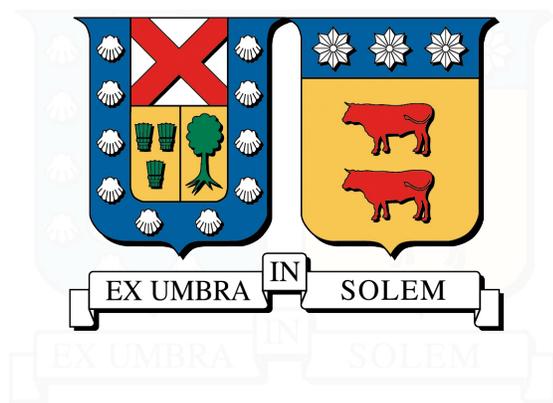


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA E INFORMÁTICA
CONCEPCION - CHILE



**TÍTULO DE MEMORIA (EL TÍTULO SÓLO PUEDE TENER UN MÁXIMO DE 3
LÍNEAS)**

ABNER ABDIEL ASTETE HERNÁNDEZ

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO INFORMÁTICO

PROFESOR GUÍA : SR. ISRAEL FIGUEROA

6 de septiembre de 2023



Mi Dedicatoria ...

[Mi mensaje]

(AGRADECIMIENTOS) [Título es opcional]

Ingrese sus agradecimientos.

Gracias!!!



Índice de Contenidos

Índice de Contenidos	v
Índice de Tablas	vii
Índice de Figuras	ix
1. Introducción	1
1.1. Planteamiento	1
1.2. Estructura de la Tesis	2
1.3. Significado de la Investigación	2
2. Investigación	3
2.1. Marco teórico	3
2.2. Bibliotecas Relevantes	3
2.3. Tecnologías Evaluadas	3
3. Objetivos	5
3.1. Objetivo General	5
3.2. Objetivos Secundarios	5
4. Solución Propuesta	7
4.1. Descripción	7
4.2. Justificaciones	8
5. Resultados	9
5.1. Benchmark	9
6. Resultados-Trabajos Futuros	11
6.1. Resultados	11
6.2. Trabajos Futuros	11
A. Primer Anexo	13
B. Segundo Anexo	15



Índice de Tablas





Índice de Figuras





Capítulo 1

Introducción

1.1. Planteamiento

La presente tesis tiene como objetivo principal proporcionar una solución eficiente para la gestión de registros en los archivos del General Transit Feed Specification (GTFS) estáticos. El problema central al que nos enfrentamos radica en la falta de una solución de código abierto que permita comparar y gestionar de manera efectiva tanto los registros nuevos como los antiguos que se generan a lo largo del tiempo en estos archivos. Esta problemática cobra especial relevancia en el contexto de la ciudad de Concepción, donde se presentan necesidades específicas relacionadas con su conjunto de datos GTFS. No obstante, la intención primordial es que cualquier individuo pueda aprovechar este componente y aplicarlo en sus proyectos, con la flexibilidad de adaptarlo fácilmente a los GTFS que estén utilizando.

El componente desarrollado en esta investigación se integra con dos bases de datos fundamentales. En la primera base de datos, se almacena información del usuario que ingresa los archivos GTFS en formato ZIP. Esta base de datos no solo sirve para el registro de los archivos, sino también para mantener un registro de quién realizó la carga, la fecha y la hora de ingreso, permitiendo así una gestión efectiva de la trazabilidad de los datos. De manera significativa, este sistema automatizará un proceso que previamente requería actualizaciones manuales, eliminando la necesidad de gestionar registros en papel o mediante hojas de cálculo como Excel.

La segunda base de datos alberga los registros previamente almacenados y es esencial para la comparación y gestión de los archivos GTFS. La capacidad de identificar tanto los registros nuevos como los eliminados en cada actualización de archivos es una característica fundamental de este componente. Una vez que se realice una actualización de los archivos GTFS, se podrá identificar claramente qué registros son nuevos y cuáles han sido eliminados.

