

UNIVERSIDAD TÉCNICA DEL NORTE



Facultad de Ingeniería en Ciencias Aplicadas
Carrera de Ingeniería en Sistemas Computacionales

**DESARROLLO DE UN SISTEMA WEB PARA LA GENERACIÓN
AUTOMÁTICA DE REVISIÓN DE LITERATURA, APLICANDO TÉCNICAS DE
INTELIGENCIA ARTIFICIAL.**

Trabajo de grado previo a la obtención del título de
Ingeniero en Sistemas Computacionales

Autor:

Alexander Patricio Pinchao Pujota

Director:

PhD. Irving Reascos

Ibarra-Ecuador



UNIVERSIDAD TÉCNICA DEL NORTE
BIBLIOTECA UNIVERSITARIA

**AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA
UNIVERSIDAD TÉCNICA DEL NORTE**

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago entrega del presente trabajo a la Universidad Técnica el Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a su disposición la siguiente información:

DATOS DE CONTACTO	
CÉDULA DE IDENTIDAD:	040193926-9
APELLIDOS Y NOMBRES:	PINCHAO PUJOTA ALEXANDER PATRICIO
DIRECCIÓN:	El Olivo – Ibarra, Profesor Rufo Simbaña & Lic. Nelson Dávila
EMAIL:	appinchaop@utn.edu.ec
TELÉFONO MÓVIL:	0980407213

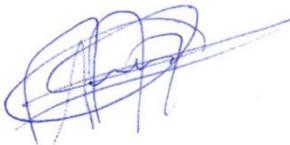
DATOS DE LA OBRA	
TÍTULO:	DESARROLLO DE UN SISTEMA WEB PARA LA GENERACIÓN AUTOMÁTICA DE REVISIÓN DE LITERATURA, APLICANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL
AUTOR (ES):	PINCHAO PUJOTA ALEXANDER PATRICIO
DIRECCIÓN:	El Olivo – Ibarra, Profesor Rufo Simbaña & Lic. Nelson Dávila
FECHA:	17/05/2021
PROGRAMA:	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
TÍTULO POR EL QUE OPTA:	INGENIERO EN SISTEMAS COMPUTACIONALES
ASESOR /DIRECTOR:	MSC. IRVING REASCOS

2. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y la desarrolló sin violar los derechos de autor de terceros, por lo tanto, la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá de la defensa de la Universidad en caso de reclamos por parte de terceros.

Ibarra, a los 25 días del mes de mayo del 2021.

AUTOR:



Alexander Patricio Pinchao Pujota

CI: 0401939269



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

Ibarra, 17 de mayo del 2021

CERTIFICACIÓN DEL DIRECTOR

Por medio del presente, yo **Irving Reascos Paredes**, certifico que el Sr. **Alexander Patricio Pinchao Pujota**, portador de la cédula de ciudadanía Nro. **040193926-9**, ha trabajado en el desarrollo del proyecto de grado denominado **“DESARROLLO DE UN SISTEMA WEB PARA LA GENERACIÓN AUTOMÁTICA DE REVISIÓN DE LITERATURA, APLICANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL”**, previo a la obtención del título de Ingeniería en Sistemas Computacionales, lo cual ha realizado en su totalidad con responsabilidad y esmero.

Es todo cuanto puedo certificar en honor a la verdad.

Atentamente,



Firmado electrónicamente por:
IRVING MARLON
REASCOS PAREDES

PhD. Irving Reascos
DIRECTOR DE TESIS

DEDICATORIA

El presente trabajo le dedico a mis padres, principalmente por su apoyo incondicional, por sus consejos, por su esfuerzo y además por confiar en mí, gracias a ellos he logrado cumplir una meta más en mi vida.

A mis hermanos quienes además de su apoyo y comprensión, han sido una motivación para seguir siempre adelante.

A mis familiares y amigos que siempre supieron estar en el momento que los necesitaba brindándome unas palabras de aliento, para todos ellos dedico esta tesis.

Patricio Pinchao.

AGRADECIMIENTO

Agradezco a mis padres y hermanos por creer en mí y por enseñarme el valor de la humildad, el sacrificio y la perseverancia.

A mi tutor Msc. Irving Reascos y opositores quienes me brindaron su apoyo para poder realizar con éxito este trabajo.

A todos los docentes que supieron brindarme sus conocimientos y consejos para que sea un excelente profesional.

Patricio Pinchao.

TABLA DE CONTENIDO

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE	i
CERTIFICACIÓN DEL DIRECTOR.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
TABLA DE CONTENIDO	vi
ÍNDICE DE FIGURAS	viii
ÍNDICE DE TABLAS	x
RESUMEN	xi
ABSTRACT.....	xii
INTRODUCCIÓN	1
Problema	1
Antecedentes:	1
Situación Actual:.....	1
Planteamiento del Problema:.....	1
Objetivos	2
Objetivo General	2
Objetivos Específicos.....	2
Alcance.....	2
Justificación	4
CAPÍTULO 1	7
Marco Teórico.	7
1.1. Proceso de revisión de literatura.....	7
1.2. Minería de texto, técnicas y software: Una breve revisión de literatura.....	9
1.2.1. Metodología	9
1.2.2. Minería de texto	11
1.2.3. Software de minería de texto.	16
1.2.4. Uso de minería de texto para realizar revisión de literatura.	19
1.3. Modelo de éxito de los sistemas de información de Delone y Mclean.....	22
CAPÍTULO 2	25
Desarrollo.....	25
2.1. Fase de Iniciación Scrum.	25
2.1.1. Definición de Roles Scrum	25
2.1.2. Definición del Product Backlog.	25
2.2. Fase 2. – Panificación y estimación	27
2.2.1. Definición de historias de usuario.	27

2.2.2.	Planificación del proyecto.....	30
2.3.	Desarrollo del proyecto	31
2.3.1.	Planificación del sprint 1.....	31
2.3.2.	Ejecución del Sprint 1	33
2.3.3.	Planificación del sprint 2.....	38
2.3.4.	Ejecución del Sprint 2	40
2.3.5.	Planificación del sprint 3.....	43
2.3.6.	Ejecución del Sprint 3	43
2.3.7.	Planificación del sprint 4.....	45
2.3.8.	Ejecución del Sprint 4	46
2.3.9.	Planificación del sprint 5.....	47
2.3.10.	Ejecución del Sprint 5	48
2.3.11.	Planificación del sprint 6.....	50
2.3.12.	Ejecución del Sprint 6	51
CAPÍTULO 3.	54
Resultados.....	54
3.1.	Sección 1: Perfil demográfico de los usuarios.....	54
3.2.	Sección 2: Variables del modelo de éxito de los sistemas de información de Delone y Mclean	55
3.2.1.	Calidad del Sistema.....	56
3.2.2.	Calidad de la información.	57
3.2.3.	Calidad del servicio	58
3.2.4.	Uso - Intensión de uso	59
3.2.5.	Satisfacción del usuario	60
3.2.6.	Beneficios obtenidos	61
CONCLUSIONES	63
RECOMENDACIONES.....	65
REFERENCIAS.....	66
ANEXOS.....	69

ÍNDICE DE FIGURAS

Figura 1: Árbol de problemas.	2
Figura 2. Arquitectura software.	4
Figura 1.1. Proceso de Búsqueda y filtrado.	10
Figura 1.2. Técnicas de clasificación del texto.	15
Figura 1.3. Modelo de éxito de los sistemas de información.	23
Figura 2.1. Diseño Login	33
Figura 2.2. Diseño Registro.....	33
Figura 2.3. Tabla de usuarios.....	34
Figura 2.4. Vista del Login.	34
Figura 2.5. Vista del Registro.....	35
Figura 2.6. Validación de formularios de login y registro.	36
Figura 2.7. Vista del listado de análisis.....	37
Figura 2.8. Tabla de análisis.....	37
Figura 2.9. Vista del listado de análisis.....	38
Figura 2.10. Tabla de archivos (files).	40
Figura 2.11. Vista de Gestión de archivos.....	40
Figura 2.12. Formulario de ingreso de archivos	41
Figura 2.13. Tabla de conceptos (concepts).	41
Figura 2.14. Diseño de la vista de gestión de conceptos.	42
Figura 2.15. Formulario de ingreso de conceptos.	42
Figura 2.16. Formulario de edición de conceptos.	42
Figura 2.17. Extracción de texto de documentos.	44
Figura 2.18. Texto después de la limpieza de datos.	44
Figura 2.19. Tabla de resultados (results).	46
Figura 2.20. Tabla de oraciones (sentences).	46
Figura 2.21. Comparación de conceptos con oraciones.	47
Figura 2.22. Diseño de la vista de Matriz de conceptos.	49
Figura 2.23. Vista de matriz de conceptos.....	49
Figura 2.24. Modal de oraciones encontradas.	50
Figura 2.25. Modificación de la vista de matriz de conceptos.....	52
Figura 2.26. Lista de las oraciones con el porcentaje de coincidencia.....	52
Figura 2.27. Actualizar análisis finalizado.....	53
Figura 3.1. Información de género.	55
Figura 3.2. Perfil de edad.	55
Figura 3.3. Calidad del sistema.....	56

Figura 3.4. Calidad de la información.	57
Figura 3.5. Calidad de servicio.	58
Figura 3.6. Uso -Intención de uso.	59
Figura 3.7. Satisfacción del usuario.	60
Figura 3.8. Beneficios obtenidos.	61
Figura 4.1. Encuesta.	69
Figura 4.2. Resultados de la encuesta.	70
Figura 4.3. Estructura final de la base de datos.	71
Figura 4.4. Descarga de Balsamic.	71
Figura 4.5. Instalación de Balsamic.	72
Figura 4.6. Interfaz de Balsamic.	72
Figura 4.7. Descarga de Python.	73
Figura 4.8. Instalación de Python.	73
Figura 4.9. Proceso de instalación de Python.	74
Figura 4.10. Descarga de Visual Studio Code.	74
Figura 4.11. Aceptar términos de licencia de Visual Code.	75
Figura 4.12. Seleccionar carpeta de instalación Visual Code.	75
Figura 4.13. Instalación de Visual Studio Code.	76
Figura 4.14. Interfaz de Visual Studio Code.	76
Figura 4.15. Descarga de Xampp.	77
Figura 4.16. Instalación de Xampp.	77
Figura 4.17. Ejecución de servicios en Xampp.	78

ÍNDICE DE TABLAS

Tabla 1: Matriz de conceptos.	11
Tabla 2.1: Roles Scrum.	25
Tabla 2.2: Estimación de esfuerzo (T-Shirt Size)	26
Tabla 2.3: Definición del Product Backlog	26
Tabla 2.4: Historia de Usuario 1 - <i>Login</i>	27
Tabla 2.5: Historia de Usuario 2 – Gestión de Análisis.	27
Tabla 2.6: Historia de Usuario 3 – Gestión de documentos.....	28
Tabla 2.7: Historia de Usuario 4 – Gestión de conceptos.	28
Tabla 2.8: Historia de Usuario 5 - Extracción del texto de los documentos.....	28
Tabla 2.9: Historia de Usuario 6 - Limpieza y preprocesamiento de datos.....	29
Tabla 2.10: Historia de Usuario 7 - Implementación del modelo de inteligencia artificial.	29
Tabla 2.11: Historia de Usuario 8 - Reporte de matriz de conceptos.	29
Tabla 2.12: Historia de Usuario 9 – Refinamiento del Reporte de matriz de conceptos v.2.....	30
Tabla 2.13: Aprobación del Sprint	30
Tabla 2.14: Sprint 1 – Desarrollo del login, registro y gestión de análisis.....	32
Tabla 2.15: Sprint 2 – Desarrollo de la gestión de documentos y conceptos.	39
Tabla 2.16: Sprint 3 – Extracción del texto, limpieza y preprocesamiento.	43
Tabla 2.17: Sprint 4 - Implementación del modelo de inteligencia artificial.	45
Tabla 2.18: Sprint 5 - Reporte de matriz de conceptos.	48
Tabla 2.19: Sprint 6 - Refinamiento del reporte de matriz de conceptos v.2.	51

RESUMEN

Actualmente, la revisión de bases de datos bibliográficas en la comunidad científica y/o educativa es un proceso arduo que conlleva mucho tiempo y esfuerzo por parte del investigador, por ello muchas investigaciones tienden a tener bajo impacto, ya que poseen información poco relevante o deficiente. Para mitigar esto es muy importante identificar la información existente en la literatura que pueda contribuir para el desarrollo y éxito de cualquier investigación de calidad.

El presente trabajo de titulación tiene como objetivo el desarrollo de una aplicación web que permita automatizar parte del proceso que se realiza en una revisión de literatura, enfocado principalmente en la generación de una matriz de conceptos a partir de artículos implementando técnicas de inteligencia artificial.

Como primera parte se realizó una revisión bibliográfica referente al procesamiento y análisis de texto con el uso de inteligencia artificial, así como también algunos métodos que se podían utilizar para dar solución al problema.

Previo al desarrollo del sistema se configuró el entorno de trabajo, en donde se usó Python como lenguaje de programación con el framework Flask, como motor de base de datos se utilizó MySQL y el IDE de Visual Studio Code. Para el desarrollo de la aplicación usamos la metodología Scrum, para lo cual se definió los roles, las historias de usuario y se planificó los distintos sprints hasta terminar con el desarrollo.

Para la validación del sistema en general se aplicó una encuesta basada en el modelo de éxito de los sistemas de información planteado por Delone y MacLen en donde se evalúa la calidad del sistema, calidad de la información, calidad del servicio, uso, intención de uso, satisfacción de uso y los beneficios netos, de lo cual se obtuvo respuestas positivas con respecto al sistema.

Palabras Claves: Revisión de literatura, matriz de conceptos, inteligencia artificial, minería de texto, Scrum.

ABSTRACT

Currently, the bibliographic databases review in the scientific and / or educational community is an arduous process that involves a lot of time and effort by the researcher, for this reason, many investigations tend to have low impact, since they have little relevant or deficient information. To mitigate this, it is very important to identify the existing information in the literature that can contribute to the development and success of any quality research.

The degree work objective is to develop a web application that allows automating part of the process that is carried out in a literature review, focused mainly on the generation of a concepts matrix from articles using artificial intelligence techniques.

As the first part, a bibliographic review was carried out regarding the processing and analysis of text with the use of artificial intelligence, as well as some methods that could be used to solve the problem.

Before developing the system, the work environment was configured, where Python was used as a programming language with the Flask framework, MySQL and the Visual Studio Code IDE were used as the database engine. For the development of the application, we used the Scrum methodology, for which the roles and user stories were defined, and the different sprints were planned until the development was finished.

For the system validation in general, a survey based on the information systems success model proposed by Delone and MacLen was applied, where the system quality, information quality, service quality, use, intention to use is evaluated, user satisfaction and net benefits, where positive responses were obtained regarding the system.

Keywords: Literature review, concept matrix, artificial intelligence, text mining, Scrum.

INTRODUCCIÓN

Problema

Antecedentes:

Cuando se realizan revisiones de la literatura una de las prioridades es la necesidad de encontrar información relevante que aporte en el desarrollo del proyecto de investigación (Calle, 2016), para lograr encontrar esta información se busca artículos y se analiza cada uno de ellos mediante una lectura comprensiva, esto conlleva mucho tiempo y esfuerzo por parte del investigador (Gonzalez, 2011), las primeras revisiones de literatura se realizaron en base a instrucciones y percepciones personales, sin la necesidad de leer mucho por la poca cantidad de documentos publicados, pero día tras día la base de conocimiento se ha ido expandiendo (Baumeister, 2013), por ello resulta más difícil realizar revisiones eficientes y productivas de un determinado tema en cortos plazos de tiempo.

Situación Actual:

Para definir el estado del arte en un determinado tema a investigar, se debe realizar un análisis crítico de varios artículos y sacar los temas relevantes de cada uno de ellos, lo que se convierte en algo tedioso que conlleva mucho tiempo (Mohammed & Ahmed, 2019), esto debido a la duplicidad de contenidos encontrados en las bases de datos bibliográficas y contextos no definidos de una manera clara, es decir que la redacción usada no es comprensible, lo que genera confusión en el lector logrando que el proceso de investigación no sea eficiente y por ende los investigadores no sean productivos.

Planteamiento del Problema:

Al realizar una búsqueda de literatura en las bases de datos bibliográficas estas nos arrojan una gran cantidad de resultados los cuales tiene una cierta coincidencia con el tema de investigación (University of Leicester, 2009). Para identificar los documentos con el contenido más relevante que aporten a la investigación se hace un análisis de forma manual, esto implica un mayor esfuerzo y tiempo para su realización, además que en algunas ocasiones los documentos no son de gran utilidad por lo cual la investigación tarde más de lo estimado.

La figura 1 representa un árbol de problemas, en donde se puede observar las causas que están directamente relacionadas con el problema central, ya que son las que determinan e influyen en la aparición del problema, además se especifica los efectos o consecuencias producidas por el problema.

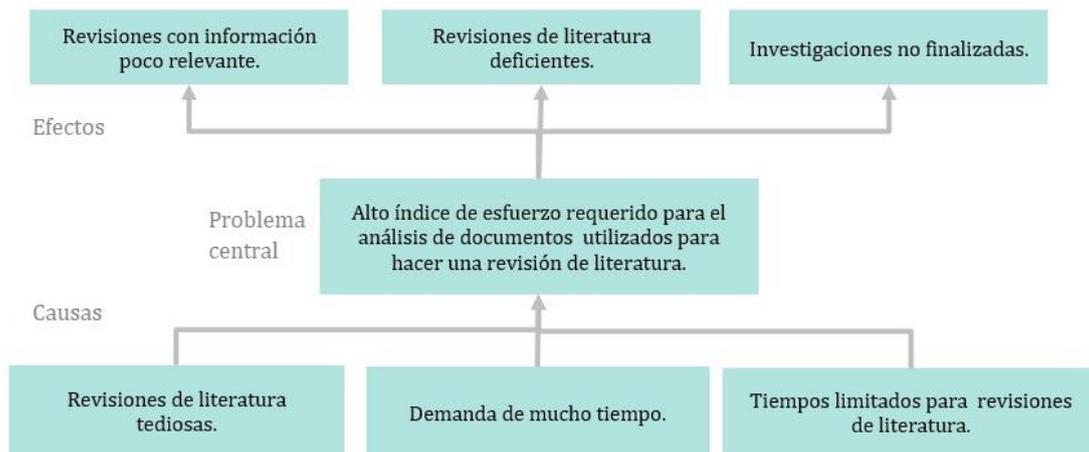


Figura 1: Árbol de problemas.

Objetivos

Objetivo General

Desarrollar una aplicación web para el análisis y extracción de información relevante de artículos de investigación aplicando técnicas de análisis de texto e inteligencia artificial.

Objetivos Específicos

- Construir un marco teórico referente al procesamiento y análisis de texto con el uso de inteligencia artificial.
- Desarrollar una aplicación web que realice el procesamiento de texto a partir de documentos *pdf*, mediante el uso de algoritmos y librerías de Python para el análisis de texto, aplicando Scrum como marco de trabajo.
- Validar los resultados basados en el modelo de éxito de sistemas de información de Delone and Mclean.

Alcance

Basado en la revisión de la literatura, se construirá un marco teórico referente al procesamiento y análisis de texto con el uso de inteligencia artificial.

Desarrollo de un sistema web para un ambiente local que permita el ingreso o carga de documentos en formato *pdf* y los conceptos a buscar. En base a esta información, el sistema realiza el procesamiento y devuelve una matriz de conceptos de los temas incluidos en los documentos.

Como este sistema es la primera versión, el análisis que realizará la aplicación será muy generalizado, sin entrar en muchos detalles, además de analizar únicamente documentos en inglés.

En esta aplicación no se considerará incluir documentos con contenido de videos, sonido y animación; solo aceptará documentos cuyo contenido sea texto e imagen, además debemos indicar que el contenido de las imágenes no será considerado.

Las técnicas de inteligencia artificial que se usarán son: la técnica de clasificación para identificar el tema o tópico de los distintos documentos a analizar, y la técnica de extracción para identificar las palabras claves. Con esta información obtenida se realizará una comparativa con los conceptos introducidos y se generará la matriz de conceptos.

Finalmente, la aplicación será evaluada usando el modelo de éxito de sistemas de información de Delone y Mclean, para lo cual se usarán cuestionarios previamente elaborados para este fin.

En la figura 2 se detalla el procesamiento interno que va a realizar la aplicación, como es la carga de información, el procesamiento y el reporte de resultados, además la figura muestra la arquitectura que se va a usar en el desarrollo, en este caso Python que es el lenguaje de programación, la metodología Scrum como marco de desarrollo y el modelo de Delone y Mclean con el cual se va a realizar la validación del sistema.

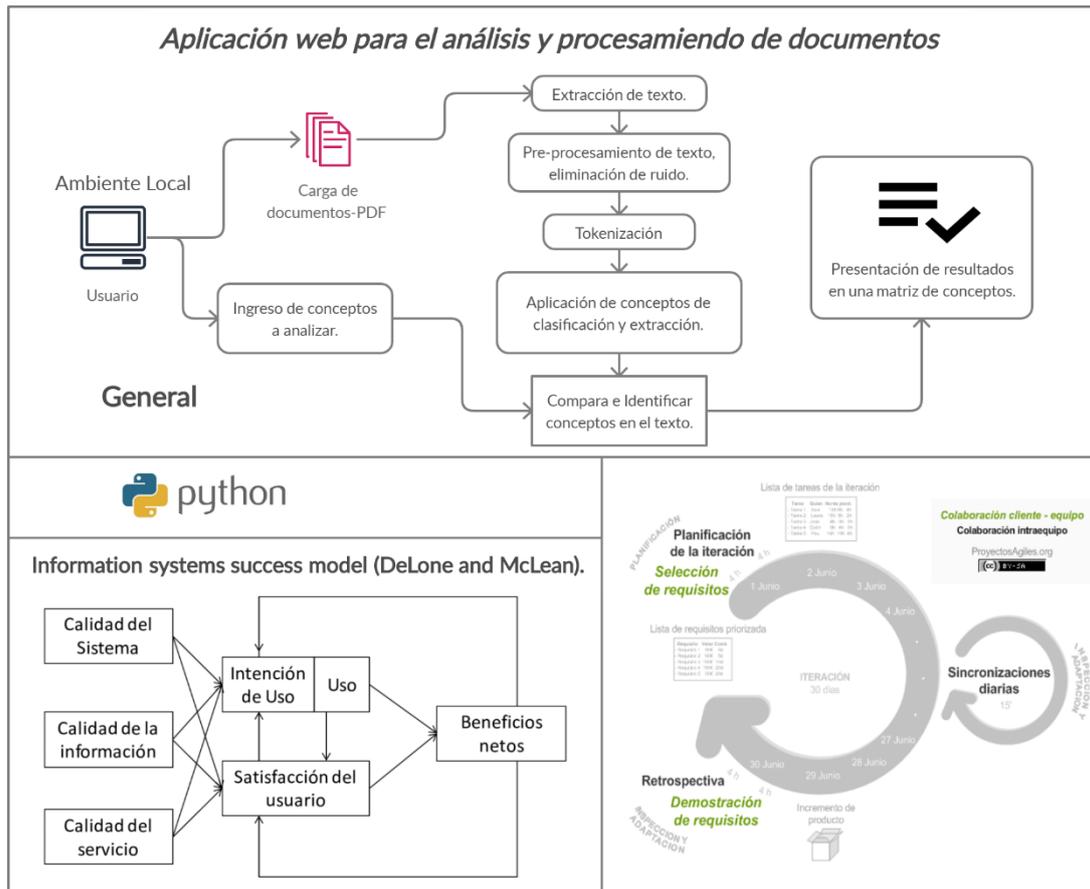


Figura 2. Arquitectura software.
Adaptado de (DeLone & McLean, 2016; Albaladejo, 2008).

Justificación

Con el presente trabajo de titulación se busca disminuir el esfuerzo en la búsqueda de información para realizar revisiones de literatura. Además de tener un enfoque hacia los objetivos de desarrollo sostenible (ODS) de las Naciones Unidas (2015) y el Plan Nacional de Desarrollo 2017 – 2021 Toda una Vida – Secretaría Técnica Planifica Ecuador (2017) en sus literales:

ODS objetivo N° 9: Industria, Innovación e Infraestructura:

Las inversiones en infraestructura (transporte, regadío, energía, tecnología de la información y las comunicaciones) son fundamentales para lograr un desarrollo sostenible y empoderar a las sociedades de numerosos países. El progreso tecnológico debe estar en la base de los esfuerzos para alcanzar los objetivos medioambientales, como el aumento de los recursos y la eficiencia energética. Sin tecnología e innovación, la industrialización no ocurrirá, y sin industrialización, no habrá desarrollo. Es necesario invertir más en productos de alta tecnología que dominen las producciones manufactureras para aumentar la eficiencia

y mejorar los servicios celulares móviles para que las personas puedan conectadas. (United Nations, 2015)

El objetivo 9.5, Aumentar la investigación científica y mejorar la capacidad tecnológica de los sectores industriales de todos los países, en particular los países en desarrollo, entre otras cosas fomentando la innovación y aumentando considerablemente, de aquí a 2030, el número de personas que trabajan en investigación y desarrollo por millón de habitantes y los gastos de los sectores públicos y privados en investigación y desarrollo. (United Nations, 2015)

9.b Apoyar el desarrollo de tecnologías, la investigación y la innovación nacional en los países en desarrollo, incluso garantizando un entorno normativo propicio a la diversificación industrial y la adición de valor a los productos básicos, entre otras cosas. (United Nations, 2015)

Plan Nacional Toda una vida Objetivo 7: Incentivar una sociedad participativa, con un Estado cercano al servicio de la ciudadanía.

Una ciudadanía activa y participativa. Promovemos la organización autónoma de la sociedad, con mejor y mayor participación ciudadana, impulsando un diálogo que va más allá del ámbito político.

Promovemos las libertades (expresión, pensamiento, culto, conciencia, asociación) para construir una sociedad sobre la base de la asociatividad y confianza interpersonal, con un Estado que garantiza los derechos (Plan Nacional de Desarrollo 2017 – 2021 Toda una Vida – Secretaría Técnica Planifica Ecuador, 2017).

Justificación Tecnológica. – La aplicación ayudará a los investigadores hacer un análisis automático de documentos con uso de inteligencia artificial y análisis de textos, mostrándole como resultado una matriz con los conceptos y documentos introducidos.

Justificación Teórica.

Con el análisis de documentos que proporcionará la aplicación, las investigaciones y revisiones de literatura se realizaran en menos tiempo y de mejor calidad, motivando de esta manera a otras personas a investigar y estar actualizados.

Justificación Ambiental.

En muchas ocasiones los investigadores imprimen los distintos artículos a analizar para leer cada uno de ellos e ir señalando de mejor manera los conceptos, esto se puede

evitar gracias al desarrollo de la aplicación ya que genera ese análisis sin la necesidad de desperdiciar muchas hojas de papel.

Justificación Social.

Con el desarrollo de la aplicación se promueve la participación de la sociedad como tal en proyectos de investigación, además de ayudar a las instituciones de educación filtrar de mejor manera los contenidos más importantes para enseñar a los estudiantes.

CAPÍTULO 1

Marco Teórico.

