



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 BASADO EN UN CHAT INTELIGENTE SOBRE CONOCIMIENTOS CIENTÍFICOS Y TÉCNICOS VINCULADOS A LOS ANDES DEL SUR

Proyecto para optar al título de
Ingeniero Civil en Informática

PROFESOR PATROCINANTE:

DR. MATHIEU VERNIER

LICENCIADO EN CIENCIAS Y TECNOLOGIAS

MASTER EN CS Y TECNOLOGIA

DOCTOR EN INFORMÁTICA

PROFESOR INFORMANTE:

M.Sc. MARIANA VILLARROEL MANFREDI

INGENIERO CIVIL INFORMÁTICO

MASTER EN EDUCACIÓN MENCIÓN EVALUACIÓN Y CURRÍCULUM

BASADO EN COMPETENCIAS

PROFESOR INFORMANTE:

DR. MAURICIO RUIZ-TAGLE MOLINA

INGENIERO CIVIL INFORMÁTICO

DOCTOR EN SOFTWARE Y SISTEMAS

ÁLVARO ANDRÉS NAVARRO GUERRERO

VALDIVIA – CHILE

2023

AGRADECIMIENTOS

En primer lugar, agradezco a mi profesor patrocinante Matthieu Vernier, por su buena disposición de integrarme a su proyecto, y también por su apoyo y orientación a lo largo de todo el desarrollo de este proyecto.

También agradecer al Laboratorio Natural Andes del Sur de la Facultad de Ciencias de la UACH por darme la oportunidad de participar en el proyecto “Conversa Con los Andes” (16ENI2-66903). Su colaboración ha sido esencial para el éxito de esta iniciativa.

Quiero dar las gracias a mi familia, especialmente a mi padre Jorge Navarro, mi madre Alicia Guerrero y mis hermanos Cristobal y Rodrigo, por haber sido mi constante apoyo. Su influencia positiva ha sido determinante en mi camino universitario.

Agradezco a todos los que han sido parte de mi tiempo en la universidad, tanto profesores como amigos. Su compañía, amistad y colaboración han hecho que esta etapa sea especial para mí.

ÍNDICE

ÍNDICE	I
ÍNDICE DE TABLAS	III
ÍNDICE DE FIGURAS	IV
RESUMEN	VI
ABSTRACT	VII
1. INTRODUCCIÓN	1
1.1 Antecedentes generales	1
1.1.1 Primeros avances del Laboratorio Natural Andes del Sur (LNAS)	1
1.2 Definición del problema	2
1.3 Colaboración con un tesista adicional	2
1.4 Objetivo General	3
1.5 Objetivos Específicos	3
1.6 Distribución de objetivos específicos por tesista.....	3
1.7 Impactos científicos y socioeconómicos	4
2. MARCO TEÓRICO	5
2.1 Estado del arte	5
2.2 Revisión sistemática	6
2.2.1 Desarrollo de Revisión Sistemática	6
2.2.2 Método de revisión.....	6
2.2.3 Resultados	8
2.1.4 Principales Hallazgos.....	10
3. ARQUITECTURA DE SOFTWARE	12
3.1 Tecnologías y herramientas utilizadas.....	12
3.1.1 Python	12
3.1.2 MariaDB.....	12
3.1.3 Discord	12
3.1.4 OpenAI.....	12
3.1.5 Langchain.....	13
3.1.6 ChromaDB	13
3.1.7 LLamaIndex	13
3.1.8 GitHub.....	13
4. DESARROLLO	14
4.1 Diseño e implementación de la base de datos	14
4.2 Plataforma Discord.....	16
4.2.1 Creación del bot	17
4.3 Diagrama de Componentes.....	20
4.4 Comandos Chat Inteligente	21
4.4.1 Comando questionDocument	22
4.4.2 Comando vectorize.....	23
4.4.3 Comando questionDB	24
5. RESULTADOS	26

5.1 Resultado Comando questionDocument	26
5.2 Resultado Comando vectorize.....	27
5.3 Resultado Comando questionDB	27
5.3.1 Prueba N°1	27
5.3.2 Prueba N°2	30
5.3.3 Prueba N°3	32
5.4 Demostración a CONAF	35
5.5 Evaluación de la usabilidad de la plataforma con participante del laboratorio.....	35
6. CONCLUSIONES	42
7. REFERENCIAS	43

ÍNDICE DE TABLAS

Tabla	Página
Tabla 1: Distribución de objetivos específicos	4
Tabla 2: Objetivos específicos y Actividades relacionadas	5
Tabla 3: Revisión de información del primer artículo	9
Tabla 4: Revisión de información del segundo artículo	9
Tabla 5: Revisión de información del tercer artículo	10

ÍNDICE DE FIGURAS

Figura	Página
Figura 1: Base de datos de LNAS	14
Figura 2: Diagrama entidad-relación Base de datos	15
Figura 3: Base de datos MariaDB	15
Figura 4: Agregar servidor	16
Figura 5: Elegir nuevo servidor	17
Figura 6: Nombre del servidor	17
Figura 7: Developer Mode	18
Figura 8: New Application	18
Figura 9: Agregar Bot	18
Figura 10: Permisos bot	19
Figura 11: Servidor bot	19
Figura 12: Token bot	20
Figura 13: Variable entorno token Bot.py	20
Figura 14: Diagrama de componentes	21
Figura 15: Flujo de sistema pregunta-respuesta	22
Figura 16: Variables función vectorize	24
Figura 17: Variables Comando questionDB	25
Figura 18: Resultados Comando questionDB.....	26
Figura 19: Resultado Comando vectorize.....	27
Figura 20: Revista documento científico Prueba N°1	27
Figura 21: Pregunta Prueba N°1 extraída del fragmento de documento	28
Figura 22: Respuesta del bot en Discord	28
Figura 23: Respuesta de ChatGPT	29
Figura 24: Extracto de la respuesta en el documento	29
Figura 25: Revista documento científico Prueba N°2	30
Figura 26: Pregunta Prueba N°2 extraída del fragmento de documento	30
Figura 27: Respuesta del bot en Discord	31
Figura 28: Respuesta de ChatGPT	31
Figura 29: Extracto de la respuesta en el documento	32
Figura 30: Revista documento científico Prueba N°3	32
Figura 31: Pregunta Prueba N°3 extraída del fragmento de documento	33
Figura 32: Respuesta del bot en Discord	33
Figura 33: Respuesta de ChatGPT	34
Figura 34: Extracto de la respuesta en el documento	34
Figura 35: Reunión CONAF	35
Figura 36: Pregunta usabilidad de la plataforma	36
Figura 37: Pregunta usabilidad comando question	37
Figura 38: Pregunta N°1 usuario y respuesta del bot comando question question	37
Figura 39: Pregunta N°2 usuario y respuesta del bot comando question	37

Figura 40: Pregunta sobre la claridad de la respuesta comando question	38
Figura 41: Pregunta usabilidad comando query_chroma	39
Figura 42: Pregunta N°1 usuario y respuesta del bot comando query_chroma	39
Figura 43: Pregunta N°2 usuario y respuesta del bot comando query_chroma	40
Figura 44: Pregunta N°3 usuario y respuesta del bot comando query_chroma	40
Figura 45: Pregunta sobre la claridad de la respuesta comando query_chroma	41

RESUMEN

El Laboratorio Natural Andes del Sur en Chile busca fortalecer el conocimiento y la conservación de las montañas del sur de Chile mediante enfoques interdisciplinarios y socio-ecológicos. A pesar de realizar análisis bibliográficos y talleres con actores locales, enfrentan desafíos en la gestión de la información. Dado esto, nace el proyecto "Conversa con los Andes", en donde se ha creado una plataforma tipo biblioteca para facilitar el acceso a información científica y técnica relacionada con los Andes del sur.

Se trabajó colaborativamente con un tesista adicional, por lo que este proyecto se distribuyó en roles específicos para abordar diversas áreas cruciales en el desarrollo de esta plataforma integral.

Para la creación de la plataforma, primero se analizó el archivo CSV que usaba el laboratorio, el cual consistía en un análisis bibliométrico de documentos relevantes al territorio de Los Andes. Este archivo almacena los metadatos más relevantes de cada documento.

Para almacenar el contenido de la información científica y técnica sobre el territorio de los andes, se diseñó e implementó una base de datos relacional en MariaDb, la cual también almacena los metadatos más relevantes.

Se extrajo el contenido de cada documento, desarrollando algoritmos de web scraping para recopilar la información de manera automática.

Se desarrolló una interfaz de usuario para facilitar la búsqueda de la información de documentos sobre el territorio de Los Andes, la cual fue la plataforma Discord, en donde se creó un servidor y un bot asociado a este. Se utilizó un canal dentro del servidor para realizar las interacciones con el bot, usando comandos desarrollados en Python.

Este bot se integró con funcionalidades avanzadas, como la búsqueda de documentos basados en criterios y la capacidad de agregar nuevos documentos de forma manual.

Se desarrolló la funcionalidad de chat inteligente, en donde se logró que el bot tenga la capacidad de interactuar con los usuarios, respondiendo preguntas en lenguaje natural. Esto se desarrolló usando técnicas de tratamiento automático del lenguaje, tales como la vectorización de información en conjunto con funciones de la librería OpenAI y una base de datos dedicada a vectores, ChromaDB.

Después se desarrolló la funcionalidad que permite realizar un análisis bibliométrico de los documentos almacenados. Este análisis proporciona información cuantitativa sobre la distribución y relevancia de la información científica y técnica almacenada.

ABSTRACT

The Southern Andes Natural Laboratory in Chile seeks to strengthen knowledge and conservation of the mountains of southern Chile through interdisciplinary and socio-ecological approaches. Despite carrying out bibliographic analyzes and workshops with local actors, facing challenges in information management. Given this, the project "Conversa con los Andes" was born, where a library-type platform has been created to facilitate access to scientific and technical information related to the southern Andes.

We worked collaboratively with an additional thesis student, so this project was distributed into specific roles to address various crucial areas in the development of this comprehensive platform.

To create the platform, the CSV file used by the laboratory was first analyzed, which consisted of a bibliometric analysis of documents relevant to the territory of Los Andes. This file stores the most relevant metadata of each document.

To store the content of scientific and technical information about the territory of the Andes, a relational database was designed and implemented in MariaDb, which also stores the most relevant metadata.

The content of each document was extracted, developing web scraping algorithms to collect the information automatically.

A user interface was developed to facilitate the search for document information about the territory of Los Andes, which was the Discord platform, where a server and a bot associated with it were created. A channel within the server was used to carry out interactions with the bot, using commands developed in Python.

This bot was integrated with advanced functionalities such as criteria-based document search and the ability to add new documents manually.

The intelligent chat functionality was developed, where the bot has the ability to interact with users, answering questions in natural language. This was developed using automatic language processing techniques, such as information vectorization in conjunction with functions from the OpenAI library and a database dedicated to vectors, ChromaDB.

Later, the functionality was developed that allows carrying out a bibliometric analysis of the stored documents. This analysis provides quantitative information on the distribution and relevance of stored scientific and technical information.

1. INTRODUCCIÓN

1.1 Antecedentes generales

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 académica a cargo del Laboratorio Natural Andes del Sur, Carla Marchant (Ladera Sur, 2022).

Es en ese contexto que surge el proyecto “Laboratorio Natural Andes del Sur de Chile” (LNAS), una iniciativa financiada por ANID en el concurso 2021 “Nodos para el Desarrollo en Investigación de Laboratorios Naturales en Chile” (Agencia Nacional de Investigación y Desarrollo [ANID], 2021) 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 transdisciplinario y de democratización del conocimiento y aportando conocimiento para el desarrollo de política pública nacional y regional”, mencionó la Directora del proyecto (Bertrán, 2021).

1.1.1 Primeros avances del Laboratorio Natural Andes del Sur (LNAS)

LNAS estructura su trabajo en tres líneas temáticas: 1) evolución y herencia geológica, 2) cambio global: interacciones bosque – suelo – recursos hídricos, 3) modos de vida y

habitar de montaña. Durante el año 2022, se realizó un análisis bibliográfico del territorio en estas tres temáticas. A través de este análisis se recopiló un total de 700 artículos científicos aproximadamente indexados en Scopus y Web of Science y publicados entre 1982 y 2021 (Resultados del análisis bibliométrico – Andes del Sur de Chile. 2022, octubre 19). Los resultados sirven como línea base para identificar los principales actores, instituciones y redes de investigación nacional e internacional asociadas a las temáticas del laboratorio.

Por otra parte, el LNAS está realizando talleres participativos con actores territoriales para mapear proyectos locales (Exitoso Mapeo Participativo en Puerto Varas: – Andes del Sur de Chile). Por ejemplo, en el taller participativo de Puerto Varas, a la reunión, participaron representantes del mundo de la montaña, como Andeshandbook, ONG como la Fundación Legado Chile, y el Sernageomin. A través de este tipo de actividades, el LNAS busca conocer y difundir los proyectos locales y crear una mejor conexión con la ciencia académica.

1.2 Definición del problema

Inicialmente, el coordinador ejecutivo del LNAS se acercó al equipo Sophia con la necesidad de poder reproducir más fácilmente el análisis bibliométrico que se realizó en 2022 a través de una herramienta de recopilación y análisis de documentos textuales. El primer análisis se realizó de manera semi-manual utilizando los motores de búsqueda de Scopus y WOS (Web Of Science) para recopilar datos manualmente, y librerías R para limpiar y extraer indicadores. Sin embargo, este análisis tiene algunas limitaciones, en particular el tiempo necesario para recopilar y limpiar los datos, y además solo integra datos científicos indexados. Para el LNAS también sería relevante incluir otras fuentes de conocimiento científico y técnico sobre el territorio.

Otra oportunidad detectada en las primeras conversaciones con LNAS, es poder construir una “biblioteca” de conocimientos sobre el territorio de Los Andes, que agrupe y organice grandes volúmenes de datos textuales habitualmente dispersos y difíciles de consultar. Esta biblioteca podría servir en el trabajo cotidiano para fortalecer y promover el conocimiento del territorio, y facilitar la creación de proyectos transdisciplinarios que involucren actores locales y la comunidad científica. A partir de esta oportunidad, nació el proyecto “Conversa con Los Andes”, que busca dotar al LNAS de una “biblioteca” de conocimientos sobre Los Andes que le pueda servir en su trabajo.

