



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

AYATORI: IMPLEMENTACIÓN EFICIENTE DE CONNECTION SCAN EN GTFS  
CON CASO DE ESTUDIO DE MOVILIDAD EN SANTIAGO

INFORME FINAL DE CC6908 PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL EN COMPUTACIÓN

FELIPE IGNACIO LEAL CERRO

PROFESOR GUÍA:  
EDUARDO GRAELLS-GARRIDO

SANTIAGO DE CHILE

2022

# Resumen

El presente informe almacena el trabajo desarrollado en la creación de una Propuesta de Memoria para optar al Título Profesional de Ingeniero Civil en Computación. A lo largo de este, se detalla la planificación en la creación de un algoritmo computacional (programado en Python) que, entregando coordenadas de inicio y término dentro de una ciudad estipulada, sea capaz de utilizar la información disponible para entregar especificaciones de rutas que unan ambos puntos geográficos. Para evaluar la utilidad y correcta implementación de esta solución, se realizará un caso de estudio para Santiago de Chile. Toda la información detallada en este informe servirá como base para el informe final del Trabajo de Título. Finalmente, se describe el trabajo realizado por adelantado para ese fin durante este semestre.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Estado del Arte</b>	<b>3</b>
2.1. Fuentes de información . . . . .	3
2.1.1. Encuesta Origen Destino . . . . .	3
2.1.2. Trazas digitales de telefonía móvil . . . . .	4
2.1.3. OpenStreetMap . . . . .	5
2.1.4. GTFS . . . . .	6
2.2. Herramientas informáticas . . . . .	7
2.2.1. Lenguaje y herramientas disponibles . . . . .	7
2.2.2. Connection Scan . . . . .	9
<b>3. Objetivos</b>	<b>10</b>
<b>4. Propuesta de Trabajo</b>	<b>12</b>
4.1. Solución Propuesta . . . . .	12
4.2. Plan de Trabajo . . . . .	13
<b>5. Trabajo Adelantado</b>	<b>14</b>
5.1. Obtención de información . . . . .	14
5.1.1. GTFS . . . . .	14
5.1.2. Encuesta Origen Destino . . . . .	16
5.1.3. OpenStreetMap . . . . .	17

5.1.4. Otras Fuentes de Información . . . . .	17
5.2. Implementaciones de CSA . . . . .	18
5.2.1. Comentarios Finales . . . . .	19
<b>Bibliografía</b>	<b>20</b>

# Índice de Ilustraciones

2.1. Resultados Generales de la Encuesta Origen Destino para Santiago de Chile (2012). . . . .	4
2.2. Mapa de Santiago en OpenStreetMap.cl. . . . .	5
2.3. Visualización de un archivo OSM exportado desde la web de OpenStreetMap. . . . .	5
2.4. Visualización de archivos del feed GTFS para Santiago. . . . .	6
2.5. Diagrama de uso de datos en formato GTFS. . . . .	7
2.6. Mapa coroplético de la distancia promedio al trabajo de los habitantes de la Región Metropolitana de Santiago. . . . .	8
4.1. Cronograma del plan de trabajo de CC6909. . . . .	13
5.1. Búsqueda en OpenStreetMap.org, vista de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. . . . .	17
5.2. Extracto de la implementación de CSA provista por el departamento de Ciencias de la Computación de Aalto University. . . . .	18

# Capítulo 1

## Introducción

En la actualidad, es normal que las grandes ciudades experimenten constantemente cambios que las hagan crecer. Este fenómeno, común a nivel mundial, está presente también en Chile. Estudiando la situación local, este proceso viene ligado a múltiples causas, algunas de estas siendo más globales (como el cambio climático), y otras más específicas, como el importante aumento de la migración interna y externa al país durante los últimos años, y la construcción de nueva infraestructura urbana. Si bien han existido ciertas condiciones que afecten el florecimiento de las ciudades (como la pandemia del COVID-19), la tendencia general de crecimiento se mantiene. Evidentemente, la planificación y buena gestión de las ciudades se ha visto afectada por el auge de estos fenómenos.

Se ha vuelto necesario hallar maneras novedosas para comprender y caracterizar mejor la vida de los habitantes de las grandes ciudades, como es el caso de la capital de nuestro país, Santiago. En este mismo contexto, una arista muy importante a considerar es la movilidad vial, la cual es un factor determinante de la calidad de vida de los habitantes de las ciudades. Estas son el hogar de una gran cantidad de personas, las cuales necesitan transportarse cada día para realizar sus jornadas regulares.

La gente puede tener a su disposición múltiples medios para movilizarse, tanto públicos como privados. Por ejemplo, se pueden mover a pie, en bicicleta, en auto, o utilizando el transporte público. Para poder caracterizar correctamente la movilidad urbana, sería útil verlo desde la perspectiva de un medio de transporte que esté disponible para toda la población, así que estudiar el uso del transporte público en Santiago resulta ser la mejor opción para este fin. La gente utiliza el MetroTren, el Metro de Santiago, y los buses de Red en múltiples combinaciones, generando una cantidad enorme de rutas diferentes para moverse por la ciudad.

Actualmente, es común encontrar registros que detallen la movilidad urbana en ciudades como Santiago, dentro y fuera del ámbito digital. Durante los últimos años, se han utilizado diversas trazas digitales para este fin, las que pueden provenir de diferentes dispositivos; por ejemplo, hablando de la telefonía móvil, son múltiples las aplicaciones de las que se puede obtener información útil, como Uber, WhatsApp, o Waze, las que utilizan el GPS del dispositivo. Sin embargo, en la actualidad, las trazas utilizadas no integran las especificaciones del transporte público disponible, por lo que se hace complejo realizar un buen análisis de

rutas. Sumado a esto, esta información se obtiene de fuentes cuyo fin original era distinto al estudio de la movilidad urbana, y al no estar enfocados en ella, estos no presentan un formato fácilmente utilizable para enlazarlos con la motivación descrita. Por esta razón, los datos requieren de ser procesados debidamente antes de ser utilizados.