1.2. Estructura de la Tesis

La tesis se organiza de la siguiente manera para abordar el problema planteado:

- **Capítulo 1:** Establece el contexto, justificación y objetivos de la investigación.
- **Capítulo 2:** Se revisará la literatura relevante en el campo de los archivos GTFS, bibliotecas relevantes y tecnologías evaluada.
- **Capítulo 3:** Detalla las problemáticas tanto de un punto de vista general como secundarias.
- **Capítulo 4:** Presenta la solución propuesta, explicando su propuesta y justificaciones.
- **Capítulo 5:** Demostración del funcionamiento y como lograr integrar el componente en otros sistemas y proyectos.
- **Capítulo 6:** Se presentan resultados y ideas para trabajos futuros.

1.3. Significado de la Investigación

Esta investigación no solo busca abordar la gestión de registros en archivos GTFS, sino que también se enfoca en aspectos como la trazabilidad de los datos, la identificación de nuevos registros y la eliminación de registros obsoletos. El registro detallado de quién carga los archivos y cuándo lo hacen añade una pequeña capa adicional de control y supervisión.

La aplicación de este componente no solo beneficiará a la ciudad de Concepción, sino que también tiene un alcance más amplio, permitiendo a otros usuarios adaptarlo a sus propios proyectos de gestión de datos de transporte público. Además, esta tesis contribuirá al enriquecimiento del conocimiento en el campo de la gestión de datos de movilidad urbana, ofreciendo una solución valiosa para las necesidades actuales en este ámbito.

Esta tesis se propone como una herramienta valiosa para optimizar la administración de registros en archivos GTFS, con un enfoque inicial en Concepción y la perspectiva de una aplicación más amplia en el ámbito de la movilidad urbana, brindando un mayor control y trazabilidad de los datos en este contexto.

Capítulo 2

Investigación



2.1. Marco teórico

La presente sección tiene como objetivo proporcionar un sólido fundamento conceptual para nuestro proyecto. A lo largo de estas páginas, exploraremos conceptos y antecedentes que sustentan la investigación y desarrollo de este proyecto. El marco teórico es esencial para comprender el contexto en el que se enmarca nuestro proyecto y para identificar las contribuciones que este aportará al campo.

2.2. Bibliotecas Relevantes

Primer párrafo.

2.3. Tecnologías Evaluadas

Primer párrafo.



Capítulo 3

Objetivos

3.1. Objetivo General

El objetivo general de este proyecto es diseñar, desarrollar e implementar un sistema eficiente de automatización para la gestión de registros en archivos GTFS, que permita la identificación automática de registros nuevos y eliminados, así como el seguimiento de la trazabilidad de los datos de manera eficaz. Este sistema busca eliminar la dependencia de procesos manuales y proporcionar una solución integral que aumente la eficiencia y la precisión en la gestión de datos de transporte público, con un enfoque inicial en la ciudad de Concepción y la capacidad de adaptarse a otros contextos y proyectos similares.

3.2. Objetivos Secundarios

- **Implementar algoritmos de comparación eficientes:** Desarrollar algoritmos efectivos que puedan comparar registros nuevos con los registros existentes en la base de datos y detectar cambios o actualizaciones.
- **Implementar la asignación de cargas a registros:** Identificar y asignar a qué carga específica pertenece cada registro en el sistema. Esto facilitará el seguimiento de la procedencia de los datos y ayudará a los usuarios a gestionar y analizar los registros de manera más precisa y organizada.
- **Visualización de registros nuevos y antiguos:** Desarrollar una funcionalidad en el sistema que permita identificar y mostrar de manera clara al usuario qué registros son nuevos y cuáles son antiguos después de cada actualización de archivos GTFS. Esto facilitará que los usuarios puedan tomar decisiones informadas basadas en los cambios detectados en los datos.
- **Facilitar la escalabilidad:** Diseñar el sistema de manera que pueda escalarse fácilmente para adaptarse a un mayor volumen de datos o a diferentes contextos de aplicación.
- **Documentar el sistema y los procedimientos:** Preparar documentación detallada que explique cómo utilizar el sistema, cómo realizar mantenimiento y cómo solucionar problemas comunes.
- **Evaluar la eficiencia del sistema:** Medir y evaluar la eficiencia y la efectividad del sistema en términos de tiempo, recursos y precisión en la gestión de registros.
- **Explorar la adaptabilidad a otros contextos:** Explorar la adaptabilidad a otros contextos: Investigar la posibilidad de adaptar el sistema a diferentes ciudades o proyectos de transporte público, evaluando su aplicabilidad en diferentes situaciones.



