

Assignment 3

Filipe Pires [85122], João Alegria [85048]

Information Retrieval

Department of Electronics, Telecommunications and Informatics

University of Aveiro

December 9, 2019

Introduction

This report was written for the discipline of 'Information Retrieval' and describes the implementation and evaluation of a ranked retrieval method that uses the indexes created with the solutions developed for the previous assignments.

We include the correction of design flaws of the delivery done prior to this one and the updates applied both to the text corpus indexation and to our class diagram. We also provide the instructions on how to run our code.

Along with the description of the solution, we also present the results of our calculations to evaluate the solution and determine its efficiency according to the metrics proposed for this last assignment (*I*). All code and documentation is present in our public GitHub project at <https://github.com/joao-alegria/RI>.

1. Re-Indexing the Corpus

The first step to index the document corpus with the format intended was to consider not only the title of each document but also the abstract

2. Ranked Retrieval of Relevant Documents

With an entirely functional index creator and a folder of generated indexes from the corpus of text documents, it was now time to develop a program capable of interpreting queries and returning the index entries most relevant to them. In this chapter we describe our implementation of a query results ranked retriever - that we called `Searcher` -, explain how we prepared it for memory limitations and present the updates done to our class diagram.

The file `Searcher.py` contains an abstract class called `Searcher` that serves as a template for the implementations of results retriever classes. For our purposes, we developed `IndexSearcher`, a class that extends from the abstract template and is capable of selecting which index files will be required to answer a given query, assigning scores to documents from the index and returning the documents considered most relevant to the query according to these scores.

3. Evaluation and Results Discussion

.....

Conclusions

After completing the assignment, we drew a few conclusions regarding our solutions and the whole concept of

The biggest challenge we faced was

From this assignment, we take

The overall perspective of our performance regarding the project is

References

1. S. Matos, *IR: Assignment 3*, University of Aveiro, 2019/20.