1.3 Colaboración con un tesista adicional

La ejecución del proyecto “Conversa con los Andes” se realizará mediante dos trabajos de tesis, siendo uno de ellos el que se describe en este proyecto. Se trabaja colaborativamente para lograr los objetivos del proyecto. Durante el transcurso del

documento se explicará detalladamente qué objetivos específicos están asociados a cada estudiante.

1.4 Objetivo General

Generar una plataforma que permita mejorar el acceso a conocimientos científicos y técnicos sobre el territorio de Los Andes del Sur mediante técnicas de análisis de lenguaje.

1.5 Objetivos Específicos

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 Distribución de objetivos específicos por tesista

Los tesistas que participaron en este proyecto son Fernando Ruiz y Álvaro Navarro. Ambos tuvieron en común el objetivo específico uno, en donde se diseñó y se implementó en conjunto la base de datos del proyecto. (ver Tabla 1).

En mi caso me correspondió implementar técnicas de Tratamiento Automático del Lenguaje para analizar y procesar los documentos (Objetivo 3), implementar una interfaz de usuario sencilla para facilitar la búsqueda de documentos, además de la creación del bot para interactuar con el usuario (Objetivo 4), y desarrollar una funcionalidad experimental de integración de un chat inteligente (Objetivo 6). Además, evalué la usabilidad de la plataforma con los usuarios del laboratorio, donde se tomó en cuenta la experiencia del usuario, la interactividad y la evaluación práctica de los comandos desarrollados en la plataforma (Objetivo 7).

Fernando se enfocó en el desarrollo de algoritmos de web scraping (Objetivo 2), manejó la integración de un bot de Discord con funcionalidades avanzadas (Objetivo 5), y la implementación de una funcionalidad para realizar un análisis bibliométrico de los documentos en la base de datos (Objetivo 8). Su enfoque técnico abarcó desde la estructuración de la información hasta la extracción de datos y la implementación de herramientas analíticas.

Tabla 1. Distribución de objetivos específicos.

Nro. O.E.	Álvaro	Fernando
1	x	x
2		x
3	x	
4	x	
5		x
6	x	
7	x	
8		x

1.7 Impactos científicos y socioeconómicos

Incremento de capacidades y conocimiento: Al contar con una base de datos de documentos científicos y técnicos específicos del territorio de Los Andes, el Laboratorio puede fortalecer su experticia y adquirir capacidades y conocimientos especializados sobre interacción del bosque, suelo y recursos hídricos, evolución y herencia geológica y modos de vida y habitar de montaña. Esto permite un mayor desarrollo de habilidades y conocimientos en el laboratorio.

Mayor visibilidad y reconocimiento: Si el Laboratorio desarrolla y promueve la plataforma "Conversa con Los Andes" de manera efectiva, esto puede aumentar su visibilidad a nivel nacional. Esto puede conducir a una mayor reputación y reconocimiento en el ámbito científico, lo que a su vez puede atraer colaboraciones, financiamiento y oportunidades de investigación adicionales para estudiantes universitarios.

Menores costos en investigación y desarrollo: Al tener acceso a una plataforma que recopila y organiza documentos científicos y técnicos relacionados con el territorio de Los Andes, el Laboratorio puede reducir los costos asociados con la búsqueda y adquisición de información relevante para sus investigaciones y proyectos. Esto permite optimizar los recursos y aumentar la eficiencia en el proceso de generación de conocimiento.

2. MARCO TEÓRICO

Con el objetivo de lograr un proceso más estructurado y transparente para abordar cada uno de los objetivos específicos del proyecto, se han definido una serie de macro actividades, las cuales se encuentran detalladas en la Tabla 2.

Tabla 2. Objetivos específicos y Actividades relacionadas.

Nro. O.E.	Actividades
1	Validación de la calidad y relevancia de los documentos. Definición de la estructura definitiva de la base de datos.
3	Investigación preliminar. Selección de técnicas de TAL. Implementación de algoritmos de TAL. Extracción de contenido. Validación de resultados.
4	Configurar canal de Discord. Crear e integrar el bot al canal de Discord.
6	Vectorización de documentos. Implementación de funciones para acceder a los vectores que contienen información. Validación de resultados.
7	Definición de los criterios de evaluación. Selección de los usuarios. Diseño y realización de pruebas de usabilidad. Análisis de los resultados de las pruebas.

2.1 Estado del arte

En este trabajo de título, buscamos solucionar principalmente el problema de la biblioteca que el LNAS quiere construir. Buscamos aportar al diseño, implementación y despliegue de la plataforma “Conversa con Los Andes” que pueda responder a distintos casos de uso, en particular servir de “biblioteca” para que los usuarios del LNAS puedan acceder más fácilmente a conocimientos científicos y técnicos sobre el territorio de Los Andes del Sur, a través de criterios de búsquedas y preguntas en lenguaje natural.

Surgen las siguientes preguntas:

¿Cuáles son las técnicas informáticas existentes para acceder a la información en colecciones de documentos? ¿Cuáles son sus conceptos principales?

Sistemas de Preguntas y Respuestas (**Question Answering Systems**): Otra rama del tratamiento automático del lenguaje que se busca diseñar e implementar sistemas capaces de responder a preguntas. La última tendencia son los sistemas de preguntas y respuestas basados en large language models (LLM) (Höhn, 2022).

¿Cuáles son las herramientas y modelos de Information Retrieval, Question Answering Systems y Semantic Search que se podrían utilizar en Conversa con Los Andes?

OpenAI ChatGPT: Ayudar a los usuarios de Conversa con los Andes a responder preguntas sobre diversos temas de los Andes, sobre territorios, investigaciones científicas, tesis de universidades.

2.2 Revisión sistemática

Se realizó una revisión sistemática para evaluar los distintos métodos relacionados con modelos de lenguaje grande (LLMs) que tenga relación con el sistema de preguntas que haría el usuario y el tipo de respuesta que se entregarán.

La revisión sistemática implica evaluar y analizar toda la investigación relevante disponible sobre una pregunta específica. Su propósito es proporcionar una evaluación fundamentada y confiable de un tema de investigación mediante el uso de un método riguroso y verificable.

2.2.1 Desarrollo de Revisión Sistemática

Objetivo y Pregunta de Revisión

Conversa con los Andes hará uso de técnicas informáticas existentes para acceder a la información en colecciones de documentos, las cuales se usarán para implementar la comunicación entre el usuario y la interfaz. El trabajo se enfocará en las técnicas de “Pregunta y respuesta” y “Búsqueda semántica”, las cuales son las más esenciales para lo que se quiere lograr. Por lo tanto, es necesario saber cómo se han utilizado estas técnicas en trabajos existentes. Es por esto que se seleccionó dar respuesta a la siguiente pregunta de investigación:

¿Cómo diseñar un sistema de preguntas-respuestas que utilice modelos de lenguaje grande (LLM) y que se base sobre búsquedas semánticas en documentos científicos y técnicos?

2.2.2 Método de revisión

Metodología

El proceso de desarrollo de la Revisión Sistemática comenzó con la elaboración de la consulta de búsqueda, la cual se diseñó teniendo en cuenta la pregunta de investigación planteada. Luego, se procedió a la selección de los artículos que cumplían con los criterios de selección establecidos y que eran considerados relevantes dentro de los resultados obtenidos de la búsqueda. Una vez seleccionados, se llevó a cabo un análisis de dichos artículos, extrayendo la información pertinente y resumiendo los hallazgos relevantes. En resumen, se siguieron estos pasos para llevar a cabo la Revisión Sistemática.

Fuentes y Cadena de búsqueda

La fuente para realizar la búsqueda fue la plataforma **Web of Science**, la cual contiene la colección de base de datos de referencias bibliográficas y citas de publicaciones. Dada a la búsqueda preliminar sobre palabras claves para obtener un conjunto de documentos relevantes, se llegó al resultado de dos palabras claves, para responder a la pregunta de investigación. Para la búsqueda en Web of Science se utilizaron palabras en inglés. Las dos palabras claves principales que se identificaron fueron las siguientes: **Question Answering Systems, Semantic Search**.

Para la primera palabra clave (Question Answering Systems) se integraron los siguientes sinónimos: Collection of documents, Automatic language processing, Natural language processing.

Para la segunda palabra clave (Semantic Search) se integró el siguiente sinónimo: Large language models.

Además, se consideraron las siguientes restricciones:

1. Solo se consideraron artículos publicados desde el año 2019 en adelante.
2. La búsqueda se hizo sólo en el resumen de los artículos.

Cadena de búsqueda

La cadena de búsqueda quedó finalmente definida:

("Question Answering systems" OR "Collection of documents" OR "automatic language processing" OR "Natural language processing") AND ("Large language models" OR "Semantic search")

Resultados y selección de estudios

1. A partir de los resultados de la búsqueda, se seleccionaron aquellos artículos que tengan relación con el siguiente criterio de inclusión:
2. Aplicación de búsqueda semántica.
3. Información sobre qué Large Language Models (LLM) se usaron.
4. Proporciona orientación sobre el diseño de preguntas y respuestas utilizando LLM.

5. Uso de técnicas de análisis semántico y procesamiento de lenguaje natural (NLP) para mejorar la exactitud.

Extracción de información de los artículos seleccionados

Con el objetivo de organizar la información recopilada sobre diseño de sistemas para acceder a la información en colecciones de documentos que usen modelos de lenguaje grandes (LLM) usando el sistema de pregunta-respuesta y semantic search, incluyendo sus características, técnicas informáticas utilizadas, se estableció una pauta de extracción de información. Dicha pauta se diseñó con el propósito de estructurar de manera sistemática los datos encontrados en relación con estos temas.

La pauta de extracción de información por artículo fue la siguiente:

1. Técnicas informáticas utilizadas
2. Uso de sistema pregunta-respuesta
3. Uso de búsqueda semántica
4. Hallazgos importantes.

Estudios incluidos y excluidos

Después de realizar la búsqueda utilizando la consulta específica en el rango de año 2019-2023, se obtuvo un conjunto de 17 resultados. Se procedió a examinar detenidamente el resumen de cada uno de ellos con el fin de evaluar si cumplían con los criterios de inclusión establecidos. Como resultado de este proceso de selección, se identificaron y seleccionaron 3 artículos de WoS y uno de arXiv, los cuales se consideraron pertinentes y se procedió a leerlos en su totalidad. Cabe destacar que la cantidad de artículos fue aumentando progresivamente desde el año 2020 al 2023.

2.2.3 Resultados

La información relevante de cada artículo seleccionado se extrajo, se analizó y posteriormente se detalló en las Tablas 3, 4 y 5.

Tabla 3. Revisión de información del primer artículo.

Autor
Zhixue Jiang (2021)
Nombre del artículo
Research on Medical Question Answering System Based on Knowledge
Técnicas informáticas utilizadas.
El sistema utiliza grafos de conocimiento para representar y organizar la información médica. Estos grafos contienen entidades de conocimiento relacionadas con enfermedades y otros aspectos médicos, y ayudan a establecer conexiones y relaciones entre diferentes entidades.
Uso de sistema pregunta-respuesta
El sistema utiliza sistemas de pregunta y respuesta para llevar a cabo diálogos entre humanos y máquinas en lenguaje natural. Estos sistemas permiten que los usuarios realicen preguntas y obtengan respuestas relevantes basadas en el conocimiento médico almacenado en el sistema.
Uso de búsqueda semántica
La búsqueda semántica se utiliza en el enfoque de "Vector Modeling", donde se representan las entidades y las preguntas en un espacio vectorial y se comparan las similitudes para encontrar las respuestas.
Hallazgos importantes
El sistema construye un grafo de conocimiento que utiliza las enfermedades como entidad central.

Tabla 4. Revisión de información del segundo artículo.

Autor
Luigi de Angelis, Francesco Baglivo, Gugliemo Arzilli, Gaetano Pierpaolo, Paolo Ferragina, Alberto Tozzi, Caterina Rizzo (2023)
Nombre del artículo
ChatGPT and the rise of a large language models: The new AI-driven infodemic threat in public health
Técnicas informáticas utilizadas.
Se menciona que varios modelos se han probado en tareas de NLP (Natural Language Processing) en el dominio científico. El NLP es una rama de la IA que se ocupa de la interacción entre las computadoras y el lenguaje humano, y se utilizan diversas técnicas informáticas como el análisis sintáctico, la clasificación de texto y la generación de lenguaje natural.
Uso de sistema pregunta-respuesta
Se menciona el uso de modelos de lenguaje grandes (LLMs) para la generación de resúmenes o recomendaciones de artículos científicos, lo cual facilita a los investigadores obtener rápidamente los puntos clave de un resultado científico. Además, se hace referencia a modelos específicos de dominio, como BioGPT y PubMedGPT, que fueron entrenados exclusivamente en resúmenes y artículos biomédicos y utilizados para la respuesta a preguntas médicas.
Uso de búsqueda semántica
No aplica.
Hallazgos importantes

El texto destaca el creciente interés en el uso de LLMs en la investigación académica. Los LLMs facilitan la generación de resúmenes y la respuesta a preguntas en el ámbito científico, especialmente en el campo biomédico.

Tabla 5. Revisión de información del tercer artículo.

Autor
Aokun Chen, Zehao Yu, Xi Yang, Yi Guo, Jiang Bian, Yonghui Wu (2023)
Nombre del artículo
Contextualized medication information extraction using Transformer-based deep learning architectures
Técnicas informáticas utilizadas.
Se desarrolló un sistema de NLP para extraer menciones de medicamentos de los textos clínicos. Esta técnica implica el procesamiento de texto y la identificación de entidades relevantes, en este caso, los nombres de los medicamentos. También se implementó un sistema de clasificación para determinar si se discuten cambios en los medicamentos o no. Esto implica la utilización de técnicas de aprendizaje automático y clasificación para asignar una etiqueta a cada evento.
Uso de sistema pregunta-respuesta
El uso de “question answering” y “semantic search” podría ser relevante en el contexto de extracción de información sobre medicamentos. Por ejemplo, la capacidad de responder preguntas específicas relacionadas con medicamentos o buscar información relevante sobre medicamentos en los textos clínicos podrían estar asociadas con técnicas de “question answering”.
Uso de búsqueda semántica
No aplica.
Hallazgos importantes
El estudio menciona la exploración de varios modelos de transformers pre-entrenados para la extracción de menciones de medicamentos y la información contextual asociada.

