

Tarea 3

En esta tarea, vamos a evaluar la eficacia de modelos de recuperación de información sobre colecciones de TREC. Vamos a usar Terrier 4.0.

La tarea, paso a paso, consiste en lo siguiente:

1. Cree una carpeta donde coloque los documentos de las colecciones Associated Press, Federal Register, Wall Street Journal (todos los años) y Ziff-Davis. Los documentos debe estar descomprimidos.
2. Indexe la carpeta completa en Terrier.
3. Vamos a usar dos modelos: Tf-Idf y BM25 y los vamos a comparar.
4. Para cada modelo, evalúe por separado los topics: 51-100, 101-150 y 151-200.
5. Verifique los resultados en `var/results/foo.eval`. Recupere precisiones promedio sobre Q para los diez primeros lugares del ranking.
6. Usando gnuplot, genere 3 plots de curvas de precisión, una por cada batch de topics. Etiquete cada curva con el modelo que corresponda.

Importante: Puede hacer la tarea en grupos de hasta 3 personas. Empiece a hacerla de inmediato, puede encontrarse con dificultades en algunos pasos.

Entregables:

1. Informe: subir un informe con los resultados obtenidos en su tarea. Indique los integrantes del grupo.

Evaluación: Las tareas serán evaluadas de acuerdo a eficacia.

Entrega: Lunes 6 de Julio, en moodle. A las 24:00 Hrs. el link de *upload* se cerrará.