera', 0.8887712359428406),  
 ('yogurt', 0.7925349473953247),  
 ('banano', 0.7829478979110718),  
 ('manzana', 0.7650518417358398),  
 ('subtropico', 0.6739339828491211),  
 ('canada', 0.6399896144866943),  
 ('usado', 0.6375434994697571)]
```

Fuente: Creación propia

La figura 26 indica las 10 palabras más similares y relacionadas a la palabra “psicología”.

Figura 26. Resultado de método con Word2Vec para palabra “psicología”

```
modelo_cargado.wv.most_similar(  
    positive=['psicologia'], topn=10)  
  
[('conepsi', 0.773409366607666),  
 ('psicologico', 0.7477157115936279),  
 ('cuasi', 0.7277992367744446),  
 ('subjetividad', 0.6863089203834534),  
 ('derivacion', 0.6861633658409119),  
 ('terceros', 0.6721749305725098),  
 ('pensadores', 0.6713626980781555),  
 ('educacionales', 0.664987325668335),  
 ('diversidades', 0.6458945274353027),  
 ('psicosociales', 0.6441888213157654)]
```

Fuente: Creación propia

La figura 27 indica las 10 palabras más similares y relacionadas a la palabra “historia”.

Figura 27. Resultado de método con Word2Vec para palabra “historia”

```
modelo_cargado.wv.most_similar(
    positive=['historia'], topn=10)

[('guerreras', 0.5881358981132507),
 ('feminista', 0.5563334822654724),
 ('musicologia', 0.5516680479049683),
 ('diversion', 0.5437859296798706),
 ('empoderarse', 0.5405762195587158),
 ('juventud', 0.5386302471160889),
 ('constatado', 0.5351422429084778),
 ('enhem', 0.5351074934005737),
 ('amor', 0.5341321229934692),
 ('resignificacion', 0.53216552734375)]
```

Fuente: Creación propia

La figura 28 indica las 10 palabras más similares y relacionadas a la palabra “matematicas”.

Figura 28. Resultado de método con Word2Vec para palabra “matemáticas”

```
modelo_cargado.wv.most_similar(
    positive=['matematicas'], topn=10)

[('gescas', 0.7219972014427185),
 ('club', 0.6565420031547546),
 ('clubes', 0.6318238973617554),
 ('amena', 0.619117021560669),
 ('objetivacion', 0.6181855201721191),
 ('estadistica', 0.6073135137557983),
 ('olimpiadas', 0.6033143997192383),
 ('apropiar', 0.6029617786407471),
 ('erm', 0.5898107886314392),
 ('biomatematicas', 0.5882513523101807)]
```

Fuente: Creación propia

La figura 29 indica las 10 palabras más similares y relacionadas a la palabra “medicos”.

Figura 29. Resultado de método con Word2Vec para palabra “medicos”

```

modelo_cargado.wv.most_similar(
    positive=['medicos'], topn=10)

[('remision', 0.8112903237342834),
 ('reumatologos', 0.775713324546814),
 ('subdiagnostico', 0.7463810443878174),
 ('precoz', 0.7410680055618286),
 ('especialista', 0.7363733649253845),
 ('reumaticas', 0.7290210127830505),
 ('generales', 0.724345862865448),
 ('comprobacion', 0.6957311630249023),
 ('spa', 0.6793832778930664),
 ('nula', 0.6671884655952454)]

```

Fuente: Creación propia

La figura 30 indica las 10 palabras más similares y relacionadas a la palabra “redes”.

Figura 30. Resultado de método con Word2Vec para palabra “redes”

```

modelo_cargado.wv.most_similar(
    positive=['redes'], topn=10)

[('inalambricas', 0.7252240180969238),
 ('niebla', 0.6841256618499756),
 ('neuronales', 0.660290002822876),
 ('configuraban', 0.658789873123169),
 ('clientelares', 0.6412293314933777),
 ('plausible', 0.6383322477340698),
 ('justo', 0.6135219931602478),
 ('parece', 0.6127933859825134),
 ('potencializan', 0.603103756904602),
 ('bancos', 0.5799424052238464)]

```

Fuente: Creación propia

La figura 31 indica las 10 palabras más similares y relacionadas a la palabra “nariño”.

Figura 31. Resultado de método con Word2Vec para palabra “nariño”

```
modelo.wv.most_similar(  
    positive=['nariño'], topn=10)  
  
[('universidad', 0.785207211971283),  
 ('departamento', 0.7190582156181335),  
 ('amazonas', 0.40693867206573486),  
 ('cosechan', 0.3996794819831848),  
 ('cauca', 0.39474451541900635),  
 ('tolima', 0.39220744371414185),  
 ('situado', 0.38814577460289),  
 ('huila', 0.3837127089500427),  
 ('renovable', 0.3831225633621216),  
 ('velan', 0.38034915924072266)]
```

Fuente: Creación propia

La figura 32 indica las 10 palabras más similares y relacionadas a la palabra “pasto”.

Figura 32. Resultado de método con Word2Vec para palabra “pasto”

```
modelo_cargado.wv.most_similar(  
    positive=['pasto'], topn=10)  
  