2.1.4 Principales Hallazgos

Dentro de los textos relacionados a cómo diseñar un sistema de preguntas-respuestas basado en búsquedas semánticas en documentos científicos y técnicos que utilice modelos de lenguaje grande (LLM), se han identificado artículos que son relevantes para el desarrollo de este tipo de sistemas. Estos hallazgos y factores claves resaltan aspectos comunes y característicos de las herramientas y enfoques utilizados en el diseño de sistemas de preguntas-respuestas basados en LLM y búsquedas semánticas.

Se ha encontrado que el uso de modelos de lenguaje grande, como GPT-3 y otros LLMs, ha demostrado un avance significativo en la capacidad de respuesta a preguntas basadas en texto. Estos modelos tienen la capacidad de comprender y generar lenguaje natural, lo que les permite interpretar las preguntas formuladas por los usuarios y generar respuestas relevantes. Su capacidad de generalización y adaptabilidad los convierte en herramientas poderosas para el diseño de sistemas de preguntas-respuestas.

LLM facilita la generación de resúmenes y la respuesta a preguntas en el ámbito científico

La importancia del modelo de lenguaje GPT-3, como LLM radica en su capacidad para facilitar la generación de resúmenes y responder preguntas en el ámbito científico. Esto tiene varias implicaciones positivas:

1. Eficiencia en la generación de resúmenes.
2. Acceso rápido y confiable a información científica.
3. Apoyo a la toma de decisiones informadas.
4. Aplicación del alcance y la difusión del conocimiento científico

Esto es relevante para el proyecto, ya que podría explorarse cómo los LLMs, como ChatGPT, pueden ser utilizados como herramientas de apoyo para recopilar evidencia o generar resúmenes de artículos científicos relacionados con la región de los Andes.

3. ARQUITECTURA DE SOFTWARE

3.1 Tecnologías y herramientas utilizadas

3.1.1 Python

El lenguaje de programación Python cuenta con una vasta colección de bibliotecas y frameworks que facilitan el desarrollo de diversas funcionalidades. En este proyecto, la utilización de librerías de OpenAI para el procesamiento de lenguaje natural, la compatibilidad de la base de datos MariaDB, y además la integración de scripts en Python junto con el bot de Discord, demuestra cómo Python permite aprovechar herramientas especializadas.

3.1.2 MariaDB

MariaDB es un sistema de gestión de bases de datos relacional de código abierto, una bifurcación de MySQL. Su principal ventaja radica en su compromiso con la comunidad y su desarrollo activo, ofreciendo mejoras continuas, características avanzadas y compatibilidad con MySQL, lo que permite una fácil migración. Se usó esta herramienta para crear la base de datos con toda la información de los documentos científicos.

3.1.3 Discord

Discord es una plataforma de comunicación en línea que combina características de chat y voz, diseñada principalmente para comunidades. La principal ventaja de utilizar bots en Discord es que permiten automatizar tareas, proporcionar funciones personalizadas y mejorar la interactividad del servidor, enriqueciendo la experiencia del usuario. Esta herramienta fue utilizada para crear la interfaz de usuario y la implementación del bot.

3.1.4 OpenAI

OpenAI es una organización de investigación en inteligencia artificial que desarrolla modelos de lenguaje avanzados, como GPT (Generative Pre-trained Transformer). Su principal ventaja radica en la capacidad de estos modelos para comprender y generar texto de manera coherente y contextual, lo que ha llevado a avances significativos en diversas aplicaciones, desde procesamiento de lenguaje natural hasta asistencia en la creación de contenido.

El uso de esta librería fue esencial para trabajar con la implementación del chat inteligente, en donde se probaron distintos modelos de GPT. Para el acceso a esta herramienta se tuvo que comprar un código serial en el sitio web de OpenAI, y se importa en el código fuente.

3.1.5 Langchain

LangChain es un framework de código abierto que permite integrar modelos de lenguaje, como GPT-3.5 y GPT-4 de OpenAI, con componentes externos para crear aplicaciones de procesamiento del lenguaje natural.

Esta herramienta fue esencial para la plataforma, ya que puede crear chatbots o sistemas de respuesta a preguntas integrando un LLM (como OpenAI) con un grupo de documentos. Esto permite que una aplicación tome el texto ingresado por el usuario, lo procese y recupere las mejores respuestas de este grupo.

Se usaron diversas funciones de langchain, tales como “TokenTextSplitter” para crear fragmentos de información y “RetrievalQA” para poder trabajar con la vectorización de documentos y el sistema de pregunta y respuesta para chat inteligente.

3.1.6 ChromaDB

ChromaDB es un sistema de almacenamiento de vectores de código abierto (base de datos de vectores) diseñado para almacenar y recuperar embeddings de vectores. Su función principal es almacenar embeddings con metadatos asociados para su uso posterior en modelos de lenguaje extensos. Esta herramienta fue utilizada para vectorizar todos los documentos y posteriormente almacenarlos localmente.

3.1.7 LLamaIndex

LlamaIndex es un framework que simplifica la creación de aplicaciones LLM (Large Language Model). Lo que proporciona es la capacidad de realizar un aumento de datos privados, lo que permite incorporarlos a los LLM para la generación de conocimiento y el razonamiento. Se usaron diversas funciones de esta herramienta, para poder vectorizar documentos, y consultar la información del contenido.

3.1.8 GitHub

GitHub es una plataforma de desarrollo colaborativo que utiliza el sistema de control de versiones Git. Su principal ventaja para nuestro grupo de tres personas radica en la facilitación de la colaboración y coordinación del proyecto. Permite el seguimiento de cambios, la resolución de conflictos y la sincronización sin problemas, mejorando la eficacia del trabajo en equipo. Esta herramienta fue útil para llevar un control del avance que se realizaba durante el transcurso del proyecto.

4. DESARROLLO

El proyecto de título se desarrolló en un servidor con sistema operativo Linux del profesor de la UACH Matthieu Vernier. Dentro de este servidor, en la carpeta "Conversa con los Andes" se encuentran elementos fundamentales como la base de datos MariaDB, el archivo principal que contiene las funcionalidades Bot.py, y archivos Python externos incorporados en Bot.py. La integración de esta carpeta con GitHub asegura que cualquier modificación realizada en la carpeta se refleje en la plataforma, garantizando así la actualización constante del equipo de trabajo.

4.1 Diseño e implementación de la base de datos

El diseño de la base de datos involucró reuniones con el Laboratorio Natural Andes del Sur (LNAS) para definir conceptos clave basados en su archivo CSV. Este archivo contiene metadatos relevantes sobre todos los documentos que poseen. Con esta información, se elaboró un diagrama entidad-relación que sirve como estructura fundamental para la base de datos del proyecto (ver Figuras 1 y 2)

A1	B	C	D	E	F	G	H	I	J	K	L	M
1	AUTORES/AS	TÍTULO	AÑO	REVISTA	DOI	CATEGORÍA	REGIÓN	COMUNAS	Enlace			
2	ADRIASOLA A;STOCKHERT B	COOLING HISTORIES AND DEFORMATION OF PLUTONIC ROCKS ALONG THE LIQUÍÑE OF	2008	REVISTA GEOLC 10.4067/	Tectónica y Peligr	Los Lagos Puerto Varas, Huai	Chile	https://www.scielo.cl/pdf/rgh/v35n1/ar				
3	ADRIASOLA A;THOMSON S;BRIX M;HI	POSTMAGMATIC COOLING AND LATE CENOZOIC DENUDATION OF THE NORTH PATAGON	2006	INTERNATI	ON 10.1007/	Mineralogía y pe	Los Lagos Puerto Varas, Huai	Chile	https://link.springer.com/article/10.1007			
4	AIUPPA A;BITETTO M;FRANCOFONTE A	CO2GAS PRECURSOR TO THE MARCH 2015 VILLARRICA VOLCANO ERUPTION	2017	GEOCHEMISTR	10.1002/	Volcanología y pe	La Araucí Pucón, Villarrica, Panguipulli		https://discovery.ucl.ac.uk/id/eprint/100			
5	ALFANO F;BONADONNA C;GURIOLI L	INSIGHTS INTO ERUPTION DYNAMICS FROM TEXTURAL ANALYSIS THE CASE OF THE MAY	2012	BULLETIN OF V	10.1007/	Volcanología y pe	Los Lagos Chaitén		https://doc.rero.ch/record/309646/files/			

Figura 1. Base de datos de LNAS.

En la Figura 1 se puede ver un extracto de la base de datos de LNAS, en donde cada fila contiene información sobre Autores/as, Título, Año, Revista, Doi, Categoría, Región, Comunas y Enlace.

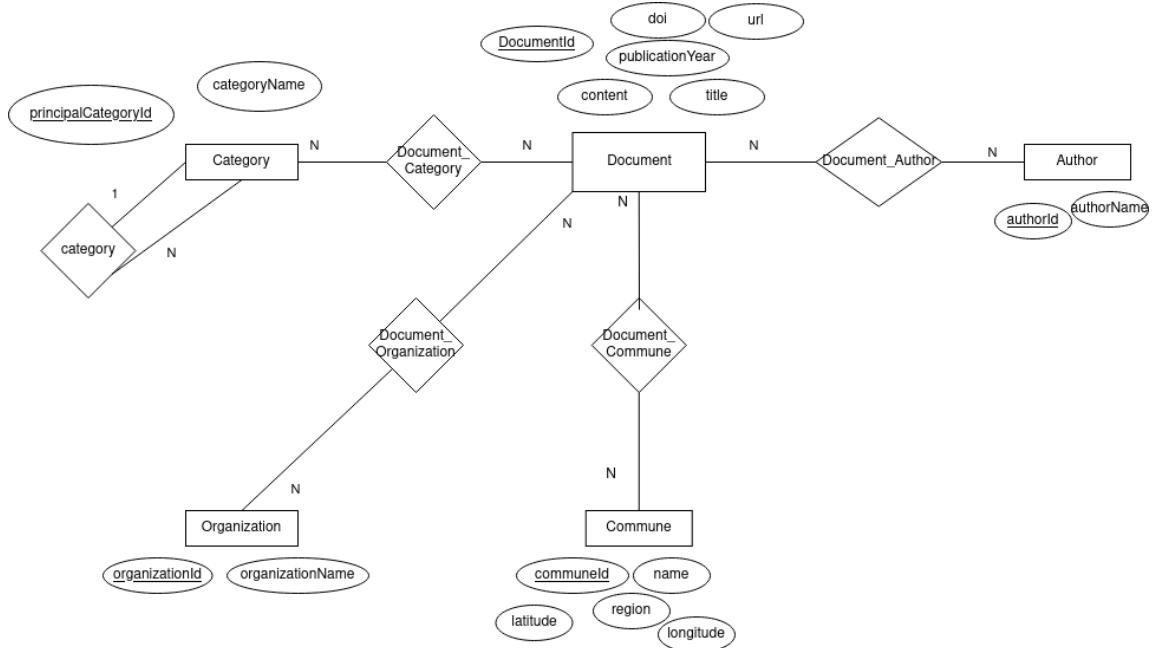


Figura 2. Diagrama entidad-relación Base de datos.

Dado el diagrama de entidad-relación, se implemento la base de datos en MariaDB para almacenar los documentos con sus respectivos atributos (ver Figura 3).

```

MariaDB [andes]> show tables;
+-----+
| Tables_in_andes |
+-----+
| Author
| Category
| Commune
| Concept
| Document
| Document_Author
| Document_Category
| Document_Commune
| Document_Concept
| Document_Organization
| Organization
+-----+
    
```

Figura 3. Base de datos MariaDB.

4.2 Plataforma Discord

Para la interfaz de usuario, en un principio se discutió si usar una página web u otra herramienta. Dado el tiempo dado para este proyecto, se decidió usar Discord ya que fue la plataforma con la que estamos más familiarizados, y tenía una buena integración con Python, lo cual nos acomodó. Además, encontramos que es una interfaz simple para el usuario, en donde solo tiene que crearse una cuenta e iniciar sesión en él, y después mediante un enlace conectarse al servidor específico que creamos en Discord.

Discord ofrece una ventaja al proporcionar una experiencia cómoda tanto en dispositivos móviles como en computadoras. Esta plataforma combina la accesibilidad de su aplicación móvil con la funcionalidad completa de su versión de escritorio, permitiendo a los usuarios participar en conversaciones desde cualquier lugar de manera fluida. La flexibilidad se extiende dada la posibilidad de interactuar con otros usuarios a través de canales que admiten tanto mensajes de texto como comunicación por voz. Ya sea compartiendo resultados, discutiendo estrategias o simplemente socializando, Discord se ajusta a lo que cada persona prefiere, por eso es una plataforma completa con distintos métodos de comunicación.

Para esto, el usuario debe tener conexión a internet. Una vez autenticado en el servidor, ya puede hacer uso de todas las funcionalidades que tiene el bot.

Creación del servidor en Discord

Una vez instalado Discord, se ingresa y se hace clic en el símbolo “+” (ver figura 4).

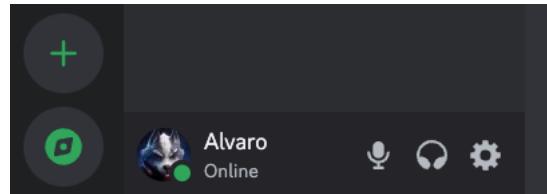


Figura 4. Agregar servidor.