En otra arista, existen registros de movilidad que, si bien fueron creados precisamente con este fin, poseen otros inconvenientes notables a la hora de utilizarlos. Por ejemplo, la Encuesta Origen Destino proporciona detalladamente una vista a los patrones de movilización de los habitantes de una ciudad determinada. Santiago ha sido foco de este instrumento en el pasado, pero la última vez que se realizó fue en el año 2012, hace una década atrás [9]. Debido a esto, la información inferida gracias a la encuesta probablemente no represente, de forma correcta, la realidad actual del transporte en Santiago.

Todo este contexto forma los lineamientos necesarios para crear una buena propuesta de trabajo de título, basada en el estudio de movilidad vial. La propuesta planteada busca implementar una manera que permita facilitar el entendimiento de la movilidad urbana en Santiago, para que pueda ser un ámbito a considerar en la toma de decisiones futuras que involucren al transporte. Para cumplir este objetivo, se planifica generar un algoritmo que permita integrar, a la data digital disponible, tanto la información del transporte público como las de otros medios de transporte (con información generada a día de hoy), y así estudiar el cómo las personas se movilizan a través de la ciudad.

A través de este estudio de rutas, se podrá generar una forma de predecir la ruta más probable que una persona llegue a utilizar al viajar por Santiago. Junto con esto, se pueden considerar diversos casos de estudio para analizar el efecto que tiene el modificar las condiciones del transporte disponible en la ciudad. Por ejemplo, se pueden estudiar los tiempos de viaje antes y después de la inauguración de un nuevo recorrido de Red, o de una nueva línea del Metro, con el fin de comprender mejor la importancia del constante avance en mejorar la infraestructura vial de las ciudades. Así, la solución resultante de este trabajo de título podrá ser utilizada por los organismos pertinentes relacionados al transporte (como el Ministerio de Transporte y Telecomunicaciones), para analizar aquellos puntos críticos de Santiago donde la gente más necesita nuevas opciones de rutas, ya sea porque la frecuencia de viajes es muy baja, o porque los tiempos estimados de viaje son considerablemente elevados, ayudando, de esta forma, a la planificación de mejoras viales en la ciudad.

# Capítulo 2

## Estado del Arte

### 2.1. Fuentes de información

Si bien existe aún espacio para innovar en nuevos análisis, la movilidad urbana en las grandes ciudades se lleva estudiando desde hace décadas. Santiago ha sido un foco importante de estos estudios, con los altos montos de inversión que se han dedicado al transporte público y a mejorar calles y carreteras durante los últimos años. Actualmente, para estudiar la movilidad, existen diferentes fuentes de información relevantes, como las que se discuten a continuación.

#### 2.1.1. Encuesta Origen Destino

La Encuesta Origen Destino es una encuesta que tiene por objetivo caracterizar patrones de viajes urbanos dentro de una ciudad determinada, a la vez de estudiar las características socioeconómicas de los viajeros. Este instrumento es encargado por el Estado de Chile, a través del Ministerio de Transporte y Telecomunicaciones y mediante la Secretaría de Planificación de Transporte (SECTRA), a organismos externos tales como universidades del país. En lo que respecta a Santiago, la última Encuesta Origen Destino fue encargada al Observatorio Social de la Universidad Alberto Hurtado, quienes la realizaron entre julio del 2012 y noviembre del 2013, considerando la información de 18.461.134 viajes diarios a través de Santiago [9]. Algunos resultados bastante destacables son que cerca del 30 % de los viajes consultados fueron realizados exclusivamente en el transporte público, y casi el 35 % de estos fueron exclusivamente a pie. El resumen general de los resultados se muestra a continuación en la figura 2.1.





Figura 2.1: Resultados Generales de la Encuesta Origen Destino para Santiago de Chile (2012).

### 2.1.2. Trazas digitales de telefonía móvil

En la actualidad, se genera un flujo considerablemente grande de información proveniente de los teléfonos móviles en Chile. Data Reportal, un portal de análisis de información relacionada con la interacción que tienen las personas con internet, realizó un reporte a inicios del 2021 [11], que arrojó que la cantidad de conexiones móviles en Chile a enero del 2021 equivalía al 132.1 % de la población del país, es decir, la cantidad de conexiones superaba a la cantidad de personas por casi un tercio de la población. Además, aproximadamente un 82 % de la población chilena tenía acceso a internet en ese año y mes.

Gracias a esto, es seguro asumir que gran parte de las personas portan un dispositivo móvil al transportarse por Santiago, como por ejemplo, un smartphone. Esta clase de teléfonos es capaz de generar trazas digitales provenientes de distintas aplicaciones, y de tener acceso a ellas, se pueden utilizar para caracterizar de mejor manera la movilidad urbana. Aplicaciones como Waze o Uber son utilizadas activamente con este mismo fin, haciendo uso del GPS del dispositivo en todo momento, pero además de estas, existen otras fuentes disponibles al cruzar la información con la geolocalización, como WhatsApp o Spotify. Con estos datos, se pueden sacar diversas conclusiones útiles para caracterizar la movilidad en Santiago.

Múltiples estudios de movilidad se han basado en información proveniente de estas fuentes. Por ejemplo, el que está actualmente siendo realizado por el profesor guía de este trabajo de título, Eduardo Graells-Garrido, del que se podrá utilizar un pipeline de análisis de movilidad en Santiago como caso de estudio [8]. Este estudio analiza la movilidad en Santiago considerando flujos de origen y destino (es decir, desde donde y hasta donde se mueven las personas), por lo que una buena forma de extender su alcance sería tener una intuición de lo que pasa en el transcurso de dicho viaje; particularmente, ser capaces de comprender las rutas utilizadas por los habitantes de la ciudad usando el transporte público.