1.1. Proceso de revisión de literatura.

Una revisión de literatura relevante es esencial cuando se desarrolla un proyecto académico, pues crea una base firme para comprender y avanzar con el conocimiento, al igual que permite cerrar áreas donde existe bastante información y descubre áreas por ser investigadas (Webster & Watson, 2002).

A continuación se describe como desarrollar una revisión de literatura según Webster y Watson (2002):

El tema es la primera parte que conforma una revisión de literatura, aunque en muchas ocasiones este no se define sino hasta que se finaliza el contenido del artículo.

A continuación se debe establecer los límites de la investigación, como el nivel de análisis que se aplicará, el período de tiempo considerado, la limitación del contexto al que se va a enfocar la investigación, además de especificar el alcance de la revisión. En esta parte se define la unidad de análisis para la revisión de literatura.

Identificando la literatura relevante

Para realizar una revisión de calidad, esta debe cubrir toda la literatura relevante sobre un determinado tema.

Recomendaciones para la búsqueda de los documentos:

- Lo primero es buscar en conferencias y revistas reconocidas ya que allí es donde se encuentra la mayoría cantidad de documentos relevantes para la investigación.
- Una vez que se tenga algunos artículos, se puede revisar las citas de cada documento para encontrar otros artículos que puede aportar significativamente en la revisión.
- En algunas ocasiones las bases de datos bibliográficas muestran que artículos más actuales han citado a los artículos resultantes de la búsqueda.

Se puede determinar que el área de análisis está relativamente completa cuando no se encuentra nuevos conceptos en el conjunto de documentos.

Estructurando la Revisión

La revisión de literatura se centra en los conceptos que son los que determinan el marco organizativo de la revisión. Para comunicar mejor al lector la información encontrada, esta se le puede presentar en tablas o gráficos para una mayor comprensión.

Tono

Hay que tratar que la revisión de literatura presente información constructiva al lector sobre lo que se ha encontrado, en el caso de hablar de algunos estudios anteriores no hay que ser muy críticos, ya que criticar es fácil y de poco valor, es mejor explicar cómo la investigación se basa en hallazgos anteriores. Hay que tomar en cuenta que toda investigación es defectuosa por ello se debe observar si los artículos están bien realizados para incluirlos en la revisión.

Tiempo

Aunque una revisión de literatura se puede comunicar de forma efectiva tanto en tiempo presente como en tiempo pasado, elegimos el presente porque el lector percibe de esta manera que la literatura es mejor, además cuando se discuten conceptos es lógico utilizar tiempo presente, y si la literatura está en tiempo presente el lector procesa más rápido la información, cabe señalar que hay algunas excepciones, como por ejemplo para citar la opinión de un autor, en ese caso se utiliza el tiempo pasado.

Desarrollo teórico del artículo

En una revisión no se necesita únicamente encontrar lo que existe de un determinado tema, sino poder identificar brechas importantes de conocimiento, lo que se necesita saber respecto a un tema específico y no se han realizado investigaciones aún, de esta forma poder motivar a los lectores a investigar y desarrollar nuevas teorías de algún área contribuyendo de esta forma a las futuras investigaciones, a pesar de que la extensión o desarrollo de nuevas teorías es una tarea muy difícil.

Evaluando la teoría

Una vez finalizada la teoría, para evaluar su calidad algunos autores comparten que ésta debe ser interesante, además de mostrar creatividad y relevancias, que a la hora de leer se comprenda con claridad los conceptos que se desea dar a conocer, puesto que los

investigadores buscan teorías. Una buena práctica es enviar el artículo a otras personas, como amigos o familiares que puedan leer y dar una retroalimentación antes de enviar para la revisión, esto ayuda a que surjan nuevas ideas o se complemente lo que se quiere expresar ya que con cada revisión el documento se refinará de mejor manera.

Creando su discusión y conclusiones

En algunos artículos de revisión de literatura no se hace mucho énfasis en las conclusiones, sin embargo, en este punto aún se puede dar a conocer más detalles y hacer énfasis en las contribuciones, como también proponer nuevos temas de investigación.

1.2. Minería de texto, técnicas y software: Una breve revisión de literatura.

1.2.1. Metodología

En base a los conceptos requeridos se estableció cinco preguntas de investigación que se presentan a continuación:

- ¿Cuáles son los métodos o técnicas para la minería de texto?
- ¿Software existente para la revisión automática de literatura?
- ¿Cuáles son los métodos o técnicas para el procesamiento natural del lenguaje?
- ¿Cuáles son los métodos o técnicas de inteligencia artificial para identificación de conceptos?

Con estas preguntas de investigación se procedió a la búsqueda de los documentos o artículos utilizados en la revisión, para lo cual se definió la cadena de búsqueda:

“(methods OR technique OR framework) AND ("Text analysis" OR "text mining" OR "natural language" OR "artificial intelligence" OR "document processing" OR "automatic generation") AND ("Literature review" OR "articles review" OR "systematic mapping studies" OR SLR) AND (Software OR application OR package)”

Esta cadena se utilizó en la base de datos Scopus, obteniendo como resultado quinientos cincuenta y cuatro documentos, posteriormente se aplicó un filtro para obtener los documentos relevantes en la investigación, para el filtrado se tomó en cuenta que los documentos deben estar en inglés o español, procurando que sean actuales además de filtrar los documentos por área temática en donde se incluía ingeniería e informática.

En primera instancia se leyó el título los artículos para definir si cumplen con los requerimientos establecidos, de este proceso de filtrado se seleccionó dieciocho documentos, luego se dio lectura del resumen de los artículos, logrando determinar doce documentos que generaron un aporte relevante en la revisión. Concluida la búsqueda se realizó una lectura comprensiva de cada documento, rescatando la información de valor que fue organizada para la presente investigación. Finalmente se elaboró una matriz de conceptos relevantes de la investigación y los artículos estudiados.

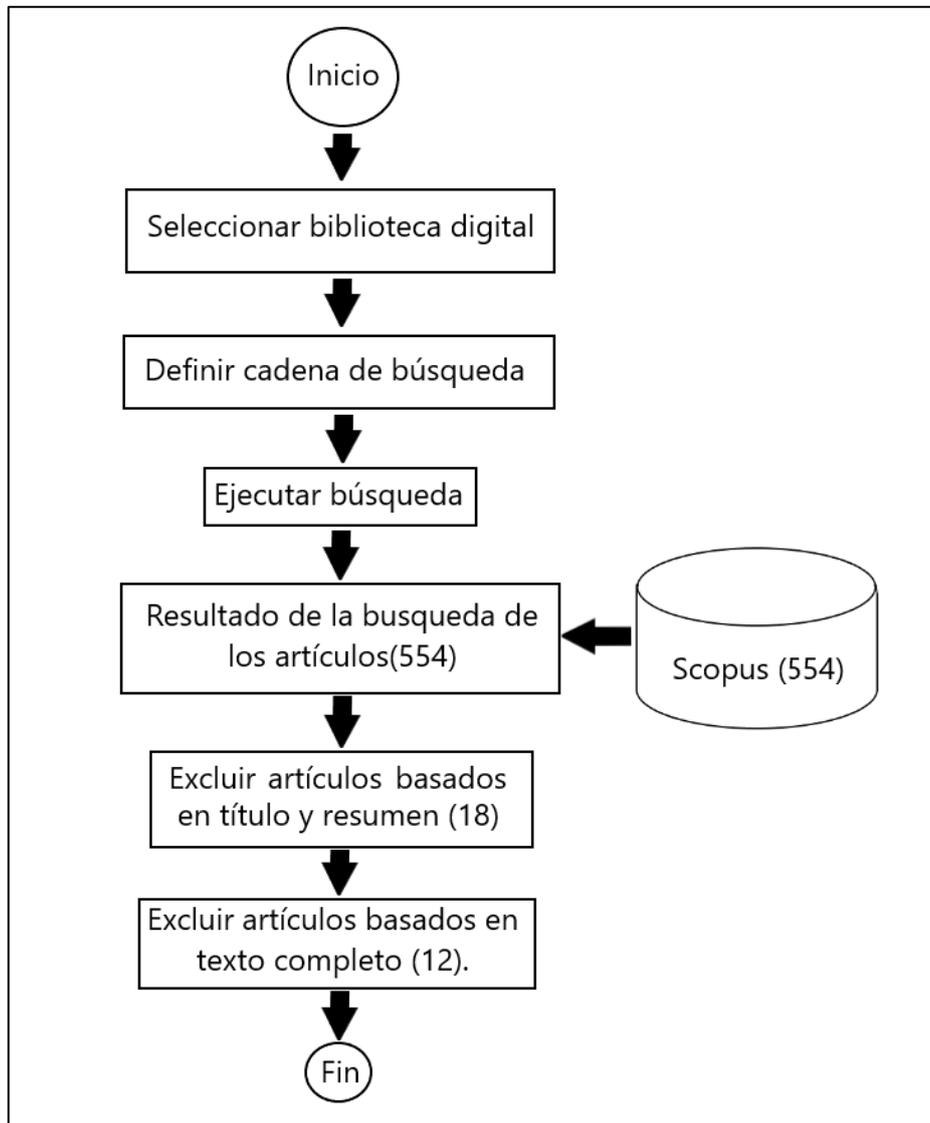


Figura 1.1. Proceso de Búsqueda y filtrado.

Tabla 1: Matriz de conceptos.

Artículos	Conceptos						
	Minería de Texto.	Técnicas de minería de texto.	Técnicas de análisis de texto.	Técnicas de clasificación de texto.	Procesamiento del lenguaje natural.	Software de minería de texto.	Software o métodos para revisión de literatura.
A Systematic Mapping Study on Text Analysis Techniques in Software Architecture (Bi et al., 2018).	X		X		X		
Natural language processing (NLP) in management research: A literature review (Kang et al., 2020).					X		
Text classification techniques: a literature review (Thangaraj & Sivakami, 2018).	X			X			
The Applications of Natural Language Processing (NLP) for Software Requirement Engineering - A Systematic Literature Review (Nazir et al., 2017).					X		
An Evaluation of Unstructured Text Mining Software (Crowsey et al., 2007).	X					X	
Buhos: A web-based systematic literature review management Software (Bustos Navarrete et al., 2018).							X
Performing literature review using text mining, Part III (Yang & Zhang, 2019).							X
Review of data, text and web mining software (Zhang & Segall, 2010).	X					X	
Literature Review of Automatic Text Summarization: Research Trend, Dataset and Method (Widyassari et al., 2019).					X		X
SLR-Tool a tool for performing systematic literature reviews (Fernández-Sáez et al., 2010).	X						X
Text Mining Software and Their Applications (Xianguang, 2014).	X					X	
Text Mining Techniques, Applications, and Challenges (Torre et al., 2018).	X	X					

A continuación, se presenta el resultado de la revisión de literatura:

1.2.2. Minería de texto

Con el avance de la tecnología y el uso masivo de internet se presencia una gran cantidad de información en formato de texto que se puede encontrar en libros, blogs,

mensajes de texto, correos electrónicos, investigaciones, periódicos y web cuyas características son datos descentralizados y difícil de analizar exhaustivamente. Las personas buscan la manera de extraer la información más relevante de estos grandes volúmenes de datos haciendo uso principalmente del software y las técnicas de minería / análisis de texto (Xiaguang, 2014).

La minería de texto describe una serie de técnicas para minar, procesar y analizar datos textuales no estructurados (Bi et al., 2018), por ende es el descubrimiento de nueva información que se extrae directamente de archivos de texto de una forma automática mediante el uso de la computadora (Crowsey et al., 2007). Es decir la minería de textos es el proceso de transformar documentos de texto en conjuntos de datos aplicando las distintas técnicas y algoritmos de minería de datos existentes (Torre et al., 2018), esto es importante porque los datos que se obtiene se utilizan en las empresas para la toma de decisiones estratégicas (Thangaraj & Sivakami, 2018), o se asigna a otras partes interesadas como al departamento de tecnologías de la información y la comunicación (TIC), en donde podrían usar estos datos para mejorar los sistemas informáticos de la empresa (Bi et al., 2018).

Dificultades para la minería de texto

Realizar el proceso de minería de texto generalmente no es una tarea fácil, mucha información contenida en el texto se pierde cuando se trata de procesar los archivos debido a varios problemas que requieren una adaptación y ciertos procesos específicos para un tratamiento correcto del texto (Torre et al., 2018).

Según estos autores, los siguientes problemas hacen que las fuentes de texto sean difíciles de tratar computacionalmente.

- **La naturaleza no estructurada del texto:** El texto todavía no tiene una estructura homogénea capaz de ser procesada automáticamente sin pérdida de información.
- **La naturaleza heterogénea y distribuida de los documentos:** Las fuentes externas de los repositorios de texto son diversas como intranets, bases de datos documentales, redes sociales, censos, sitios web, informes corporativos, publicaciones, correo electrónico, etc.
- **El multilingüismo:** No solo en diferentes conjuntos de documentos, sino también en la misma colección de textos
- **La dependencia del contexto en el análisis del texto:** Hay que tener en cuenta el dominio de la aplicación, lo que implica el uso de diccionarios, tesauros, u ontologías específicas para el contexto actual del texto.

- **La ambigüedad del lenguaje natural:** La identificación de frases ambiguas y las palabras en textos de lenguaje natural es un problema abierto en el área de minería de textos, ya que no son fáciles de detectar

Preparación de datos

Cuando se realiza minería de texto, hay que tener en cuenta que para obtener mejores resultados se deben preparar los datos adecuadamente, por ello se recomienda realizar una limpieza de los datos o preprocesamiento de estos para obtener una forma semiestructurada que se puede procesar de mejor manera. Aunque la minería de texto busca encontrar relaciones entre datos no estructurados, esto no elimina la necesidad de preparación de datos si el usuario desea lograr un alto nivel de fiabilidad y extraer conceptos útiles (Crowsey et al., 2007).

Técnicas de análisis de texto.

A continuación, se describen algunas técnicas de análisis de texto (Bi et al., 2018), se usa la terminología en inglés debido a que es más conocida.

a) Search and Information Retrieval (SIR) - Búsqueda y recuperación de información.

Es una de las técnicas de análisis de texto que más se ha usado en el procesamiento de información de arquitectura textual. Con SIR se realiza una recuperación y almacenamiento de información mediante la búsqueda en los motores de búsqueda predefinidos con el uso de palabras clave que coinciden con los requisitos de los usuarios.

b) Natural Language Processing (NLP) - Procesamiento natural del lenguaje.

El procesamiento natural del lenguaje (NLP) permite descubrir nuevo conocimiento extrayendo información valiosa automáticamente de los documentos de texto (Nazir et al., 2017). NLP emplea técnicas escalables basadas en estadísticas las cuales permiten realizar procesamiento, comprensión y búsqueda texto de manera eficiente (Bi et al., 2018), El NLP es asistido por computadora y su principal función es analizar y comprender matemáticamente el lenguaje humano, permitiéndoles de esta forma a los académicos extraer fácilmente información importante presente en conjuntos de datos textuales, y de este modo evitar el esfuerzo para realizar este proceso manual. Con el surgimiento de la Inteligencia Artificial (IA) se han implementado nuevas aplicaciones del NLP como la traducción automática, coincidencia de patrones, análisis de sentimientos y el reconocimiento de voz, además de mejorar las prácticas comerciales (Kang et al., 2020). Los estudios realizados acerca de NLP abarcan teorías y métodos que permiten una comunicación efectiva entre humanos y computadoras en lenguaje natural. El objetivo principal de NLP es traducir

el lenguaje humano (o natural) en comandos que pueden ser ejecutado por computadoras, consta de dos enfoques: Natural language understanding (NLU) - Comprensión del lenguaje natural y Natural language generation (NLG) - Generación de lenguaje natural (Kang et al., 2020).

c) Clustering (Ct) – Agrupación.

Es una técnica de aprendizaje no supervisado, que se utiliza para dividir los elementos de datos en grupos.

d) Web Mining (WM) - Minería Web

WM se centra en descubrir y analizar recursos de información en las páginas web, en la mayoría de los casos con el objetivo de mejorar la experiencia del usuario, o para realizar alguna investigación en particular.

e) Classification (Cf) – Clasificación.

Es una técnica de aprendizaje supervisado que se puede usar para clasificar términos, párrafos, o documentos usando métodos clasificación.

f) Information Extraction (IE) - Extracción de información.

IE ayuda en la extracción de la información de documentos de forma automática, conformando entidades o eventos e identificando las relaciones entre cada entidad.

g) Concept Extraction (CE) - Extracción de conceptos.

CE es un proceso para agrupar palabras y frases semánticamente similares.

Técnicas de minería de texto.

A continuación, se describen algunas técnicas de minería de texto (Torre et al., 2018):

1) Técnicas de agrupación

Localiza grupos similares mediante una función de similitud, estos grupos pueden ser de documentos, sobre temas, conceptos clave, oraciones, términos o secuencias. Dependiendo el tipo de agrupamiento seleccionado, la manera de localizar los elementos comunes puede variar.

- **Agrupamiento difuso** En el área de minería de texto se han utilizado algoritmos de agrupación difusos, debido principalmente a la gestión adecuada de datos incompletos / ruidosos, esto mejora la comprensibilidad de los patrones y el manejo de clústeres superpuestos.
- **Reglas de asociación difusas** Permiten la extracción de conocimiento mediante reglas de asociación.

2) Técnicas de clasificación

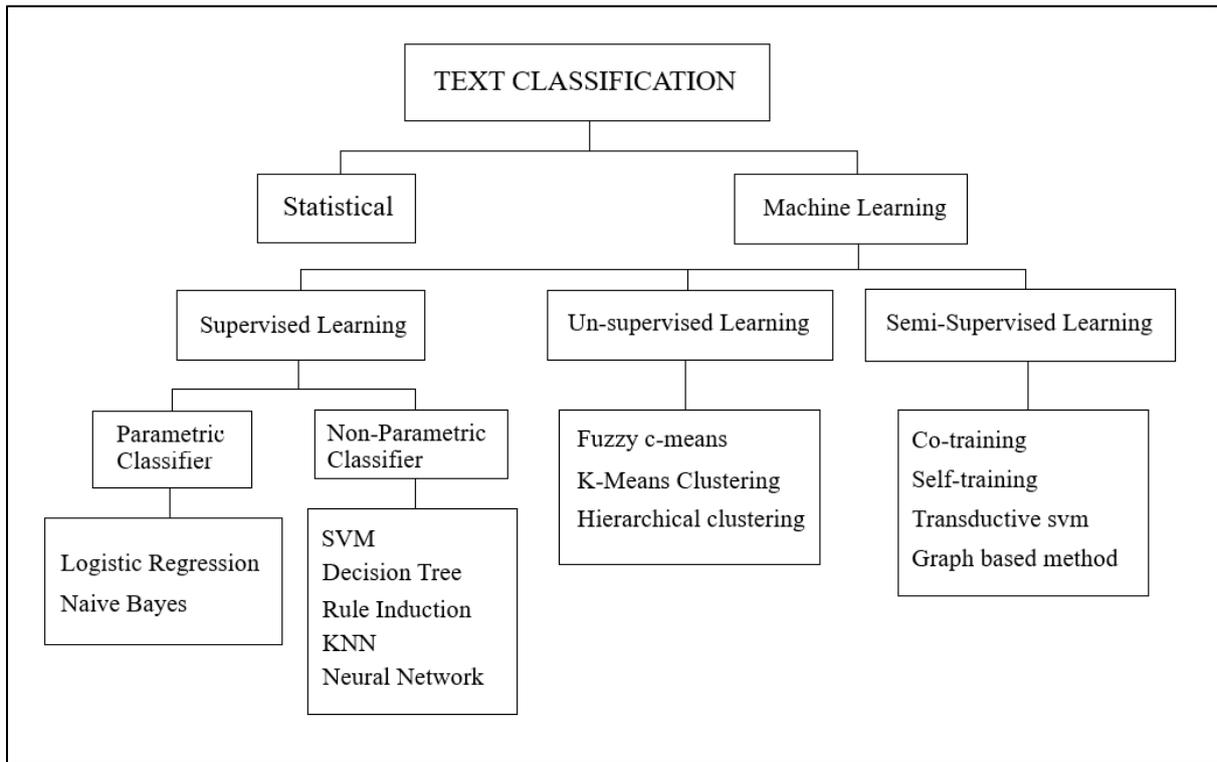


Figura 1.2. Técnicas de clasificación del texto.
Adaptado de (Thangaraj & Sivakami, 2018).

Se dividen principalmente en aplicaciones estadísticas y de aprendizaje automático o machine learning (ML). A continuación, se presentan algunas de las técnicas de clasificación de basado en el trabajo de Thangaraj & Sivakami (2018), ver figura 1.2.

Enfoque estadístico (Statistical)

Las técnicas estadísticas son procesos puramente matemáticos, que se ejecutan de manera similar a un programa de computadora, siguiendo una serie de instrucciones dadas sin ninguna habilidad propia, además son ineficientes para grandes conjuntos de datos, estas técnicas se pueden aplicar para sistemas de detección de fraude.

Enfoque de aprendizaje de máquinas (Machine learning)

Debido al aumento en el volumen de datos, la velocidad y la variedad, surgió la necesidad de automatizar las técnicas de procesamiento de texto. Estas técnicas de clasificación de texto con un enfoque de aprendizaje automático se dividen en tres categorías: clasificación de texto supervisada, clasificación de texto sin supervisión y clasificación de texto semi supervisada.

1. Clasificación de texto supervisada

Es el más costoso ya que requiere intervención humana y esto en grandes cantidades de datos resulta muy difícil. Se divide en dos categorías:

- a) **Modelos paramétricos:** En esta categoría se encuentra la regresión logística y el clasificador Naive Bayes.
- b) **Modelos no paramétricos:** En esta categoría está el algoritmo Support Vector Machine (SVM), los árboles de decisión, la regla de inducción, el algoritmo K-Nearest Neighbour (k-NN) y las redes neuronales artificiales.
- **Redes neuronales** Las redes neuronales están dentro de los modelos no paramétricos y son útiles para datos de series de tiempo ya que funcionan bien con datos incompletos o faltantes.
- **Redes neuronales difusas** Esta técnica ha sido revelada como una función de autoaprendizaje y enfoque no lineal para la clasificación de patrones y minería.

2. Clasificación de texto sin supervisión

No se proporcionan datos para la capacitación del sistema, al principio resulta complejo, pero a medida que se ingresan los datos el algoritmo se refina, este modelo es más adecuado para datos de gran volumen. Hay que tener en cuenta que la eficiencia de las iteraciones es primordial para la capacidad computacional efectiva de estos algoritmos.

3. Clasificación de texto semi-supervisada.

Es una combinación entre las técnicas de aprendizaje supervisadas y no supervisadas, para el entrenamiento se usan pocos datos etiquetados y muchos datos no etiquetados, a pesar de no ser iterativo aumenta la eficiencia del clasificador, entre sus técnicas se encuentra auto entrenamiento o autoaprendizaje, co-entrenamiento, modelos generativos y métodos basados en gráficos.

1.2.3. Software de minería de texto.

De la revisión de literatura se encontraron los siguientes softwares:

Leximancer [<https://info.leximancer.com/>]

Es un software que extrae los conceptos, además muestra la relación que hay entre esos conceptos junto con la intensidad de relación. La interfaz de usuario es basada en Java

y ofrece algunas características que permiten a un usuario manipular las etapas en el proceso de minería de texto (Crowsey et al., 2007).

SAS Enterprise Miner [https://www.sas.com/en_us/software/enterprise-miner.html]

Este software usa un enfoque de minería llamado SEMMA, que representa una lista de pasos secuenciales comenzando por muestreo, exploración, modificación, modelado y evaluación. SAS ofrece características para comparar los resultados de los diferentes tipos de modelado a través del análisis estadístico y en términos comerciales permitiendo la integración entre el grupo de modelado estadístico, los gerentes de negocios y el departamento de tecnologías de la información.

SAS permite cargar texto en una variedad de formatos de archivo, primero los almacena externamente para convertirlos a un conjunto de datos SAS a través de un macro predefinido, esto hace que la aplicación sea compleja y consuma más tiempo en este paso del proceso de minería de texto. Su interfaz es poco intuitiva, aunque ofrece muchas de las mismas características que otros productos (Crowsey et al., 2007).

- **SAS Text Miner** Es un complemento de SAS Enterprise Miner para realizar análisis estadísticos simples, análisis exploratorios de datos textuales, agrupamiento y modelado predictivo de datos textuales (Zhang & Segall, 2010).

Clementine [https://spss-clementine.software.informer.com/]

Admite muchos formatos de texto, como: texto plano, PDF, HTML, Microsoft Office y Archivos de texto XML, su arquitectura abierta le permite unirse con otras aplicaciones de análisis de texto. En la extracción de texto se puede utilizar algunos métodos con el fin de limitar el alcance de una extracción, por ese motivo produce más términos y frases informativas en sus resultados.

Incluye dos funciones principales, una es el reconocimiento y extracción de sentimientos del texto, esto gracias a los archivos de diccionario creados en Text Mining Builder. La segunda función, detecta correlaciones entre cosas, tales como personas / eventos, o enfermedades / genes (Crowsey et al., 2007).

Polyanalyst [https://www.megaputer.com/polyanalyst/]

Desarrollado por Megaputer, tiene una funcionalidad de procesar bases de datos a gran escala, tiene una baja curva de aprendizaje y se incluyen tutoriales paso a paso, además se puede integrar el análisis en texto estructurado como no estructurado (Crowsey et al., 2007).

PolyAnalyst permite crear, importar y administrar taxonomías, y realizar una categorización automatizada de registros de texto entre taxonomías existentes, esta característica de categorización de taxonomías es útil en grandes colecciones de documentos no estructurados (Zhang & Segall, 2010).

Clarabridge [https://www.clarabridge.com/]

Permite realizar análisis de datos y procesamiento paralelo de grandes cantidades de datos, mientras realiza el procesamiento se reconoce las entidades, relaciones, secciones, encabezados, temas, tablas y otros datos. Esta información se almacena en metadatos vinculándolos directamente a su fuente, adicional a esto, puede contener grandes cantidades de datos y mantener alto rendimiento y requiere una cantidad mínima de codificación de usuarios y procesos que se lo puede realizar sin intervención de alguna persona (Crowsey et al., 2007).

Compare Suite [https://comparesuite.com/]

Es un software de minería de texto desarrollado por AKS-Labs, permite realizar una comparación entre documento o archivos en distintos formatos, para ello realiza un procesamiento de carácter por carácter, palabra por palabra o por palabras clave. También se pueden comparar dos carpetas para ver los archivos que contiene cada una y revisar los cambios que se han realizado permitiendo realizar un informe después de la comparación (Zhang & Segall, 2010).

TextAnalyst [https://www.analyst.ru/index.php?lang=eng&dir=content/downloads/]

Tiene una sección para el tratamiento del texto y análisis semántico, utiliza una red semántica que le permite determinar la importancia de un concepto de texto realizando un análisis de su conexión con otros conceptos en el mismo texto, implementando también algoritmos semejantes a los utilizados para el análisis de texto en el cerebro humano (Zhang & Segall, 2010).

Su principal ventaja es que puede extraer automáticamente la red semántica de texto sin el diccionario de temas especializado y con ello puede ejecutar el análisis de texto basado en la tecnología de red neuronal (Xiaguang, 2014).

Según (Xiaguang, 2014), las funciones principales de TextAnalyst son filtrar el significado del texto, navegación del cuerpo, temas de estructuración, agrupación y resumen del texto.

