**PROJECT DESCRIPTION**

**TOPIC:** Understanding text in health domain

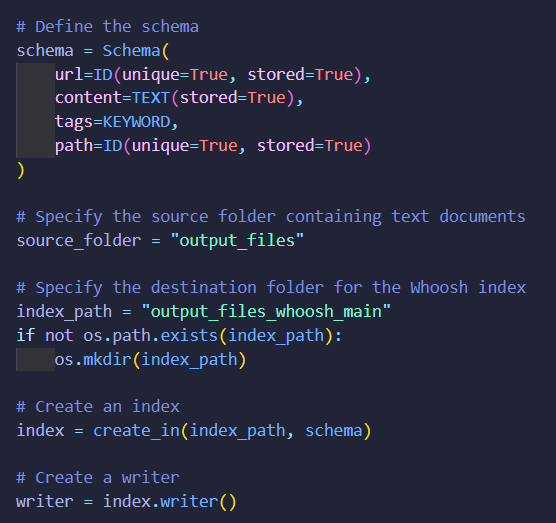
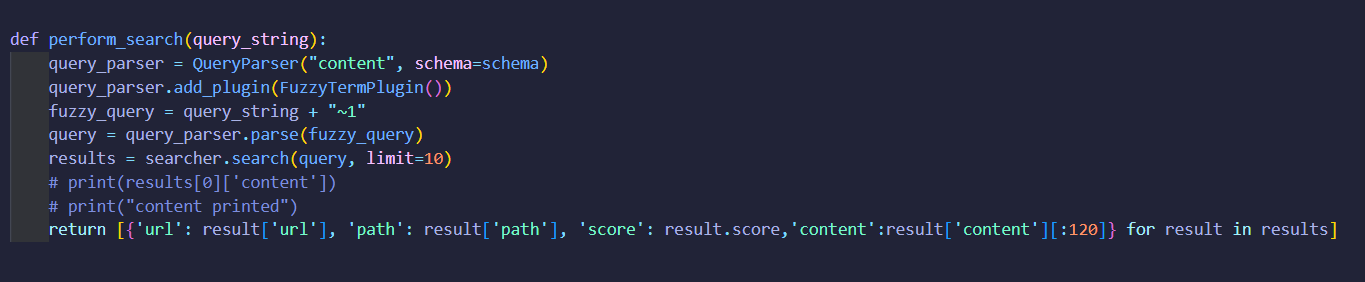
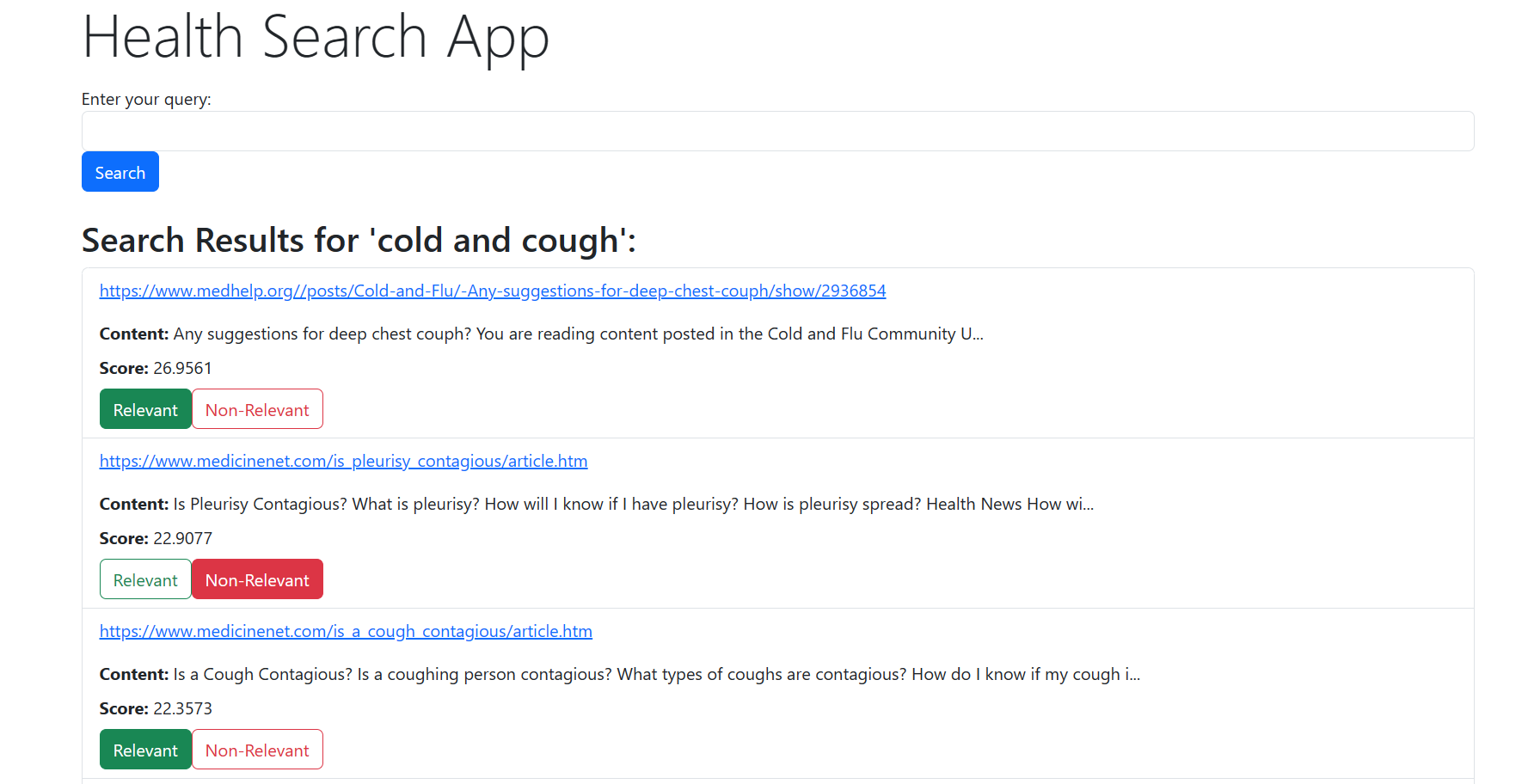
**TEAM MEMBERS:**