### 2.1.3. OpenStreetMap

OpenStreetMap es una herramienta online que consiste en un mapa del mundo, creado por una comunidad de personas provenientes de distintos países, el cual es de uso libre bajo una licencia abierta. Para nuestro país, la Fundación OpenStreetMap Chile es la organización sin fines de lucro que se encarga de apoyar a la comunidad a nivel local. Gracias a este proyecto comunitario, es posible obtener información cartográfica de Santiago para utilizar en los estudios de movilidad urbana [2]. La información provista, exportable en un archivo de extensión propia similar a HTML (OSM), incluye la latitud y longitud de los lugares relevantes contenidos en el área especificada, los nombres de las calles y de lugares relevantes, como colegios; también incluye los paraderos de buses. En la figura 2.2, se observa el mapa del Gran Santiago, en el que se pueden notar las carreteras más destacadas de la ciudad, como la Circunvalación Américo Vespucio y la Autopista Central. Por otro lado, un ejemplo de exportación de los datos en archivo OSM se puede ver en la figura 2.3.

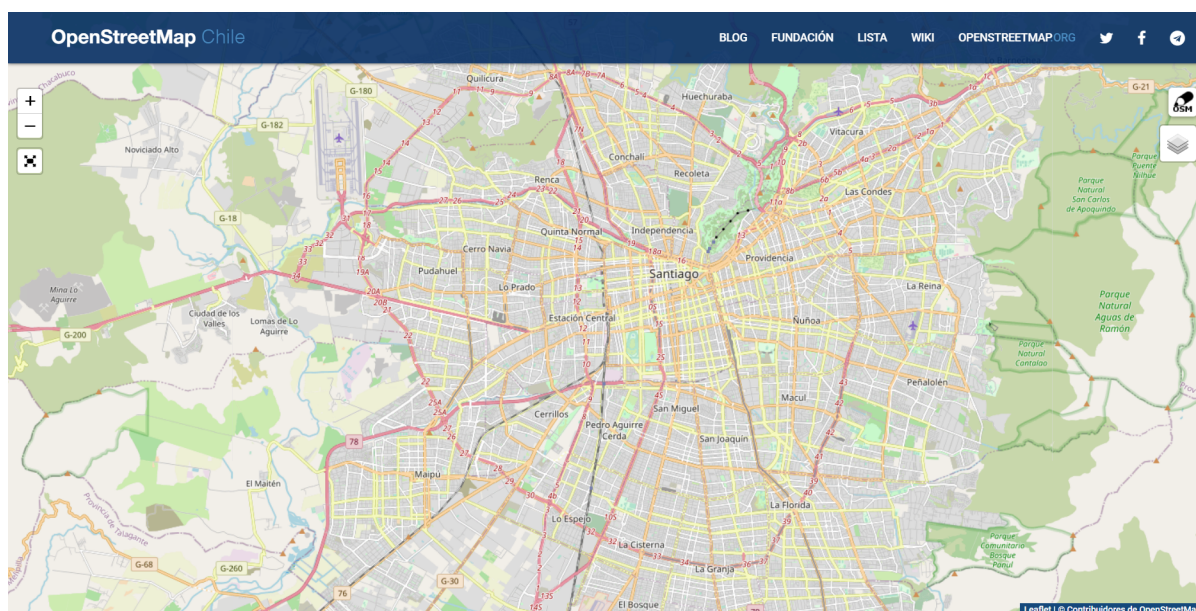


Figura 2.2: Mapa de Santiago en OpenStreetMap.cl.

```
<node id="1175388726" visible="true" version="2" changeset="127431613" timestamp="2022-10-12T18:46:19Z" user="dintrans_g" uid="237777" lat="
<tag k="attribution" v="Coordinación de Transportes de Santiago"/>
<tag k="bus" v="yes"/>
<tag k="highway" v="bus_stop"/>
<tag k="is_in:city" v="Santiago"/>
<tag k="name" v="Abate Molina / Gorbea"/>
<tag k="public transport" v="platform"/>
<tag k="ref" v="PA578"/>
<tag k="ref_name" v="Abate Molina / Gorbea"/>
<tag k="route_ref" v="H17 PM;H17"/>
<tag k="source" v="http://www.transantiago.cl"/>
<tag k="ts_calle" v="Gorbea"/>
<tag k="ts_codigo" v="L-20-7-OP-25"/>
<tag k="ts_desde" v="Avenida España"/>
<tag k="ts_hacia" v="Abate Molina"/>
<tag k="ts_orientacion" v="OP"/>
</node>
<node id="1234856346" visible="true" version="11" changeset="112695366" timestamp="2021-10-19T12:08:19Z" user="dintrans_g" uid="237777" lat="
<node id="1245530364" visible="true" version="3" changeset="119078311" timestamp="2022-03-29T16:26:47Z" user="dintrans_g" uid="237777" lat="
<tag k="addr:housenumber" v="2485"/>
<tag k="addr:street" v="Toesca"/>
</node>
<node id="1245530366" visible="true" version="2" changeset="36287198" timestamp="2015-12-31T17:40:50Z" user="Maximi89" uid="432825" lat="
<tag k="addr:housenumber" v="2655"/>
<tag k="addr:street" v="Toesca"/>
</node>
```

Figura 2.3: Visualización de un archivo OSM exportado desde la web de OpenStreetMap.

## 2.1.4. GTFS

Las Especificaciones Generales del Suministro de datos para el Transporte público, o en inglés, General Transit Feed Specification (GTFS), son un tipo de especificaciones ampliamente utilizado para definir y trabajar sobre datos de transporte público en las grandes ciudades. Este instrumento consiste en una serie de archivos de texto, recopilados en un archivo ZIP, de manera tal que cada archivo modela un aspecto específico de la información del transporte público, como paradas, rutas, viajes y horarios. En la figura 2.4, se muestra el cómo se ven los archivos en GTFS, mostrando, como ejemplo, la información de los paraderos disponibles.

