

Exercício para seleção aluno especial - Curso "Deep Learning aplicado a sistemas de buscas", FEEC-Unicamp, primeiro semestre de 2023.

Como tivemos muitas inscrições de alunos especiais, precisaremos realizar um processo seletivo. Neste ano, propomos desenvolver um pequeno projeto de Information Retrieval com a liberdade de utilizar o chatGPT para auxiliá-lo em todas as fases do projeto, i) para aprender sobre IR, ii) para codificá-lo em Python, e iii) para escrever o relatório.

Project Title: Building a Simple Information Retrieval System using BM25 and GPT-3 and evaluated in the CISI collection.

Project Overview:

The aim of this project is to introduce students to the basics of Information Retrieval (IR) and the BM25 ranking algorithm. The solution to this project will help us to select the students to attend the course on deep learning applied to search. The students will build a simple IR system that retrieves relevant documents from a pre-defined corpus based on the user's query. The students will also use the chatGPT to assist in all phases of the project. The coding will be done using Google Colab notebooks and the deliverables will be placed in github account.

Project Outcomes:

Upon completion of the project, students will have gained the following skills:

Understanding of the BM25 ranking algorithm

Familiarity with Information Retrieval systems

Use of GitHub to store the notebooks and documentation

Experience with chatGPT for help in find solutions and coding

Exposure to Google Colab notebooks for code execution and presentation.

Project Requirements:

Google Colab Notebook

GitHub account

chatGPT for help on all aspects of the project

CISI collection: http://ir.dcs.gla.ac.uk/resources/test_collections/cisi/

Familiarity with Python programming

Project Deliverables:

A github repository with:

- link shared with the instructors to evaluate the project contents
- all Google Colab notebooks and associated functions and libraries if any, with the code for the IR system
- A brief report describing the implementation details, results, how to test the IR system and how chatGPT helped you with the project.

Schedule:

The results should be submitted by 22nd February 2023 using this form:

<https://forms.gle/p3B83a1YYLFJtQkM9>

Evaluation:

The project will be evaluated based on the following criteria:

Completion of the project deliverables (partial is OK to submit)

Implementation accuracy

Quality of the report

Code readability and commenting

How chatGPT was used