



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 TEMA DE MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN COMPUTACIÓN

Felipe Leal Cerro

PROFESOR GUÍA:
Eduardo Graells-Garrido

SANTIAGO DE CHILE
2022

1. Introducción

En la época en la que vivimos, es normal que las grandes ciudades experimenten constantemente cambios que las hagan crecer. Estudiando la situación local en Chile, este fenómeno 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, 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. 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.

2. Situación Actual

En la actualidad, la movilidad urbana es estudiada utilizando información proveniente de múltiples fuentes. De manera digital, una de ellas es el registro de uso de teléfonos móviles, el que presenta datos que poseen una resolución bastante fina en términos espaciales y temporales. Esta información se suele complementar, también, con la proveniente de otras fuentes oficiales, tales como los resultados del Censo o de encuestas solicitadas por el gobierno, como la Encuesta Origen Destino [4], la cual tiene por objetivo caracterizar patrones de viajes urbanos y características socioeconómicas de los viajeros. Además, ciertas herramientas que han sido generadas por la comunidad suelen ser de gran aporte en la obtención de información, tal como la iniciativa de OpenStreetMap [1].

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 [3]. 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. El problema principal de esta clase de estudios radica en que, como fue mencionado, no suelen integrar concretamente los medios de transporte utilizados en el viaje, y específicamente, no integran datos pertinentes al transporte público de Santiago. A causa de esto, se presenta un importante obstáculo para poder caracterizar correctamente el transporte a través de la ciudad. Sin embargo, la información de uso del transporte público sí está disponible desde otras fuentes.

Existe un tipo de especificaciones ampliamente utilizado para definir y trabajar sobre datos de transporte público, llamado Especificaciones Generales del Suministro de datos para el Transporte público (General Transit Feed Specification, o GTFS en inglés). Este 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:

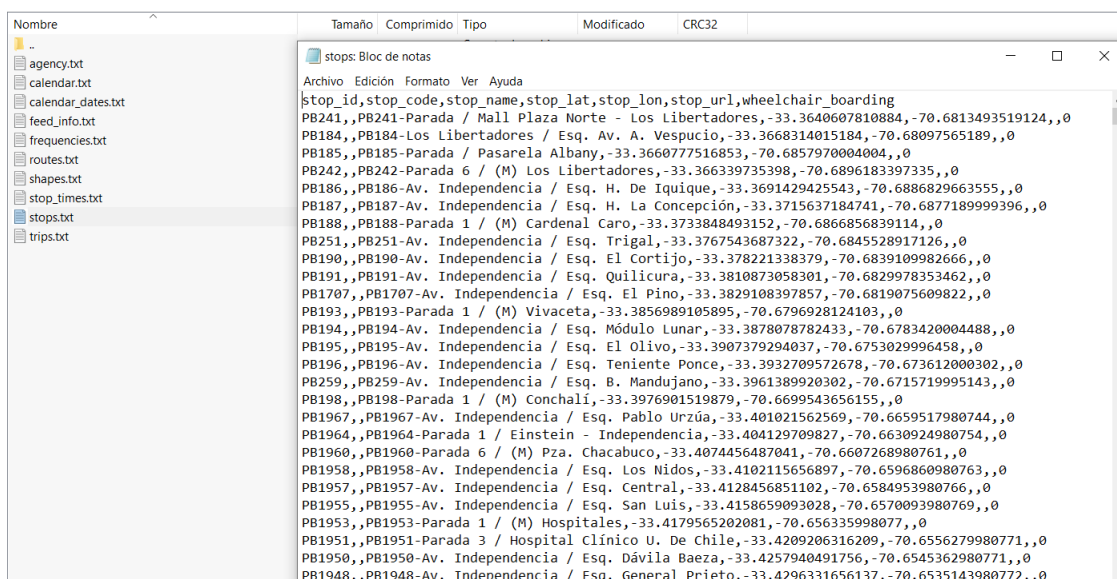


Figura 1: Visualización de archivos del feed GTFS para Santiago. El archivo 'stops.txt' entrega información de los paraderos disponibles.

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

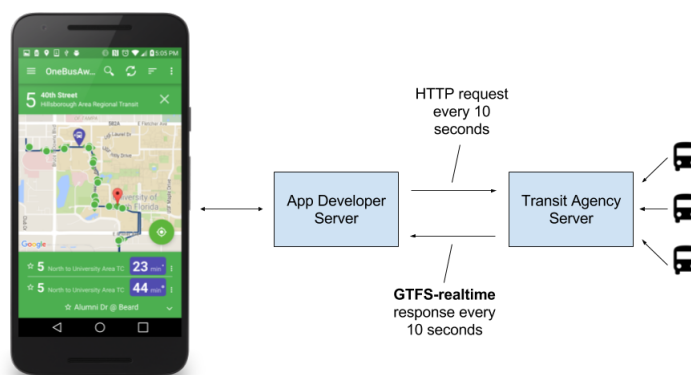


Figura 2: Diagrama de uso de datos en formato GTFS.

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. Precisamente, en esa posibilidad, se enmarcan los objetivos de este proyecto.

3. 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 [3].
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.

4. 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 [2], un algoritmo que toma como entrada una posición de partida y una posición de destino, y entrega como resultado una secuencia de vehículos (como trenes o buses) que un viajero debería tomar para llegar desde la posición de partida hasta la de destino. Así, se propone basar la solución en este algoritmo, tal 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 [3]. Además de esto, la DTPM tiene disponible, en su página web, la GTFS vigente hasta agosto de 2022 [5], 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.

5. Plan de Trabajo (Preliminar)

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.

Referencias

- [1] *Fundación OpenStreetMap Chile*. Información disponible en <https://www.openstreetmap.cl>. Revisado el 2022/09/27.
- [2] Dibbelt, Julian, Thomas Pajor, Ben Strasser y Dorothea Wagner: *Connection Scan Algorithm*. ACM Journal of Experimental Algorithmics, 23(1.7):1–56, 2018.
- [3] Graells-Garrido, Eduardo, Daniela Opitz y Francisco Rowe: *A Generalisable Data Fusion Framework to Infer Mode of Transport Using Mobile Phone Data*. Paper presentado para su publicación, 2022.
- [4] Hurtado, Universidad Alberto: *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?mf=3253> (2012).
- [5] Transporte Público Metropolitano, Directorio de: *GTFS Vigente*. Disponible en <https://www.dtpm.cl/index.php/gtfs-vigente> (2022/08/27).