VisualText [https://visualtext.org/]

Permite a cada usuario crear su propio analizador de texto, si se requiere el procesamiento de voz se debe convertir primero en texto para realizar el procesamiento de texto. Se puede usar el software para combatir el terrorismo, espionaje narcótico, proliferación nuclear, filtrado de documentos, calificación de prueba y automática codificación (Zhang & Segall, 2010).

WordStat [https://provalisresearch.com/products/content-analysis-software/]

Es un módulo de software de análisis de texto que se puede usar para estudiar información textual tales como entrevistas, respuestas a preguntas abiertas, artículos de revistas, comunicaciones electrónicas, etc. También se usa para categorizar automáticamente el texto usando un enfoque de diccionario, incorpora muchos análisis de datos y herramientas gráficas usadas para ver las relaciones entre el contenido del documento y explorar la información en variables categóricas o numéricas (Zhang & Segall, 2010).

Para realizar un mayor análisis utiliza listas cruzadas, palabras clave y complejos métodos estadísticos. WordStat realiza varias funciones: soporta múltiples idiomas y realiza procesamiento de texto, análisis de frecuencia de palabras unitarias, extracción de características, búsqueda de palabras clave y análisis simbiótico y clasificación automática de texto (Xiaguang, 2014).

1.2.4. Uso de minería de texto para realizar revisión de literatura.

La realización de una revisión de literatura es difícil de ejecutar, debido a que en primera instancia requieren la búsqueda de cientos de miles de artículos, además de requerir la identificación de los documentos que tiene información relevante dependiendo la pregunta de investigación. Todos los pasos de la revisión se deben realizarse de forma precisa, transparente y registrada. Por lo tanto, desarrollar una herramienta de software requieren unificar varios tipos de investigaciones, teórico, así como empírico, para llegar a una conclusión, esta herramienta debe ser flexible y permitir realizar revisiones de múltiples formas, garantizando transparencia y rigor al registrar cada paso (Bustos Navarrete et al., 2018).

Pese a la dificultad de desarrollo, el uso de herramientas de software puede facilitar de modo significativo la realización del proceso completo de revisión de literatura. El software más especializado en el proceso de revisión de literatura está diseñado para cumplir con los

requisitos de la salud, las ciencias médicas y la ingeniería de software (Bustos Navarrete et al., 2018), a continuación, se presenta algunos de estos aplicativos y algoritmos.

TextRank

Es un algoritmo que ayuda al proceso de revisión de la literatura, siendo su objetivo principal el resumir un texto dado, basado en la importancia de las oraciones (Yang & Zhang, 2019).

Yang y Zhang (2019) afirman que, para realizar el resumen basada en minería de texto, se categorizan un trabajo académico o documento en tres niveles:

- **Nivel superior:** El título y las palabras clave indexadas proporciona la información de nivel superior, que el investigador evalúa para decidir si continúa leyendo.
- **Segundo nivel:** Se puede considerar de segundo nivel el resumen de la publicación, hay que tomar en cuenta que el resumen puede ser confuso o poco claro debido a los diferentes estilos de escritura de los autores.
- **Tercer nivel:** Si el investigador encuentra relevante la información de segundo nivel, él puede acceder a leer el texto completo, que es la información de tercer nivel.

La forma más rápida de crear el resumen es usar las N primeras frases más relevantes que se pueden presentar en el orden de su aparición en el texto o en el orden de su importancia (Yang & Zhang, 2019).

Buhos [https://www.buhos.org/]

Es una aplicación web que gestiona todo el proceso de revisiones de literatura, se puede usar de manera local a través de un servidor interno, como también de manera distribuida (Bustos Navarrete et al., 2018).

Bustos Navarrete, (2018) menciona las siguientes características de Buhos:

- **Multiplataforma:** La aplicación puede ejecutarse en cualquier sistema operativo en el que se pueda compilar Ruby en su versión 2.4, utilizando el código fuente.
- **Trabajo colaborativo:** Cada revisión de literatura se realiza con la guía de un administrador, además de puede enviar mensajes de forma interna maximizando la transparencia en el proceso de decisión.
- **Internacionalización:** A pesar de que las versiones completas disponibles están en inglés y español, se puede realizar una adaptación del software a varios idiomas.

- **Etapas flexibles de análisis:** La revisión se realiza en etapas como la búsqueda de documentos, selección del título y resumen, selección de referencias, revisión de texto completo y por último los informes.
- **Integración con varios sistemas de metadatos científicos:** En el sistema se puede guardar las búsquedas realizadas en varias bases de datos bibliográficas, además de las referencias disponibles en software de referencia bibliográfica.
- **Repositorio de archivos:** Para cada revisión se puede alojar diferentes archivos, que se pueden ver a través de una vista en línea de archivos PDF y ODF, para evaluar la calidad de cada documento y para la extracción de datos, realizando todo el proceso en línea.
- **Múltiples formas de análisis:** Se puede personalizar los criterios de inclusión, exclusión y los criterios de evaluación de calidad de los estudios.
- **Múltiples salidas:** Se puede exportar tanto los registros de cada etapa en archivos BibTeX o Excel, como también la lista de documentos con sus respectivas relaciones en GraphML (formato de archivo para gráficos basado en XML). Adicional a ello se puede acceder a un diagrama de flujo PRISMA (presentación de revisiones sistemáticas y metaanálisis), que desglosa el flujo de información durante las diversas etapas del proceso.

SLR-Tool [<http://ta.mdx.ac.uk/slr/about/>]

Es una herramienta de software con interfaces disponibles en inglés y español, a diferencia de otras herramientas, es gratis y ayuda en el proceso de revisión de literatura, reduciendo significativamente el esfuerzo requerido para dicha revisión (Fernández-Sáez et al., 2010).

En cuanto a la funcionalidad de SLR-Tool, presenta las siguientes ventajas (Fernández-Sáez et al., 2010):

- En cada fase del proceso de revisión de literatura, puede almacenar los datos de las actividades realizadas.
- Se puede refinar las búsquedas usando minería de texto.
- Permite realizar un esquema de clasificación lo que ayuda al investigador a realizar síntesis y análisis de datos.
- Agrupa los distintos documentos utilizando técnicas de minería de texto.

- Se puede exportar los datos en archivos de Excel en formatos de tabla o gráfico, además de permitir la exportación de las referencias de los documentos en formatos aceptados por los paquetes bibliográfico como EndNote, BibTeX y Ris.

EPPI-Reviewer [https://eppi.ioe.ac.uk/EPPIReviewer-Web]

Es una herramienta web que permite a los investigadores desarrollar una revisión de literatura colaborativa, brindando la posibilidad de realizar gestión bibliográfica, además de gestionar los criterios de inclusión y exclusión. Su enfoque principal es la presentación de resultados en donde se puede generar informes y aplicar algunas técnicas de metaanálisis (Fernández-Sáez et al., 2010).

Resumen automático

El resumen automático es una parte indispensable de conocimiento sobre procesamiento del lenguaje natural. Se puede definir como el proceso de presentar de forma automática uno o más documentos de texto con menos de la mitad del texto original, manteniendo el contenido de la información principal, es decir sin perder el significado de los documentos iniciales, mediante el uso de una máquina. Aunque las investigaciones registradas acerca de resumen automático comenzaron desde hace mucho tiempo, hasta la actualidad no hay un sistema capaz de generar resúmenes con un margen de calidad como los realizados por profesionales o humanos (Widyassari et al., 2019).

1.3. Modelo de éxito de los sistemas de información de DeLone y Mclean

La misión de los sistemas de información (SI) es el diseño, entrega, uso e impacto efectivo de las distintas tecnologías en las organizaciones y la sociedad, esto debido a que las organizaciones que optan por invertir en IS para solucionar una necesidad o mejorar una oportunidad de negocio, requieren asegurarse del éxito de sus inversiones. Por ende, la determinación del valor de del éxito o rendimiento de los sistemas de información es un aspecto muy importante, aunque las formas de evaluación van cambiando dependiendo el contexto en donde se aplique los SI, así como el propósito por el cual se los usa y el impacto que ha tenido la tecnología en los últimos años (DeLone & McLean, 2016).

Con el constante avance de la tecnología, se puede evidenciar un gran crecimiento de la información, la cual está disponible y accesible para las personas, esto ha cambiado la dinámica de poder de las relaciones entre las distintas organizaciones ya sean grandes o pequeñas, entre ciudadanos y sus gobiernos, por lo que la medición del éxito de SI se ha vuelto más difícil. Se ha vuelto difícil porque cada vez se usa con mayor frecuencia los sistemas de información y aumenta los registros de usuarios, aumentando las posibilidades

del propósito de un SI y la definición de sus partes interesadas. En esencia sigue siendo simple porque hay elementos clave que no cambian, como la calidad de la información, la calidad del sistema, el uso y los resultados, que permite realizar la medición del éxito (DeLone & McLean, 2016).

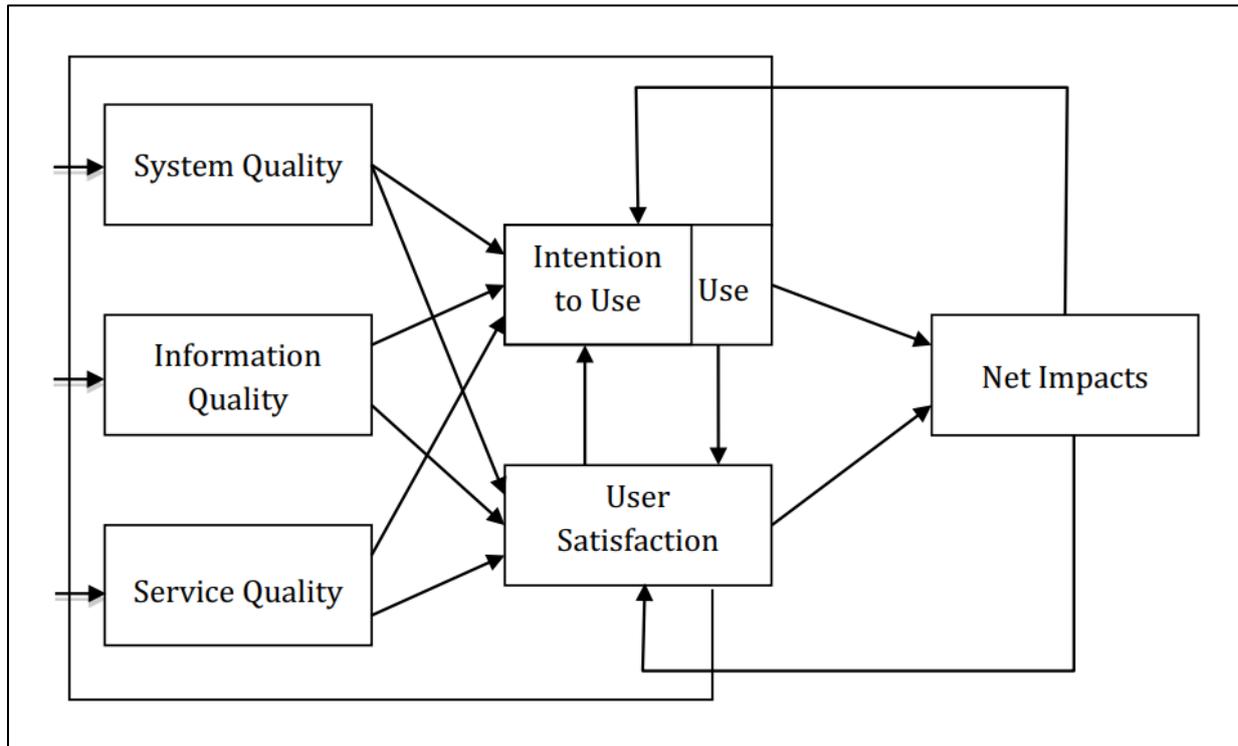


Figura 1.3. Modelo de éxito de los sistemas de información.
Adaptado de (DeLone & McLean, 2016).

Según DeLone y McLean (2016) las variables de éxito son las siguientes figura 1.3:

- **Calidad del sistema:** Son las características requeridas de un sistema de información como, la usabilidad, flexibilidad del sistema, confiabilidad del sistema, entre otras, además de las características del sistema como la percepción, sofisticación, flexibilidad y tiempos de respuesta.
- **Calidad de la información:** Los resultados requeridos del sistema, es decir, con que precisión, calidad, concisión, entre otros aspectos, se genera los informes por ejemplo de gestión o de páginas web.
- **Calidad del servicio:** En esta parte se refiere a que tan efectivo es el soporte a los usuarios que brinda la organización del sistema de información.
- **Uso:** A qué nivel usan los empleados y clientes las distintas funcionalidades de un sistema de información y la forma como lo usan.

- **Satisfacción del usuario:** Que tan cómodos o satisfechos se sienten los usuarios con los informes, las distintas páginas web y los servicios de asistencia.
- **Impactos netos:** El grado de contribución de un sistema de información para el crecimiento de las personas, grupos u organizaciones.

Como se evidencia claramente en la figura 4, la calidad de la información, del sistema y del servicio influirá de forma positiva o negativa en el uso, la intensidad de uso y la satisfacción del usuario, es decir, si el sistema presenta una buena calidad en los aspectos mencionados el usuario tendrá la intención de usar el sistema, posterior a ello, hará uso del sistema experimentando que tan satisfactorio es su uso y esto lo conllevará a obtener los beneficios netos, en el caso de que los factores principales del sistema no tengan una buena calidad, la persona que haga uso del sistema no estará satisfecha y por ende no se obtendrán los beneficios netos del sistema como tal.

CAPÍTULO 2

Desarrollo

El desarrollo de la aplicación informática para la generación automática de la matriz de conceptos se lo realizará usando la metodología Scrum como marco de desarrollo, como lenguaje de programación se utilizó Python con algunas librerías que facilitan el procesamiento de documentos y la extracción del texto, además de implementar técnicas y métodos de inteligencia artificial para la extracción de los conceptos.

2.1. Fase de Iniciación Scrum.

2.1.1. Definición de Roles Scrum

Para iniciar con el desarrollo del sistema se realiza la asignación de roles y responsabilidades a todos los involucrados en este proceso. A continuación, en la Tabla 2.1 se detalla la asignación del equipo de trabajo.

Tabla 2.1: Roles Scrum.

ROL	NOMBRE	FUNCIÓN
Product Owner	MSc. Irving Reascos	Evaluar que las funcionalidades estén acorde a los requerimientos.
Scrum Máster	MSc. Alexander Guevara	Promueve y apoya el uso adecuado de la metodología, además de verificar el avance del desarrollo del sistema.
Equipo Scrum	Sr. Patricio Pinchao	Desarrollar el sistema, con base en los requerimientos.

2.1.2. Definición del Product Backlog.

En esta parte se define la lista de las funcionalidades requeridas para el producto final, especificando la prioridad de cada historia de usuario y la estimación con la ayuda del método de *T-Shirt Size – Effort Estimation* (Zanker et al., 2010), representado en la tabla 2.2.

Tabla 2.2: Estimación de esfuerzo (T-Shirt Size)

Tamaño	Rango de trabajo estimado
S	20 – 30 horas
M	30 – 50 horas
L	50 – 70 horas
XL	70 – 100 horas

A continuación, se muestra el *product backlog* con los requerimientos establecidos y una breve descripción de cada uno (Tabla 2.3).

Tabla 2.3: Definición del Product Backlog

ID	Historia de usuarios	Prioridad	Estimación	Descripción
1	Login/inicio de sesión y Registro.	Alta	L	El sistema debe permitir el registro de usuarios e inicio de sesión con control de sesiones.
2	Gestión de Análisis.	Media	S	El usuario puede crear un análisis para generar una matriz de conceptos.
3	Gestión de documentos.	Media	M	El sistema debe permitir subir documentos (pdf) para procesarlos.
4	Gestión de conceptos.	Media	M	El sistema debe permitir agregar conceptos para buscar en los documentos.
5	Extracción del texto de los documentos.	Alta	M	Una vez cargado los documentos el sistema debe extraer el texto para buscar la información relevante.
6	Limpieza y preprocesamiento de datos.	Alta	S	Para poder comparar el texto con los conceptos se necesita eliminar el ruido.
7	Implementación del modelo de inteligencia artificial.	Alta	XL	El sistema debe realizar una búsqueda de los conceptos presentes en el texto mediante el uso de inteligencia artificial.
8	Reporte de matriz de conceptos.	Alta	L	El sistema debe generar automáticamente una matriz de conceptos donde se especifique en que artículos están los conceptos introducidos.
9	Refinamiento del reporte de matriz de conceptos v.2	Alta	M	El sistema debe permitir cargar más archivos y concepto una vez finalizado un análisis.

2.2. Fase 2. – Panificación y estimación

2.2.1. Definición de historias de usuario.

En las historias de usuario se especifican los requerimientos que el usuario necesita como parte del sistema.

Tabla 2.4: Historia de Usuario 1 - *Login*

HISTORIA DE USUARIO	
Número: 1	Usuario: Usuario
Nombre historia: Login/inicio de sesión y Registro.	
Prioridad: Alta	Estimación: L
Descripción: <ul style="list-style-type: none">• Como usuario quiero que el sistema me permita registrarme y los datos se almacenen en la base de datos para luego poder iniciar sesión.	
Pruebas de aceptación: <ul style="list-style-type: none">• Mostrar un mensaje de error si el correo ya está registrado.• Mostrar un mensaje de error si el usuario ingreso mal las credenciales.• Los parámetros de usuario y contraseña son requeridos.• Control de sesiones.	

Tabla 2.5: Historia de Usuario 2 – Gestión de Análisis.

HISTORIA DE USUARIO	
Número: 2	Usuario: Usuario
Nombre historia: Gestión de Análisis.	
Prioridad: Media	Estimación: S
Descripción: <ul style="list-style-type: none">• Como usuario del sistema quiero crear varios análisis, poder ver cuál es su estado (en proceso o finalizados), editar y eliminar.	
Pruebas de aceptación: <ul style="list-style-type: none">• Mostrar una tabla con los análisis creados y su estado.	

Tabla 2.6: Historia de Usuario 3 – Gestión de documentos.

HISTORIA DE USUARIO	
Número: 3	Usuario: Usuario
Nombre historia: Gestión de documentos.	
Prioridad: Media	Estimación: M
Descripción:	
<ul style="list-style-type: none"> • Como usuario del sistema quiero cargar por cada análisis los artículos usados en la revisión de literatura para que el sistema procese la información de cada documento. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • Los documentos que se puedan procesar en el sistema deben ser en formato PDF. • El sistema mostrará la lista de documentos cargados. • Se puede borrar algún documento que no se requiera para el procesamiento. • No se puede guardar documentos con el mismo nombre. 	

Tabla 2.7: Historia de Usuario 4 – Gestión de conceptos.

HISTORIA DE USUARIO	
Número: 4	Usuario: Usuario
Nombre historia: Gestión de conceptos.	
Prioridad: Media	Estimación: M
Descripción:	
<ul style="list-style-type: none"> • Como usuario del sistema quiero agregar múltiples conceptos por cada análisis, así como editar y eliminar en el caso de no requerir alguno para que el sistema realice una búsqueda de los conceptos dentro de los artículos. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • Se puede agregar más de un concepto. • Se puede borrar los conceptos agregados por error. • No se debe guardar conceptos repetidos. 	

Tabla 2.8: Historia de Usuario 5 - Extracción del texto de los documentos.

HISTORIA DE USUARIO	
Número: 5	Usuario: Usuario
Nombre historia: Extracción del texto de los documentos.	
Prioridad: Alta	Estimación: M
Descripción:	
<ul style="list-style-type: none"> • Como usuario quiero que el sistema extraiga todo el contenido textual de forma automática de cada documento para posteriormente procesar tal información. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • El sistema solo procesará el texto de los documentos. • Se trabajará cada documento de forma individual sin mezclar la información. 	

Tabla 2.9: Historia de Usuario 6 - Limpieza y preprocesamiento de datos.

HISTORIA DE USUARIO	
Número: 6	Usuario: Usuario
Nombre historia: Limpieza y preprocesamiento de datos.	
Prioridad: Alta	Estimación: S
Descripción:	
<ul style="list-style-type: none"> • Como usuario quiero que el sistema realice una limpieza de datos en el texto de los artículos eliminando el ruido, como referencias, saltos de línea, etc. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • El texto resultante no debe perder su estructura. • No se deben eliminar los signos de puntuación. 	

Tabla 2.10: Historia de Usuario 7 - Implementación del modelo de inteligencia artificial.

HISTORIA DE USUARIO	
Número: 7	Usuario: Usuario
Nombre historia: Implementación del modelo de inteligencia artificial.	
Prioridad: Alta	Estimación: XL
Descripción:	
<ul style="list-style-type: none"> • Como usuario quiero que el sistema realice una búsqueda de forma automática de los conceptos que estén incluidos en cada documento aplicando inteligencia artificial para obtener mejor información en la búsqueda. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • El sistema debe encontrar las relaciones que hay entre una oración o párrafo con respecto a los conceptos introducidos de cada documento. 	

Tabla 2.11: Historia de Usuario 8 - Reporte de matriz de conceptos.

HISTORIA DE USUARIO	
Número: 8	Usuario: Usuario
Nombre historia: Reporte de matriz de conceptos.	
Prioridad: Alta	Estimación: L
Descripción:	
<ul style="list-style-type: none"> • Como usuario quiero poder visualizar en el sistema una matriz de conceptos resultante del procesamiento de los documentos y conceptos introducidos por cada análisis, especificando en que artículo se encontró cada concepto. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • La información debe estar presentada de una forma ordenada y compresible para el usuario. • Se debe presentar el porcentaje de coincidencia de cada concepto con cada documento. • En el caso de haber coincidencia mostrar las tres primeras oraciones con la coincidencia más alta. 	

Tabla 2.12: Historia de Usuario 9 – Refinamiento del Reporte de matriz de conceptos v.2.

HISTORIA DE USUARIO	
Número: 9	Usuario: Usuario
Nombre historia: Refinamiento del reporte de matriz de conceptos v.2.	
Prioridad: Alta	Estimación: M
Descripción:	
<ul style="list-style-type: none"> • Como usuario quiero poder visualizar en el sistema una matriz de conceptos resultante del procesamiento de los documentos y conceptos introducidos por cada análisis, especificando en que artículo se encontró cada concepto, para tener un esquema general de la investigación que se esté realizando. 	
Pruebas de aceptación:	
<ul style="list-style-type: none"> • La información debe estar presentada de una forma ordenada y compresible para el usuario. • Se debe presentar el porcentaje de coincidencia de cada concepto con cada documento. • En el caso de haber coincidencia mostrar las tres primeras oraciones con la coincidencia más alta. 	

2.2.2. Planificación del proyecto

A continuación, en la tabla 2.13 se asigna cada historia de usuarios a los Sprint, especificando la estimación en horas basandonos en la tabla presentada anteriormente y definiendo las fechas de iniciación y finalización del sprint.

Tabla 2.13: Aprobación del Sprint

ID	Historia de usuarios	Estimación dificultad	Sprint número	Fecha
1	Login/inicio de sesión y Registro.	L	Sprint 1	(29/06/2020)
2	Gestión de Análisis.	S		al (10/07/2020)
3	Gestión de documentos.	M	Sprint 2	(13/07/2020)
4	Gestión de conceptos.	M		al (24/07/2020)
5	Extracción del texto de los documentos.	M	Sprint 3	(27/07/2020)
6	Limpieza y preprocesamiento de datos.	S		al (07/08/2020)
7	Implementación del modelo de inteligencia artificial.	XL	Sprint 4	(10/08/2020) al (21/08/2020)
8	Reporte de matriz de conceptos.	L	Sprint 5	(24/08/2020) al (11/09/2020)
9	Refinamiento del reporte de matriz de conceptos v.2	M	Sprint 6	(14/09/2020) al (25/09/2020)

2.3. Desarrollo del proyecto

Definido anteriormente los roles, responsabilidades y requisitos necesarios; se continuó con el desarrollo del proyecto, ejecutando en iteraciones cada uno de los sprints en las fechas especificadas, hasta conseguir el desarrollo completo del sistema.

Por cada iteración o sprint se evidencia el diseño, desarrollo e implementación hasta finalizar con todos los sprints.

Para el desarrollo del proyecto se hizo uso de distintas herramientas de software tanto para el diseño como para la codificación, en este caso para el diseño y prototipado de las vistas del sistema se utilizó Balsamiq en la versión 4.0.25, como lenguaje de programación se usó Python 3.9.2, el IDE de desarrollo fue Visual Studio Code en su versión 1.55.2, además del Flask 1.1.2 como framework de desarrollo y por último como motor de base de datos MySQL 15.1, el proceso de instalación se puede evidenciar en los anexos.

2.3.1. Planificación del sprint 1

En el sprint 1 se contempla el desarrollo del registro de usuarios en el sistema e inicio de sesión con las debidas credenciales, controlando el acceso no autorizado, la creación, edición, consulta de los análisis por cada usuario.

Tabla 2.14: Sprint 1 – Desarrollo del login, registro y gestión de análisis.

Sprint 1			
Fecha inicio: 29/06/2020			
Fecha fin: 10/07/2020			
ID	Historia de Usuario	Tarea	Horas
1	Login/inicio de sesión y Registro.	Investigación del Framework Flask.	8
		Creación del repositorio del Proyecto en GitHub.	2
		Preparación del entorno de desarrollo.	5
		Creación del proyecto con Flask.	2
		Diseño de la vista de inicio de sesión y registro.	4
		Creación de la tabla users en la base de datos.	3
		Rutas de navegación.	1
		Desarrollo de la vista de registro.	5
		Creación del método de registro de usuarios.	3
		Desarrollo de la vista de inicio de sesión.	5
		Creación del método de inicio de sesión.	3
		Encriptación de la contraseña.	4
		Validación de los campos en inicio de sesión y registro.	3
		Pruebas de funcionalidad.	4
		Total	52
2	Gestión de Análisis.	Diseño de la vista del listado de análisis.	3
		Creación de la tabla analysis en la base de datos.	4
		Rutas de navegación.	1
		Desarrollo del método de consulta y listado de análisis.	6
		Desarrollo de la vista del listado de análisis.	4
		Desarrollo del método de creación de análisis.	4
		Desarrollo del método de eliminación de análisis.	3
		Pruebas de funcionalidad.	4
		Total	29
	Reuniones	Planificación del Sprint uno.	5
		Sprint Review.	2
		Sprint Retrospective.	2
		Total	9
Total			90

2.3.2. Ejecución del Sprint 1

A continuación, se especifica algunas tareas del Sprint 1

Diseño de la vista de inicio de sesión y registro.

En las figuras 2.1 y 2.2 se muestra el diseño de las vistas del login y registro respectivamente, en donde se realizó un diseño simple y fácil de usar mediante el uso de la herramienta de balsamic.

