

Assignment 1 - Search engine Query Parser

March 18, 2021

1 Description: Help the historian :D

In this assignment you are required to help Mr. Ahmed the historian with a simple search engine to help him search in some documents about famous characters from the islamic golden age. The documents and their index file are provided you are required to implement a query parser using the NLP techniques discussed in the tutorials and the lecture to parse the input query and retrieve the matched document.

2 Input Description

When you pull this assignment from the github course repo you will have the following:

- 10.txt files numbered from 0 to 9
- pickle file with name index.pkl, saving the index in dictionary format as follows:
index[word]=(doc number, line number)

3 Requirement

you are required to parse the following queries and return the matched documents.

queries

1. "Who were famous women?"
2. "Poems and poets"
3. "What is the scientific work of islamic golden age? "
4. "Ancient Spain"
5. "A Famous Muslim Engineer"
6. "A woman with a kingdom"
7. "How were the prayer times determined?"
8. "A theologian"

4 Assessment

After processing the following queries and retrieving the required documents answer the following questions:

1. which query returned in exactly one document?
2. which query returned almost all the documents?
3. which queries returned around 80% relevant results (more than one)?

4. which queries returned somehow irrelevant results?
5. why do you think some queries were successful in returning relevant results and others are not?
6. if you read any of the documents which character was totally new for you, what new information did you get?