

Sistemas Inteligentes

Trabajo 7 – Recuperación de Información

Profesor: Alejandro Figueroa

Ayudantes: Alexander Espina - Daniel Palomera

Fecha de publicación: 14/09/15

Fecha de entrega: 09/11/2015

Aspectos generales

- Poner en práctica los conceptos básicos de indexación.

Objetivos

Confrontar al alumno con un motor de indexación, los pasos relacionados con esta tarea, y las métricas para evaluar el desempeño de estos motores.

Desarrollo

En esta tarea utilizaremos como motor de recuperación de información Xapian:

<http://xapian.org/>

El objetivo de esta tarea es primero indexar todas las respuestas etiquetadas en la tarea anterior, y realizar búsquedas sobre esta colección con las preguntas.

Para indexar, el alumno debe considerar el texto de la respuesta, la id de la pregunta a la cual pertenece, y si es o no mejor respuesta. Después de indexar todas las respuestas, el alumno debe utilizar los títulos de las preguntas para buscar en el repositorio creado. Para esto debe solicitar a Xapian que le devuelva diez resultados (como el motor de búsqueda Google). Calcular la precisión, recall, F1-Score para este sistema. Además del MRR de la mejor respuesta. Posteriormente, hacer este análisis enviando los cuerpos y los títulos más los cuerpos al motor. Compare los resultados. Comente qué observa. ¿Hay alguna relación entre el largo del título y el desempeño de Xapian? ¿Hay relación entre el tipo de pregunta informacional/no-informacional y el desempeño de Xapian? ¿Y con respecto a las respuestas encontradas?

Nota:

Para el caso de los alumnos trabajando con Alexander Espina. Estos deben acercarse al ayudante para obtener el conjunto respectivo de datos. En vez de hacer un análisis informacional o no, pueden hacerlo para cada una de las cinco categorías de emociones.