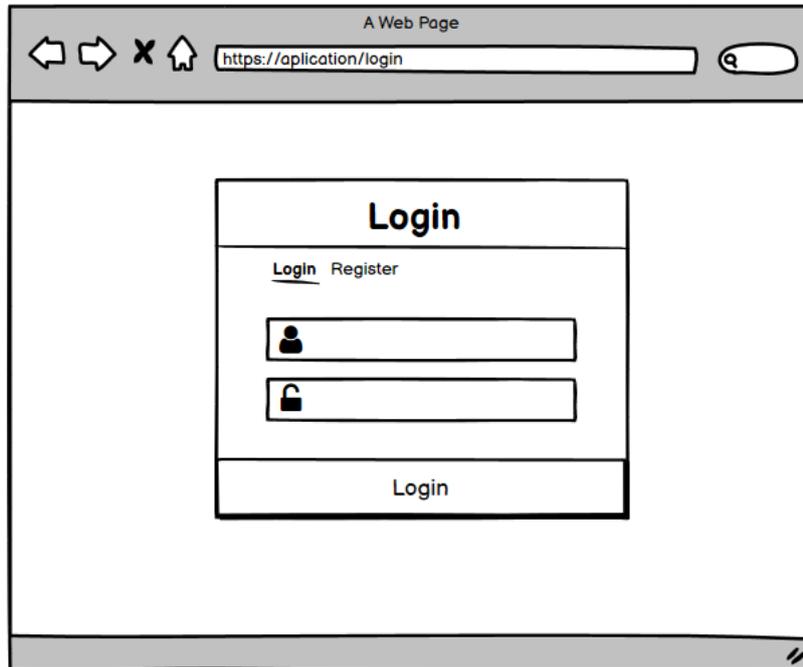


Figura 2.1. Diseño Login

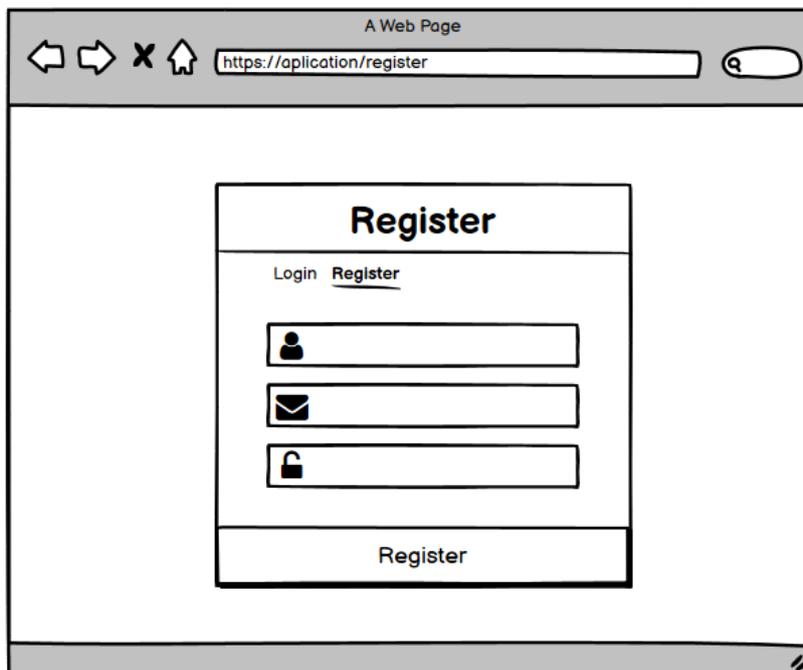


Figura 2.2. Diseño Registro.

Creación de la tabla *users* en la base de datos.

En la figura 2.3 se muestra la tabla *users* ya creada en la base de datos MySQL.

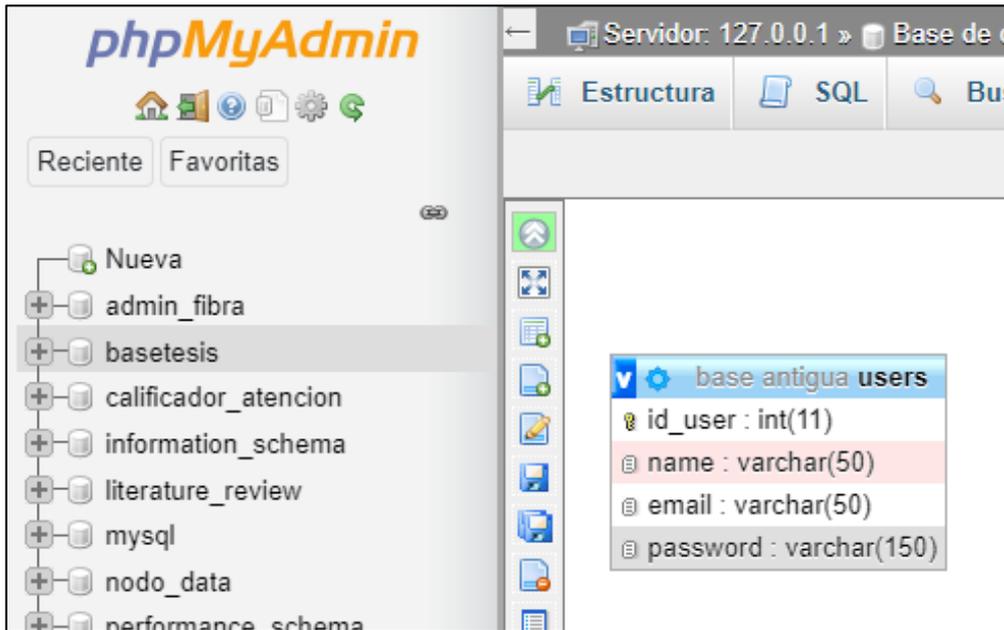


Figura 2.3. Tabla de usuarios.

Desarrollo de la vista de registro.

La vista del registro se creó conforme al prototipo diseñado con los distintos estilos y contenido figura 2.4.

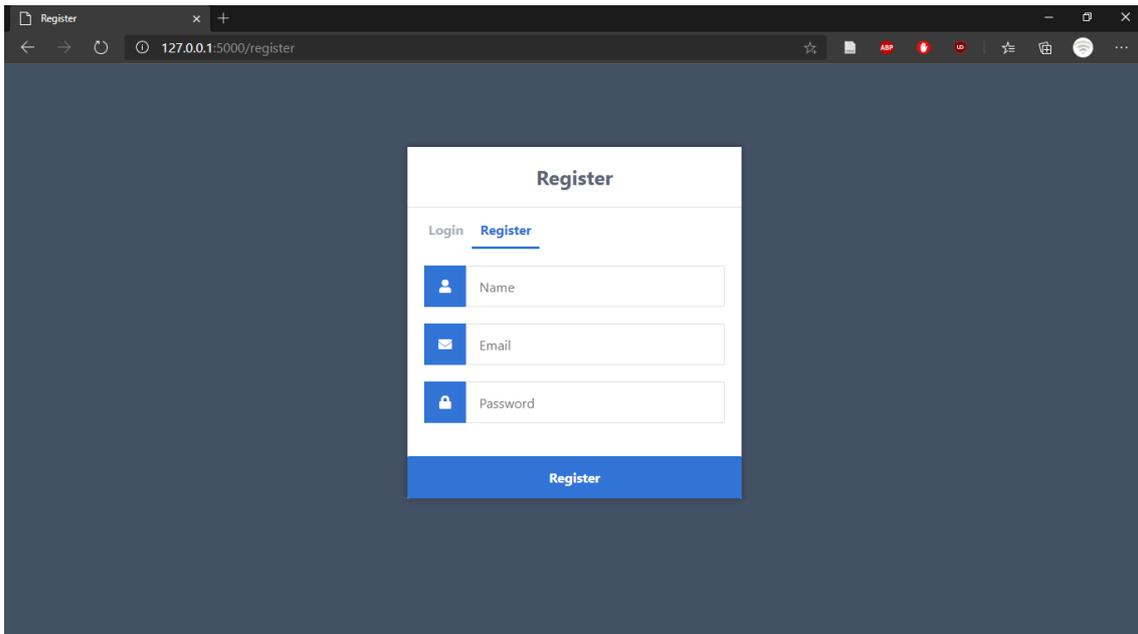


Figura 2.4. Vista del Login.

Desarrollo de la vista de inicio de sesión.

La vista del login se creó conforme al prototipo diseñado con los distintos estilos y contenido figura 2.5.

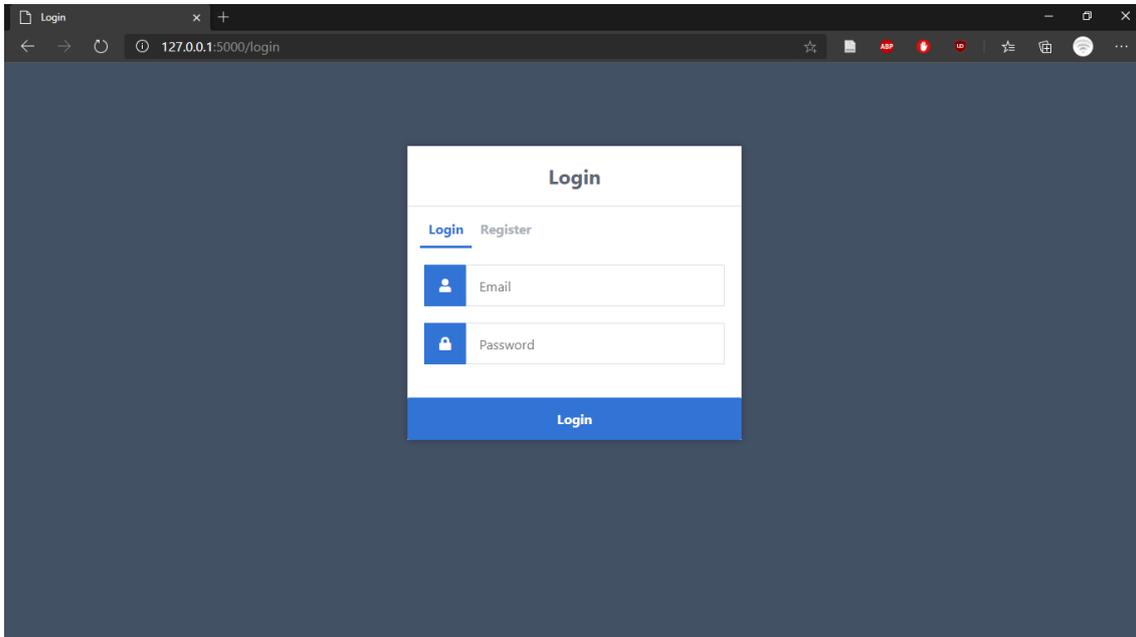


Figura 2.5. Vista del Registro.

Validación de los campos en inicio de sesión y registro.

En la figura 2.6 se puede notar las advertencias y mensajes de error confirmando la validación de los campos de **login** y **registro** que deben ser obligatorios, no se puede registrar dos usuarios con el mismo nombre y las credenciales de acceso deben ser correctas.

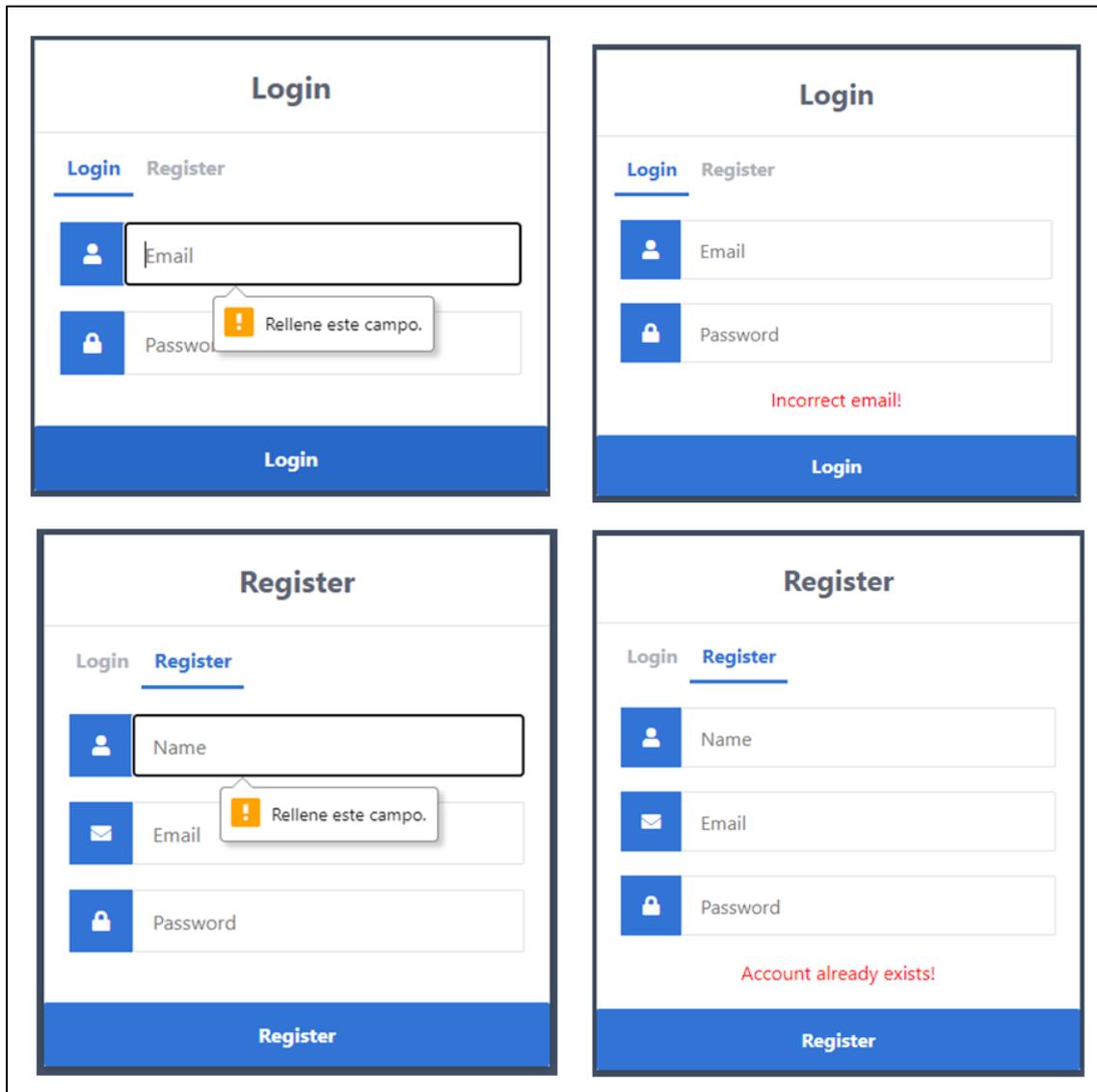


Figura 2.6. Validación de formularios de login y registro.

Diseño de la vista del listado de análisis.

Se realiza el diseño de la vista de análisis en donde se muestra la lista de análisis realizados con su respectivo estado y las opciones de crear nuevo, buscar, ver y eliminar figura 2.7.

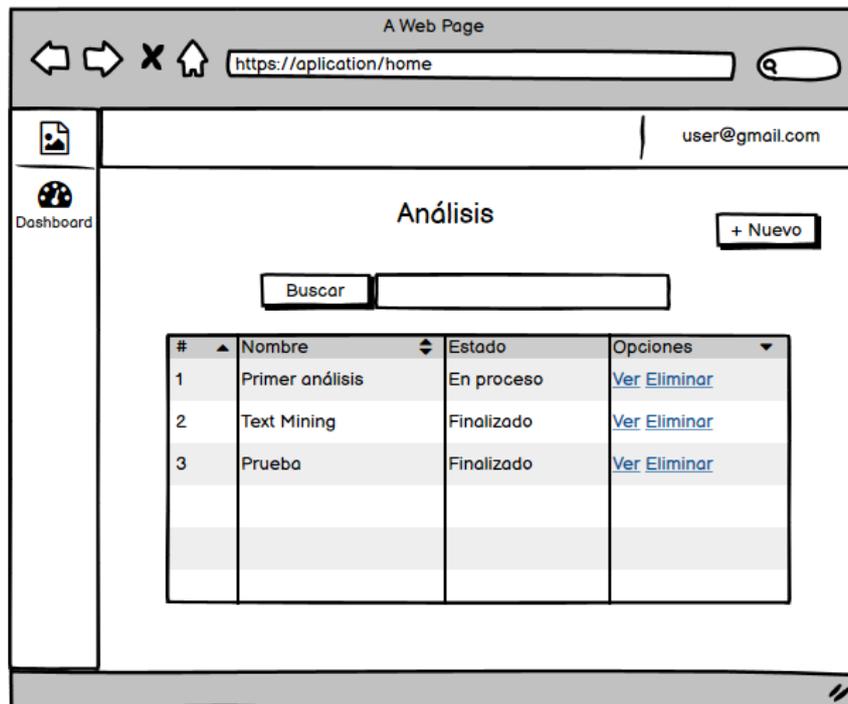


Figura 2.7. Vista del listado de análisis.

Creación de la tabla analysis en la base de datos.

En la figura 2.8 se muestra la tabla para gestionar los distintos análisis creados por el usuario.

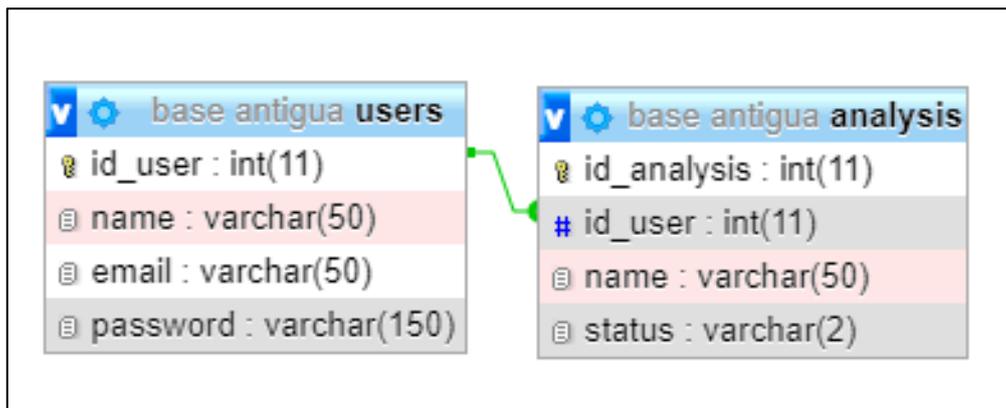


Figura 2.8. Tabla de análisis.

Desarrollo de la vista del listado de análisis.

En la figura 2.9 se muestra la vista desarrollada en donde se listan los análisis creados por cada usuario con sus respectivos estados y opciones.

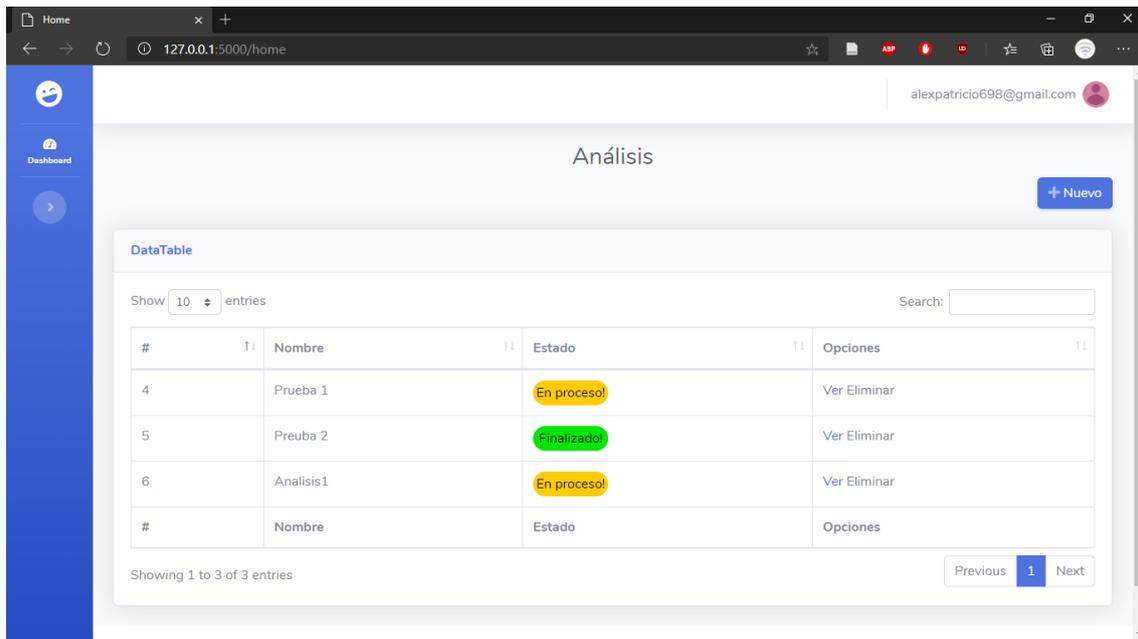


Figura 2.9. Vista del listado de análisis.

2.3.3. Planificación del sprint 2

En el sprint 2, se especifican las tareas para el cumplimiento de las historias de usuario de gestión de documentos y gestión de conceptos.

Tabla 2.15: Sprint 2 – Desarrollo de la gestión de documentos y conceptos.

Sprint 2			
Fecha inicio: 13/07/2020			
Fecha fin: 24/07/2020			
ID	Historia de Usuario	Tarea	Horas
3	Gestión de documentos.	Investigación de manejo de documentos en Python.	4
		Configurar el directorio de almacenamiento de los archivos.	2
		Implementación de las librerías necesarias.	3
		Creación de la tabla files en la base de datos relacionada con la tabla analysis .	4
		Diseño de la vista de gestión de archivos.	4
		Rutas de navegación.	1
		Crear el formulario de ingreso y selección de archivos.	4
		Desarrollo del método de consulta y listado de documentos.	3
		Desarrollo del método de eliminación de archivos.	3
		Desarrollo del método de carga de archivos.	4
		Validación de ingreso de documentos pdf.	3
Pruebas de funcionalidad.	6		
		Total	41
4	Gestión de conceptos.	Creación de la tabla concepts en la base de datos relacionada con la tabla analysis .	3
		Diseño de la vista de gestión de conceptos.	4
		Rutas de navegación.	1
		Crear el formulario de ingreso de conceptos.	3
		Crear el formulario de edición de conceptos.	2
		Desarrollo del método de consulta y listado de conceptos.	3
		Desarrollo del método de eliminación de conceptos.	2
		Desarrollo del método de creación de conceptos.	4
		Desarrollo del método de edición de conceptos.	4
		Validación de los formularios de ingreso y edición de conceptos.	3
		Pruebas de funcionalidad.	5
		Total	34
	Reuniones	Planificación del Sprint Dos.	5
		Revisión Sprint.	2
		Retrospectiva del Sprint.	2
		Total	9
		Total	84

2.3.4. Ejecución del Sprint 2

A continuación, se listan algunas tareas realizadas del Sprint 2.

Creación de la tabla files en la base de datos relacionada con la tabla analysis.

En la base de datos se debe almacenar los archivos por cada análisis que se realice por lo cual se crea una tabla de **files** relacionada con la tabla de **analysis** figura 2.10.

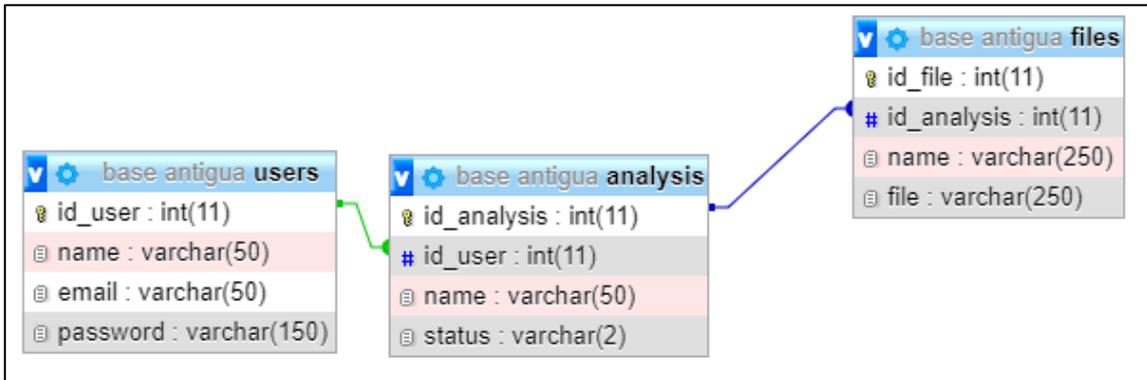


Figura 2.10. Tabla de archivos (files).

Diseño de la vista de gestión de archivos.

La figura 2.11 detalla el diseño de la vista para listar los archivos, cargados y que permita ingresar nuevos archivos, eliminar y ver.

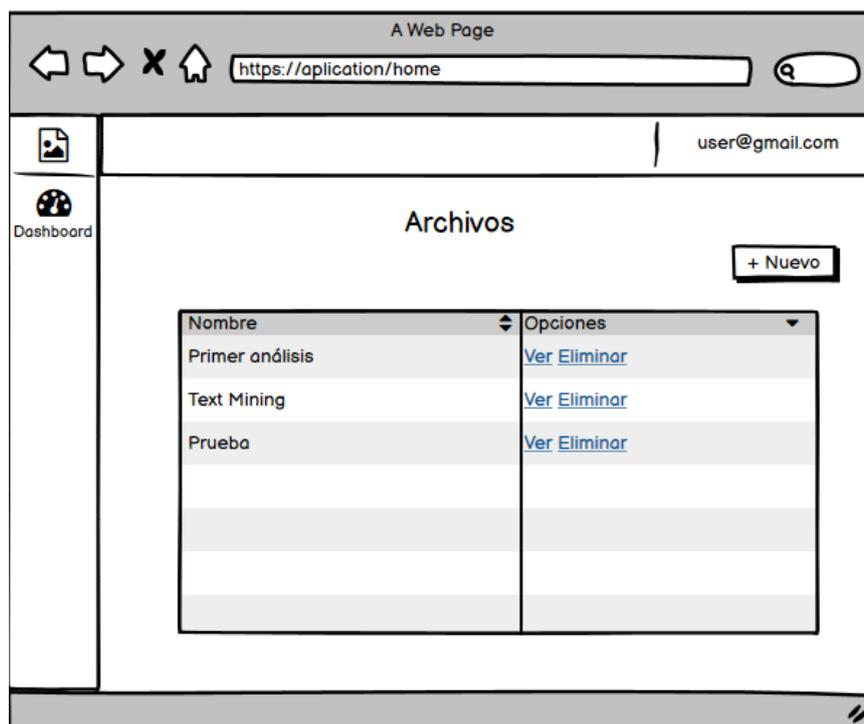


Figura 2.11. Vista de Gestión de archivos.

Crear el formulario de ingreso y selección de archivos.

En la figura 2.12 se puede evidenciar la vista del formulario de carga de archivos al sistema.



Figura 2.12. Formulario de ingreso de archivos

Creación de la tabla *concepts* en la base de datos relacionada con la tabla *analysis*.

Al igual que la tabla *files*, la tabla *concepts* debe estar relacionada con cada análisis que el usuario realiza figura 2.13.

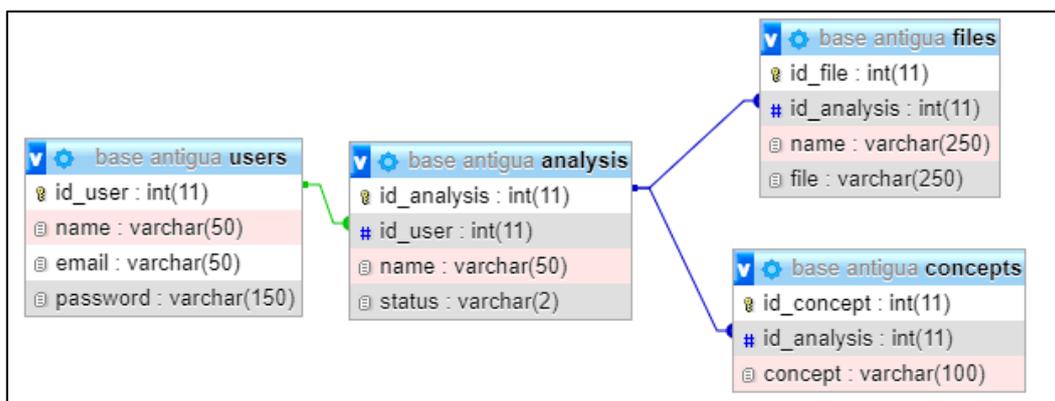


Figura 2.13. Tabla de conceptos (concepts).

