



Universidad Austral de Chile

Facultad de Ciencias de la Ingeniería
Escuela de Ingeniería Civil en Informática

CONVERSA CON LOS ANDES: DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE INFORMACIÓN SOBRE CONOCIMIENTOS CIENTÍFICOS Y TÉCNICOS VINCULADOS A LOS ANDES DEL SUR

Proyecto para optar al título de
Ingeniero Civil en Informática

PROFESOR PATROCINANTE:
MATTHIEU VERNIER
DOCTOR EN INFORMÁTICA

PROFESOR INFORMANTE:
MAURICIO RUIZ-TAGLE MOLINA
INGENIERO CIVIL EN INFORMÁTICA
DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

PROFESOR INFORMANTE:
MARIANNA VILLARROEL MANFREDI
INGENIERO CIVIL EN INFORMÁTICA
MAGÍSTER EN EDUCACIÓN MENCIÓN EVALUACIÓN Y
CURRÍCULUM BASADO EN COMPETENCIAS

FERNANDO ALONSO RUIZ LEAL

VALDIVIA – CHILE
2023

AGRADECIMIENTOS

Agradezco a Dios por guiarme en cada paso de este recorrido académico y por brindarme la fortaleza necesaria para superar los desafíos.

A mi familia, pilar fundamental de mi vida, les expreso mi profundo agradecimiento. A mis padres, quienes con amor incondicional y apoyo constante me alentaron a seguir adelante, incluso en los momentos más difíciles. A mi hermana y su familia por ser una fuente constante de inspiración y orgullo.

A los profesores, quienes compartieron su conocimiento y experiencia, les estoy agradecido por su dedicación y por inspirarme a alcanzar metas más altas. Sus enseñanzas han dejado una huella significativa en mi formación académica y personal.

Agradezco a aquellos amigos que conocí durante mi carrera, quienes se convirtieron en parte esencial de mi red de apoyo.

Finalmente, agradezco a todos aquellos que, de alguna manera, contribuyeron a mi crecimiento y desarrollo durante estos años de estudio. Cada experiencia, ya sea desafiante o gratificante, ha sido una oportunidad para aprender y crecer.

ÍNDICE

1. INTRODUCCIÓN.....	1
1.1 Laboratorio Natural Andes del Sur de Chile	1
1.2 Estructura y avances del Laboratorio Natural Andes del Sur.....	2
1.3 Problemática del Laboratorio Natural Andes del Sur.....	2
1.4 “Conversa con Los Andes”: Hacia la creación de una biblioteca.....	2
1.5 Objetivo general y específicos	3
1.6 Colaboración en un proyecto conjunto.....	4
2. REVISIÓN SISTEMÁTICA	5
2.1 Objetivo y pregunta de revisión	5
2.2 Método de revisión... ..	5
2.3 Resultados... ..	7
2.4 Principales hallazgos y factores claves	10
3. DISEÑO DEL SISTEMA “CONVERSA CON LOS ANDES”	12
3.1 Diagrama de casos de uso	12
3.2 Descripción de actores.	13
3.3 Casos de uso	13
3.3.1 Añadir nuevos documentos a partir de un archivo CSV	13
3.3.2 Buscar documentos a partir de criterios	15
3.3.3 Generar nuevo CSV con datos de cybertesis.....	16
3.3.4 Añadir un nuevo documento manualmente.....	17
3.3.5 Extraer indicadores a partir de los datos.....	19
3.4 Diseño de la base de datos	20
3.5 Diagrama de componentes	25
3.6 Diagrama de clases	26
3.6 Diagrama de secuencia	28
4. IMPLEMENTACIÓN.....	32
4.1 Tecnologías utilizadas.....	32
4.1.1 Python	32
4.1.2 MariaDB.....	32
4.1.3 Discord.py.....	32
4.1.4 BeautifulSoup.....	32
4.1.5 Asyncio.....	32
4.1.6 GitHub.....	33
4.1.7 spacy.....	33
4.2 ¿Por qué utilizar Discord?.....	33
4.3 Implementación base de datos	34
4.4 Implementación de funcionalidades y desarrollo.....	35
4.4.1 añadir nuevos documentos a partir de un archivo CSV	36
4.4.2 Añadir un nuevo documento manualmente	39
4.4.3 Obtener indicadores a partir de los datos	40
4.4.4 Obtener documentos a partir de parámetros	44
4.4.5 Generar un nuevo CSV con datos de cybertesis.....	49
4.5 Flujo de trabajo durante la implementación	54

5. REUNIÓN CON CONAF.....	56
6. ENCUESTA AL ACTOR DEL LABORATORIO.....	58
6.1 Diseño de la encuesta	58
6.2 Resultados de la encuesta por funcionalidad	58
6.3 Acciones post encuesta	64
7. CONCLUSION.....	66
8. REFERENCIAS	67

ÍNDICE DE TABLAS

TABLA	PÁGINA
1 Agrupación palabras claves	6
2 Revisión de información del primer artículo	7
3 Revisión de información del segundo artículo	8
4 Revisión de información del tercer artículo	9
5 Revisión de información del cuarto artículo	10
6 Descripción de actores	13
7 Caso de uso añadir nuevos documentos a partir de un archivo CSV	14
8 Curso normal de los eventos añadir nuevos documentos a partir de un archivo CSV...	14
9 Curso alternativo añadir nuevos documentos a partir de un archivo CSV.....	15
10 Caso de uso buscar documentos a partir de criterios.	15
11 Curso normal de los eventos buscar documentos a partir de criterios..	16
12 Curso alternativo buscar documentos a partir de criterios.	16
13 Generar nuevo CSV con datos de cybertesis.	16
14 Curso normal de los eventos generar nuevo CSV con datos de cybertesis	17
15 Caso de uso añadir un nuevo documento manualmente.	17
16 Curso normal de los eventos añadir un nuevo documento manualmente	18
17 Curso alternativo añadir un nuevo documento manualmente.....	18
18 Caso de uso extraer indicadores a partir de los datos.....	19
19 Curso normal de los eventos caso de uso extraer indicadores a partir de los datos.....	19
20 Diccionario de datos tabla Autor	23
21 Diccionario de datos tabla Commune	23
22 Diccionario de datos tabla Document	24
23 Diccionario de datos tabla Category.....	24
24 Diccionario de datos tabla Organization	25

ÍNDICE DE FIGURAS

FIGURA	PÁGINA
1 Diagrama de casos de uso	12
2 Diagrama entidad relación	21
3 Diagrama relacional	22
4 Diagrama de componentes.....	26
5 Diagrama de clases	27
6 Diagrama de secuencia agregar nuevos documentos a partir de CSV.....	28
7 Diagrama de secuencia agregar documento manualmente	29
8 Diagrama de secuencia generar un nuevo CSV con datos de cybertesis.....	30
9 Diagrama de secuencia buscar documentos a partir de criterios.....	30
10 Diagrama de secuencia extraer indicadores a partir de los datos	31
11 Código para crear tabla Author	34
12 Código para crear tabla Document	34
13 Tablas implementadas.....	35
14 Formato esperado de un archivo CSV parte 1.....	36
15 Formato esperado de un archivo CSV parte 2.....	36
16 Formato esperado de un archivo CSV parte 3.....	37
17 Ejecución del comando !addDocument.....	38
18 Bot respondiendo con los documentos que no pudieron ser ingresados.....	38
19 Bot respondiendo cuando se adjunta un archivo CSV invalido.....	39
20 Ejecución comando !addManualDocument	40
21 Bot mostrando el numero de documentos en la base de datos.....	41
22 Bot mostrando los autores más prolíficos.....	41
23 Bot mostrando la distribución de los documentos por laboratorio temático.....	42
24 Bot mostrando las organizaciones más activas.....	43
25 Bot mostrando la evolución de la producción de documentos por años.....	43
26 Bot mostrando resultados de ingresar volcán como palabra clave.....	45
27 Bot mostrando los documentos que pertenecen al laboratorio temático 1.....	45
28 Bot mostrando los documentos que fueron publicanos el año 2005.....	46
29 Bot mostrando documentos en un rango de años.....	46
30. Bot mostrando documentos pertenecientes a una categoría.....	47
31 Bot mostrando documentos con una mezcla de parámetros.....	48
32 Bot mostrando un error.....	48
33 Página principal de cybertesis, con el menú despegable.....	50
34 Lista de tesis relacionadas a una escuela.....	51
35 Página con información de tesis especifica.....	51
36 String de búsqueda.....	53
37 Palabras claves adaptadas de un string de búsqueda.....	53
38 Bot mostrando el resultado de el comando !getThesis.....	54
39 Reunión llevada a cabo en CONAF.....	56
40 Twitter de CONAF informando de la reunión	57
41 Resultado de encuesta comando !search.....	59

42	Resultado de encuesta comando !search parte 2.....	60
43	Resultado de encuesta comando !analysis.....	60
44	Resultado de encuesta comando !analysis parte 2.....	61
45	Resultado de encuesta comando !addDocument	61
46	Resultado de encuesta comando !addDocument parte 2.....	62
47	Resultado de encuesta comando !addManualDocument.....	62
48	Resultado de encuesta comando !getThesis.....	63
49	Resultado de encuesta comando !getThesis parte 2.....	63
50	Cambio de título.....	64
51	Modificaciones comando search.....	65

RESUMEN

Conversa con Los Andes aborda la necesidad del Laboratorio Natural Andes del Sur de Chile de gestionar de manera eficiente documentos científicos y técnicos relacionados con el territorio en el que se ubica el laboratorio. A través de un bot en Discord, la plata-forma simplifica el ingreso de estos documentos a una base de datos mediante comandos específicos que han sido implementados.

Una vez almacenados, se puede acceder a la información mediante funcionalidades como la búsqueda avanzada y la extracción de indicadores de la base de datos. Estas características, desarrolladas a través de un proceso de diseño e implementación, buscan satisfacer las necesidades del laboratorio al ofrecer una gestión eficiente de la información.

Adicionalmente, el bot tiene la capacidad de identificar nuevos documentos de interés para el laboratorio mediante el scraping de datos desde la página de tesis de la Universidad Austral.

Estas funcionalidades, integradas en la plataforma, contribuyen a crear una herramienta integral que mejora el acceso a conocimientos científicos y técnicos sobre Los Andes del Sur.

ABSTRACT

"Conversa con Los Andes" addresses the need of the Southern Chile Andes Natural Laboratory to efficiently manage scientific and technical documents related to the territory where the laboratory is located. Through a Discord bot, the platform simplifies the input of these documents into a database using specific commands that have been implemented.

Once stored, information can be accessed through features such as advanced search and extraction of indicators from the database. These features, developed through a design and implementation process, aim to meet the laboratory's needs by providing efficient information management.

Additionally, the bot has the ability to identify new documents of interest to the laboratory by scraping data from the thesis page of the Austral University.

These features, integrated into the platform, contribute to creating a comprehensive tool that enhances access to scientific and technical knowledge about the Southern Andes.

1. INTRODUCCIÓN

1.1 Laboratorio Natural Andes del Sur de Chile

Chile es un país que cuenta con una gran cantidad de montañas, las cuales representan el 64% de su territorio. Estas formaciones geográficas son especialmente vulnerables a los efectos del cambio climático y otros procesos de transformación global, por lo que resultan de gran interés para la investigación científica (Flores, 2012). “La montaña tiene muchas singularidades a escala nacional y planetaria. Una de ellas es que estamos insertos en socio-ecosistemas de volcanes activos, bosques, glaciares, cuencas hidrográficas y asentamientos humanos. Es necesario pensar la montaña de manera integrada” señala la Dra. Carla Marchant, académica a cargo del Laboratorio Natural Andes del Sur (Laboratorio Natural Andes Del Sur De Chile: El Desafío De Recopilar Información Científica En Montañas únicas En El Mundo, 2022).

Es en este contexto que surge el proyecto “Laboratorio Natural Andes del Sur de Chile”, una iniciativa financiada por ANID en el concurso 2021 “Nodos para el Desarrollo en Investigación de Laboratorios Naturales en Chile”, y ejecutada por la Universidad Austral de Chile. Un Laboratorio Natural se describe como un territorio con características naturales únicas que entregan ventajas comparativas para el desarrollo de ciencia y tecnología de alto nivel e impacto planetario”. La convocatoria de ANID buscó identificar y desarrollar laboratorios naturales en cinco macro-categorías: Desierto, montaña, cielos, antártico, océanos, subantártico (Bertrán, 2021). El Laboratorio Natural Andes del Sur, constituye una plataforma territorial de montaña con incidencia en un área donde habitan 250.000 personas en 40 poblados y espacios rurales de 28 comunas de las regiones de La Araucanía, Los Ríos y Los Lagos.

El objetivo del Laboratorio Natural Andes del Sur consiste en fortalecer y promover el conocimiento, la investigación de frontera, valoración intercultural y conservación de los territorios de montaña del sur de Chile a través de una aproximación socio-ecológica, inter y transdisciplinaria, que permita generar ciencia de punta para el país y el mundo, y con impacto en el desarrollo local (Proyecto – Andes Del Sur De Chile, 2022). “Otro de los principales desafíos que los Laboratorios Naturales deben abordar es lograr transformar la manera en que hoy se hace ciencia en Chile, avanzando hacia modelos participativos con un enfoque transdisciplinar y de democratización del conocimiento y aportando conocimiento para el desarrollo de política pública nacional y regional”, mencionó la directora del proyecto (Bertrán, 2021).

1.2. Estructura y avances del Laboratorio Natural Andes del Sur

El Laboratorio Natural Andes del Sur estructura su trabajo en tres laboratorios temáticos: 1) evolución y herencia geológica, 2) cambio global: interacciones bosque – suelo – recursos hídricos, 3) modos de vida y habitar de montaña. A lo largo del año 2022, se dedicó a la identificación y análisis de toda la producción científica indexada presente en el territorio donde se ubica el laboratorio natural.

Mediante este análisis, se recopiló un total de cerca de ochocientos artículos científicos indexados en Scopus y Web of Science, publicados entre 1982 y 2021. Estos resultados constituyen una línea base que facilita la identificación de los actores principales, instituciones y redes de investigación a nivel nacional e internacional vinculadas a las temáticas abordadas por el laboratorio.

1.3. Problemática del Laboratorio Natural Andes del Sur

En su compromiso por fomentar la investigación interdisciplinaria y la preservación de los territorios montañosos en el sur de Chile, el Laboratorio Natural Andes del Sur se enfrenta a un desafío crucial: la carencia de un sistema de información que centralice y gestione la información con la que cuenta, especialmente aquella contenida en el archivo CSV con el que cuentan. Este archivo es vital, ya que alberga la identificación de todos los documentos científicos de interés para el laboratorio, abarcando sus laboratorios temáticos.

La ausencia de un repositorio central de conocimientos impide al laboratorio aprovechar su capacidad para fortalecer y promover el conocimiento del territorio de Los Andes. Asimismo, es relevante explorar e integrar otras fuentes de conocimiento científico y técnico relacionadas con el territorio. Ampliar la perspectiva más allá de las publicaciones indexadas en revistas permitiría al laboratorio enriquecer aún más su capacidad de investigación.

1.4. “Conversa con Los Andes”: Hacia la creación de una biblioteca

El proyecto tiene como objetivo desarrollar la plataforma "Conversa con Los Andes" para el Laboratorio Natural Andes del Sur. Esta plataforma facilitará a los usuarios del laboratorio el acceso a conocimientos científicos y técnicos sobre el territorio de los Andes del Sur, actuando como una biblioteca virtual especializada. La plataforma contendrá una base de datos que albergará información de documentos científicos y técnicos exclusivamente relacionados con el territorio donde se encuentra el laboratorio natural. La obtención de información de los documentos se llevará a cabo mediante el archivo en formato CSV que posee el laboratorio. Este archivo incluye en una columna el enlace de descarga de cada documento, permitiendo la descarga automática de los mismos. Además,

se busca incorporar las tesis de la Universidad Austral de Chile que aborden aspectos del territorio para ampliar la información disponible.

Se almacenará toda la información de los documentos identificados por el laboratorio, incluyendo títulos, autores y otros detalles, así como su contenido, en una base de datos. Este proceso tiene como objetivo facilitar la gestión de los datos relevantes sobre el territorio de los Andes.

La plataforma, presentada a través de un canal de Discord, contará con un bot integrado que ofrecerá diversas funcionalidades para interactuar con la información almacenada. Además, se explorará la inclusión de un chat inteligente capaz de interpretar preguntas en lenguaje natural y proporcionar respuestas basadas en la información de la base de datos. De esta manera, la plataforma se consolidará como una herramienta integral de estilo biblioteca virtual para el Laboratorio Natural Andes del Sur.

1.5. Objetivo general y específicos

El objetivo general del proyecto es: Generar una plataforma que permita mejorar el acceso a conocimientos científicos y técnicos sobre el territorio de Los Andes del Sur.

Los objetivos específicos que guiarán el desarrollo de esta plataforma son:

1. Diseñar e implementar una base de datos que almacene información científica y técnica sobre el territorio de los Andes.
2. Desarrollar algoritmos de *web scraping* para recopilar de manera automática documentos científicos y técnicos relevantes del territorio de Los Andes a partir del análisis bibliométrico existente y/o nuevas fuentes de documentos.
3. Aplicar técnicas de Tratamiento Automático del Lenguaje usando vectorización para analizar y procesar los documentos recopilados mediante una base de datos vectorial.
4. Implementar una interfaz de usuario sencilla y un bot que facilite la búsqueda de documentos científicos y técnicos sobre el territorio de Los Andes.
5. Integrar al bot de Discord funcionalidades avanzadas.
6. Desarrollar una funcionalidad experimental de integración de un chat inteligente que tenga la capacidad de interactuar con los usuarios para responder preguntas basadas en la información disponible.
7. Evaluar la usabilidad de la plataforma "Conversa con Los Andes" con usuario representante del laboratorio.
8. Implementar en la plataforma "Conversa con Los Andes" una funcionalidad que permita realizar un análisis bibliométrico de los documentos de la base de datos.

1.6 Colaboración en un proyecto conjunto

La ejecución del proyecto "Conversa con Los Andes" implica la colaboración de dos participantes, cada uno enfocado en aspectos específicos para alcanzar los objetivos establecidos. En este documento, se detallarán los objetivos específicos asociados a mi contribución en el proyecto, los cuales abarcan:

1. Diseñar e implementar una base de datos que almacene información científica y técnica sobre el territorio de los Andes.
2. Desarrollar algoritmos de *web scraping* para recopilar de manera automática documentos científicos y técnicos relevantes del territorio de Los Andes a partir del análisis bibliométrico existente y/o nuevas fuentes de documentos.
3. Integrar al bot de discord funcionalidades avanzadas
4. La implementación de una funcionalidad de análisis bibliométrico.

2. REVISION SISTEMÁTICA

El propósito central de la plataforma Conversa con los Andes es desarrollar una plataforma de gestión del conocimiento de un territorio en específico. Para lograr esto, es fundamental adquirir un conocimiento del dominio en el que la plataforma se insertará.

Con el fin de explorar y comprender la gestión del conocimiento y/o inteligencia territorial, así como examinar las posibles soluciones tecnológicas utilizadas para su manejo en diversos contextos, se llevará a cabo una revisión sistemática.

2.1 Objetivo y pregunta de revisión

El objetivo principal de esta revisión sistemática es investigar y comprender la gestión del conocimiento y la inteligencia en el ámbito territorial, así como examinar las diferentes soluciones tecnológicas utilizadas para su manejo. Con este fin, resulta fundamental adquirir un conocimiento claro de las definiciones asociadas a estos conceptos. Asimismo, se busca obtener una visión general de las prácticas actuales y las opciones disponibles para la gestión del conocimiento en diversos contextos. Para abordar este objetivo, se plantea la siguiente pregunta de investigación:

¿Qué es la gestión del conocimiento y/o inteligencia territorial y qué plataformas o herramientas existen para llevar a cabo esta gestión?

2.2 Método de revisión

Metodología

En base a la pregunta de investigación, lo primero que se hizo fue crear el *string* de búsqueda que se tuvo que aplicar para obtener los artículos. Luego de realizar esta búsqueda, se seleccionaron los artículos que contenían información relevante bajo ciertos criterios de selección expuestos más adelante. Finalmente se analizaron y resumieron los resultados. A continuación, se detalla el proceso del desarrollo de la revisión sistemática.

Fuentes y estrategias de búsqueda

La fuente para realizar la búsqueda fue la librería digital de la ISI Web Of Science (<http://www.isiknowledge.com>).