[('ciudad', 0.5812232494354248),  
 ('municipio', 0.5763037204742432),  
 ('juan', 0.49123457074165344),  
 ('gerentes', 0.4473351836204529),  
 ('subzona', 0.43632400035858154),  
 ('subzonas', 0.4348103404045105),  
 ('piedras', 0.43003690242767334),  
 ('rio', 0.4298369288444519),  
 ('jamondino', 0.42734017968177795),  
 ('acontecen', 0.42417120933532715)]
```

Fuente: Creación propia

En las figuras 33, 34, 35 y 36 se pueden observar los cuatro cuadrantes en donde se pueden visualizar las palabras del corpus de proyectos de investigación que incluyen proyectos docentes, proyectos estudiantiles y trabajos de grado. Entre más cercanas en

distancia están las palabras, es más posible que se relacionen de manera semántica o en un contexto determinado, entonces recordando que se está en el ámbito universitario puede existir gran cantidad de temas de cualquier programa de formación. Es de mencionar que para efectos de visualización en el presente proyecto se visualiza un total de aproximadamente 200 palabras para las figuras antes mencionadas, pese a que no son muchas palabras se logra encontrar relaciones de calidad. Para más palabras la relación es evidente, pero dada la complejidad en cuestiones de visualización, sólo se observan 200 palabras. Estas palabras al igual que las anteriores (que hacían uso del algoritmo de Word2Vec) tampoco están con tildes dado que fueron preprocesadas en la etapa de preparación de datos.

En la figura 33 se puede observar posibles relaciones semánticas entre las palabras:

- ✓ derecho – laboral
- ✓ trabajo – reconocimiento
- ✓ personas – principios

En la figura 34 se puede observar posibles relaciones semánticas entre las palabras:

- ✓ entidades - instituciones - academicas - ministerio - promover
- ✓ impacto - condiciones
- ✓ empresas - comercio - turismo – emprendimiento

En la figura 35 se puede observar posibles relaciones semánticas entre las palabras:

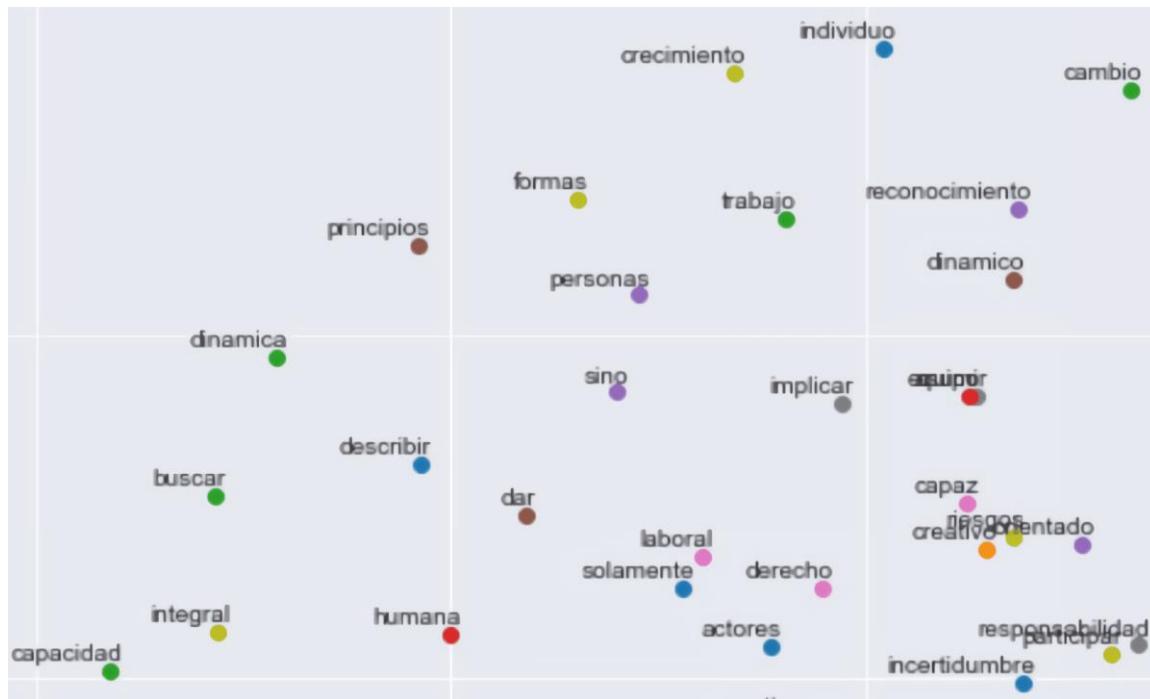
- ✓ adolescentes - escolares - escolar - bullying
- ✓ nivel - mundial
- ✓ aportar - contribuir
- ✓ relaciones - económicos
- ✓ social – sociales

En la figura 36 se puede observar posibles relaciones semánticas entre las palabras:

- ✓ lado - humano
- ✓ programa - diseño - industrial - artes - tecnologicos
- ✓ investigacion - curriculo
- ✓ desarrollo - proyecto - estudios - docente - grado
- ✓ universidad – Nariño

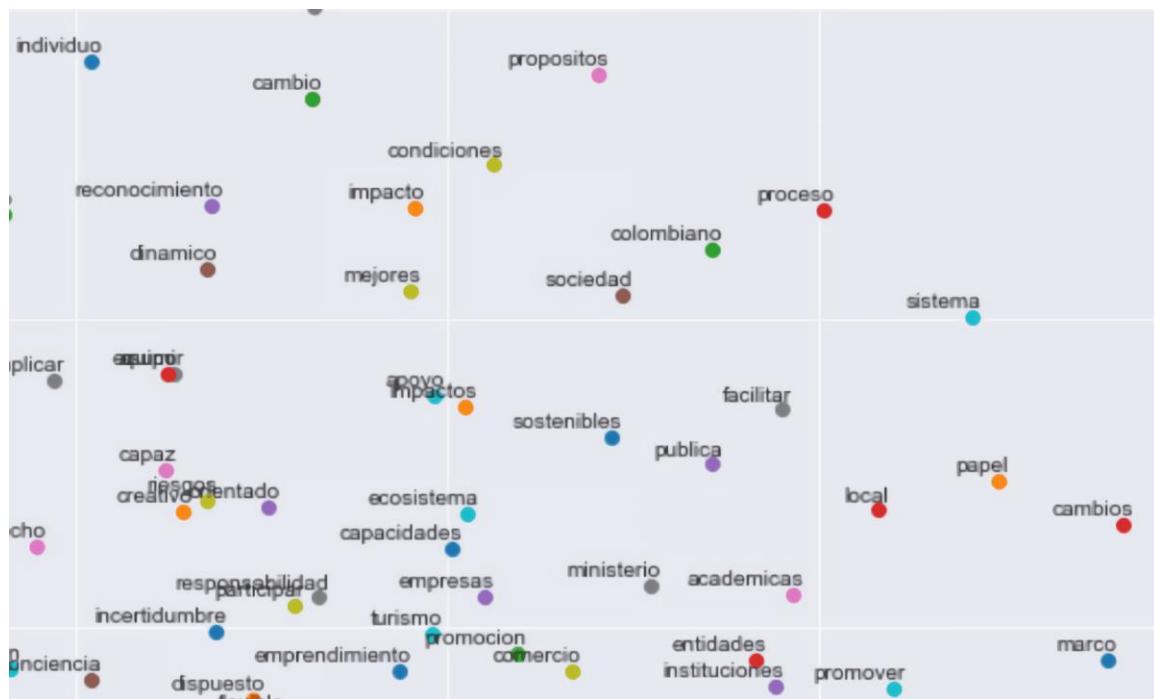
✓ desarrollar – propuesta

Figura 33. Primer cuadrante palabras semánticamente relacionadas



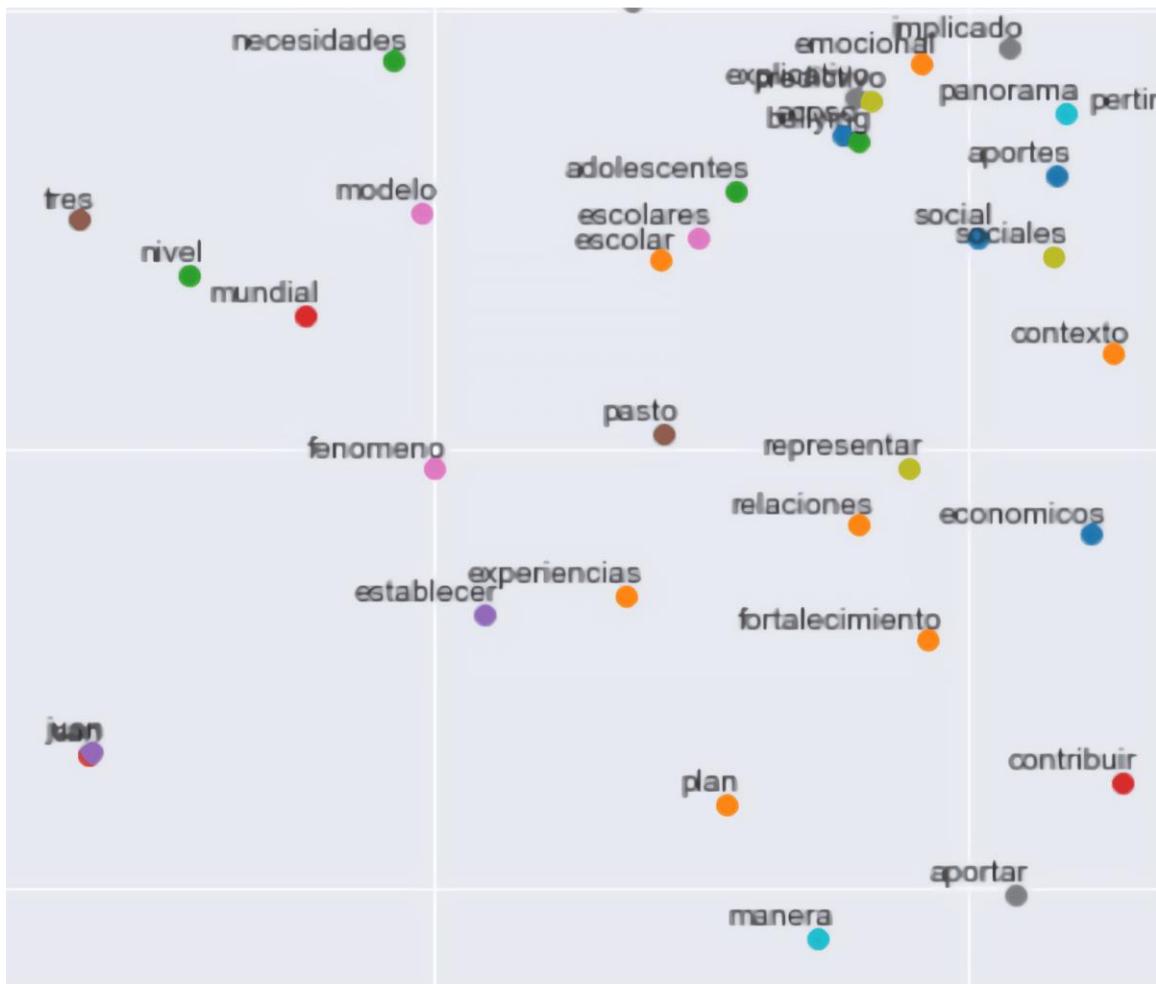
Fuente: Creación propia

Figura 34. Segundo cuadrante palabras semánticamente relacionadas



Fuente: Creación propia

Figura 35. Tercer cuadrante palabras semánticamente relacionadas



Fuente: Creación propia

Figura 36. Cuarto cuadrante palabras semánticamente relacionadas



Fuente: Creación propia

6.7.3 Desarrollo unificado de Ontología y Machine Learning. Para esta fase se integra la Ontología y el Machine learning brindando potencia, efectividad y poder semántico para optimizar tiempos, recursos, y para tener mayores posibilidades de encontrar resultados exitosos y satisfactorios a determinadas búsquedas en THAQHAÑA.

Esto se logró llevando los vectores que Doc2Vec generó a Elasticsearch; este último ayudó en los procesos de rankeo al tener velocidad, escalabilidad y al ser un motor de análisis distribuido que favorece en la búsqueda y en la indexación de los proyectos de investigación.

Después se crearon scripts para gestionar las consultas de los proyectos de investigación para la Ontología con SPARQL la cual se apoya del modelo Word2Vec entrenado para agregar a la búsqueda palabras adicionales que se relacionan a las solicitadas y así encontrar investigaciones relacionadas a determinada consulta. De igual forma con Doc2Vec se logró inferir vectores a partir de un conjunto de palabras suministradas, luego como resultado parcial se presentan las investigaciones que se relacionan a dichos vectores inferidos. Finalmente se hace la unión de los resultados obtenidos en la consulta

SPARQL y el algoritmo Doc2Vec, así el ranking final de una búsqueda mostrará resultados acordes, coherentes, exitosos y satisfactorios a lo solicitado con la capacidad adicional de recomendar documentos que pueden ser de utilidad e interés para el usuario.

Además se construye el gestor de búsquedas y el gestor de resultados, que como se observó en la fase de arquitectura, estos son integrados y están constantemente interactuando en la capa de lógica con las capas de presentación y de datos.

6.8 IMPLEMENTACIÓN Y DESPLIEGUE DEL MOTOR DE BÚSQUEDA SEMÁNTICO

El motor de búsqueda semántico ya está implementado y desplegado, para ello se utilizó Flask, manejando el patrón de arquitectura de software: MVC (Modelo Vista Controlador) en donde se realizaron los modelos soportados con PostgreSQL, los controladores en lenguaje de Flask para finalmente realizar las vistas con HTML y CSS de la mano de Bootstrap Studio. Esto se puede observar en la figura 37.

Figura 37. Implementación de Thaqhaña: motor de búsqueda semántico



Fuente: Creación propia

6.9 REALIZACIÓN DE PRUEBAS FINALES Y EVALUACIÓN DEL RENDIMIENTO DEL MOTOR DE BÚSQUEDA SEMÁNTICO

En esta etapa se describen las pruebas aplicadas al motor de búsqueda cuyo objetivo fue el de medir si el motor de búsqueda encuentra resultados satisfactorios a diversas

búsquedas y si lo hace de manera efectiva siguiendo determinados lineamientos. Así las cosas, se definieron 23 tipos de casos de prueba, los cuales se califican como E (exito) o F (fracasado) según el resultado obtenido en el proceso. Se realizaron 10 pruebas unitarias para comprobar el correcto funcionamiento del motor de búsqueda y sus componentes, la calificación para cada prueba se categorizó entre E (exito) o F (fracasado). También se presentan 10 pruebas de aceptación que van asociadas a cada prueba unitaria, la calificación para cada prueba se categorizó entre E (exito) o F (fracasado).

Las pruebas de aceptación fueron definidas en las historias de usuario en el análisis de Scrum (Ver anexo A).

6.9.1 Casos de prueba. Se definieron 23 tipos de casos de prueba, los cuales son calificados como E (Exitoso) o F (Fracasado) según el resultado obtenido en el proceso. De estos casos de prueba, se presentan los 3 más destacados, los demás se anexan (ver anexo E.)

- ✓ **Caso de prueba 1:** El objetivo de esta prueba fue verificar el correcto funcionamiento del motor de búsqueda cuando se realizan diferentes consultas, como la mencionada en la tabla 15.

Tabla 15. Caso de prueba 1

Búsqueda realizada	“Proyectos de investigación sobre psicología”
Resultado búsqueda	<p>Búsqueda de Documentos</p> <p>1 Propiedades psicométricas de una prueba para medir el pensamiento crítico a estudiantes de pregrado de Colombia, México y Chile.</p> <p>2 Comportamiento proecológico y austero en el personal administrativo, en el marco del Sistema de Gestión Ambiental de la Universidad de Nariño</p> <p>3 Efectos de un programa de intervención basado en el debate crítico sobre el pensamiento crítico de los estudiantes de psicología de la Universidad de Nariño</p> <p>4 Empoderamiento comunitario de los factores psicosociales asociados al proceso de participación estudiantil en la universidad de Nariño sede Pasto.</p> <p>5 Estrategias de aprendizaje mediado para el desarrollo de la modicabilidad cognitiva y la disminución de la disruptión escolar en adolescentes de 13 a 16 años en situación de deprivación sociocultural de la Institución Educativa San Juan Bosco de Pasto - Colombia</p>
Calificación	E

Fuente: Creación propia

- ✓ **Caso de prueba 2:** El objetivo de esta prueba fue verificar el correcto funcionamiento del motor de búsqueda cuando se realizan diferentes consultas, como la mencionada en la tabla 16.

Tabla 16. Caso de prueba 2

Búsqueda realizada	“Proyectos de investigación sobre minería de datos”
Resultado búsqueda	<p>Búsqueda de Documentos</p> <p>1 APLICACIÓN DE LA MINERÍA DE DATOS EN LA DETECCIÓN DE PATRONES DE DESEMPEÑO EN LAS COMPETENCIAS GENÉRICAS DE LAS PRUEBAS SABER PRO 2012, 2013 y 2014 DE LOS ESTUDIANTES DE LOS PROGRAMAS PROFESIONALES DE LA UNIVERSIDAD DE NARIÑO</p> <p>2 CONSTRUCCIÓN DE UN REPOSITORIO LIMPIO DE DATOS PARA LA DETECCIÓN DE PATRONES DE EVENTOS ERUPTIVOS DEL VOLCÁN GALERAS CON TÉCNICAS DE MINERÍA DE DATOS</p> <p>3 APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA EL DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 5º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO.</p> <p>4 Desarrollo de una metodología de visualización interactiva y eficaz de información en Big Data.</p> <p>5 DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 11º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO CON TÉCNICAS DESCRIPTIVAS DE MINERÍA DE DATOS</p>
Calificación	E

Fuente: Creación propia

- ✓ **Caso de prueba 3:** El objetivo de esta prueba fue verificar el correcto funcionamiento del motor de búsqueda cuando se realizan diferentes consultas, como la mencionada en la tabla 17.

Tabla 17. Caso de prueba 3

Búsqueda realizada	“Proyectos de investigación sobre física”
Resultado búsqueda	 <p>1 Plan de Mejoramiento y Actualización Pedagógica de los Laboratorios de Física II, Física Fundamental II y Física III del Programa de Física de la Universidad de Nariño</p> <p>2 Inclusión de la Universidad de Nariño a la colaboración LAGO (Large Aperture Gamma Ray Burst Observatory) a través del Diseño, Construcción, e Instalación de un Tanque - Detector Cherenkov.</p> <p>3 Aplicación de funciones de Green a problemas con valores de frontera asociados con ecuaciones diferenciales ordinarias lineales con coeficientes variables no homogéneas de segundo orden</p> <p>4 Revisión del estado actual del curso Termodinámica y su disponibilidad de equipos de laboratorio.</p> <p>5 Estudio de las anomalías del experimento Large Hadron Collider beauty (LHCb) y los modelos mínimos para resonancias pesadas neutras</p>
Calificación	E

Fuente: Creación propia

6.9.2 Pruebas unitarias. Se plantearon 10 pruebas unitarias para comprobar el exitoso desempeño de cada funcionalidad presente en el motor de búsqueda semántico, la calificación para cada prueba se categorizó entre E (exitsa) o F (fracasada). En la tabla 38 se muestra la recopilación de las 10 pruebas unitarias creadas.

Tabla 18. Pruebas Unitarias

Id prueba	Descripción	Calificación
PU-001 Búsqueda de proyectos docentes.	Comprueba que el ingreso y el procesamiento de la búsqueda junto con la visualización de la respuesta sobre los proyectos docentes se realicen correctamente.	E
PU-002 Búsqueda de proyectos estudiantiles.	Comprueba que el ingreso y el procesamiento de la búsqueda junto con la visualización de la respuesta sobre los proyectos estudiantiles se realicen correctamente.	E
PU-003	Comprueba que el ingreso y el procesamiento de la búsqueda junto con la visualización de la respuesta	E

Id prueba	Descripción	Calificación
Búsqueda de trabajos de grado.	sobre los trabajos de grado se realicen correctamente.	
PU-004 Descarga de proyectos docentes.	Comprueba que se descarguen los proyectos docentes de manera local correctamente.	E
PU-005 Visualización en línea de proyectos docentes.	Comprueba que se visualicen en línea los datos más relevantes de los proyectos docentes.	E
PU-006 Descarga de proyectos estudiantiles.	Comprueba que se descarguen los proyectos estudiantiles de manera local correctamente.	E
PU-007 Visualización en línea de proyectos estudiantiles.	Comprueba que se visualicen en línea los datos más relevantes de los proyectos estudiantiles.	E
PU-008 Descarga de trabajos de grado.	Comprueba que se descarguen los trabajos de grado de manera local correctamente.	E
PU-009 Visualización en línea de trabajos de grado.	Comprueba que se visualicen en línea los datos más relevantes de los trabajos de grado.	E
PU-0010 Usuario administrador	Comprueba que el administrador tenga todos los permisos para gestionar usuarios, datos y procesos dentro de la aplicación.	E

Fuente: Creación propia

6.9.3 Pruebas de aceptación. Se presentan 10 pruebas de aceptación que van asociadas a cada prueba unitaria, la calificación para cada prueba se categorizó entre E (exitosa) o F (fracasada). Por cada una de estas pruebas se especificaron preguntas sobre funciones específicas que se puede realizar en el motor. Estas pruebas de aceptación se definieron desde las historias de usuario que se encuentran en el análisis de Scrum (Ver anexo A).

En las tablas: 39, 40, 41, 42, 43, 44, 45, 46, 47 y 48 se visualizan las diferentes pruebas de aceptación.

Tabla 19. Pruebas de aceptación para PU-001 Búsqueda de proyectos docentes

Prueba de aceptación	Descripción de prueba	Calificación
Validación de datos.	¿Se valida el ingreso de datos nulos? ¿Se valida que la longitud máxima de caracteres para búsqueda ingresada no exceda 50 caracteres?	E
Búsqueda de proyecto de investigación sobre un proyecto docente.	¿Muestra el resultado de búsqueda que se desea sobre proyectos docentes? Cuando en el resultado de la búsqueda no se encuentran los proyectos docentes, ¿se muestran proyectos cercanamente parecidos o relacionados?	E
	¿Muestra mensaje de error si no encontró resultados exactos o cercanamente parecidos o relacionados?	E
	Al hacer nueva búsqueda de proyectos docentes, ¿se actualizan adecuadamente los resultados?	E

Fuente: Creación propia

Tabla 20. Pruebas de aceptación para PU-002 Búsqueda de proyectos estudiantiles

