



## Ciclo de Coloquios 2017

### Charla técnica



El Departamento de Informática (DI) de la Universidad Técnica Federico Santa María tiene el agrado de invitar a la comunidad a un nuevo coloquio, a realizarse en la sala de reuniones del DI Campus San Joaquín (Santiago), el **Martes 11 de Abril a las 12:00**. La charla se transmitirá por videoconferencia al auditorio Claudio Matamoros (F106) en Casa Central (Valparaíso).

### Título

## To Index or not to Index? Text Search Engines based on String Compression.

### Expositores



#### Dr. Diego Arroyuelo

*Sub-Director, DI Campus San Joaquín*

#### Mini-Bio

Diego Arroyuelo es Profesor Auxiliar del Departamento de Informática, UTFSM. Sus áreas de interés incluyen estructuras de datos, análisis de algoritmos y compresión de datos. Sus principales aportes han sido en el área eficiencia de motores de búsqueda, búsqueda en texto comprimido, y representación sucinta de árboles. Su experiencia en la industria, incluye indexación y compresión de textos para Yahoo!. El profesor mantiene contacto con los principales grupos de investigación a nivel mundial en su área, y ha publicado resultados en conferencias y revistas

internacionales. Además, ha sido miembro del comité de programa de diversas conferencias internacionales, así como también revisor de importantes revistas del área.

### Resumen

El objetivo de las estructuras de datos sucintas y comprimidas es representar estructuras de datos usando la menor cantidad posible de espacio de memoria. Como consecuencia, permiten manipular grandes conjuntos de datos completamente en memoria principal, sin la necesidad de acceder a memoria secundaria. El objetivo de esta charla es mostrar avances en el uso de estructuras de datos sucintas, tanto en la teoría como en la práctica. En particular, se discutirá sobre aplicaciones en recuperación de información, donde se tiene una colección de documentos sobre la que se quiere buscar para satisfacer las necesidades de información de un usuario. En la charla se plantea reemplazar toda la maquinaria de indexación común en los motores de búsqueda por un método de compresión de strings (en este caso, la colección de texto) que es capaz de proveer toda la funcionalidad necesaria. En otras palabras, la compresión de strings es capaz de reemplazar a un índice invertido (tradicionalmente usado para lograr eficiencia), a un índice posicional (necesario para generar ranking de los resultados), y a un índice directo (que permite extraer snippets y páginas en caché), todo esto ocupando espacio proporcional al del string comprimido. Se mostrarán resultados experimentales que indican que ésta es una idea promisoriosa. Además, se mostrarán las líneas principales de investigación que se están siguiendo, para lograr avances tanto teóricos como prácticos que lleven a la transferencia tecnológica de los resultados.

### Lugar y Fecha

11 de Abril de 2017, 12:00

Sala de Reuniones, Departamento de Informática, UTFSM, Campus San Joaquín.

La charla se transmitirá en videoconferencia al Auditorio Claudio Matamoros (F-106).