Se decidió realizar la búsqueda en inglés ya que se obtenía un mayor número de resultados. Las palabras claves que se identificaron, a partir de la pregunta de investigación, fueron las siguientes: *Territorial knowledge, Territorial information, Territorial data, Territorial intelligence, Management, Practice, Data management, Territories*.

Una vez definidas las palabras claves, éstas fueron clasificadas en grupos como se puede apreciar en la Tabla 1.

Tabla 1. Agrupación palabras claves.

Grupo 1: Territorial	Grupo 2: Management	Grupo 3: Data Management	Grupo 4: Territories
territorial knowledge	management	data management	territories
territorial information	practice		
territorial data			
territorial intelligence			

En la columna Grupo 1, se encuentran los términos relacionados con el conocimiento, la información, los datos y la inteligencia territorial. Estos conceptos están estrechamente relacionados y se refieren a aspectos específicos relacionados con los territorios.

En la columna Grupo 2, se encuentran los términos *management* y *practice*. Estos términos pueden tener un significado más general y no están necesariamente limitados al ámbito territorial, pero se incluyen en la agrupación para considerar cualquier contexto relacionado con la gestión y la práctica en relación con los territorios.

En la columna Grupo 3, se encuentra el término *data management*, que se refiere específicamente a la gestión de datos. Este término puede estar relacionado tanto con los territorios como con otros ámbitos.

En la columna Grupo 4, se encuentra el término *territories*, que se refiere a los territorios en general. Este término puede ser relevante para comprender el contexto general de la búsqueda y puede estar relacionado con otros conceptos presentes en las otras categorías.

Solo se consideraron artículos publicados desde el año 2000 en adelante y la búsqueda se hizo sólo en el resumen de los artículos.

Cadena de búsqueda

La cadena de búsqueda quedó finalmente definida:

AB= (((("territorial knowledge" OR "territorial information" OR "territorial data" OR "territorial intelligence")
AND
("management" OR "practice"))) OR ("data management" AND "territories"))

Criterios de selección

A partir de los resultados de la búsqueda, se seleccionaron aquellos artículos que hablaran sobre la gestión del conocimiento en los territorios o sobre la inteligencia territorial o sistemas que gestionan algún tipo de conocimiento territorial. Todos los artículos que no trataran sobre estas temáticas no se seleccionaron.

Extracción de información

Se definió una pauta de extracción de información con la finalidad de estructurar la información encontrada sobre la gestión del conocimiento en los territorios, así como de aplicaciones usadas para realizar esta gestión. La pauta de extracción de información por artículo fue la siguiente:

- Descripción/para qué sirve la gestión del conocimiento (inteligencia) territorial
- Aplicaciones propuestas/descritas: Los modelos teóricos o prácticos relacionados con la gestión del conocimiento y/o inteligencia territorial que se presentan en el artículo.
- Herramientas y funcionalidades: Las herramientas y funcionalidades específicas de las aplicaciones mencionadas en el artículo que se utilizan para la gestión del conocimiento en los territorios.
- Aspectos claves

Estudios incluidos y excluidos

Al realizar la búsqueda, con la cadena definida, se obtuvieron un total de 52 resultados. Se leyó el resumen de todos ellos para así poder aplicar los criterios de selección definidos. Finalmente, 48 fueron descartados quedando 4 artículos para resumir y extraer la información relevante para la investigación.

2.3. Resultados

La información relevante, de cada artículo seleccionado, se extrajo, se analizó y posteriormente se detalló en la Tabla 2, Tabla 3, Tabla 4 y Tabla 5.

Tabla 2. Revisión de información del primer artículo.

Artículo
Garcia-Madurga, M.-A., Grillo-Mendez, A.-J., & Esteban-Navarro, M.-A. (2020).
Título
Territorial Intelligence, a Collective Challenge for Sustainable Development: A Scoping Review
Descripción/para qué sirve la gestión del conocimiento (inteligencia) territorial
Obtener, analizar y valorar información y conocimientos sobre el entorno destinados a promover el desarrollo de un territorio, prestando especial atención a la identificación de oportunidades y evaluación de riesgos. Es una forma de inteligencia colectiva desarrollada sobre y alrededor de un territorio para pensar y actuar sobre él.
Aplicaciones propuestas/descritas

No se presentan aplicaciones.
Herramientas y funcionalidades
No se mencionan herramientas.
Aspectos claves
Encargada de la gestión coordinada de la información y el conocimiento generado por los actores políticos, sociales y económicos para evitar su pérdida y lograr un aprovechamiento y adecuación óptimos,

Tabla 3. Revisión de información del segundo artículo.

Artículo
Foronda-Robles, C., & Galindo-Perez-de-Azpillaga, L. (2021).
Título
Territorial intelligence in rural areas: The digitization of non-profit associations through social media
Descripción/para qué sirve la gestión del conocimiento (inteligencia) territorial
Oportunidad para actores que quieran anticipar e incorporar procesos adecuados a un modelo de desarrollo que pueda redefinir las políticas públicas y lograr una mayor cohesión territorial. Sirve para poder entender las estructuras territoriales y la dinámica que existe entre ellas. Ha evolucionado de una visión estrictamente económica a una enfocada en el desarrollo sostenible y la comunidad, y se ha introducido un paradigma de colaboración en lugar de competencia
Aplicaciones propuestas/descritas
No se presentan aplicaciones.
Herramientas y funcionalidades
No se mencionan herramientas.
Aspectos claves
Planificación integral y el diseño de instrumentos básicos para la organización y estructura del territorio. Se destaca la importancia de la colaboración y la participación de la comunidad en la gestión del conocimiento territorial y la promoción del desarrollo sostenible.

Tabla 4. Revisión de información del tercer artículo.

Artículo
Corchado, J. M., Chamoso, P., Hernández, G., Gutierrez, A. S. R., Camacho, A. R., González-Briones, A., Pinto-Santos, F., Goyenechea, E., Garcia-Retuerta, D., Alonso-Miguel, M., Hernandez, B. B., Villaverde, D. V., Sanchez-Verdejo, M., Plaza-Martínez, P., López-Pérez, M., Manzano-García, S., Alonso, R. S., Casado-Vara, R., Tejedor, J. P. (2021)
Título
Deepint.net: A Rapid Deployment Platform for Smart Territories
Descripción/para qué sirve la gestión del conocimiento (inteligencia) territorial.
Ciudad inteligente, es un entorno que utiliza tecnologías innovadoras para hacer que las redes y los servicios sean más flexibles, efectivos y sostenibles con el uso de tecnologías de la información, digitales y de telecomunicaciones, mejorando las operaciones de la ciudad en beneficio de sus ciudadanos.
Aplicaciones propuestas/descritas.
Deepint.net, plataforma para gestionar e interpretar datos de forma eficiente y sencilla. Gestiona los datos que se generan en las ciudades inteligentes
Herramientas y funcionalidades.
Permite aplicar técnicas de IA a grandes volúmenes de datos de manera rápida. Integración de datos multi fuentes Mecanismos de tratamiento de datos. Detección automática del tipo de datos para mejorar análisis.
Aspectos claves.
Hoy en día, los territorios son los mayores productores de datos y todos los principales sectores pueden extraer conocimientos y beneficiarse considerablemente de estos

Tabla 5. Revisión de información del cuarto artículo.

Artículo
IVarona-Taborda, M.-A., Mosquera-Ramirez, J.-C., Medina-Moreno, C.-A., Lemus-Munoz, D.-F., Munoz-Hernandez, C.-J., & Arias-Iragorri, C.-G. (2021).
Título
Business Intelligence for the Programs of the Secretaries of Health, Education and Planning in a Territorial Entity.
Descripción/para qué sirve la gestión del conocimiento (inteligencia territorial)
No se menciona.
Aplicaciones propuestas/descritas
El modelo implementado en el documento tiene como objetivo la integración y análisis de los datos de los programas de las secretarías de salud, educación y planeación en un ente territorial en Colombia.
Herramientas y funcionalidades
Integración de datos de diferentes fuentes, la transformación de datos en información útil para la toma de decisiones, la generación de informes y paneles de control en tiempo real, y la posibilidad de análisis y visualización de datos. Se utilizó Ralph Kimball con un modelo de topología en estrella en el Datamart, usando MySQL como bd. Se hizo un sistema ETL con la herramienta Pentaho que permite la obtención y extracción, transformación y carga de datos en el Datamart. Cubos, informes y paneles de control se obtienen con la gestión de herramientas (Pentaho y Power BI).
Aspectos claves
No se menciona

2.4 Principales hallazgos y factores claves.

Si bien los artículos seleccionados fueron pocos debido a que existe poca investigación sobre la gestión del conocimiento territorial e inteligencia territorial, los hallazgos obtenidos proporcionan una buena visión sobre este campo. A pesar de la limitada cantidad de estudios disponibles, se ha logrado identificar ciertos aspectos clave relacionados con la gestión del conocimiento y la inteligencia territorial.

Definición de gestión del conocimiento y la inteligencia en el ámbito territorial: Los artículos mencionan que la gestión del conocimiento y la inteligencia territorial se refieren a la capacidad de adquirir, generar, almacenar, compartir y utilizar el conocimiento en el contexto de un territorio específico. Estos conceptos están orientados a mejorar la toma de decisiones, promover el desarrollo sostenible y fomentar la colaboración entre los actores involucrados.

Objetivos de la gestión del conocimiento y/o inteligencia territorial: Los estudios revisados señalan que los objetivos principales de la gestión del conocimiento y la inteligencia territorial son: analizar y valorar información y conocimientos sobre el entorno territorial, identificar oportunidades y evaluar riesgos, promover el desarrollo de un territorio, facilitar la toma de decisiones informadas, y mejorar la cooperación y la participación de los actores territoriales.

Soluciones tecnológicas para la gestión del conocimiento territorial: Los artículos revisados mencionan plataformas y herramientas utilizadas para llevar a cabo la gestión del conocimiento en los territorios. Entre las soluciones tecnológicas identificadas se encuentran: Deepint.net y el modelo de inteligencia de negocios implementado, ambas soluciones tienen por objetivo gestionar la información que se genera, el primero (Deepint.net) gestiona la información que se genera en las ciudades utilizando sensores mientras que el segundo es para gestionar la información de distintos departamentos que hace difícil su gestión al estar distribuida la información.

Factores clave en la gestión del conocimiento territorial: Durante el análisis de los artículos (los dos primeros), se identificaron varios factores clave que influyen en la efectividad de la gestión del conocimiento en los territorios. El principal factor clave es la colaboración de los actores territoriales.

La revisión sistemática sobre la gestión del conocimiento en los territorios ha resultado fundamental para el proyecto de desarrollo de la plataforma "Conversa con los Andes", ya que se ha logrado adquirir un conocimiento claro de las definiciones asociadas a la gestión del conocimiento y la inteligencia territorial, así como obtener una visión general de las opciones disponibles en diversos contextos.

3. DISEÑO DEL SISTEMA “CONVERSA CON LOS ANDES”

El diseño del sistema "Conversa con Los Andes" se basó en un enfoque colaborativo, fomentando interacciones cercanas tanto con los usuarios del Laboratorio Andes del Sur como con los participantes involucrados en el desarrollo del proyecto

3.1 Diagrama de casos de uso

El proceso de identificación de las funcionalidades del sistema se llevó a cabo mediante discusiones, reuniones y una comprensión directa de las necesidades del laboratorio. Este proceso se materializó en un diagrama de casos de uso, proporcionando una representación visual de las interacciones entre los usuarios y el sistema. El diagrama, que se muestra en la Figura 1, resume de manera gráfica las funcionalidades avanzadas que el bot de Discord debe tener y sirve como guía en la implementación de las funcionalidades

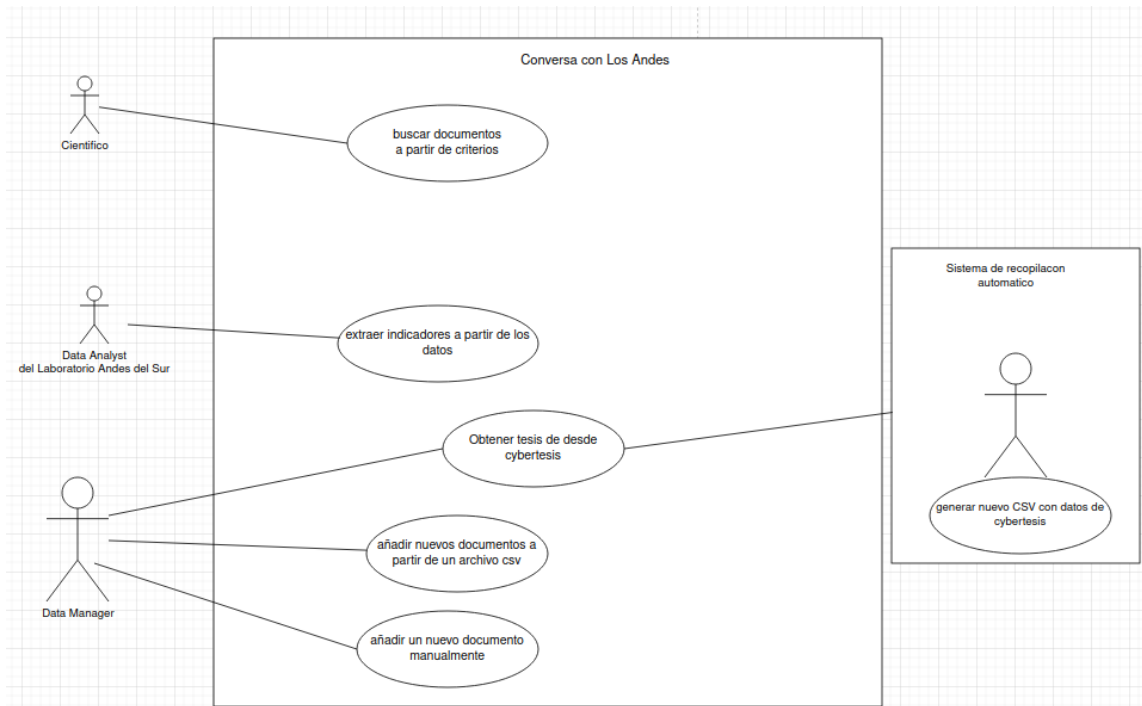


Figura 1. Diagrama de casos de usos.

3.2 Descripción de actores

En la tabla 6 se muestra la descripción de los actores

Tabla 6. Descripción de actores

Actor	Descripción
Científico	Este actor corresponde a los científicos pertenecientes al laboratorio natural que harán uso de la aplicación para buscar documentos
Data analyst	Miembro del laboratorio encargado de realizar análisis de datos, búsqueda y recopilación de información científica
Data manager	Encargado de gestionar y organizar los datos con los que cuenta el laboratorio natural.

3.3 Casos de uso

En el contexto de un bot de Discord, los casos de uso ofrecen descripciones detalladas de las interacciones entre los usuarios y el sistema. La representación visual de estas interacciones se aprecia en el Diagrama de Casos de Uso presentado anteriormente, proporcionando una visión general y estructurada de las funcionalidades clave del sistema. En contraste con interfaces más complejas, este entorno se caracteriza por su simplicidad, basada en comandos directos y respuestas claras. Cada caso de uso, detallado en esta sección, desglosa minuciosamente las acciones específicas que los usuarios pueden realizar al invocar comandos. Estos casos de uso sirven como guía para el desarrollo del sistema, asegurando una implementación que cumple con las expectativas y los objetivos definidos para el proyecto.

3.3.1 Añadir nuevos documentos a partir de un archivo CSV.

En la Tabla 7, Tabla 8 y Tabla 9 se puede ver el caso de uso de añadir nuevos documentos a partir de un archivo CSV.

Tabla 7. Caso de uso añadir nuevos documentos a partir de un archivo CSV

Caso de Uso	Añadir nuevos documentos a partir de un archivo CSV
Actores	Data Manager
Tipo	Primario
Precondición	El usuario tiene acceso al canal de discord.
Postcondición	La información del CSV se inserta correctamente en la base de datos.
Propósito	Permitir a los usuarios cargar y agregar documentos a la base de datos del sistema mediante archivos CSV.
Descripción	Este caso de uso posibilita la incorporación de datos desde un archivo CSV a la base de datos del sistema, incluyendo el contenido de los documentos identificados en el CSV.

Tabla 8. Curso normal de los eventos añadir nuevos documentos a partir de un archivo CSV

Acción de los actores	Respuesta del sistema
1) Este caso de uso comienza cuando un usuario invoca el comando de agregar documentos proporcionando un archivo CSV.	<p>2) El sistema valida el archivo CSV</p> <p>3)El sistema muestra un mensaje indicando que el archivo CSV es valido</p> <p>4)El sistema guarda la información en la base de datos</p> <p>5)El sistema muestra un mensaje indicando que la información del CSV fue guardada en la base de datos</p> <p>6) El sistema descarga y guarda el contenido de los documentos en segundo plano</p> <p>7)El sistema muestra que se están descargando los documentos y guardando la información en segundo plano</p> <p>8) El sistema muestra un mensaje de que los documentos fueron descargados mostrando una lista de documentos que no se pudieron descargar</p>

Tabla 9. Curso alternativo añadir nuevos documentos a partir de un archivo CSV

Acción de los actores	Respuesta del sistema
1.1) Un usuario invoca el comando de agregar documentos proporcionando un archivo CSV con columnas invalidas	<p>2.1) El sistema reconoce que el archivo CSV no cumple con las columnas requeridas.</p> <p>2.2) El sistema muestra un mensaje indicando que el archivo CSV no es válido y solicita al usuario que ingrese el comando con un archivo válido.</p>

3.1.2 Buscar documentos a partir de criterios.

En la Tabla 10, Tabla 11 y Tabla 12 se puede ver el caso de uso de buscar documentos a partir de criterios.

Tabla 10. Caso de uso buscar documentos a partir de criterios.

Caso de Uso	Buscar documentos a partir de criterios
Actores	Científico
Tipo	Primario
Precondición	El usuario tiene acceso al canal de Discord.
Postcondición	
Propósito	Buscar documentos en la base de datos que cumplan parámetros.
Descripción	Permite buscar información en la plataforma utilizando criterios como título, nombre del autor, línea temática u otros atributos relevantes. Esta función permite a los usuarios encontrar documentos específicos que se ajusten a sus necesidades e intereses.

Tabla 11. Curso normal de los eventos buscar documentos a partir de criterios.

Acción de los actores	Respuesta del sistema
1)Un usuario en el canal de Discord invoca el comando de búsqueda de documentos seguido de una cadena de búsqueda	<p>2) El sistema analiza los parámetros y ejecuta clausulas SQL necesarias para realizar la búsqueda de documentos.</p> <p>3) El sistema devuelve los resultados de la búsqueda al usuario en el canal de Discord de una manera legible</p>

Tabla 12. Curso alternativo buscar documentos a partir de criterios.

Acción de los actores	Respuesta del sistema
1.1) El usuario ingresa parámetros incorrectos o no válidos en el canal de Discord.	2.1) El sistema envía un mensaje al canal de Discord informando al usuario sobre el error.

3.3.3 Generar nuevo CSV con datos de cybertesis

En la Tabla 13, Tabla 14 y Tabla 15 se puede ver el caso de uso de generar nuevo CSV con datos de cybertesis

Tabla 13. Generar nuevo CSV con datos de cybertesis

Caso de Uso	Generar nuevo CSV con datos de cybertesis
Actores	Desencadenado por el actor Data Manager al llamar la funcionalidad de obtener tesis desde cybertesis en el canal de Discord.
Tipo	Primario
Precondición	El actor llama a la funcionalidad de obtener tesis desde cybertesis en el canal de Discord.
Postcondición	Se genera un nuevo archivo CSV con datos desde cybertesis.
Propósito	Generar un documento CSV con información obtenida y analizada desde cybertesis
Descripción	Este caso de uso posibilita al laboratorio natural conocer cuáles son las tesis que pueden ser de interés para ellos.

Tabla 14. Curso normal de los eventos generar nuevo CSV con datos de cybertesis

Acción de los actores	Respuesta del sistema
1) El usuario invoca el comando de obtener de tesis a través del canal de Discord.	<p>2) Se inicia un proceso de <i>scraping</i> que analiza automáticamente tesis desde la página web cybertesis.uach.cl.</p> <p>3) El sistema devuelve al canal de Discord un archivo CSV con los resultados de los documentos identificados en el <i>scraping</i></p>

3.3.4 Añadir un nuevo documento manualmente

En la Tabla 15, Tabla 16 y Tabla 17 se puede ver el caso de uso añadir un nuevo documento manualmente

Tabla 15. Caso de uso añadir un nuevo documento manualmente

Caso de Uso	Añadir un nuevo documento manualmente
Actores	Data Manager
Tipo	Primario
Precondición	El usuario tiene acceso al canal de Discord.
Postcondición	Los datos del documento se insertan correctamente en la base de datos.
Propósito	Permitir a los usuarios cargar y agregar documentos a la base de datos del sistema de manera unitaria.
Descripción	Este caso de uso posibilita la incorporación de una nueva manera de ingresar documentos, pero solo de manera unitaria y sin depender de un archivo CSV.

