

Temas de Memoria/Tesis

Otoño 2010

Sebastian Ríos

srios@dii.uchile.cl

Requisitos Postulantes

- ✦ Gusto por la Minería de Datos (algoritmos)
- ✦ Saber trabajar con Bases de datos (mysql, postgres, oracle, etc.) y con ganas de ensuciarse las manos (mucho ;)
- ✦ No tenerle miedo a programar
- ✦ Saber usar o aprender rápidamente a usar software de analysis de datos como SPSS, SAS, Rapidminer, etc.
- ✦ Ser proactivo e independiente
- ✦ Buscar realizar un trabajo de excelencia

Como postular

- ✦ Mandar un mail a srios@dii.uchile.cl en que digas lo siguiente:
 - ✦ Que buscas? memoria o tesis (que programa de magister)
 - ✦ Copia el nombre del tema o temas que te interesan
 - ✦ Explica por que quieres tomar ese tema
 - ✦ Explica que competencias crees tu que tienes para tomar este tema

1. DOCODE - Detector de Copia

- Application y Evaluación de Modelos de Lenguaje para recuperación de información en grandes bases de datos.
- Diseño e Integración de la Arquitectura DOCODE-Engine SNA con DOCODE-Engine y DOCODE-ASP (Ing. de Software, SOA, JAVA, Javascript, Applets)

Temas de Memoria/Tesis

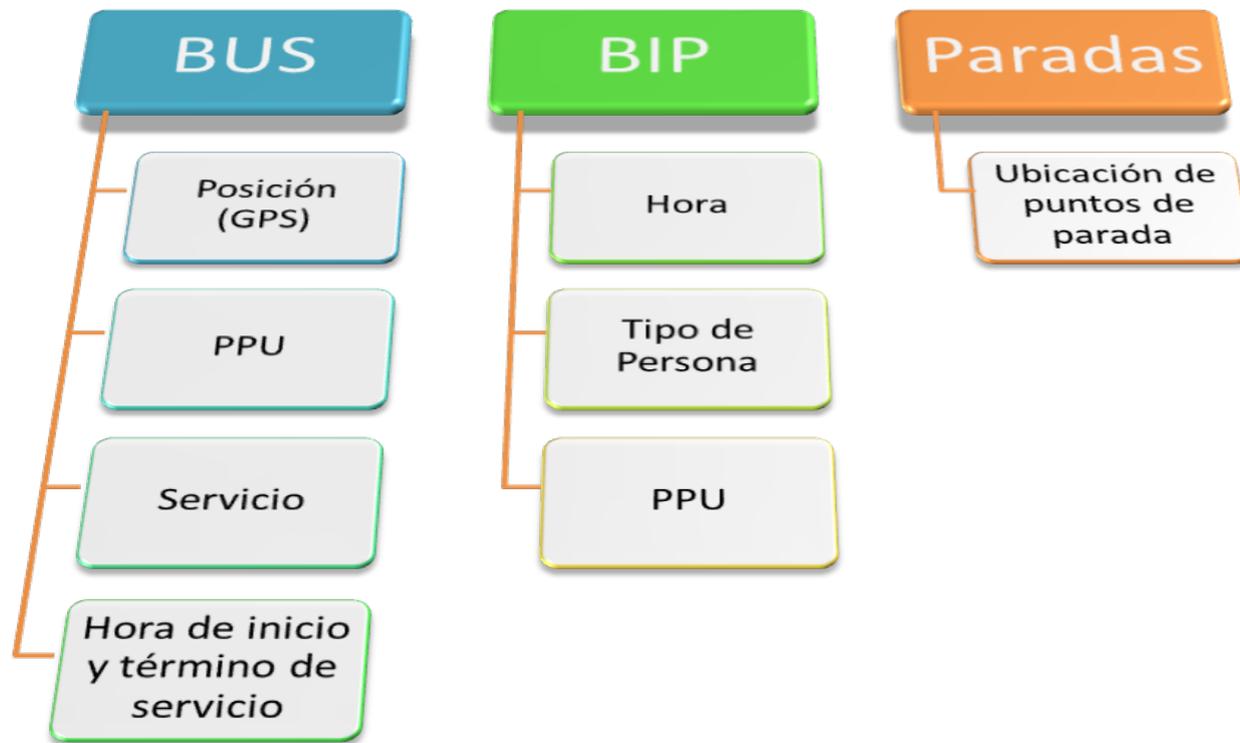
Colaboración DII - Ing. Civil - Transantiago