Se abrirá una ventana con opciones, en donde se elige si se crea un servidor propio o ya se tiene una invitación. En este caso como somos los creadores, se elige la opción “Create My Own” (ver Figura 5)

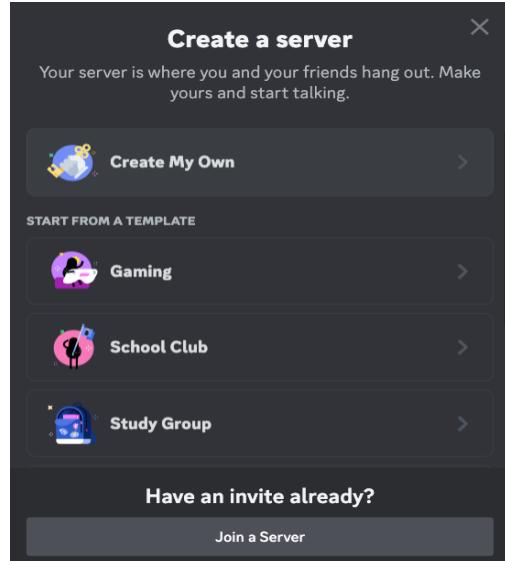


Figura 5. Elegir nuevo servidor.

Por último, se le asigna un nombre al nuevo servidor, y se hace clic en “Create” (ver Figura 6). Una vez realizado este paso, el servidor se encuentra creado y ya se puede seguir con la creación del bot e integrarlo al servidor.

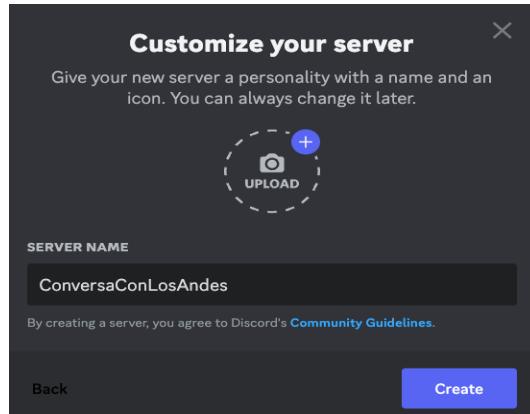


Figura 6. Nombre del servidor.

4.2.1 Creación del bot

Para integrar el bot, primero se debe acceder a configuración avanzada de la cuenta, y activar **Developer Mode**, para crear bots en el modo desarrollador de Discord (ver Figura 7).

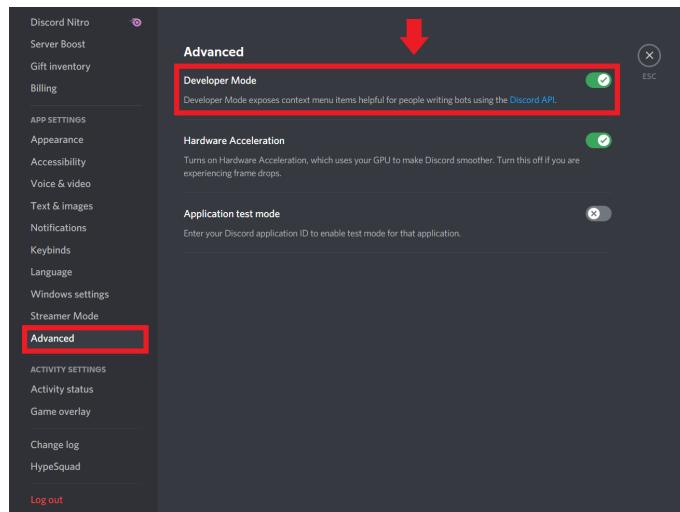


Figura 7. Developer Mode.

Se ingresó al Developer Portal desde el enlace de “Discord API” de la Figura 7, donde se elige Applications y se hace click en New Application (ver Figura 8).

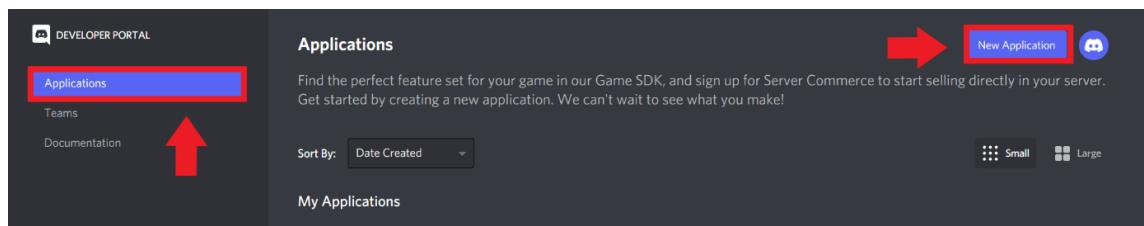


Figura 8. New Application.

Una vez añadida la nueva aplicación, se pudo agregar el bot (ver Figura 9).

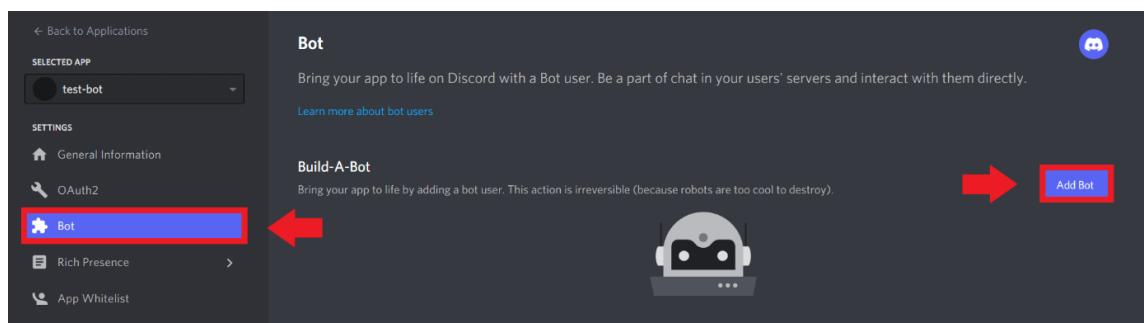


Figura 9. Agregar Bot.

Ya creado el bot, se eligieron los permisos que tendrá (ver Figura 10).

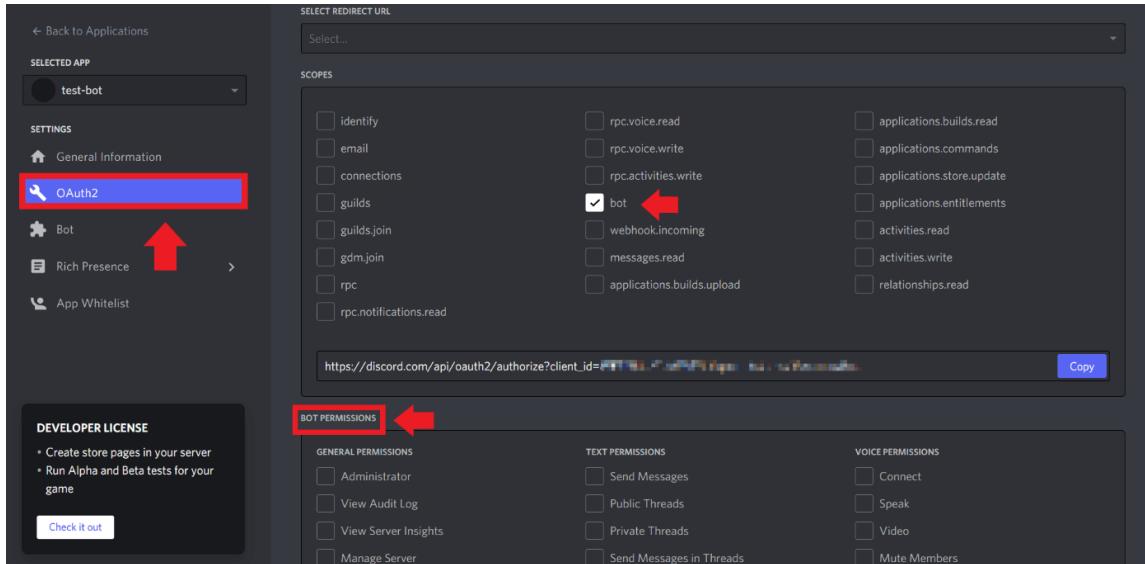


Figura 10. Permisos Bot.

Para integrar el bot al servidor deseado en Discord, se copió el enlace (ver Figura 10) y se pegó en algún navegador web, en donde entrega la opción para elegir el servidor, y se elige el ya creado anteriormente (ver Figura 11).

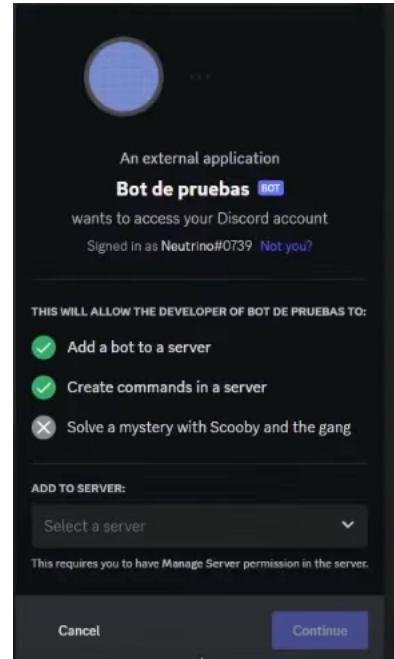


Figura 11. Servidor Bot.

Finalmente, en la pestaña “Bot” se genera un Token, el cual nos sirve para agregarlo en el archivo (Bot.py) que tendrá las funcionalidades, con este token se logra la comunicación del servidor de Discord y Bot.py. Este token se guardó como variable de entorno llamada “DISCORD_TOKEN” (ver Figuras 12 y 13).

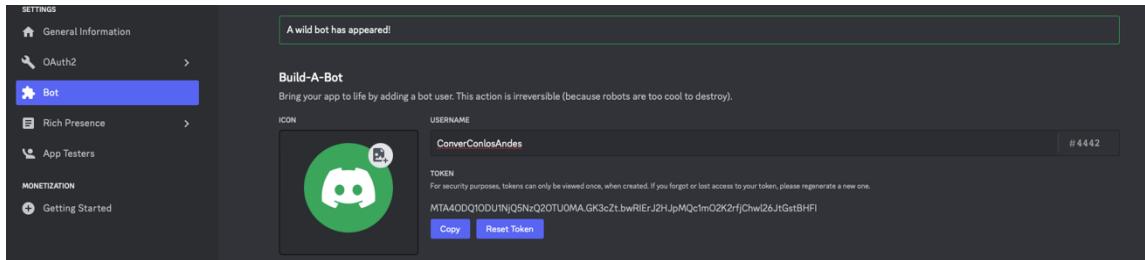


Figura 12. Token Bot.

```
TOKEN = os.environ.get('DISCORD_TOKEN')
```

Figura 13. Variable entorno token Bot.py.

Al finalizar los pasos mencionados, se puede activar el bot ejecutando el archivo bot.py. Una vez encendido el bot, se puede hacer uso de todas las funcionalidades que están implementadas en bot.py.

¿Qué es un bot?

Un bot es un programa informático diseñado para realizar tareas automáticas, repetitivas o específicas en representación de los usuarios. En el contexto de Discord, un bot se refiere a un tipo de aplicación automatizada que funciona dentro de los canales de texto de Discord.

4.3 Diagrama de Componentes

Se elaboró un diagrama de componentes que permite una visualización clara de los elementos fundamentales. Estos componentes centrales comprenden una base de datos MariaDB, encargada de suministrar información al bot de Discord. La función del bot es procesar la información obtenida de la base de datos, canalizando el flujo de datos. También se incorporó un componente para el sistema de recopilación automática, diseñado para realizar web scraping de los documentos. Una vez procesada, la información se presenta al usuario final a través del canal de Discord, proporcionando resultados comprensibles y accesibles. Este diseño significa que la base de datos, el bot de Discord y cómo los usuarios ven la información están conectados de manera fluida. (ver Figura 114).

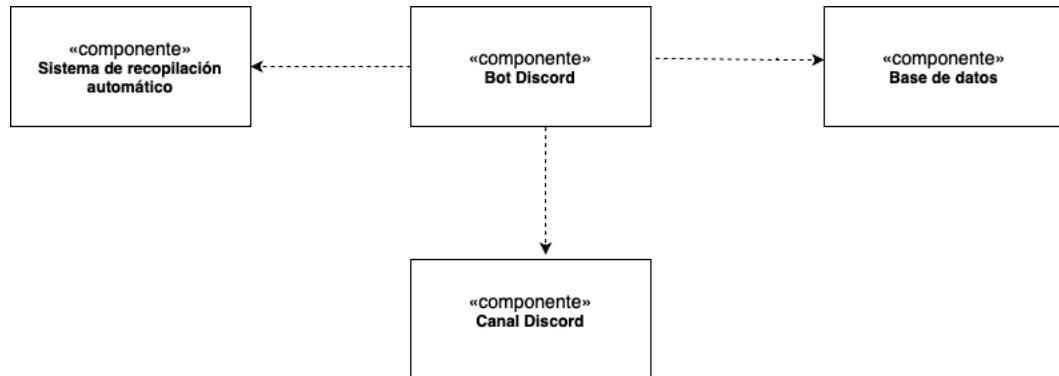


Figura 14. Diagrama de componentes.

4.4 Comandos Chat Inteligente

¿Qué es un embedding?

Es una representación vectorial de una palabra o una secuencia de palabras. Estos vectores son espacios matemáticos en los cuales las palabras o frases están representadas como puntos. Esta representación vectorial captura la semántica y la relación entre las palabras.

¿Cómo se construye el chat inteligente?

Para desarrollar un chat inteligente fue necesario entender en qué consistía cada paso que se ejecuta desde la pregunta del usuario hasta la respuesta final. El primer paso consiste en obtener la información que se desea consultar, en este caso son cadenas de textos.

El segundo paso es generar un índice, esto significa implementar una estructura de datos que permita consultar los datos, esto se logra creando embeddings, para representar numéricamente el significado de los datos que se almacenan.

Una vez que los datos están indexados, se decide si almacenar o no el índice, esta práctica es útil para evitar volver a indexarlos. Lo que, en este caso, se decidió almacenar el índice, para así tener una base de datos de forma local.

Finalmente se hace uso de algún LLM(Large Language Model), como por ejemplo GPT, para pasarle el contenido del índice junto con la pregunta del usuario vectorizada, para que GPT calcule la mayor similitud entre vectores y entregue la respuesta a la pregunta (ver Figura 15).

