



Computer Science Sessions

*Escuela de Ingeniería en Computación.
Facultad de Ciencias Físicas y Matemáticas.
Universidad Central de Chile.*

A la comunidad universitaria

"Minería de Datos en Motores de Búsqueda"

Dr. Marcelo Mendoza Rocha

**Profesor Asociado de Ciencia de la
Computación, Departamento de Ciencias de la
Computación, Facultad de Ciencias.
Universidad de Valparaíso, Playa Ancha,
Valparaíso, Chile**

Lugar: Sala 317

Hora: 11:45 - 13:15

02 de Abril del 2008

"MINERÍA DE DATOS EN MOTORES DE BÚSQUEDA"

Resumen

La Web es un gran espacio de información donde muchos recursos como documentos, imágenes u otros contenidos multimediales pueden ser accesados. En este contexto, varias tecnologías de la información han sido desarrolladas para ayudar a los usuarios a satisfacer sus necesidades de búsqueda en la Web, y las más usadas de estas son los motores de búsqueda. Los motores de búsqueda permiten a los usuarios encontrar recursos formulando consultas y revisando una lista de respuestas.

Uno de los principales desafíos para la comunidad de la Web es diseñar motores de búsqueda que permitan a los usuarios encontrar recursos semánticamente conectados con sus consultas. El gran tamaño de la Web y la vaguedad de los términos más comúnmente usados en la formulación de consultas es un gran obstáculo para lograr este objetivo.

En este trabajo proponemos explorar las selecciones de los usuarios registradas en los *logs* de los motores de búsqueda para aprender cómo los usuarios buscan y también para diseñar algoritmos que permitan mejorar la precisión de las respuestas recomendadas a los usuarios. Comenzaremos explorando las propiedades de estos datos. Esta exploración nos permitirá determinar la naturaleza dispersa de estos datos. Además presentaremos modelos que nos ayudarán a entender cómo los usuarios buscan en los motores de búsqueda.

Luego, exploraremos las selecciones de los usuarios para encontrar asociaciones útiles entre consultas registradas en los *logs*. Concentraremos los esfuerzos en el diseño de técnicas que permitirán a los usuarios encontrar mejores consultas que la consulta original. Como una aplicación, diseñaremos métodos de reformulación de consultas que ayudarán a los usuarios a encontrar términos más útiles mejorando la representación de sus necesidades.

Usando términos de documentos construiremos representaciones vectoriales para consultas. Aplicando técnicas de *clustering* podremos determinar grupos de consultas similares. Usando estos grupos de consultas, introduciremos métodos para recomendación de consultas y documentos que nos permitirán mejorar la precisión de las recomendaciones.

Finalmente, diseñaremos técnicas de clasificación de consultas que nos permitirán encontrar conceptos semánticamente relacionados con la consulta original. Para lograr esto, clasificaremos las consultas de los usuarios en directorios Web. Como una aplicación, introduciremos métodos para la mantención automática de los directorios.