Information Retrieval

Practical session n°3: Weighting

Create a directory named *practice3*. In this directory, create a new file named *practice3_report.txt*.
During the practical session, for each exercise, copy-paste in this file some outputs of your program, showing that you complete the exercise and it works correctly. Add some explanations.
At the end of the practical session, copy-paste the source code of your program(s) in the directory *practice3*.
Compress the directory in a file named *practice3_YourTeamName.zip* (or *.tar*, *.gz*, *.rar*, etc.) (e.g.: *practice3_ VictorAlbertJulesIsaac.zip*).
Upload this compressed file (one file / team) on the website of the course. **Deadline October 16th** (first try); **October 23th** (final version).

Exercise 1: Increasing (again) the size of the collection

Download the collection of documents *Practice_03_data.zip* on the website of the course.
It contains 9,804 documents stored in a single file (76.4MB).
Index this collection using your indexing program (cf. Practical session n°2: dictionary, postings lists, *df*, *tf*).

Exercise 2: Collection Statistics

Compute again the collection statistics (cf. Exercise 2, Practical Session 2), without stop-words neither stemmer.

Exercise 3: Collection Statistics using stop-words and stemmer

Refresh the index, removing stop-words and applying Porter's stemmer.
Compute again the statistics of the exercise n°2.

Exercise 4: SMART *ltn* weighting

Compute a weighted index based on SMART *ltn* weighting function.

Exercise 5: Ranked Retrieval (*ltn* weighting)

Compute the score of each document for the query « web ranking scoring algorithm », using the index based on SMART *ltn* weighting function. Print the list of the ten most relevant documents, and their relevance score.

Exercise 6: SMART *ltc* weighting

Compute a weighted index based on SMART *ltc* weighting function.

Exercise 7: Ranked Retrieval (*ltc* weighting)

Compute the score of each document for the query « web ranking scoring algorithm », using the index based on SMART *ltc* weighting function. Print the list of the ten most relevant documents, and their relevance score.

Exercise 8: *BM25* weighting

Compute a weighted index based on *BM25* weighting function.

Exercise 9: Ranked Retrieval (*BM25* weighting)

Compute the score of each document for the query « web ranking scoring algorithm », using the index based on *BM25* weighting function. Print the list of the ten most relevant documents, and their relevance score.