Existen diversas técnicas para calcular la similitud entre vectores, una de las comunes en este contexto es la similitud coseno.

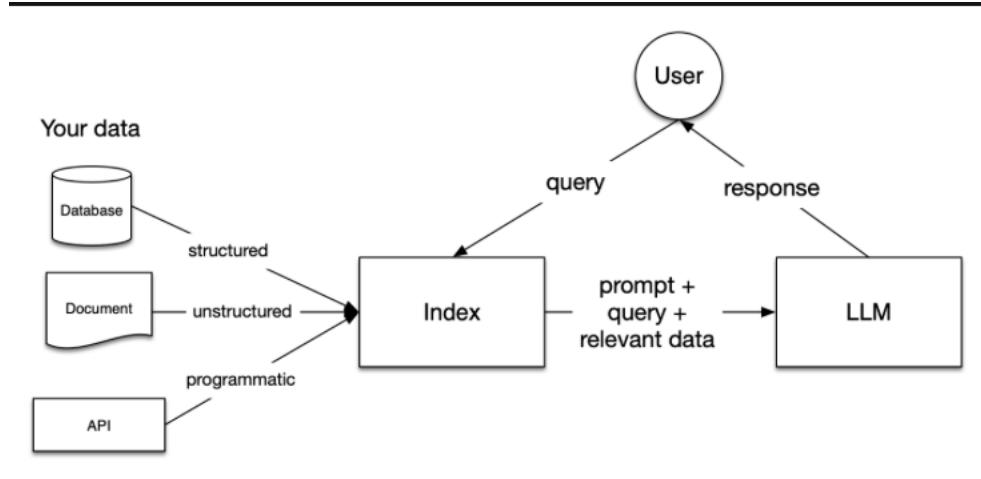


Figura 15. Flujo de sistema pregunta-respuesta.

Dado el objetivo específico: “**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**”, se desarrollaron tres comandos para lograr la interacción del usuario con la información de los documentos.

4.4.1 Comando questionDocument

!questionDocument <pregunta>.<titulo_documento>

El comando questionDocument recibe dos parámetros, la pregunta del usuario y el título del documento, entre los dos parámetros debe haber un punto para hacer diferenciación de la variable pregunta y título. La implementación del comando "questionDocument" en el sistema permite a los usuarios realizar consultas en lenguaje natural sobre el contenido de un documento específico, obteniendo respuestas precisas y relevantes.

Este proceso comienza con la búsqueda del documento en la base de datos MariaDB a partir del título proporcionado, extrayendo luego su contenido para su posterior procesamiento. Para optimizar la eficiencia en consultas repetidas del mismo documento, el sistema verifica la existencia de un índice de vectores almacenado en el servidor. En caso de ausencia, se procede a vectorizar el contenido del documento, generando un índice que se almacena para futuras consultas, evitando así la necesidad de repetir el proceso de vectorización.

Para evitar escribir el título cada vez que se quiera preguntar por el mismo documento más de una vez, se implementó que solo se escriba el título la primera vez que se consulta un documento.

La clave de este flujo comienza en la vectorización del contenido, lo cual permite utilizar la librería "llamaindex" para enviar la consulta a la librería OpenAI, utilizando el modelo "GPT-3.5 turbo". Este enfoque aprovecha la capacidad de GPT-3.5 turbo para comprender y responder en lenguaje natural, proporcionando respuestas contextualmente relevantes a las preguntas formuladas por los usuarios.

La combinación de MariaDB, la vectorización y el modelo de lenguaje de OpenAI, facilitó la experiencia de búsqueda interactiva, enriqueciendo la interacción del usuario con la plataforma "Conversa con Los Andes".

4.4.2 Comando vectorize

!vectorize

La implementación del comando "questionDocument" se generó a partir de la preocupación de evitar la vectorización repetida al consultar documentos. La forma en que se estaba vectorizando con "llamaindex" resultaba más tediosa para almacenar múltiples documentos. Por esta razón, se tomó la decisión de emplear ChromaDB, una base de datos vectorial, para almacenar todos los documentos vectorizados. En este caso, se optó por utilizar la base de datos Chroma de tipo local, almacenada en disco, prescindiendo de servicios en la nube y evitando la necesidad de adquirir almacenamiento de vectores a través de un proveedor de servicios.

El comando "vectorize" se ejecuta sin parámetros y procede a vectorizar los documentos de diez en diez, ya que intentar vectorizar todos los documentos de una vez causaba colapsos en la memoria RAM del servidor. Este proceso implica extraer el contenido de los documentos directamente de la base de datos MariaDB y almacenar los vectores en un archivo SQL de Chroma, que sirve como la base de datos vectorial. Una ventaja identificada al vectorizar y guardar con Chroma es que, a pesar de hacerlo de manera incremental por lotes de diez, la información se añade sin reemplazarse.

El propósito principal de este comando es utilizarlo solo una vez para vectorizar todos los documentos almacenados, a menos que se desee agregar más documentos en el futuro.

A continuación, se detallan dos variables más relevantes de la función **vectorize**, como se puede ver en un extracto de código en la Figura 16.

text_splitter

Contiene los parámetros para abordar los fragmentos de textos que se generarán al vectorizar el contenido de los documentos. El parámetro **chunk_size** especifica el tamaño máximo de cada fragmento (chunk) de texto. En este caso se ha configurado en 300, esto significa que el texto será dividido en fragmentos de 300 tokens máximo.

Por otra parte, el parámetro **chunk_overlap** establece la cantidad de tokens que se superponen entre fragmentos adyacentes. Esto significa que los últimos veinte tokens de un fragmento se repetirán al comienzo del siguiente fragmento. Esto ayuda a garantizar que no se pierda información en ocasiones debido a la división.

vectordb

Esta variable crea un objeto de tipo Chroma haciendo uso de su función “from_texts” que es especificó para poder trabajar con variables de tipo string, que es el tipo de variable que se usó para almacenar el contenido de los documentos. Recibe como parámetro los diez documentos almacenados en la variable texts, el embedding de tipo OpenAI, y la ubicación de la carpeta donde se almacenará la base de datos vectorial.

```
text_splitter = TokenTextSplitter(chunk_size=300, chunk_overlap=20)
texts = text_splitter.split_text(buffer.read())
vectordb = Chroma.from_texts(texts, embedding=embedding, persist_directory=persist_directory)
```

Figura 16. Variables función vectorize.

4.4.3 Comando questionDB

!questionDB <pregunta>

El comando questionDB recibe un parámetro, que es la pregunta del usuario, se procesa la pregunta para vectorizarla y va a buscar la respuesta a la base de datos de vectores almacenada en el servidor, donde se calcula la mayor similitud entre vectores para entregar la respuesta al usuario.

En la Figura 17 se puede ver el código de la función questionDB, en donde se mencionan algunas variables más relevantes:

vectordb: Se crea un objeto de tipo Chroma y se le indica en qué lugar se va almacenar la base de datos de vectores, además del tipo de embedding que se usó para vectorizar el contenido de los documentos, que en este caso se usó el modelo de OpenAI.

RetrievalQA.from_chain_type: Esta variable proveniente de la librería **langchain** está diseñada para tareas de “Question Answering” donde la respuesta se recupera de un contexto determinado. Es útil cuando existe un conjunto de documentos del que se desea extraer información.

temperature: Este parámetro controla la aleatoriedad de las respuestas generadas por el modelo de OpenAI GPT-3.5-turbo. Este valor va de cero a uno, siendo uno el más alto y

hace que las respuestas sean más creativas, pero también menos coherentes, mientras que un valor más bajo como 0.4 hace que las respuestas sean más deterministas y coherentes.

openai_api_key=TOKEN_OPENAI: Se refiere a la clave de API de OpenAI que se proporciona para autenticar y utilizar el modelo GPT-3.5-turbo. Esta clave es necesaria para acceder a las distintas funciones que ofrece OpenAI.

max_tokens: Establece el límite máximo de tokens para las respuestas generadas por el modelo. Si la respuesta excede este límite, se truncará. Esto fue útil para el desarrollo de la función, ya que un token equivale aproximadamente a cuatro caracteres en inglés, y Discord soporta hasta 2048 caracteres por mensaje. Por lo tanto, al fijar el máximo de tokens en 512, se multiplica por cuatro, dando un resultado de 2048 caracteres.

```
@commands.command(name='questionDB')
async def questionDB(self,ctx,*, question):
    embedding = OpenAIEMBEDDINGS(openai_api_key=TOKEN_OPENAI)
    persist_directory = 'db'

    vectordb = Chroma(persist_directory=persist_directory, embedding_function=embedding)
    qa = RetrievalQA.from_chain_type(llm=ChatOpenAI(temperature=0.4,
                                                    openai_api_key=TOKEN_OPENAI, model_name="gpt-3.5-turbo",
                                                    max_tokens=512), chain_type="stuff", retriever=vectordb.as_retriever())

    result = qa({"query": question})
    await ctx.send(result["result"])

    del embedding
    del vectordb
    del qa
    del result
    gc.collect()
```

Figura 17. Variables Comando questionDB.

5. RESULTADOS

A continuación, se detallan los resultados obtenidos de cada comando implementado para el **chat inteligente** en la plataforma. Además, se compartirá la experiencia adquirida al realizar demostraciones de cómo funciona la plataforma para diferentes personas. Asimismo, se llevó a cabo una evaluación de la usabilidad en colaboración con un representante del laboratorio, con el objetivo de garantizar que la plataforma satisfaga de manera eficiente las necesidades de los usuarios.

5.1 Resultado Comando questionDocument

Dada las múltiples pruebas de este comando, se notó un gran acierto en la información de la respuesta entregada, siendo clara y precisa. Resultó útil tenerla en la plataforma para los casos que los usuarios quieran centrarse en un solo documento y extraer la mayor cantidad de información. A continuación, se muestran algunas de las pruebas realizadas (ver Figura 18).

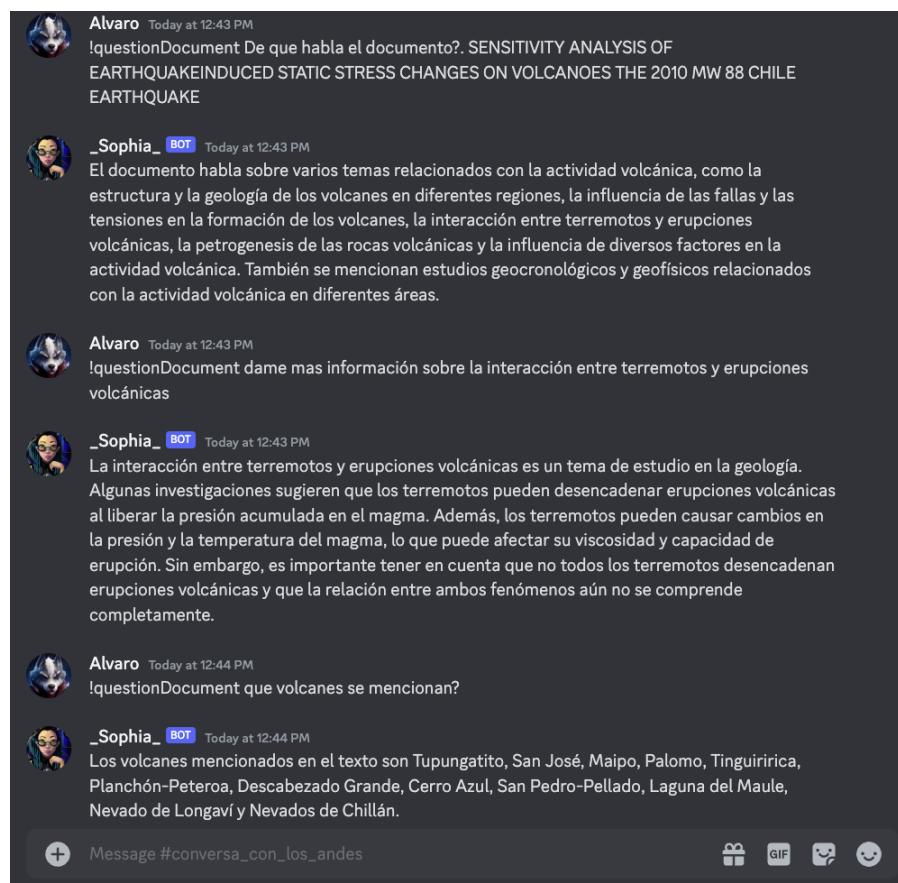


Figura 18. Resultados Comando questionDB.

5.2 Resultado Comando vectorize

El resultado del comando se presenta en forma de un mensaje en Discord que confirma la exitosa vectorización de 10 documentos. Se ejecuta las veces que sean necesarias para abarcar todos los documentos, como se mencionó anteriormente en el desarrollo del comando. Este mensaje indica que los documentos en formato string, han sido convertidos en representaciones vectoriales (ver Figura 19)

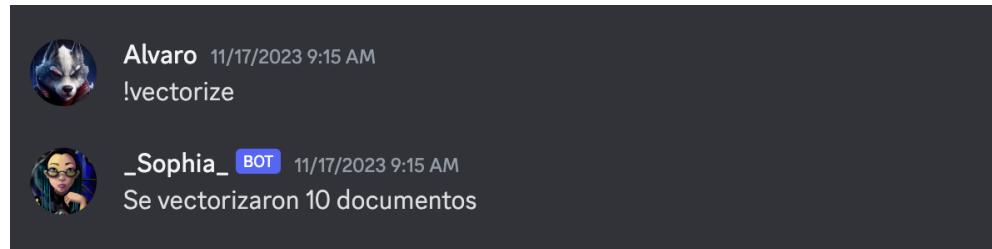


Figura 19. Resultado Comando vectorize.

5.3 Resultado Comando questionDB

Del total de pruebas realizadas, se detallan tres pruebas con el objetivo de verificar la exactitud de la información proporcionada por el comando "questionDB". Estas pruebas se estructuraron en cuatro pasos clave:

1. Se seleccionó un documento al azar de la base de datos, y a partir de su contenido se formuló una pregunta específica.
2. Se realizó la pregunta al bot de Discord.
3. Se realizó la pregunta a ChatGPT.
4. Se buscó fragmentos de la respuesta del bot en el documento de la base de datos, para comparar la respuesta con la información del documento.