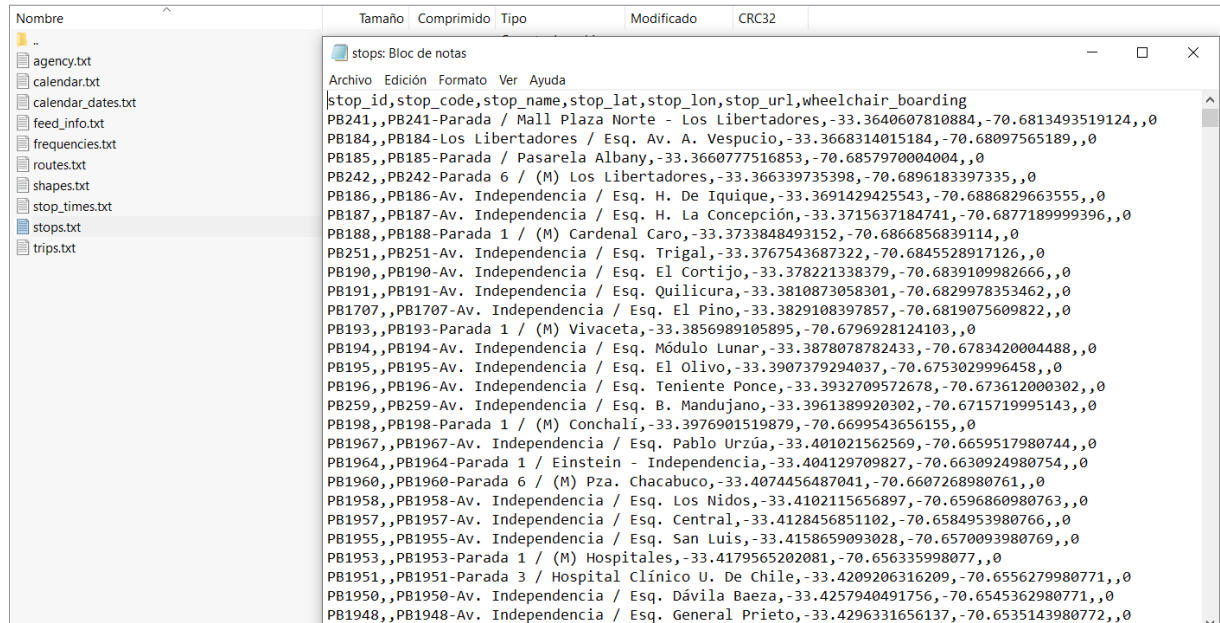


Figura 2.4: Visualización de archivos del feed GTFS para Santiago.

Siguiendo este formato, los operadores de transporte pueden almacenar y publicar la información pertinente a sus sistemas, para que esta sea utilizada por los programadores para crear aplicaciones que permitan entregarle esta información a los usuarios. El Directorio de Transporte Público Metropolitano (DTPM), organismo dependiente del Ministerio de Transportes y Telecomunicaciones, y cuya misión es mejorar la calidad del sistema de transporte público en Santiago, tiene disponible públicamente esta información, y la actualiza periódicamente [4]. El flujo de información en el que estos datos son utilizados se detalla en el diagrama mostrado en la figura 2.5.

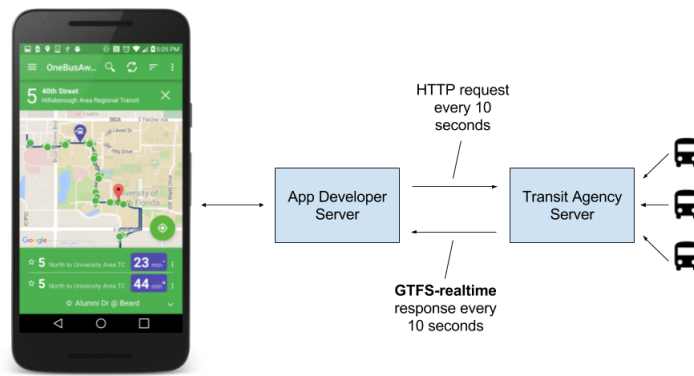


Figura 2.5: Diagrama de uso de datos en formato GTFS.

De esta manera, la información del transporte público puede ser utilizada para múltiples fines. Por ejemplo, personas naturales pueden acceder a ella para planificar sus viajes (complementando con otras opciones de movilización), y además, entidades gubernamentales pueden verificarla para estudiar el uso de medios de transporte. Por lo tanto, estos datos prueban ser una fuente indispensable para el trabajo futuro.

## 2.2. Herramientas informáticas

Además de definir las fuentes de datos que son posibles de utilizar en el estudio, es necesario identificar ciertas herramientas digitales que pueden ser utilizadas para desarrollar el análisis y experimentación. Evidentemente, existen múltiples alternativas, como las que se detallarán a continuación.

### 2.2.1. Lenguaje y herramientas disponibles

Como lenguaje de programación a usar, una excelente alternativa es Python. Este lenguaje de alto nivel es ampliamente utilizado por la legibilidad de su código y por ser concebido como un lenguaje de propósito general, siendo la base para desarrollar diversas aplicaciones de todo tipo. Ejemplos de estas son Instagram, Spotify, Netflix, entre otras. Además, se utilizarán librerías adicionales para poder trabajar la información, tales como Pyrosm, la que permite la lectura de archivos provenientes de OpenStreetMap [13].



Con respecto a otras herramientas disponibles, el profesor Eduardo Graells-Garrido ha creado un repositorio en GitHub llamado AVES: Análisis y Visualización, Educación y Soporte [7]. Este repositorio contiene datos y código desarrollado por el profesor en su trabajo, con un enfoque importante en la visualización de información. Dentro de esta misma arista, el repositorio contiene múltiples herramientas para visualizar datos geográficos, las que pueden ser de gran ayuda en el desarrollo de este trabajo de título. Por ejemplo, para visualizar este tipo de datos, se permite crear un mapa coroplético (o *choropleth map* en inglés) con las herramientas disponibles. En la figura 2.6, se muestra un mapa coroplético generado con las herramientas del repositorio, que representa, en promedio, la distancia que deben recorrer los habitantes de la Región Metropolitana para llegar a sus trabajos.

