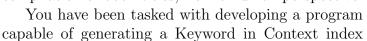# Quiz 1: Keyword Indexing Quest

Topics: sorting, searching - **A MUST use starter code**

Problem Statement

In computer science, algorithms for searching and sorting play crucial roles in both theoretical and practical applications. For instance, the binary search algorithm is a classical example of a straightforward algorithm that operates with sub-linear complexity. Similarly, Mergesort exemplifies an effective $O(nlogn)$ sorting algorithm on average, based on comparisons. Beyond these, Keyword in Context Indexing stands out as a method that significantly enhances the efficiency of searching through textual data, such as a compilation of book titles, from a human perspective.

You have been tasked with developing a program capable of generating a Keyword in Context index for a given set of book titles, while considering a specified list of words that should be disregarded. **In this Keyword in Context index, each title should be listed for every significant keyword it contains, with the entire index being alphabetically ordered by these keywords.** Any word not listed in the "words to ignore" list is considered a potential keyword.

For instance, if the words to ignore are "`an`, `in`, `for`, `with`, `of`, `on`, `the`," and the provided list of titles includes:

- `Rise of the Machines`
- `An Unseen Enemy Within`
- `On the Brink of Discovery`
- `In Search of Lost Time`

The generated Keyword in Context index would be the following:

```
        on the BRINK of discovery
  on the brink of DISCOVERY
      an unseen ENEMY within
    in search of LOST time
     rise of the MACHINES
              RISE of the machines
           in SEARCH of lost time
in search of lost TIME
           an UNSEEN enemy within
  an unseen enemy WITHIN
```

Input Format

The input file, which will be given as the *first command-line argument*, will consist of a series of lines. The string '...' will serve as a delimiter separating the list of words to ignore from the titles list. Each word to be ignored will be presented in lowercase, on a separate line, and will not exceed 10 characters. Titles will be on separate lines, they may include both uppercase and lowercase letters, and words within titles are separated by whitespace. No title will contain more than 15 words.

There will not be more than 50 words to ignore, no more than 200 titles, and the combined length of titles and words to ignore will not exceed 10,000 characters. The input will only include the characters 'a'–'z', 'A'–'Z', and whitespace.

Sample Input:

```
it
is
to
in
a
the
...
It is Time to Measure Time in a Timely Manner
The Time is Right
```

Output Format

Output the Keyword in Context index of the given book titles, **listing each title once for every keyword it contains**, **sorted alphabetically by the keyword**. If a word appears more than once in a title, it is considered a potential keyword for each occurrence. The keyword should be in uppercase, with the rest of the title in lowercase. **Titles with the same keyword should follow the same order as they appeared in the input**. When multiple instances of a word are keywords in the same title, they should be capitalized from left to right.

Sample Output:

```
it is time to measure time in a timely MANNER
it is time to MEASURE time in a timely manner
the time is RIGHT
it is TIME to measure time in a timely manner
it is time to measure TIME in a timely manner
the TIME is right
it is time to measure time in a TIMELY manner
```

> **Important Rules**
>
> - **You MUST use this starter code. Do not change any part of the given starter codes, those are given to ensure that you pass the unit tests in autograding. Only complete the TODOs.**
>
> - Test your codes using the Tur[6]Bo Grader and finally submit them via submit.cs.hacettepe.edu.tr using the same format given below:
>
>   - **<studentID>.zip**
>   * **Quiz1.java**

Academic Integrity Policy

All work **must be done individually**. You are encouraged to discuss the given problems with your classmates, but these discussions should be carried out in an abstract way. That is, discussions related to a particular solution to a specific problem (either in actual code or in pseudocode) **will not be tolerated**. In short, turning in someone else's work (including work available on the internet, or generated by the AI tools), in whole or in part, as your own will be considered as **a violation of academic integrity**. Please note that the former condition also holds for the material found on the web as everything on the web has been written by someone else.

> **!** The submissions will be subjected to a similarity check. Any submissions that fail the similarity check will not be graded and will be reported to the ethics committee as a case of academic integrity violation, which may result in the suspension of the involved students.