5.3.1 Prueba N°1



Figura 20. Revista documento científico Prueba N°1.

Extracto del documento (documentId = 7, año 2017) de la base de datos MariaDB para formular la pregunta (Brent V. Alloway et al., 2017) (ver Figura 21).

The ability to recognize, correlate and characterize tephra of mafic to intermediate-silicic compositions that have been subjected to intense post-depositional weathering and disturbance in temperate to tropical climate environments is a significant obstacle to tephra studies in many proximal to distal volcanic settings. Not only does the intense post-depositional pedogenic weathering effectively mask fine-grained and/or thin tephra inter-beds within soil-dominated cover-beds, it potentially compromises the morphological expression of the tephra as well as its constituent geochemistry and grain-size characteristics. All attributes of which are fundamental in tephra correlation. The characterization of such tephra already susceptible to post-depositional alteration can be further complicated by compositional heterogeneity (that is, upward changes in tephra composition as an eruption progresses). Such changes may reflect sequential surface discharge of either a compositionally segregated magma body in the sub-volcanic system or a sudden magmatic recharge event into an already fractionating body. The principal objective of this study is to describe the stratigraphy, distribution, age and geochemistry of a prominent tephra marker (here formally named Lepu /C19e Tephra).

Figura 21. Pregunta Prueba N°1 extraída del fragmento de documento.

El documento de la Figura 21 menciona el estudio de un tipo específico de tefra (piroclasto) llamada **Lepue Tephra**.

Pregunta formulada: **Dame información sobre Lepu Tephra.**

Pregunta al bot de Discord (ver Figura 22).

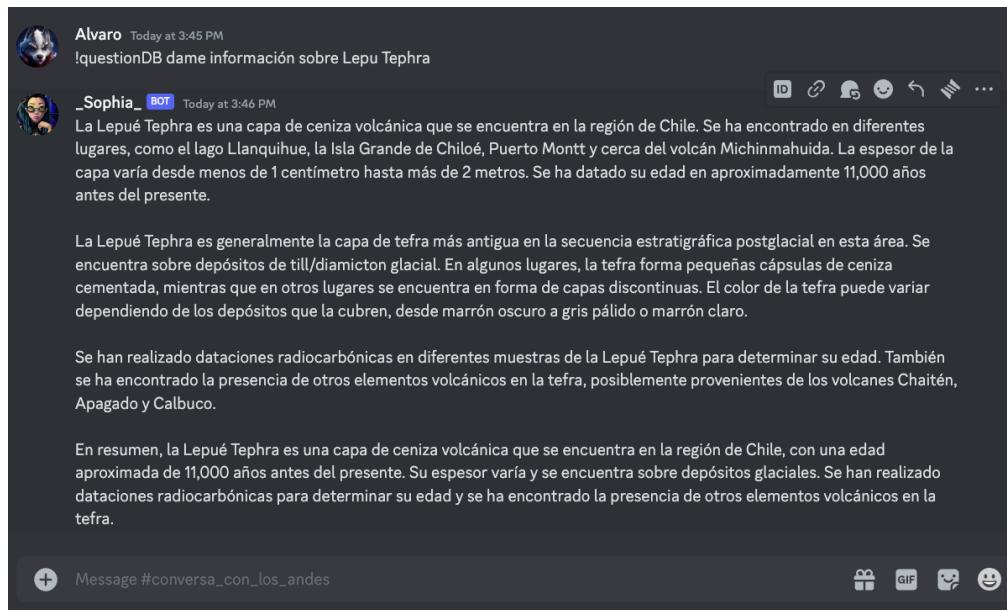


Figura 22. Respuesta del bot en Discord.

Pregunta a ChatGPT (ver Figura 23).

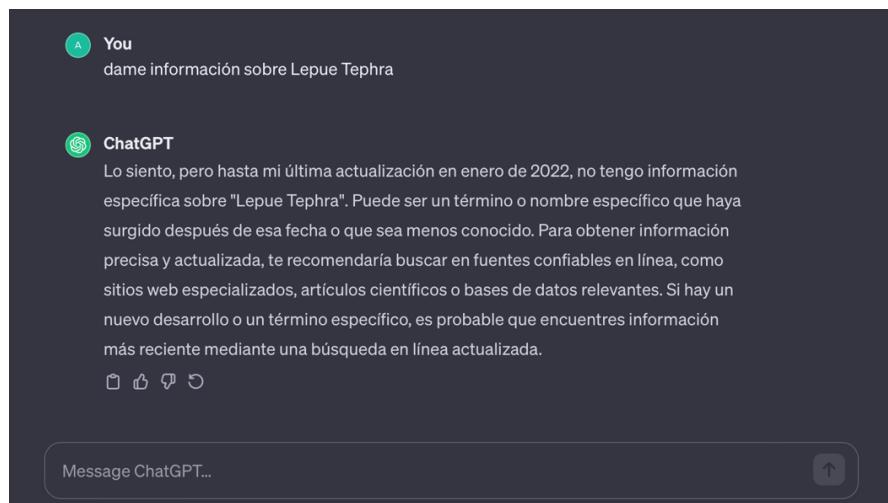


Figura 23. Respuesta de ChatGPT.

Búsqueda de la respuesta del bot en el documento de la base de datos (ver Figura 24).

(indicated in red). The positions of calibrated radiocarbon ages are indicated on the right. The inset shows a prominent macroscopic tephra encapsulated within clay and silty clays between 14.52 and 14.83 mcd and is accompanied by a prominent spike in magnetic susceptibility (916.8-m mol/C01(SI)). Unfortunately, no radiocarbon dates were obtained directly associated with the upper and lower contacts of this tephra but bracketing radiocarbon samples from 12.94 and 17.01 mcd yielded calibrated ages of 10 040 and 12 260 cal a BP, respectively.

Figura 24. Extracto de la respuesta en el documento.

Se puede ver que la respuesta del bot fue clara y con suficiente información, en comparación con la respuesta de ChatGPT, el cual no pudo encontrar información sobre el tema específico. Esta verificación se refuerza comparando el resultado de la respuesta del bot, con el contenido del documento, como el extracto de la Figura 24, en donde una de las respuestas (Edad de la Lepue Tephra, 11000 años aproximadamente) se encuentra en la frase destacada (Ages of 10 040 and 12 260).

5.3.2 Prueba N°2



Figura 25. Revista documento científico Prueba N°2.

Extracto del documento (documentId = 8, año 2017) de la base de datos MariaDB para formular la pregunta (Brent V. Alloway et al., 2017) (ver Figura 26).

coherent tephrostratigraphic framework. In order to identify VCha-sourced tephra and distinguish these from eruptive deposits from other volcanic centres, we utilize glass shard major- and trace-element compositions using grain-discrete electron microprobe (EMP) and laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) techniques, respectively (see Supplementary Information 1.0). Ultimately this comprehensive eruptive inventory

Figura 26. Pregunta Prueba N°2 extraída del fragmento de documento.

El documento de la Figura 26 menciona la técnica **Glass shard major** para el estudio de la tefra volcánica.

Pregunta formulada: **¿Para qué se usa la técnica glass shard major? En el contexto de identificar tefra VCha.**

Pregunta al bot de Discord (ver Figura 27).

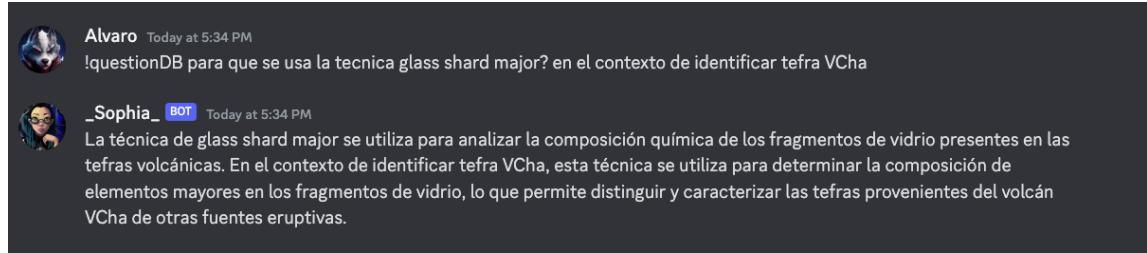


Figura 27. Respuesta del bot en Discord.

Pregunta a ChatGPT (ver Figura 28).

The screenshot shows a ChatGPT interface. A user asks about the 'glass shard major' technique. ChatGPT responds that it's a specific technique used to identify volcanic ash (tephra) from VChá. It then provides a general overview of how glass shards are analyzed in volcanic ash studies, mentioning microscopy, chemistry, radiocarbon dating, and satellite imagery. A message input field at the bottom is visible.

You
para que se usa la tecnica glass shard major? en el contexto de identificar tefra VChá

ChatGPT
Hasta mi última actualización en enero de 2022, no tengo información específica sobre una técnica llamada "glass shard major" en el contexto de identificar tefra volcánica (VChá). Sin embargo, puedo ofrecerte información general sobre cómo se realizan análisis de tefra volcánica.

En el estudio de tefra volcánica, una técnica comúnmente utilizada es el análisis de fragmentos vítreos o "glass shards". Los glass shards son fragmentos de material vítreo, generalmente de cenizas volcánicas, que se forman durante una erupción volcánica. Estos fragmentos pueden contener información valiosa sobre la composición y la historia de las erupciones volcánicas.

Para identificar y caracterizar tefra volcánica, se pueden emplear técnicas analíticas como:

1. **Microscopía de dispersión de luz:** Permite examinar la morfología y el tamaño de los glass shards bajo un microscopio.
2. **Análisis químico:** Se puede realizar mediante técnicas como la microsonda electrónica o la espectroscopía de energía dispersiva de rayos X (EDS) para determinar la composición química de los glass shards.
3. **Datación por radiocarbono o métodos isotópicos:** Puede proporcionar información sobre la edad de las capas de tefra y la frecuencia de las erupciones.
4. **Análisis de imagen por satélite:** Puede utilizarse para mapear la distribución de tefra volcánica en una región.

Message ChatGPT...

Figura 28. Respuesta de ChatGPT.

Búsqueda de la respuesta del bot en el documento de la base de datos (ver Figura 29).

coherent tephrostratigraphic framework. In order to identify VCha-sourced tephra and distinguish these from eruptive deposits from other volcanic centres, we utilize glass shard major- and trace-element compositions using grain-discrete electron microprobe (EMP) and laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) techniques, respectively (see Supplementary Information 1.0). Ultimately this comprehensive eruptive inventory

Figura 29: Extracto de la respuesta en el documento.

Se puede ver que la respuesta del bot hizo mención sobre “caracterizar las tefras provenientes del volcán VCha (Chaitén) de otras fuentes eruptivas”, en comparación con la respuesta de ChatGPT, el cual a pesar de que si supo lo que era la técnica de “glass shard major”, no pudo dar una respuesta en el contexto del volcán VCha. Esta verificación se refuerza comparando el resultado de la respuesta del bot, con el contenido del documento, como se destaca en el extracto de la Figura 29, donde ambas coinciden en identificar material piro clástico dentro del volcán Chaitén proveniente de otros volcanes.

5.3.3 Prueba N°3

Geochemistry, Geophysics, Geosystems®

Research Article |  Free Access

Dynamics of Outgassing and Plume Transport Revealed by Proximal Unmanned Aerial System (UAS) Measurements at Volcán Villarrica, Chile

Emma J. Liu , Kieran Wood, Emily Mason, Marie Edmonds, Alessandro Aiuppa, Gaetano Giudice, Marcello Bitetto, Vincenzo Francofonte, Steve Burrow, Thomas Richardson ... See all authors 

First published: 28 December 2018 | <https://doi.org/10.1029/2018GC007692> | Citations: 32

Figura 30. Revista documento científico Prueba N°3.

Extracto del documento (documentId = 188, año 2018) de la base de datos MariaDB para formular la pregunta (Emma J. Liu et al., 2018) (ver Figura 31).

Abstract

Volcanic gas emissions are intimately linked to the dynamics of magma ascent and outgassing, and, on geological timescales, constitute an important source of volatiles to the Earth's atmosphere. Measurements of gas composition and flux are therefore critical to both volcano monitoring and to determining the contribution of volcanoes to global geochemical cycles. However, significant gaps remain in our global inventories of volcanic emissions, (particularly for CO₂, which requires proximal sampling of a concentrated plume) for those volcanoes where the near-vent region is hazardous or inaccessible. Unmanned Aerial Systems (UAS) provide a robust and effective solution to proximal sampling of dense volcanic plumes in extreme volcanic environments. Here, we present gas compositional data acquired using a gas sensor payload aboard a UAS flown at Volcán Villarrica, Chile. We compare UAS –

Figura 31. Pregunta Prueba N°3 extraída del fragmento de documento.

El documento de la Figura 31 menciona **el uso de UAS (Aeronaves No Tripuladas)** en el volcán Villarrica para el muestreo proximal de la densidad de la pluma volcánica.

Pregunta formulada: **Dame información sobre UAS volado en el Volcán Villarrica.**

Pregunta al bot de Discord (ver Figura 32).

Alvaro Today at 6:36 PM
!questionDB dame información sobre UAS volado en el Volcán Villarrica

Sophia_BOT Today at 6:36 PM
El Volcán Villarrica en Chile ha sido utilizado como laboratorio natural para probar y estudiar el uso de Sistemas Aéreos No Tripulados (UAS, por sus siglas en inglés) en la medición de emisiones volcánicas. En un estudio reciente, se utilizó un prototipo comercial de sensor (Airgraph Aeris) montado en un cuadricóptero DJI Phantom 3 Pro para mapear el dióxido de azufre (SO₂) en grandes escalas en la pluma del volcán.

Los resultados de este estudio revelaron un comportamiento periódico de desgasificación previamente no reconocido en el Volcán Villarrica. Además, demostraron la validez de los UAS para estudios de alta resolución y monitoreo rutinario de emisiones volcánicas. El estudio también sugiere una lista de verificación previa al vuelo para garantizar la operación segura de los UAS.