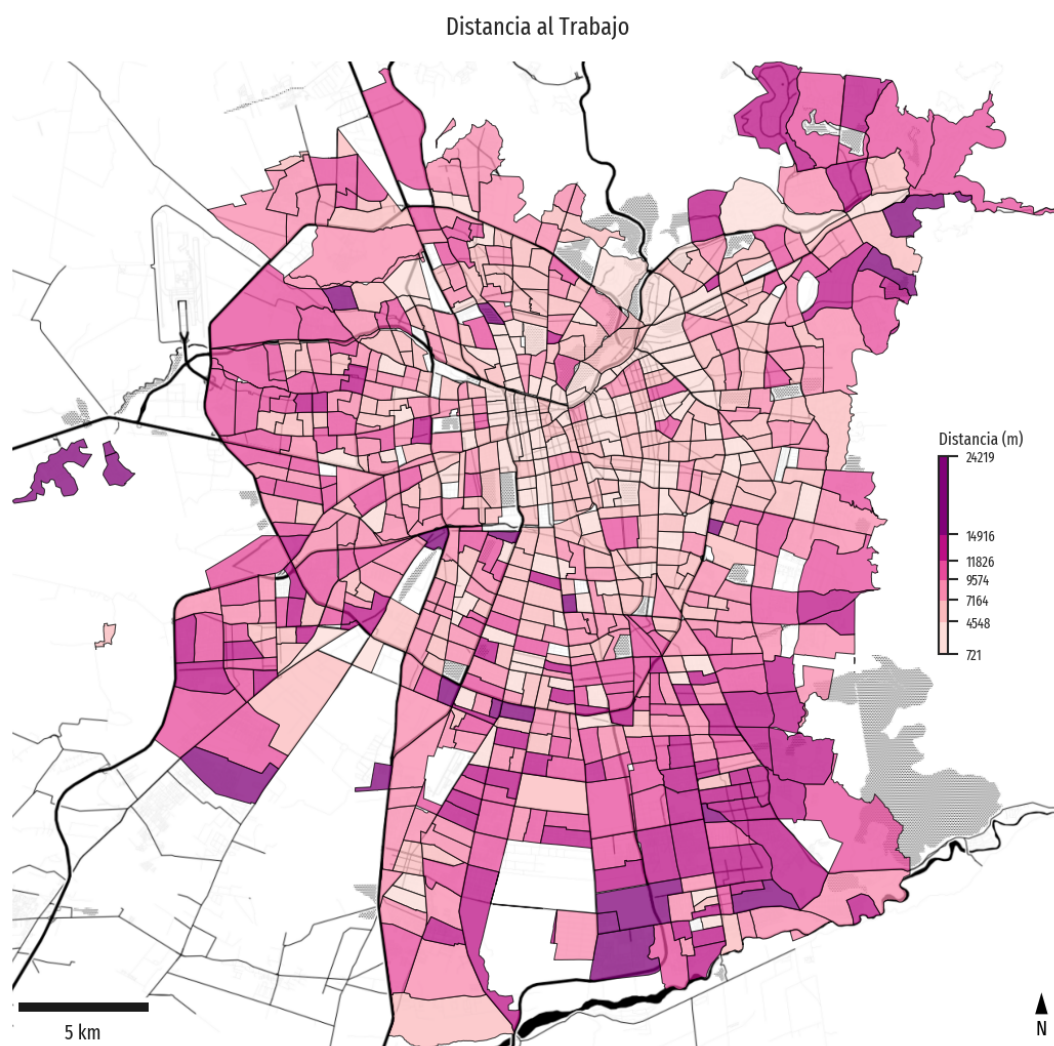


Figura 2.6: Mapa coroplético de la distancia promedio al trabajo de los habitantes de la Región Metropolitana de Santiago.

### 2.2.2. Connection Scan

Connection Scan Algorithm (CSA) es un algoritmo desarrollado para responder, de manera eficiente, consultas relacionadas a información espacio-temporal [5]. Este algoritmo es capaz de optimizar los tiempos de viaje entre dos puntos determinados de origen y destino, los cuales recibe como entrada, y, siendo alimentado por distintas fuentes de información de transporte, entrega como resultado una secuencia de vehículos (como trenes o buses) que un viajero debería tomar para llegar al destino desde el origen establecido.

Por ejemplo, para llegar desde la esquina de Almirante Latorre con la Avenida Libertador Bernardo O'Higgins (Alameda) hasta la Facultad de Ciencias Físicas y Matemáticas, el algoritmo podría entregar como resultado que se debe caminar una cantidad determinada de minutos a través de Almirante Latorre para llegar a pie, o el recorrido de Red que debe tomarse para llegar más rápido. Evidentemente, esto depende de asumir ciertos supuestos, como la velocidad promedio de caminata de una persona, o la disponibilidad ininterrumpida de dicho recorrido de autobús.

CSA puede tomar en consideración variables como los retrasos esperados de ciertos vehículos, para obtener la ruta más probable entre los dos puntos entregados. Dada su arquitectura, Connection Scan se muestra como una buena base para desarrollar una propuesta de solución.

# Capítulo 3

## Objetivos

Para poder proceder con el trabajo de título, es necesario estipular correctamente cuáles serán los objetivos a abarcar. A continuación, se detalla el objetivo general y los objetivos específicos de esta memoria. Además, se define el criterio de evaluación que se usará para medir el cumplimiento de estos objetivos.

### Objetivo General

El objetivo general de este trabajo de título consiste en implementar un método que permita modelar y caracterizar las rutas utilizadas por las personas al movilizarse a través de Santiago, considerando diversos medios de transporte. Se creará un algoritmo basado en Connection Scan Algorithm con este fin. Gracias a esto, se podrán averiguar las rutas más utilizadas, lo que permitirá predecir las rutas futuras de los habitantes, y así facilitar el estudio de movilidad vial.