Diseño de la vista de gestión de conceptos.

A continuación en la figura 2.14 se evidencia el diseño de la vista de gestión de conceptos.

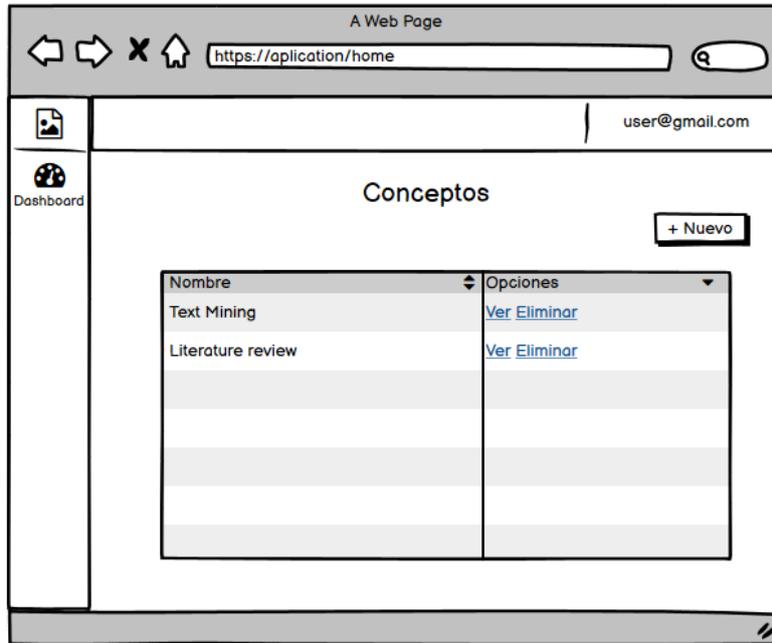


Figura 2.14. Diseño de la vista de gestión de conceptos.

Crear el formulario de ingreso de conceptos.

La figura 2.15 detalla el formulario de ingreso de los conceptos.



Figura 2.15. Formulario de ingreso de conceptos.

Crear el formulario de edición de conceptos.

En el caso de que se haya ingresado mal algún concepto puede eliminarlo o en tal caso editarlo, para lo cual se desarrollo el formulario de edición de conceptos figura 2.16.

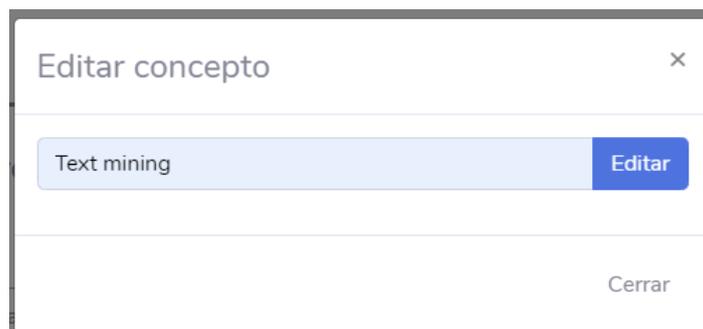


Figura 2.16. Formulario de edición de conceptos.

2.3.5. Planificación del sprint 3

En el sprint 3 se detallan las tareas para obtener el texto de cada documento cargado por el usuario, además del preprocesamiento del texto, para posteriormente realizar la comparativa o búsqueda de los conceptos.

Tabla 2.16: Sprint 3 – Extracción del texto, limpieza y preprocesamiento.

Sprint 3			
Fecha inicio: 27/07/2020			
Fecha fin: 07/08/2020			
ID	Historia de Usuario	Actividad	Horas
5	Extracción del texto de los documentos.	Investigación de extracción de texto en Python.	8
		Implementación de las librerías necesarias.	4
		Creación de rutas de prueba para la extracción.	1
		Método para consulta de documentos por análisis.	4
		Método de extracción de texto de cada página por cada documento.	8
		Almacenamiento temporal del texto.	4
		Pruebas de funcionalidad	10
		Total	39
6	Limpieza y preprocesamiento de datos.	Eliminación de correos electrónicos del texto.	4
		Eliminación de direcciones web del texto.	3
		Eliminación de citas bibliográficas del texto.	4
		Eliminación de referencias del texto.	5
		Eliminación de caracteres especiales.	3
		Pruebas de funcionalidad.	8
		Total	27
	Reuniones	Planificación del Sprint tres.	5
		Revisión Sprint.	2
		Retrospectiva del Sprint.	2
		Total	9
Total			75

2.3.6. Ejecución del Sprint 3

Debido que la principal tarea del sprint 3 es la extracción del texto y la limpieza o preprocesamiento del mismo, a continuación se presenta los resultados ya aplicando las distintas librerías y utilizando rutas de prueba.

Extracción de texto por cada documento.

Se extrae sólo el texto de los documentos cargados por el usuario, en la figura 2.17 se puede evidenciar que el texto se presenta igual que en un documento, es decir que incluye citas, referencias, saltos de línea, entre otros.

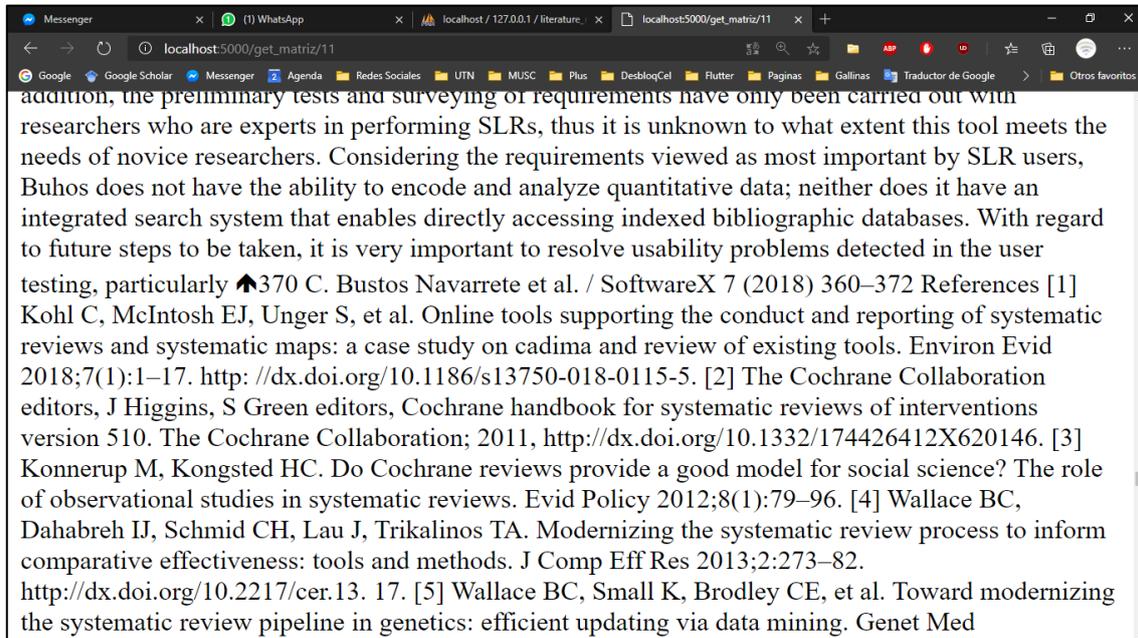


Figura 2.17. Extracción de texto de documentos.

Limpieza y preprocesamiento de datos.

En la figura 2.18 ya se puede notar la diferencia del texto original, en donde el texto está sin ruido pero sin perder su estructura sintáctica, esto con el objetivo de tener un mejor procesamiento para la búsqueda o comparación de los conceptos.

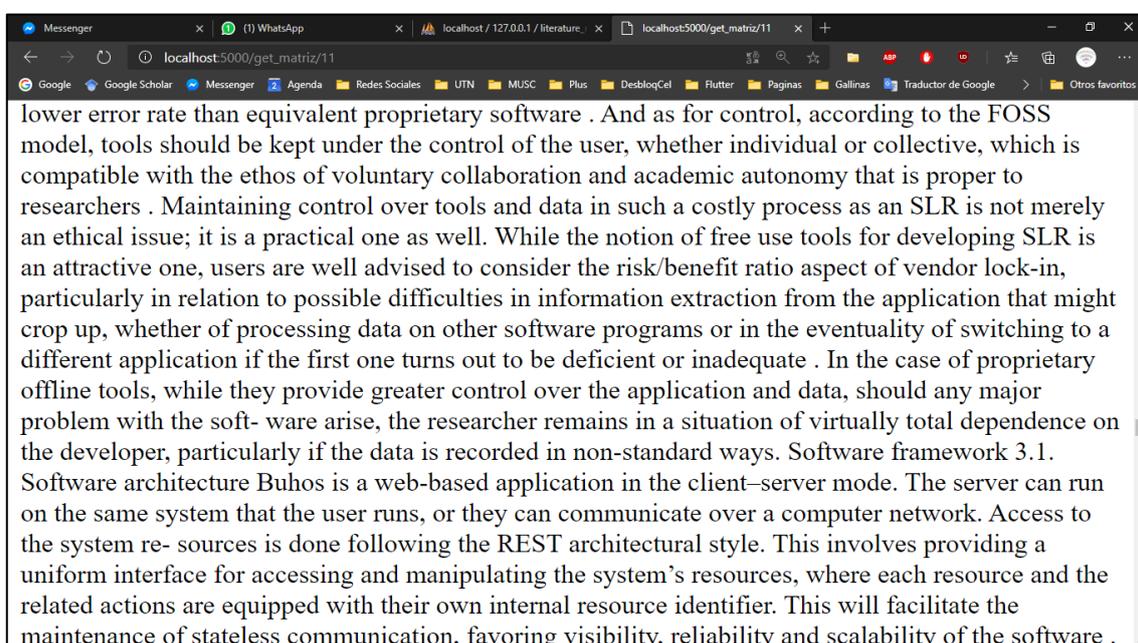


Figura 2.18. Texto después de la limpieza de datos.

2.3.7. Planificación del sprint 4

En el sprint 4 se desarrolla una de las historias de usuario de mayor prioridad, la implementación del modelo de inteligencia artificial para la comparación entre los conceptos y las oraciones de cada documento, obteniendo el porcentaje de coincidencia.

Tabla 2.17: Sprint 4 - Implementación del modelo de inteligencia artificial.

Sprint 4			
Fecha inicio: 10/08/2020			
Fecha fin: 21/08/2020			
ID	Historia de Usuario	Actividad	Horas
7	Implementación del modelo de inteligencia artificial.	Creación de la tabla results en la base de datos relacionada con las tablas files y concepts .	3
		Creación de la tabla sentences en la base de datos relacionada con la tabla results	3
		Creación de rutas para pruebas.	1
		Investigación de modelos y librerías de inteligencia artificial en Python.	10
		Implementación de las librerías necesarias.	4
		Descargar un modelo de minería de texto.	10
		Cargar el modelo de minería de texto al sistema.	8
		Separar el texto en oraciones.	4
		Realizar una limpieza del texto.	6
		Obtener un vector numérico de cada oración.	5
		Realizar una limpieza de cada concepto.	4
		Obtener un vector numérico de cada concepto.	5
		Implementar un método para encontrar la similitud entre los vectores de las oraciones y conceptos.	8
		Pruebas de funcionalidad.	10
		Total	81
	Reuniones	Planificación del Sprint Cuatro.	5
		Sprint Review.	2
		Sprint Retrospective.	2
		Total	9
Total			90

2.3.8. Ejecución del Sprint 4

A continuación se detallan algunas actividades realizadas en el sprint 4.

Creación de la tabla results en la base de datos relacionada con las tablas files y concepts.

La figura 2.19 muestra la tabla de resultados en donde se va a almacenar el porcentaje de coincidencia entre un concepto y un archivo.

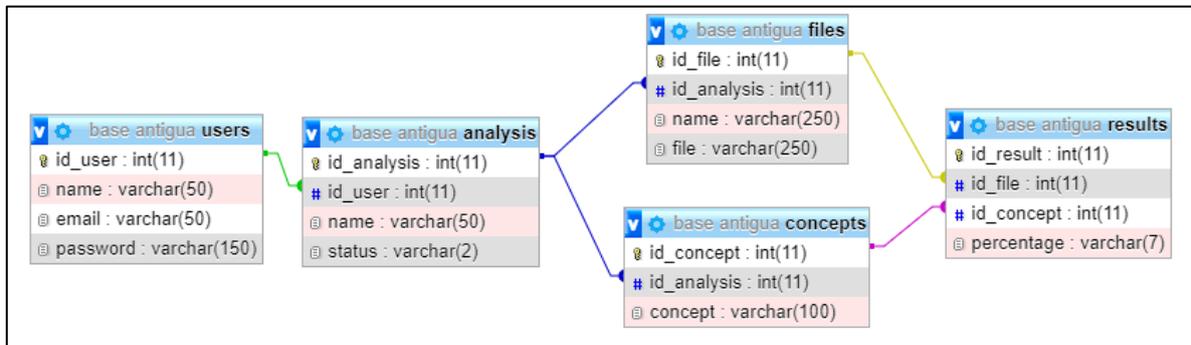


Figura 2.19. Tabla de resultados (results).

Creación de la tabla sentences en la base de datos relacionada con la tabla results.

Además de la tabla de resultados se requería una tabla para almacenar las oraciones con el mayor porcentaje de coincidencia de cada resultado figura 2.20.

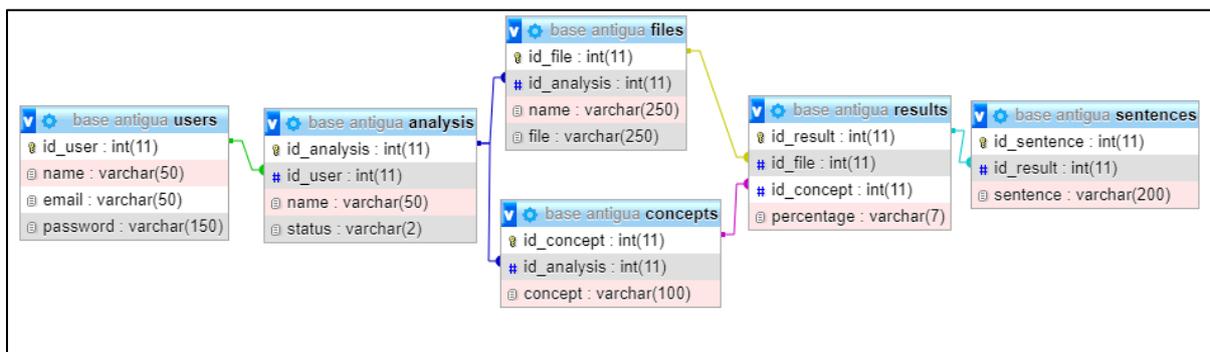


Figura 2.20. Tabla de oraciones (sentences).

Implementar el coseno de similitud entre los vectores de las oraciones y conceptos.

Se desarrollo el método aplicando el coseno de similitud entre cada concepto y cada oraciones de los documentos, en la figura 2.21 se observa el resultado de la comparación de un concepto con las oraciones de un archivo.

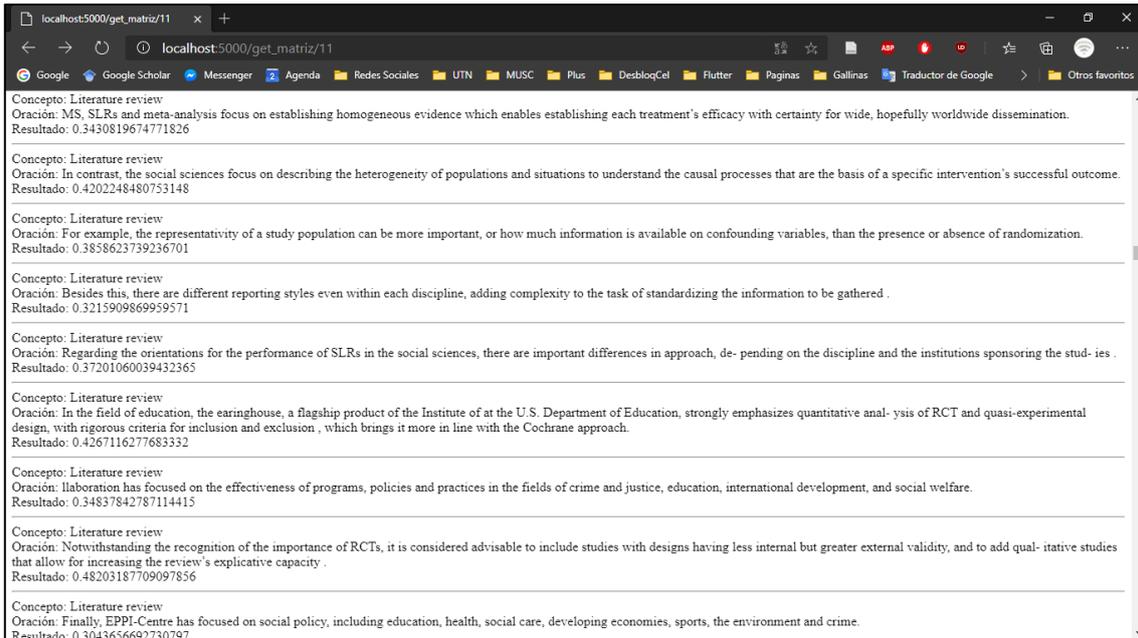


Figura 2.21. Comparación de conceptos con oraciones.

2.3.9. Planificación del sprint 5

En el sprint 5 se desarrolla la presentación de resultados en una matriz o tabla, especificando si un concepto esta presente en un archivo o no.

Tabla 2.18: Sprint 5 - Reporte de matriz de conceptos.

Sprint 5			
Fecha inicio: 24/08/2020			
Fecha fin: 11/09/2020			
ID	Historia de Usuario	Actividad	Horas
8	Reporte de matriz de conceptos.	Integración del método de extracción de texto con el modelo de inteligencia artificial.	10
		Procesar los documentos y conceptos por cada análisis	6
		Calcular el porcentaje de coincidencia de cada concepto con cada texto o documento.	8
		Obtener las tres primeras oraciones con mayor coincidencia.	5
		Almacenar los resultados del análisis en la base de datos.	5
		Diseño de la vista de la matriz de conceptos resultante.	4
		Rutas de navegación.	1
		Desarrollo de la vista de la matriz de conceptos con las oraciones con mayor similitud.	5
		Desarrollo del método de consulta y organización de información para la matriz de conceptos.	6
		Controlar que un análisis finalizado ya no permita generar nuevamente la matriz de conceptos.	3
Pruebas de funcionalidad.	10		
		Total	63
	Reuniones	Planificación del Sprint Cinco.	5
		Sprint Review.	2
		Sprint Retrospective.	2
		Total	9
Total			72

2.3.10. Ejecución del Sprint 5

A continuación se presenta algunas tareas descritas en el Sprint 5.

Diseño de la vista de la matriz de conceptos resultante.

La figura 2.22 muestra el diseño como tal de la matriz resultante de cada análisis.

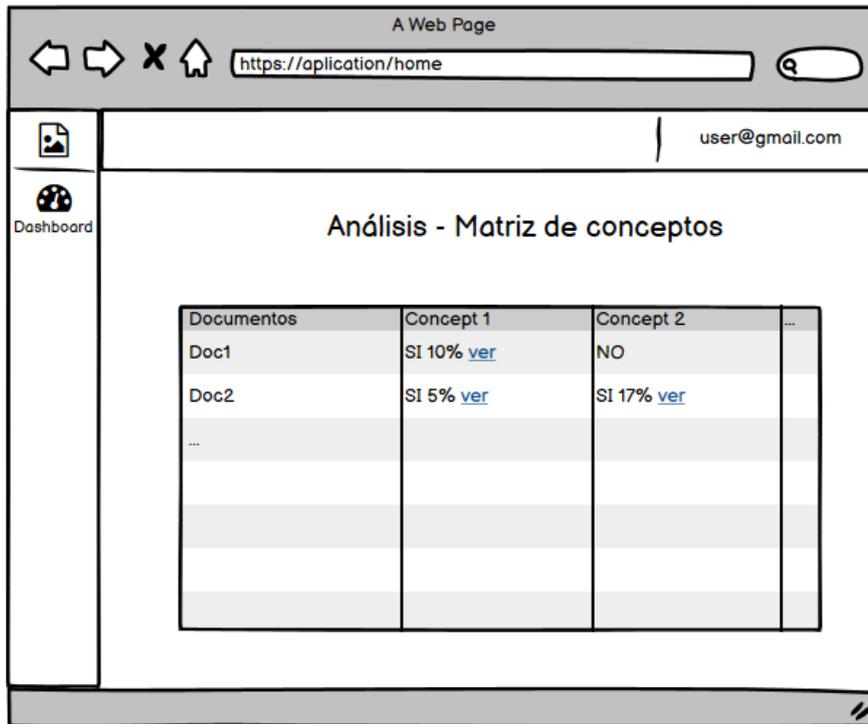


Figura 2.22. Diseño de la vista de Matriz de conceptos.

Desarrollo de la vista de la matriz de conceptos con las oraciones con mayor similitud.

La vista de la matriz generada por cada análisis, detallando si los conceptos están presentes en cada archivo y de ser así mostrando el porcentaje de coincidencia se evidencia en la figura 2.23.

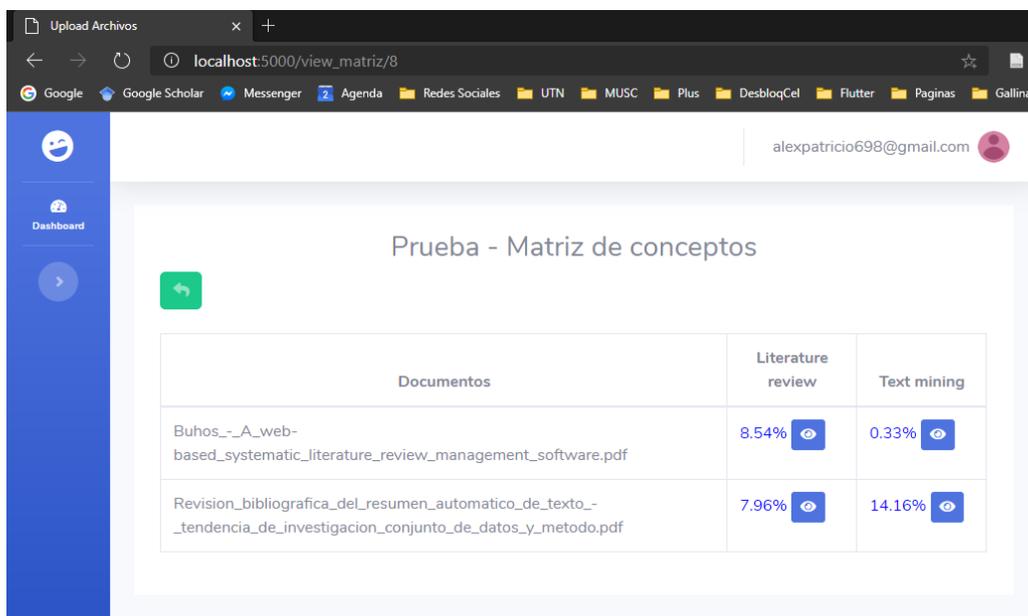


Figura 2.23. Vista de matriz de conceptos.

En la figura 2.24 se muestra el modal en el cual aparecen las oraciones con mayor coincidencia de cada resultado.

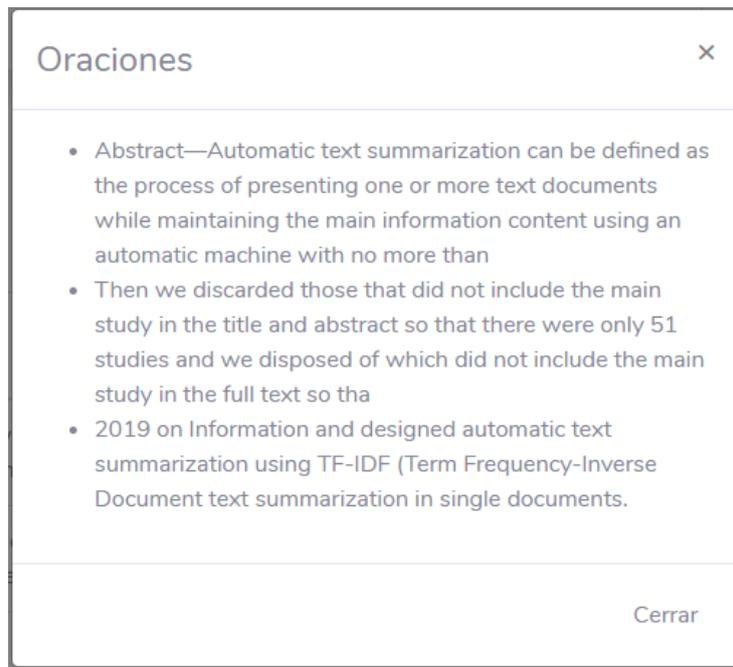


Figura 2.24. Modal de oraciones encontradas.

2.3.11. Planificación del sprint 6

Por último, en el sprint 6 se realiza las actualizaciones del reporte de la matriz de conceptos.

Tabla 2.19: Sprint 6 - Refinamiento del reporte de matriz de conceptos v.2.

Sprint 6			
Fecha inicio: 14/09/2020			
Fecha fin: 25/09/2020			
ID	Historia de Usuario	Actividad	Horas
9	Refinamiento del reporte de matriz de conceptos v.2.	Validar ingreso de conceptos en inglés.	4
		Limitar el tamaño de los archivos que se permite cargar al sistema.	3
		Modificación de la vista de matriz de conceptos.	2
		Guardar todas las oraciones que coincidan con los conceptos.	6
		Desarrollo de la vista para mostrar en orden descendente las oraciones con el porcentaje de coincidencia.	4
		Permitir cargar más archivos una vez realizado el análisis.	6
		Permitir cargar más conceptos una vez realizado el análisis.	5
		Procesar documentos y conceptos agregados después del análisis.	6
		Pruebas de funcionalidad.	10
Total			46
Reuniones		Planificación del Sprint Cinco.	5
		Sprint Review.	2
		Sprint Retrospective.	2
		Total	9
Total			55

2.3.12. Ejecución del Sprint 6

A continuación se presenta algunas tareas descritas en el Sprint 6.

Modificación de la vista de matriz de conceptos.

La figura 2.25 muestra el diseño como tal de la matriz resultante de cada análisis.



Figura 2.25. Modificación de la vista de matriz de conceptos.

Desarrollo de la vista para mostrar en orden descendente las oraciones con el porcentaje de coincidencia.

La vista en donde se muestra las oraciones con el porcentaje de coincidencia calculado con respecto a un determinado concepto se muestra en la figura 2.26.