Tabla 16. Curso normal de los eventos añadir un nuevo documento manualmente

Acción de los actores	Respuesta del sistema
1) Este caso de uso comienza cuando un usuario invoca el comando addManualDocument proporcionando los detalles del documento como argumento.	<p>2) El sistema valida los parámetros</p> <p>3) El sistema guarda la información en la base de datos</p> <p>4) El sistema descarga y guarda el contenido de los documentos en segundo plano.</p> <p>5) El sistema muestra un mensaje indicando que el documento fue agregado a la base de datos.</p>

Tabla 17. Curso alternativo añadir un nuevo documento manualmente

Acción de los actores	Respuesta del sistema
1.1) Usuario invoca el comando addManualDocument proporcionando detalles incorrectos del documento como argumento.	<p>2.1) El sistema reconoce que los parámetros no son validos</p> <p>2.2) El sistema muestra un mensaje indicando que los datos ingresados no cumplen con el formato o la cantidad de columnas requeridas.</p>

3.3.5 Extraer indicadores a partir de los datos

En la Tabla 18 y Tabla 19 se puede ver el caso de uso extraer indicadores a partir de los datos

Tabla 18. Caso de uso extraer indicadores a partir de los datos

Caso de Uso	Extraer indicadores a partir de los datos
Actores	<i>Data analyst</i>
Tipo	Primario
Precondición	El usuario tiene acceso al canal de Discord.
Postcondición	
Propósito	Obtener información detallada sobre los datos presentes en la base de datos.
Descripción	Este caso de uso tiene como objetivo darle a conocer al actor diferentes indicadores, como pueden ser la cantidad de documentos que existen en la base de datos, que autores han escrito más, etc.

Tabla 19. Curso normal de los eventos caso de uso extraer indicadores a partir de los datos

Acción de los actores	Respuesta del sistema
1) El usuario invoca el comando de análisis bibliométrico a través del sistema.	2) Se ejecutan consultas SQL específicas para obtener información sobre la base de datos de. 3) El sistema presenta al usuario los resultados del análisis en el formato adecuado (Total de Documentos en la Base de Datos, Autores más prolíficos y el número de documentos asociados, distribución de los documentos por laboratorios temáticos, organizaciones más activas y el número de documentos asociados y la evolución de la producción de documentos a lo largo de los años.

3.4 Diseño de la base de datos

El diseño de la base de datos surgió como respuesta a la necesidad de gestionar los documentos identificados como útiles para el laboratorio, guardados en un archivo CSV.

En un esfuerzo por garantizar la idoneidad del diseño, se mantuvo un proceso de retroalimentación constante con los miembros del Laboratorio Natural Andes del Sur. A pesar de que no tenían conocimientos profundos en informática. Se presentó visualmente el progreso del modelo de base de datos en reuniones regulares, permitiendo una comprensión más clara de cómo el sistema gestionaría y aprovecharía el archivo con el cuentan, así como el papel que desempeñarían las tablas de la base de datos en dicho proceso.

El diseño de la base de datos fue concebido para su implementación en una base de datos relacional. Esto se debe a la eficacia de las bases de datos relacionales para gestionar datos y establecer relaciones entre entidades, lo cual se alinea de manera óptima con la estructura relacional de los datos contenidos en el archivo CSV.

Uno de los elementos principales en el diseño de la base de datos es lograr la representación visual a través del diagrama entidad-relación. Este diagrama proporciona una visión esquemática de cómo las entidades se relacionan entre sí, ofreciendo una herramienta valiosa para comprender la estructura y la interconexión de los datos. La claridad proporcionada por este tipo de representación visual facilita la comunicación y comprensión. La Figura 2 muestra el diagrama entidad-relación alcanzado

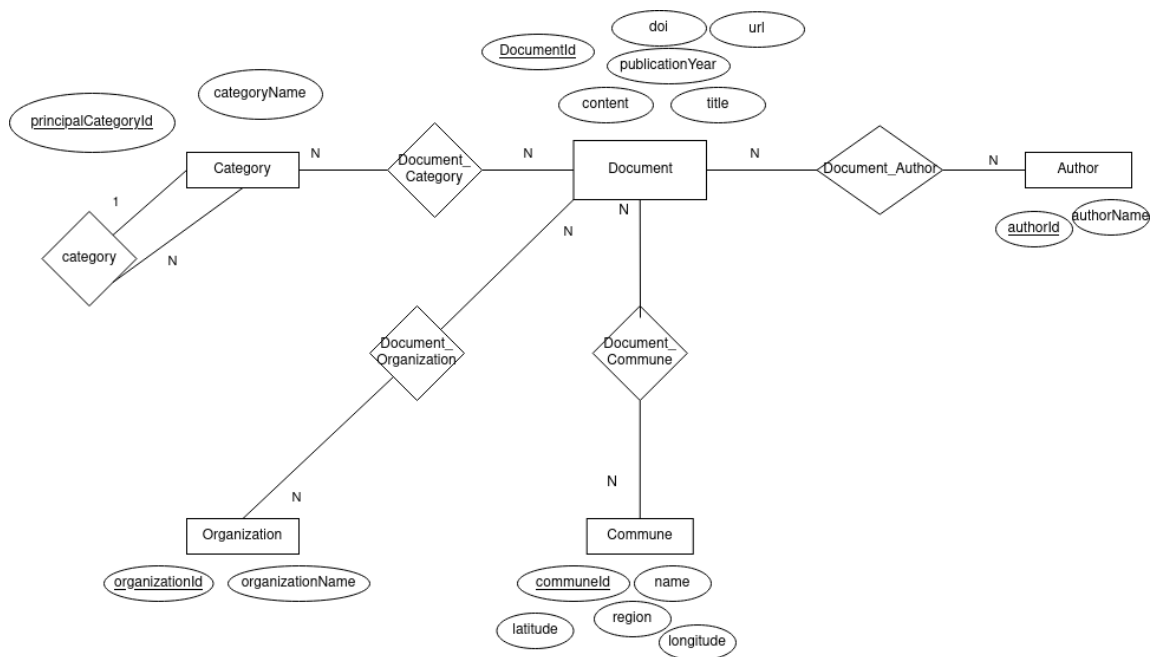


Figura 2. Diagrama entidad relación.

De igual manera, para complementar la representación visual a través del diagrama entidad-relación, se elaboró un diagrama relacional que muestra la estructura y las relaciones entre las tablas de la base de datos. Este diagrama relacional proporciona una perspectiva adicional, detallando la forma en que las tablas están vinculadas y cómo se establecen las relaciones clave primaria y foránea. La Figura 3 presenta el diagrama relacional correspondiente al diseño de la base de datos.

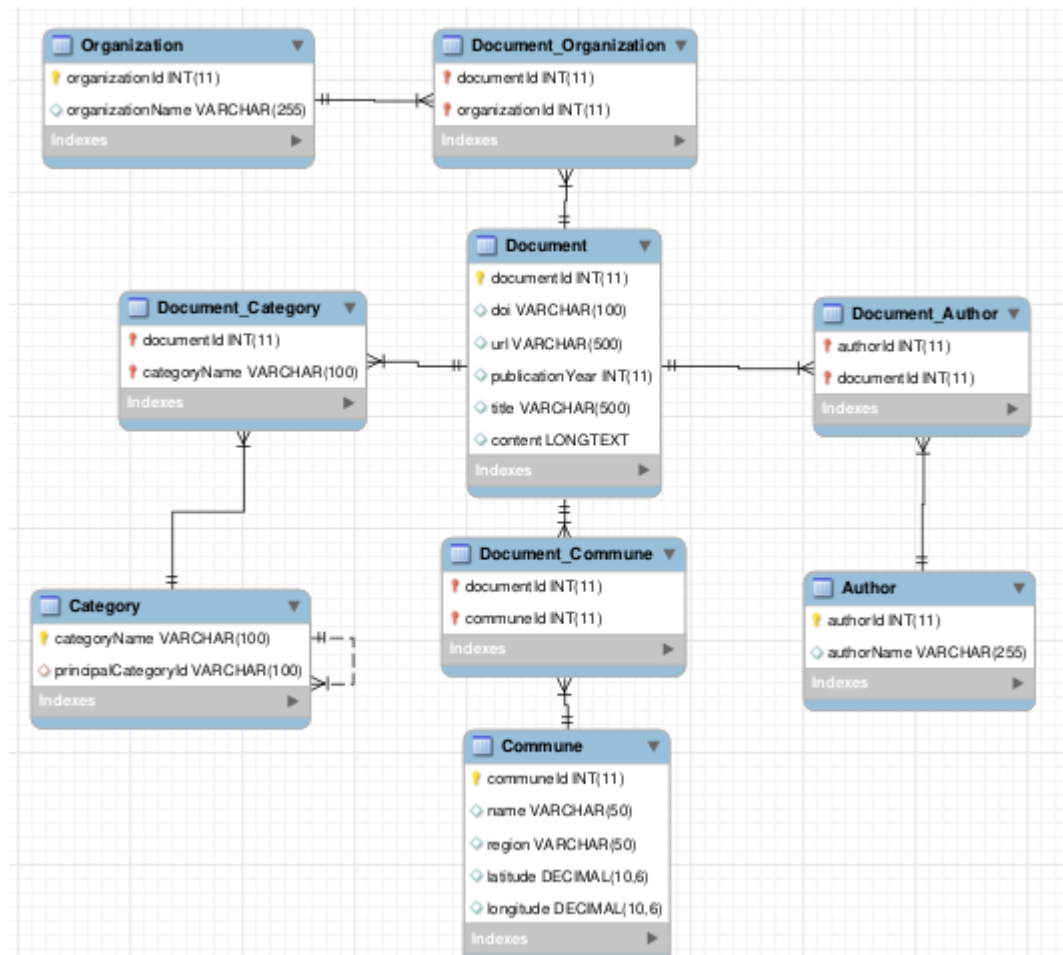


Figura 3. Diagrama relacional.

Además de los diagramas presentados, otro componente esencial en el diseño de la base de datos es el diccionario de datos. Este actúa como una guía que ayuda a comprender de manera más profunda la información almacenada en la base de datos y cómo se organizan las tablas. Resulta fundamental tanto en el momento de implementar la base de datos por parte de los desarrolladores como en momentos posteriores, ya que proporciona una referencia útil en caso de olvidar detalles específicos de los datos almacenados.

A continuación, en la Tabla 20, Tabla 21, Tabla 22, Tabla 23 y Tabla 24 se presenta el diccionario de datos de cada tabla.

Tabla 20. Diccionario datos tabla Author.

Nombre Columna	¿Es una clave?	Tipo datos	Ejemplo	Descripción	¿Es obligatoria?
authorId	si	INT(11)	1	Identificador único para cada autor. Se incrementa automáticamente con cada nuevo autor agregado a la tabla.	Sí, es obligatoria
authorName	no	VARCHAR(255)	John Doe	Es el nombre del autor de un documento.	No

Tabla 21. Diccionario datos tabla Commune.

Nombre Columna	¿Es una clave?	Tipo datos	Ejemplo	Descripción	¿Es obligatoria?
communeId	Sí, es la clave primaria.	INT	1	Identificador único para cada comuna. Se incrementa automáticamente con cada nueva comuna agregada a la tabla.	Sí, es obligatoria.
name	No	VARCHAR(50)	“Valdivia”	El nombre de la comuna.	No.
region	No	VARCHAR(50)	Los Ríos	La región a la que pertenece la comuna.	No.
latitude	No	VARCHAR(50)	-15.0	La latitud geográfica de la comuna.	No.
longitude	No	VARCHAR(50)	-40.33	La longitud geográfica de la comuna.	No

Tabla 22. Diccionario datos tabla Document.

Nombre Columna	¿Es una clave?	Tipo datos	Ejemplo	Descripción	¿Es obligatoria?
documentId	Si	INT	1	Identificador único para cada documento. Se incrementa automáticamente con cada nuevo documento agregado a la tabla.	Sí, es obligatoria.
doi	No	VARCHAR(100)	10.12345/abcde	El Identificador de Objeto Digital (DOI) del documento, si está disponible.	No.
url	No	VARCHAR(500)	https://example.com/document	La URL del documento en línea, si está disponible.	No
publicationYear	No	INT	2023	El año de publicación del documento.	No
title	No	VARCHAR(500)	El magma del volcan Villarrica	El título del documento.	No
content	No	TEXT	Contenido del documento...	El contenido principal del documento, almacenado como texto.	No

Tabla 23. Diccionario datos tabla Category.

Nombre Columna	¿Es una clave?	Tipo datos	Ejemplo	Descripción	¿Es obligatoria?
categoryName	Si, es la clave primaria	VARCHAR(100)	Volcanología y peligro volcánico	El nombre de la categoría. Es la clave primaria de la tabla y se espera que sea único para cada categoría.	Sí, es obligatoria.
principalCategoryId	No	VARCHAR(100)	LT1: EVOLUCIÓN Y HERENCIA GEOLÓGICA	La categoría principal a la que pertenece esta categoría.	No

Tabla 24. Diccionario datos tabla Organization.

Nombre Columna	¿Es una clave ?	Tipo datos	Ejemplo	Descripción	¿Es obligatoria?
organizationId	Si	INT(11)	1	Identificador único para cada organización. Se incrementa automáticamente con cada nueva organización agregada a la tabla	Si
organizationName	No	VARCHAR(255)	BULLETIN OF VOLCANOLOGY	El nombre de la organización.	No

3.5 Diagrama de componentes

Con el propósito de brindar una visión clara de la arquitectura del sistema, se ha elaborado un diagrama de componentes, el cual se presenta en la Figura 4. En este esquema, se puede apreciar que el componente principal es el componente bot, que actúa como el núcleo central de la plataforma. Este componente no opera de manera aislada, sino que se integra con otros tres componentes: El componente Canal Discord, el componente Sistema de recopilación automático y el componente Base de datos.

El componente sistema de recopilación automático es el que se encarga de conectarse al sitio web donde están alojadas las tesis de la Universidad Austral y extraer información específica sobre estas para construir un archivo CSV con documentos que cumplan los lineamientos del laboratorio natural.

El componente Canal Discord, sirve como el medio donde los usuarios interactúan con el bot. En este canal, los usuarios pueden enviar mensajes, comandos y realizar diversas interacciones que el bot procesará.

El componente Base de Datos sirve como el almacén central para la información relevante en la plataforma "Conversa con Los Andes". Es responsable de almacenar y organizar los datos necesarios para el funcionamiento del sistema.

La interacción entre estos componentes sirve para coordinar y desencadenar las diversas funcionalidades en respuesta a las interacciones de los usuarios en el servidor de Discord.

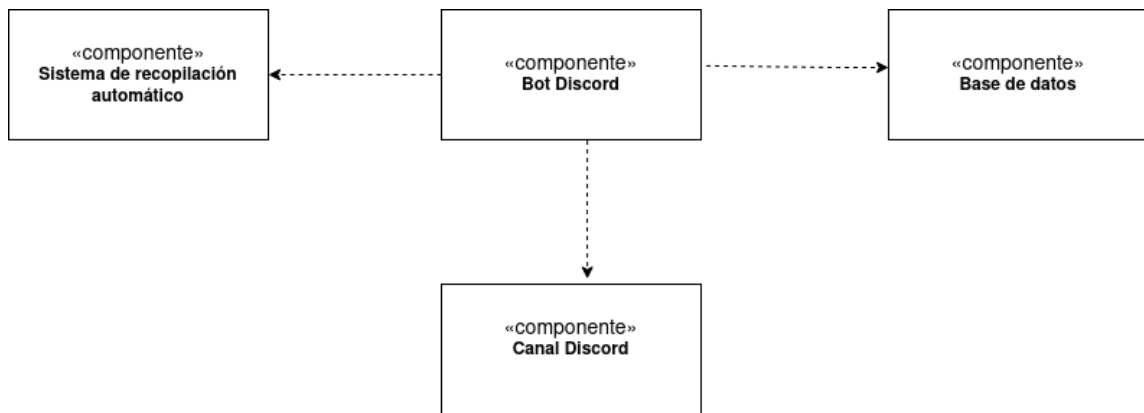


Figura 4. Diagrama de componentes.

3.6 Diagrama de clases

Se menciona que el componente principal es el componente Bot. Este componente está implementado en una clase llamada de la misma manera y se encarga de gestionar las interacciones del usuario en el canal de Discord, procesar comandos y coordinar la ejecución de las funcionalidades del sistema. Los métodos dentro de esta clase representan las capacidades del bot y son invocados en respuesta a los comandos emitidos por los usuarios.

La clase Bot opera en conjunto con otras clases que dan vida al sistema Conversa Con Los Andes. Entre estas clases se incluye la clase databaseConnector, que tiene como objetivo gestionar la conexión con la base de datos. Su estructura se orienta hacia la consolidación de todas las operaciones relacionadas con la base de datos, garantizando que solo esta clase se comuniquen directamente con la misma. Estos métodos, como connect() para establecer la conexión inicial, insertDocument() para agregar documentos a la base de datos, close() para cerrar la conexión de manera segura cuando sea necesario, executeQuery() para ejecutar consultas a la base de datos, y insertDocumentData() para agregar el contenido de un documento a la base de datos, son fundamentales para el funcionamiento del sistema.

La clase Scraper destaca por dos métodos fundamentales. El método downloadDocument() se encarga de la descarga y almacenamiento de documentos desde diversas fuentes, mientras que el método scrapeThesis() es el encargado de comunicarse con el componente Sistema de recopilación automático.

Por otro lado, la clase Tools proporciona utilidades generales, como la lectura de archivos CSV con el método readCSV(), y la extracción del contenido de cada documento PDF mediante el método readPdf().

A continuación, en la Figura 5 se presenta el diagrama de clases que representa la estructura del sistema. Este diagrama ofrece una visión de las clases, sus métodos principales y las relaciones entre ellas.

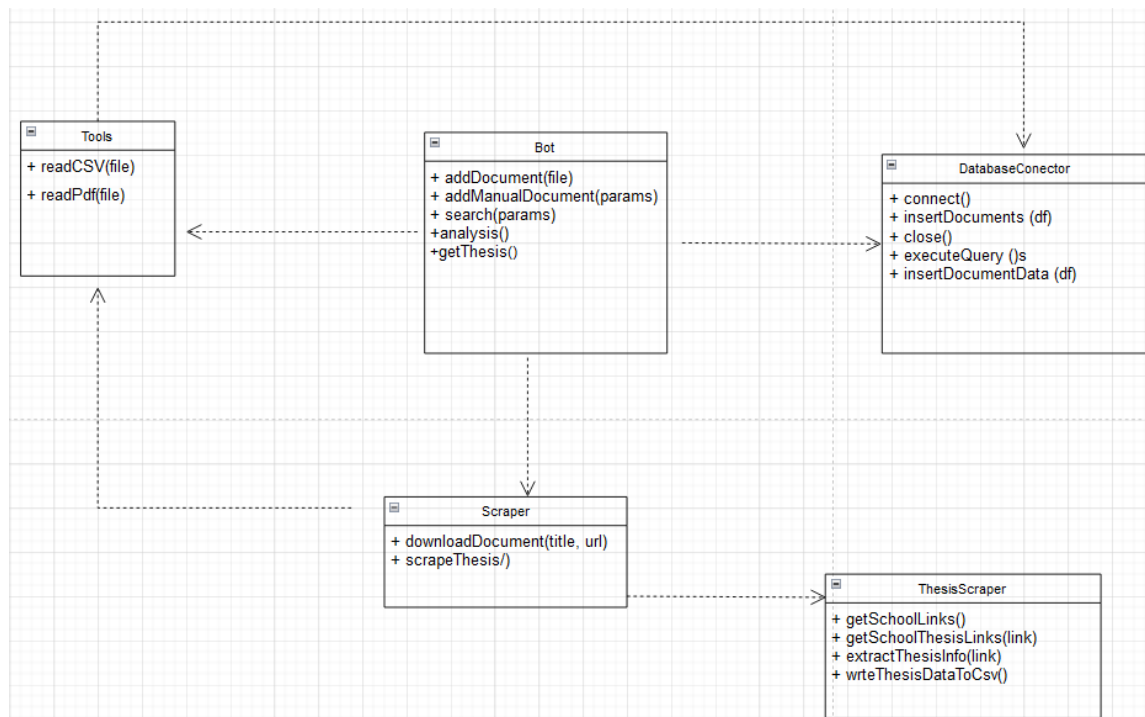


Figura 5. Diagrama de clases

El componente Sistema de recopilación automático está implementado mediante una clase llamada ThesisScraper; esta clase es donde está automatizado el proceso de obtener y generar un archivo CSV con datos desde cybertesis

El metodo getSchoolLinks() se encarga de obtener enlaces que dirigen hacia las distintas escuelas.