Prueba de aceptación	Descripción de prueba	Calificación
Validación de datos.	¿Se valida el ingreso de datos nulos?	E
	¿Se valida que la longitud máxima de caracteres	E

Prueba de aceptación	Descripción de prueba	Calificación
Búsqueda de proyecto de investigación sobre un proyecto estudiantil.	para búsqueda ingresada no exceda 50 caracteres? ¿Muestra el resultado de búsqueda que se desea sobre proyectos estudiantiles?	E
	Cuando en el resultado de la búsqueda no se encuentran los proyectos estudiantiles, ¿se muestran proyectos cercanamente parecidos o relacionados?	E
	¿Muestra mensaje de error si no encontró resultados exactos o cercanamente parecidos o relacionados?	E
	Al hacer nueva búsqueda de proyectos estudiantiles, ¿se actualizan adecuadamente los resultados?	E

Fuente: Creación propia

Tabla 21. Pruebas de aceptación para PU-003 Búsqueda de trabajos de grado

Prueba de aceptación	Descripción de prueba	Calificación
Validación de datos.	¿Se valida el ingreso de datos nulos? ¿Se valida que la longitud máxima de caracteres para búsqueda ingresada no exceda 50 caracteres?	E
Búsqueda de proyecto de investigación sobre un trabajo de grado.	¿Muestra el resultado de búsqueda que se desea sobre trabajos de grado? Cuando en el resultado de la búsqueda no se encuentran los trabajos de grado, ¿se muestran trabajos cercanamente	E

Prueba de aceptación	Descripción de prueba	Calificación
	parecidos o relacionados? ¿Muestra mensaje de error si no encontró resultados exactos o cercanamente parecidos o relacionados?	E
	Al hacer nueva búsqueda de trabajos de grado, ¿se actualizan adecuadamente los resultados?	E

Fuente: Creación propia

Tabla 22. Pruebas de aceptación para PU-004 Descarga de proyectos docentes

Prueba de aceptación	Descripción de prueba	Calificación
Descarga de proyecto de investigación sobre un proyecto docente.	¿Se descarga el proyecto docente de manera local?	E
	¿Se descarga en formato PDF para su visualización apropiada?	E
	¿Se descarga el proyecto docente en su totalidad?	E
	¿Se pueden descargar múltiples proyectos docentes?	E
	¿El tiempo de descarga es óptimo?	E

Fuente: Creación propia

Tabla 23. Pruebas de aceptación para PU-005 Visualización en línea de proyectos docentes

Prueba de aceptación	Descripción de prueba	Calificación
Visualización en línea de proyecto de investigación sobre un proyecto docente.	¿Se pueden visualizar los detalles más importantes y relevantes del proyecto docente sin necesidad de descargarlo?	E

Fuente: Creación propia

Tabla 24. Pruebas de aceptación para PU-006 Descarga de proyectos estudiantiles

Prueba de aceptación	Descripción de prueba	Calificación
Descarga de proyecto de investigación sobre un proyecto estudiantil.	¿Se descarga el proyecto estudiantil de manera local?	E
	¿Se descarga en formato PDF para su visualización apropiada?	E
	¿Se descarga el proyecto estudiantil en su totalidad?	E
	¿Se pueden descargar múltiples proyectos estudiantiles?	E
	¿El tiempo de descarga es óptimo?	E

Fuente: Creación propia

Tabla 25. Pruebas de aceptación para PU-007 Visualización en línea de proyectos estudiantiles

Prueba de aceptación	Descripción de prueba	Calificación
Visualización en línea de proyecto de investigación sobre un proyecto estudiantil.	¿Se pueden visualizar los detalles más importantes y relevantes del proyecto estudiantil sin necesidad de descargarlo?	E

Fuente: Creación propia

Tabla 26. Pruebas de aceptación para PU-008 Descarga de trabajos de grado

Prueba de aceptación	Descripción de prueba	Calificación
Descarga de proyecto de investigación sobre un trabajo de grado.	¿Se descarga el trabajo de grado de manera local?	E
	¿Se descarga en formato PDF para su visualización apropiada?	E
	¿Se descarga el trabajo de grado en su totalidad?	E

Prueba de aceptación	Descripción de prueba	Calificación
	¿Se pueden descargar múltiples trabajos de grado?	E
	¿El tiempo de descarga es óptimo?	E

Fuente: Creación propia

Tabla 27. Pruebas de aceptación para PU-009 Visualización en línea de trabajos de grado

Prueba de aceptación	Descripción de prueba	Calificación
Visualización en línea de proyecto de investigación sobre un trabajo de grado.	¿Se pueden visualizar los detalles más importantes y relevantes del trabajo de grado sin necesidad de descargarlo?	E

Fuente: Creación propia

Tabla 28. Pruebas de aceptación para PU-010 Usuario administrador

Prueba de aceptación	Descripción de prueba	Resultado
Gestión de usuarios.	¿Se puede gestionar los usuarios correctamente?	E
Gestión de datos	¿Se puede gestionar los datos correctamente?	E
Gestión de procesos	¿Se puede gestionar los procesos correctamente?	E

Fuente: Creación propia

6.9.4 Pruebas de comparación. Finalmente en esta sección se demuestra la comparación entre THAQHAÑA y el buscador manual. Por ejemplo en figura 38 no se encontraron resultados para la búsqueda sobre proyectos de investigación de “física”, en figura 39 no se encontraron resultados para la búsqueda sobre proyectos de investigación de “amenaza volcánica” y en figura 40 no se encontraron resultados para la búsqueda sobre proyectos de investigación de “álgebra”. Mientras que en los casos de prueba expuestos anteriormente sobre THAQHAÑA, se demostró su eficiencia y efectividad.

Figura 38. Resultados para la búsqueda “física” en el buscador manual

Buscar por

No se encontraron registros asociados a la búsqueda

Fuente: Creación propia

Figura 39. Resultados para la búsqueda “amenaza volcanica” en el buscador manual

Buscar por

No se encontraron registros asociados a la búsqueda

Fuente: Creación propia

Figura 40. Resultados para la búsqueda “algebra” en el buscador manual

Buscar por

No se encontraron registros asociados a la búsqueda

Fuente: Creación propia

La tabla 49 indica la comparación final entre THAQHAÑA y buscador manual mediante variables que funcionan como indicadores clave a modo de pregunta. Se respondió a las preguntas mediante respuesta binaria de SI/NO, donde los resultados arrojan que THAQHAÑA tiene un rendimiento general del 100% mientras que el buscador manual del 18.75%

Tabla 29. Pruebas de comparación entre motor de búsqueda semántico y buscador manual

Variable a medir	THAQHAÑA: Motor de búsqueda semántico	Buscador manual sisinfoviis
¿El número de resultados es pertinente y apropiado?	SI	NO

Variable a medir	THAQHĀNA: Motor de búsqueda semántico	Buscador manual sisinfoviis
¿Los resultados son de calidad?	SI	NO
¿Se encuentra la búsqueda solicitada y cuando no exista se muestra error?	SI	NO
¿El manejo de la herramienta es fácil?	SI	SI
¿El manejo de la herramienta es intuitivo?	SI	SI
¿El manejo de la herramienta es interactivo?	SI	NO
¿Las vistas de los diseños de la herramienta son agradables al usuario?	SI	NO
¿La herramienta es eficiente y efectiva?	SI	NO
¿Los resultados en respuesta a las búsquedas realizadas en la herramienta son satisfactorios?	SI	NO
¿Se permite la búsqueda múltiple en la herramienta?	SI	SI
¿No existen muchas categorías para realizar las búsquedas lo que vuelve tedioso y dispendioso para el usuario?	SI	NO
¿Funcionan todos los botones en la interacción y dinamismo de la herramienta?	SI	NO
¿La herramienta no se demora actualizando o refrescando?	SI	NO

Variable a medir	THAQHĀÑA: Motor de búsqueda semántico	Buscador manual sisinfoviis
¿La herramienta no se demora en la búsqueda?	SI	NO
¿Los espacios en el servidor para la herramienta están actualizados y funcionan adecuadamente?	SI	NO
¿No se abren riesgos potenciales al hacer click en un resultado determinado?	SI	NO
¿La forma de realizar la búsqueda en dicha herramienta es la más óptima?	SI	NO

Fuente: Creación propia

7. DISCUSIÓN

- ✓ THAQHAÑA aporta un grado de optimización, originalidad e innovación frente a la forma en que se realiza las búsquedas actualmente en VIIS y con respecto al estado del arte, porque se integran las ontologías con Machine Learning mediante los scripts mencionados donde la ontología está bien configurada y los algoritmos bien entrenados. Además de esto, los vectores de las palabras se están gestionando con Elasticsearch, lo que ahorra un importante consumo de memoria, la cual además es otra razón de porque el motor de búsqueda es muy rápido y preciso.
- ✓ Las pruebas realizadas en su totalidad para el manejo de búsquedas, interacción con la herramienta mediante descargas y visualizaciones y demás funcionalidades extra fueron exitosas y otorgaron constancia y evidencia de la realización satisfactoria del proyecto.
- ✓ A pesar de que se utilizaron pocos datos que fueron los proyectos de investigación que se encuentran en el Sistema de Investigación de VIIS (porque faltaron todos los proyectos desarrollados a nivel de la Universidad) se obtuvieron buenos resultados en el motor de búsqueda. De este modo, la Ontología y los modelos de Machine learning se comportaron de la forma esperada con pocos datos.

8. CONCLUSIONES

- ✓ Con la culminación de este trabajo de investigación se obtuvo THAQHAÑA, es un motor de búsqueda semántico basado en una Ontología y en un Modelo de Machine Learning para proyectos de investigación de la Universidad de Nariño. Mediante el desarrollo exitoso de las etapas del proyecto se solucionó el problema formulado, se cumplieron los objetivos planteados y se obtuvieron resultados satisfactorios. De este modo, esta herramienta facilitó la búsqueda exitosa de proyectos de investigación que han elaborado docentes y estudiantes en la Universidad de Nariño.
- ✓ En las etapas de apropiación del conocimiento e instalación y configuración de las herramientas, bibliotecas y tecnologías se adquirió un dominio de las diversas temáticas manejadas y esto aparte de contribuir con el desarrollo del proyecto, propendió a la formación personal de los investigadores así como también realizó destacados aportes para el grupo de investigación GRIAS (Grupo de Investigación Aplicado en Sistemas) y para la Universidad de Nariño en general.
- ✓ Las etapas de recolección, extracción y preparación de proyectos de investigación fueron sumamente importantes porque actuaron como antesala a modo de insumo para THAQHAÑA. En este orden de ideas es correcto afirmar que sin estas no se podría haber logrado el desarrollo de THAQHAÑA.
- ✓ Methontology fue la metodología que se integró al proyecto, la cual fue fundamental en la construcción de la Ontología porque permitió ejecutar unas fases y tareas puntuales con un orden, entendimiento y exactitud en los procesos.
- ✓ La Ontología integrada con Machine Learning de la mano de los algoritmos de Elasticsearch evidenciaron la aplicabilidad correcta de la semántica con efectividad porque se obtuvieron resultados concretos y acordes a las búsquedas realizadas optimizando los recursos disponibles en memoria y velocidad. Esto es porque los algoritmos de aprendizaje de máquina, específicamente de procesamiento de lenguaje natural como Word2Vec y Doc2Vec trabajan con redes neuronales, las cuales fueron entrenadas con las palabras del corpus de proyectos de investigación, adaptándolas al contexto y encontrando las diversas relaciones semánticas entre las mismas. Así mismo la Ontología actuó como una gran red semántica cuyas instancias, de la mano de clases, relaciones y atributos, interactuaron bajo el esquema de triples que maneja RDF y que consulta SPARQL para extraer todo el conocimiento del dominio de los proyectos de investigación.

- ✓ Scrum fue una metodología muy útil para el proyecto porque permitió dar un control y manejo de tiempos, y orden de ejecución junto con agilidad en los distintos procesos realizados. Así mismo su facilidad para adaptarse a cambios fue clave en el proyecto dado que en muchas ocasiones se presentaban situaciones externas al proyecto por las cuales se debía otorgar lapsos de tiempo adicional para los sprints.
- ✓ Haber usado Flask para la etapa de implementación y despliegue del motor de búsqueda semántico fue muy efectivo dada su facilidad de manejo, seguridad y adaptabilidad esto es porque Flask al ser un micro framework se enfoca en proporcionar lo necesario para permitir el funcionamiento de cualquier tipo de aplicación en muy poco tiempo.
- ✓ Finalmente se concluye que se realizó exitosamente la publicación de artículos en revistas así como en memorias de eventos, encuentros y capítulos de libro, con ponencias y pósters en los respectivos congresos. Estos valiosos aportes contribuyeron a la formación personal y profesional de los investigadores, contribuyeron al crecimiento del grupo de investigación GRIAS (Grupo de Investigación Aplicado en Sistemas) y dejaron en alto el nombre de la Universidad de Nariño porque expandieron las fronteras del conocimiento, generaron nuevas herramientas útiles e innovadoras y propendieron al cambio para transformar continuamente a la sociedad (Ver Anexo D).

9. RECOMENDACIONES

- ✓ Se recomienda acoplar el motor de búsqueda a nuevos dominios, es decir a otras Universidades, adaptándose a los esquemas o modelos que en estas se tenga.
- ✓ Dado que el Procesamiento de Lenguaje Natural es general y tiene aplicabilidad en muchos campos, sería innovador integrar a esta herramienta otro tipo de documentos o tópicos que no sólo se relacionen al campo investigativo o a proyectos de investigación sino a diversos temas del mundo de la revisión sistemática de la literatura, de este modo se tendrían múltiples motores de búsqueda semántico, uno enfocado para cada tema.