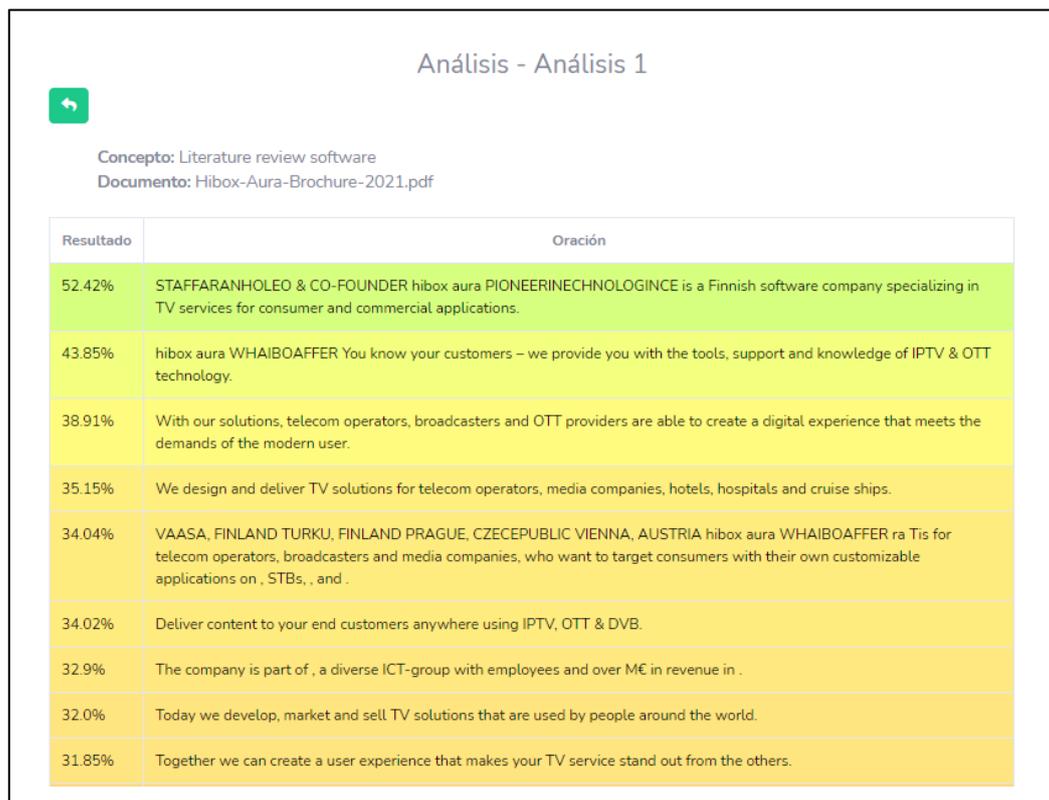


Figura 2.26. Lista de las oraciones con el porcentaje de coincidencia.

Permitir cargar más archivos y conceptos una vez realizado el análisis.

En la figura 2.27 se muestra la vista donde se puede agregar archivos y conceptos a un análisis ya finalizado, y permite actualizar la matriz de conceptos.

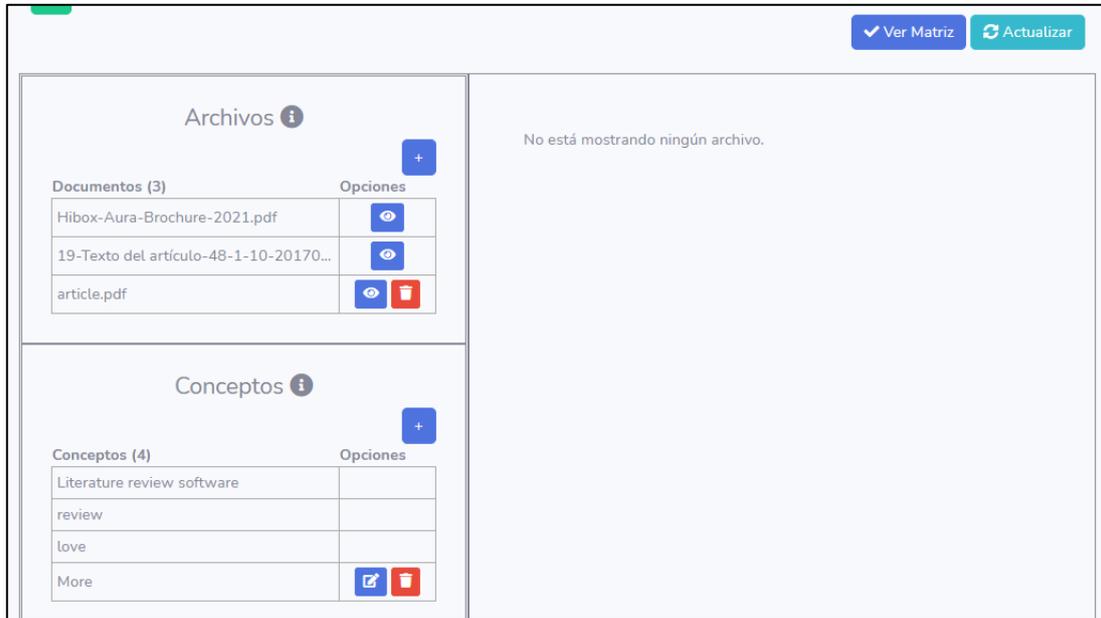


Figura 2.27. Actualizar análisis finalizado.

CAPÍTULO 3.

Resultados

Una vez finalizada la etapa de desarrollo se subió temporalmente la aplicación a un servidor para que los usuarios puedan acceder con mayor facilidad y hagan uso del sistema. Los usuarios estuvieron conformados por algunos estudiantes y docentes de la Universidad Técnica del Norte, la mayor parte de los usuarios con un rango de edades entre 18 y 25 años.

Después de un cierto tiempo de uso del sistema, se diseñó una encuesta con base en un cuestionario realizado por Ojo, A. (2017) el cual incluye las variables de éxito de los sistemas de información planteado en el modelo de Delone y Mclean que es un modelo para medir las distintas dimensiones que conforman a un sistema.

Las preguntas de la encuesta se diseñaron para obtener datos confiables utilizando preguntas cerradas, las cuales se dividieron en el perfil demográfico, calidad del Sistema, calidad de la información, calidad del servicio, uso - intensidad de uso, satisfacción del usuario y beneficios obtenidos.

La encuesta se aplicó a las personas que hicieron uso del mismo, de un total de 15 a quienes se les realizó la invitación, el 73% hicieron uso del sistema, además respondieron la encuesta.

Este capítulo se centra en la validación del sistema con base en un análisis e interpretación de los datos recopilados a partir de la encuesta aplicada a los usuarios.

3.1. Sección 1: Perfil demográfico de los usuarios

En esta sección se detalla el perfil demográfico de los usuarios comprendido únicamente por el sexo, y la edad.

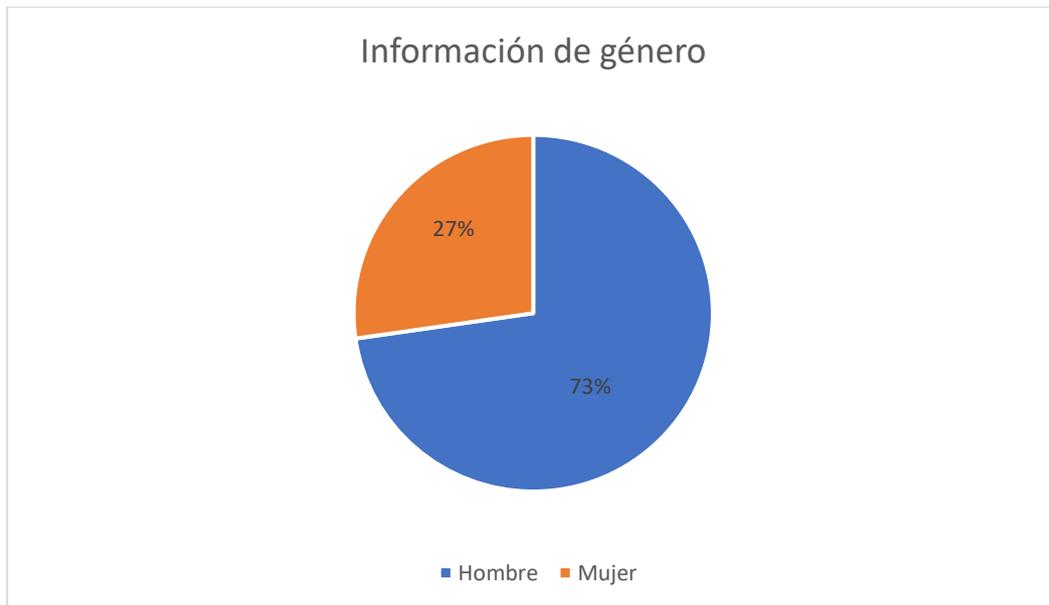


Figura 3.1. Información de género.

La Figura 3.1 muestra que la mayor parte de los que usaron el sistema y respondieron la encuesta eran hombres, ya que del total de encuestados, el 27% eran mujeres y el 73% eran hombres.

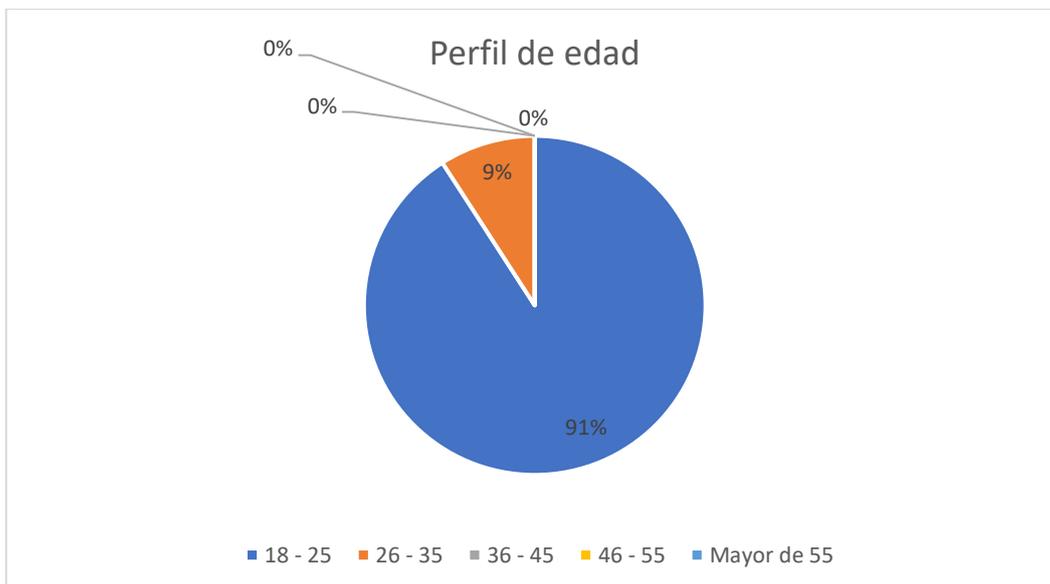


Figura 3.2. Perfil de edad.

El perfil de edad se muestra en la figura 3.2, en donde se evidencia que la mayor concentración de encuestados corresponde al grupo de edad de 18 a 25 años, esto debido a que el sistema fue usado en su mayoría por estudiantes universitarios.

3.2. Sección 2: Variables del modelo de éxito de los sistemas de información de Delone y Mclean

En la siguiente sección se presenta la síntesis de respuestas relacionadas con las 6 dimensiones del modelo de DeLone y McLean: Calidad del sistema, Calidad de la información, Calidad del servicio, Uso e intención de uso, Satisfacción del usuario y beneficios obtenidos. Se solicitó específicamente a los encuestados que indicaran si están de acuerdo o en desacuerdo con las afirmaciones presentadas en cada sección, para las respuestas se usó la escala de Likert con una numeración de 1 a 5, donde 1 significa en total desacuerdo y 5 corresponde a una total aceptación. A continuación, se presenta un análisis más detallado de las diferentes variables.

3.2.1. Calidad del Sistema

La calidad del sistema es una característica que abarca el funcionamiento y la forma en que procesa la información, se debe tomar en cuenta la accesibilidad, el tiempo de respuesta, además de la facilidad de uso.

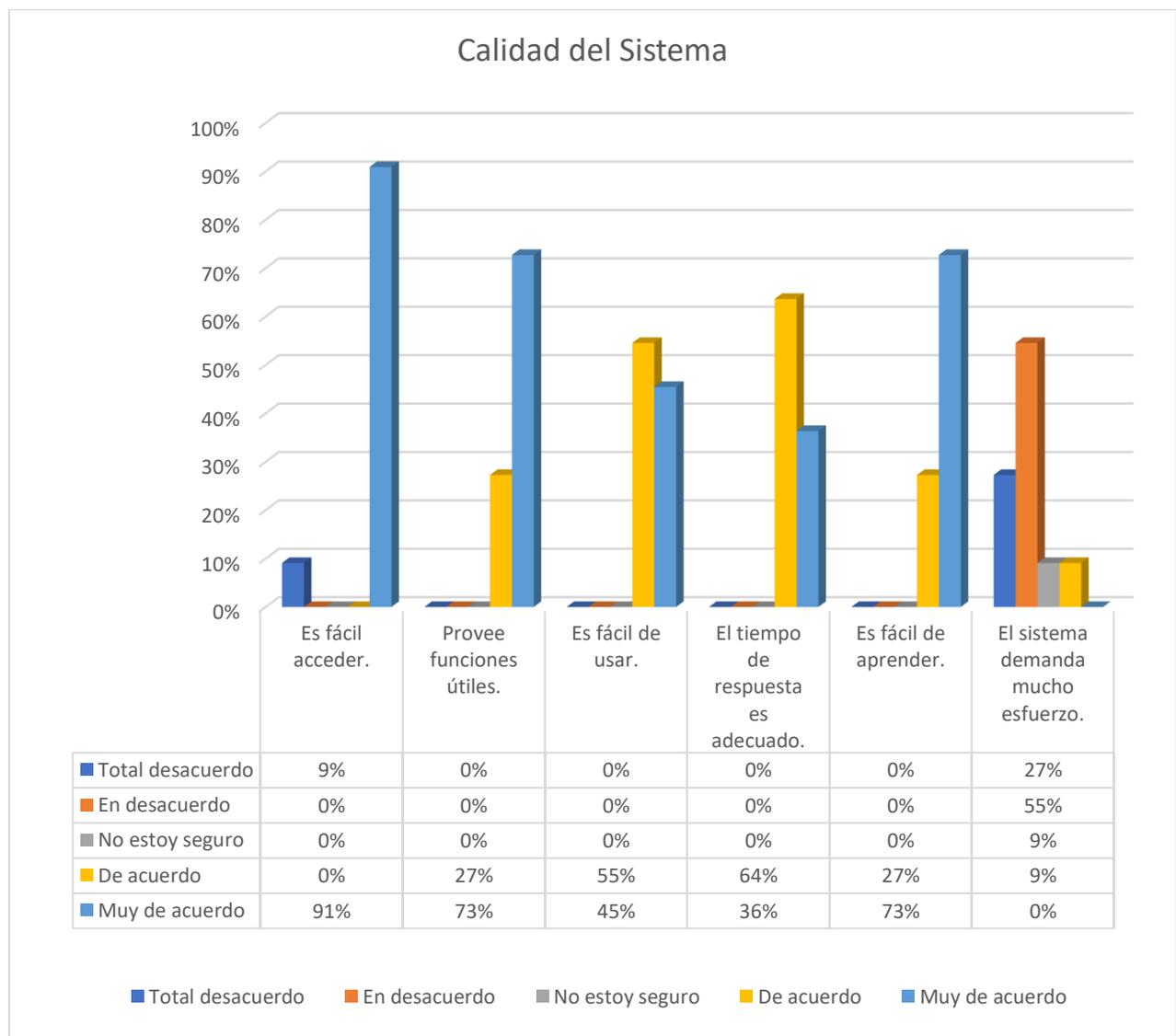


Figura 3.3. Calidad del sistema.

En la figura 3.3 se muestra las afirmaciones de la encuesta acerca de la calidad del sistema, donde se puede notar que los usuarios respondieron en su mayoría positivamente, lo que da a entender que el sistema tiene una buena calidad. El 91% de los encuestados estuvo de acuerdo que el sistema es fácil de acceder, solo el 9% respondió que está en total desacuerdo. En cuanto a las afirmaciones; el sistema provee funciones útiles, es fácil de usar, tiempo de respuesta adecuado y es fácil de aprender, las respuestas variaron entre de acuerdo y muy de acuerdo. Únicamente en la última afirmación se obtuvieron respuestas negativas esto debido a que el sistema en sí no demanda de mucho esfuerzo.

3.2.2. Calidad de la información.

La calidad de la información se refiere a como se genera la información o como se muestran los resultados en el sistema, la precisión, la confiabilidad, la cantidad de información, la relevancia, si es adecuada y si está actualizada.

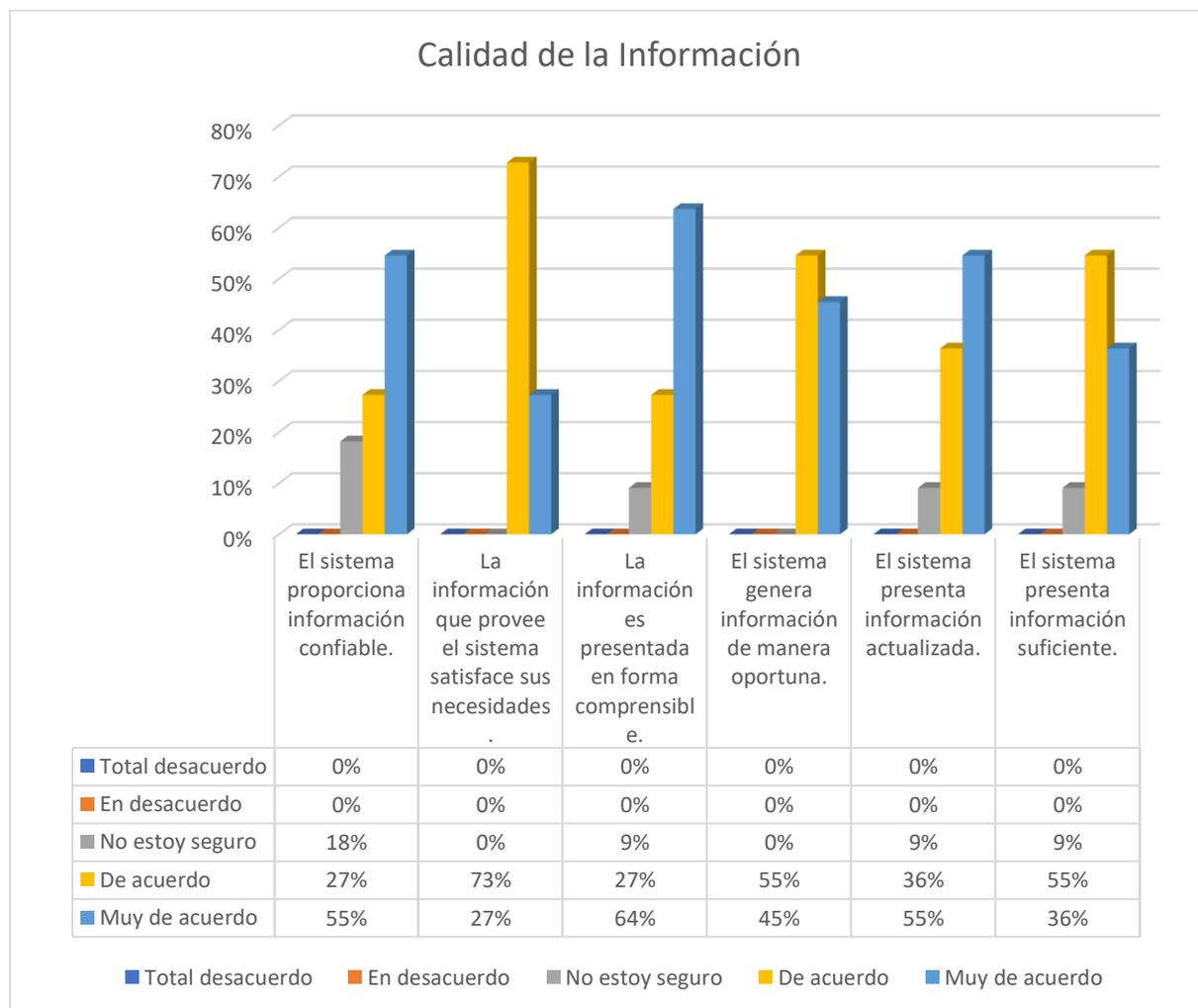


Figura 3.4. Calidad de la información.

Según la figura 3.4 los encuestados en su mayoría están satisfechos con la información y el contenido del sistema. Un 18% de los encuestados indico que no está seguro si la información que proporciona el sistema es confiable, le resto de los encuestados estuvo de acuerdo (27%) y muy de acuerdo (55%). Con respecto a la satisfacción de las necesidades con la información presentada el 73% estuvo de acuerdo y el 27% muy de acuerdo. El 9% de las repuestas obtenidas fueron neutrales con base en la comprensibilidad de la información, si es actual y suficiente, mientras que el resto de respuestas para las mismas afirmaciones fueron del 91% de aceptación es decir estaba de acuerdo y muy de acuerdo respectivamente, lo que significa que los usuarios aprueban la calidad de la información y en general el sistema está organizado correctamente para facilitar su uso.

3.2.3. Calidad del servicio

La calidad de servicio se relaciona con la efectividad del soporte brindado a los usuarios, como también la disponibilidad del sistema para tener acceso en cualquier momento, además evidenciar si el sistema cumple o no con las funcionalidades para las que fue diseñado.

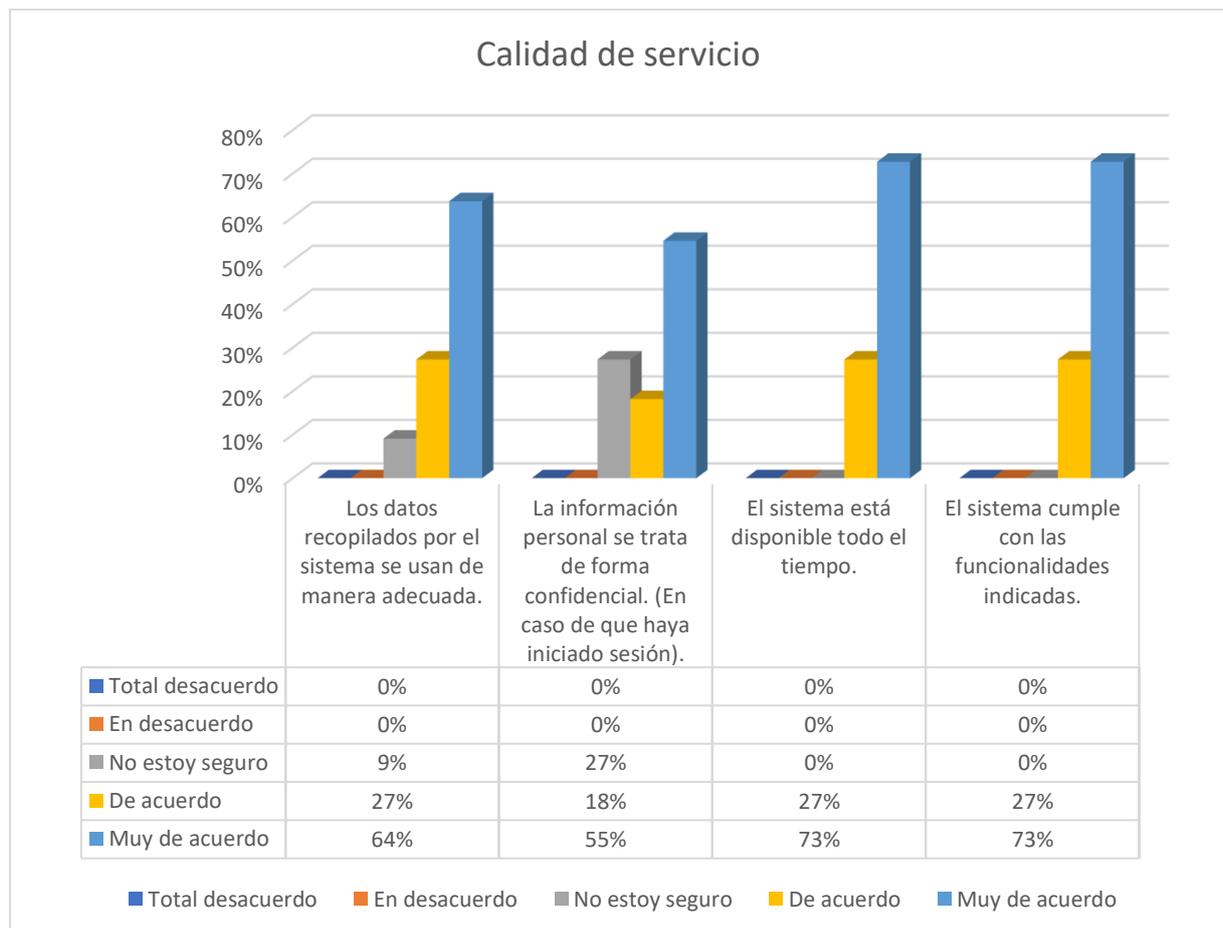


Figura 3.5. Calidad de servicio.

Para la validación de la variable de calidad de servicio en la encuesta se incluyeron las afirmaciones presentes en la figura 3.5. Las afirmaciones; los datos recopilados por el sistema se usan de manera adecuada y la información personal se trata de forma confidencial, tuvieron algunas respuestas neutrales siendo un 9% y 27% respectivamente y los demás encuestados están de acuerdo, además que todos los encuestados esta de acuerdo y muy de acuerdo sobre la disponibilidad del sistema y que el sistema cumple con las funcionalidades indicadas.

3.2.4. Uso - Intensión de uso

En esta parte hace referencia a la intención de uso que tienen los usuarios, dependiendo de los beneficios que el sistema les brinde como es el incremento de la productividad, mejoras en el desempeño de su trabajo que en este caso es la realización de revisiones de literatura.