El método getSchoolThesisLinks() utiliza los enlaces obtenidos en getSchoolLinks() para acceder a los enlaces individuales de cada tesis asociada a diferentes escuelas.

El método extractTesisInfo() emplea los enlaces obtenidos en getSchoolThesisLinks() para extraer información detallada de cada tesis. Para obtener información que se encuentra en el contenido cada uno documento, es necesario descargar este documento y esto se hace mediante el método downloadAndSavePdf(), que descarga cada tesis

detectada en `extractTesisInfo()` para analizar el contenido en busca de condiciones específicas.

El método `writeThesisDataToCsv` toma datos de tesis extraídos y los escribe en un archivo CSV.

3.7 Diagrama de secuencia

Los diagramas de secuencia representan una visión del sistema, ilustrando la interacción entre sus distintas partes durante la ejecución. Estos diagramas sirven como herramientas para comprender el comportamiento de las funciones y la colaboración entre clases u objetos. Al proporcionar una representación visual clara de las interacciones entre las clases, estos diagramas sirven como guías esenciales para el desarrollo del código.

A continuación, en la Figura 6 se presenta el diagrama de secuencia de la funcionalidad de agregar nuevos documentos a partir de un archivo CSV.

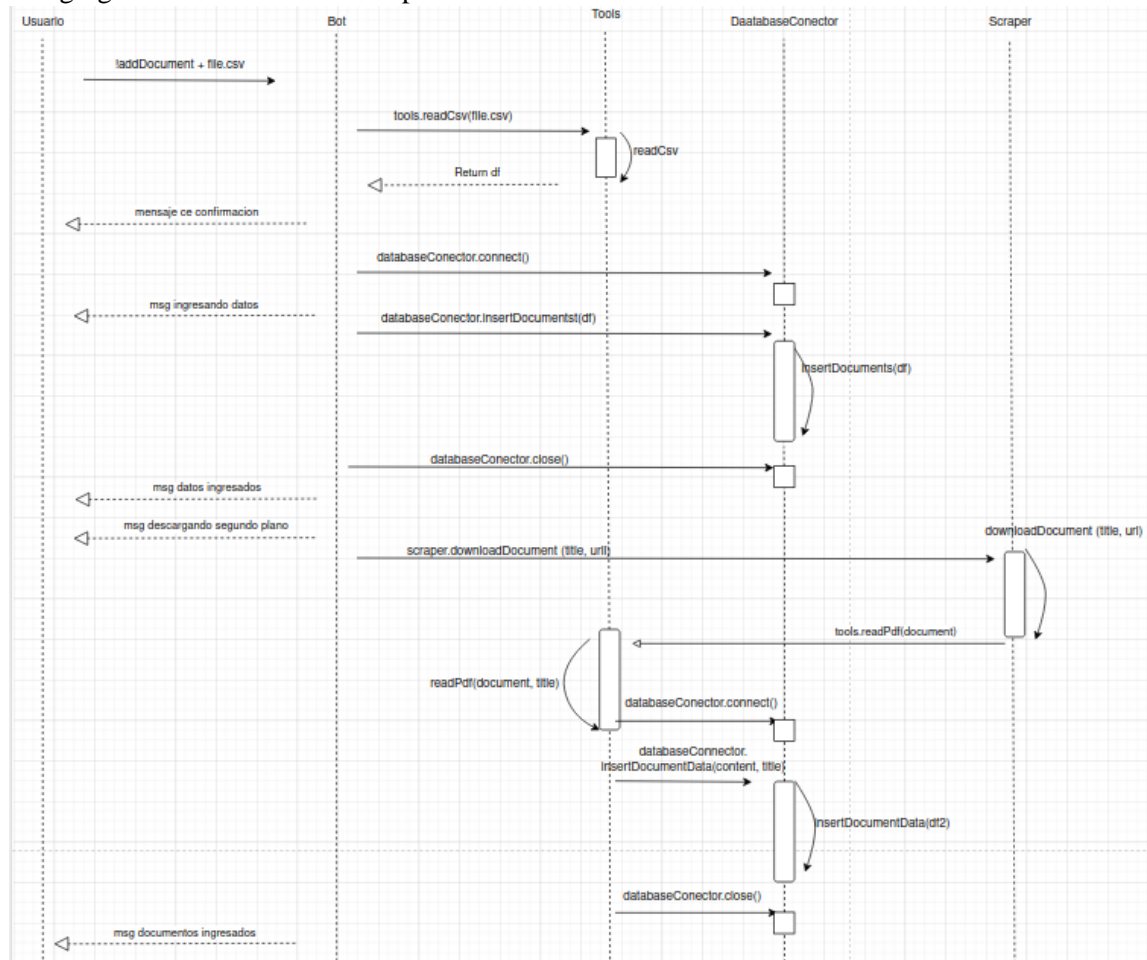


Figura 6 Diagrama de secuencia agregar nuevos documentos a partir de CSV.

En la Figura 7 se muestra el diagrama de secuencia de la funcionalidad agregar un documento manualmente

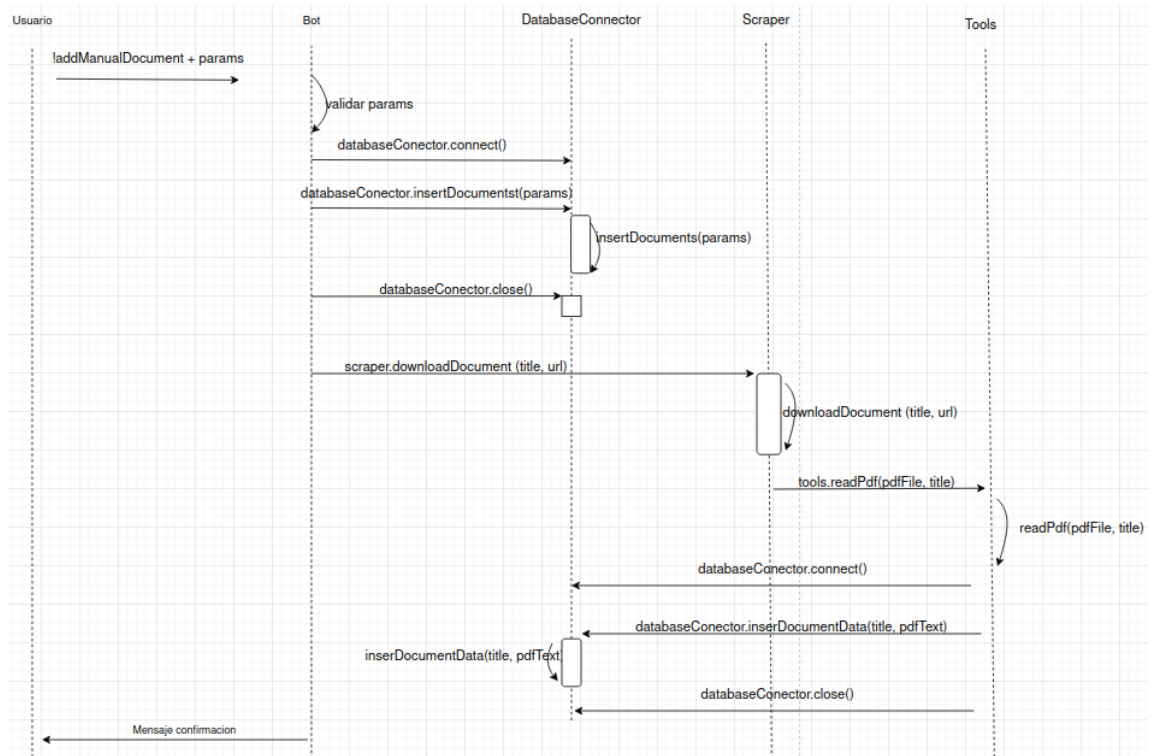


Figura 7. Diagrama de secuencia agregar documento manualmente

En la figura 8 se muestra el diagrama de secuencia del caso de uso de generar un nuevo CSV con datos de cybertesis.

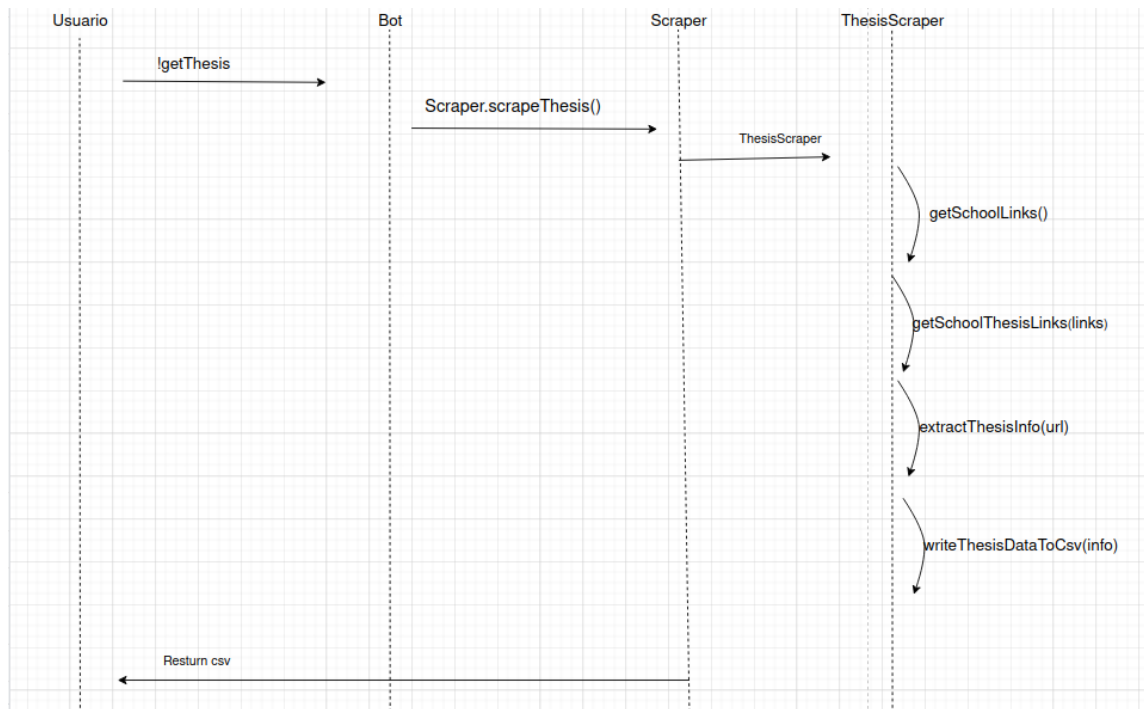


Figura 8: Diagrama de secuencia generar un nuevo CSV con datos de cybertesis.

En la figura 9 se muestra el diagrama de secuencia del caso de uso buscar documentos a partir de criterios

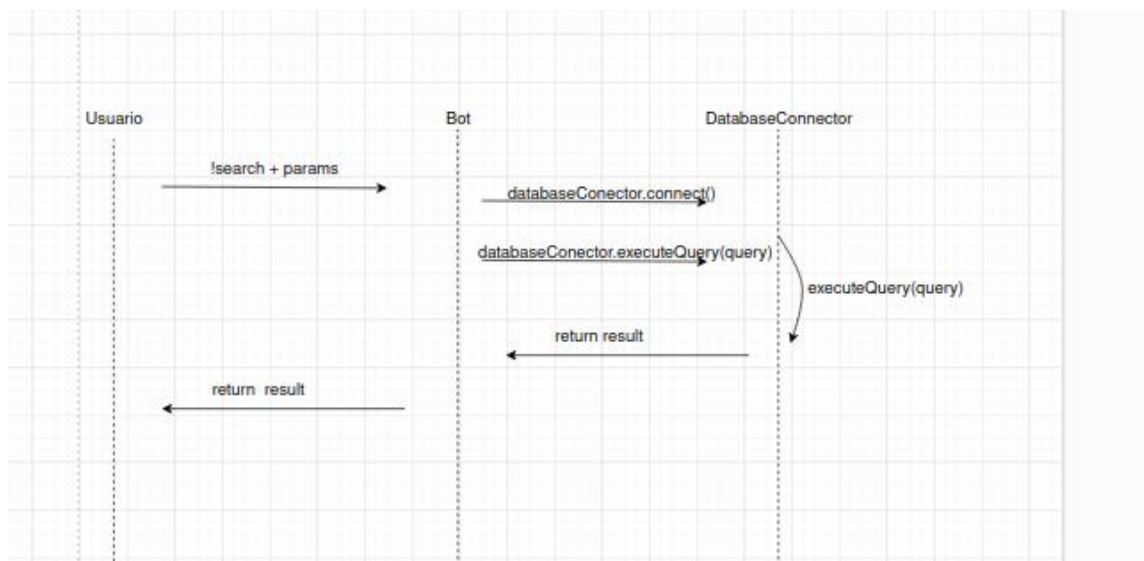


Figura 9: Diagrama de secuencia buscar documentos a partir de criterios.

En la figura 10 se muestra el diagrama de secuencia del caso extraer indicadores a partir de los datos.

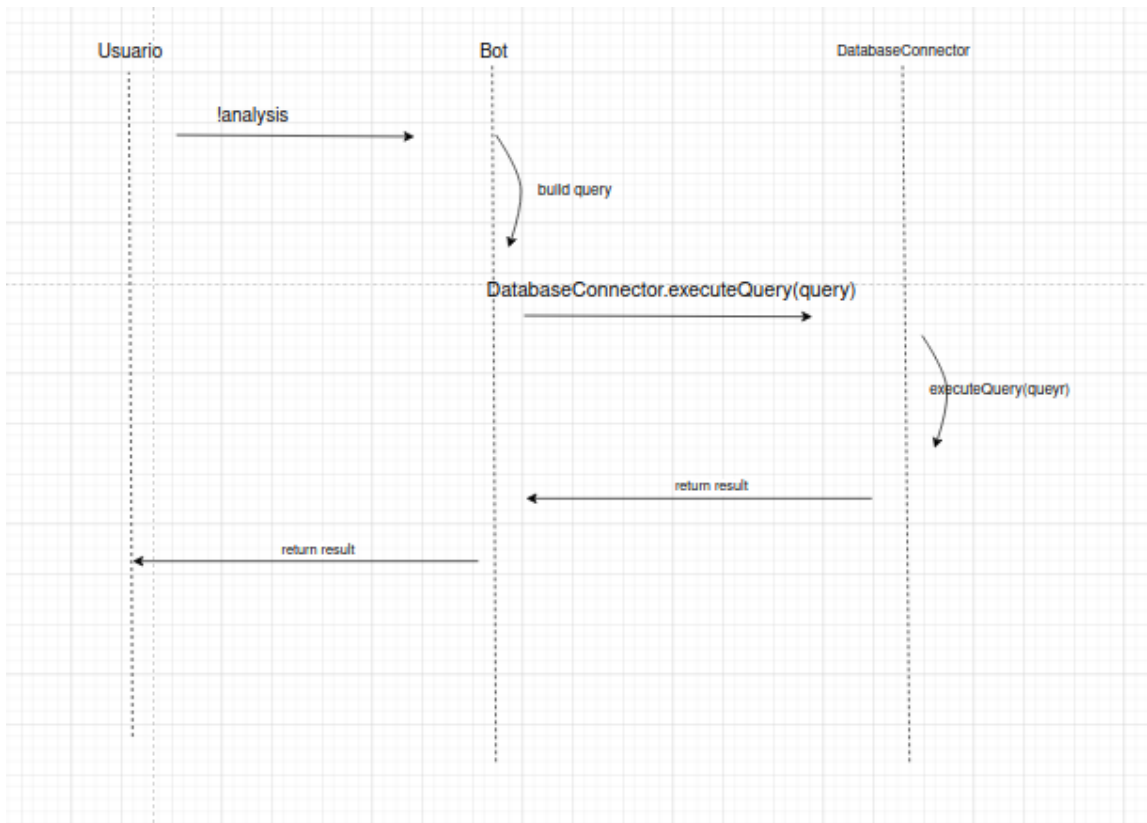


Figura 10: Diagrama de secuencia extraer indicadores a partir de los datos

4. IMPLEMENTACIÓN

4.1 Tecnologías utilizadas

A continuación, se procederán a presentar las principales tecnologías utilizadas durante la implementación del sistema.

4.1.1 Python

Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el *machine learning*. Los desarrolladores utilizan Python porque es eficiente y fácil de aprender, además de que se puede ejecutar en muchas plataformas diferentes. (Amazon Web Services, s.f.)

4.1.2 MariaDB

MariaDB es una versión modificada de MySQL. MariaDB fue creada por el equipo de desarrollo original de MySQL debido a problemas de licencia y distribución después de que Oracle Corporation adquiriera MySQL. (Amazon Web Services, s.f.)

4.1.3 Discord.py

Discord.py es una librería que permite utilizar la API de Discord mediante python utilizando sintaxis de *async/await*, la cual permite crear código para un bot. (Discord.py, s.f.).

4.1.4 Beautiful Soup

Con Beautiful Soup los desarrolladores pueden acceder fácilmente a los elementos del documento HTML o XML, realizar búsquedas específicas y realizar manipulaciones en la estructura del árbol, lo que simplifica el proceso de extracción y análisis de datos web. (Beautiful Soup, s.f.).

4.1.5 Asyncio

Asyncio es una biblioteca para escribir código concurrente utilizando la sintaxis *async/await*. Se emplea como base para diversos marcos asincrónicos en Python que ofrecen servidores de red y web de alto rendimiento, bibliotecas de conexión a bases de datos, colas de tareas distribuidas, entre otros. Asyncio suele ser ideal para código de red estructurado de alto nivel y operaciones ligadas a la entrada/salida. (Asyncio, s.f.).

4.1.6 GitHub

GitHub es una plataforma de hospedaje de código para el control de versiones y la colaboración. Permite que personas trabajen juntos en proyectos desde donde sea. (GitHub, s.f.).

4.1.7 spacy

spaCy es una biblioteca de procesamiento de lenguaje natural Python diseñada específicamente con el objetivo de ser una biblioteca útil para implementar sistemas listos para producción. (spaCy, s.f.).

4.2 ¿Por qué utilizar Discord?

MidJourney, un destacado laboratorio de investigación independiente enfocado en diseño, infraestructura humana e inteligencia artificial, ha capturado la atención gracias a su sistema excepcional, el cual tiene la capacidad de convertir descripciones textuales en imágenes sorprendentemente realistas. Este sistema opera a través de un bot en la plataforma Discord y ha ganado su lugar en el ámbito de la generación de imágenes a partir de texto.

Inspirado por el éxito y la eficacia de MidJourney, se ha tomado este innovador sistema como referencia al seleccionar Discord como la plataforma principal para implementar las funcionalidades que se esperan en nuestro sistema.

Siguiendo esta orientación, además, Discord se destaca por sus ventajas adicionales que contribuyen a la eficacia y facilidad de uso de nuestro proyecto.

Las ventajas de esta plataforma son:

- Cualquiera puede crear un servidor de Discord de forma gratuita y se establece y mantiene en la propia infraestructura de Discord, lo que lo hace fácil de usar.
- Los servidores son privados y las personas solo pueden unirse a través de enlaces de invitación. Los usuarios tienen la capacidad de unirse a hasta 100 servidores diferentes, tratando cada servidor como una comunidad independiente.
- Los servidores pueden tener canales de texto (con la capacidad de compartir imágenes, vídeos, etc.) y canales de voz. Además, los propietarios de los servidores pueden personalizar los canales según sus necesidades. Existen numerosas aplicaciones y bots disponibles para mejorar la funcionalidad del canal.

4.3 Implementación base de datos

La implementación se llevó a cabo utilizando MariaDB versión 10.6.12. Dado que las tablas ya estaban identificadas durante la fase de diseño mediante el uso de diagramas de entidad-relación, el diagrama relacional y el diccionario de datos. El proceso de implementación consistió en plasmar esas estructuras en la base de datos.

