



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

PROPUESTA DE MEMORIA 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 usará información de Santiago de Chile como caso de estudio. Toda la información detallada en este informe servirá como base para el informe final de la Memoria de Título. Finalmente, se describe el trabajo realizado por adelantado para ese fin.

# 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. Trazas digitales de telefonía móvil . . . . .	3
2.1.2. Encuesta Origen Destino . . . . .	4
2.1.3. OpenStreetMap . . . . .	5
2.1.4. GTFS . . . . .	5
2.2. Herramientas informáticas . . . . .	7
2.2.1. Lenguaje y herramientas disponibles . . . . .	7
2.2.2. Connection Scan . . . . .	7
<b>3. Objetivos</b>	<b>8</b>
<b>4. Propuesta de Trabajo</b>	<b>10</b>
4.1. Solución Propuesta . . . . .	10
4.2. Plan de Trabajo . . . . .	11
<b>5. Trabajo Adelantado</b>	<b>12</b>
<b>Bibliografía</b>	<b>13</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 archivos del feed GTFS para Santiago. El archivo 'stops.txt' entrega información de los paraderos disponibles. . . . .	6
2.4. Diagrama de uso de datos en formato GTFS. . . . .	6
4.1. Cronograma del plan de trabajo de CC6909. . . . .	11

# 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 los efectos de la pandemia del COVID-19), 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. Evidentemente, la planificación y buena gestión de las ciudades se ha visto afectada por el auge de estos fenómenos, y se ha vuelto necesario hallar maneras novedosas para comprender y caracterizar mejor la vida en las urbes, tales como Santiago.

En este mismo contexto, una arista muy importante a considerar es la movilidad y el transporte a través de la ciudad, dado que lugares como Santiago son el hogar de una gran cantidad de personas, las cuales necesitan transportarse cada día para realizar sus jornadas regulares. Según el sector donde se encuentren y la disponibilidad personal, la gente tiene a su disposición múltiples medios para movilizarse; por ejemplo, a pie, en bicicleta, en auto, o utilizando el transporte público. En específico, estudiar el uso del transporte público en Santiago resulta ser un experimento bastante útil para caracterizar la movilidad urbana, pues permite analizar las rutas que suelen usar las personas para desplazarse, en base a un medio de transporte disponible para toda la población.

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 la especificación del transporte público disponible en la ciudad, 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.

La propuesta planteada busca implementar una manera que permita, a través del estudio de las rutas utilizadas por las personas al viajar por Santiago, predecir las que sean más probables de utilizar. Para esto, se planifica generar un algoritmo que permita integrar, a la data digital disponible, tanto la información del transporte público como la de otros medios de transporte. Esto busca facilitar el entendimiento de la movilidad urbana en Santiago, para que pueda ser considerado en la toma de decisiones futuras que involucren al transporte.

# 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 muy 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:

#### 2.1.1. 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 [7], 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 [5]. 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.

### 2.1.2. 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 [6]. Algo bastante destacable es que más del 30 % de los viajes consultados fueron realizados exclusivamente en el transporte público. El resumen de los resultados está indicado en la siguiente imagen:



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



### 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 [1].

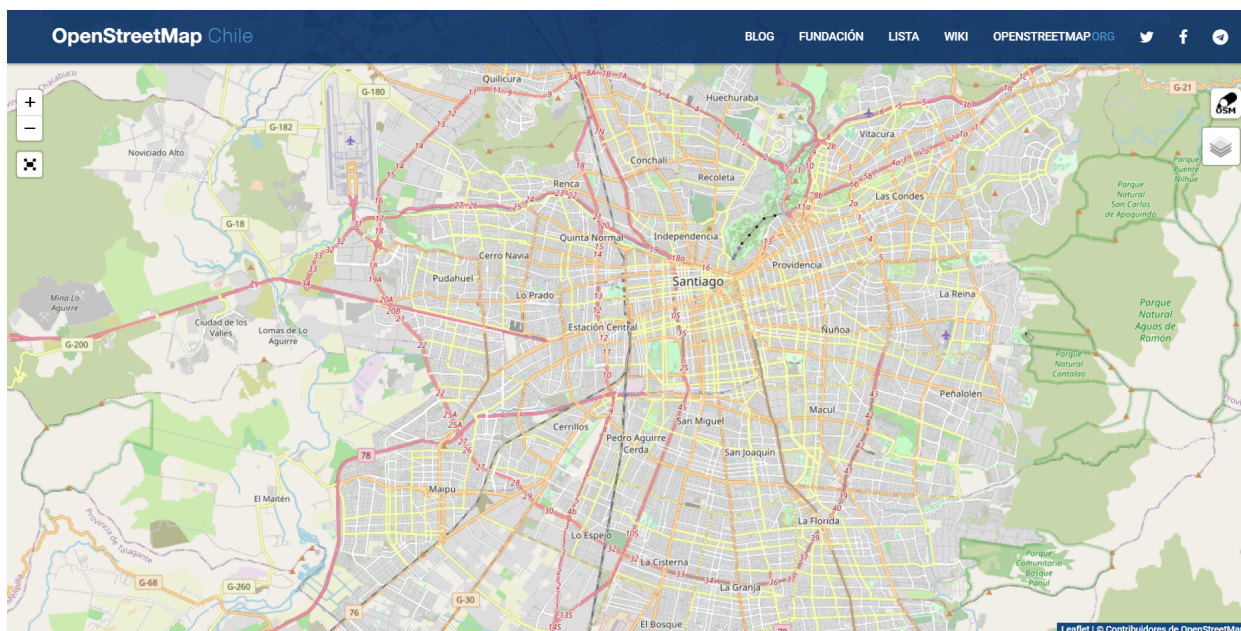


Figura 2.2: Mapa de Santiago en OpenStreetMap.cl.

### 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 siguiente figura, se muestra el cómo se ven los archivos en GTFS, especificando la información del transporte en Santiago:

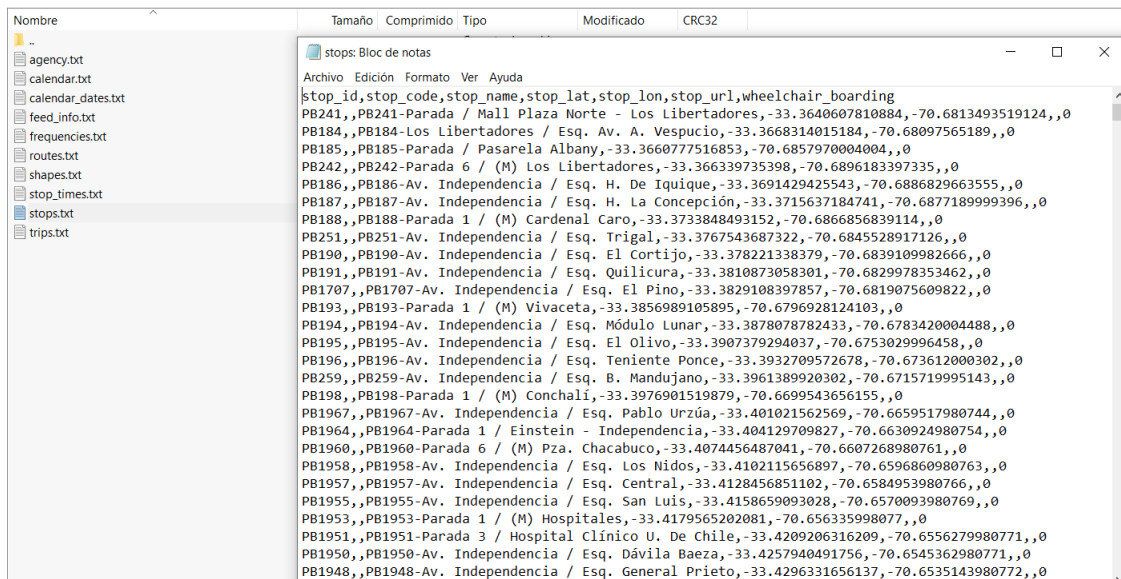


Figura 2.3: Visualización de archivos del feed GTFS para Santiago. El archivo 'stops.txt' entrega información de los paraderos disponibles.