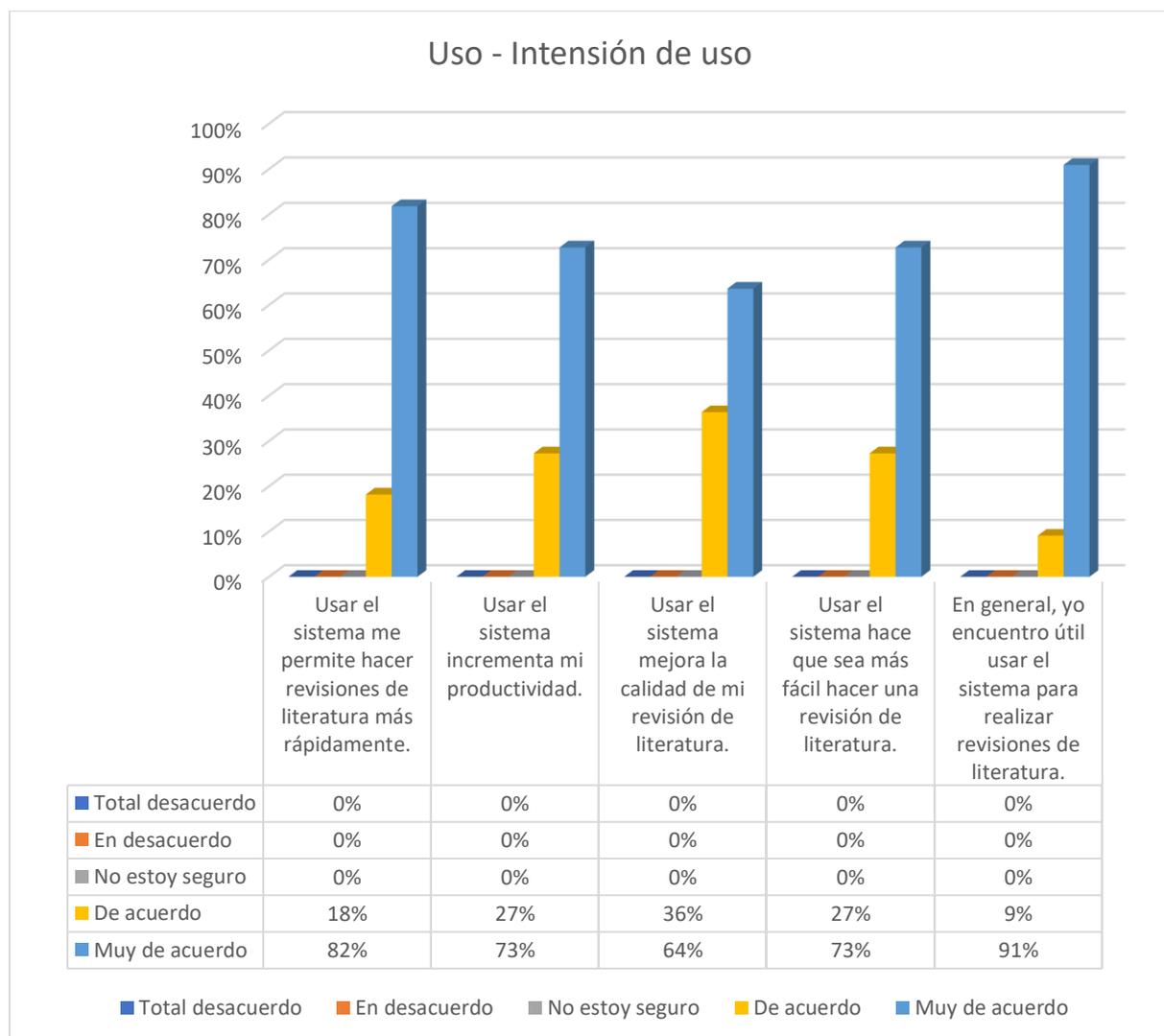


Figura 3.6. Uso -Intención de uso.

En esta sección los encuestados en definitiva tienen la intención de usar el sistema, ya que como se muestra en la figura 3.6, las respuestas son positivas en todas las afirmaciones, en donde el 100% de las respuestas está dividido entre de acuerdo y muy de acuerdo.

3.2.5. Satisfacción del usuario

Con esta variable se trata de identificar que tan satisfechos se sienten los usuarios con el sistema, si están de acuerdo que el sistema presenta la información necesaria, si cumple con lo que ellos esperaban, además de que si piensa en seguir usando el sistema.

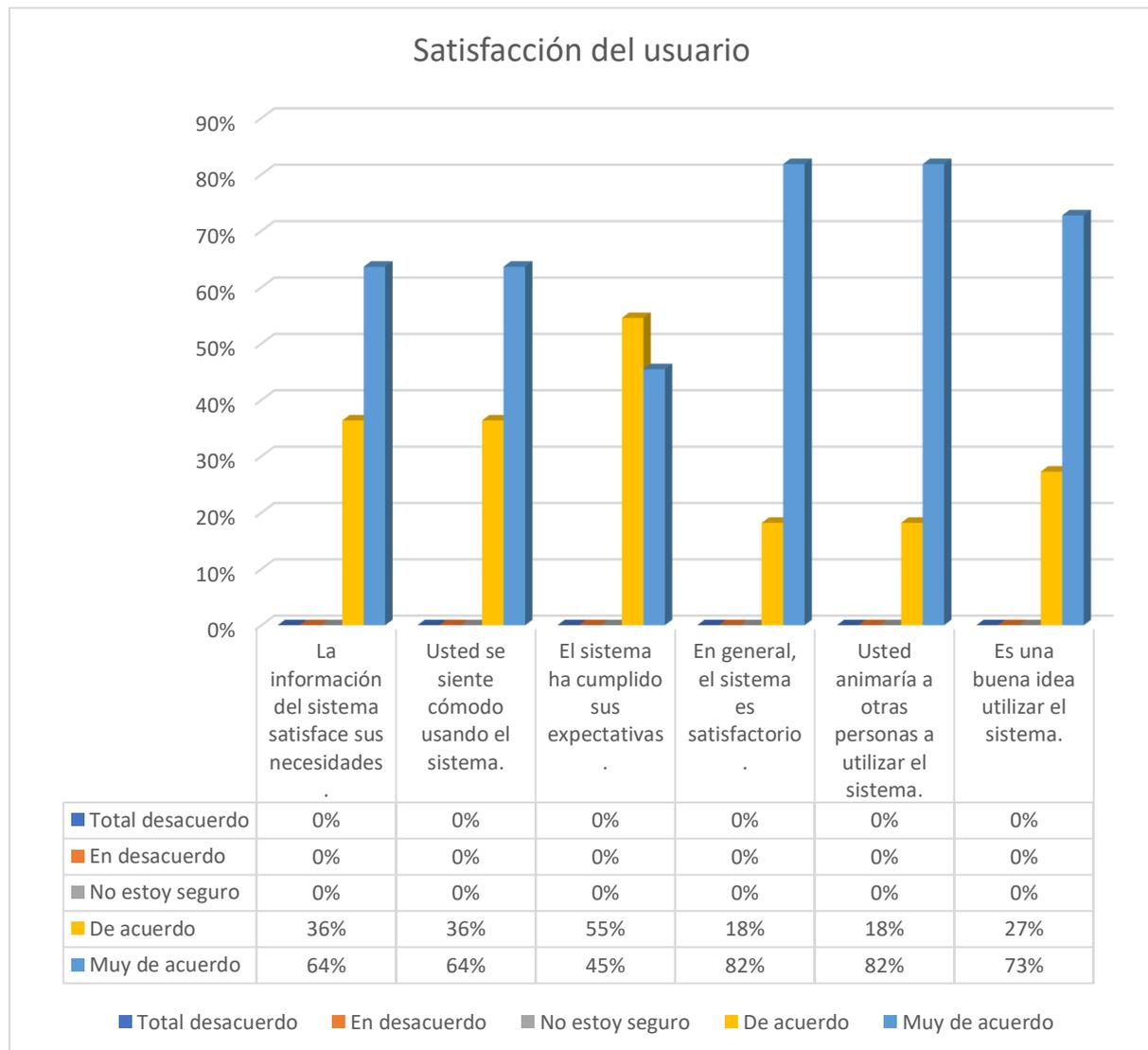


Figura 3.7. Satisfacción del usuario.

De igual manera que en la anterior sección los encuestados respondieron positivamente, lo que da a entender que el uso del sistema fue satisfactorio para ellos. En la primera afirmación que se especifica si el sistema satisface sus necesidades, se tiene un 35%

de respuestas que están de acuerdo y un 64% de respuestas que están muy de acuerdo. Los usuarios se sienten cómodos al usar el sistema, ya que el 36% de encuestados están de acuerdo con esta afirmación y el 64% están muy de acuerdo. En cuanto a las expectativas de los usuarios con respecto al sistema fueron las que esperaban por lo que se tiene un 55% y 45% que están de acuerdo y muy de acuerdo respectivamente. Debido a la satisfacción de los usuarios, ellos respondieron de forma muy positiva a la afirmación; es buena idea utilizar el sistema, y por ende ellos pueden animar a otras personas para que hagan uso del sistema.

3.2.6. Beneficios obtenidos

Esta variable se refiere a que beneficios consiguen los usuarios al utilizar el sistema, es decir la contribución que les brinda el sistema para su desarrollo tanto personal como social.

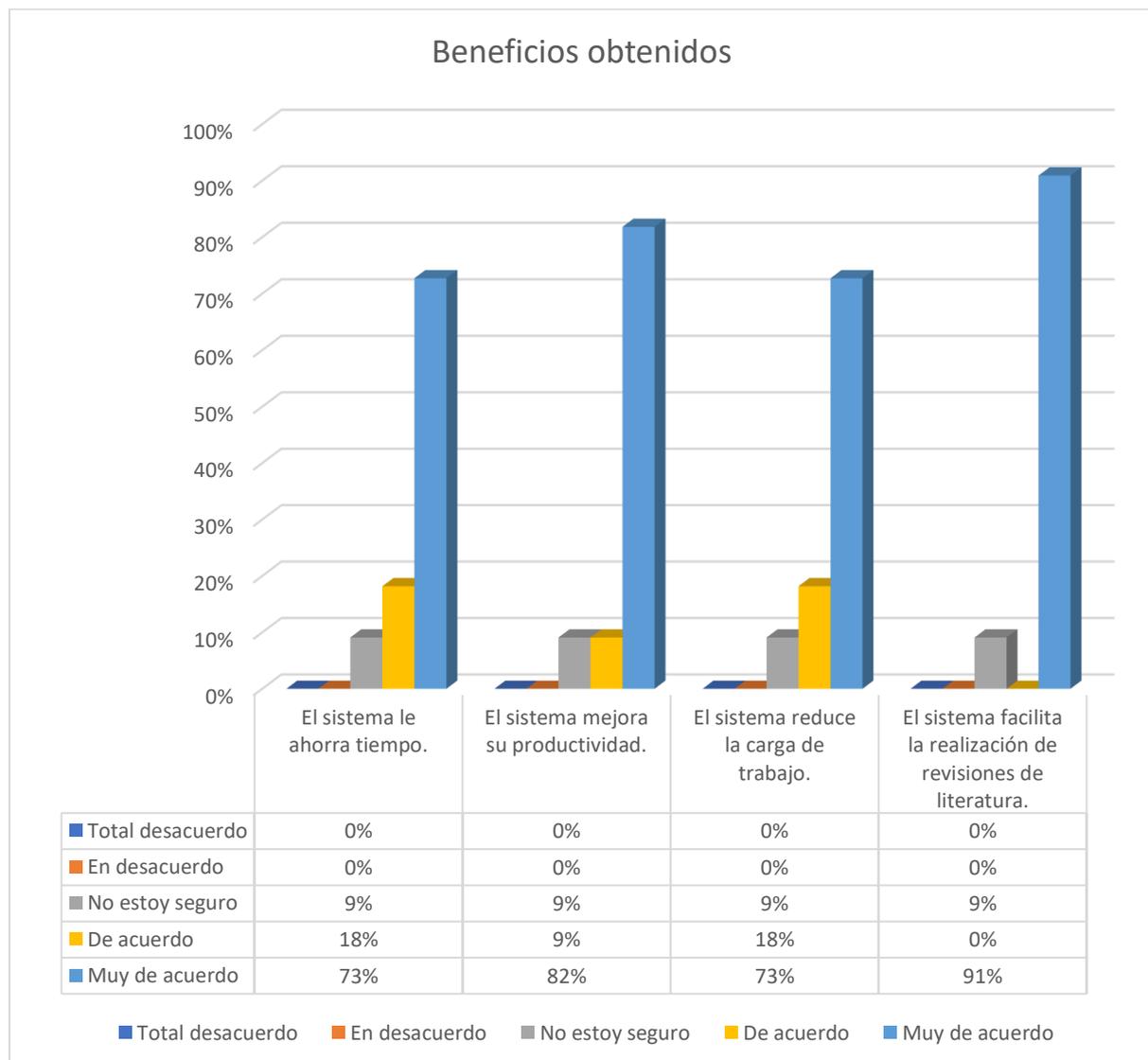


Figura 3.8. Beneficios obtenidos.

En las afirmaciones de esta sección que se muestran en la figura 3.8, el 9% de los encuestados no estaban seguros de los beneficios que se obtienen a partir del sistema, tal vez porque requieren usar el sistema más tiempo para poder definir bien que beneficios en sí les brinda el sistema, el resto de encuestados respondieron positivamente estando en su mayoría muy de acuerdo con las afirmaciones presentadas.

El modelo planteado por Delone y Mclean para medir el éxito de los sistemas de información ayudo a evaluar la efectividad del sistema que se desarrolló, tras aplicar la encuesta a los usuarios, se obtuvo resultados positivos en las distintas categorías presentes en el modelo;

En cuanto a la calidad del sistema, el 91% de los usuarios estuvieron muy de acuerdo que se puede acceder fácilmente al sistema, además de estar de acuerdo y muy de acuerdo que el sistema es fácil de usar y los tiempos de respuesta son los adecuados, los usuarios indicaron que la calidad de la información es muy aceptable, ya que la información que se presenta es bastante precisa, confiable y relevante para las investigaciones que se realizan, de igual forma en la calidad del servicio los encuestados estuvieron muy de acuerdo en que el sistema cumple con las funcionalidades especificadas y que presenta una alta disponibilidad, un 27% de los encuestados indico que no estaba seguro en cuanto al manejo de la información confidencial, ante ello se puede resaltar que el sistema trabaja con normas de seguridad, control de sesiones y encriptación de las contraseñas por lo cual la información está segura, como parte final los encuestados tiene la intención de seguir usando el sistema, ya que están muy satisfechos con los beneficios obtenidos, como son la mejora de productividad, la reducción de la carga de trabajo y en definitiva el ahorro de tiempo.

CONCLUSIONES

- Se cumplió el total de los objetivos planteados en el inicio de esta investigación, como son la elaboración del marco referencial aunque cabe mencionar que en la búsqueda bibliográfica no se encontró mucha información referente al tema, pero se trató de obtener la información más esencial. Además de ello se cumplió el objetivo de desarrollar del sistema como tal y realizar la validación del sistema con el modelo de Delone y Mclean.
- El sistema que se desarrolló en una primera versión presenta la información en modo general y si ayuda a realizar las revisiones de literatura de una forma más efectiva y a tener una mejor organización de la información.
- Flask es un potente marco de trabajo (Framework) de desarrollo web que usa como lenguaje de programación Python, con el que se pueden desarrollar aplicaciones bastante robustas debido a la gran cantidad de métodos y librerías que se pueden utilizar principalmente librerías de inteligencia artificial que son de código abierto.
- Gracias a la metodología Scrum se desarrolló el sistema de una forma muy fácil y organizada. Se definió principalmente las historias de usuario reflejando los requerimientos del sistema levantados previamente y luego se estableció los sprint con su determinado tiempo y tareas especificadas.
- Con el modelo de éxito de los sistemas de información planteado por Delone y Mclean se realizó la validación del sistema una vez finalizado su desarrollo, obteniendo resultados mayormente positivos en todas las secciones especificadas en las encuestas, además se evidenció el gran interés que presentaron los usuarios.
- El sistema desarrollado permite el registro e inicio de sesión de los usuarios, una vez iniciado sesión se puede gestionar varios análisis que se vayan creando, en cada análisis se puede cargar documentos y conceptos para generar una matriz de conceptos que muestre si los conceptos están presentes en los artículos y de ser así, muestre el porcentaje de coincidencia y las oraciones encontradas.
- La inteligencia artificial ha permitido que las máquinas sean capaces de ejecutar algunos procesos similares a los que el hombre realiza. Aunque aún no se ha logrado que las máquinas piensen como un ser humano.

- El procesamiento del lenguaje natural que está inmerso en la inteligencia artificial tiene varias aplicaciones, desde el análisis del texto, como el reconocimiento de sentimientos, la comparación de conceptos, entre otros. Esto permite realizar tareas de una forma automática mejorando el rendimiento en el trabajo que se esté realizando.

RECOMENDACIONES

- Para la búsqueda de información se debe definir que es lo que se quiere buscar, en tal caso plantear algunas preguntas de investigación y a partir de ahí generar una cadena de búsqueda que filtre aquellos documentos que puedan dar respuesta a esas preguntas, además se debe realizar la búsqueda en bases de datos reconocidas, con el fin de obtener la mayor cantidad de información que aporte significativamente a la investigación que se esté realizando.
- Para realizar un proyecto semejante al que se presenta en esta investigación como primer punto se debe investigar que hay en la actualidad, las herramientas y librerías que nos pueden facilitar el trabajo y con esto comenzar a realizar una estructura general de la aplicación identificando toda la arquitectura a usar y en consecuencia centrarse en las funcionalidades más importantes del sistema.
- Siempre es recomendable el uso de alguna metodología que ayude a organizar y planificar las actividades, con el fin de obtener productos de mejor calidad, como en este caso que se usó Scrum como marco de desarrollo. Aunque principalmente se debe investigar y leer bien la documentación de la metodología para poder aplicarla de una forma correcta siguiendo cada uno de los pasos y buenas prácticas, así se podrá aprovechar al máximo la metodología.
- Para la creación de nuevas aplicaciones es recomendable utilizar tecnologías que nos faciliten el desarrollo, como por ejemplo los frameworks que disponen de una variedad de librerías que se pueden utilizar para crear aplicaciones robustas y manejar la información con mayor seguridad.
- En un futuro se podría agregar nuevas funcionalidades al sistema como es la de permitir el análisis de documentos en otros idiomas, y tener un área de administración del sistema en general, donde se pueda visualizar los clientes registrados, mostrar algún cuadro estadístico de los análisis que se realizan, entre otras cosas.
- Para la validación de los sistemas desarrollados, es recomendable usar el modelo planteado por Delone y McLean ya que este modelo permite medir el éxito de los sistemas de información.

REFERENCIAS

- Baumeister, R. F. (2013). Writing a Literature Review. *The Portable Mentor*, 119-132.
https://doi.org/10.1007/978-1-4614-3994-3_8
- Bi, T., Liang, P., Tang, A., & Yang, C. (2018). A systematic mapping study on text analysis techniques in software architecture. *Journal of Systems and Software*, 144, 533-558.
<https://doi.org/10.1016/j.jss.2018.07.055>
- Bustos, C., Morales, M. G., Salcedo, P., & Díaz, A. (2018). Buhos: A web-based systematic literature review management software. *SoftwareX*, 7, 360-372.
<https://doi.org/10.1016/j.softx.2018.10.004>
- Calle, L. (2016). *Metodologías para hacer la revisión de literatura de una investigación*.
[https://www.researchgate.net/publication/301748735_Metodologias_para_hacer_la_r
evision_de_literatura_de_una_investigacion](https://www.researchgate.net/publication/301748735_Metodologias_para_hacer_la_revision_de_literatura_de_una_investigacion)
- University of Leicester. (2009). *Doing a literature review—University of Leicester* [Page].
[https://www2.le.ac.uk/offices/ld/all-resources/writing/writing-resources/literature-
review](https://www2.le.ac.uk/offices/ld/all-resources/writing/writing-resources/literature-review)
- United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. Retrieved from:
[https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%2
0for%20Sustainable%20Development%20web.pdf](https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf)
- Plan Nacional de Desarrollo 2017 – 2021 Toda una Vida – Secretaría Técnica Planifica Ecuador. (2017). Recuperado de [https://www.planificacion.gob.ec/plan-nacional-de-
desarrollo-2017-2021-toda-una-vida/](https://www.planificacion.gob.ec/plan-nacional-de-desarrollo-2017-2021-toda-una-vida/)
- Crowsey, M. J., Ramstad, A. R., Gutierrez, D. H., Paladino, G. W., & White Jr., K. P. (2007). *An evaluation of unstructured text mining software*. 2007 IEEE Systems and Information Engineering Design Symposium, SIEDS.
<https://doi.org/10.1109/SIEDS.2007.4373985>

- DeLone, W. H., & McLean, E. R. (2016). Information Systems Success Measurement. *Foundations and Trends® in Information Systems*, 2(1), 1-116.
<https://doi.org/10.1561/29000000005>
- Fernández, A. M., Bocco, M. G., & Romero, F. P. (2010). *SLR-Tool a tool for performing systematic literature reviews*. 2, 157-166.
- Gonzalez, F. E. (2011). *Dificultades en la realización de trabajos de investigación: Cómo afrontarlas*.
https://www.academia.edu/13097082/Dificultades_en_la_realizaci%C3%B3n_de_trabajos_de_investigaci%C3%B3n_c%C3%B3mo_afrontarlas
- Mohammed, I. S., & Ahmed, M. (2019). *Problems of academic literature review and writing: the way forward*. 16, 11-26.
- Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, 7(2), 139-172. <https://doi.org/10.1080/23270012.2020.1756939>
- Nazir, F., Butt, W. H., Anwar, M. W., & Khan Khattak, M. A. (2017). The applications of natural language processing (NLP) for software requirement engineering—A systematic literature review. *Lecture Notes in Electrical Engineering*, 424, 485-493.
https://doi.org/10.1007/978-981-10-4154-9_56
- Ojo, A. I. (2017). *Validation of the DeLone and McLean Information Systems Success Model*. 23(1), 7. <https://doi.org/10.4258/hir.2017.23.1.60>
- Thangaraj, M., & Sivakami, M. (2018). Text classification techniques: A literature review. *Interdisciplinary Journal of Information, Knowledge, and Management*, 13, 117-135.
<https://doi.org/10.28945/4066>
- Torre, C., D. Sánchez, Blanco, I., & Martín-Bautista, M. J. (2018). Text mining: Techniques, applications, and challenges. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 26(4), 553-582.
<https://doi.org/10.1142/S0218488518500265>

- Webster, J., & Watson, R. T. (2002). *Guest Editorial: Analyzing the Past to Prepare for the Future: Writing a literature Review*. 11.
- Widyassari, A. P., Noersasongko, E., Syukur, A., Affandy, Fanani, A. Z., & Basuki, R. S. (2019). *Literature review of automatic text summarization: Research trend, dataset and method*. 491-496. <https://doi.org/10.1109/ICOIACT46704.2019.8938454>
- Xianguang, Y. (2014). *Text mining software and their applications*. 902-905. <https://doi.org/10.1109/IMCCC.2014.189>
- Yang, D., & Zhang, A. (2019). *Performing literature review using text mining, Part III: Summarizing articles using TextRank*. 3186-3190. <https://doi.org/10.1109/BigData.2018.8622408>
- Zhang, Q., & Segall, R. S. (2010). Review of data, text and web mining software. *Kybernetes*, 39(4), 625-655. <https://doi.org/10.1108/03684921011036835>
- Albaladejo, X. (2008). Qué es SCRUM. *Proyectos Ágiles*. <https://proyectosagiles.org/que-es-scrum/>
- Zanker, M., Nica, M., Profile, S., Peischla, B., Zanker, M., Nica, M., & Schmid, W. (2010). *Constraint-Based Recommendation for Software Project Effort Estimation*.

ANEXOS

Grado de satisfacción del sistema de revisión de literatura.

Estamos realizando una investigación acerca de los beneficios obtenidos y el grado de satisfacción en el uso del Sistema de revisión de literatura, para lo cual nos gustaría saber su opinión acerca de este sistema en 6 dimensiones: Calidad del sistema, Calidad de la información, Calidad del servicio, Uso e intención de uso, Satisfacción del usuario y beneficios obtenidos. La encuesta le tomará pocos minutos y sus respuestas son totalmente anónimas.

Para las respuestas se ha usado una escala de 1 a 5, dónde 1 significa en total desacuerdo y 5 corresponde a una total aceptación.

* Obligatorio

1. Sexo *

- Mujer
- Hombre

2. Edad *

- 18 - 25
- 26 - 35
- 36 - 45
- 46 - 55
- Mayor de 55

Figura 4.1. Encuesta.

Grado de satisfacción del sistema de revisión de literatura.

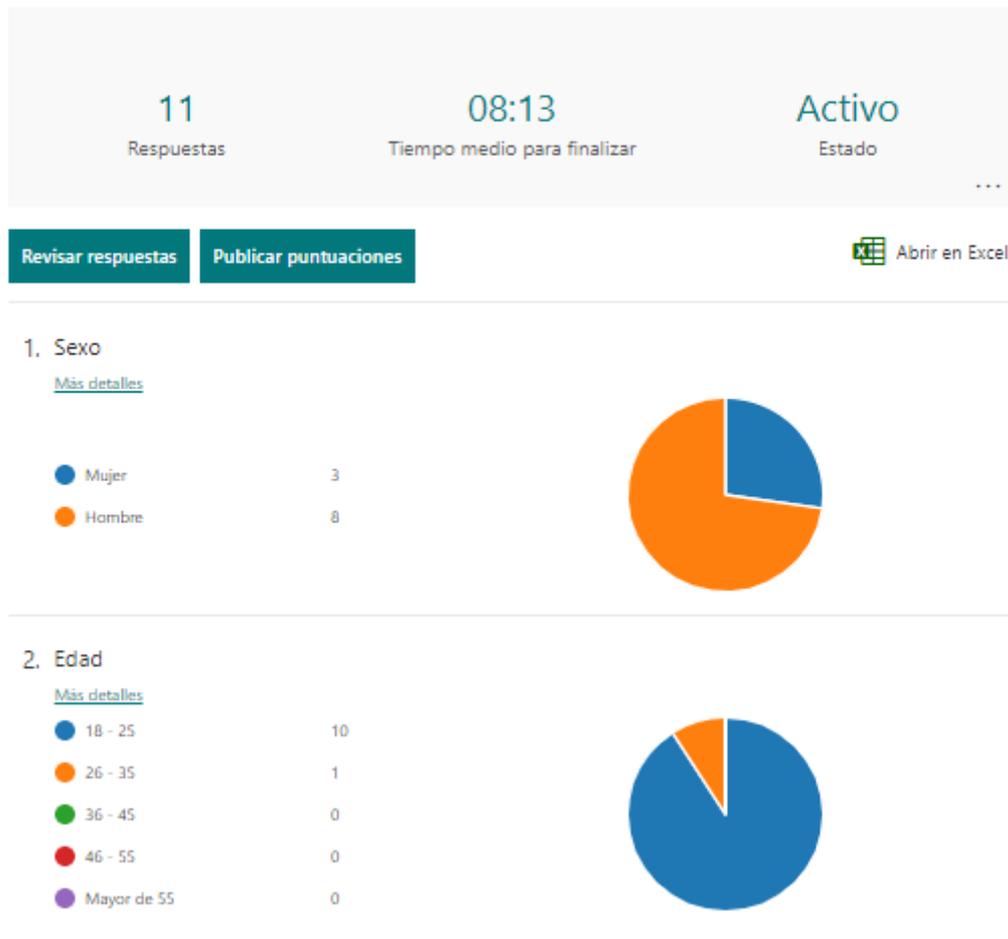


Figura 4.2. Resultados de la encuesta.

