**Recuperação de Informação / Information Retrieval**

**2020/2021 MEI/MIECT, DETI, UA**

# Assignment 3

Submission deadline: **14 January 2021**

For this assignment, you will continue extending your previous indexing and retrieval methods. Use the latest dataset available here: ai2-semanticscholar-cord-19/historical_releases.html

1. Implement the SPIMI approach in your indexing method.

2. Extend your indexer to store term positions.
   Write the resulting index to file using the following format (one term per line):
   term:idf;doc_id:term_weight:pos1,pos2,pos3,…;doc_id:term_weight:pos1,pos2,pos3,…

3. Extend your ranked retrieval method to account for query term proximity, that is, consider the shortest text span that includes all (or most, or several) query words and use this information for adapting the ranking score.
   3.1. Evaluate your retrieval methods (using the same queries as in Assignment 2) with and without query term proximity boost.

**Instructions:**

- Use Python or Java (in this case, manage your project with Maven)
- **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 compile and run correctly
- Submit your assignment by the due date using Moodle