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 [2]. El flujo de información en el que estos datos son utilizados se detalla en el siguiente diagrama:

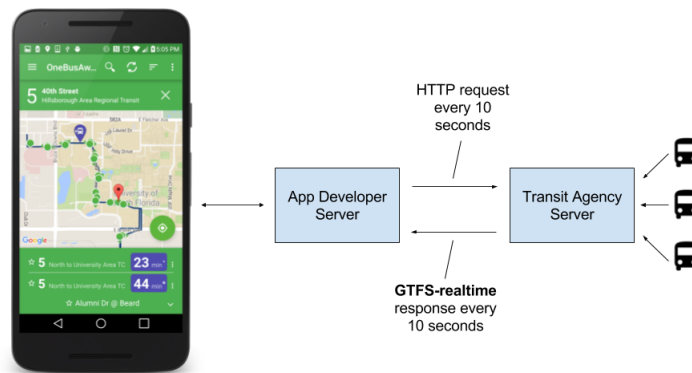


Figura 2.4: Diagrama de uso de datos en formato GTFS.

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**

Una vez definidas las fuentes de datos a utilizar en el estudio, es necesario estipular las herramientas digitales que serán utilizadas para desarrollarlo.

### **2.2.1. Lenguaje y herramientas disponibles**

En el desarrollo del trabajo, se utilizará el lenguaje de programación 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.

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 [4]. 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 serán de gran ayuda en el desarrollo de este trabajo de título.

### **2.2.2. Connection Scan**

Connection Scan Algorithm (CSA) es un algoritmo desarrollado para responder, de manera eficiente, consultas relacionadas a información espacio-temporal [3]. 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. 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 una manera que permita predecir las rutas utilizadas por las personas al movilizarse a través de Santiago. Con esto, se busca facilitar el estudio de movilidad de las personas a través de la ciudad.

### 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 trazas digitales actualmente utilizadas.
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 [5].
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 utilizar el algoritmo para determinar la probabilidad de que una persona elija uno o más recorridos de transporte público para ir desde un punto A hasta un punto B en Santiago.

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 [3]. 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 [5]. Además de esto, la DTPM tiene disponible, en su página web, la GTFS vigente hasta noviembre de 2022 [2], y la web de OpenStreetMap Chile tiene disponible un mapamundi de uso libre bajo una licencia abierta [1]. Toda esta información será considerada por el algoritmo para predecir, el cual estará implementado en Python.

## 4.2. Plan de Trabajo

1. Obtener la información de transportes en Santiago, proveniente de GTFS, de las trazas de telefonía móvil, y de OpenStreetMap.
2. Estudiar la implementación de Connection Scan 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.

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															
Escribir memoria															

Figura 4.1: Cronograma del plan de trabajo de CC6909.

# Capítulo 5

## Trabajo Adelantado

En esta sección, se describirá el trabajo práctico que se adelantará durante CC6908-Introducción al Trabajo de Título, con el fin de comenzar desde ya a desarrollar la solución que plantea esta propuesta de memoria. Se planea comenzar pronto con esta tarea, y así tenerla realizada para el informe final del curso.



# Bibliografía

- [1] Fundación openstreetmap chile. Información disponible en <https://www.openstreetmap.cl>. Revisado el 2022/09/27.
- [2] Directorio de Transporte Público Metropolitano. Gtfs vigente. Disponible en <https://www.dtpm.cl/index.php/gtfs-vigente> (2022/11/05).
- [3] Julian Dibbelt, Thomas Pajor, Ben Strasser, and Dorothea Wagner. Connection scan algorithm. *ACM Journal of Experimental Algorithmics*, 23(1.7):1–56, 2018.
- [4] Eduardo Graells-Garrido. Aves: Análisis y visualización, educación y soporte. Repositorio disponible en <https://github.com/zorzalerrante/aves> (2022/09/23).
- [5] 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.
- [6] 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).
- [7] Data Reportal. Digital 2021 report for chile. Disponible en <https://datareportal.com/reports/digital-2021-chile> (2021/02/11).