

# Análisis y Recuperación de Información

Cursada 2013

## Trabajo práctico para la aprobación de la cursada

- 1 Considere dos motores de búsqueda A y B que utilizan la misma colección de documentos. Cada motor de búsqueda retorna los 30 documentos mejor rankeados para cada búsqueda, ordenados por relevancia con respecto a la misma. Suponiendo que existen 12 documentos relevantes en la colección para la búsqueda Q, y que los resultados arrojados por ambos buscadores son los siguientes:

A: + + - + - - - + - - - + - - - - - - - + - - - - - - - +  
B: - + + - - - + - - + + - + - - - + - - + - - + - + - - - - -

Donde + indica un documento relevante en la posición dada y - un documento irrelevante

- a Dibujar el gráfico de precisión/recall para ambos motores
  - b ¿Qué motor arroja mejores resultados? Explicar
- 2 Utilizar la aplicación provista por la cátedra para indexar la carpeta "movies" provista.
  - 3 Utilizar la aplicación luke de Apache, para inspeccionar el índice creado. Analizar si la información indexada es la esperada. En caso de no serlo, indicar cuál es el problema que presenta
  - 4 Inspeccione el código provisto de la aplicación provista por la cátedra. Describa qué cambiaría para solucionar el problema presentado en el punto 3.
  - 5 ¿Qué modificaría de la aplicación provista para que las películas en las que participa *Robert De Niro* aparezcan mejor posicionadas en el ranking de resultados?
  - 6 ¿Cómo modificaría la aplicación para realizar una búsqueda avanzada por género, actores y director?

**Fecha de entrega:** 14 de junio de 2013

**Consultas:** marcelo.armentano@isistan.unicen.edu.ar

# Análisis y Recuperación de Información

*Cursada 2013*

## Trabajo práctico para la aprobación de la materia (final)

- 1 Parsear los documentos html proporcionados por la cátedra para que los documentos Lucene contengan los siguientes campos
  - Género
  - Título
  - Imagen de tapa
  - Reparto (actores)
  - Director
  - Argumento
  - IMDB rating
  - IMDB link
- 2 Implementar los cambios sugeridos en el trabajo práctico de cursada relativos a los puntos 4, 5 y 6
- 3 Mejorar la presentación de los resultados de las búsquedas mostrando el título y la imagen de tapa de las películas

**Fecha de entrega:** 9 de agosto de 2013

**Consultas:** [marcelo.armentano@isistan.unicen.edu.ar](mailto:marcelo.armentano@isistan.unicen.edu.ar)