- ✓ Se recomienda incorporar el manejo de UX (User Experience) en el motor de búsqueda, para observar y analizar que tanto ha gustado la herramienta a los usuarios finales.
- ✓ Se recomienda realizar otro trabajo en el que se analizarán los archivos log sobre las búsquedas realizadas en el motor de búsqueda, los cuales graban información de manera secuencial en la base de datos.
- ✓ Se recomienda realizar pruebas con más datos para ver cómo se comporta THAQHAÑA ante un tamaño masivo de la información. Esto se debe a que los datos utilizados fueron todos los proyectos de investigación que se encuentran en el Sistema de Investigación de VIIS, pero todos los proyectos desarrollados a nivel de la Universidad no están dentro de THAQHAÑA.
- ✓ Se propone hacer el acoplamiento de THAQHAÑA en otras universidades y en diversos entornos no sólo académicos, determinando así la estructura de los modelos para la Ontología y Machine Learning con sus posibles variantes.
- ✓ Se sugiere realizar un análisis de lo que buscan los usuarios, analizando los registros de búsquedas, descargas, almacenando todo en la base de datos, para luego aplicar minería de datos con toda la información para posiblemente determinar aspectos como: "¿A qué semestres pertenecen las personas que hacen consultas sobre astronomía?" o "¿De qué edades son las personas que hacen consultas sobre psicología?". Para este futuro trabajo se podría utilizar perfectamente toda la magia que hay detrás de Machine Learning.

- ✓ También se propone incorporar en la página del buscador una vista con su respectiva base de datos que permita calificar y comentar el motor de búsqueda con el fin de observar y analizar cómo los usuarios están calificando a THAQHAÑA, así como realizar sus opiniones y si están satisfechos o no, determinando así la usabilidad de THAQHAÑA.

BIBLIOGRAFÍA

- AMARAL, Carlos, LAURENT, Dominique, MARTINS, André, MENDES, Alfonso & PINTO, Cláudia. Design and Implementation of a Semantic Search Engine for Portuguese. [En línea]. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.4090&rep=rep1&type=pdf>
- ANDREONI, Antonella, BALDACCI Maria, BIAGONI, Stefania, CARLESI, Carlo, CASTELLI, Donatella, PAGANO, Pasquale, PETERS, Carol & PISANI, Serena. The ERCIM Technical Reference Digital Library. En: D-Lib Magazine. Vol. 5. No. 12. (Diciembre, 1999). [En línea]. Disponible en: <http://www.dlib.org/dlib/december99/peters/12peters.html>
- Apache Software Foundation. Apache Jena Fuseki. [En línea]. Disponible en: <https://jena.apache.org/documentation/fuseki2>
- ARAUJO, Joaquín. ¿Qué es Docker? ¿Qué son los contenedores? y ¿Por qué no usar VMs? [En línea]. Disponible en: <https://platzi.com/tutoriales/1432-docker/1484-guia-del-curso-de-docker>
- AUCAPÍÑA, Yolanda & PLAZA, C. Buscador semántico universitario: Caso de estudio Universidad de Cuenca. Cuenca, 2018, 200p. Trabajo de grado (Tesis previa a la obtención del Título de Ingeniero en Sistemas). Universidad de Cuenca. Facultad de Ingeniería. Ingeniería de Sistemas. [En línea]. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/30291>
- BACULIMA, Jhon & CAJAMARCA, Marcelo. Diseño e Implementación de un Repositorio Ecuatoriano de Datos Enlazados Geoespaciales. Cuenca, 2014, 131p. Trabajo de grado (Tesis de Grado previa a la obtención del Título: Ingeniero de Sistemas). Universidad de Cuenca. Facultad de Ingeniería. Ingeniería de sistemas. [En línea]. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/19876>
- BARBERÁ, Consuelo, MILLET, Mercé & TORRES, Emiliano. Estudio del buscador semántico Swoogle. [En línea]. Disponible en: <https://www.uv.es/etomar/trabajos/swoogle/swoogle.pdf>
- BENAVIDES, Mauricio & GUERRERO, Jimmy. UMAYUX: un modelo de software de gestión de conocimiento soportado en una Ontología dinámica débilmente acoplado con un gestor de base de datos. San Juan de Pasto, 2014, 145p. Trabajo de grado (Trabajo de grado presentado como requisito parcial para optar al título de Ingeniero de Sistemas). Universidad de Nariño. Facultad de Ingeniería. Ingeniería de Sistemas. [En línea]. Disponible en: <http://sired.udnar.edu.co/2030>
- BUDHIRAJA, Amar. A simple explanation of document embeddings generated using Doc2Vec. [En línea]. Disponible en: <https://medium.com/@amarbudhiraja/understanding-document-embeddings-of-doc2vec-bfe7237a26da>
- BUSTOS, Gabriel. Prototipo de un sistema de integración de recursos científicos, diseñado para su funcionamiento en el espacio de los datos abiertos enlazados para

mejorar la colaboración, la eficiencia y promover la innovación en Colombia. Bogotá, 2015. Tesis de Maestría. Universidad Nacional de Colombia. Facultad de Ingeniería. Ingeniería de Sistemas e Industrial. [En línea]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/55245>

CAMACHO, María. Incorporación de un buscador semántico en la plataforma LdShake para la selección de patrones educativos. Barcelona, 2013, 76p. Trabajo de grado. Universidad Pompeu Fabra. Escuela Superior Politécnica UPF. Ingeniería de Telemática. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/22172>

CHALLENGER, Ivet, DÍAZ, Yanet & BECERRA, Roberto. El lenguaje de programación Python. En: Ciencias Holguín. Vol. XX. No. 2. (Junio, 2014), p. 1-13. [En línea]. Disponible en: www.redalyc.org/articulo.oa?id=181531232001

CHECA, Diego & ROJAS, Oscar. ONTOLOGÍA PARA LOS SISTEMAS HOLÓNICOS DE MANUFACTURA BASADOS EN LA UNIDAD DE PRODUCCIÓN. En: Revista Colombiana de Tecnologías de Avanzada. Vol. 1. No. 23. (Noviembre, 2013), p. 134-141. [En línea]. Disponible en: http://revistas.unipamplona.edu.co/ojs_viceinves/index.php/RCTA/article/view/2334

CLASSORA. Sacando provecho a la Web Semántica: SPARQL. [En línea]. Disponible en: <http://blog.classora.com/2012/11/05/sacando-provecho-a-la-web-semantica-sparql>

CODINA, Lluís & CRISTÒFOL, Rovira. La Web Semántica. En: Jesús Tramullas (coord.). Tendencias en documentación digital. Guijón: Trea, 2006. p. 9-54. [En línea]. Disponible en: <http://eprints.rclis.org/8899>

DE PEDRO, A. Buscadores Semánticos, para qué sirven. Usos en la AAPP. [En línea]. Disponible en: <http://www.alejandrodepedro.es/buscadores-semanticos-el-paso-al-30>

EDWARDS, Gavin. Machine Learning An Introduction. [En línea]. Disponible en: <https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0>

EFIGENIA, Ana & CANTOR, Sandoval. USO DE ONTOLOGÍAS Y WEB SEMÁNTICA PARA APOYAR LA GESTIÓN DEL CONOCIMIENTO. En: Ciencia e Ingeniería Neogranadina. Vol. 17 No. 2. (Diciembre, 2007), p.111-129. [En línea]. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/2512191.pdf>

Elasticsearch B.V. Elasticsearch. [En línea]. Disponible en: <https://www.elastic.co/es/what-is/elasticsearch>

FAZZINGA, Bettina, GIANFORME, Giorgio, GOTTLÖB, Georg & LUKASIEWICZ, Thomas. Semantic Web Search Based On Ontological Conjunctive Queries. En: SSRN Electronic Journal. [En línea]. Disponible en: https://www.researchgate.net/publication/326473981_Semantic_Web_Search_Based_on_Ontological_Conjunctive_Queries

FLORES, Pedro & PORTILLO, Julio. ELABORACIÓN DE PROPUESTA DE GUÍA DE IMPLEMENTACIÓN DE SCRUM PARA EMPRESA SALVADOREÑA, UN CASO DE ESTUDIO. Antiguo Cuscatlán, 2017, 117p. Trabajo de grado (MAESTRO EN ARQUITECTURA DE SOFTWARE). Universidad Don Bosco. Arquitectura de Software.

- [En línea]. Disponible en: <http://rd.udb.edu.sv:8080/jspui/bitstream/11715/1264/1/documento.pdf>
- FLÓREZ, Héctor. Construcción de ontologías OWL. En: VÍNCULOS. Vol. 4. No. 1. (Diciembre, 2007), p. 19-34. [En línea]. Disponible en: <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/4112>
- GALLO, Manuel, FABRE, Ernesto & GALLO, Manuel. ¿Qué es un buscador? [En línea]. Disponible en: http://media.axon.es/pdf/98234_1.pdf
- GARCÍA, Francisco. Web Semántica y Ontologías. [En línea]. Disponible en: https://www.researchgate.net/publication/267222548_Web_Semantica_y_Ontologias
- IGLESIAS, Daniela, MEJÍA, Omar, NIETO, Julio, SÁNCHEZ, Steven & MORENO, Silvia. Construcción de un buscador ontológico para búsquedas semánticas de proyectos de maestría y doctorado. En: Investigación y Desarrollo en TIC. Vol. 7. No. 1. (Mayo, 2017), p. 7-13. [En línea]. Disponible en: <https://revistas.unisimon.edu.co/index.php/identic/article/view/2501>
- Kit de herramientas de lenguaje natural. [En línea]. Disponible en: <https://www.nltk.org>
- LAMY, Jean. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. En: Artificial Intelligence in Medicine. Vol. 80. (Agosto, 2017), p. 11-28. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0933365717300271>
- LINCOLN, Matthew. Uso de SPARQL para acceder a datos abiertos enlazados. [En línea]. Disponible en: <https://programminghistorian.org/es/lecciones/sparql-datos-abiertos-enlazados>
- LOZANO, Adolfo. Ontologías en la Web Semántica. [En línea]. Disponible en: <http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Ontologias%20en%20la%20Web%20Semantica.pdf>
- MARTÍN, Javier. Swotti buscador de opiniones. [En línea]. Disponible en: <https://loogic.com/swotti-buscador-de-opiniones>
- MORENO, Carlos & SÁNCHEZ, Yakeline. Prototipo de buscador semántico aplicado a la búsqueda de libros de Ingeniería de Sistemas y Computación en la biblioteca Jorge Roa Martínez de la Universidad Tecnológica de Pereira. Pereira, 2012, 66p. Trabajo de grado. Universidad Tecnológica de Pereira. Facultad de Ingenierías: Eléctrica, Electrónica, Física y Ciencias de la Computación. Ingeniería De Sistemas y Computación. [En línea]. Disponible en: <http://repositorio.utp.edu.co/dspace/bitstream/11059/2671/1/0057565M843.pdf>
- MOURIÑO, M. Clasificación multilingüe de documentos utilizando machine learning y la wikipedia. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/tesis?codigo=150295>
- MUÑOZ, José. Introducción a flask. [En línea]. Disponible en: <https://plataforma.josedomingo.org/pledin/cursos/flask/curso/u05/>

NDLTD. Networked Digital Library of Theses and Dissertations. [En línea]. Disponible en: <http://www.ndltd.org>

PEDRAZA, Rafael, CODINA, Lluís & CRISTÒFOL, Rovira. Web semántica y ontologías en el procesamiento de la información documental. En: El profesional de la información. Vol. 16. No. 6. (Noviembre, 2007), p. 569-579. [En línea]. Disponible en: <https://repositori.upf.edu/handle/10230/13141>

PEREZ, Fernando & GRANGER, Brian. Project Jupyter. [En línea]. Disponible en: <https://jupyter.org>

PostgreSQL. [En línea]. Disponible en <https://www.postgresql.org/about/>

ROCCA, Joseph. A simple introduction to Machine Learning. [En línea]. Disponible en: <https://towardsdatascience.com/introduction-to-machine-learning-f41aabc55264>

SHETTY, Badreesh. Natural Language Processing (NLP) for Machine Learning. [En línea]. Disponible en: <https://towardsdatascience.com/natural-language-processing-nlp-for-machine-learning-d44498845d5b>

SHPERBER, Gidi. A gentle introduction to Doc2Vec. [En línea]. Disponible en: <https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e>

Sistema de Información de Investigaciones. [En línea]. Disponible en: <http://sisinfoviis.udnar.edu.co>

SpaCy 101: todo lo que necesita saber. [En línea]. Disponible en: <https://spacy.io/usage/spacy-101>

TABARES, John & JIMÉNEZ, Jovani. Ontología para el proceso evaluativo en la educación superior. En: Revista Virtual Universidad Católica del Norte. Vol. 1. No. 42. (Agosto, 2014), p. 68-79. [En línea]. Disponible en: <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/495>

UMPIÉRREZ, Francisco. SPARQL Interpreter. Las Palmas de Gran Canaria, 2014, 65p. Trabajo de grado (Trabajo Final de Grado en Ingeniería Informática). Universidad de Las Palmas de Gran Canaria. Escuela Ingeniería Informática. Ingeniería Informática. [En línea]. Disponible en: https://nanopdf.com/download/0701044000000000pdf_pdf

VELÁSQUEZ, Torcoroma, PUENTES, Andrés & GUZMÁN, Jaime. Ontologías: una técnica de representación de conocimiento. En: Avances en Sistemas e Informática. Vol. 8. No. 2. (Julio, 2011), p. 211-216. [En línea]. Disponible en: <https://revistas.unal.edu.co/index.php/avances/article/view/26750>

WolframAlpha Computational Intelligence. [En línea]. Disponible en: <https://www.wolframalpha.com>

ANEXOS

ANEXO A. ANÁLISIS SCRUM

En este anexo se presentan los componentes y el análisis producto de usar la metodología ágil Scrum en la etapa de desarrollo del motor de búsqueda. En esta etapa se realizó una adaptación a Scrum donde se tomaron partes que se consideraron relevantes en el desarrollo y estas se integraron con la experiencia del equipo de desarrollo del proyecto para su exitoso desenlace.

HISTORIAS DE USUARIO

Las historias de usuario definen en lenguaje común los requisitos que se identificaron relevantes para el proyecto. Se trata de una descripción básica de las funcionalidades del motor de búsqueda y de cómo estas son percibidas por los usuarios finales. Los aspectos importantes de la historia de usuario son:

- 1 Id historia: Identificador único para cada historia de usuario.
- 2 Título: campo para el nombre de la historia de usuario.
- 3 Descripción: campo para representar en palabras concretas y claras qué es lo que desea el usuario.
- 4 Prioridad: campo para determinar el grado de prioridad de dicha historia de usuario.
- 5 Estimación en tiempo (días): campo que indica el tiempo estimado para cumplir con dicho requerimiento.
- 6 Pruebas de aceptación: campo que indica las pruebas con criterios mínimos para dar por cumplida la historia de usuario.
- 7 Dependencias: campo que indica las historias de usuario de las cuales depende la historia de usuario actual.

