Recuperação de Informação / Information Retrieval

2017/2018 MIECT/MEI, DETI, UA

Assignment 2

For this assignment you will create a simple searcher to work with the index created in assignment 1.

1. Create an index reader that loads the index from disk.

2. Create a class implementing a Boolean retrieval method.
   a. This method should assume disjunctive ("OR") queries: all words in a query should be combined using the OR operator.

   b. Implement two scoring options for the results:
      i. Number of words in the query that appear in the document
      ii. Total frequency of query words in the document

3. Process the queries (file 'cranfield.queries.txt') and retrieve the results for each query.
   a. Write the sorted results to a text file using the following tabular format:

query_id	doc_id	doc_score
1	211	21
1	33	17
(…)		
100	76	19

Note:

Your assignment will be evaluated in terms of: modelling, class diagram, code structure, organization and readability, correct use of data structures, submitted results, and report. See suggestions and submission instructions below.

Suggestions:

- Write **modular** code
- Favour **efficient** data structures
- Add **comments** to your code
- Follow the **submission instructions**

Submission instructions:

- To manage your project please use **Maven** (preferably) or Netbeans
- At each submission, include a small **Report** including:
    - Your project's **class diagram**
    - A description of each class and main methods, identifying where these are called
    - A block diagram and a high-level (but sufficiently detailed) description of the overall processing pipeline (data flow diagram)
    - Complete instructions on how to run your code, including any parameters that need to be changed
    - A list of any external libraries that are needed to run the code
    - Efficiency measures: total indexing time; maximum amount of memory used during indexing; total index size on disk
    - A short commentary/assessment of your own work, describing features or implementation decisions that you consider the most relevant/positive (or otherwise)
- Make sure you **include your name and student number** in the code and in the report.
- Make sure all your programs compile and run correctly.
- Submit your assignment by the due date using Moodle.