### Objetivos Específicos

Para alcanzar el objetivo general del trabajo, se han definido los siguientes objetivos específicos:

1. Establecer una estructura de datos, generada en base a la información proveniente de las diversas fuentes mencionadas.
2. Desarrollar un algoritmo que permita vincular datos de GTFS con lo aprendido del resto de las fuentes.
3. Verificar que la implementación de dicho algoritmo sea eficiente.
4. Una vez el algoritmo sea correctamente implementado, integrarlo a un pipeline existente de análisis de datos de movilidad urbana [8].

5. Utilizar los resultados del análisis de datos para caracterizar, concretamente, el uso de modos de transporte en Santiago.

## **Evaluación**

Para asegurar el éxito del proyecto, el criterio que se usará para evaluar que el resultado del trabajo cumpla su objetivo, es:

- El usuario final deberá ser capaz de, haciendo uso del algoritmo propuesto como solución, identificar patrones de movilidad a través de Santiago de Chile.

El usuario final del proyecto será cualquier individuo u organización que requiera información sobre cómo movilizarse en Santiago. Por ejemplo, una persona que desee saber la mejor forma de movilizarse desde Padre Hurtado hasta Vitacura, o una entidad gubernamental (como la DTPM) que requiera estudiar las rutas más utilizadas por la gente para repartirles los recursos correspondientes.



# Capítulo 4

## Propuesta de Trabajo

### 4.1. Solución Propuesta

El algoritmo propuesto para solucionar el problema necesita ser capaz de mostrar cómo se conectan las personas con los distintos medios de transporte disponibles, mostrando las distintas posibilidades que tienen para llegar a su destino, y de esta forma, ser capaz de entregar la ruta más probable. Por esta razón, la mejor opción resulta ser basar dicho algoritmo en alguna herramienta que sea capaz de mostrar estas conexiones, tal como Connection Scan [5]. De esta forma, se puede desarrollar un algoritmo que determine la probabilidad de que una persona elija uno o más medios de transporte para ir de un punto A a un punto B, dentro de Santiago.

Para obtener la información, los datos originales que están disponibles provienen de las trazas digitales de telefonía celular [8]. Además de esto, se utilizará información de las diversas fuentes presentadas en la sección 2.1. Toda esta información será considerada por el algoritmo para predecir, el cual estará implementado en Python.

Posterior a la implementación, se realizará un caso de estudio para analizar la movilidad en Santiago. La finalidad es probar la efectividad y eficiencia de la solución desarrollada, y utilizar sus resultados para comparar la situación en la ciudad bajo diferentes condiciones. El caso de estudio definido consiste en estudiar las rutas disponibles en Santiago, antes y después de la implementación de la Línea 3 del Metro (la última en inaugurarse al momento del desarrollo de este trabajo de título).

## 4.2. Plan de Trabajo

1. Obtener la información de transportes en Santiago, proveniente de GTFS, de encuestas y estudios como la Encuesta Origen Destino, de las trazas de telefonía móvil, y de OpenStreetMap.
2. Estudiar implementaciones de Connection Scan Algorithm (CSA) en Python.
3. Desarrollar una estructura de datos que permita la operación de un algoritmo basado en Connection Scan.
4. Implementar el algoritmo propuesto como solución.
5. Estudiar el pipeline de análisis de datos y los resultados alcanzables en su estado actual.
6. Integrar el algoritmo al pipeline y verificar diferencias en los resultados alcanzables.
7. Utilizar los resultados para inferir situaciones que permitan caracterizar la movilidad urbana en Santiago.
8. Realizar el Caso de Estudio definido.

En la siguiente imagen, se especifica la planificación del trabajo a lo largo de las 15 semanas que componen el semestre próximo, donde se desarrollará el ramo CC6909-Trabajo de Título. El cronograma incluye todos los pasos descritos arriba, más la tarea de escribir la memoria en sí.

	Cronograma CC6909														
	Semanas														
Pasos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Paso 1															
Paso 2															
Paso 3															
Paso 4															
Paso 5															
Paso 6															
Paso 7															
Paso 8															
Escribir memoria															

Figura 4.1: Cronograma del plan de trabajo de CC6909.

# Capítulo 5

## Trabajo Adelantado

Para adelantar trabajo previo al siguiente semestre, donde se desarrollará este trabajo de título, se decidió enfrentar introductoriamente los dos primeros pasos del Plan de Trabajo definido en la sección anterior. Esto significa comenzar con el proceso de obtención de información a utilizar para alimentar el algoritmo propuesto como solución (junto a su análisis), y además, estudiar implementaciones de CSA para inspirar la construcción de dicho algoritmo. Todo lo relacionado se irá guardando en un repositorio almacenado en GitHub [10].

### 5.1. Obtención de información

#### 5.1.1. GTFS

Ingresando al sitio web del Directorio de Transporte Público Metropolitano, se puede obtener la información completa del GTFS, en su versión más actualizada [4]; al momento de la entrega de este informe, la última versión fue lanzada el 10 de diciembre de 2022.

La información está contenida en diferentes archivos de texto, con sus valores separados por comas (similar a un CSV). Cada archivo concentra un área específica de los datos:

- **Agency:** entrega la información de las diferentes agencias de transporte que alimentan el GTFS. En este caso, se encuentra la Red Metropolitana de Movilidad (que engloba a todos los buses Red, antiguamente Transantiago), el Metro de Santiago, y EFE Trenes de Chile.
- **Calendar Dates:** especifica fechas especiales que alteran el funcionamiento habitual de los recorridos que varían por día. Para la última versión, este archivo no contiene información (hasta la versión anterior, se incluían las modificaciones de recorridos de buses Red para el feriado del 8 de diciembre).
- **Calendar:** especifica los diferentes recorridos que varían por día, con su tiempo de validez. Acá se especifican los recorridos de Red para los días laborales (lunes a viernes), para los sábados, y para los domingos.

- **Feed Info:** información de la entidad que publica el GTFS.
- **Frequencies:** listado que, para todos los viajes de los recorridos disponibles, incluye sus tiempos de inicio y de término, y el *headway* o tiempo de espera estimado entre vehículos.
- **Routes:** contiene el identificador de cada ruta existente, su agencia, ubicación de origen y destino.
- **Shapes:** lista las diferentes 'formas' de los viajes de cada recorrido. Esto incluye el identificador de cada viaje (el recorrido y si acaso es de ida o retorno), y las latitudes y longitudes para cada secuencia posible.
- **Stop Times:** incluye las horas estimadas de llegada para que cada recorrido incluido en el GTFS llegue a cada parada incluida en su trayecto.
- **Stops:** contiene los identificadores, nombres, latitud y longitud de cada parada de transporte.
- **Trips:** contiene todos los viajes diferentes que realiza cada recorrido, señalando el nombre del recorrido, sus días de funcionamiento, si es de ida o retorno, y su dirección de destino.

Gracias a esta información, es posible acceder detalladamente a la información necesaria para considerar el transporte público en la solución propuesta. Con esto, las rutas calculadas podrán considerar viajes en microbuses y en las diferentes líneas del Metro de Santiago, pero además, también viajes realizados utilizando el MetroTren, tanto en su trayecto hacia Nos (sector de San Bernardo) como hacia Rancagua. Así, abarcando tres medios de transporte distintos que están disponibles en Santiago.

Estudiando los módulos detenidamente, se concluye que, para poder ser más específicos con las rutas a calcular, se debiera tener en consideración la fecha y hora para las cuales se hará la consulta. Si bien existen algunos recorridos disponibles transversalmente durante la semana, según el horario pueden estar más o menos disponibles; incluso, hay trayectos que solo funcionan en un horario específico ciertos días de la semana. Por ejemplo, todos los recorridos **N** de Red son recorridos nocturnos (como I08N, la versión nocturna del recorrido I08), los cuales operan únicamente entre las 00:00 y las 05:00 horas. A consecuencia de esto, lo ideal sería que el algoritmo desarrollado como solución fuese capaz de considerar información temporal, además de la espacial.

### 5.1.2. Encuesta Origen Destino

La información arrojada por la Encuesta Origen Destino, subida en la página web del Programa de Vialidad y Transporte Urbano [9], está disponible en diferentes formatos. Existen múltiples informes en PDF que grafican la información obtenida, pero además, la información también está disponible en un archivo ACCDB (de Microsoft Access DataBase). Este contiene todas las Tablas donde se almacenan las respuestas posibles de cada sección de la encuesta, y además, una sección de Consultas, que trabaja dichas tablas realizando cruces y/o distribuciones para poder concluir información.

Dado que la Encuesta Origen Destino tiene por objetivo caracterizar socioeconómicamente a la población, existen múltiples consultas que fueron realizadas a la base de datos con ese fin. Por ejemplo, el Nivel Educacional por Sector, la Distribución sexo/edad, y el Tamaño de los Hogares por Sector. Sin embargo, existen un par de consultas realizadas que pueden ser de utilidad para el proyecto,:

- **Personas por comuna:** cantidad de personas entrevistadas por cada comuna de la RM.
- **Vehículos por comuna:** cantidad de vehículos referentes a las personas encuestadas, por cada comuna de la RM.
- **Viajes Laboral Normal por Motorización:** viajes distribuidos por las características de motorización del vehículo en un día laboral corriente.

Estas consultas servirán principalmente para hacer los estudios de movilidad, una vez el algoritmo esté creado. Existen algunas tablas de datos que también pueden ser de ayuda para el proyecto, tales como **Modo** (que representa el modo de transporte utilizado por la persona), o **Tiempo medio** (promedio de la duración de un viaje normal declarado por la persona encuestada). Además de esto, se podrá utilizar información previamente inferida por la SECTRA (Secretaría de Transportes), que se encuentra detallada en los informes de resultados.

### 5.1.3. OpenStreetMap

Como fue mencionado anteriormente, la web de OpenStreetMap Chile permite acceder a una visualización del mapamundi, pudiendo ingresar en el link de la página especificaciones de latitud y longitud para buscar un punto específico en el mapa [2]. Por otro lado, la web de OpenStreetMap Global permite hacer lo mismo, pero además cuenta con una barra de búsqueda que gestiona consultas específicas de lugares, pudiendo buscar, por ejemplo, “Beauchef 850, Santiago”, mostrando la Facultad de Ciencias Físicas y Matemáticas [6].



Figura 5.1: Búsqueda en OpenStreetMap.org, vista de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

De esta forma, se puede cruzar esta información con la información entregada por el GTFS para ver dónde se ubican las diferentes paradas de los recorridos del transporte público. Otro uso práctico de estos datos es el análisis visual de la ubicación de las diferentes vías principales que tiene la Región Metropolitana, o la cercanía a alguna estación del Metro de Santiago desde una ubicación dada. Como se vio en la sección 2.1.3, esta información

### 5.1.4. Otras Fuentes de Información

Además de las fuentes mencionadas en esta sección, existen otras fuentes destacables que considerar para el proyecto. Específicamente, estudiar trazas de telefonía móvil puede ser de enorme utilidad, de tener acceso a ellas de alguna forma. Por otro lado, hará falta considerar también la información de las ciclovías disponibles en la RM, si se desea considerar a las bicicletas como medios de transporte válidos dentro del alcance del algoritmo; una idea es considerar registros en línea, como el que se encuentra en la web de Bicinet Chile [1]. El estudio y análisis de estas y otras fuentes adicionales se dejará como trabajo para el siguiente semestre.

## 5.2. Implementaciones de CSA

Para poder basar el desarrollo del algoritmo propuesto como solución, se procede a investigar diferentes implementaciones de Connection Scan Algorithm.

En el sitio web de Papers with Code, un portal que recopila códigos desarrollados sobre la idea central de diferentes papers, existen varias implementaciones realizadas de Connection Scan [14], basadas en el paper que lo define [5]. Investigando los repositorios enlazados, el que más destaca es ULTRA: UnLimited TRAnsfers for Multimodal Route Planning [12], un framework desarrollado en C++ para realizar planeamiento de viajes viales considerando diferentes medios de transporte. Este framework considera CSA, junto con otros algoritmos, para entregar respuestas. Si bien la implementación presentada está en un lenguaje diferente a Python, sirve como inspiración para comprender de forma práctica cómo funciona el algoritmo.