A continuación, en la figura 11 y Figura 12 se presenta el código de cómo fueron creadas las tablas en la base de datos.

```
CREATE TABLE `Author` (  
  `authorId` int(11) NOT NULL AUTO_INCREMENT,  
  `authorName` varchar(255) DEFAULT NULL,  
  PRIMARY KEY (`authorId`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_general_ci;
```

Figura 11: Código para crear tabla Author.

```
CREATE TABLE `Document` (  
  `documentId` int(11) NOT NULL AUTO_INCREMENT,  
  `doi` varchar(100) DEFAULT NULL,  
  `url` varchar(500) DEFAULT NULL,  
  `publicationYear` int(11) DEFAULT NULL,  
  `title` varchar(500) DEFAULT NULL,  
  `content` text DEFAULT NULL,  
  PRIMARY KEY (`documentId`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_general_ci;
```

Figura 12: Código para crear tabla Document.

En la figura 13 se muestran todas las tablas implementadas en la base de datos.

Tables_in_andes
Author
Category
Commune
Concept
Document
Document_Author
Document_Category
Document_Commune
Document_Concept
Document_Organization
Organization

Figura 13: Tablas implementadas.

4.4 Implementación de funcionalidades y desarrollo

La implementación del sistema Conversa con Los Andes se llevó a cabo mediante una cuidadosa traducción de los elementos de diseño en comandos ejecutables para el bot de Discord. Durante este proceso, se implementaron las funcionalidades a través de comandos específicos, para invocar cada comando en el canal de Discord se utiliza el símbolo “!”:

- `!addDocument`: Permite añadir nuevos documentos a la plataforma a partir de un archivo csv.
- `!search`: Implementa la funcionalidad de búsqueda de documentos, utilizando parámetros como año, comuna, región, categoría y laboratorio temático.
- `!addManualDocument`: funcionalidad para agregar un nuevo documento manualmente
- `!analysis`: Extrae indicadores relevantes a partir de los datos almacenados en la base de datos.
- `!getThesis`: Genera un archivo CSV con datos recopilados de la plataforma cybertesis.uach.cl.

Durante el proceso de implementación del código, se seleccionó el caso de uso de añadir nuevos documentos a partir de un archivo CSV como punto de partida. Adoptando un enfoque iterativo, se dedicó un período de dos semanas al desarrollo y codificación para llevar a cabo este caso de uso. Posteriormente, se presentaron los resultados obtenidos a los miembros del laboratorio, lo que facilitó una iteración ágil con la oportunidad de recibir comentarios y realizar ajustes según las necesidades y expectativas. Luego, se siguió el mismo proceso con los otros casos de uso.

Se garantizó la coherencia entre el diseño realizado en la sección correspondiente y la implementación del sistema, utilizando los casos de uso y secuencia como guía principal durante este proceso de implementación

4.4.1 Añadir nuevos documentos a partir de un archivo CSV

La implementación de la funcionalidad para agregar nuevos documentos desde un archivo CSV mediante el comando !addDocument tiene como propósito principal facilitar la entrada de datos a partir de un archivo en el canal de Discord, para luego ser incorporados a la base de datos. Este proceso consta de varias etapas.

En la etapa inicial, se lleva a cabo la verificación del formato del archivo CSV mediante el método readCSV de la clase Tools y la biblioteca pandas. El sistema busca validar la presencia de columnas esenciales como 'LT', 'AUTORES/AS', 'TÍTULO', 'AÑO', 'REVISTA', 'DOI', 'CATEGORÍA', 'REGIÓN', 'COMUNAS' y 'Enlace'. Si el formato es correcto, se emite un mensaje indicando la validez del archivo, asegurando coherencia antes de proceder. En caso de problemas, se notifica al usuario.

A continuación, en la Figura 14, figura 15 y Figura 16 se muestra un ejemplo visual del formato esperado en un archivo CSV bien formateado

	A	B	C
1	LT	AUTORES/AS	TÍTULO
2	LT1	ADRIASOLA A;STOCKHERT B	COOLING HISTORIES AND DEFORMATION OF PLUTONIC ROCKS ALONG THE LIQ
3	LT1	ADRIASOLA A;THOMSON S;BRIX M;HERVÉ F;STÖCKHERT B	POSTMAGMATIC COOLING AND LATE CENOZOIC DENUDATION OF THE NORTH P
4	LT1	AIUPPA A;BITETTO M;FRANCOFONTE V;VELASQUEZ G;PARRA O A	CO2GAS PRECURSOR TO THE MARCH 2015 VILLARRICA VOLCANO ERUPTION
5	LT1	ALFANO F;BONADONNA C;GURIOLI L	INSIGHTS INTO ERUPTION DYNAMICS FROM TEXTURAL ANALYSIS THE CASE OF
6	LT1	ALFANO F;BONADONNA C;VOLENTIK A;CONNOR C;WATT S;PYLÉ	TEPHRA STRATIGRAPHY AND ERUPTIVE VOLUME OF THE MAY 2008 CHAITÉN V
7	LT1	ALFANO F;BONADONNA C;WATT S;CONNOR C;VOLENTIK A;PYLÉ	RECONSTRUCTION OF TOTAL GRAIN SIZE DISTRIBUTION OF THE CLIMACTIC P
8	LT1	ALLOWAY B;MORENO P;PEARCE N;DE POL-HOLZ R;HENRIQUEZ *	STRATIGRAPHY AGE AND CORRELATION OF LEPUÉ TEPHRA A WIDESPREAD C
9	LT1	ALLOWAY B;PEARCE N;MORENO G;JARA I;DE POL-HOLZ R;OUTP	AN 18000 YEARLONG ERUPTIVE RECORD FROM VOLCAN CHAITEN VOLCANO NO
10	LT1	ALLOWAY B;PEARCE N;VILLAROSA G;QUTES V;MORENO P	MULTIPLE MELT BODIES FED THE AD 2011 ERUPTION OF PUYEHUECORDON CA

Figura 14. Formato esperado de un archivo CSV parte 1.

	D	E	F	G
1	AÑO	REVISTA	DOI	CATEGORÍA
2	2008	REVISTA GEOLOGICA DE CHILE	10.4067/S0716-02082008000100002	Tectónica y Peligro Sísmico
3	2006	INTERNATIONAL JOURNAL OF EARTH SCIENCES	10.1007/s00531-005-0027-9	Mineralogía y Petrología;Tectónica
4	2017	GEOCHEMISTRY, GEOPHYSICS, GEOSYSTEMS	10.1002/2017GC006892	Volcanología y peligro volcánico
5	2012	BULLETIN OF VOLCANOLOGY	10.1007/s00445-012-0648-3	Volcanología y peligro volcánico
6	2011	BULLETIN OF VOLCANOLOGY	10.1007/s00445-010-0428-x	Volcanología y peligro volcánico
7	2016	BULLETIN OF VOLCANOLOGY	10.1007/s00445-016-1040-5	Volcanología y peligro volcánico
8	2017	JOURNAL OF QUATERNARY SCIENCE	10.1002/jqs.2976	Volcanología y peligro volcánico
9	2017	QUATERNARY SCIENCE REVIEWS	10.1016/j.quascirev.2017.05.011	Volcanología y peligro volcánico
10	2015	SCIENTIFIC REPORTS	10.1038/srep17589	Volcanología y peligro volcánico

Figura 15. Formato esperado de un archivo CSV parte 2.

	H	I	J
1	REGIÓN	COMUNAS	ENLACE
2	Los Lagos	Puerto Varas, Hualaihué, Cochamó	https://www.scielo.cl/pdf/rch/v35n1/art02
3	Los Lagos	Puerto Varas, Hualaihué, Cochamó	http://www.vliz.be/imisdocs/publications/2
4	La Araucanía, Los Ríos	Pucón, Villarrica, Panguipulli	https://discovery.ucl.ac.uk/id/eprint/10076
5	Los Lagos	Chaitén	https://doc.rero.ch/record/309646/files/44
6	Los Lagos	Chaitén	https://cdr.lib.unc.edu/downloads/w0892m
7	Los Lagos	Chaitén	http://pure-pai.bham.ac.uk/ws/files/28420
8	Los Lagos	Chaitén	https://repositorio.uchile.cl/bitstream/han
9	Los Lagos	Chaitén	https://repositorio.uchile.cl/bitstream/han
10	Los Lagos, Los Ríos	Lago Ranco, Río Bueno, Puyehue	https://www.nature.com/articles/srep1758

Figura 16. Formato esperado de un archivo CSV parte 3.

Con el formato del archivo CSV validado, el sistema procede a la inserción de datos en la base de datos utilizando el método insertDocuments() de la clase databaseConnector. Se itera sobre las filas del DataFrame generado, ejecutando consultas para verificar y agregar registros según sea necesario. La clase databaseConnector() gestiona la integridad referencial, asegurando que no se creen registros duplicados y que las relaciones en la base de datos se mantengan coherentes. Una vez completada la inserción, se emite un mensaje en el canal de Discord indicando que la información del archivo CSV se ha guardado correctamente en la base de datos.

Hasta este punto, la base de datos solo contiene información esencial de los documentos. Para almacenar el contenido de los documentos, se inicia una descarga en segundo plano de los documentos asociados a cada entrada en el archivo CSV. Mientras se lleva a cabo este proceso, se emite un mensaje en el canal de Discord indicando que se están descargando documentos y guardando información en segundo plano.

Se utiliza el método downloadDocument() de la clase Scraper; este método, luego de descargar un documento, utiliza el método readPdf() de la clase Tools para extraer el contenido y guardarlo en la base de datos mediante databaseConnector. Este proceso de descarga y almacenamiento del texto de los documentos en la base de datos se realiza mediante un manejo asíncrono utilizando la biblioteca asyncio. Esta implementación permite evitar bloquear el hilo principal del programa, permitiendo la ejecución simultánea de otras operaciones mientras se realiza este proceso. Durante la descarga, se manejan posibles errores y, tras completarse, el sistema informa al usuario sobre el resultado, además de notificar si algunos documentos no pudieron descargarse correctamente, detallando los títulos.

A continuación, se presenta una serie de imágenes que capturan la respuesta del bot al comando !addDocument en el canal de Discord. Estas ilustraciones que se pueden ver en la Figura 17, figura 18 y Figura 19 proporcionan una representación visual del proceso que incluye la verificación del formato del archivo CSV, la inserción de datos en la base de datos, la descarga asíncrona de documentos asociados.

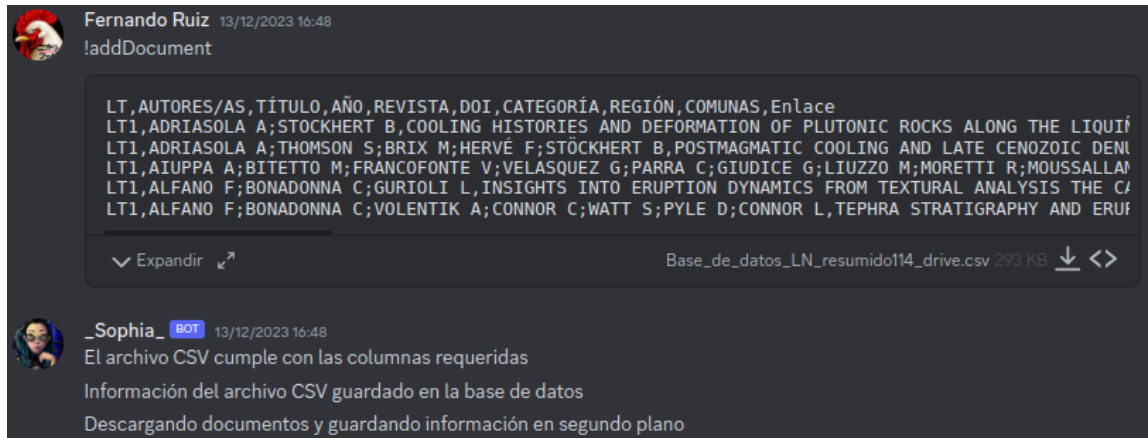


Figura 17. Ejecución del comando !addDocument

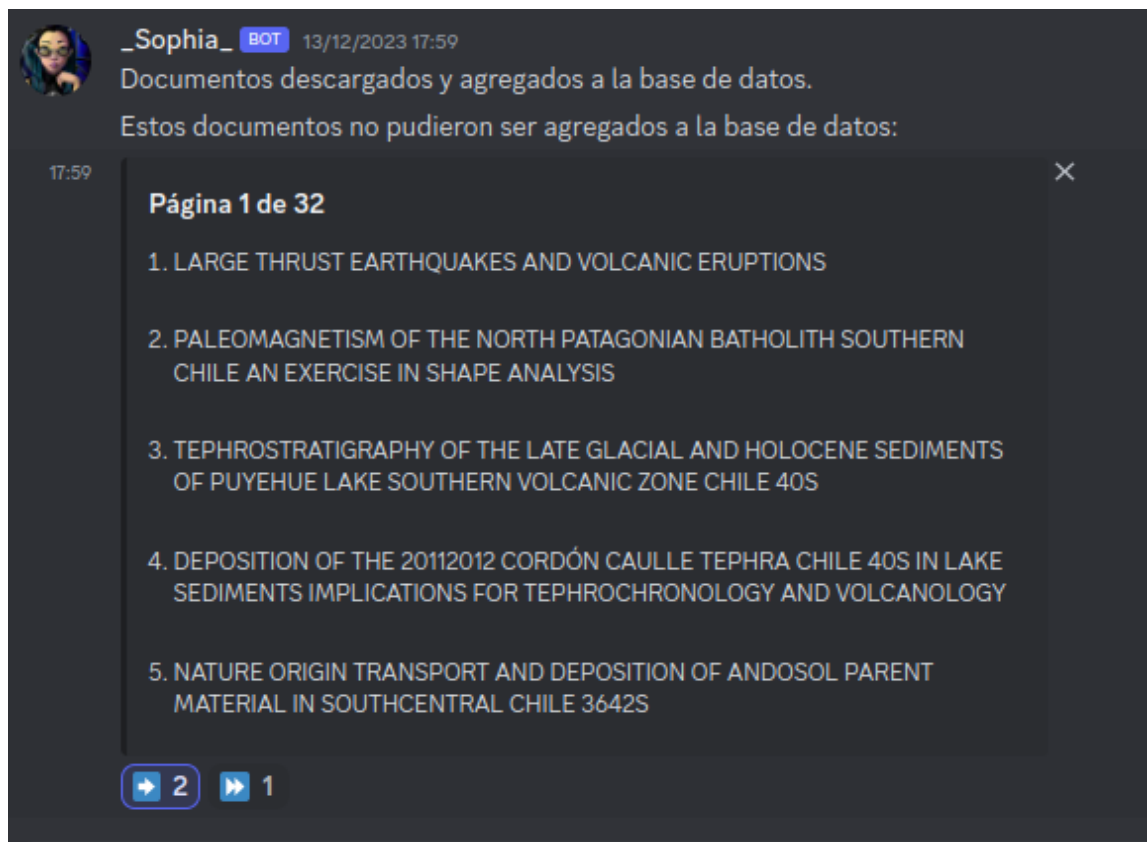


Figura 18. Bot respondiendo con los documentos que no pudieron ser ingresados

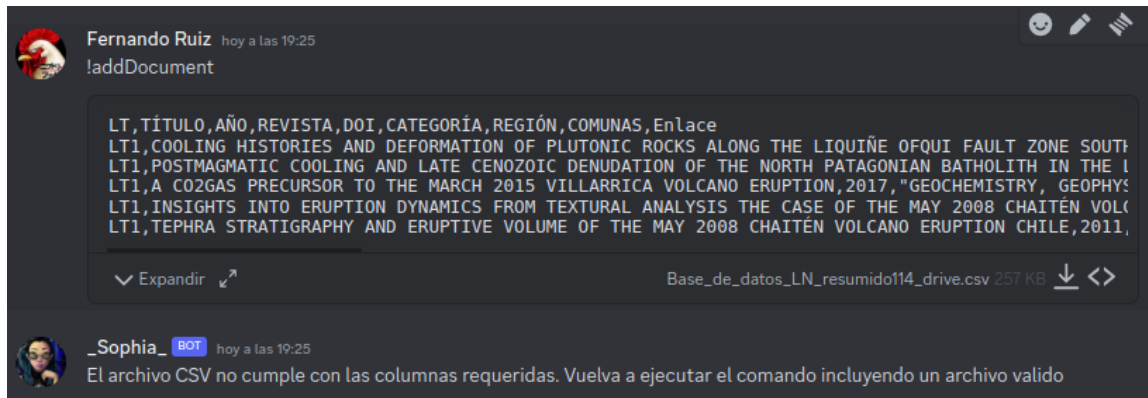


Figura 19. Bot respondiendo cuando se adjunta un archivo CSV invalido

4.4.2 Añadir un nuevo documento manualmente

El propósito del comando `!addManualDocument` es posibilitar que los usuarios ingresen manualmente datos relacionados con documentos en el canal de Discord. Este comando resulta útil cuando los usuarios desean ingresar documentos de manera individual sin depender de un archivo CSV adjunto, brindando una opción ágil para almacenar un documento en la base de datos.

El proceso inicia con la recepción de los datos ingresados manualmente. Estos datos, que incluyen el laboratorio temático, autores/as, título, año, revista, doi, categoría, región, comunas y enlace, constituyen la misma estructura de información que se recibe en el comando `!addDocument` a través de archivos CSV.

Se realiza una verificación para asegurar la coherencia en la entrada de datos. Si hay errores, se emite un mensaje en el servidor de Discord indicando el problema, alertando a los usuarios para corregir la entrada antes de proceder. Esta verificación garantiza que los datos sean adecuados y cumplan con la estructura esperada.

Después de verificar los datos, se procede a crear un `DataFrame` que organiza la información de manera estructurada. A continuación, se utiliza la clase `databaseConnector()` para establecer conexión con la base de datos y posteriormente insertar los datos en ella. El método principal para agregar la información estructural del documento a las tablas correspondientes se logra mediante el método `insertDocuments()`.

Una vez que la información se ha insertado en la base de datos se emite un mensaje en el servidor de Discord indicando que la información se ha guardado en la base de datos. Luego se aprovecha la misma lógica utilizada en `addDocument` para la fase de descarga en segundo plano de documentos utilizando la biblioteca `asyncio`. El método `downloadDocument()` de la clase `Scraper`, diseñada para gestionar descargas, se utiliza aquí para descargar el documento. Durante el proceso de descarga, se manejan posibles errores, como problemas de conexión o la indisponibilidad de la URL. Si ocurre un error

en la descarga del documento, se envía una notificación al usuario alertando que el documento no pudo ser descargado.

La ejecución de este comando se ve reflejado en la Figura 20.

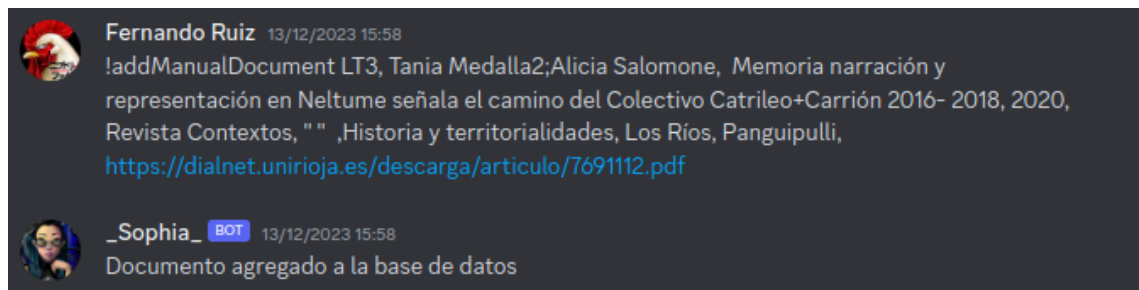


Figura 20. Ejecución comando !addManualDocument

4.4.3 Obtener indicadores a partir de los datos

La implementación del comando !analysis busca ofrecer a los usuarios una visión detallada de los datos presentes en la base de datos, abordando aspectos como la cantidad total de documentos, la evolución temporal de la producción, los autores más prolíficos, la distribución en laboratorios temáticos y las revistas destacadas.

Para obtener esta información, se utilizan diversas consultas SQL a través del método executeQuery() de la clase databaseConnector. La identificación de autores prolíficos se realiza mediante consultas basadas en tablas de relación entre documentos y autores.

La distribución de documentos en laboratorios temáticos se visualiza con gráficos de torta generados mediante la biblioteca Matplotlib en Python.

La consulta SQL para obtener las organizaciones más activas sigue un enfoque similar al utilizado para identificar autores prolíficos, pero ahora se realiza en la tabla de relación entre documento y organizaciones.

Se realiza una verificación para asegurar la coherencia en la entrada de datos. Si hay errores, se emite un mensaje en el servidor de Discord indicando el problema, alertando a los usuarios para corregir la entrada antes de proceder. Esta verificación garantiza que los datos sean adecuados y cumplan con la estructura esperada.