Tabla A1. Historias de usuario

ID HIS TO RIA	TÍTULO	DESCRIPCIÓN	PRIORIDA D	ESTIMACIÓN EN TIEMPO (DÍAS)	PRUEBAS DE ACEPTACIÓN	DEPENDENC IAS
HU-001	Búsqueda de proyectos de investigación docentes.	Como quiero realizar búsquedas de proyectos de investigación sobre proyectos docentes.	ALTA	30	✓ Validación de datos. ✓ Búsqueda de proyecto de investigación sobre un proyecto docente.	
HU-002	Búsqueda de proyectos de investigación estudiantiles.	Como quiero realizar búsquedas de proyectos de investigación sobre proyectos estudiantiles.	ALTA	30	✓ Validación de datos. ✓ Búsqueda de proyecto de investigación sobre un proyecto estudiantil.	
HU-003	Búsqueda de proyectos de investigación de trabajos de grado.	Como quiero realizar búsquedas de proyectos de investigación sobre trabajos de grado.	ALTA	30	✓ Validación de datos. ✓ Búsqueda de proyecto de investigación sobre un trabajo de grado.	
HU-004	Descarga de proyectos de investigación docentes.	Como quiero descargar proyectos de investigación sobre proyectos docentes.	MEDIA	10	✓ Descarga de proyecto de investigación sobre un proyecto docente.	HU-001

HU-005	Visualización en línea de proyectos de investigación docentes.	Como usuario quiero visualizar en línea proyectos de investigación sobre proyectos docentes.	MEDIA	10	✓	Visualización en línea de proyecto de investigación sobre un proyecto docente.	HU-001
HU-006	Descarga de proyectos de investigación estudiantiles.	Como usuario quiero descargar proyectos de investigación sobre proyectos estudiantiles.	MEDIA	10	✓	Descarga de proyecto de investigación sobre un proyecto estudiantil.	HU-002
HU-007	Visualización en línea de proyectos de investigación estudiantiles.	Como usuario quiero visualizar en línea proyectos de investigación sobre proyectos estudiantiles.	MEDIA	10	✓	Visualización en línea de proyecto de investigación sobre un proyecto estudiantil.	HU-002
HU-008	Descarga de proyectos de investigación de trabajos de grado.	Como usuario quiero descargar proyectos de investigación sobre trabajos de grado.	MEDIA	10	✓	Descarga de proyecto de investigación sobre un trabajo de grado.	HU-003
HU-009	Visualización en línea de proyectos de investigación de trabajos de grado.	Como usuario quiero visualizar en línea proyectos de investigación sobre trabajos de grado.	MEDIA	10	✓	Visualización en línea de proyecto de investigación sobre un trabajo de grado.	HU-003

HU-010	Usuario administrador	Como administrador . quiero tener control sobre el manejo de los datos, usuarios y procesos del motor de búsqueda.	ALTA	15	✓ Gestión de usuarios. ✓ Gestión de datos. ✓ Gestión de procesos.
---------------	-----------------------	--	------	----	---

Fuente: Creación propia

PILA DEL PRODUCTO

En la pila del producto se registra una lista con las historias de usuario que se desea obtener, ordenadas por la prioridad con se definieron previamente con su tiempo respectivo, de igual forma se contempla el sprint al que está asociado cada historia de usuario.

Tabla A2. Pila del producto

ID HISTORIA	PRIORIDAD	ESTIMACIÓN EN TIEMPO (DÍAS)	ID SPRINT
HU-001	ALTA	30	1
HU-002	ALTA	30	2
HU-003	ALTA	30	3
HU-010	ALTA	15	4
HU-004	MEDIA	10	5
HU-005	MEDIA	10	5
HU-006	MEDIA	10	6
HU-007	MEDIA	10	6
HU-008	MEDIA	10	7
HU-009	MEDIA	10	7

Fuente: Creación propia

SEGUIMIENTO Y ANÁLISIS DE SPRINTS

Los sprints hacen referencia a las iteraciones, ciclos o intervalos de tiempo donde se crea un incremento terminado del producto el cual es potencialmente entregable. Los sprints son trabajados como pequeños proyectos que aportan al proyecto general. Cuentan con una definición de lo que se va a construir, un diseño y un plan flexible que guía el desarrollo del trabajo y del producto resultante.

A continuación se muestra el seguimiento que se realizó para cada uno de los sprints después de la evaluación del equipo de trabajo antes, durante y después del desarrollo. Se evidencia cada sprint asociado a una historia de usuario correspondiente especificando la finalidad de dicho sprint, sus artefactos, el análisis y las decisiones tomadas. Así mismo se evidencia el seguimiento para cada uno de los sprints, se realizaron gráficas basadas en cálculos de días vs horas con funciones matemáticas para horas estimadas (siendo que del total de días de duración estimada del sprint cada día se estimaba trabajar 8 horas) y horas reales del sprint.

Tabla A3. Sprint 1

SPRINT No. 1

FINALIDAD	Permitir búsquedas de proyectos docentes.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-001

ANÁLISIS DEL SPRINT

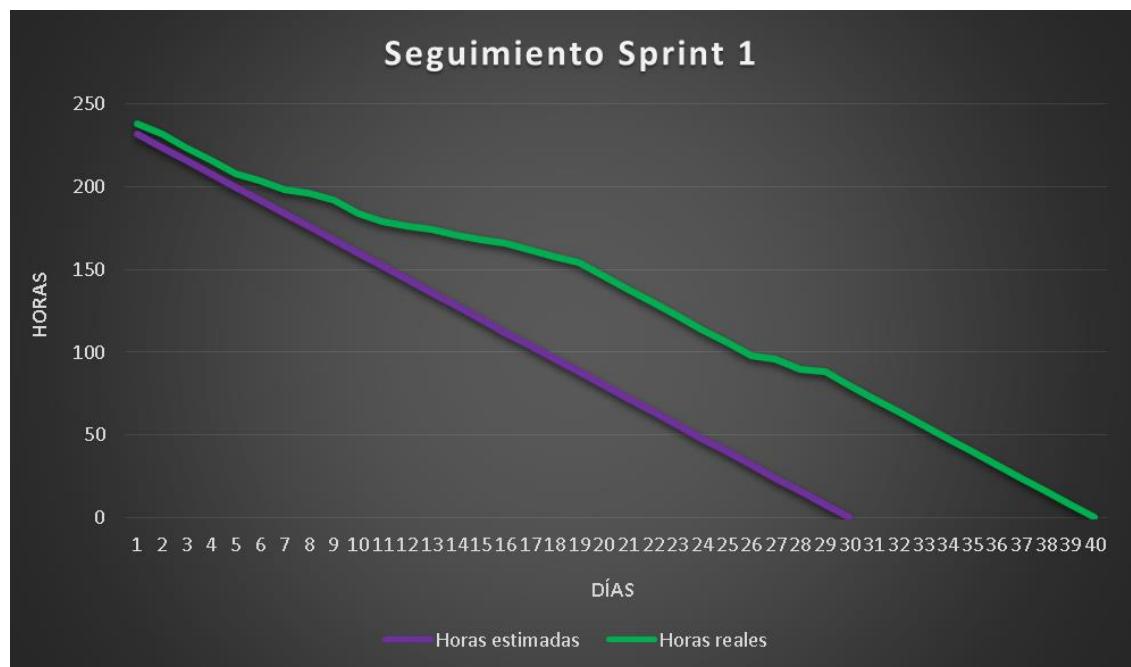
- ✓ **Se observa que las pruebas de aceptación fueron completadas exitosamente.**
- ✓ **Se presentaron situaciones externas al proyecto que no permitieron terminar el sprint en el tiempo estimado.**

DECISIONES

- ✓ **Dar un lapso adicional de tiempo para poder culminar exitosamente el sprint.**
- ✓ **Retroalimentar los procesos del sprint y continuar con el siguiente sprint.**

Fuente: Creación propia

Figura A1. Seguimiento Sprint 1



Fuente: Creación propia

Tabla A4. Sprint 2

SPRINT No. 2

FINALIDAD	Permitir búsquedas de proyectos estudiantiles.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-002

ANÁLISIS DEL SPRINT

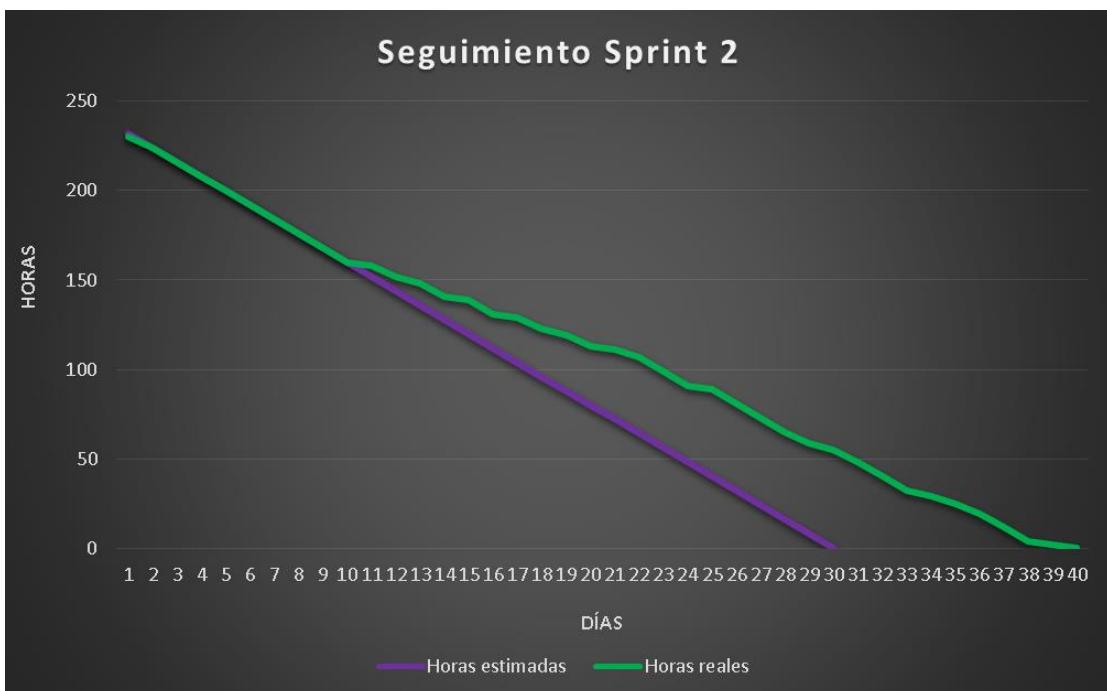
- ✓ **Se observa que las pruebas de aceptación fueron completadas exitosamente.**
- ✓ **Se presentaron situaciones externas al proyecto que no permitieron terminar el sprint en el tiempo estimado.**

DECISIONES

- ✓ **Dar un lapso adicional de tiempo para poder culminar exitosamente el sprint.**
- ✓ **Retroalimentar los procesos del sprint y continuar con el siguiente sprint.**

Fuente: Creación propia

Figura A2. Seguimiento Sprint 2



Fuente: Creación propia

Tabla A5. Sprint 3

SPRINT No. 3	
FINALIDAD	Permitir búsquedas de trabajos de grado.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-003

ANÁLISIS DEL SPRINT

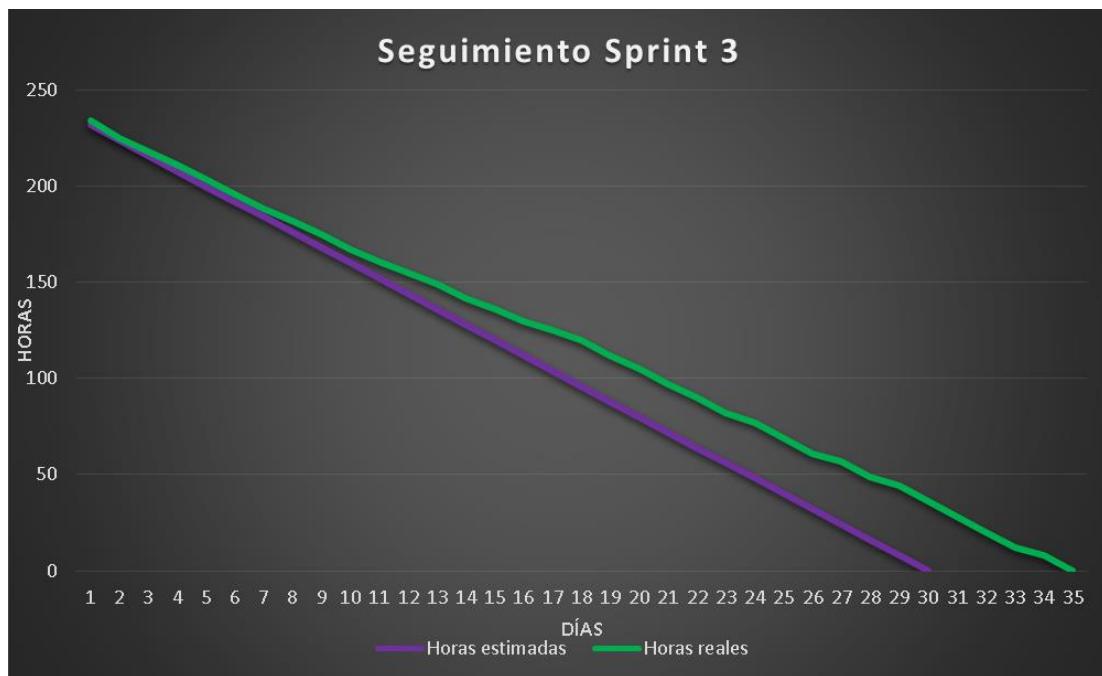
- ✓ **Se observa que las pruebas de aceptación fueron completadas exitosamente.**
- ✓ **Se presentaron situaciones externas al proyecto que no permitieron terminar el sprint en el tiempo estimado.**

DECISIONES

- ✓ **Dar un lapso adicional de tiempo para poder culminar exitosamente el sprint.**
- ✓ **Retroalimentar los procesos del sprint y continuar con el siguiente sprint.**

Fuente: Creación propia

Figura A3. Seguimiento Sprint 3



Fuente: Creación propia

Tabla A6. Sprint 4

SPRINT No. 4

FINALIDAD	Configurar administrador.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-010

ANÁLISIS DEL SPRINT

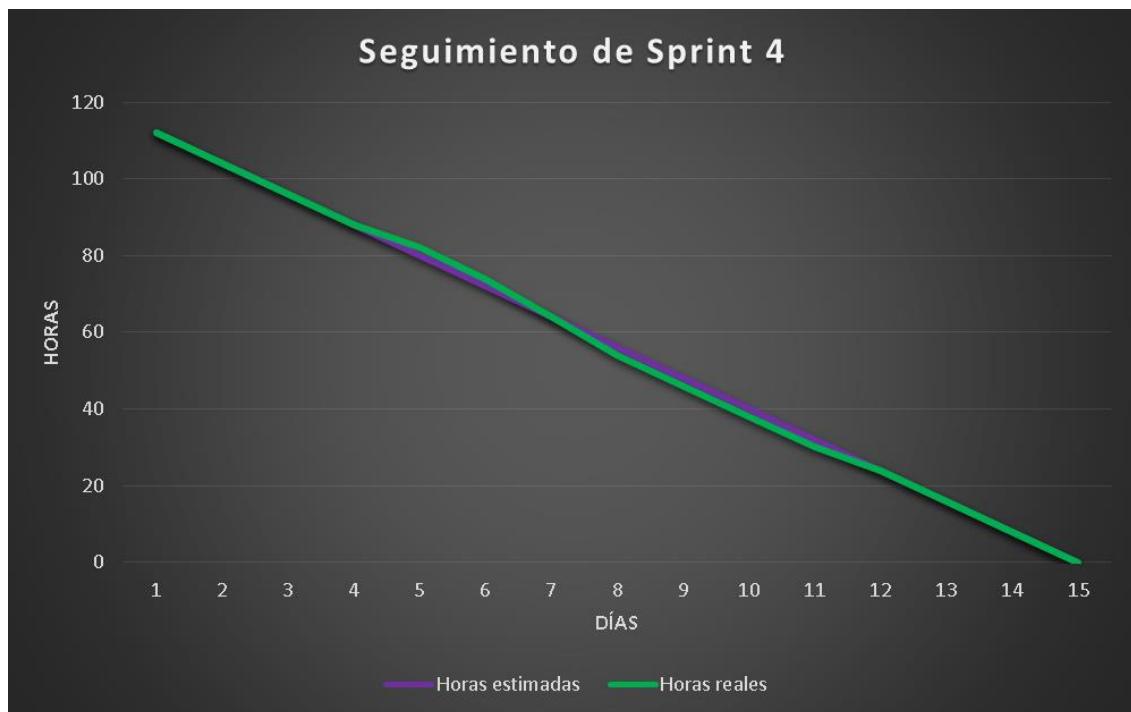
- ✓ Se observa que las pruebas de aceptación fueron completadas exitosamente.
- ✓ El tiempo que se estimó fue muy acorde al tiempo de realización del sprint.

DECISIONES

- ✓ Retroalimentar los procesos del sprint y continuar con el siguiente sprint.

Fuente: Creación propia

Figura A4. Seguimiento Sprint 4



Fuente: Creación propia

Tabla A7. Sprint 5

SPRINT No. 5

FINALIDAD	Permitir descarga y visualización en línea de proyectos docentes.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-004, HU-005

ANÁLISIS DEL SPRINT

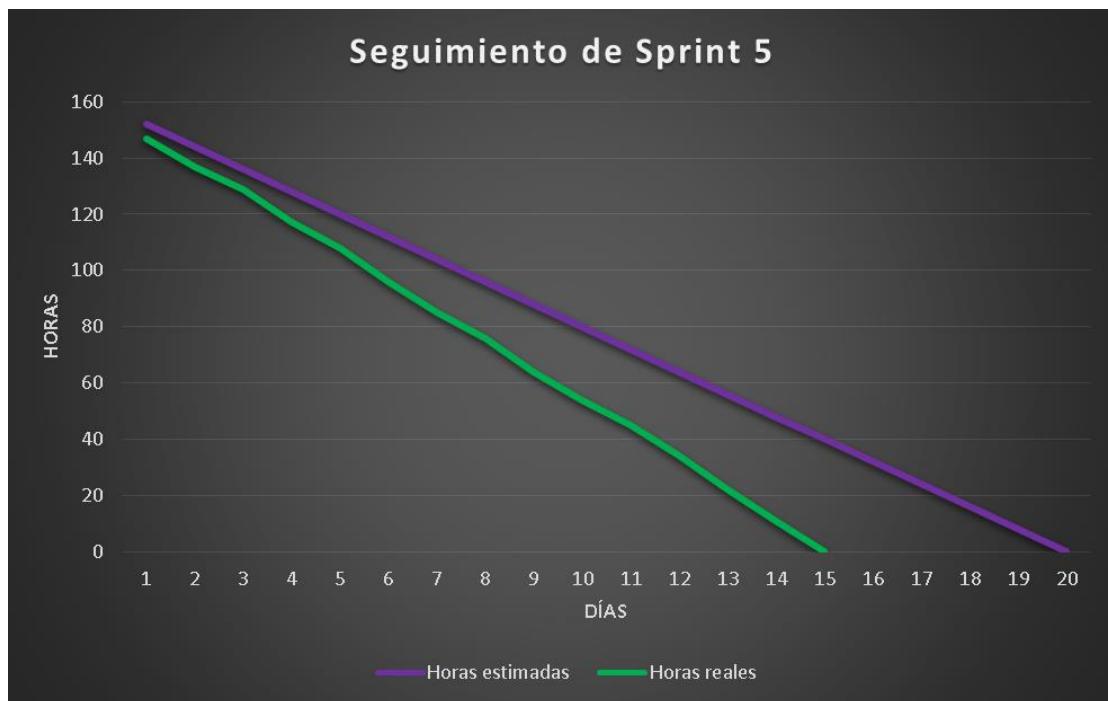
- ✓ **Se observa que las pruebas de aceptación fueron completadas exitosamente.**
- ✓ **Se sobreestimó el tiempo para el sprint dado que el tiempo para completarlo fue mucho menor respecto al tiempo estimado.**

DECISIONES

- ✓ **Retroalimentar los procesos del sprint y continuar con el siguiente sprint.**

Fuente: Creación propia

Figura A5. Seguimiento Sprint 5



Fuente: Creación propia

Tabla A8. Sprint 6

SPRINT No. 6

FINALIDAD	Permitir descarga y visualización en línea de proyectos estudiantiles.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-006, HU-007

ANÁLISIS DEL SPRINT

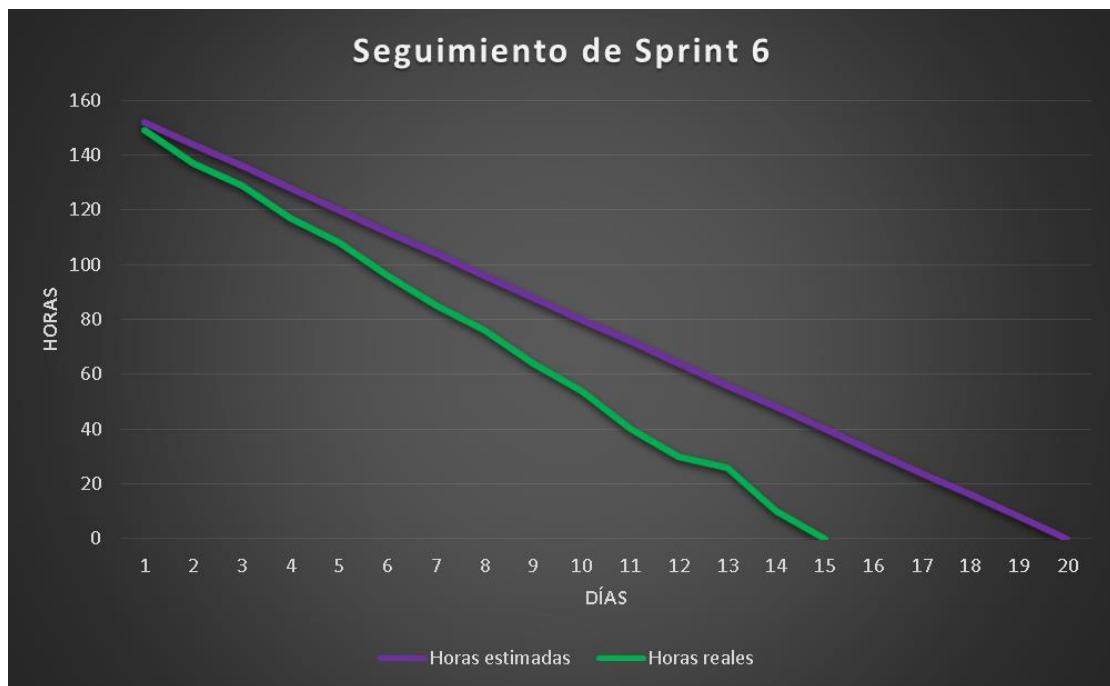
- ✓ **Se observa que las pruebas de aceptación fueron completadas exitosamente.**
- ✓ **Se sobreestimó el tiempo para el sprint dado que el tiempo para completarlo fue mucho menor respecto al tiempo estimado.**

DECISIONES

- ✓ **Retroalimentar los procesos del sprint y continuar con el siguiente sprint.**

Fuente: Creación propia

Figura A6. Seguimiento Sprint 6



Fuente: Creación propia

Tabla A9. Sprint 7

SPRINT No. 7

FINALIDAD	Permitir descarga y visualización en línea de trabajos de grado.
ARTEFACTOS	Gráfico de seguimiento de sprint Tabla general de sprint
HISTORIAS DE USUARIO	HU-008, HU-009

ANÁLISIS DEL SPRINT

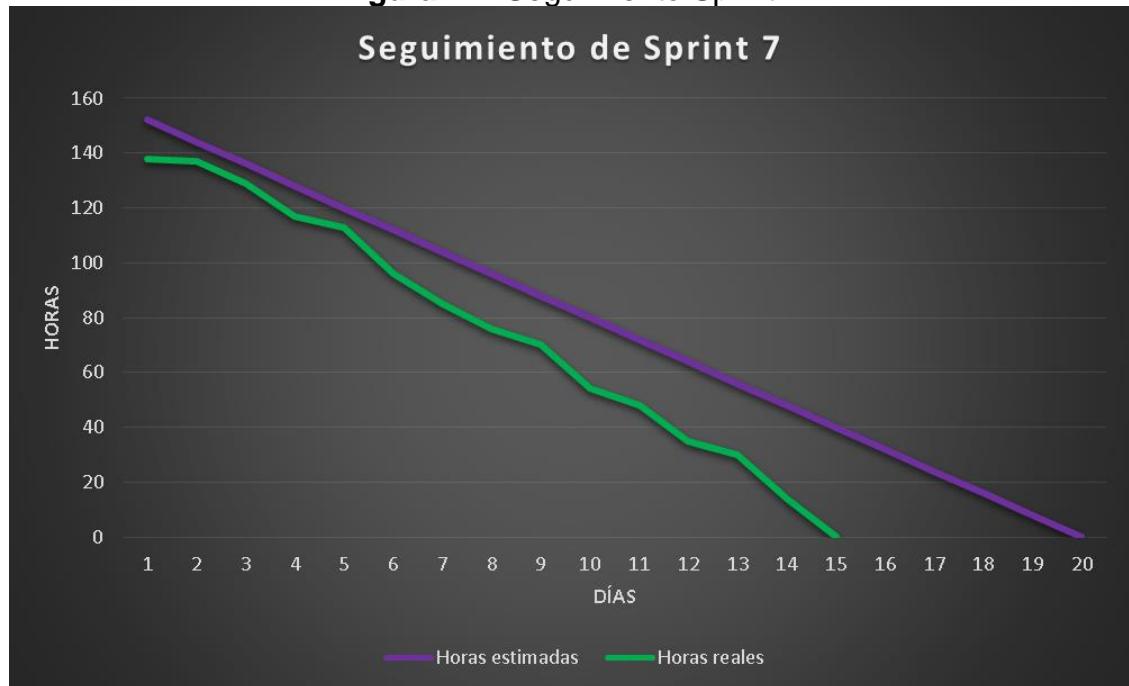
- ✓ Se observa que las pruebas de aceptación fueron completadas exitosamente.
- ✓ Se sobreestimó el tiempo para el sprint dado que el tiempo para completarlo fue mucho menor respecto al tiempo estimado.

DECISIONES

- ✓ Retroalimentar los procesos del sprint y continuar con el siguiente sprint.

Fuente: Creación propia

Figura A7. Seguimiento Sprint 7



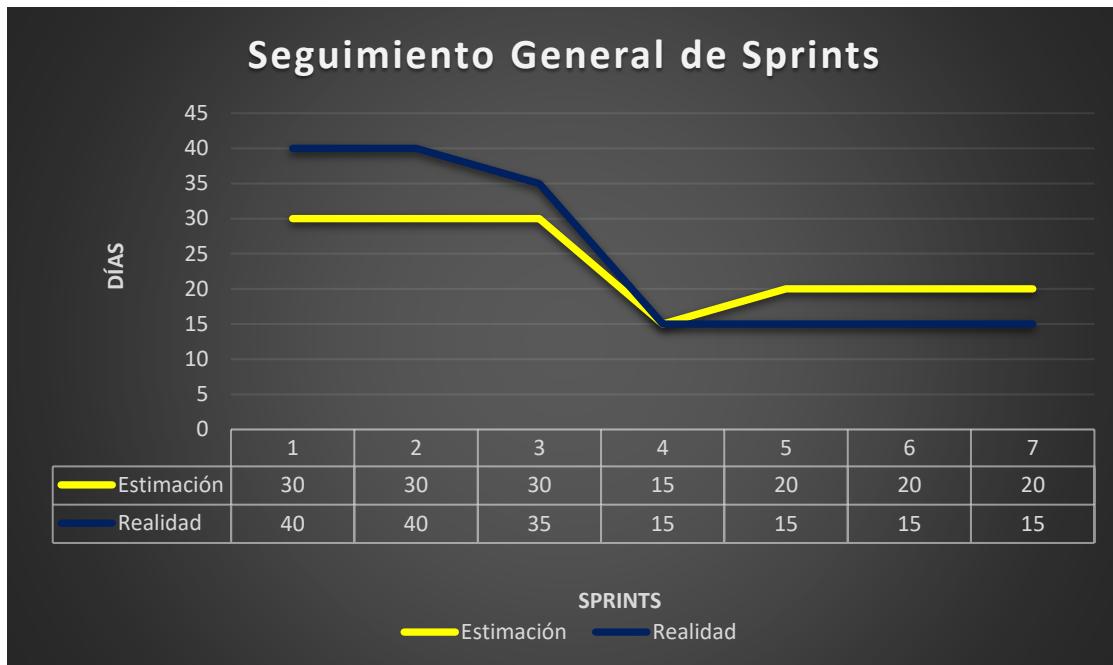
Fuente: Creación propia

RESULTADO GENERAL DE SPRINTS

Con el equipo de desarrollo se planteó determinar un resultado general de los sprints gráficamente. Para lograr esto se tomaron los tiempos en los cuales se estimaba que terminaría cada sprint y los tiempos en los cuales realmente terminaron.

Esto tuvo como objetivo conocer en qué punto estaban los sprints, llevar un control del proceso y determinar la calidad de cada uno de estos a lo largo del desarrollo del proyecto. Lo ideal era siempre estar igual o por debajo de la línea de tiempos de estimación de cada sprint.

Figura A12. Seguimiento general de sprints



Fuente: Creación propia

De este modo se puede observar que la línea de tiempo que se utilizó a lo largo del desarrollo fue diferente de la línea de tiempo de estimación, pues se generaron diversos valles y picos en cada uno de los sprints, aun así el desarrollo del proyecto en términos generales fue muy bueno, óptimo y de calidad porque algunos sprints terminaron antes de lo previsto y permitieron equilibrar los tiempos de entrega del producto final, así mismo se culminaron todas las fases de manera exitosa.

ANEXO B. CLASES, ATRIBUTOS Y MÉTODOS

Tabla B1. Clase Grupo_investigacion

Clase
class Grupo_investigacion(Thing)
Atributos Tabla B2.
Métodos Tabla B3.

Fuente: Creación propia

Tabla B2. Atributos clase Grupo_investigacion

Atributos	
Tipo	Nombre
Int	id_grupo_investigacion
String	nombre_grupo_investigacion
String	clasificacion_grupo_investigacion
String	area_grupo_investigacion
String	correo_grupo_investigacion

Fuente: Creación propia

Tabla B3. Métodos clase Grupo_investigacion

Métodos
def get_id_grupo_investigacion(self)
def set_id_grupo_investigacion(self,id_grupo_investigacion)
def get_nombre_grupo_investigacion(self)
def set_nombre_grupo_investigacion(self,nombre_grupo_investigacion)
def get_clasificacion_grupo_investigacion(self)
def set_clasificacion_grupo_investigacion(self,clasificacion_grupo_investigacion)
def get_area_grupo_investigacion(self)
def set_area_grupo_investigacion(self,area_grupo_investigacion)
def get_correo_grupo_investigacion(self)
def set_correo_grupo_investigacion(self,correo_grupo_investigacion)
def relation_gi_tiene_li(self,li)
def relation_gi_tiene_docente(self,docente)
def relation_gi_tiene_estudiante(self,estudiante)

Fuente: Creación propia

Tabla B4. Clase Linea_investigacion

Clase
class Linea_investigacion(Grupo_investigacion)
Atributos Tabla B5.
Métodos Tabla B6.

Fuente: Creación propia

Tabla B5. Atributos clase Linea_investigacion

Atributos	
Tipo	Nombre
Int	id_linea_investigacion
String	nombre_linea_investigacion

Fuente: Creación propia

Tabla B6. Métodos clase Linea_investigacion

Métodos
def get_id_linea_investigacion(self)
def set_id_linea_investigacion(self, id_linea_investigacion)
def get_nombre_linea_investigacion(self)
def set_nombre_linea_investigacion(self, nombre_linea_investigacion)
def relation_li_tiene_pi(self,pi)

Fuente: Creación propia

Tabla B7. Clase Investigador

Clase
class Investigador(Thing)
Atributos Tabla B8.
Métodos Tabla B9.

Fuente: Creación propia

Tabla B8. Atributos clase Investigador

Atributos	
Tipo	Nombre
Int	id_investigador
String	nombres_investigador
String	apellidos_investigador
String	codigo_investigador

String	cedula_investigador
String	correo_investigador

Fuente: Creación propia

Tabla B9. Métodos clase Investigador

Métodos
<code>def get_id_investigador(self)</code>
<code>def set_id_investigador(self,id_investigador)</code>
<code>def get_nombres_investigador(self)</code>
<code>def set_nombres_investigador(self,nombres_investigador)</code>
<code>def get_apellidos_investigador(self)</code>
<code>def set_apellidos_investigador(self,apellidos_investigador)</code>
<code>def get_codigo_investigador(self)</code>
<code>def set_codigo_investigador(self,codigo_investigador)</code>
<code>def get_cedula_investigador(self)</code>
<code>def set_cedula_investigador(self,cedula_investigador)</code>
<code>def get_correo_investigador(self)</code>
<code>def set_correo_investigador(self,correo_investigador)</code>
<code>def relation_investigador_es_docente(self,docente)</code>
<code>def relation_investigador_es_estudiante(self,estudiante)</code>
<code>def relation_investigador_es_ie(self,ie)</code>

Fuente: Creación propia

Tabla B10. Clase Docente

Clase
<code>class Docente(Investigador)</code>
Atributos Tabla B11.
Métodos Tabla B12.

Fuente: Creación propia

Tabla B11. Atributos clase Docente

Atributos	
Tipo	Nombre
Int	id_docente

Fuente: Creación propia

Tabla B12. Métodos Docente

Métodos
<code>def get_id_docente(self)</code>