Por otro lado, Richard Darst, de Aalto University (Finlandia), ha desarrollado un paquete para Python 3.8+ para analizar datos en formato GTFS, llamado gtfspsy [3]. Este módulo permite, entre otras cosas, importar feeds a una base de datos SQLite para hacer consultas de forma eficiente, y considerar distancias reales a pie entre paradas de transporte público utilizando información de OpenStreetMap, además de estar planteado como una adaptación de CSA (aunque el algoritmo no se encuentre expuesto). Por estas razones, se destaca como la mejor opción para basar la solución.

```
class ConnectionScan(AbstractRoutingAlgorithm):
    """
    A simple implementation of the Connection Scan Algorithm (CSA) solving the first arrival problem
    for public transport networks.

    http://iti11www.iti.uni-karlsruhe.de/extra/publications/dpsw-isftr-13.pdf
    """

    def __init__(self, transit_events, seed_stop, start_time,
                 end_time, transfer_margin, walk_network, walk_speed):
        """
        Parameters
        -----
        transit_events: list[Connection]
        seed_stop: int
            index of the seed node
        start_time: int
            start time in unixtime seconds
        end_time: int
            end time in unixtime seconds (no new connections will be scanned after this time)
        transfer_margin: int
            required extra margin required for transfers in seconds
        walk_speed: float
            walking speed between stops in meters / second
        walk_network: networkx.Graph
            each edge should have the walking distance as a data attribute ("d_walk") expressed in meters
        """
        AbstractRoutingAlgorithm.__init__(self)
        self.seed = seed_stop
        self.connections = transit_events
        self.start_time = start_time
        self.end_time = end_time
        self.transfer_margin = transfer_margin
        self.walk_network = walk_network
        self.walk_speed = walk_speed
```

Figura 5.2: Extracto de la implementación de CSA provista por el departamento de Ciencias de la Computación de Aalto University.

Como se puede observar en la imagen, el algoritmo considera ciertos parámetros necesarios

al entregar el cálculo de la mejor ruta. Entre ellos se encuentran los tiempos de inicio y término, la cantidad de conexiones necesarias, y la velocidad de caminata esperada.

Con estas inspiraciones como base, se espera poder desarrollar correctamente una implementación de CSA en Python para desarrollar el algoritmo propuesto como solución de este proyecto de título.

### **5.2.1. Comentarios Finales**

El trabajo adelantado realizado durante este semestre se muestra como un avance fructífero para lo que será el desarrollo de la solución final durante Otoño 2023. Los resultados obtenidos de trabajar los dos primeros puntos del Plan de Trabajo son útiles para ir formando una idea, tanto de la Estructura de Datos que se deberá crear para considerar toda la información que alimentará al algoritmo (Punto 3), como para su propia implementación (Punto 4). Se espera que el desarrollo del proyecto durante el semestre de CC6909-Trabajo de Título proceda tal como fue planificado en la Carta Gantt presentada en la sección 4.2 de este informe.



# Bibliografía

- [1] Bicineta Chile. Mapa de Ciclovías de la Región Metropolitana. Disponible en <https://www.bicineta.cl/cicloviias>. Revisado el 2022/12/13.
- [2] Fundación OpenStreetMap Chile. Mapa de OpenStreetMap Chile. Información disponible en <https://www.openstreetmap.cl>. Revisado el 2022/12/13.
- [3] Richard Darst. gtfspy. Repositorio disponible en <https://github.com/CxAalto/gtfspy>. Revisado el 2022/12/13.
- [4] Directorio de Transporte Público Metropolitano. GTFS Vigente. Disponible en <https://www.dtpm.cl/index.php/gtfs-vigente> (2022/12/10). Revisado el 2022/12/13.
- [5] Julian Dibbelt, Thomas Pajor, Ben Strasser, and Dorothea Wagner. Connection Scan Algorithm. *ACM Journal of Experimental Algorithmics*, 23(1.7):1–56, 2018.
- [6] OpenStreetMap (Global). Mapa de OpenStreetMap. Información disponible en <https://www.openstreetmap.org>. Revisado el 2022/12/13.
- [7] Eduardo Graells-Garrido. Aves: Análisis y Visualización, Educación y Soporte. Repositorio disponible en <https://github.com/zorzalerrante/aves> (2022/09/23). Revisado el 2022/12/13.
- [8] Eduardo Graells-Garrido, Daniela Opitz, and Francisco Rowe. A Generalisable Data Fusion Framework to Infer Mode of Transport Using Mobile Phone Data. Paper presentado para su publicación, 2022.
- [9] Universidad Alberto Hurtado. Actualización y recolección de información del sistema de transporte urbano, IX Etapa: Encuesta Origen Destino Santiago 2012. Encuesta origen destino de viajes 2012. Disponible en <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mfn=3253> (2012). Revisado el 2022/12/13.
- [10] Felipe Leal. CC6909-Ayatori (Repositorio del Trabajo de Título). Repositorio disponible en <https://github.com/Lysorek/CC6909-Ayatori> (2022/12/04). Revisado el 2022/12/13.
- [11] Data Reportal. Digital 2021 Report for Chile. Disponible en <https://datareportal.com/reports/digital-2021-chile> (2021/02/11). Revisado el 2022/12/13.
- [12] Jonas Sauer. ULTRA: UnLimited TRAnsfers for Multimodal Route Planning. Repositorio disponible en <https://github.com/kit-algo/ULTRA>. Revisado el 2022/12/13.

- [13] Henrikki Tenkanen. Pyrosm: Read OpenStreetMap data from Protobuf files into GeoDataFrame with Python, faster. Repositorio disponible en <https://github.com/HTenkanen/pyrosm>. Revisado el 2022/12/13.
- [14] Papers with Code. Connection Scan Algorithm implementations. Disponible en <https://cs.paperswithcode.com/paper/connection-scan-algorithm>. Revisado el 2022/12/13.