El análisis de la evolución de la producción de los documentos a lo largo de los años se presenta en un gráfico de barras utilizando nuevamente la biblioteca Matplotlib, representando años en el eje X y el número de documentos en el eje Y.

Todos estos resultados se organizan en páginas accesibles a través del canal de Discord, asegurando una presentación clara y gradual para facilitar la exploración por parte de los usuarios

En la Figura 21, Figura 22, figura 23, Figura 24 y Figura 25 se puede apreciar el resultado de ejecutar el comando !analysis.

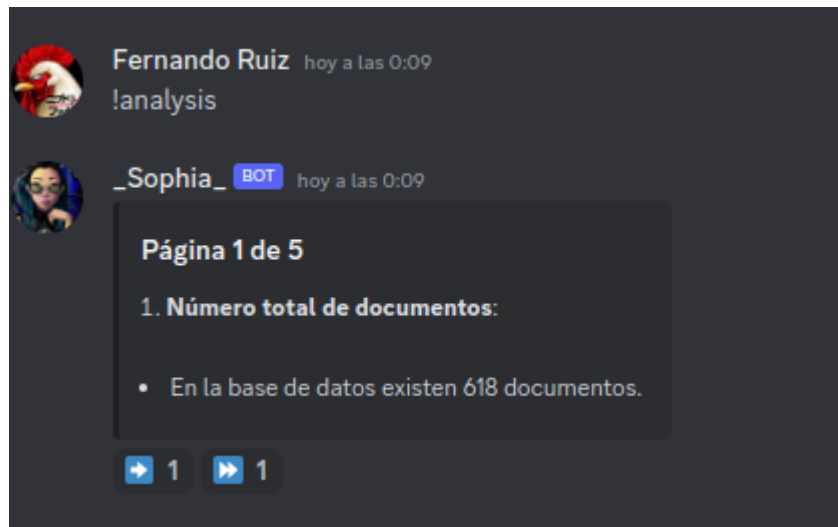


Figura 21. Bot mostrando el numero de documentos en la base de datos.

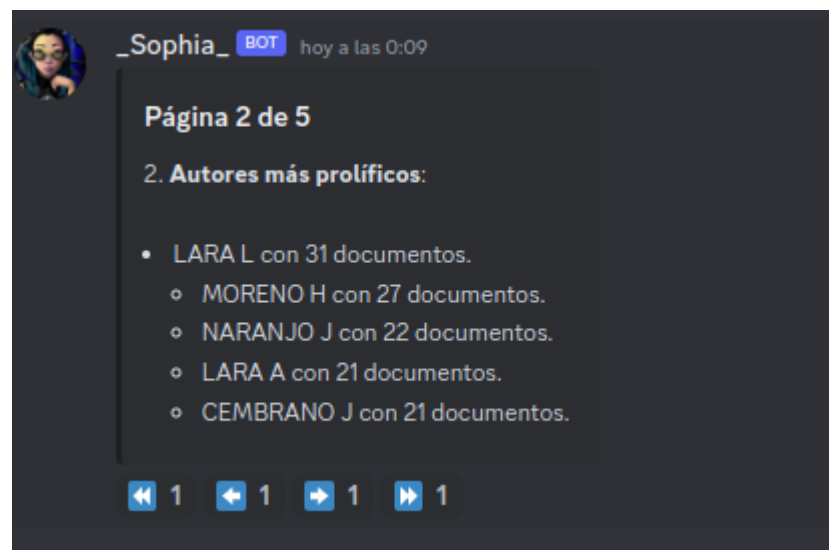


Figura 22. Bot mostrando los autores más prolíficos.



Figura 23. Bot mostrando la distribución de los documentos por laboratorio temático.

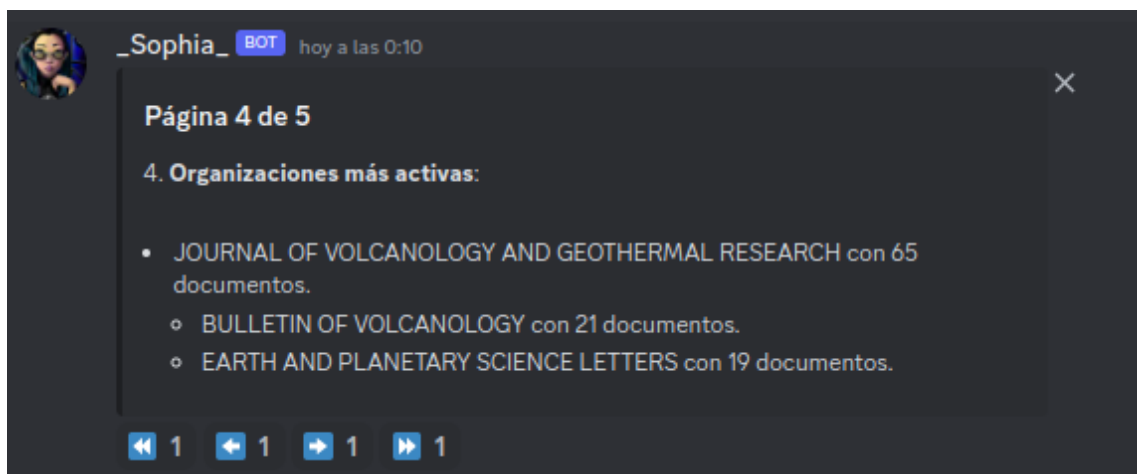


Figura 24. Bot mostrando las organizaciones más activas.



Figura 25. Bot mostrando la evolución de la producción de documentos por años

4.4.4 Obtener documentos a partir de parámetros

El comando !search ha sido desarrollado con el objetivo de permitir a los usuarios realizar búsquedas avanzadas de documentos académicos y científicos en el servidor de Discord. Su función principal es buscar los documentos presentes en la base de datos que cumplan con criterios específicos proporcionados por el usuario. Mediante la implementación de métodos y de las clases diseñadas, este comando procesa cuidadosamente los parámetros de búsqueda del usuario y ejecuta consultas SQL dinámicas

En la fase inicial, los parámetros de búsqueda proporcionados por el usuario se descomponen para analizar y estructurar la entrada, facilitando la identificación clara de criterios como lo son el año de publicación, categoría del documento, ciudad, región, laboratorio y palabras clave. En caso de que se asigne un parámetro que no existe o esté mal escrito, se envía un mensaje al canal de Discord informando al usuario sobre el error, solicitando verificar los parámetros.

Con los parámetros ya procesados, el sistema procede a construir dinámicamente una consulta SQL adaptada a las necesidades de búsqueda. Además, incorpora lógica para ordenar los resultados basándose en la relevancia, considerando la cantidad de coincidencias de palabras claves en el contenido del documento, en caso de que este parámetro haya sido ingresado.

Este comando, además, permite mezclar los parámetros que se ingresan, con el propósito de tener una mejor calidad de resultados

Por ejemplo, si el usuario ingresa los parámetros "year: 2015" y "laboratory: LT1", la consulta generada es la siguiente:

```
SELECT Document.title, Document.url FROM Document INNER JOIN
Document_Category ON Document.documentId =
Document_Category.documentId INNER JOIN Category ON
Document_Category.categoryName = Category.categoryName WHERE
Category.principalCategoryId = 'LT1' AND Document.publicationYear
= 2015;
```

En este ejemplo, la consulta busca documentos pertenecientes al laboratorio temático uno (LT1) y publicados en el año 2015

Los resultados obtenidos de la base de datos son procesados y presentados de manera clara en el canal de Discord mediante un sistema de paginación, facilitando la exploración

En la Figura 26, Figura 27, Figura 28, Figura 29, Figura 30, Figura 31 y Figura 32 se muestran varios ejemplos con distintos parámetros dados al comando !search.

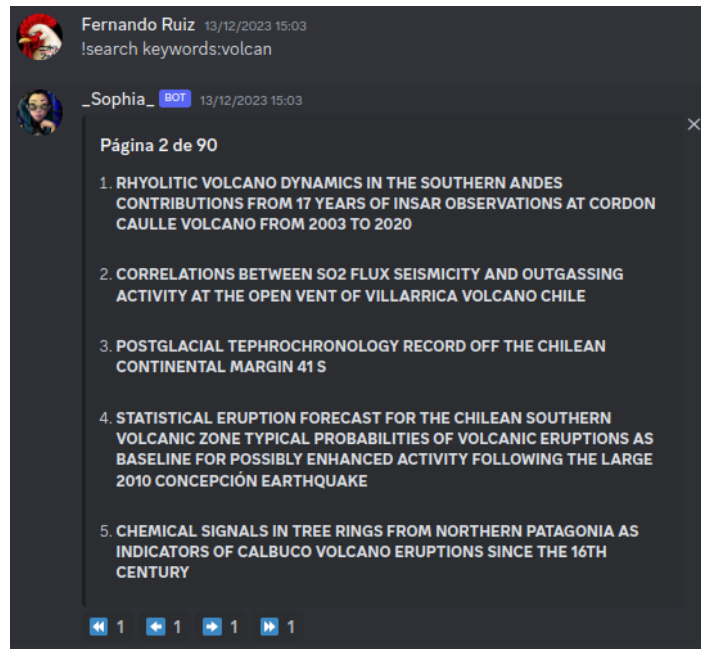


Figura 26. Bot mostrando resultados de ingresar volcán como palabra clave.

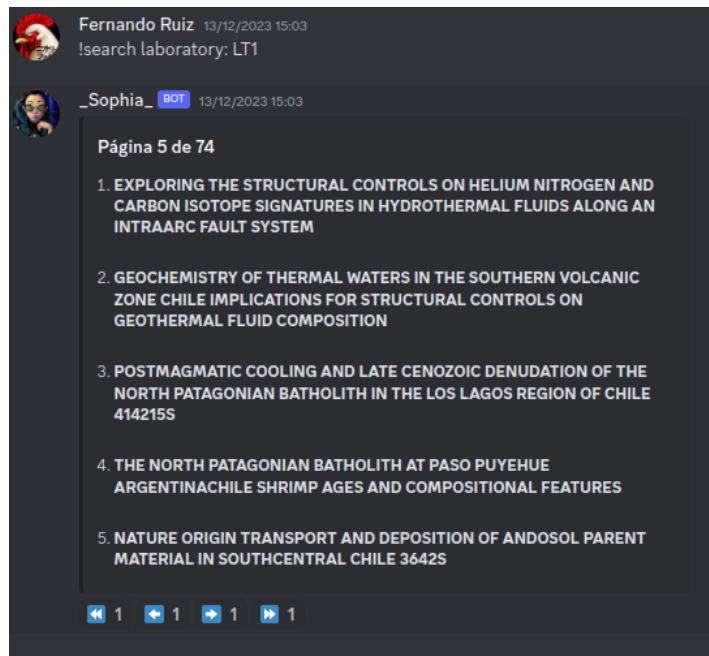


Figura 27. Bot mostrando los documentos que pertenecen al laboratorio temático 1.

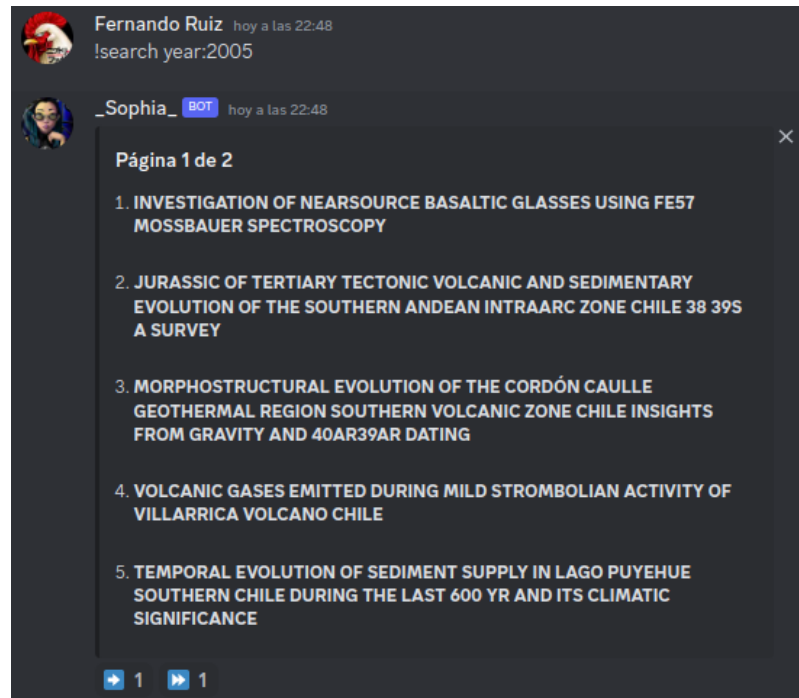


Figura 28. Bot mostrando los documentos que fueron publicanos el año 2005.

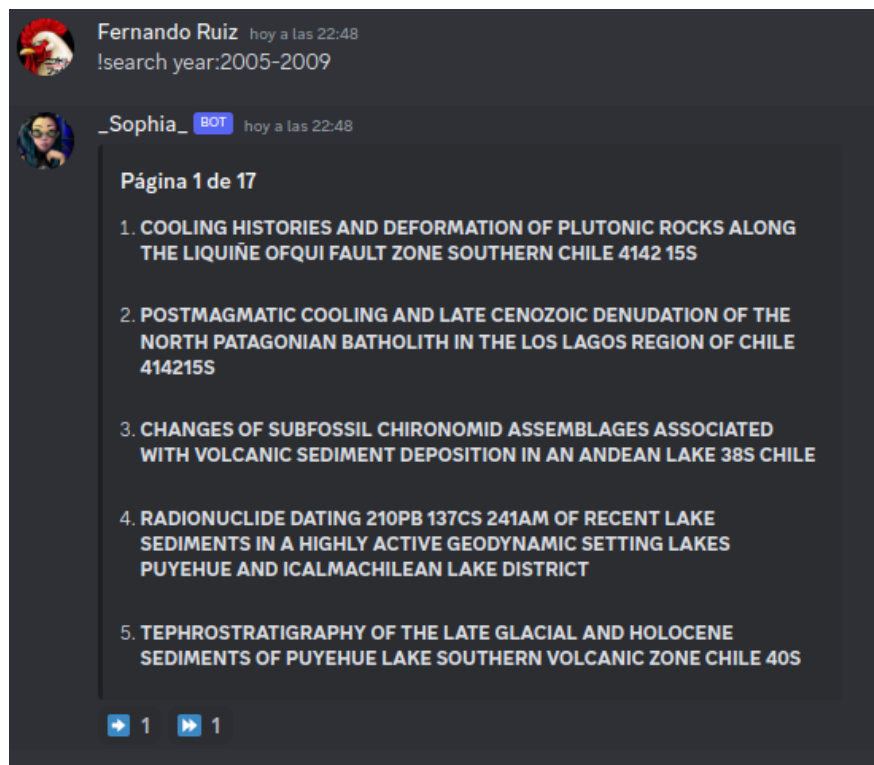


Figura 29. Bot mostrando documentos en un rango de años.

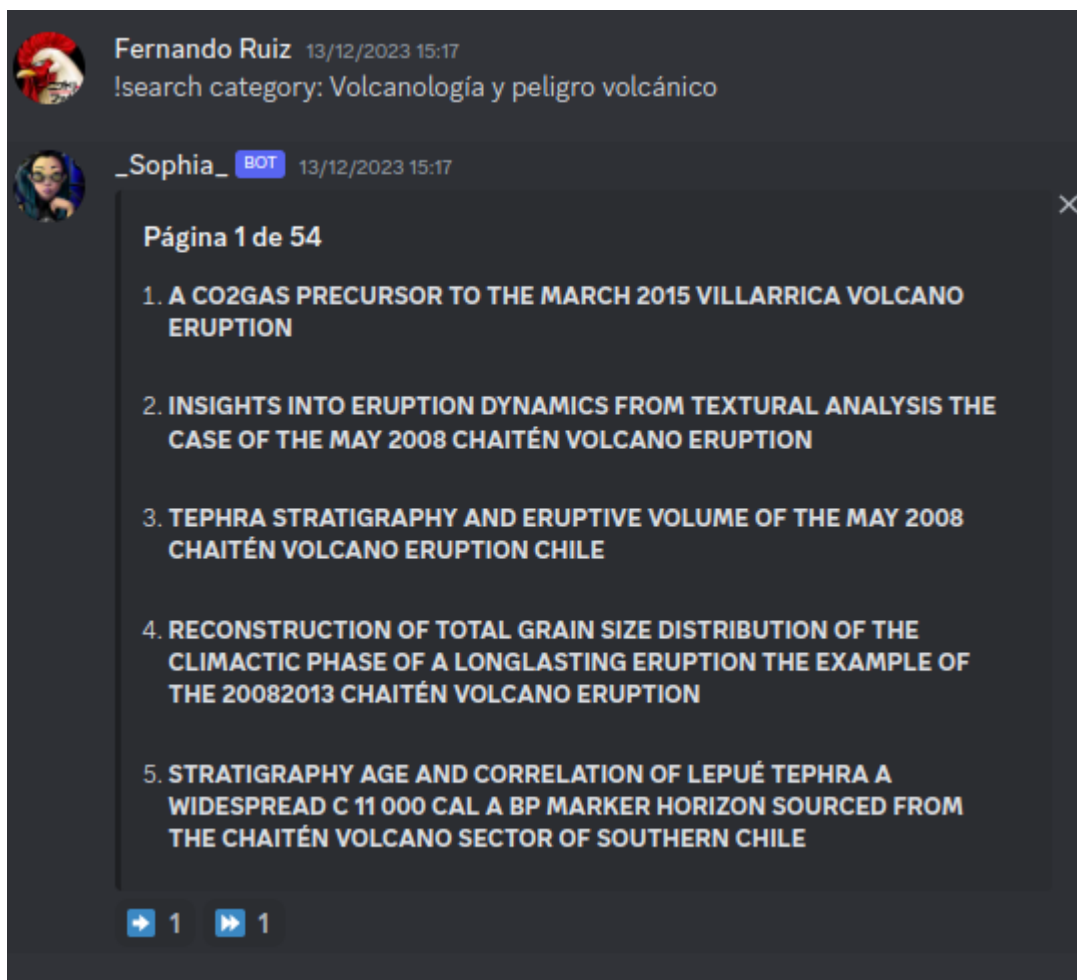


Figura 30. Bot mostrando documentos pertenecientes a una categoría.

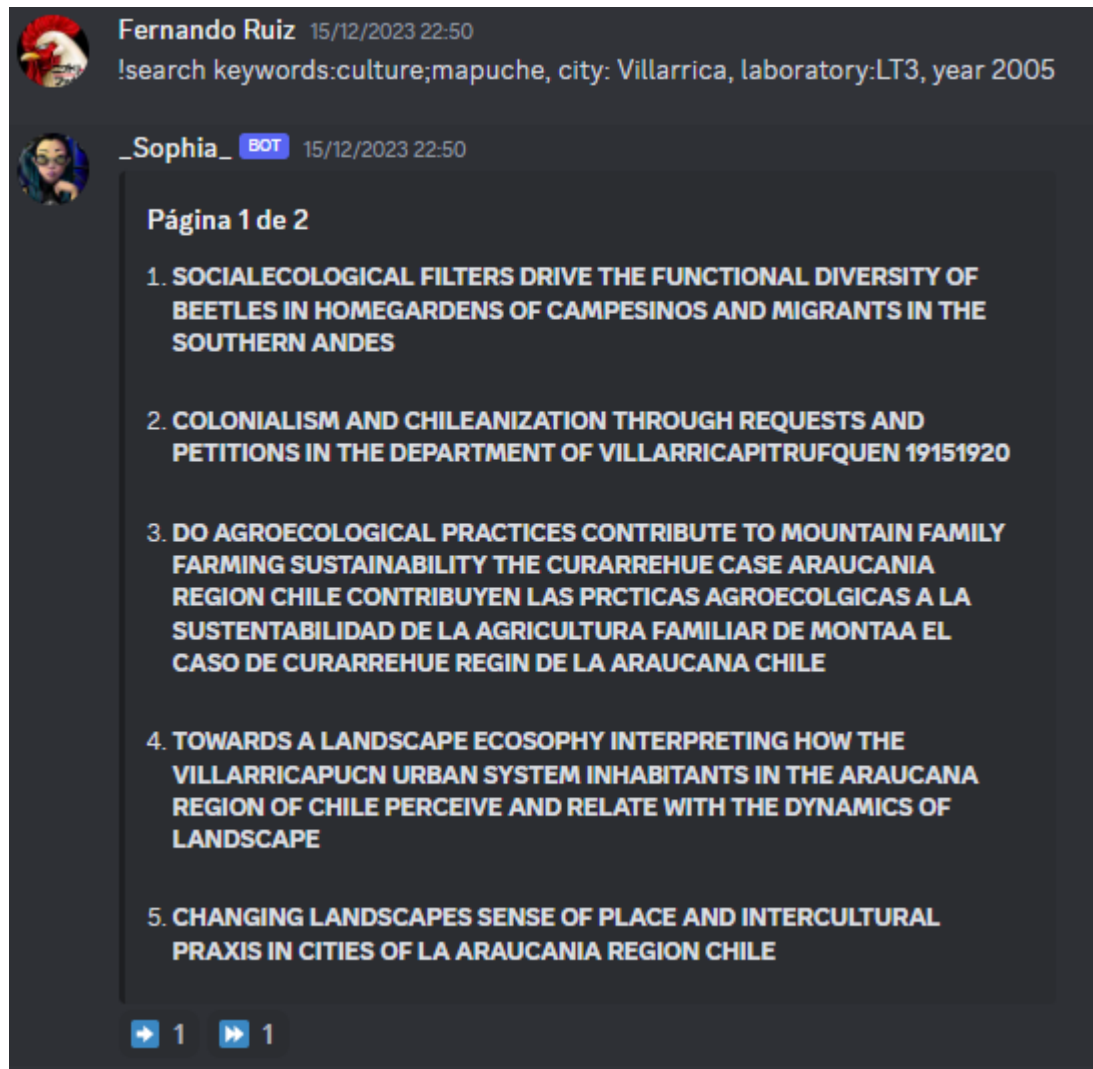


Figura 31. Bot mostrando documentos con una mezcla de parámetros.