2. Ejecutamos el instalador, presionamos en el botón de instalar y esperamos que finalice.

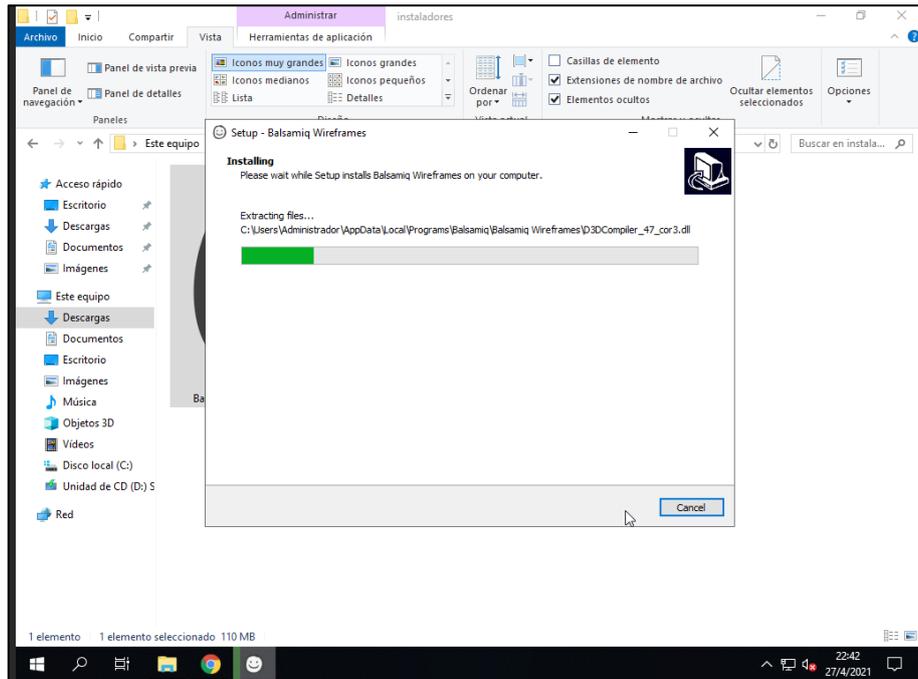


Figura 4.5. Instalación de Balsamic.

3. Una vez finalizada la instalación ya podemos ejecutar el programa.

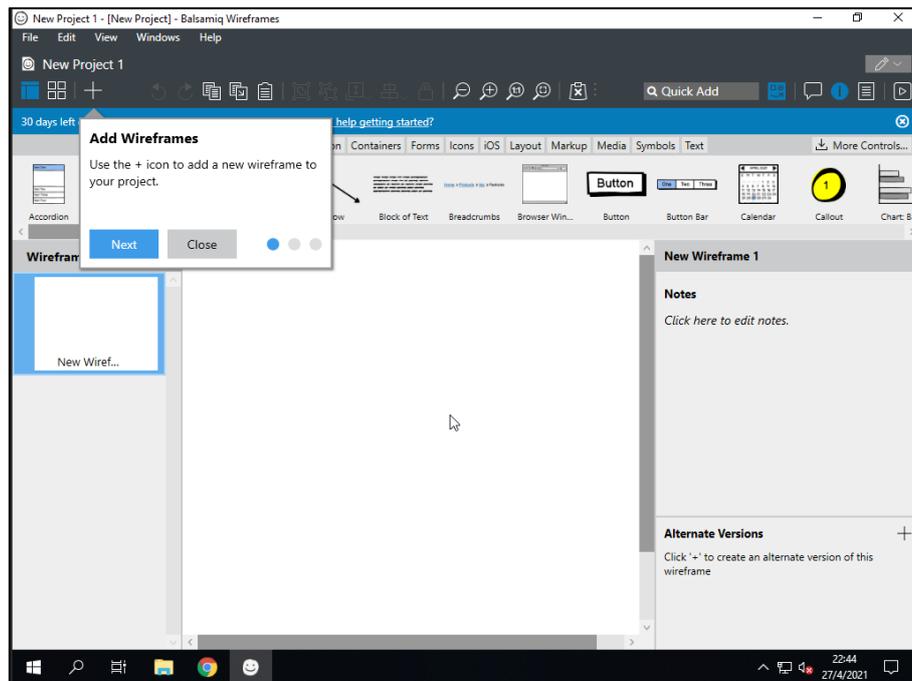


Figura 4.6. Interfaz de Balsamic.

Instalación del lenguaje de programación Python

1. Ingresamos a la página de python y descargamos el instalador.

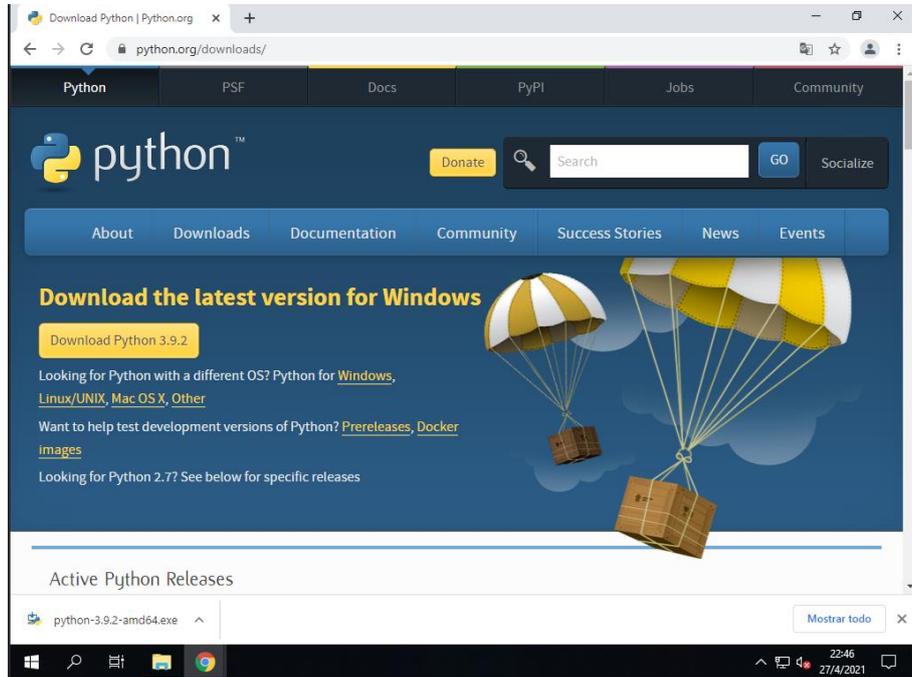


Figura 4.7. Descarga de Python.

2. En esta parte presionamos en instalar ahora.

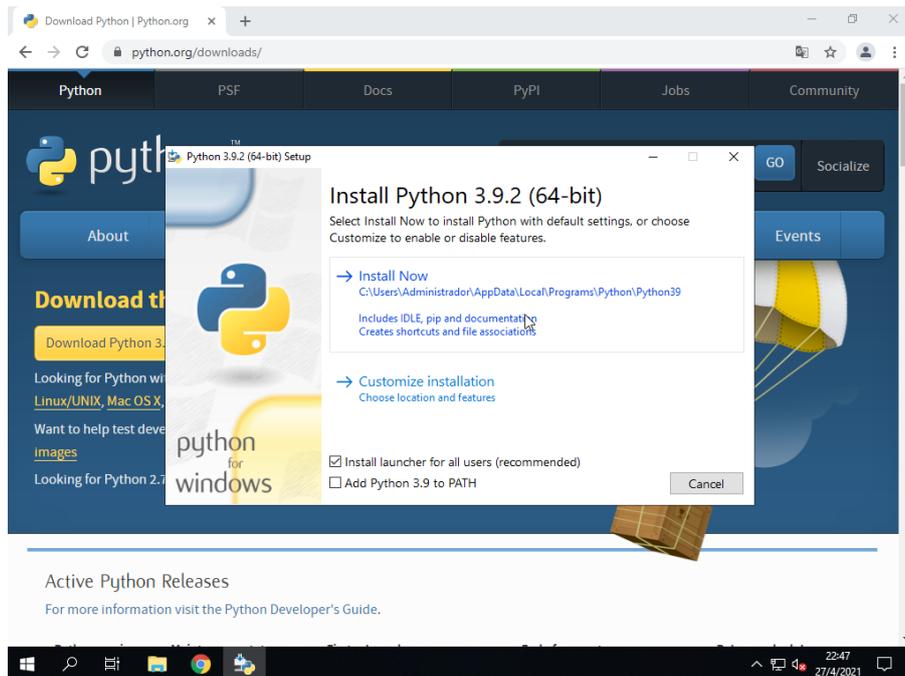


Figura 4.8. Instalación de Python.

3. Finalmente esperamos que finalice la instalación.

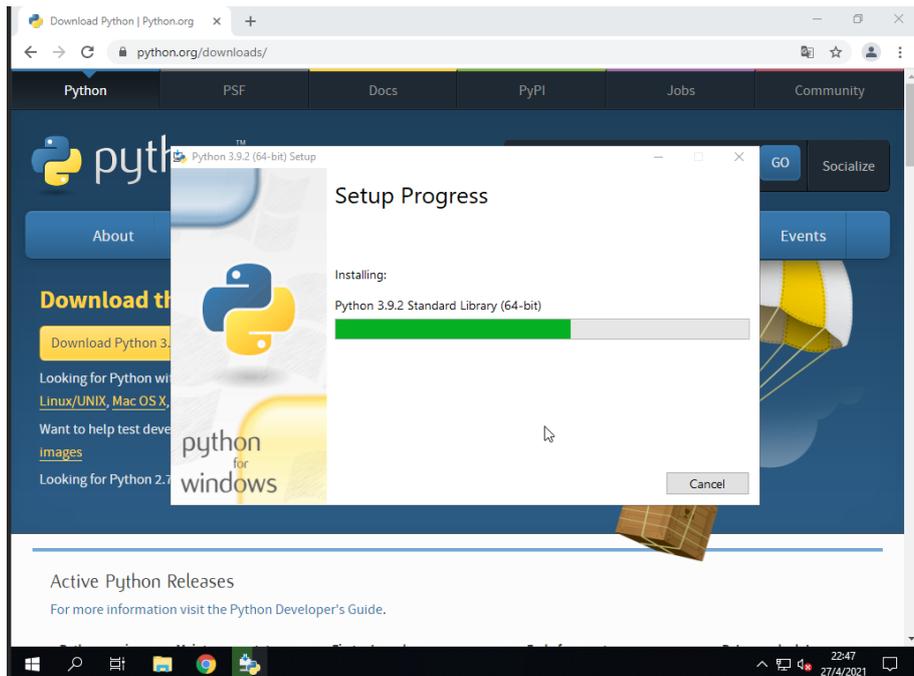


Figura 4.9. Proceso de instalación de Python.

Instalación de Visual Studio Code.

1. Descargamos el instalador de la página oficial de visual studio code.

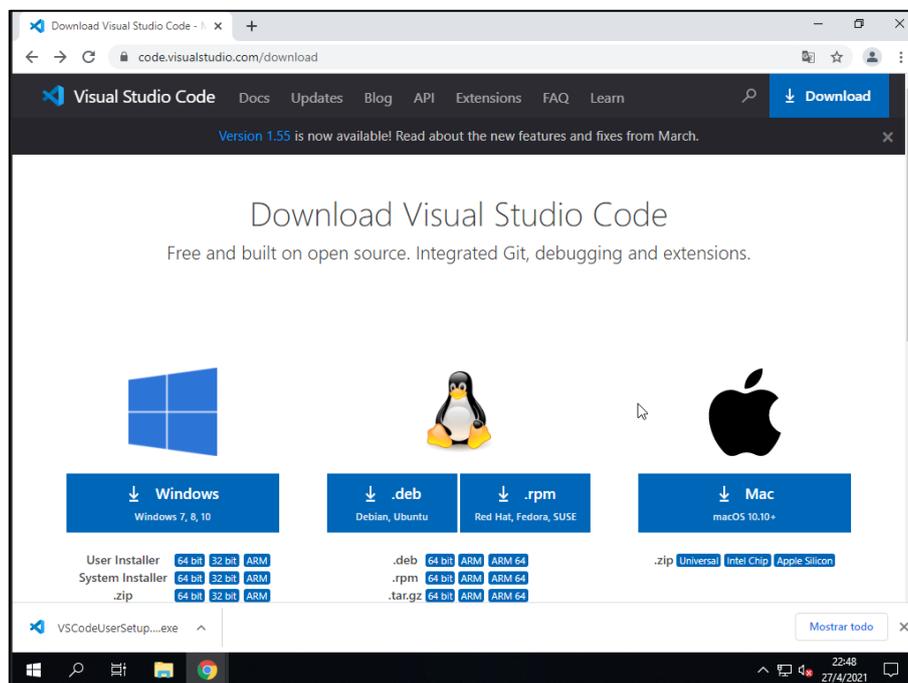


Figura 4.10. Descarga de Visual Studio Code.

2. Aceptamos los términos de la licencia.

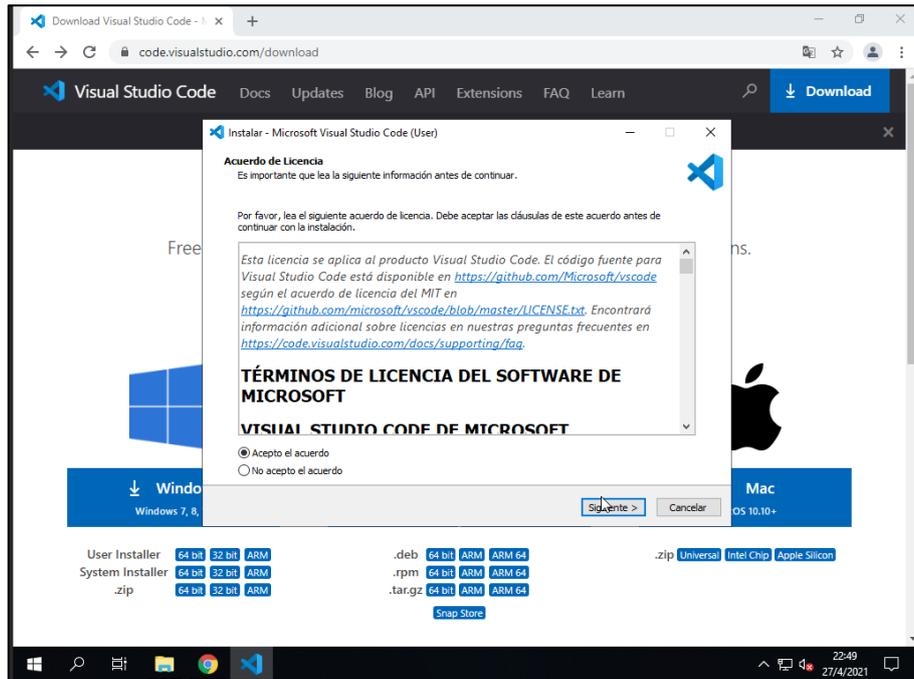


Figura 4.11. Aceptar términos de licencia de Visual Code.

3. Seleccionamos el directorio de instalación.

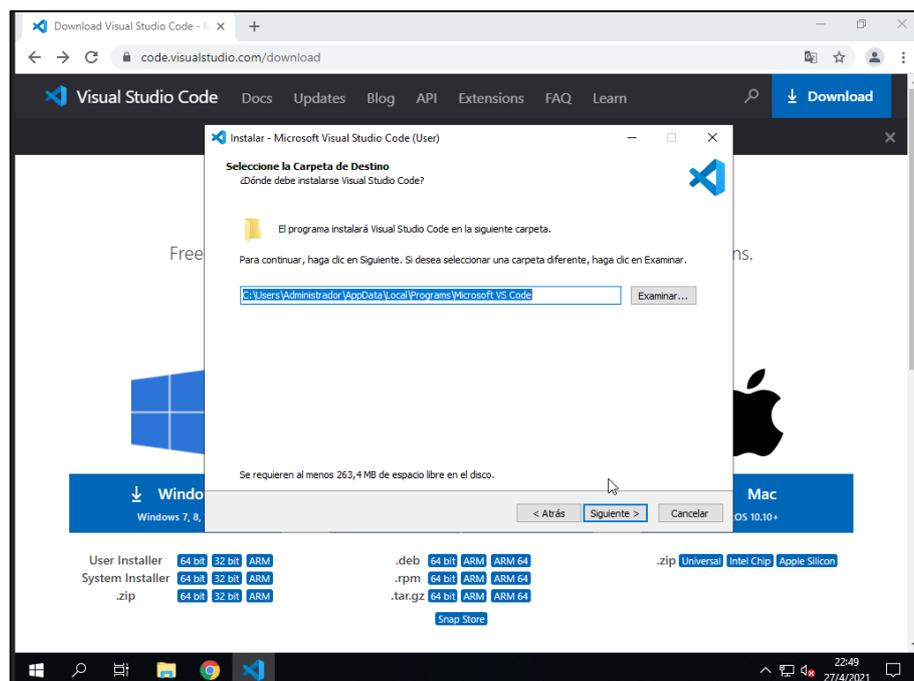


Figura 4.12. Seleccionar carpeta de instalación Visual Code.

4. Esperamos que finalice la instalación.

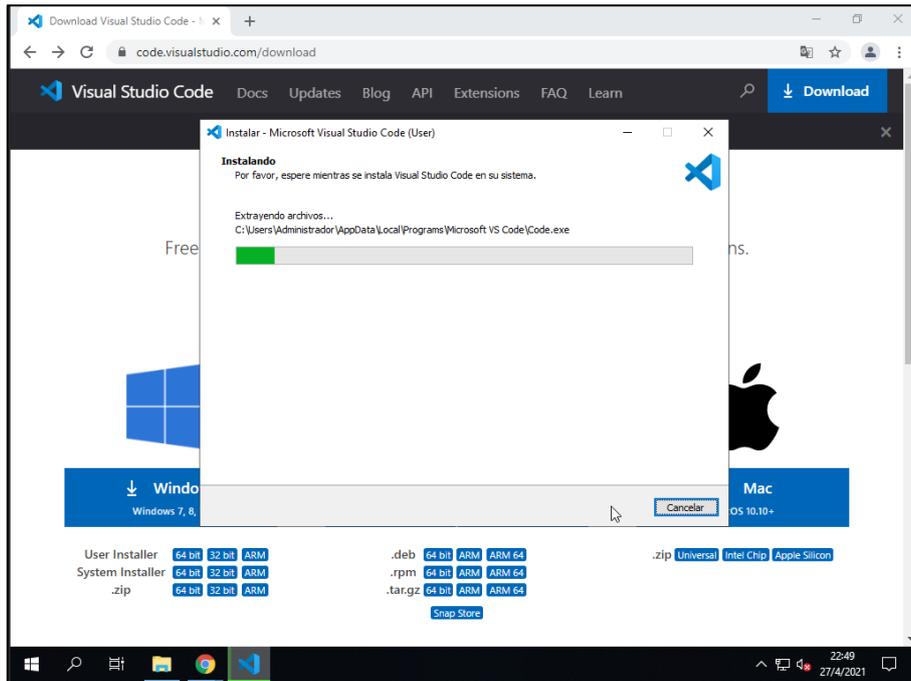


Figura 4.13. Instalación de Visual Studio Code.

5. Al final ya podemos ejecutar el ID de desarrollo e instalar algunas librerías si se prefiere.

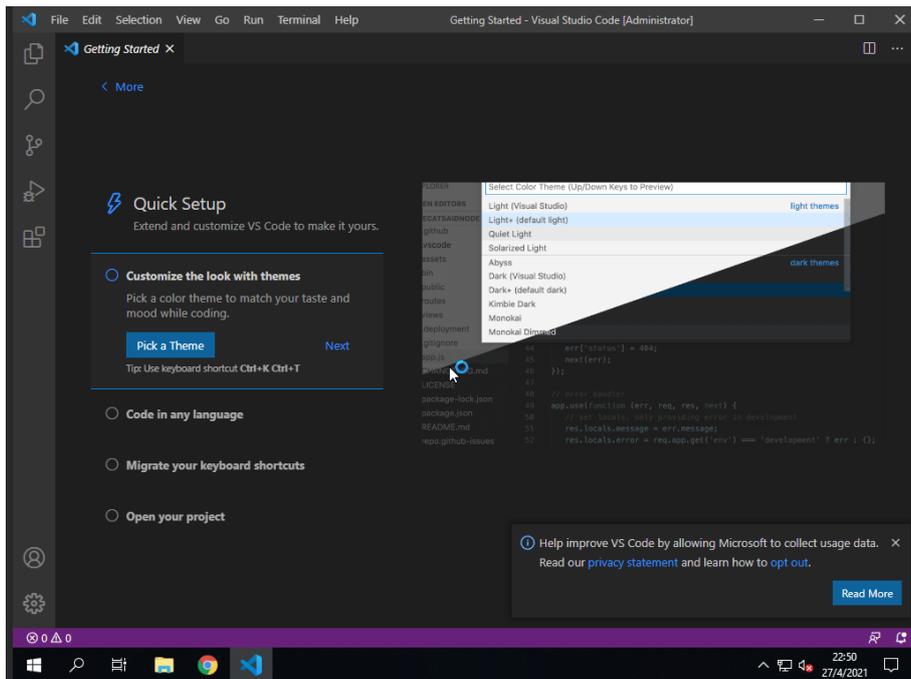


Figura 4.14. Interfaz de Visual Studio Code.

Instalación de MySQL

1. Como es un entorno de desarrollo, por facilidad se instala XAMPP en donde viene incluido MySQL, por lo cual descargamos el instalador de la página de Xampp.

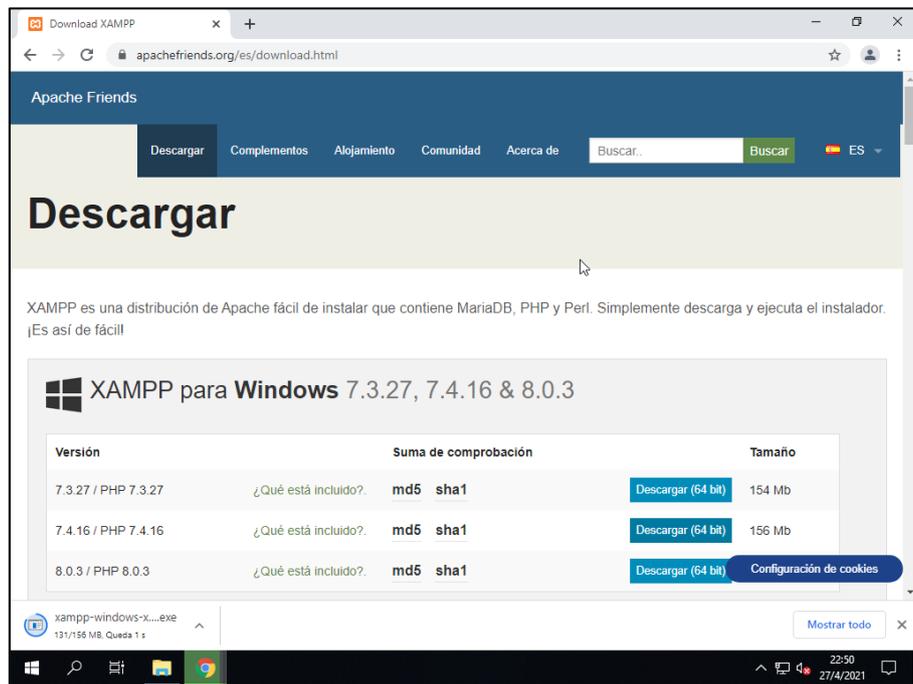


Figura 4.15. Descarga de Xampp.

2. Ejecutamos el instalador y esperamos a que finalice la instalación.

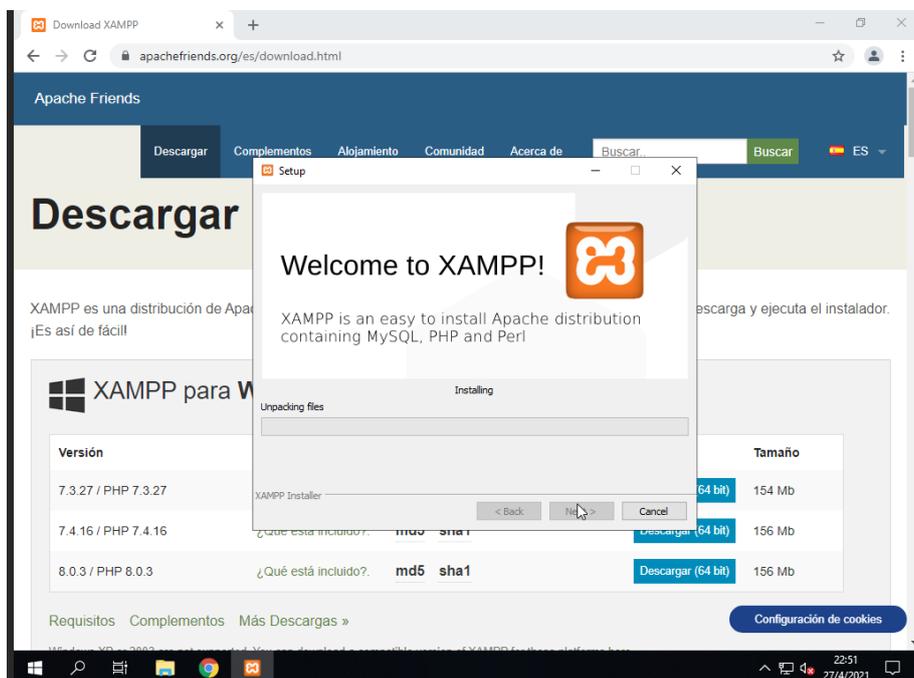


Figura 4.16. Instalación de Xampp.

- Una vez finalizada la instalación, se ejecuta el programa para iniciar los servicios de Apache y MySQL.

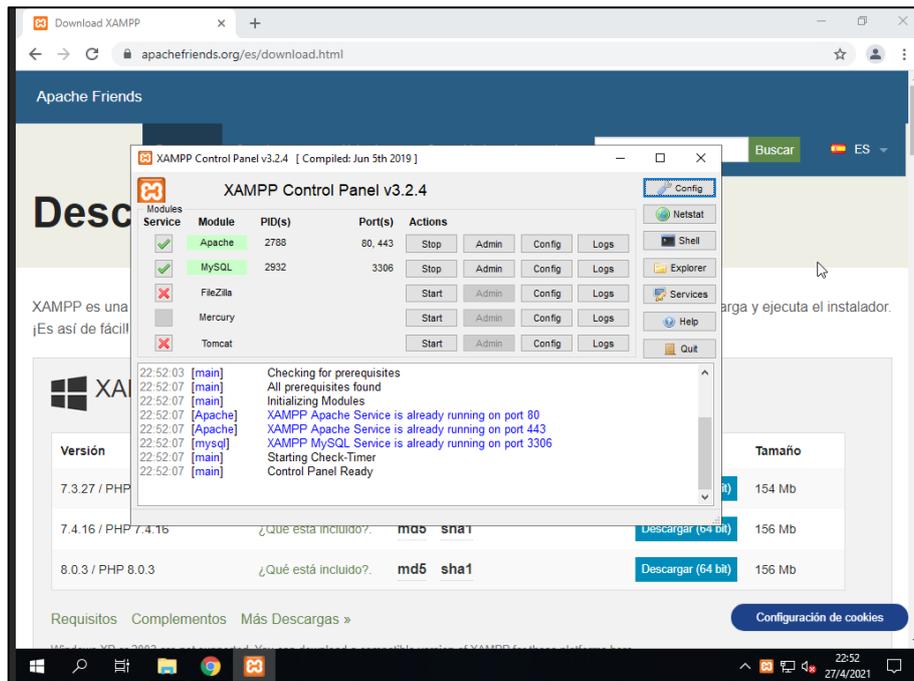


Figura 4.17. Ejecución de servicios en Xampp.

Código fuente:

Diríjase al siguiente enlace:

<https://github.com/AlexanderPatricio/revision-literatura>