```

def set_id_docente(self, id_docente)
def relation_docente_es_autor_pi(self,pi)
def relation_docente_asesora_pi(self,pi)

```

Fuente: Creación propia

Tabla B13. Clase Estudiante

Clase
class Estudiante(Investigador)
Atributos Tabla B14.
Métodos Tabla B15.

Fuente: Creación propia

Tabla B14. Atributos clase Estudiante

Atributos				
<table> <tr> <th>Tipo</th> <th>Nombre</th> </tr> <tr> <td>Int</td> <td>id_estudiante</td> </tr> </table>	Tipo	Nombre	Int	id_estudiante
Tipo	Nombre			
Int	id_estudiante			

Fuente: Creación propia

Tabla B15. Métodos clase Estudiante

Métodos
def get_id_estudiante(self)
def set_id_estudiante(self, id_estudiante)
def relation_estudiante_es_autor_pi(self,pi)

Fuente: Creación propia

Tabla B16. Clase Investigador_externo

Clase
class Investigador_externo(Investigador)
Atributos Tabla B17.
Métodos Tabla B18.

Fuente: Creación propia

Tabla B17. Atributos clase Investigador_externo

Atributos				
<table> <tr> <th>Tipo</th> <th>Nombre</th> </tr> <tr> <td>Int</td> <td>id_investigador_externo</td> </tr> </table>	Tipo	Nombre	Int	id_investigador_externo
Tipo	Nombre			
Int	id_investigador_externo			

Fuente: Creación propia

Tabla B18. Métodos clase Investigador_externo

Métodos
<code>def get_id_investigador_externo(self)</code>
<code>def set_id_investigador_externo(self, id_investigador_externo)</code>
<code>def relation_ie_es_autor_pi(self,pi)</code>

Fuente: Creación propia

Tabla B19. Clase Palabra

Clase
<code>class Palabra(Thing)</code>
Atributos Tabla B20.
Métodos Tabla B21.

Fuente: Creación propia

Tabla B20. Atributos clase Palabra

Atributos	
Tipo	Nombre
<code>Int</code>	<code>id_palabra</code>
<code>String</code>	<code>descripcion_palabra</code>
<code>String</code>	<code>lema_palabra</code>
<code>String</code>	<code>tipo_palabra</code>
<code>String</code>	<code>concepto_palabra</code>

Fuente: Creación propia

Tabla B21. Métodos clase Palabra

Métodos
<code>def get_id_palabra(self)</code>
<code>def set_id_palabra(self,id_palabra)</code>
<code>def get_descripcion_palabra(self)</code>
<code>def set_descripcion_palabra(self,descripcion_palabra)</code>
<code>def get_lema_palabra(self)</code>
<code>def set_lema_palabra(self,lema_palabra)</code>
<code>def get_tipo_palabra(self)</code>
<code>def set_tipo_palabra(self,tipoo_palabra)</code>
<code>def get_concepto_palabra(self)</code>
<code>def set_concepto_palabra(self,concepto_palabra)</code>
<code>def relation_palabra_sinonimo_palabra(self,palabra)</code>
<code>def relation_palabra_conecta_palabra(self,palabra)</code>

Fuente: Creación propia

Tabla B22. Clase Proyecto_investigacion

Clase
class Proyecto_investigacion(Thing)
Atributos Tabla B23.
Métodos Tabla B24.

Fuente: Creación propia

Tabla B23. Atributos clase Proyecto_investigacion

Atributos		
		Nombre
	Int	id_proyecto_investigacion
	String	titulo_proyecto_investigacion
	String	resumen_proyecto_investigacion
	String	palabra_clave1
	String	palabra_clave2
	String	palabra_clave3
	String	palabra_clave4
	String	palabra_clave5
	String	estado_proyecto_investigacion
	String	tipo_proyecto_investigacion

Fuente: Creación propia

Tabla B24. Métodos clase Proyecto_investigacion

Métodos
def get_id_proyecto_investigacion(self)
def set_id_proyecto_investigacion(self, id_proyecto_investigacion)
def get_titulo_proyecto_investigacion(self)
def set_titulo_proyecto_investigacion(self, titulo_proyecto_investigacion)
def get_resumen_proyecto_investigacion(self)
def set_resumen_proyecto_investigacion(self, resumen_proyecto_investigacion)
def get_estado_proyecto_investigacion(self)
def set_estado_proyecto_investigacion(self, estado_proyecto_investigacion)
def get_tipo_proyecto_investigacion(self)
def set_tipo_proyecto_investigacion(self, tipo_proyecto_investigacion)
def get_palabras_clave(self)
def set_palabras_clave(self, palabras_clave)
def relation_pi_tiene_palabra(self, palabra)

Fuente: Creación propia

Tabla B25. Clase Universidad

Clase
class Universidad(Thing)
Atributos Tabla B26.
Métodos Tabla B27.

Fuente: Creación propia

Tabla B26. Atributos clase Universidad

Atributos	
Tipo	Nombre
Int	id_universidad
String	nombre_universidad

Fuente: Creación propia

Tabla B27. Métodos clase Universidad

Métodos
def get_id_universidad(self)
def set_id_universidad(self, id_universidad)
def get_nombre_universidad(self)
def set_nombre_universidad(self, nombre_universidad)
def relation_universidad_tiene_facultad(self,facultad)
def relation_universidad_tiene_viiis(self,viis)

Fuente: Creación propia

Tabla B28. Clase Facultad

Clase
class Facultad(Universidad)
Atributos Tabla B29.
Métodos Tabla B30.

Fuente: Creación propia

Tabla B29. Atributos clase Facultad

Atributos	
Tipo	Nombre
Int	id_facultad
String	nombre_facultad

Fuente: Creación propia

Tabla B30. Métodos clase Facultad

Métodos
<code>def get_id_facultad(self)</code>
<code>def set_id_facultad(self, id_facultad)</code>
<code>def get_nombre_facultad(self)</code>
<code>def set_nombre_facultad(self, nombre_facultad)</code>
<code>def relation_facultad_tiene_departamento(self,departamento)</code>

Fuente: Creación propia

Tabla B31. Clase Departamento

Clase
<code>class Departamento(Facultad)</code>
Atributos Tabla B32.
Métodos Tabla B33.

Fuente: Creación propia

Tabla B32. Atributos clase Departamento

Atributos	
Tipo	Nombre
Int	id_departamento
String	nombre_departamento

Fuente: Creación propia

Tabla B33. Métodos clase Departamento

Métodos
<code>def get_id_departamento(self)</code>
<code>def set_id_departamento(self, id_departamento)</code>
<code>def get_nombre_departamento(self)</code>
<code>def set_nombre_departamento(self, nombre_departamento)</code>
<code>def relation_departamento_tiene_programa(self,programa)</code>
<code>def relation_departamento_tiene_gi(self,gi)</code>

Fuente: Creación propia

Tabla B34. Clase Programa

Clase
<code>class Programa(Departamento)</code>
Atributos Tabla B35.
Métodos Tabla B36.

Fuente: Creación propia

Tabla B35. Atributos clase Programa

Atributos	
Tipo	Nombre
Int	id_programa
String	nombre_programa

Fuente: Creación propia

Tabla B36. Métodos clase Programa

Métodos
<code>def get_id_programa(self)</code>
<code>def set_id_programa(self, id_programa)</code>
<code>def get_nombre_programa(self)</code>
<code>def set_nombre_programa(self, nombre_programa)</code>
<code>def relation_programa_tiene_estudiante(self,estudiante)</code>
<code>def relation_programa_tiene_docente(self,docente)</code>

Fuente: Creación propia

Tabla B37. Clase VIIS

Clase
<code>class VIIS(Thing)</code>
Atributos Tabla B38.
Métodos Tabla B39.

Fuente: Creación propia

Tabla B38. Atributos clase VIIS

Atributos	
Tipo	Nombre
Int	id_VIIS
String	nombre_VIIS

Fuente: Creación propia

Tabla B39. Métodos clase VIIS

Métodos
<code>def get_id_VIIS(self)</code>
<code>def set_id_VIIS(self, id_VIIS)</code>
<code>def get_nombre_VIIS(self)</code>