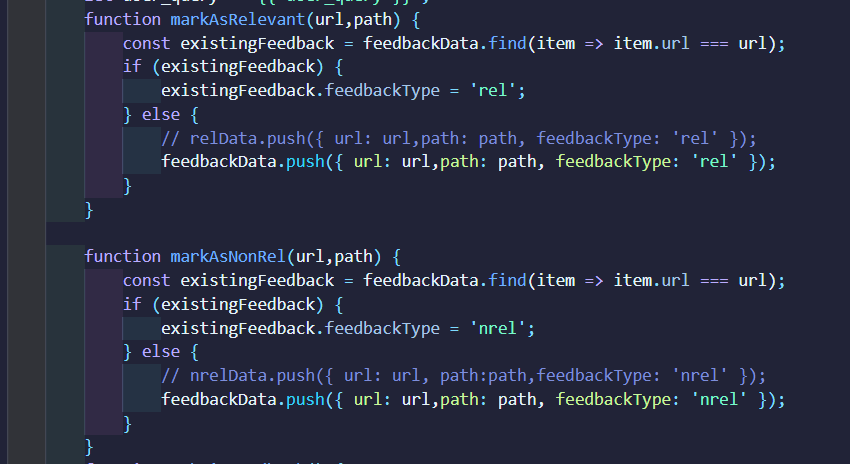
* K.Y Vaishnavi (S20210010109)
* Kavya Shridhar Lolla (S20210010117)

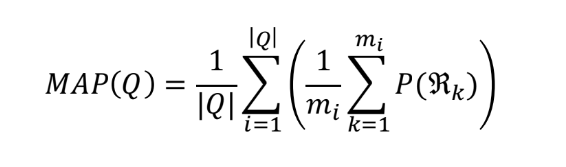
**TASKS:**

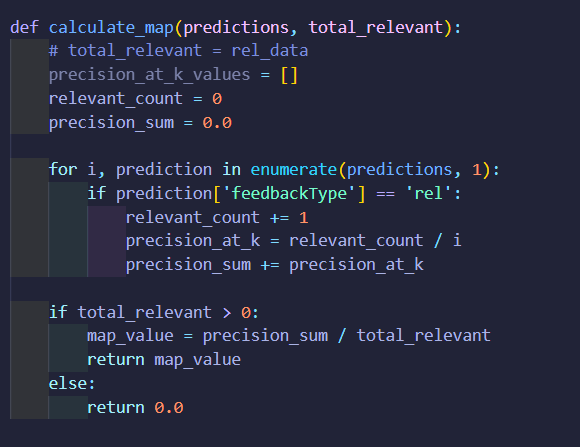
* Collecting documents
  + Creating code for crawling the web for appx. 100,000 links
  + Creating code for scraping all web pages and getting only the relevant information
* Indexing the documents
* Searching
* Ranking, scoring
* Relevance feedback
* Assessment components

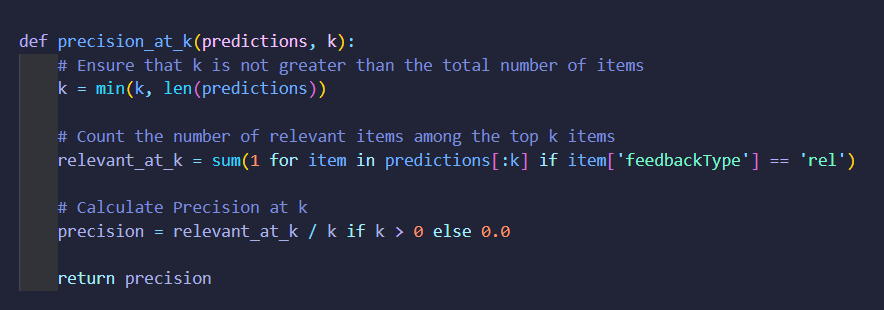
**DETAILS:**

* Indexing- written in file “whoosh-indexing.py”
  + schema defines the different fields in which the data from each document is stored in the index directory.
* Querying and searching
  + Done using inbuilt whoosh features
* The scoring is again handled by whoosh
  + They use the BM25 algorithm for scoring and ranking documents based on query
* The relevance feedback is received from the user and taken into account for re-ranking and calculation of assessment components
* Functions for handling relevant and non-relevant retrieved documents.



* On submitting the feedback, the assessment components are calculated
* The assessment components used are:
  + MAP- Mean Average Precision
    - Formula:



* Precision @ k (k=3):

OUTPUT:

