Computer Science Apprenticeship (CAP)

An-Najah National University

Information Retrieval Course - 2025-2026 (FALL)

Assignment-1: Creating and Querying Positional Indexes

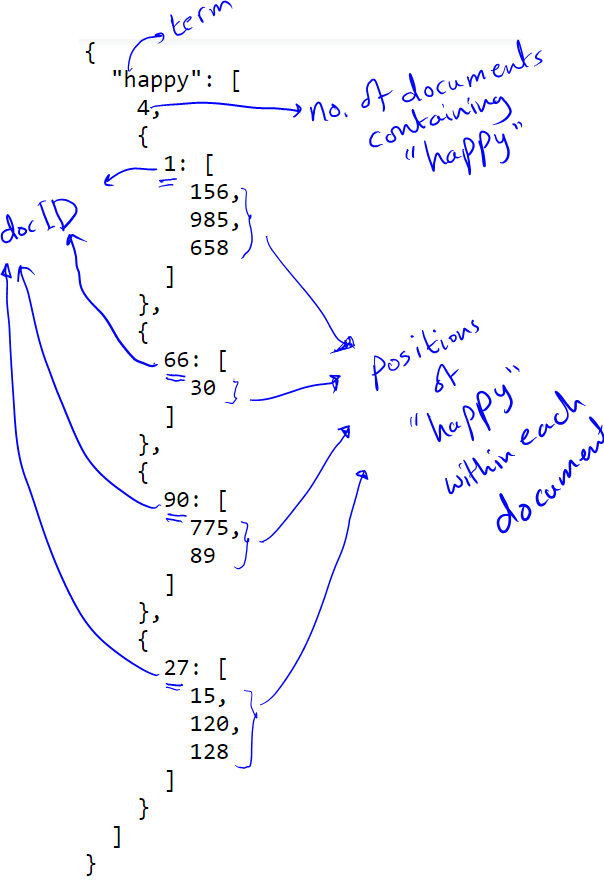
Instructor: Dr. Hamed Abdelhaq

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In this assignment, students have to write a C++ application to generate a positional inverted index from a given collection of documents stored in a folder. The generated index should be eventually stored in a file (pos\_inverted\_index.json), having each dictionary term placed in a single line. Each line in the file should obey the following json format:

{"happy": [4, {1 : [156, 985, 658]}, {66 : [30]}, {90 : [775, 89]}, {27 : [15, 120, 128]} ] }

The fields of this json elements are specified as follows:



In addition, you will create a file (docId\_filePath\_mapping.csv) containing the mapping between the document ids used in the index and the relative paths of corresponding files.

Finally, once the index is created and loaded into memory, prompt user to enter a phrase query, then use the created index to retrieve the relative paths of all relevant documents.

Note:

1. while parsing the documents, ignore English stop words and other words having a length less than 3 characters.
2. Try to write organized and well-documented code