```

def set_nombre_VIIS(self, nombre_VIIS)
def relation_viis_tiene_convocatoria(self,convocatoria)
def relation_viis_tiene_investigador(self,investigador)
def relation_viis_adscribe_gi(self,gi)
def relation_viis_tiene_pi(self,pi)

```

Fuente: Creación propia

Tabla B40. Clase Convocatoria

Clase
class Convocatoria(VIIS)
Atributos Tabla B41.
Métodos Tabla B42.

Fuente: Creación propia

Tabla B41. Atributos clase Convocatoria

Atributos	
Tipo	Nombre
Int	id_convocatoria
String	nombre_convocatoria
Int	anio_convocatoria
String	tipo_convocatoria

Fuente: Creación propia

Tabla B42. Métodos clase Convocatoria

Métodos
def get_id_convocatoria(self)
def set_id_convocatoria(self, id_convocatoria)
def get_nombre_convocatoria(self)
def set_nombre_convocatoria(self, nombre_convocatoria)
def get_tipo_convocatoria(self)
def set_tipo_convocatoria(self, tipo_convocatoria)
def get_anio_convocatoria(self)
def set_anio_convocatoria(self, anio_convocatoria)
def relation_convocatoria_tiene_pi(self,pi)
def relation_convocatoria_dirigida_investigador(self,investigador)

Fuente: Creación propia

ANEXO C. DICCIONARIO DE DATOS

En un comienzo se diseñó un MER (Modelo Entidad Relación) con 16 tablas, las cuales fueron producto de la normalización, donde se llegó hasta la cuarta forma normal, pero en los sprints con los actores del sistema se observó que esto sería inevitablemente ineficiente, porque para las búsquedas se dificultaría realizar joins para poder llegar a un determinado resultado, también se tornaba ineficiente en el momento de alimentar a la Ontología. De modo tal que se construyó una big table denominada “Investigaciones” con todos los datos importantes de los proyectos de investigación.

Tabla C1. Entidad Investigaciones

ATRIBUTO	TIPO	TAMAÑO	PK	DESCRIPCIÓN
id	INTEGER		PK	Identificador único autoincremental.
id_investigacion	VARCHAR	8		Identificador de la investigación. Se diferencia del atributo “id” porque es un código asignado en el Sistema de Investigaciones de la Universidad de Nariño.
titulo_investigacion	VARCHAR	10000		Título de la investigación
resumen_investigacion	TEXT			Resumen de la investigación
estado_investigacion	VARCHAR	100		Estado en el que se encuentra la investigación
palabra_clave1	VARCHAR	500		Primera palabra clave de la investigación
palabra_clave2	VARCHAR	500		Segunda palabra clave de la investigación
palabra_clave3	VARCHAR	500		Tercera palabra clave de la investigación
palabra_clave4	VARCHAR	500		Cuarta palabra clave de la investigación
palabra_clave5	VARCHAR	500		Quinta palabra clave de la investigación

convocatoria	VARCHAR	100	Datos de la convocatoria a la cual se presentó dicha investigación para ser financiada
tipo_convocatoria	VARCHAR	100	Tipo de convocatoria que identifica si la investigación es un proyecto estudiantil o trabajo de grado
anio_convocatoria	INTEGER		Año de convocatoria
codigo_autor1	INTEGER		Identificador único del primer autor de la investigación
nombres_autor1	VARCHAR	100	Nombres del primer autor de la investigación
apellidos_autor1	VARCHAR	100	Apellidos del primer autor de la investigación
programa_autor1	VARCHAR	500	Programa al que pertenece el primer autor de la investigación
facultad_autor1	VARCHAR	500	Facultad a la que pertenece el primer autor de la investigación
departamento_autor1	VARCHAR	500	Departamento al que pertenece el primer autor de la investigación
grupo_investigacion1	VARCHAR	500	Grupo de investigación al que pertenece el primer autor de la investigación
linea_investigacion1	VARCHAR	500	Línea de investigación que trabaja el primer autor de la investigación
codigo_autor2	INTEGER		Identificador único del segundo autor de la investigación

nombres_autor2	VARCHAR	100	Nombres del segundo autor de la investigación
apellidos_autor2	VARCHAR	100	Apellidos del segundo autor de la investigación
programa_autor2	VARCHAR	500	Programa al que pertenece el segundo autor de la investigación
facultad_autor2	VARCHAR	500	Facultad a la que pertenece el segundo autor de la investigación
departamento_autor2	VARCHAR	500	Departamento al que pertenece el segundo autor de la investigación
grupo_investigacion2	VARCHAR	500	Grupo de investigación al que pertenece el segundo autor de la investigación
linea_investigacion2	VARCHAR	500	Línea de investigación que trabaja el segundo autor de la investigación
codigo_autor3	INTEGER		Identificador único del tercer autor de la investigación
nombres_autor3	VARCHAR	100	Nombres del tercer autor de la investigación
apellidos_autor3	VARCHAR	100	Apellidos del tercer autor de la investigación
programa_autor3	VARCHAR	500	Programa al que pertenece el tercer autor de la investigación
facultad_autor3	VARCHAR	500	Facultad a la que pertenece el tercer

			autor de la investigación
departamento_autor3	VARCHAR	500	Departamento al que pertenece el tercer autor de la investigación
grupo_investigacion3	VARCHAR	500	Grupo de investigación al que pertenece el tercer autor de la investigación
linea_investigacion3	VARCHAR	500	Línea de investigación que trabaja el tercer autor de la investigación
codigo_autor4	INTEGER		Identificador único del cuarto autor de la investigación
nombres_autor4	VARCHAR	100	Nombres del cuarto autor de la investigación
apellidos_autor4	VARCHAR	100	Apellidos del cuarto autor de la investigación
programa_autor4	VARCHAR	500	Programa al que pertenece el cuarto autor de la investigación
facultad_autor4	VARCHAR	500	Facultad a la que pertenece el cuarto autor de la investigación
departamento_autor4	VARCHAR	500	Departamento al que pertenece el cuarto autor de la investigación
grupo_investigacion4	VARCHAR	500	Grupo de investigación al que pertenece el cuarto autor de la investigación
linea_investigacion4	VARCHAR	500	Línea de investigación de que

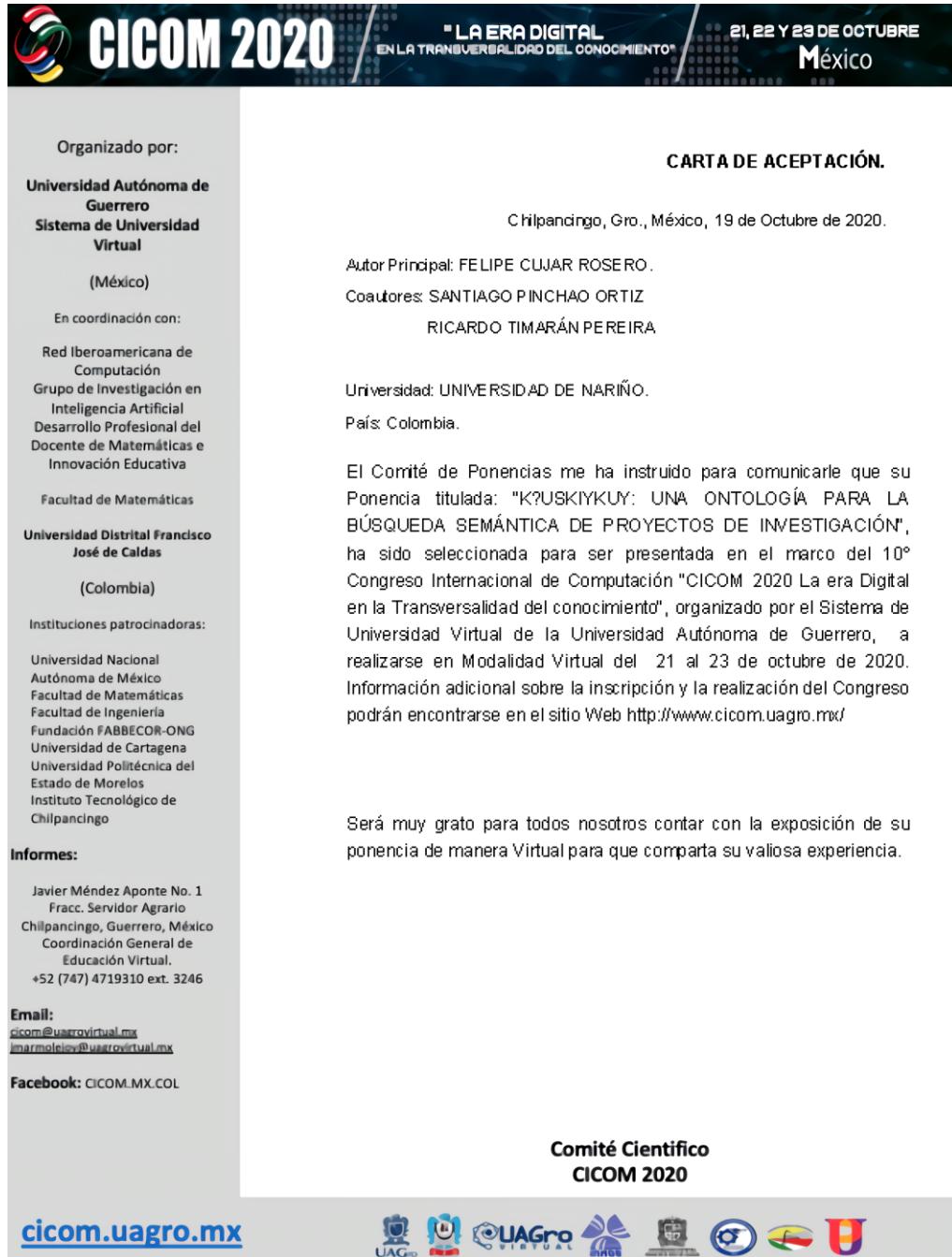
asesor1	VARCHAR	100	trabaja el cuarto autor de la investigación
asesor2	VARCHAR	100	Nombre de la primera persona que asesora el proyecto de investigación: estudiantil o trabajo de grado.
asesor3	VARCHAR	100	Nombre de la segunda persona que asesora el proyecto de investigación: estudiantil o trabajo de grado.
asesor4	VARCHAR	100	Nombre de la tercera persona que asesora el proyecto de investigación: estudiantil o trabajo de grado.
tipo_investigacion	VARCHAR	20	Nombre de la cuarta persona que asesora el proyecto de investigación: estudiantil o trabajo de grado.
corpus	TEXT		Es el tipo de investigación que puede ser proyecto estudiantil, proyecto docente o trabajo de grado.
corpus_palabras	TEXT		Es el corpus que resultó de unificar todos los datos de los proyectos de investigación.
corpus_lemas	TEXT		Es el atributo que tiene el corpus preprocesado.

Fuente: Creación propia

ANEXO D. LOGROS – PUBLICACIONES - PONENCIAS - ARTÍCULOS - POSTERS - VIIS

CICOM (Congreso Internacional de Computación - México)

Figura D1. Carta de aceptación ponencia CICOM (México)



CICOM 2020 / **“LA ERA DIGITAL EN LA TRANSVERSAVIDAD DEL CONOCIMIENTO”** / **21, 22 Y 23 DE OCTUBRE** / **México**

Organizado por:

Universidad Autónoma de Guerrero
Sistema de Universidad Virtual
(México)

En coordinación con:

Red Iberoamericana de Computación
Grupo de Investigación en Inteligencia Artificial
Desarrollo Profesional del Docente de Matemáticas e Innovación Educativa

Facultad de Matemáticas

Universidad Distrital Francisco José de Caldas
(Colombia)

Instituciones patrocinadoras:

Universidad Nacional Autónoma de México
Facultad de Matemáticas
Facultad de Ingeniería
Fundación FABECOR-ONG
Universidad de Cartagena
Universidad Politécnica del Estado de Morelos
Instituto Tecnológico de Chilpancingo

Informes:

Javier Méndez Aponte No. 1
Fracc. Servidor Agrario
Chilpancingo, Guerrero, México
Coordinación General de Educación Virtual.
+52 (747) 4719310 ext. 3246

Email:
cicom@uagrovirtual.mx
imarmolejo@uagrovirtual.mx

Facebook: CICOM.MX.COL

CARTA DE ACEPTACIÓN.

Chilpancingo, Gro., México, 19 de Octubre de 2020.

Autor Principal: FELIPE CUJAR ROSERO.

Coautores: SANTIAGO PINCHAO ORTIZ
RICARDO TIMARÁN PEREIRA

Universidad: UNIVERSIDAD DE NARIÑO.

País: Colombia.

El Comité de Ponencias me ha instruido para comunicarle que su Ponencia titulada: "K?USKIYKUY: UNA ONTOLOGÍA PARA LA BÚSQUEDA SEMÁNTICA DE PROYECTOS DE INVESTIGACIÓN", ha sido seleccionada para ser presentada en el marco del 10º Congreso Internacional de Computación "CICOM 2020 La era Digital en la Transversalidad del conocimiento", organizado por el Sistema de Universidad Virtual de la Universidad Autónoma de Guerrero, a realizarse en Modalidad Virtual del 21 al 23 de octubre de 2020. Información adicional sobre la inscripción y la realización del Congreso podrán encontrarse en el sitio Web <http://www.cicom.uagro.mx/>

Será muy grato para todos nosotros contar con la exposición de su ponencia de manera Virtual para que comparta su valiosa experiencia.

Comité Científico
CICOM 2020

cicom.uagro.mx

UAG UAGro VIRTUAL UAGro UAGro UAGro UAGro

Figura D2. Certificado presentación ponencia CICOM (México)



Figura D3. Artículo publicado en Revista Tlamati Sabiduría, Vol. 11, No. 3

tlamati.uagro.mx/t113e/t113e.pdf

- | + 150% ▾

Tlamati Sabiduría, Volumen 11 Número Especial 3 (2020)

Revista
Tlamati
Sabiduría

10º Congreso Internacional de Computación CICOM
Ciudad de México, 21, 22 y 23 de octubre 2020
Resúmenes.

K'USKIYKUY: Una ontología para la búsqueda semántica de proyectos de investigación

Felipe Cujar Rosero

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia
felipe.cujar@gmail.com

Santiago Pinchao Ortiz

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia
sangeeky@outlook.com

Mateo Guerrero Restrepo

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia
jimaguere@udenar.edu.co

Ricardo Timarán Pereira

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia
ritimar@udenar.edu.co

Enlace Revista Formato Htm: <http://tlamati.uagro.mx/revista113e.htm>

Enlace Articulos Formato Pdf: <http://tlamati.uagro.mx/t113e/t113e.pdf>

Figura D4. Artículo publicado en Memorias del evento de CICOM 2020



CICOM 2020
10° Congreso Internacional de computación México - Colombia
Octubre 21, 22 y 23 de 2020, Ciudad de México

K'USKIYKUY: Una ontología para la búsqueda semántica de proyectos de investigación

Felipe Cujar Rosero

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia

felipecujar@gmail.com

Mateo Guerrero Restrepo

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia

jimaguere@udenar.edu.co

Santiago Pinchao Ortiz

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia

sangeeky@outlook.com

Ricardo Timarán Pereira

Universidad de Nariño
Departamento de Sistemas
Pasto, Colombia

ritimar@udenar.edu.co

RESUMEN

En este artículo se presentan los primeros resultados del proyecto de investigación que tuvo como objetivo la construcción de una ontología como base de conocimiento para que soporte la búsqueda semántica de los proyectos de investigación de tipo docente y estudiantil financiados mediante convocatoria por el sistema de investigaciones de la Universidad de Nariño. Se utilizó para la construcción de la ontología denominada K'USKIYKUY, la metodología Methontology, que está compuesta por cinco fases que incluye: especificación, conceptualización, formalización,

investigación y obtención de conocimiento gracias a la estructura y semántica que posee.

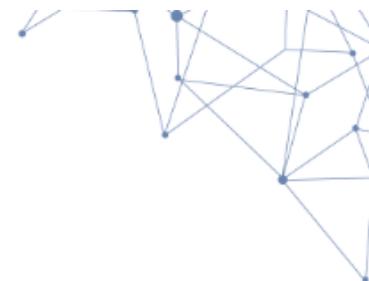
ABSTRACT

This paper presents the first results of the research project whose objective was the construction of an ontology as a knowledge base to support the semantic search of teaching and student research projects financed through a call by the research system of the University of Nariño. It was used for the construction of the ontology called K'USKIYKUY, the Methontology methodology,

Enlace Publicación: http://www.cicom2020.uagro.mx/?page_id=915

EXPOPOSTER

Figura D5. Carta de Aceptación EXPOPOSTER (Colombia)



Pasto, 17 de Octubre 2020

Estimado(a):

FELIPE CUJAR ROSERO
UNIVERSIDAD DE NARIÑO

Cordial saludo

EL comité organizador de **EXPOPOSTER 2020 ENCUENTRO ACADÉMICO REDIS NODO SUR OCCIDENTE**, que se coordinará desde la ciudad de Pasto (Nariño) –los días 4, 5 y 6 de noviembre de 2020, se complace en informarle que su contribución con ID 1020 titulada "MOTOR DE BÚSQUEDA INTELIGENTE SEMÁNTICO BASADO EN UNA ONTOLOGÍA CON MACHINE LEARNING PARA PROYECTOS DE INVESTIGACIÓN DENTRO DE LA UNIVERSIDAD DE NARIÑO." ha sido aceptada para su presentación en el encuentro académico en la modalidad de trabajo de grado.

La versión final de su video y documento deberá ser elaborada siguiendo las indicaciones y especificaciones que se encuentran en la plataforma expoposter.udener.edu.co.

Esperamos contar con su participación en el encuentro y que se convierta en una de sus mejores experiencias en el ámbito académico y personal.

Atentamente,

Comité Organizador ExpoPóster 2020

expoposter@udener.edu.co

Teléfonos: 7244309 – 7311449 ext. 2005



Figura D6. Certificado presentación EXPOPOSTER (Colombia)



CACIED

Figura D7. Carta de aceptación presentación CACIED (Colombia)



San Juan de Pasto, septiembre 16 de 2019

Felipe Cujar Rosero
Universidad de Nariño
Colombia

Cordial saludo,

El comité organizador del 4^{to} Congreso Andino de Computación, Informática y Educación CACIED 2019, que se realizará en la ciudad de San Juan de Pasto (N)- Colombia, entre el 6 y el 8 de noviembre de 2019, se place en informarle que su contribución identificada con el ID 6 titulada: **MOTOR DE BÚSQUEDA INTELIGENTE DE INFORMES DE INVESTIGACIÓN BASADO EN RECURSOS SEMÁNTICOS**, ha sido aceptada para su presentación en el congreso en la modalidad de **Póster**.

La versión final de su artículo/poster deberá ser elaborada siguiendo las indicaciones y especificaciones que se encuentran en la plataforma del congreso.

De antemano le damos la bienvenida a San Juan de Pasto y esperamos que su estancia en el evento se convierta en una de sus mejores experiencias en el ámbito académico y personal.

Atentamente,

Comité Organizador CACIED 2019
cacied2019@udenar.edu.co
info_cacied2019@udenar.edu.co



Universidad de Nariño



Universidad
Mariana



UNIVERSIDAD
CESMAG
Centro Universitario de Sistemas y
Ciberespacio Magdalena

Figura D8. Certificado PRIMER PUESTO para presentación CACIED (Colombia)



Financiamiento de parte de VIIS (Vicerrectoría de Investigación e Interacción Social)

Figura D9. Acuerdo de aprobación para financiar el proyecto (parte 1)

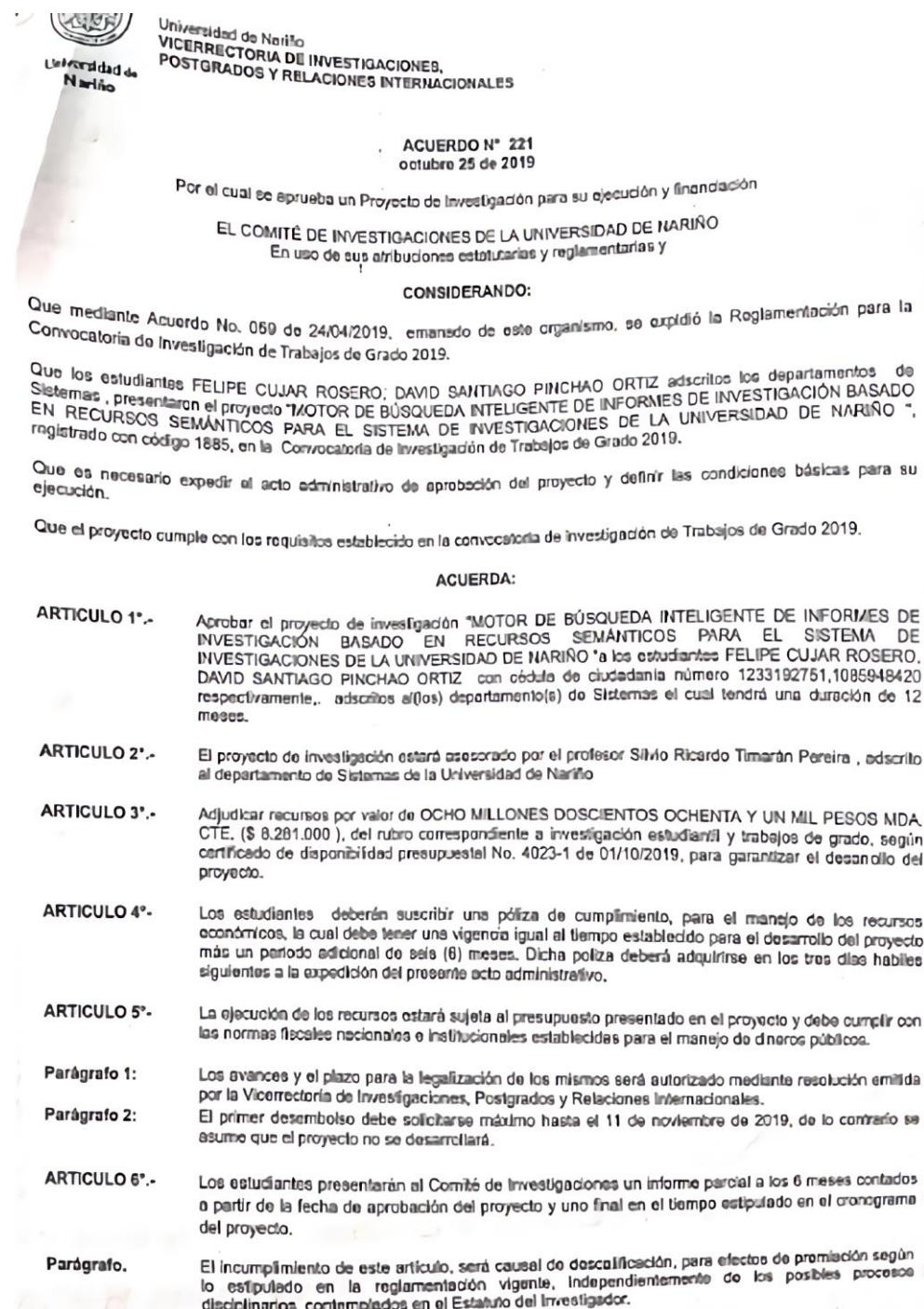


Figura D10. Acuerdo de aprobación para financiar el proyecto (parte 2)



Universidad de Nariño
VICERRECTORÍA DE INVESTIGACIONES,
POSTGRADOS Y RELACIONES INTERNACIONALES

ACUERDO N° 221
octubre 25 de 2019

ARTICULO 7°.- Se iniciará con la ejecución de recursos una vez esté firmada el acta cumplimiento, por los estudiantes y asesor del proyecto.

ARTICULO 8°.- La Vicerrectoría de Investigaciones, Postgrados y Relaciones Internacionales, y los departamentos Sistemas anotarán lo de su cargo.

COMUNIQUESE Y CÚMPLASE

Dado en San Juan de Pasto, a los 26 días del mes de octubre de 2019.

A handwritten signature in black ink.

OSCAR EDUARDO CHECA CORAL
Presidente

A handwritten signature in black ink.

MARÍA CONSTANZA CABRERA DULCE
Secretaria

SOFTHARD 2021

Figura D.11 Certificado de presentación SOFTHARD (Colombia)



2nd International Conference on Natural Language Processing and Computational Linguistics (NLPCL 2021) (Vancouver, Canadá)

Figura D.12 Carta de aceptación NLPCL 2021 (Vancouver, Canadá)

Notification Mail - 2nd International Conference on Natural Language Processing & Computational Linguistics (NLPCL 2021) ▾ Recibidos x

 nlpcl@ccsit2021.org 8 abr 2021 2:55 (hace 1 día)   

para mí, felipecujar, santipinchi, ritimar, jimaguere ▾

 inglés ▾ > español ▾ Traducir mensaje Desactivar para: inglés x

Dear Author,

First of all, thank you very much for submitting your paper to NLPCL 2021 to be held in May 29~30, 2021, Vancouver, Canada. Based upon the reviewer's reports, we are pleased to inform you that your paper has been **ACCEPTED** by the conference and will be included in the proceedings published by [Computer Science Conference Proceedings](#) in [Computer Science & Information Technology \(CS & IT\) series \(Confirmed\)](#).

Congratulations on your excellent work!

Figura D.13 Certificado presentación NLPCL 2021 (Vancouver, Canadá)



Enlace Publicación: <https://aircconline.com/csit/abstract/v11n7/csit110709.html>

Enlace Publicación: <http://airccse.org/csit/V11N07.html>

Enlace Programación: <https://www.ccsit2021.org/ps/nlpcl-ps.html>

Enlace Papers aceptados: <https://www.ccsit2021.org/nlpcl/papers.html>

LACCEI International Multi-conference for Engineering, Education and Technology 2021 (Argentina)

Figura D.14 Carta de aceptación LACCEI International Multi-conference for Engineering, Education and Technology 2021 (Argentina)

LATIN AMERICAN AND CARIBBEAN CONSORTIUM OF ENGINEERING INSTITUTIONS
CONSORCIO DE ESCUELAS DE INGENIERÍA DE LATINOAMÉRICA Y DEL CARIBE

LACCEI
Latin American and Caribbean Consortium of Engineering Institutions

2021 Executive Board:

President
Luis Fernando Martínez Arconada
Ecole Nationale d'Ingenieurs de Tarbes,
France

Past President
Miguel Angel Sosa
Universidad Tecnológica Nacional,
Argentina

Executive Director
Maria Mercedes Larondo Petrie
Florida Atlantic University,
United States

VP of Finance
Roberto Lorán
Universidad Ana G. Mendoza,
Puerto Rico

VP of Initiatives
Laura Romero Robles
Instituto Tecnológico de Monterrey,
México

VP of Membership
Claudio Camilo González Clavijo
Universidad Nacional Abierta y a Distancia,
Colombia

At Large Members:
Stella Batalama
Florida Atlantic University,
United States

Boca Raton, FL, USA, June 13, 2021

Felipe Cujar Rosero
Universidad de Nariño
Colombia

Re: Invitation to participate as student finalist at 2021 LACCEI International Multi-Conference of Engineering, Education and Technology – Virtual Edition. **July 19 – 23,** 2021

Dear Felipe,

We take great pleasure in informing you that your student full paper titled **"Thaqhaña: Un Motor De Búsqueda Semántico Basado En Una Ontología Y En Un Modelo Entrenado Con Machine Learning Para Soporte A La Investigación"** has been **ACCEPTED** for presentation as **FINALIST** of the Student Research Paper Competition in the 2021 LACCEI International Multi-Conference of Engineering, Education and Technology.

Considering this, we would like to formally invite you to present your work in the student finalists' sessions during our conference, that this time will be held in a totally virtual, interactive, and synchronous way, on July 19 – 23, 2021.

You have until Friday June 18th to confirm your participation, sending an email to admin@laccei.org. If no email is sent, we will not include your paper in the finalist session.

The mission of the LACCEI conference is to bring together teachers, researchers, students, industrialists and professionals, particularly from the Americas, to share experiences and promote collaboration between the different actors and institutions. Our conference is taking the lead in our hemisphere and we are confident that your work will contribute to achieving our goals in the 2021 conference.

To be included in the program and the conference proceedings at least one of the authors of the paper accepted must register for the conference, pay the registration and present the work at the virtual conference. For more information, please visit www.laccei.org/conference.

Also register for the conference as soon as possible to take advantage of a discounted price visit: www.laccei.org/blog/information-registration/

We humbly thank you for your participation in this grand event.

Sincerely,


Maria M. Larondo Petrie, PhD
LACCEI Executive Director



LACCEI, Inc.
Florida Atlantic University - FAU
777 Glades Road, EE-308
Boca Raton, Florida, USA
33431-0991
Web: www.laccei.org
Email: admin@laccei.org
Skype: laccei
Tel: +1 561 297 3899
Fax: +1 561 297 1111

ANEXO E. PRUEBAS

Se anexan casos de prueba, pruebas unitarias y pruebas de aceptación.

Tabla E.1 Caso de prueba 4

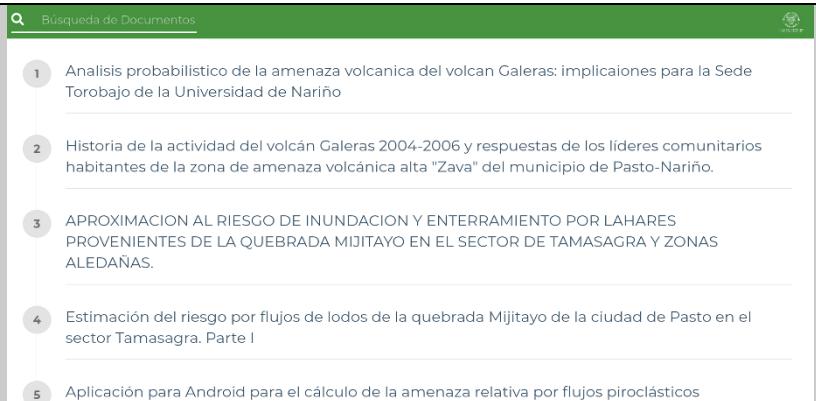
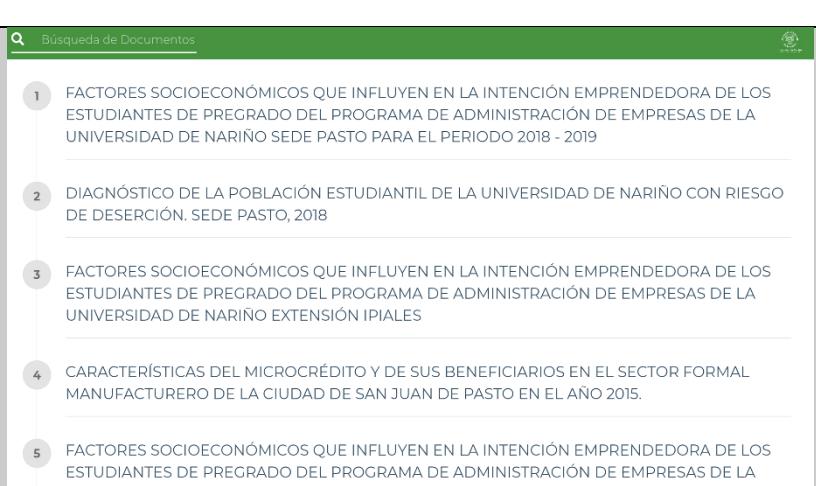
Búsqueda realizada	“Proyectos de investigación sobre amenaza volcánica”
Resultado búsqueda	Calificación
	E
Fuente: Creación propia	

Tabla E.2 Caso de prueba 5

Búsqueda realizada	“Proyectos de investigación sobre economía”
Resultado búsqueda	Calificación
	E

Calificación	E
Fuente: Creación propia	

Tabla E.3 Caso de prueba 6

Búsqueda realizada	“Proyectos de investigación sobre cultivos”
Resultado búsqueda	<p>Búsqueda de Documentos</p> <ul style="list-style-type: none"> 1 EVALUACIÓN DEL POTENCIAL PRODUCTIVO Y LA RELACION ENTRE COMPONENTES DE RENDIMIENTO DE 20 LINEAS DE HABA (Vicia faba) EN EL MUNICIPIO DE PASTO 2 EFECTO DE ALGUNOS ABONOS VERDES SOBRE EL COMPORTAMIENTO AGRONOMICO DE CULTIVOS TRANSITORIOS EN DOS ANDISOLES DEL MUNICIPIO DE PASTO 3 EFECTO DE TRES TIPOS DE ENMIENDAS, SOBRE LA PRODUCTIVIDAD DEL CULTIVO DE ALFALFA (Medicago Sativa L) EN UN SUELO Typic haplustands 4 Evaluación por componentes de rendimiento y ajuste de sistemas y densidades de siembra en arvejas arbustivas con gen afila (Pisum sativum L) 5 EVALUACIÓN DEL EFECTO DE SIETE ABONOS VERDES SOBRE EL RENDIMIENTO DE TRES CULTIVOS TRANSITORIOS EN DOS ANDISOLES DEL MUNICIPIO DE PASTO DEPARTAMENTO DE NARIÑO.

Calificación	E
Fuente: Creación propia	

Tabla E.4 Caso de prueba 7

Búsqueda realizada	“Proyectos de investigación sobre idiomas”
Resultado búsqueda	<p>Búsqueda de Documentos</p> <ul style="list-style-type: none"> 1 Working Memory Training and its Effect on Second Language Low Performance Students. 2 Learners satisfaction and Language Teacher Education programs at Universidad de Nariño 3 DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 11º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO CON TÉCNICAS DESCRIPTIVAS DE MINERÍA DE DATOS 4 NIVEL DE INGLÉS DE ESTUDIANTES Y DOCENTES EN LOS PROGRAMAS Y FACULTADES DE LA UNIVERSIDAD DE NARIÑO, SEDE PASTO, A TRAVÉS DE UNA PRUEBA INTERNACIONAL ESTANDARIZADA PARA DEFINIR UNA POLÍTICA DE LA LENGUA EXTRANJERA, INGLÉS Y UN PLAN DE MEJORAMIENTO 5 MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO

Calificación	E
Fuente: Creación propia	

Tabla E.5 Caso de prueba 8