Otoño 2010

Sebastian Ríos

srios@dii.uchile.cl

Información



Resultados

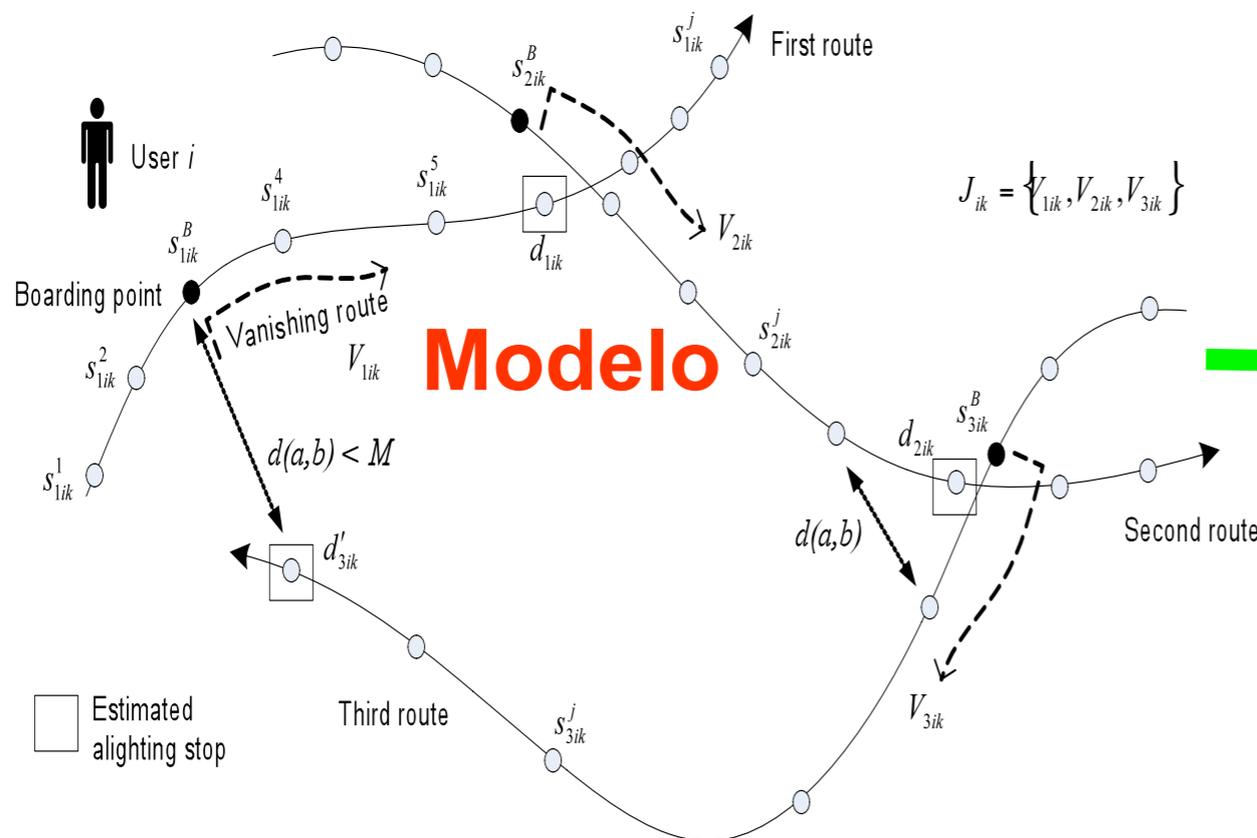
Estructura de flujos

Nivel de servicio

Detección de usuarios en riesgo

Performance de operadores

...



