



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

IDENTIFICACIÓN DE CONTENIDO MULTIMEDIA RELEVANTE A PARTIR DE  
EVENTOS UTILIZANDO SU INFORMACIÓN SOCIAL

TESIS PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN

MAURICIO DANIEL QUEZADA VEAS

PROFESORA GUÍA:  
BÁRBARA POBLETE LABRA

MIEMBROS DE LA COMISIÓN:  
SERGIO OCHOA DE LORENZI  
MAURICIO MARÍN CAIHUAN

SANTIAGO DE CHILE  
DICIEMBRE 2012



# Resumen

asdf



*Jason Funk disipa patitos*

# Agradecimientos

asdf

# Índice general





# Capítulo 1

## Introducción

Al igual que en el buffet de un restaurante, por mucho que se quisieran comer todos los platos favoritos, es imposible comer todo lo que uno quisiera por razones obvias. Una posibilidad es probar un poco de cada comida, para luego saber qué es lo más delicioso y comer hasta hartarse.

Pero, ¿qué hacer si hay demasiados platos y no se conocen todos? de alguna manera hay que saber cuáles hay que probar, si el objetivo es comer lo mejor posible. Un amigo puede recomendar una u otra comida, lo cual puede servir para orientarse. Entonces se pueden escoger pequeñas muestras de acuerdo a las recomendaciones.

Pasando a un contexto diferente, supóngase que este gran buffet es la Web y los distintos platos corresponden a contenido publicado en ella. Por lo tanto, dada la gran cantidad de información disponible, se hace necesario poder encontrar lo más atractivo de acuerdo a la preferencia del usuario o usuarios. Se está haciendo otra suposición importante con esta analogía, y es que se está considerando que la información es íntegramente para ser *consumida*, y no, por ejemplo, para generar más contenido, o para ser utilizada por máquinas, etc. Dentro de este contexto se plantea la pregunta de cómo seleccionar el contenido más atractivo dentro de todo lo que hay disponible en un momento dado.

Siguiendo el razonamiento de la analogía, una manera de poder seleccionar sólo el contenido más “atractivo” (de acuerdo a las preferencias del usuario), es probar un poco de cada uno. Diversos esfuerzos han sido hechos para este propósito, entre ellos está el de generar resúmenes automáticos<sup>1</sup> a partir de uno o múltiples documentos. Luego, utilizando las recomendaciones de otros usuarios es posible ordenar estos resultados de acuerdo a la relevancia que éstos les dan.

---

<sup>1</sup>??

- 1.1. Contexto
- 1.2. Motivación
- 1.3. Contribuciones
- 1.4. Alternativas analizadas
- 1.5. Objetivos
  - 1.5.1. Objetivo general
  - 1.5.2. Objetivos específicos
- 1.6. Descripción general de la solución
- 1.7. Resultados obtenidos

# Capítulo 2

## Antecedentes

2.1. Conceptos involucrados

2.2. Soluciones existentes

## Capítulo 3

### Especificación del Problema

- 3.1. Descripción detallada
- 3.2. Relevancia de una solución
- 3.3. Características de calidad
- 3.4. Criterios de aceptación

# Capítulo 4

## Descripción de la Solución

- 4.1. Desafíos técnicos
- 4.2. Metodología de desarrollo
- 4.3. Casos de estudio
- 4.4. Validación

# Capítulo 5

## Conclusiones

5.1. Resumen del trabajo realizado

5.2. Objetivos alcanzados

5.3. Relevancia del trabajo realizado

5.4. Trabajo futuro

# Capítulo 6

## Unclassified D:

### 6.1. Twitter

Twitter es un servicio que permite conectar a personas mediante mensajes cortos, rápidos y frecuentes. Estos mensajes son publicados en el perfil del usuario que los emite, pueden ser vistos directamente por los seguidores de este usuario o ser vistos directamente en el perfil o buscándolos mediante una funcionalidad que provee el servicio. Además, un usuario puede *seguir* a otros para poder ver en su *timeline* los mensajes de todos a quienes sigue.

### 6.2. Metodología de obtencion del dataset

Se describe a continuación el proceso diseñado para la obtención de datos para alimentar al sistema implementado.

Las etapas de generación del Dataset son las siguientes:

- Recolección de eventos (noticias y conciertos);
- Enriquecimiento de los eventos existentes mediante tweets; e
- Identificación de documentos a partir de los tweets por cada evento.

Se recolectaron datos (eventos y tweets) desde el 19 de noviembre de 2012 hasta XXXXXXXXXXXX todos los días desde la medianoche hasta que el procedimiento terminaba exitosamente.

#### 6.2.1. Recolección de eventos

Se consideraron dos tipos de eventos para el sistema: noticias y conciertos musicales. Los conciertos incluyen festivales de varios artistas.