Búsqueda realizada	“Proyectos de investigación sobre álgebra”
Resultado búsqueda	<p> Búsqueda de Documentos</p> <p>1 Subálgebras de Mishchenko-Fomenko en $U(gl_n)$ y secuencias regulares</p> <p>2 Fractales, matemáticas y algoritmos</p> <p>3 Una mirada computacional a la Teoría Algebraica de Códigos</p> <p>4 Estudio Teórico y Computacional de la Ecuación Diferencial Parcial Elíptica de Poisson</p> <p>5 Análisis teórico y computacional de las propiedades de matrices para ciertas discretizaciones numéricas.</p>
Calificación	E
Fuente:	Creación propia

Tabla E.6 Caso de prueba 9

Búsqueda realizada	“Proyectos de investigación sobre química”
Resultado búsqueda	<p> Búsqueda de Documentos</p> <p>1 Evaluación de la actividad antimicrobiana de Peperomia subspathulata y Cordia acuta</p> <p>2 SÍNTESIS, CARACTERIZACIÓN ESTRUCTURAL, ACTIVIDAD ANTIFÚNGICA Y RELACIONES CUANTITATIVAS DE ESTRUCTURA-ACTIVIDAD DE NUEVOS 1,2,4-TRIAZOLES</p> <p>3 Estudio químico y actividad antioxidante in vitro de la fracción polifenólica (antocianinas) aislada de los granos, hojas de mazorca y corona del maíz morado (<i>Zea Mays L</i>) cultivado en el Departamento de Nariño.</p> <p>4 DESARROLLO DE NANOCATALIZADORES METÁLICOS CON Pt, Ni Y Co PARA LA ELECTRO-OXIDACIÓN DE ACETALDEHIDO</p> <p>5 ESTUDIO COMPUTACIONAL DEL EQUILIBRIO TAUTOMÉRICO EN 1,2,4-TRIAZOLES-3-AMINO-5-SUSTITUIDOS</p>
Calificación	E
Fuente:	Creación propia

Tabla E.7 Caso de prueba 10

Búsqueda realizada	“Proyectos de investigación sobre ciencias agrícolas”
--------------------	--

Resultado búsqueda	Búsqueda de Documentos
	1 Procesos de acompañamiento en el aula realizado por los docentes tutores en el marco del Programa Todos a Aprender 2.0
	2 EVALUACIÓN DE VARIABLES AGRONÓMICAS MEDIANTE VEHÍCULOS AÉREOS NO TRIPULADOS EN DOS ESPECIES DE IMPORTANCIA AGRONÓMICA EN EL MUNICIPIO DE PASTO, DEPARTAMENTO DE NARIÑO.
	3 CONTRIBUCIONES DE LAS FACULTADES DE EDUCACIÓN, CIENCIAS HUMANAS Y AGRÍCOLAS DE LA UNIVERSIDAD DE NARIÑO, PARA LA GENERACIÓN DE ACCIONES DE PAZ EN LA REGIÓN.
	4 ALMACENAMIENTO DE CARBONO EN EL SUELO Y LA BIOMASA ARBOREA EN DIFERENTES SISTEMAS DE USO DEL SUELO DE LA ZONA ALTO ANDINA, DEPARTAMENTO DE NARIÑO.
	5 LOS GERROMORPHA DEL PIEDEMONTE SUR DE LA AMAZONIA (PUTUMAYO, COLOMBIA)

Calificación

E

Fuente: Creación propia

Tabla E.8 Caso de prueba 11

Búsqueda realizada	“Proyectos de investigación sobre inteligencia de negocios”
	Búsqueda de Documentos
Resultado búsqueda	
	1 MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO
	2 ANÁLISIS DE FUNCIONALIDAD DE LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS DE SOFTWARE LIBRE SPAGO BI ORIENTADA AL USO EN LAS PYMES DE LA REGIÓN
	3 SITAPP: UNA APLICACIÓN INTELIGENTE PARA DISPOSITIVOS MÓVILES DEL SISTEMA DE RUTAS DE TRANSPORTE URBANO DEL MUNICIPIO DE PASTO
	4 Negocios inmobiliarios en la morfología urbana para ciudad intermedia: Caso de estudio ciudad de pasto 2019
	5 UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS PRUEBAS SABER 5 DE LAS INSTITUCIONES EDUCATIVAS DE LOS MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO

Calificación

E

Fuente: Creación propia

Tabla E.9 Caso de prueba 12

Búsqueda realizada	“Proyectos de investigación sobre psicología”
Resultado búsqueda	<p>1 PROPIEDADES PSICOMÉTRICAS DE UNA PRUEBA PARA MEDIR EL PENSAMIENTO CRÍTICO A ESTUDIANTES DE PREGRADO DE COLOMBIA, MÉXICO Y CHILE.</p> <p>Resumen: El proyecto tiene por objetivo determinar las propiedades psicométricas de una prueba para medir pensamiento crítico a estudiantes de pregrado de Colombia, México y Chile. Se utilizará el paradigma cuantitativo de investigación, con un tipo de estudio instrumental ya que aplica los principios y técnicas propios de la psicometría encaminados a la construcción, validación y estandarización (propiedades psicométricas) de pruebas. El presente estudio permitirá la construcción de un instrumento sensible a la población latina y específicamente a los estudiantes de pregrado de Colombia, México y Chile, aportando a la construcción de nuevo conocimiento, específicamente en la psicología educativa, cognitiva, medición y evaluación y psicometría. Asimismo, se aportará a la formación de recursos humanos, consolidación de redes de investigación, construcción de cooperación internacional y consolidación de grupos de investigación.</p> <p>Palabras Clave: Estandarización, habilidades de pensamiento, Evaluación, construcción de prueba, Pensamiento crítico.</p> <p>Autores:</p> <ul style="list-style-type: none"> • Christian Alexander Zambrano Guerrero • Ana Karen Ceballos Mora • Sonia Betancourt Zambrano <p>Facultad: Ciencias Humanas. Programa: Psicología. Grupos de Investigación: Psicología y Salud .Psicología y Salud. Líneas de Investigación: LÍNEA DE INVESTIGACIÓN DESARROLLO DEL PENSAMIENTO ,Medición y Evaluación Psicológica. Tipo Convocatoria: Docente Año Convocatoria: 2014</p> <p style="text-align: right;">Investigaciones Relacionadas </p>

Calificación

E

Fuente: Creación propia

Tabla E.10 Caso de prueba 13

Búsqueda realizada	“Proyectos de investigación sobre minería de datos”
Resultado búsqueda	<p>1 APLICACIÓN DE LA MINERÍA DE DATOS EN LA DETECCIÓN DE PATRONES DE DESEMPEÑO EN LAS COMPETENCIAS GENÉRICAS DE LAS PRUEBAS SABER PRO 2012, 2013 Y 2014 DE LOS ESTUDIANTES DE LOS PROGRAMAS PROFESIONALES DE LA UNIVERSIDAD DE NARIÑO</p> <p>Resumen: Los estudios que se han realizado hasta el momento a nivel nacional, en el marco de las pruebas Saber Pro, se basan en información procesada mediante análisis estadístico, donde fundamentalmente se consideran variables y relaciones primarias, sin tener en cuenta las verdaderas interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento más complejo de los datos, lo cual es posible con la Minería de Datos. Por otra parte, en la Universidad de Nariño no se han planteado investigaciones que analicen el desempeño de los estudiantes de los diferentes programas profesionales que ofrece esta institución en las competencias genéricas de las pruebas SABER PRO 2012, 2013 y 2014 utilizando técnicas estadísticas o de minería de datos. En este proyecto de investigación se plantea utilizar técnicas de minería de datos para detectar patrones de desempeño en las competencias genéricas de las pruebas Saber Pro 2012, 2013 y 2014, presentadas por los estudiantes de los programas profesionales de la Universidad de Nariño, a partir de los datos sociodemográficos, económicos, académicos e institucionales almacenados en las bases de datos del ICFES. La metodología a seguir será CRISP-DM, la guía más ampliamente empleada en el desarrollo de proyectos de minería de datos, que contempla seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación. El conocimiento generado permitirá soportar la toma de decisiones de las directivas universitarias, con el fin de mejorar la calidad de la educación en la Universidad de Nariño.</p> <p>Palabras Clave: Programas profesionales, Competencias genéricas, Minería de datos, Pruebas Saber Pro, Patrones de Desempeño.</p> <p>Autores:</p> <ul style="list-style-type: none"> • Silvio Ricardo Timarán Pereira • Arsenio Hidalgo Troya • Javier Caicedo Zambrano <p>Facultades: Ciencias Exactas y Naturales, Ingeniería. Programas: Matemáticas y Estadística, Sistemas. Grupo de Investigación: GRIAS. Línea de Investigación: Detección de patrones con técnicas de minería de datos. Tipo Convocatoria: Docente Año Convocatoria: 2014</p> <p style="text-align: right;"><small>Investigaciones Relacionadas </small></p>

Calificación

E

Fuente: Creación propia

Tabla E.11 Caso de prueba 14

Búsqueda realizada	“Proyectos de investigación sobre inteligencia de negocios”
Resultado búsqueda	<p>1 MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO</p> <p>Resumen: En este proyecto de investigación se plantea desarrollar una herramienta web libre de Inteligencia de Negocios para el Descubrimiento de Conocimiento, orientada a las PYME del Departamento de Nariño. Esta herramienta estará basada en el lenguaje estadístico R, que permitirá a los usuarios de estas empresas, realizar análisis estadístico y detección de patrones útiles de manera fácil e intuitiva y visualizar los resultados de manera gráfica. El conocimiento descubierto servirá de soporte a las PYME de la región, para la toma de decisiones acertadas y oportunas que potencien su crecimiento y el nivel de competitividad. MAKIPURAY, como se denomina esta herramienta, será probada y evaluada con el repositorio de datos sobre el rendimiento académico en las competencias genéricas de las pruebas SaberPro, presentadas por los estudiantes de programas profesionales de las Universidades del Departamento de Nariño entre los años 2011 y 2014. Esta herramienta será desarrollada por un estudiante investigador del grupo GRIAS del Departamento de Sistemas, de la Facultad de Ingeniería de la Universidad de Nariño.</p> <p>Palabras Clave: PYME, Lenguaje R, Descubrimiento de Conocimiento, Inteligencia de Negocios, Herramienta Web.</p> <p>Autores: • William Geovanny Imbacuán Tutistar</p> <p>Asesor: Silvio Ricardo . Facultad: Ingeniería. Programa: Ingeniería de Sistemas. Grupo de Investigación: GRIAS. Línea de Investigación: Inteligencia de Negocios. Tipo Convocatoria: Trabajos de Grado Año Convocatoria: 2015</p> <p>Investigaciones Relacionadas </p>
Calificación	E

Fuente: Creación propia

Tabla E.12 Caso de prueba 15

Búsqueda realizada	“Proyectos de investigación sobre psicología”
Resultado búsqueda	<p>1 PROPIEDADES PSICOMÉTRICAS DE UNA PRUEBA PARA MEDIR EL PENSAMIENTO CRÍTICO A ESTUDIANTES DE PREGRADO DE COLOMBIA, MÉXICO Y CHILE.</p> <p>Resumen: El proyecto tiene por objetivo determinar las propiedades psicométricas de una prueba para medir pensamiento crítico a estudiantes de pregrado de Colombia, México y Chile. Se utilizará el paradigma cuantitativo de investigación, con un tipo de estudio instrumental ya que se realizó una validación y estandarización del instrumento sensible a la persona, aportando a la construcción de una evaluación y psicometría. Asimismo, se evaluará la construcción de la prueba para medir el pensamiento crítico.</p> <p>Palabras Clave: Estandarización, validación, estandarización, instrumento sensible a la persona, evaluación y psicometría. Asimismo, se evaluará la construcción de la prueba para medir el pensamiento crítico.</p> <p>Autores:</p> <ul style="list-style-type: none"> • Christian Alexander Zambrano • Ana Karen Ceballos Mora • Sonia Betancourt Zambrano <p>Facultad: Ciencias Humanas.</p> <p>Programa: Psicología.</p> <p>Grupos de Investigación: Psicología y Salud ,Psicología y Salud.</p> <p>Líneas de Investigación: LÍNEA DE INVESTIGACIÓN DESARROLLO DEL PENSAMIENTO ,Medición y Evaluación Psicológica.</p> <p>Tipo Convocatoria: Docente</p> <p>Año Convocatoria: 2014</p> <p>Abriendo file.pdf</p> <p>Ha elegido abrir: <input checked="" type="radio"/> file.pdf el cual es un: Formato de documento portátil (PDF) (220 KB) de: blob:</p> <p>¿Qué debería hacer Firefox con este archivo?</p> <p><input type="radio"/> Abrir con Visor de documentos (predeterminada) <input checked="" type="radio"/> Guardar archivo <input type="checkbox"/> Repetir esta decisión de ahora en adelante para este tipo de archivos.</p> <p>Cancelar Aceptar</p> <p>Investigaciones Relacionadas</p>

Calificación

E

Fuente: Creación propia

Tabla E.13 Caso de prueba 16

Búsqueda realizada

“Proyectos de investigación sobre minería de datos”

Resultado búsqueda

1 APLICACIÓN DE LA MINERÍA DE DATOS EN LA DETECCIÓN DE PATRONES DE DESEMPEÑO EN LAS COMPETENCIAS GENÉRICAS DE LAS PRUEBAS SABER PRO 2012, 2013 Y 2014 DE LOS ESTUDIANTES DE LOS PROGRAMAS PROFESIONALES DE LA UNIVERSIDAD DE NARIÑO

Resumen:

Los estudios que se han realizado hasta el momento a nivel nacional, en el marco de las pruebas Saber Pro, se basan en información procesada mediante técnicas de minería de datos, sin tener en cuenta que el análisis se realiza descubriendo utilizando un tratado teórico que no es el que se aplica en la Universidad de Nariño ni en las pruebas Saber Pro. En este trabajo se analizan los resultados de los diferentes programas profesionales de la Universidad de Nariño en los años 2012, 2013 y 2014 utilizando las técnicas de minería de datos que se aplicaron en las pruebas Saber Pro 2012, 2013 y 2014, presentando los resultados sociodemográficos, el diseño metodológico y las estrategias de evaluación y implementación de los programas profesionales universitarios, con el fin de establecer las competencias genéricas que se evalúan en las pruebas Saber Pro.

Palabras Clave: Programas profesionales, Minería de datos, Desempeño.

Autores:

- Silvio Ricardo Timarán Pereira
- Arsenio Hidalgo Troya
- Javier Caicedo Zambrano

Facultades: Ciencias Exactas y Naturales, Ingeniería.
Programas: Matemáticas y Estadística, Sistemas.
Grupo de Investigación: CRIAS.
Línea de Investigación: Detección de patrones con técnicas de minería de datos.
Tipo Convocatoria: Docente
Año Convocatoria: 2014

Abriendo file.pdf

Has elegido abrir:

file.pdf
el cual es un: Formato de documento portátil (PDF) (326 KB)
de: blob:

¿Qué debería hacer Firefox con este archivo?

Abrir con Visor de documentos (predeterminada) Guardar archivo
 Repetir esta decisión de ahora en adelante para este tipo de archivos.

Cancelar Aceptar

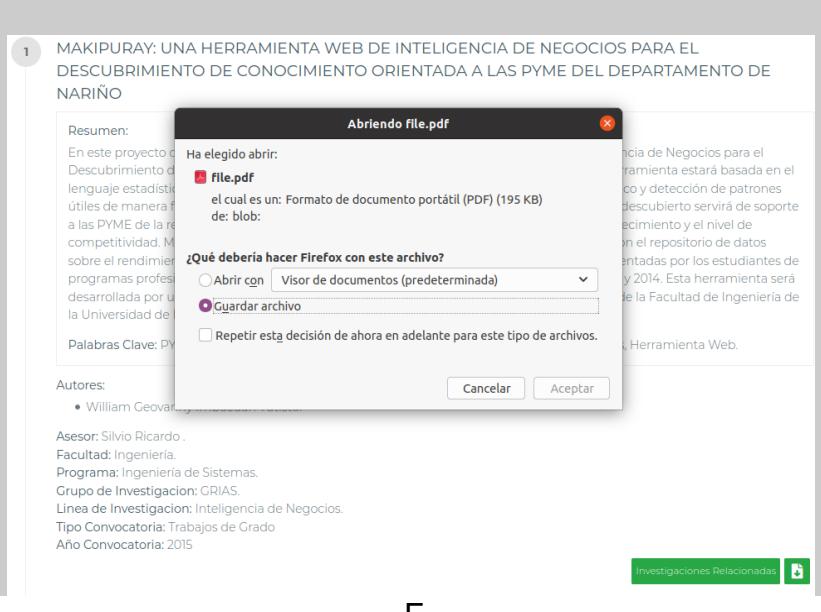
Investigaciones Relacionadas

Calificación

E

Fuente: Creación propia

Tabla E.14 Caso de prueba 17

Búsqueda realizada	“Proyectos de investigación sobre inteligencia de negocios”
Resultado búsqueda	 <p>MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO</p> <p>Resumen:</p> <p>En este proyecto se propone la creación de una herramienta web de inteligencia de negocios para el descubrimiento de conocimiento orientada a las PYME del Departamento de Nariño. La herramienta se basará en el uso de lenguaje estadístico y de aprendizaje automático para analizar y extraer información útil de manera eficiente y precisa, contribuyendo así a la competitividad de las PYME. Se evaluarán diferentes estrategias y tecnologías para mejorar el rendimiento de los sistemas de inteligencia de negocios y optimizar la ejecución de los programas profesionales. La herramienta se desarrollará por un equipo de investigadores y estudiantes de la Universidad de Nariño, con la supervisión de un profesor de la Facultad de Ingeniería.</p> <p>Palabras Clave: PYME, inteligencia de negocios, herramienta web, descubrimiento de conocimiento.</p> <p>Autores:</p> <ul style="list-style-type: none"> • William Geovanni <p>Asesor: Silvio Ricardo.</p> <p>Facultad: Ingeniería.</p> <p>Programa: Ingeniería de Sistemas.</p> <p>Grupo de Investigación: GRIAS.</p> <p>Línea de Investigación: Inteligencia de Negocios.</p> <p>Tipo Convocatoria: Trabajos de Grado</p> <p>Año Convocatoria: 2015</p> <p>Investigaciones Relacionadas </p>

Calificación

E

Fuente: Creación propia

Tabla E.15 Caso de prueba 18

Búsqueda realizada	“Proyectos de investigación sobre ciencias agrícolas”
Resultado búsqueda	
	<p>6 EVALUACIÓN DE LA ACTIVIDAD MICROBIANA EN UN SISTEMA PRODUCTIVO GANADERO EN EL MUNICIPIO DE PASTO</p> <p>7 EVALUACIÓN DEL POTENCIAL PRODUCTIVO Y LA RELACION ENTRE COMPONENTES DE RENDIMIENTO DE 20 LINEAS DE HABA (VICIA FABA) EN EL MUNICIPIO DE PASTO</p> <p>8 ANÁLISIS DE FACTORES DE APRENDIZAJE EN SEGURIDAD ALIMENTARIA CON COMUNIDADES RURALES EN EL MUNICIPIO DE PASTO, DEPARTAMENTO DE NARIÑO.</p> <p>9 EVALUACIÓN DE UN SISTEMA DOMICILIARIO DE TIPO NO CONVENCIONAL PARA EL MEJORAMIENTO DE LA CALIDAD DE AGUA DE CONSUMO HUMANO EN POBLACIÓN VULNERABLE, MUNICIPIO DE PASTO</p> <p>10 EFECTO DE ALGUNOS ABONOS VERDES SOBRE EL COMPORTAMIENTO AGRONOMICO DE CULTIVOS TRANSITORIOS EN DOS ANDISOLES DEL MUNICIPIO DE PASTO</p>

Calificación

E

Fuente: Creación propia

Tabla E.16 Caso de prueba 19

Búsqueda realizada	“Proyectos de investigación sobre química”
Resultado búsqueda	
	<p>6 ESTUDIO COMPUTACIONAL DE LA TAUTOMERÍA CETO-ENÓLICA EN BETA-CETOAMIDAS</p> <p>7 EVALUACIÓN DE LA REACCIÓN DE ALFA-ALQUILACIÓN DE LA L-PROLINA CON HALOGENUROS DE BENCILo COMO: BROMURO DE BENCILo, CLORURO DE 4-METOXBENCILo Y CLORURO DE 4-METILBENCILo</p> <p>8 ESTUDIO QUÍMICO DE LOS COMPONENTES BIOACTIVOS (BETALÁINAS) DE EXTRACTOS AISLADOS DEL FRUTO DE DOS ESPECIES DE CACTUS: ACTIVIDAD ANTIOXIDANTE Y ESTABILIDAD DE SUS MICROENCAPSULADOS.</p> <p>9 ACTIVIDAD ANTIBACTERIAL DE POLIFENOLES DEL AGUACATE (PERSEA AMERICANA VAR. HASS) SOBRE HELICOBACTER PYLORI AISLADO DE PACIENTES DE UNA ZONA DE ALTO RIESGO DE CÁNCER GASTRICO DEL DEPARTAMENTO DE NARIÑO.</p> <p>10 EVALUACIÓN DE DERIVADOS DE LA L-PROLINA COMO POTENCIALES ORGANOCATALIZADORES EN LA ALFA-ALQUILACIÓN DE ALDEHÍDOS MEDIANTE CATÁLISIS ASIMÉTRICA</p> <p style="text-align: right;">« 1 2 3 4 5 ... 18 » 1 / 18 Ir</p>

Calificación

E

Fuente: Creación propia

Tabla E.17 Caso de prueba 20

Búsqueda realizada	“Proyectos de investigación sobre cultivos”
Resultado búsqueda	<p>6 EFECTO DEL INCREMENTO DEL DIÓXIDO DE CARBONO CO₂, SOBRE LA PRODUCTIVIDAD DE LA REMOLACHA FORRAJERA (BETA VULGARIS) CULTIVADA EN MICRO TÚNELES EN LA GRANJA EXPERIMENTAL DE BOTANA, UNIVERSIDAD DE NARIÑO.</p> <p>7 ÍNDICE DE CIUDAD CREATIVA PARA LA CIUDAD DE PASTO</p> <p>8 EVALUACIÓN DE VARIABLES AGRONÓMICAS MEDIANTE VEHÍCULOS AÉREOS NO TRIPULADOS EN DOS ESPECIES DE IMPORTANCIA AGRONÓMICA EN EL MUNICIPIO DE PASTO, DEPARTAMENTO DE NARIÑO.</p> <p>9 EFECTO DE DIFERENTES NIVELES DE QUINUA (CHENOPODIUM QUINOA) ADICIONADA A LA DIETA SOBRE EL CRECIMIENTO DE TILAPIA ROJA (OREOCHROMIS SP)</p> <p>10 FERTILIZACIÓN NPK DE PHYSALIS PERUVIANA ASOCIADA CON SU POTENCIAL DE CRECIMIENTO, NUTRICIONAL Y PRODUCTIVO</p>

« 1 2 3 4 5 ... 10 » 1 / 10 Ir

Calificación

E

Fuente: Creación propia

Tabla E.18 Caso de prueba 21

Búsqueda realizada	“Proyectos de investigación sobre psicología	
Resultado búsqueda		
	<p>1 PROPIEDADES PSICOMÉTRICAS DE UNA PRUEBA PARA MEDIR EL PENSAMIENTO CRÍTICO A ESTUDIANTES DE PREGRADO DE COLOMBIA, MÉXICO Y CHILE.</p> <p>1 EFECTOS DE UN PROGRAMA DE INTERVENCIÓN BASADO EN EL DEBATE CRÍTICO SOBRE EL PENSAMIENTO CRÍTICO DE LOS ESTUDIANTES DE PSICOLOGÍA DE LA UNIVERSIDAD DE NARIÑO</p> <p>2 COMPORTAMIENTO PROECOLÓGICO Y AUSTERO EN EL PERSONAL ADMINISTRATIVO, EN EL MARCO DEL SISTEMA DE GESTIÓN AMBIENTAL DE LA UNIVERSIDAD DE NARIÑO</p> <p>3 CARACTERIZACIÓN DE LA CONVIVENCIA ESCOLAR DE LAS INSTITUCIONES EDUCATIVAS EN NARIÑO</p> <p>4 AUTOGESTIÓN INSTITUCIONAL FRENTE AL RIESGO VOLCÁNICO DEL GALERAS EN LA INSTITUCIÓN EDUCATIVA SAN BARTOLOMÉ DEL MUNICIPIO DE LA FLORIDA (NARIÑO - COLOMBIA)</p> <p>5 EFECTOS DEL PENSAMIENTO CRÍTICO EN ESTUDIANTES UNIVERSITARIOS NARIÑO</p> <p>4 EMPODERAMIENTO Y AUTOGESTIÓN INSTITUCIONAL FRENTE AL RIESGO VOLCÁNICO DEL GALERAS EN LA INSTITUCIÓN EDUCATIVA SAN BARTOLOMÉ DEL MUNICIPIO DE LA FLORIDA (NARIÑO - COLOMBIA)</p>	<p>CRÍTICO A</p> <p>nto crítico a con un tipo de ucción, cción de un o y Chile, edición y de</p> <p>nto crítico.</p> <p>Psicológica.</p> <p>lacionadas B</p> <p>VO, EN EL</p> <p>SOBRE EL</p> <p>Cerrar</p>

Calificación

E

Fuente: Creación propia

Tabla E.19 Caso de prueba 22

Búsqueda realizada	“Proyectos de investigación sobre minería de datos”	
<p>Resultado búsqueda</p> <p>1 APLICACIÓN DE LA MINERÍA DE DATOS EN LA DETECCIÓN DE PATRONES DE DESEMPEÑO EN LAS COMPETENCIAS GENÉRICAS DE LAS PRUEBAS SABER PRO 2012, 2013 y 2014 DE LOS ESTUDIANTES DE LOS PROGRAMAS PROFESIONALES DE LA UNIVERSIDAD DE NARIÑO</p> <p>2 APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA EL DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 5º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO</p> <p>3 DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 11º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO CON TÉCNICAS DESCRIPTIVAS DE MINERÍA DE DATOS</p> <p>4 MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO</p> <p>5 UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS PRUEBAS SABER 5 DE LAS INSTITUCIONES EDUCATIVAS DE LOS MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO</p> <p>6 IMARAÑA: UNA HERRAMIENTA WEB DE ANALÍTICA DE DATOS BASADA EN EL LENGUAJE R PARA SOPORTAR LA TOMA DE DECISIONES EN LAS PYMES DEL DEPARTAMENTO DE NARIÑO</p>	<p>EMPEÑO EN E LOS NARIÑO</p> <p>basan en</p> <p>se pueden</p> <p>otra parte, en</p> <p>de los</p> <p>SABER PRO</p> <p>antea utilizar</p> <p>ebas Sabor Pro</p> <p>la partir de los</p> <p>CFES. La</p> <p>mería de datos,</p> <p>elado,</p> <p>ctivas</p>	<p>X</p> <p>Resumen</p> <p>Los estu informa primaria descubri la Unive diferentes 2012, 2013 técnicas 2012, 2013 datos so metodos que con evaluaci universi</p> <p>Palabras Desempeñ</p> <p>Autores:</p> <ul style="list-style-type: none"> • Silvid • Arser • Jayvie <p>Facultad</p> <p>Programa</p> <p>Grupo de</p> <p>Línea de</p> <p>Tipo Conv</p> <p>Año Conv</p> <p>CONSTITU</p> <p>DE EVE</p> <p>APLICA</p> <p>ASOCIA</p> <p>DE LAS</p> <p>DESAR</p> <p>INFOR</p> <p>lacionadas</p> <p>PATRONES</p> <p>DATOS</p> <p>FACTORES</p> <p>ESTUDIANTES</p> <p>DE</p>

Calificación

E

Fuente: Creación propia

Tabla E.20 Caso de prueba 23

Búsqueda realizada	“Proyectos de investigación sobre inteligencia de negocios”
Resultado búsqueda	<p>MAKIPURAY: UNA HERRAMIENTA WEB DE INTELIGENCIA DE NEGOCIOS PARA EL DESCUBRIMIENTO DE CONOCIMIENTO ORIENTADA A LAS PYME DEL DEPARTAMENTO DE NARIÑO</p> <p>IMARAÑA: UNA HERRAMIENTA WEB DE ANALÍTICA DE DATOS BASADA EN EL LENGUAJE R PARA SOPORTAR LA TOMA DE DECISIONES EN LAS PYMES DEL DEPARTAMENTO DE NARIÑO</p> <p>ANÁLISIS DE FUNCIONALIDAD DE LA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS DE SOFTWARE LIBRE SPAGO BI ORIENTADA AL USO EN LAS PYMES DE LA REGIÓN</p> <p>APLICACIÓN DE LA MINERÍA DE DATOS EN LA DETECCIÓN DE PATRONES DE DESEMPEÑO EN LAS COMPETENCIAS GENÉRICAS DE LAS PRUEBAS SABER PRO 2012, 2013 Y 2014 DE LOS ESTUDIANTES DE LOS PROGRAMAS PROFESIONALES DE LA UNIVERSIDAD DE NARIÑO</p> <p>DESCUBRIMIENTO DE FACTORES ASOCIADOS AL DESEMPEÑO ACADÉMICO EN LAS PRUEBAS SABER 11º DE LOS ESTUDIANTES DE LAS INSTITUCIONES EDUCATIVAS DEL DEPARTAMENTO DE NARIÑO CON TÉCNICAS DESCRIPTIVAS DE MINERÍA DE DATOS</p> <p>UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS PRUEBAS SABER 5 DE LAS INSTITUCIONES EDUCATIVAS DE LOS MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE NARIÑO</p>

Calificación

E

Fuente: Creación propia