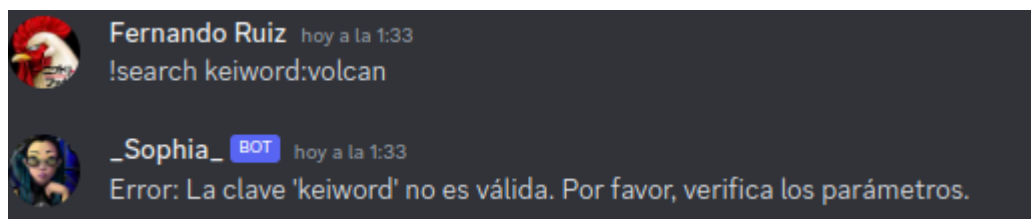


Figura 32. Bot mostrando un error.

4.4.5 Generar nuevo CSV con datos de cybertesis

Para construir el archivo CSV que alberga la producción científica relevante para el Laboratorio Natural Andes del Sur, el equipo del laboratorio realizó un proceso de búsqueda de documentos que hablen específicamente del territorio donde el laboratorio está ubicado. En este proceso se aplicaron *strings* de búsqueda diseñados específicamente para cada laboratorio temático. Estos *strings* fueron cuidadosamente elaborados considerando las palabras clave asociadas a cada área de interés y se aplicaron en plataformas como Scopus y Web of Science. La información de los documentos que cumplían con los criterios de búsqueda se descargó y, a continuación, se sometió a una revisión manual por parte de miembros del laboratorio.

Cada documento fue evaluado detenidamente para determinar su relevancia con respecto a los laboratorios temáticos establecidos. Este proceso, que combina la precisión de las búsquedas automatizadas utilizando *strings* de búsqueda en conjunto con la intervención humana en la revisión, culminó en el archivo CSV con el que el laboratorio cuenta.

Lo anteriormente descrito se realizó para obtener los documentos indexados en base de datos bibliográficas. Sin embargo, reconociendo el valor de las tesis universitarias como valiosas contribuciones al conocimiento científico, hasta el momento, el laboratorio no cuenta con una forma automatizada para obtener las tesis de la Universidad Austral de Chile (UACH). En este contexto, la implementación del programa de scraping tiene como objetivo proporcionar una solución para la adquisición automatizada de tesis universitarias de la UACH. Este scraping busca "imitar" la primera etapa realizada por los miembros del laboratorio natural, donde se obtienen documentos aplicando *strings* de búsqueda.

Es fundamental destacar que, al igual que con los documentos indexados, las tesis obtenidas mediante scraping también están sujetas a un proceso de revisión manual por parte del filtro humano

Navegación para Acceder a una Tesis en el Portal

Con el fin de comprender mejor el proceso de acceso a las tesis en el portal de la Universidad Austral de Chile, se presenta a continuación una descripción paso a paso:

1. El proceso de acceso a las tesis en el portal de la Universidad Austral de Chile se inicia en la página principal, disponible en la URL <http://cybertesis.uach.cl/>. Los usuarios pueden seleccionar la escuela de su interés en el menú desplegable. Como se aprecia en la Figura 33.



Este portal tiene como principal objetivo facilitar el acceso a las [Tesis en Texto Completo](#) de Pregrado y Postgrado que han sido desarrolladas por estudiantes de la Universidad Austral de Chile.

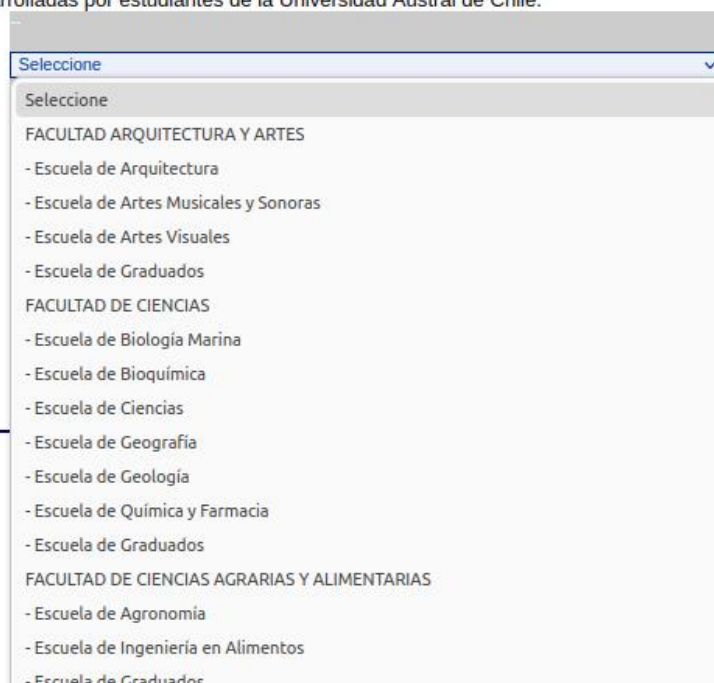



Figura 33. Página principal de cybertesis, con el menú desplegable.

2. Al elegir la escuela deseada, se despliega una nueva página que presenta una lista de tesis académicas asociadas a esa escuela, como se aprecia en la Figura 34.

Resultados de búsqueda

72 títulos encontrados para: **Licenciado Geografía Planificación Territorial OR Geógrafo AND FORMATO ELECTRÓNICO AND Facultad Ciencias**

Limited by: Tesis ^(x)

 SIBUACH

Ordenar por: Fecha Publicación ▾

Colección: Colección ▾

Siguiente

1. Transformación de los sistemas alimentarios desde una perspectiva histórica Caso de estudio en Comunidades mapuche-pehuenche de la comuna de Curarrehue (Tesis sólo en formato electrónico)

Por Marambio Aránguiz, Ashly Judith, Marchant Santiago, Carla , prof. guía, Universidad Austral de Chile. Facultad de Ciencias. Escuela de Geografía
Valdivia : UACH, 2022.
Clasif.: FC

Agregar a Mi Lista

2. Prácticas recientes de movilidad en el pericentro del Gran Santiago : El caso de la comuna de Renca. (Tesis sólo en formato electrónico)

Por Bustamante Bustos, Diego Felipe, Maturana Miranda, Francisco , prof. guía, Universidad Austral de Chile. Facultad de Ciencias
Valdivia : UACH, 2022.
Clasif.: FC

Agregar a Mi Lista

3. Percepción y gestión de riesgo volcánico : Incidencia de los volcanes Osorno y Calbuco en la localidad de Ensenada, comuna de Puerto Varas, Chile (Tesis sólo en formato electrónico)

Por Alegría Zúñiga, Camila Fernanda, Contreras Fuentes, Patricio , prof. guía, Vergara Pinto, Francisca , prof. co-patroc., Universidad Austral de Chile. Facultad de Ciencias.
Escuela de Geografía
Valdivia : UACH, 2022.
Clasif.: FC

Agregar a Mi Lista

4. Análisis comparativo de la composición florística y retención hídrica de dos tipos de turberas chilotas y su rol como reservorios de agua en la Cuenca del río Chepu, Región de Los Lagos, Chile (Tesis sólo en formato electrónico)

Figura 34. Lista de tesis relacionadas a una escuela.

3. Dentro de esta lista, los usuarios pueden hacer clic en la tesis de su interés, lo que abrirá una nueva página con la información detallada de la tesis, incluido el enlace de descarga del documento, como se puede apreciar en la figura 35.

Búsqueda

Mi Cuenta

Horario Bibliotecas

General

Avanzada

Historial de búsquedas

Búsqueda:

Palabra Clave Materia

ÓNICO AND .AW=Facultad Ciencias

Editar búsqueda

Volver a resultados

Siguiente

Catálogo Electrónico del Sistema de Bibliotecas UACH

Información del título

Solicitudes

Ver otros títulos del mismo autor

- Marambio Aránguiz, Ashly Judith
- Marchant Santiago, Carla , prof. guía
- Universidad Austral de Chile. Facultad de Ciencias. Escuela de Geografía

Materias

- COMUNIDADES INDÍGENAS -- CURARREHUE -- CHILE -- SISTEMA ALIMENTARIO -- EVOLUCION TECNOLÓGICA
- COMUNIDADES INDÍGENAS -- CURARREHUE -- CHILE -- ALIMENTACION -- HISTORIA
- HISTORIA AGRARIA -- MODERNIZACION -- CURARREHUE -- CHILE
- CURARREHUE -- IX REGION -- CHILE
- Tesis Geógrafo

Transformación de los sistemas alimentarios desde una perspectiva histórica Caso de estudio en Comunidades mapuche-pehuenche de la comuna de Curarrehue (Tesis sólo en formato electrónico)

Por Marambio Aránguiz, Ashly Judith, Marchant Santiago, Carla , prof. guía, Universidad Austral de Chile. Facultad de Ciencias. Escuela de Geografía

Valdivia : UACH, 2022.

Materias

- COMUNIDADES INDÍGENAS -- CURARREHUE -- CHILE -- SISTEMA ALIMENTARIO -- EVOLUCION TECNOLÓGICA
- COMUNIDADES INDÍGENAS -- CURARREHUE -- CHILE -- ALIMENTACION -- HISTORIA
- HISTORIA AGRARIA -- MODERNIZACION -- CURARREHUE -- CHILE
- CURARREHUE -- IX REGION -- CHILE
- Tesis Geógrafo

Descripción:

101 h.

Pub. Elec.:

31/07/2023

URL:

<http://cybertesis.uach.cl/tesis/uach/2022/fcm299t/doc/fcm299t.pdf>

Agregar a Mi Lista

Información de copia(s)

Clasificación	Estado	Colección	
FC M299t 2022	Revisando existencia	Tesis	Agregar a Mi Lista

Figura 35. Página con información de tesis específica.

Lo anterior describe la forma convencional en que los usuarios navegan para acceder a tesis. Ahora, para mejorar la eficiencia y automatizar este proceso, se ha implementado un sistema de scraping utilizando la clase TesisScraper. Este proceso se encarga de realizar las mismas acciones de manera automatizada, desde la obtención de enlaces de escuelas hasta la descarga y procesamiento de documentos PDF, como se detalla a continuación.

Para generar un nuevo CSV con datos de cybertesis, implementamos la clase TesisScraper en el componente del sistema de recopilación automático. Esta clase se vale de diversas bibliotecas, como requests para las solicitudes HTTP, BeautifulSoup (beautifulsoup4) para el análisis estructurado del HTML, PyPDF2 para la manipulación de archivos PDF, y spacy con el modelo de lenguaje español para el procesamiento de texto.

El método getSchoolLinks() se enfoca en obtener enlaces que dirigen a las diferentes escuelas desde el portal <http://cybertesis.uach.cl/>. El proceso inicia con la realización de solicitudes HTTP para acceder a la página principal del sitio. Una vez en la página principal, se procede a identificar el menú desplegable y, a partir de este, se extraen los enlaces asociados a las escuelas de interés. De esta manera, se establece un mecanismo automatizado para recopilar y catalogar enlaces específicos que conducen a las distintas escuelas dentro del portal.

Posteriormente, la labor del método getSchoolThesisLinks() consiste en recuperar los enlaces que conducen a las tesis vinculadas a una escuela específica. Este procedimiento implica realizar solicitudes HTTP a la página correspondiente a la escuela, examinar el contenido HTML en busca de los enlaces a las tesis y, mediante el uso de expresiones regulares, identificar patrones que faciliten la extracción de información del código HTML de la página web. La funcionalidad del método permite recorrer todas las páginas, garantizando así la obtención completa de los enlaces de tesis asociadas a la escuela.

Luego, se opera sobre cada enlace específico mediante el método extractThesisInfo(). Se realizan solicitudes HTTP para obtener la página de la tesis, donde se extraen datos clave como el título, autor y año de publicación. Además, se busca el enlace al archivo PDF asociado y se llama al método downloadAndSavePdf() para la descarga y análisis del contenido del PDF.

En el ámbito de la descarga, downloadAndSavePdf() administra la obtención del archivo PDF vinculado a una tesis específica, evaluando si cumple con condiciones particulares. Este proceso se basa en la identificación de palabras clave presentes en las primeras 15 páginas del PDF, donde se encuentra el resumen de las tesis. Estas palabras clave fueron configuradas mediante *matchers* de spaCy.

Estos *matchers* se establecieron utilizando conjuntos de palabras clave adaptadas y ajustados a partir de los strings de búsqueda originales utilizados por el laboratorio durante la fase inicial, uno de estos strings de búsqueda se puede apreciar en la Figura 36. Dichos ajustes incluyeron cambios, como la traducción a nombres en español para mejorar la efectividad de la búsqueda. Un ejemplo de un conjunto de palabras clave ajustadas se presenta en la Figura 37.

Palabras claves Laboratorio Natural Andes del Sur:

LT1: Evolución y Herencia Geológica

Chile AND ((Southern AND Andes) OR Lonquimay OR Llaima OR Sollipulli OR Villarrica OR Lanin OR Quetrupillan OR (Mocha AND Choshuencho) OR (Carran AND Los AND Venados) OR Puyehue OR (Cordon AND Caulle) OR Antillanca OR Qsomo OR Calbuco OR Yate OR Hornopiren OR Huegui OR Michinmahuida OR Chaiten OR Corcovado) AND (Geology* OR orogenesis* OR subduction* OR geodynamics* OR magmatism* OR erosion* OR geological hazards OR volcanism* OR paleontology* OR sedimentology* OR geochemistry* OR petrology* OR geomorphology* OR stratigraphy* OR mudflow* OR landslide* OR earthquake* OR fault* OR geothermal* OR energy* OR entalphy* OR fossil* OR eruption* OR volcano* OR tephra* OR tephrochronology* OR (Liquiñe AND Ofqui) OR fjord* OR geoheritage* OR geosite* OR geotourism*)

Figura 36: String de búsqueda.

```
secondPartKeywordsLT1 = ["Geología", "orogénesis", "subducción", "geodinámica",  
"magmatismo", "erosión", "riesgos geológicos", "volcanismo",  
"paleontología", "sedimentología", "geoquímica", "petrología",  
"geomorfología", "estratigrafía", "flujo de lodo", "deslizamiento",  
"terremoto", "falla", "geotermia", "energía", "entalpía", "fósil",  
"erupción", "volcán", "tefra", "tefrocrónología", "Liquiñe-Ofqui",  
"fiordo", "geopatrimonio", "sitio geológico", "geoturismo"]
```

Figura 37: Palabras claves adaptadas de un *string* de búsqueda.

Los *matchers* actúan como filtros que identifican la presencia de palabras clave específicas en el texto del PDF. Dependiendo de las condiciones satisfechas, la tesis se asigna a un laboratorio temático específico. Este enfoque permite una búsqueda adaptada a los requisitos del laboratorio.

Para concluir, el método `writeThesisDataToCsv()` estructura la información recolectada y la organiza en un archivo CSV. Finalmente, cuando es creado este archivo CSV, se envía al canal de Discord para presentárselo al usuario.

Este archivo estructurado incluye columnas como título, autor, año, comunas, regiones y enlace, proporcionando así un registro detallado para que los miembros del laboratorio puedan iterar sobre este documento y realizar un filtro manual de la información. Una vez validado el documento, el usuario puede ingresar el archivo a la base de datos utilizando el comando `!addDocument` para que la información esté disponible para su utilización en las distintas funcionalidades que ofrece el bot.

En la Figura 38 se puede observar el comportamiento del bot al llamar el comando `!getThesis`.

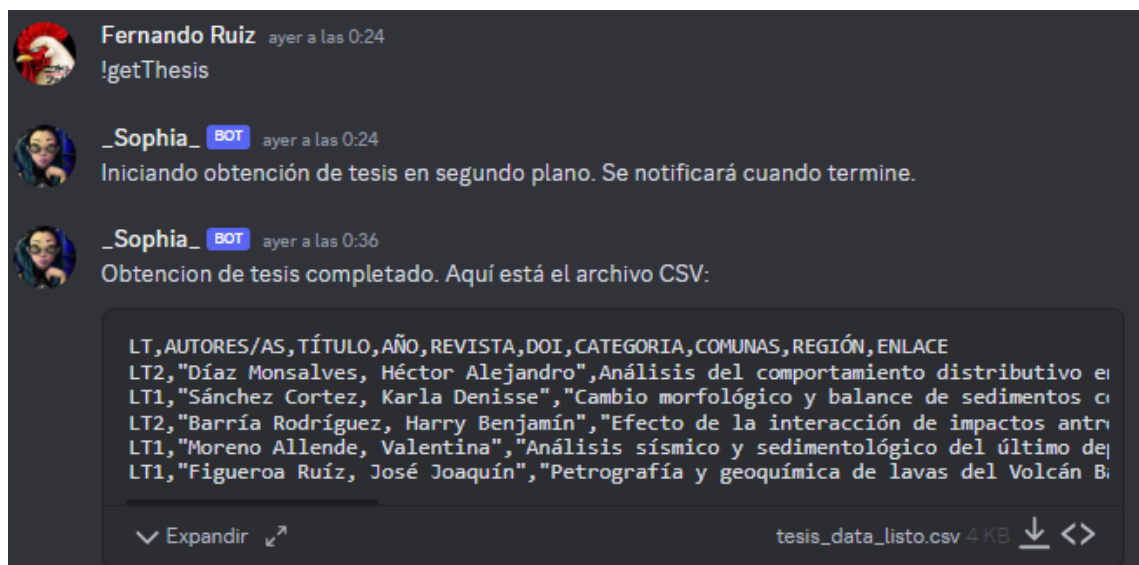


Figura 38. Bot mostrando el resultado de el comando `!getThesis`.

4.5 Flujo de trabajo durante la implementación

Se estableció un flujo de trabajo utilizando un repositorio en GitHub como punto central para el almacenamiento y gestión del código fuente del proyecto. En dicho repositorio se encuentra el código fuente, actuando como el lugar principal para el desarrollo y la gestión. Este enfoque facilita la colaboración, el seguimiento de cambios y la eficiente gestión del desarrollo, permitiendo a los colaboradores acceder y contribuir al código de manera coordinada.

En la rutina diaria, se avanzó en el desarrollo de las funcionalidades. Tras la finalización de una tarea, se ejecutó un "push" hacia el repositorio de GitHub para consolidar los cambios. En el servidor que hospeda el bot de Discord, se utilizó el comando "git pull" para actualizar el código, incorporando las últimas modificaciones. Este enfoque facilitó la sincronización del código.

El servidor que hospeda el bot opera bajo el sistema Linux con el kernel 5.15. El procesador, un AMD EPYC 7713 64-Core Processor, destaca por su arquitectura de 64 bits, estableciendo la base del rendimiento del entorno de ejecución. El almacenamiento en disco ofrece una capacidad total de 25 GB. El sistema de memoria RAM cuenta con 957 MB.

Para asegurar la disponibilidad continua del bot en el servidor, se utiliza el siguiente comando en el terminal del servidor donde este alojado el bot: `nohup python3 bot.py`.

Esta instrucción inicia la ejecución del bot, desencadenando la comunicación con Discord y activando todas las funcionalidades implementadas. Este enfoque asegura que el bot

permanezca activo y disponible para los usuarios en el servidor de Discord, proporcionando una experiencia continua y sin interrupciones.

5. REUNIÓN CON CONAF

Uno de los hallazgos clave de la revisión sistemática fue la importancia de la colaboración entre los actores territoriales para lograr una gestión efectiva del conocimiento del territorio, aspecto que se tiene en cuenta en la Plataforma Conversa con Los Andes.

En línea con esta premisa, se llevó a cabo una sesión con el equipo de CONAF Valdivia con el propósito de exponer el funcionamiento de las funciones clave de la plataforma. A lo largo de esta reunión, se ilustraron de manera práctica las características fundamentales del sistema, ofreciendo una visión detallada de su operación.

Esta interacción directa posibilitó que todos los participantes comprendieran más a fondo el potencial del proyecto, generando un espacio propicio para abordar discusiones constructivas acerca de posibles colaboraciones y adaptaciones del sistema para satisfacer las necesidades específicas de CONAF Valdivia. La disposición y actitud positiva mostradas por los representantes de CONAF durante la demostración resaltaron la utilidad percibida de la plataforma y establecieron las bases para la exploración de futuros vínculos colaborativo. En las Figura 39 y Figura 40 se pueden observar imágenes de la reunión que tuvo lugar.



Figura 39. Reunión llevada a cabo en CONAF.



Figura 40. Twitter de CONAF informando de la reunión.

6. ENCUESTA AL ACTOR DEL LABORATORIO

Con el objetivo de asegurar la efectividad de las funciones implementadas, se llevó a cabo una fase de validación mediante la realización de una encuesta dirigida a un actor clave del Laboratorio Natural Andes del Sur. El objetivo principal de esta encuesta fue obtener retroalimentación directa y específica sobre las funcionalidades propuestas, así como evaluar su utilidad y eficacia, según la percepción de aquel que estará directamente involucrado en su uso.

Con el propósito de optimizar el proceso de evaluación y obtener una perspectiva integral, se realizó una encuesta unificada dirigida a un actor que desempeña las funciones de tres roles dentro del laboratorio natural. Este actor, representando simultáneamente al Científico, Data Analyst y Data Manager, participó en la evaluación de las funcionalidades de la aplicación.

6.1 Diseño de la encuesta

La encuesta se diseñó generando preguntas de alternativa para cada funcionalidad implementada en la plataforma. Cada pregunta abordó aspectos como la facilidad de uso, la claridad de los resultados y, además, se incluyó una sección de pregunta abierta para recibir comentarios detallados y observaciones sobre cada comando probado.

6.2 Resultados de la encuesta por funcionalidad

Los resultados de aplicar la encuesta se pueden apreciar en la Figura 41, Figura 42, Figura 43, Figura 44, Figura 45, Figura 46, Figura 47, Figura 48 y la Figura 49.

¿Cómo calificarías la facilidad de uso del comando !search?

1 respuesta



¿La información devuelta por el comando !search es clara y fácil de interpretar?

1 respuesta

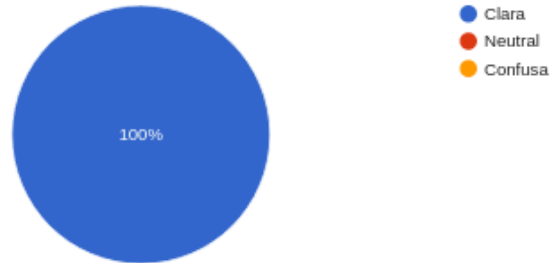
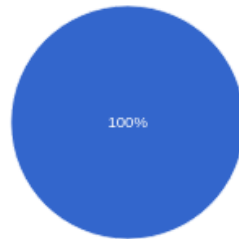


Figura 41. Resultado de encuesta comando !search

¿El comando !search devuelve resultados relevantes según los parámetros proporcionados?

 Copiar

1 respuesta



- Sí, los resultados son relevantes
- Sí, en su mayoría, pero a veces los resultados no son tan relevantes.
- No, los resultados no son precisos ni relevantes la mayoría de las veces.

¿Hay alguna observación o comentario que te gustaría compartir sobre tu experiencia con el comando !search?

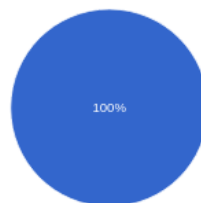
1 respuesta

Es importante la aclaración de que usando el comando buscando por palabras claves, los primeros documentos que muestra son los que mencionan más esa palabra clave. También el hacer la búsqueda en español e inglés.

Figura 42. Resultado de encuesta comando !search parte 2.

¿La información devuelta por el comando !analysis es clara y fácil de interpretar?

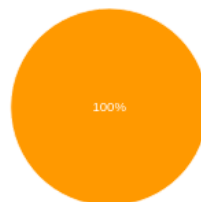
1 respuesta



- Clara
- Neutral
- Confusa

¿Cómo calificarías la facilidad de uso del comando !analysis?

1 respuesta



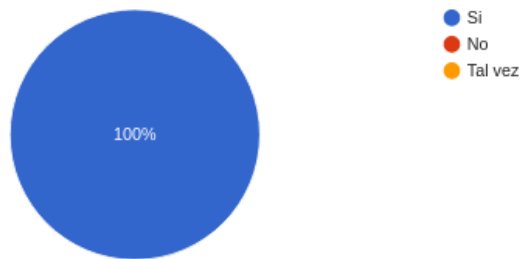
- Difícil
- Normal
- Fácil

Figura 43. Resultado de encuesta comando !analysis.

¿Consideras que los resultados del comando !analysis son de utilidad?

 Copiar

1 respuesta



¿Hay alguna observación o comentario que te gustaría compartir sobre tu experiencia con el comando !analysis?

1 respuesta

Cambio del título resultado: "las organizaciones más activas" en realidad son las revistas donde más se publica

Figura 44. Resultado de encuesta comando !analysis parte 2.

¿Cómo calificarías la facilidad de uso del comando !addDocument?

1 respuesta



¿La información devuelta por el comando !addDocument es clara y fácil de interpretar?

1 respuesta

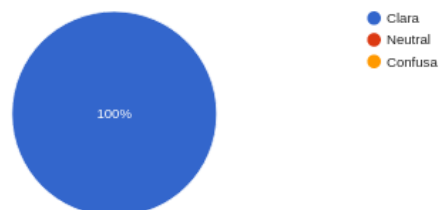


Figura 45. Resultado de encuesta comando !addDocument

¿Hay alguna observación o comentario que te gustaría compartir sobre tu experiencia con el comando !addDocument?

1 respuesta

Fácil de usar y con respuestas claras, sólo tendremos que trabajar en los posibles errores al ingresar los documentos.

Figura 46. Resultado de encuesta comando !addDocument parte 2.

¿Cómo calificarías la facilidad de uso del comando !addManualDocument?

 Copiar

1 respuesta



¿Hay alguna observación o comentario que te gustaría compartir sobre tu experiencia con el comando !addManualDocument?

1 respuesta

El comando !addManualDocument es fácil de usar, solo hay que saber bien "la formula" para ingresar el documento (uso de las comas o punto y comas para separar la información)

Figura 47. Resultado de encuesta comando !addManualDocument.

¿Cómo calificarías la facilidad de uso del comando !getTesis?

 Copiar

1 respuesta



Qué opinas sobre la manera en que se devuelven los resultados de las tesis en el archivo CSV? ¿Encuentras que la presentación es clara y fácil de comprender?

 Copiar

1 respuesta

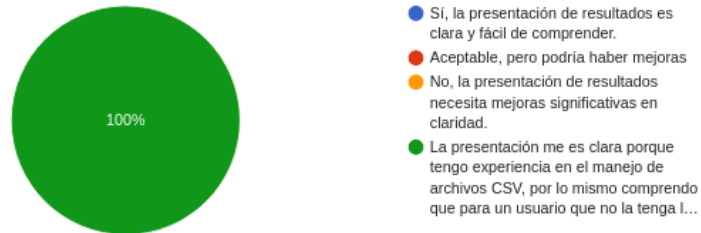


Figura 48. Resultado de encuesta comando !getThesis.

A primera vista, ¿consideras que las tesis seleccionadas podrían ser de utilidad para el laboratorio?

 Copiar

1 respuesta



¿Hay alguna observación o comentario que te gustaría compartir sobre tu experiencia con el comando !getTesis?

1 respuesta

Es necesario este filtro humano antes de asegurar de que sean pertenecientes a información del territorio del Laboratorio Natural

Figura 49. Resultado de encuesta comando !getThesis parte 2.

6.3 Acciones post encuesta

Basándonos en los resultados obtenidos de la encuesta, se implementaron ajustes y mejoras específicas en la plataforma para abordar las sugerencias y optimizar la experiencia del usuario. La retroalimentación recopilada se consideró invaluable para la iteración y el perfeccionamiento continuo del proyecto.

Para el comando `!analysis` y como se menciona en la Figura 44, se cambió el título de organizaciones más activas a revistas que han publicado la mayor cantidad de documentos como se puede ver en la Figura 50.

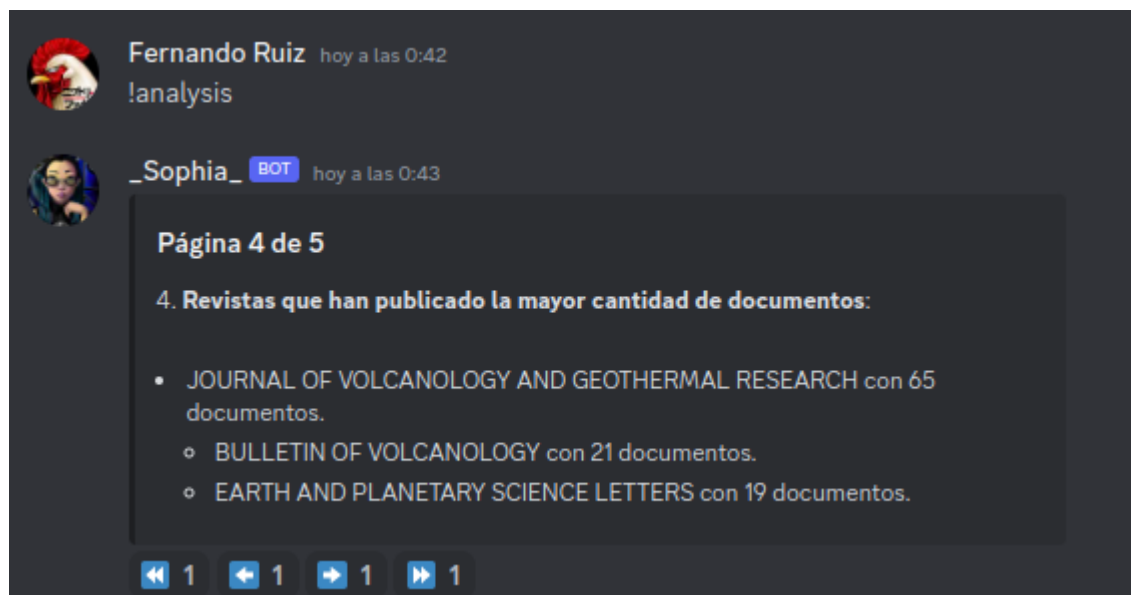


Figura 50. Cambio de título.

Para el comando `!search`, el bot ahora menciona que los resultados al ingresar keywords se ordenan por coincidencia de palabra clave, además que se pueden realizar búsqueda utilizando *keywords* tanto en inglés como en español, como se puede apreciar en la Figura 51, todo esto en base a lo dicho en la encuesta en la Figura 42.

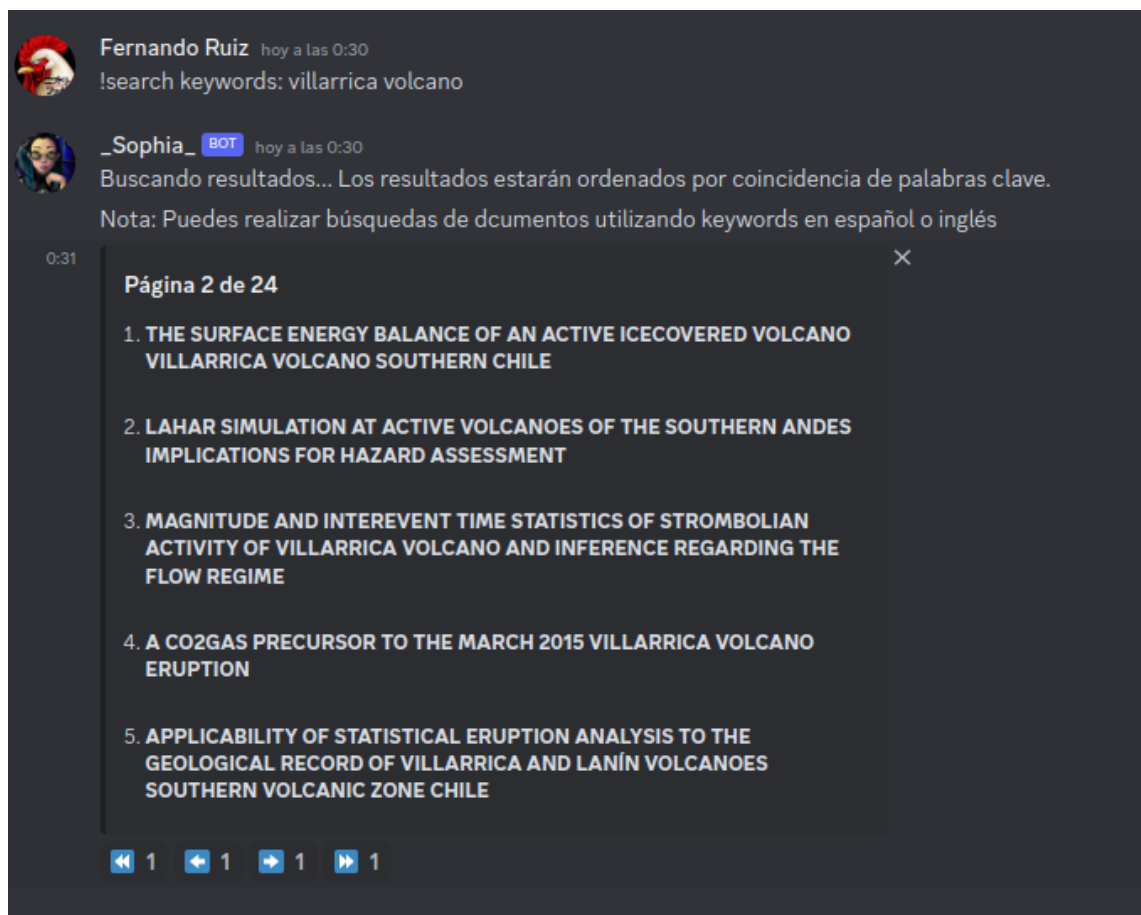


Figura 51. Modificaciones comando search.

7. CONCLUSION

La implementación exitosa de las funcionalidades propuestas ha permitido alcanzar los objetivos establecidos para el proyecto, orientados a mejorar el acceso a conocimientos científicos y técnicos sobre el territorio de Los Andes del Sur. A continuación, se presenta un análisis crítico de los resultados obtenidos en relación con cada objetivo específico.

Diseño e implementación de una base de datos: La base de datos diseñada e implementada satisface la necesidad fundamental de almacenar información científica y técnica sobre el territorio de los Andes. Su estructura posibilita la gestión eficiente de los documentos identificados por el laboratorio,

Desarrollo de algoritmos de web scraping: El objetivo específico de "Desarrollo de algoritmos de web scraping" se ha logrado mediante la automatización y recopilación de datos de Cybertesis. En este proceso, se extrae y procesa información clave relacionada con las tesis para identificar documentos que podrían ser útiles al laboratorio a través de un archivo CSV. Todo esto se ha implementado en la funcionalidad de 'Generar nuevo CSV con datos de Cybertesis'. Es importante destacar que se debe aplicar un filtro humano para garantizar que los documentos identificados en esta funcionalidad cumplan al cien por ciento con los requisitos necesarios para poder ser utilizados por los miembros del laboratorio.

Integración de funcionalidades avanzadas al bot de Discord: El comando !search brinda a los usuarios la capacidad de realizar búsquedas avanzadas, permitiendo criterios múltiples y proporcionando resultados claros y estructurados, el comando !addManualDocument proporciona una alternativa ágil y flexible para ingresar documentos a la base de datos, al igual que el comando !addManualDocument,

Implementación de Funcionalidad de Análisis Bibliométrico: Este objetivo específico, que surge de la necesidad de los miembros del laboratorio de acceder y comprender los documentos almacenados, se materializa a través del comando !analysis. Esta herramienta es esencial para obtener indicadores bibliométricos detallados, facilitando la comprensión y exploración profunda de la base de datos.

Se logró cumplir el objetivo principal al crear una plataforma que facilita el acceso a conocimientos científicos y técnicos sobre Los Andes del Sur. La implementación exitosa de la base de datos, los algoritmos de web scraping y las funciones mejoradas en Discord permite a los miembros del laboratorio hacer uso efectivo de esta plataforma.

6. REFERENCIAS

Asyncio. (s.f.). Recuperado el 15 de diciembre, 2023, desde <https://docs.python.org/3/library/asyncio.html>

Amazon Web Services. Accedido el 15 de Diciembre, 2023, desde <https://aws.amazon.com/es/what-is/python/>

Beautiful Soup. (s.f.). Recuperado el 15 de diciembre, 2023, desde <https://pypi.org/project/beautifulsoup4/>

Bertrán, A. (2021,). UACH adjudicó Nodo de Laboratorios Naturales ANID 2021. Accedido el 1 de junio, 2023, desde <https://diario.uach.cl/uach-adjudico-nodo-de-laboratorios-naturales-anid-2021/>

Corchado, J. M., Chamoso, P., Hernández, G., Gutierrez, A. S. R., Camacho, A. R., González-Briones, A., Pinto-Santos, F., Goyenechea, E., Garcia-Retuerta, D., Alonso-Miguel, M., Hernandez, B. B., Villaverde, D. V., Sanchez-Verdejo, M., Plaza-Martínez, P., López-Pérez, M., Manzano-García, S., Alonso, R. S., Casado-Vara, R., Tejedor, J. P. (2021). Deepint.net: A Rapid Deployment Platform for Smart Territories. *Sensors*, 21(1), 1-23.

Discord.py. (s.f.). Recuperado el 15 de Diciembre, 2023, desde <https://discordpy.readthedocs.io/en/stable/index.html>

Flores, J. P. (2012). Diagnóstico nacional de montaña. Food and Agriculture Organization of the United Nations. Accedido el 18 de Mayo, 2023 desde <https://www.fao.org/3/CA0124ES/ca0124es.pdf>

Foronda-Robles, C., & Galindo-Perez-de-Azpillaga, L. (2021). Territorial intelligence in rural areas: The digitization of non-profit associations through social media. *Technology in Society*, 64, 1-9.

Garcia-Madurga, M.-A., Grillo-Mendez, A.-J., & Esteban-Navarro, M.-A. (2020). Territorial Intelligence, a Collective Challenge for Sustainable Development: A Scoping Review. *Social Sciences-Basel*, 9(7), 1-8.

GitHub. (s.f.). Recuperado el 15 de diciembre, 2023, desde <https://docs.github.com/es/get-started/quickstart/hello-world>

IVarona-Taborda, M.-A., Mosquera-Ramirez, J.-C., Medina-Moreno, C.-A., Lemus-Munoz, D.-F., Munoz-Hernandez, C.-J., & Arias-Iragorri, C.-G. (2021). Business Intelligence for the Programs of the Secretaries of Health, Education and Planning in a

Territorial Entity. Revista Facultad de Ingenieria, Universidad Pedagogica y Tecnologica de Colombia, 30(58), 1-14

Laboratorio Natural Andes del Sur. (2022), Laboratorio Natural Andes del Sur de Chile: El desafío de recopilar información científica en montañas únicas en el mundo. Accedido el 1 de junio, 2023 desde <https://laderasur.com/articulo/laboratorio-natural-andes-del-sur-de-chile-el-desafio-de-recopilar-informacion-cientifica-en-montanas-unicas-en-el-mundo/>

Proyecto – Andes del Sur de Chile. (2022,). Laboratorio Natural Andes del Sur. Accedido el 6 de junio, 2023 desde <https://andesdelsurlab.cl/proyecto/>

spaCy. (s.f.). Recuperado el 15 de diciembre, 2023, desde <https://medium.com/datos-y-ciencia/comenzando-con-spacy-para-procesamiento-de-lenguaje-natural-e8cf24a18a5a>