El Volcán Villarrica es un estratovolcán parcialmente glaciar ubicado en la Zona Volcánica del Sur de los Andes de Chile. Esta zona es relativamente pobre en carbono en comparación con otras zonas volcánicas en el mundo. El volcán es históricamente uno de los más activos en Chile y alberga un lago de lava abierto en su cumbre. Esto lo convierte en un lugar ideal para probar y estudiar el uso de UAS en la medición de emisiones volcánicas.

En el estudio mencionado, se utilizaron dos tipos de UAS instrumentados: un cuadricóptero Vulcan Black Widow modificado para llevar un equipo de medición en tiempo real y resistir las condiciones ambientales desafiantes en la pluma volcánica, y un prototipo comercial de sensor (Airgraph Aeris) que se acopló al cuadricóptero DJI Phantom 3 Pro.

En resumen, el estudio demuestra el potencial de los UAS para medir y monitorear de manera segura y eficiente las emisiones volcánicas en el Volcán Villarrica y otros volcanes inaccesibles. Sin embargo, no se proporcionan detalles específicos sobre los

Message #conversa_con_los_andes

Figura 32. Respuesta del bot en Discord.

Pregunta a ChatGPT (ver Figura 33).

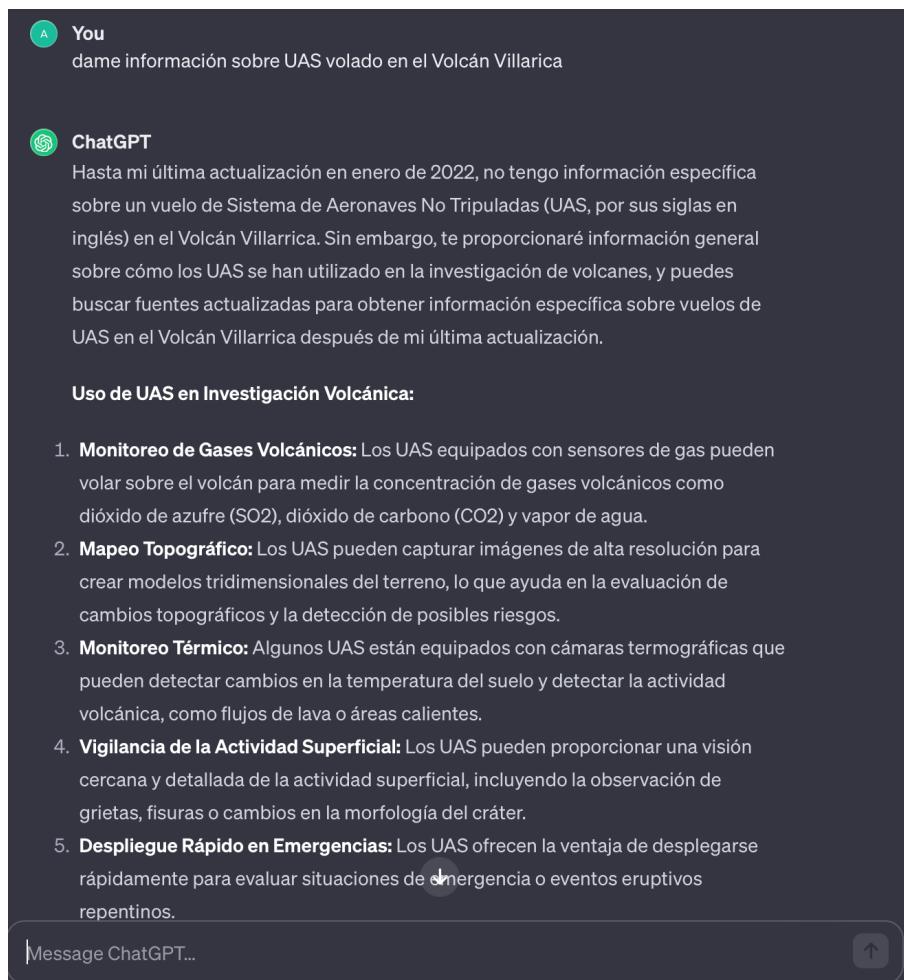


Figura 33. Respuesta de ChatGPT.

Búsqueda de la respuesta del bot en el documento de la base de datos (ver Figura 34).

© 2018 American Geophysical Union. All rights reserved.
Figure 1: Location and volcanological setting ; (a) Volcán Villarrica is the easternmost in a chain of three volcanoes (Lanín, Quetrupillan) striking obliquely to the main N-S axis of volcanism; (b, c) The magma level was extremely low in the conduit during the measurement campaign, with the lake surface only visible as several pixels in aerial imagery; (d) UAS were launched from a sheltered plateau on the northern rim of the crater, with the semi-permanent multi-GAS station visible on the eastern rim; (e) Location map of the region, showing the position of UV camera. The green shaded region delimits the extent of the national park. Inset: Aerial map of the summit region shown in (d). The summit crater is ~200 m in diameter; (f) Two instrumented multi-rotor vehicles were used in this campaign, the Vulcan octocopter with multi-GAS (left) and DJI Phantom 3 Pro with Aeris gas sensor (right); (g) Vulcan UAS in flight on 20/03/18.

Figura 34. Extracto de la respuesta en el documento.

Dada la respuesta del bot que menciona la utilización de dos tipos de UAS instrumentados: un cuadricóptero Vulcan Black Widow para el equipo de medición y un prototipo

comercial de sensor Airgraph Aeris que se acopló al cuadricóptero DJI Phantom 3 Pro, se puede verificar que la respuesta se encuentra en el fragmento destacado del documento en la Figura 34, lo cual se puede apreciar una alta coincidencia en la respuesta.

5.4 Demostración a CONAF

Se llevó a cabo una reunión estratégica con el equipo de CONAF Valdivia, motivada por su interés en nuestro proyecto y el deseo de explorar posibles beneficios futuros. Durante esta sesión, se organizó una demostración integral en la que participaron usuarios del Laboratorio Natural Andes del Sur (LNAS), incluyendo al profesor Matthieu Vernier, ambos tesis y representantes de CONAF. La demostración tuvo como objetivo ilustrar de manera práctica el funcionamiento del bot, exhibiendo sus funcionalidades clave y proporcionando una visión detallada de cómo podría contribuir a las necesidades específicas del equipo de CONAF. Esta interacción directa permitió a todos los participantes comprender mejor el potencial del proyecto y abrió un espacio para discusiones constructivas sobre posibles colaboraciones y adaptaciones del sistema para satisfacer las necesidades particulares de CONAF Valdivia (ver Figura 35).



Figura 35. Reunión CONAF.

5.5 Evaluación de la usabilidad de la plataforma con participante del laboratorio

Se realizó una prueba de usabilidad, donde nos reunimos con el usuario representante del laboratorio Javiera y junto al tesista Fernando, con el objetivo de evaluar la eficacia y la facilidad de uso de las funcionalidades del chat inteligente. Esta prueba se hizo con constante comunicación a través del canal de voz dentro del servidor Discord. Inicialmente, se guió al participante a través de los pasos necesarios para acceder a la plataforma Discord y unirse al servidor específico. Posteriormente, se proporcionó un manual de uso que incluía instrucciones sobre los comandos disponibles. Esto permitió

una interacción más fluida y efectiva con el chat inteligente, asegurando una mejor comprensión de las funciones disponibles.

Esta prueba de usabilidad se hizo con los nombres antiguos de los comandos, ya que la petición de cambio de nombre se hizo después de la prueba.

Plataforma

Primero se realizó la pregunta sobre qué tan cómodo se siente usando la plataforma Discord (ver Figura 36).



Figura 36. Pregunta usabilidad de la plataforma.

Su respuesta en la Figura 36 fue que se sentía normal, ya que, el ambiente donde trabaja LNAS no es común usar la plataforma Discord. A pesar de que no se está muy familiarizado, no fue problema ingresar a la plataforma.

Comando question (questionDocument)

Después se le preguntó sobre la dificultad de usar el comando “question”, en una escala de menor a mayor dificultad, donde uno es lo más fácil. (ver Figura 37).

¿Fue fácil entender cómo formular una pregunta usando el comando !question?

1 respuesta

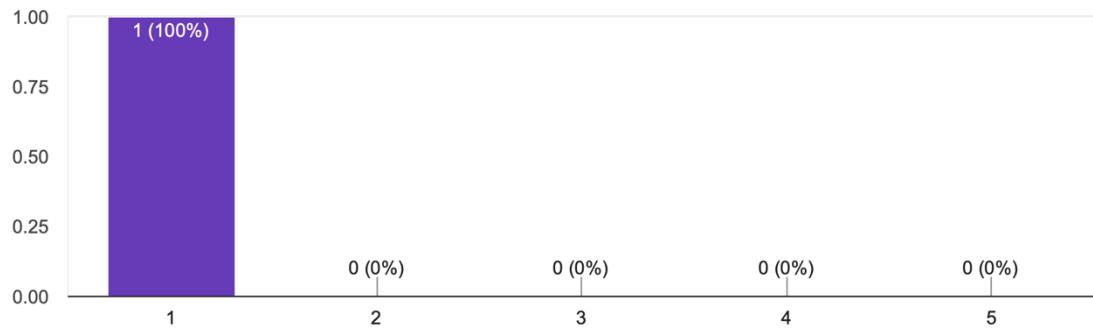


Figura 37. Pregunta usabilidad comando question.

Dada la respuesta entregada en la Figura 37, el representante comentó que no hubo problema en entender cómo usar el comando, y su experiencia fue fluida. En las figuras 38 y 39 se presentan dos de las preguntas que realizó Javiera.

javivovu 12/13/2023 3:53 PM
!question de qué habla este documento?. TEPHRA STRATIGRAPHY AND ERUPTIVE VOLUME OF THE MAY 2008 CHAITÉN VOLCANO ERUPTION CHILE

Sophia BOT 12/13/2023 3:53 PM
Este documento habla sobre el volcán Chaitén y la erupción que tuvo lugar en mayo de 2008. Describe la cronología de los eventos eruptivos, la dispersión de las cenizas y los diferentes depósitos estratigráficos resultantes de la erupción. También se mencionan las características de los diferentes tipos de capas y su espesor, así como las condiciones meteorológicas en las que se depositaron. Además, se proporciona información sobre la actividad sísmica y las explosiones asociadas con la erupción.

Figura 38. Pregunta N°1 usuario y respuesta del bot comando question.

javivovu 12/13/2023 3:54 PM
!question qué tan activo es el volcán Chaitén?

Sophia BOT 12/13/2023 3:54 PM
El volcán Chaitén actualmente no está entregando una cantidad significativa de cenizas a las áreas pobladas. Sin embargo, existe el riesgo de remobilización de cenizas o un aumento potencial en la actividad volcánica, lo que podría representar un riesgo para las poblaciones expuestas.

Figura 39. Pregunta N°2 usuario y respuesta del bot comando question.

También se le preguntó a Javiera si el comando question le agrega valor a la experiencia de interactuar con la plataforma, su respuesta fue positiva ya que consideró que poder

realizar preguntas específicas a un solo documento es de gran ayuda para extraer el contenido más relevante.

Respecto a la respuesta del bot, se le preguntó qué tan clara y fácil de interpretar fue la información entregada (ver Figura 40). En una escala de menos a más claridad, su respuesta fue positiva, indicando que no tuvo problemas en comprender y analizar las respuestas entregadas.

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

1 respuesta

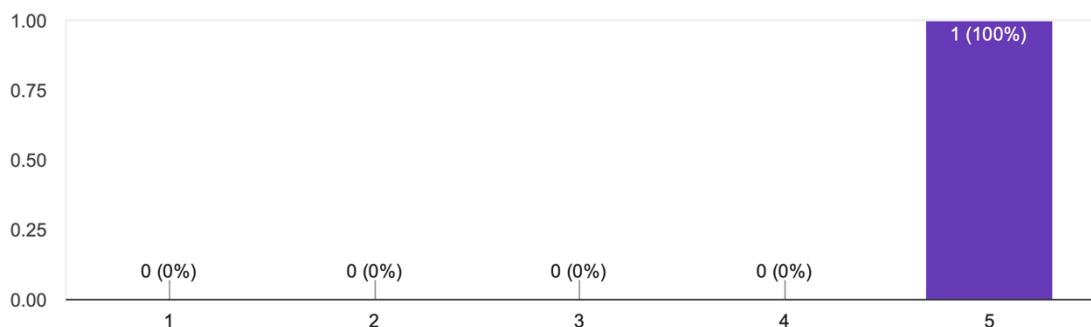


Figura 40. Pregunta sobre la claridad de la respuesta comando question.

Comando **query_chroma (questionDB)**

Al igual que el comando question, se le preguntó sobre la dificultad de usar el comando **query_chroma**, en una escala de menor a mayor dificultad, donde uno es lo más fácil. (ver Figura 41).

¿Fue fácil entender cómo usar el comando !query_chroma?

1 respuesta

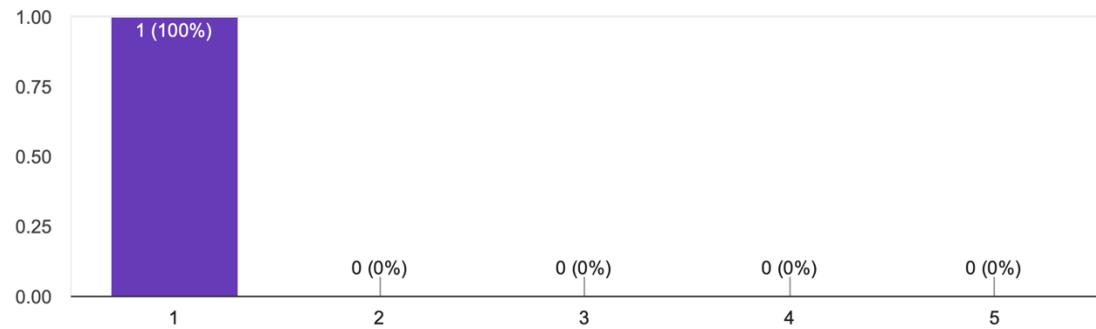


Figura 41. Pregunta usabilidad comando query_chroma.

No se encontraron problemas por parte de Javiera en entender el funcionamiento del comando **query_chroma**.