Minería de Datos en Transporte Público

- ✦ Base de datos de transantiago es gigantesca y hacer los experimentos necesarios en tiempo razonable es prácticamente imposible sin aplicar alguna técnicas para acelerar el procesamiento, se proponen dos temas al respecto:
 1. Diseño e Implementación de un heurísticas y algoritmos para acelerar el procesamiento en la base de datos de Transantiago
 2. Paralelización de algoritmos de minería de datos usando MapReduce a grafos de gran escala.

Minería de Datos en Transporte Público (2)

- ✦ Caracterización y Búsqueda de Transbordos Masivos con información de la base de datos de Transantiago
- ✦ Mejoramiento de la Calidad de los datos de Transantiago, mediante el uso de minería de datos
 - ✦ Por la naturaleza del transporte público muchas veces en la BD de Transantiago faltan algunos atributos, los usuarios no siempre viajan usando las mismas tarjetas Bip!, hay viajeros ocasionales, etc. Se pretende mejorar la calidad de los datos mediante el uso de técnicas de minería de datos para completar información faltante o eliminar información errónea. De este modo mejorar los datos para su análisis posterior

Minería de Datos en Transporte Público (3)

- Implementación de algoritmos de búsqueda de subestructuras frecuentes en grafos de gran escala usando minería de datos. (graphmining)

Minería de Datos en ENTEL

- Estos temas requieren pasar gran parte del tiempo en la empresa, pero con horarios flexibles
- Se requiere ir formal al trabajo
- Se paga \$150.000 al mes por 6 meses

- ✦ Predicción de Churn para el producto NGN en ENTEL (Frc. Barrientos)
- ✦ Minería de datos sobre base de datos de reclamos para mejorar atención al cliente empresa (text mining)
- ✦ Diseño e Implementación de un algoritmo para mejorar el producto NGN para empresas (canasta de productos)

Minería de Datos en ENTEL

- **IMPLEMENTACION/AUTOMATIZACION DE UN TABLERO DE CONTROL PARA EL AREA DE ASEGURAMIENTO DE INGRESOS DE UNA EMPRESA DE TELECOMUNICACIONES:**

LO DESEADO

- Sandra nos esta ayudando llevando a Teradata y al DWH las bases que usamos y procesamos manualmente cada mes (con su costo y tiempos respectivos).
- Se nos ha pedido automatizar los KPI, dejando de hacer informes manuales y liberando nuestro tiempo para gestionar los hallazgos que señalan los informes con los responsables.
- Eventualmente diseñar y llevar a una web intranet (ya tenemos el espacio) el Tablero de control automatizado, linkeado a algún repositorio desde donde mostrar kpi actualizados de acuerdo a los ciclos de actualización de las fuentes de información.

Minería de Datos Sobre Comunidades Virtuales

- Desarrollo e Implementación de algoritmos para encontrar expertos en una comunidad virtual (H. Alvarez)
- Desarrollo e Implementación de algoritmos para encontrar comunidades en una comunidad virtual (vacante)
- Diseño e implementación de algoritmos para analizar la evolución histórica del propósito de una comunidad virtual (vacante, tenemos trabajo previo)
- Analisis de Redes Sociales sobre una comunidad virtual (memoria, aplicar en nueva comunidad, vacante)

Minería de Datos Sobre Comunidades Virtuales (2)

- Diseño e Implementación de algoritmos para analizar la autoregulación de una comunidad virtual (memoria, vacante)
- Evaluación del impacto de técnicas para fomentar la participación de los miembros en una comunidad virtual. (memoria, vacante)

Con Cesar (HCI)

- Temas investigaciÃ“n conjunto cesar, Felipe, seba
- Cambios para el administrador!
- La ruta:
- 1. USO de mineria de datos para Caracterizar el trabajo de Los administradores de comunidades virtuales
- Queremos encontrar los patrones de administraciÃ“n. Cuanto rato modera,
- cuantos post modera, etc. Luego, que palabras hacen que modere, que
- factores hace que modere; i.e. Un usuario conflictivo o si el
- intercambio de post se realiza entre muy pocos usuarios y en poco
- tiempo.

Minería Semántica de Datos

- Uso de frases para mejorar Concept-based text mining process.