Capítulo 4

Solución Propuesta

4.1. Descripción

Para abordar el desafío de la gestión de archivos GTFS, nuestra solución se basa en la flexibilidad y adaptabilidad para satisfacer las necesidades de cualquier persona que desee actualizar sus datos GTFS. El enfoque principal fue crear un módulo que pudiera conectarse a una base de datos existente. Es importante destacar que algunas partes de esta solución se diseñaron específicamente para demostrar su aplicabilidad en un entorno del mundo real.

Parte del proceso de desarrollo implicó establecer una relación efectiva entre los archivos GTFS. Aunque Google proporciona un estándar, es esencial adaptarlo a las necesidades particulares de Concepción. En esta línea, creamos un modelo que nos permite registrar tanto el archivo ZIP como el usuario que lo cargó.

Con el contexto del proyecto en mente, pasamos a explicar las herramientas que utilizamos en su implementación. Para la creación de la API, optamos por el uso del framework FastAPI, que se basa en el lenguaje Python. Además, empleamos SQLAlchemy como ORM para realizar consultas a la base de datos directamente desde el código y utilizamos pandas para interactuar con los archivos "Txt" que manejamos.

El proceso de actualización y gestión de archivos GTFS se divide en varios pasos clave:

- **Identificación y Validación del ZIP:** Verificamos y validamos el archivo ZIP cargado para asegurarnos de que cumple con los requisitos necesarios.
- **Extracción y Validación de los Txt:** Extraemos los archivos "Txt" del archivo ZIP y los validamos para garantizar que sean compatibles con el formato GTFS.
- **Transformación y Validación de los Dataframes:** Convertimos los datos de los archivos "Txt" en dataframes que puedan ser procesados y validados.
- **Verificación de Registros Antiguos en la Base de Datos:** Comparamos los registros en los dataframes con los registros antiguos almacenados en la base de datos para identificar las actualizaciones.
- **Verificación de Registros Nuevos en la Base de Datos:** Detectamos y registramos los registros nuevos que se agregarán a la base de datos.
- **Guardar ZIP Utilizado durante la Carga:** Mantenemos un registro del archivo ZIP utilizado en la carga para futuras referencias.

Una vez completado el proceso, presentamos a los usuarios una vista que muestra los registros que fueron eliminados y los nuevos registros que se agregaron. Esta visualización proporciona una comprensión clara de los cambios realizados en la base de datos GTFS.

Esta solución no solo automatiza la gestión de datos de transporte público, sino que también permite una fácil adaptación a diferentes contextos y necesidades. Con la flexibilidad y las herramientas adecuadas, nuestro proyecto ofrece una solución sólida para la gestión de archivos GTFS en Concepción y más allá.

4.2. Justificaciones

Primer párrafo.



Capítulo 5

Resultados

5.1. Benchmark

Primer párrafo.





Capítulo 6

Resultados-Trabajos Futuros

6.1. Resultados

Primer párrafo.

6.2. Trabajos Futuros

Primer párrafo.





Anexo A

Primer Anexo

Aqui van los anexos, puede agregar todos los que necesite y separarlos por chapters.





Anexo B

Segundo Anexo

También puede hacer archivos separados para hacerlo