En las figuras 42, 43 y 44 se presentan tres de las preguntas que realizó Javiera.

javivovu 12/13/2023 3:41 PM
!query_chroma que estudios sobre glaciares se mencionan?

Sophia BOT 12/13/2023 3:41 PM
Se mencionan varios estudios sobre glaciares:

1. "Inventario de glaciares entre las cuencas de los ríos Bío Bío y Petrohue: su relación con el volcanismo activo: caso Volcán Lonquimay" - Tesis de licenciatura de A. Rivera, Universidad de Chile, 1989.
2. "Ice thickness measurements on the Southern Patagonia Icefield" - A. Rivera y G. Casassa, en "The Patagonian Icefields: a unique natural laboratory for environmental and climate change studies", 2002.
3. "Mediciones de espesores en glaciares de Chile centro-Sur" - A. Rivera, G. Casassa y C. Acuña, Rev. Invest. Geogr., 2001.
4. "Use of remotely-sensed and field data to estimate the contribution of Chilean glaciers to eustatic sea-level rise" - A. Rivera, C. Acuña, G. Casassa y F. Bown, 2002.

Además, se mencionan otros estudios relacionados con los glaciares en la región de Patagonia y en Chile en general, como "Ice elevation and areal changes of glaciers from the Northern Patagonia Icefield, Chile" y "Valdivia 1960. Entre aguas y escombros".

Figura 42. Pregunta N°1 usuario y respuesta del bot comando query_chroma.

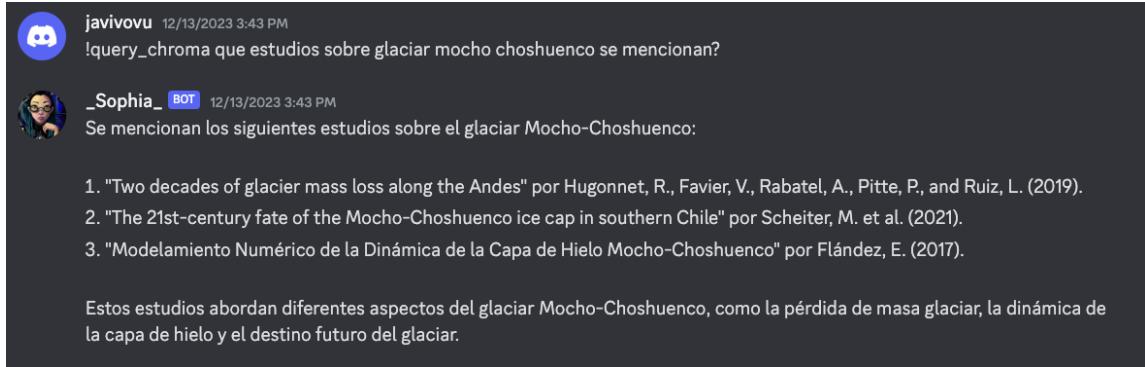


Figura 43. Pregunta N°2 usuario y respuesta del bot comando query_chroma.

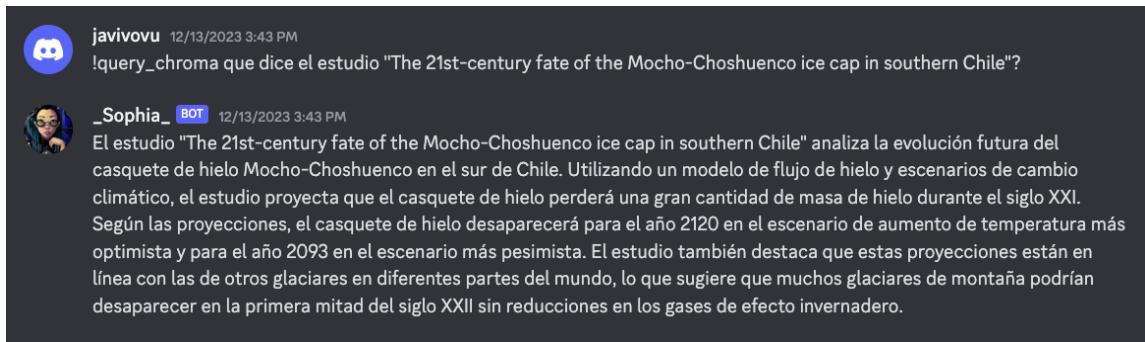


Figura 44: Pregunta N°3 usuario y respuesta del bot comando query_chroma.

Se le preguntó sobre la claridad y facilidad de interpretar la respuesta del bot, en una escala de menos a más claridad (ver Figura 45). Su respuesta fue positiva, mencionando que quedó conforme con las respuestas proporcionadas, siendo este comando útil para cuando se quieren hacer preguntas generales y no solo de un documento en específico. También consideró que este comando le agrega valor a la experiencia de interactuar con la plataforma.

¿La información devuelta por el comando !query_chroma es clara y fácil de interpretar?
1 respuesta

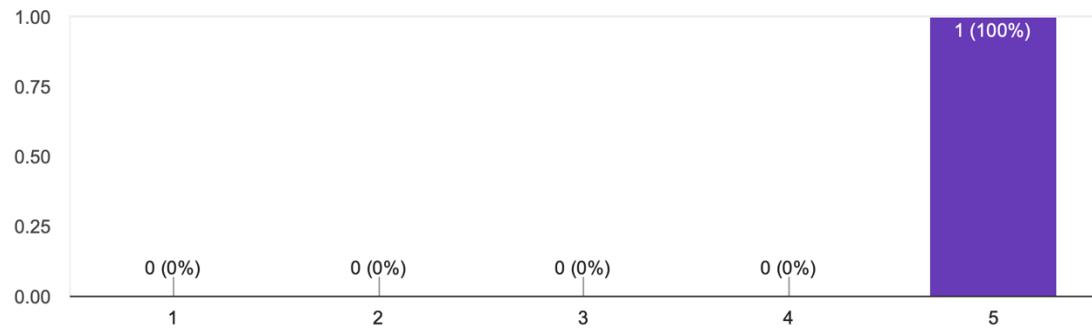


Figura 45. Pregunta sobre la claridad de la respuesta comando query_chroma.

Javiera tuvo una sugerencia al terminar de usar ambos comandos (question y query_chroma), sobre cambiar el nombre del comando query_chroma, para que sea parecido al de question pero haciendo la diferencia que “**question**” es para un documento en específico, mientras que “**query_chroma**” es para consultar la base de datos en general.

Dada esta sugerencia, se decidió hacer el cambio de nombre de ambos comandos:

question: **questionDocument**

query_chroma: **questionDB**

Finalmente, se le preguntó a Javiera cómo evalúa el funcionamiento del bot en general, su respuesta fue que sin estar familiarizada con el uso de Discord, le pareció cómodo, fácil y rápido de usar. También mencionó que el funcionamiento es claro si viene acompañado de una buena explicación, por ello es relevante proporcionar un buen manual de usuario.

6. CONCLUSIONES

La plataforma Discord demostró ser cómoda para usuarios que no la habían utilizado previamente, destacando su facilidad de uso. La presencia de un entorno comunitario, que permite compartir resultados, tanto en canales de texto como a través de comunicación por voz, añadió un valor significativo a la experiencia de los usuarios.

La utilidad del comando **questionDocument** se destacó en sus respuestas al enfocarse en un solo documento, proporcionando información más precisa y específica. Los beneficios de esta aproximación incluyen una mayor precisión en las respuestas, facilitando la obtención de información relevante y detallada.

La funcionalidad del comando **vectorize** fue de gran apoyo para tener todos los documentos de la base de datos vectorizados almacenados localmente, reduciendo tanto el tiempo como el costo monetario en comparación a si se estuviera vectorizando cada vez que se ejecuta el programa. Además, con la ventaja que, si se quieren agregar nuevos documentos, se pueden añadir a los vectores existentes sin tener que reemplazarlos.

El comando **questionDB** tuvo un buen desempeño en la entrega de respuestas a preguntas sobre el contenido de toda la base de datos de vectores. Dado el control de calidad realizado en las respuestas, comparando el bot de Discord y ChatGPT, se notó una gran diferencia entre las respuestas, ya que en la mayoría de las pruebas respondía mejor el bot, entregando respuestas específicas del contenido de los documentos.

La experiencia de realizar demostraciones con distintos usuarios y pruebas de usabilidad con un representante de LNAS, proporcionaron valiosas percepciones sobre la facilidad de uso, la comprensión de las funciones y la interactividad general.

Dado el funcionamiento de la plataforma **Conversa Con los Andes**, se espera que los beneficios generados tengan un alcance significativo y contribuyan al desarrollo de LNAS y a la generación de impacto en la sociedad. Entregando mayor conocimiento científico y aumentando el interés de participar en proyectos para los estudiantes universitarios del área de ciencias.

Finalmente se pudo cumplir con el objetivo general de LNAS “Generar una plataforma que permita mejorar el acceso a conocimientos científicos y técnicos sobre el territorio de Los Andes del Sur mediante técnicas de análisis de lenguaje”, ya que a pesar de que fue un desarrollo experimental de chat inteligente, este logró poder satisfacer las necesidades de acceder a la información científica de una forma más fácil.

7. REFERENCIAS

Agencia Nacional De Investigación y Desarrollo. (2021). *Desarrollo de Nodos de Laboratorios Naturales*. <https://anid.cl/redes-estrategia-y-conocimiento/desarrollo-de-nodos-de-laboratorios-naturales/>

Aokun Chen, Zehao Yu, Xi Yang, Yi Guo, Jiang Bian, Yonghui Wu (2023). *Contextualized medication information extraction using Transformer-based deep learning architectures*. (n.d.).
<https://www.sciencedirect.com/science/article/pii/S1532046423000916>

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

Brent V. Alloway, Nick J.G. Pearce , Patricio I. Moreno , Gustavo Villarosa , Ignacio Jara , Ricardo De Pol-Holz , Valeria Outes (2017). *An 18,000 year-long eruptive record from Volcán Chaitén, northwestern Patagonia: Paleoenvironmental and hazard-assessment implications*. Quaternary Science Reviews.
<https://www.sciencedirect.com/science/article/abs/pii/S0277379117300549>

Brent V. Alloway, Patricio I. Moreno, Nick J. G. Pearce, Ricardo De Pol-Holz, William I. Henríquez, Oscar H. Pesce, Esteban Sagredo, Gustavo Villarosa, Valeria Outes. (2017). *Stratigraphy, age and correlation of Lepué Tephra: a widespread c. 11 000 cal a BP marker horizon sourced from the Chaitén Sector of southern Chile*. Journal of Quaternary Science.
<https://onlinelibrary.wiley.com/doi/10.1002/jqs.2976>

ChromaDB Documentation. Accedido el 1 de Octubre, 2023, desde
<https://docs.trychroma.com/usage-guide>

Diagnóstico de las montañas de Chile. (2015). ResearchGate. Accedido el 18 de Mayo, 2023, desde
https://www.researchgate.net/publication/281408166_Diagnostico_de_las_montanas_de_Chile

Emma J. Liu, Kieran Wood, Emily Mason, Marie Edmonds, Alessandro Aiuppa, Gaetano Giudice, Marcello Bitetto, Vincenzo Francofonte, Steve Burrow, Thomas Richardson, Matthew Watson, Tom D. Pering, Thomas C. Wilkes, Andrew J. S. McGonigle, Gabriela Velasquez, Carlos Melgarejo, Claudia Bucarey (2018). *Dynamics of Outgassing and Plume Transport Revealed by Proximal Unmanned Aerial System (UAS) Measurements at Volcán Villarrica, Chile*.
<https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018GC007692>

Exitoso Mapeo Participativo en Puerto Varas: – Andes del Sur de Chile. (n.d.). Laboratorio Natural Andes del Sur. Accedido el 4 de Junio, 2023, desde <https://andesdelsurlab.cl/mapeo-participativo-puerto-varas/>

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>

Höhn, F. (2022). *Natural Language Question Answering Systems – Get Quick Answers To Concrete Questions*. SEEBURGER Blog. Accedido el 2 de Junio, 2023, desde <https://blog.seeburger.com/natural-language-question-answering-systems-get-quick-answers-to-concrete-questions/>

Ladera 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/>

LLamaindex Documentation. Accedido el 10 de julio, 2023, desde <https://docs.llamaindex.ai/en/stable/>

LongChain Documentation. Accedido el 10 de julio, 2023, desde https://python.langchain.com/docs/get_started/introduction

Luigi de Angelis, Francesco Baglivo, Gugliemo Arzilli, Gaetano Pierpaolo, Paolo Ferragina, Alberto Tozzi, Caterina Rizzo (2023). *ChatGPT and the rise of large Language models: the new AI-driven infodemic threat in public health.* (2023, April 25). Frontiers. Accedido el 20 de Junio, 2023, desde <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1166120/full>

Nasr, M., Karam, A., Atef, M., Boles, K., Samir, K., & Raouf, M. (2020). "Natural Language Processing: Text Categorization and Classifications." - Wiktionary. Accedido el 18 de Mayo, 2023, desde https://www.researchgate.net/publication/345647920_Natural_Language_Processing_Text_Categorization_And_Classifications

OpenAI Documentation. Accedido el 1 de julio, 2023, desde <https://platform.openai.com/docs/api-reference/authentication>

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

Resultados del análisis bibliométrico – Andes del Sur de Chile. (2022). Laboratorio Natural Andes del Sur. Accedido el 4 de Junio, 2023, desde <https://andesdelsurlab.cl/resultado-analisis-bibliometrico/>

"Transformer Models. (n.d.). " - Wiktionary. Accedido el 4 de Junio, 2023, desde <https://amatriain.net/blog/transformer-models-an-introduction-and-catalog-2d1e9039f376/?fbclid=IwAR3XWbd7E12GUFi8HneUpE5O22MiK2pkAjQaJgp7dKTVhteNnhP2kaf5IY4>

Zhixue Jiang (2021). *Research on Medical Question Answering System Based on Knowledge Graph.* (n.d.). ResearchGate. Accedido el 20 de Junio, 2023, desde https://www.researchgate.net/publication/348854825_Research_on_Medical_Question_Answering_System_Based_on_Knowledge_Graph