

Recuperação de Informação / Information Retrieval

2021/2022 MEI/MIECT, DETI, UA

Assignment 2

Submission deadline: **22 December 2021**

For this assignment, you will create a weighted (tf-idf) indexer and a ranked retrieval method. Use the datasets from assignment 1 (start with the smaller one).

1. Extend your indexer to apply term weighting and implement the following ranking methods.
 - 1.1. Vector space ranking with tf-idf weights. Use the *lnc.ltc* indexing schema as default, allowing other combinations.
 - 1.2. BM25 ranking. Use $k_1=1.2$ and $b=0.75$ as default parameters.
 - 1.3. Add a method to write the resulting index to file. Use the following format, or a similar one (one term per line):
term:idf;doc_id:term_weight;doc_id:term_weight;...
2. Process each query in the file 'queries.txt' and retrieve the sorted results for each query, using each ranking method.
 - 2.1. Save the top 100 results for each query to a file (one file per ranking method).

Instructions:

- **Modelling**, code **structure**, **organization** and **readability** will be considered when grading your project
- **Comment** your code; and make sure you include your name and student number
- Write **modular** code
- Favour **efficient** data structures
- Use **parameters**, preferably through the command line
- Make sure all your programs run correctly
- Submit your assignment by the due date using Moodle