

GitHub: <https://github.com/rohansikder/MultithreadedIndexingAPI>

A multithreaded indexing API in Java 19+ that allows a word index to be created from an e-book or a URL.

This Java Program is created with Java 19 using the new Virtual Threads, To Process eBooks specified by the user. which shows all words that are in the eBook along with the definition and pages numbers that the word appeared on.

The words are filtered using Googles top 1000 Common word, but this can be changed by the menu

How to run program:

Please below command in CMD. The dictionary and common words should be in the same directory:

java --enable-preview -cp ./indexer.jar ie.atu.sw.Runner

```
C:\Users\Rohan\workspace\MultithreadedIndexingAPI>java --enable-preview -cp ./indexer.jar ie.atu.sw.Runner
-[-0;33m
*****
*      ATU - Dept. Computer Science & Applied Physics      *
*                                                           *
*      Virtual Threaded Text Indexer                        *
*                                                           *
*      Rohan Sikder - G00389052                             *
*                                                           *
*****
(1) Specify Text File
(2) Configure Dictionary - (Not Required - Dictionary provided)
(3) Configure Common Words - (Not Required - Google Top 1000 Word's provided)
(4) Specify Output File - (Not Required - Default will be output.txt)
(5) Execute
(6) Quit
-[-0;37mSelect Option [1-6]>
```

Menu Options:

1. Specify Text File - Please enter eBook directory into here.
2. Configure Dictionary - (Not Required - Dictionary provided) - Please enter dictionary directory into here.
3. Configure Common Words - (Not Required - Google Top 1000 Word's provided) - Please enter common words dictionary here.
4. Specify Output File - (Not Required - Default will be output.txt) - Please enter output file name here.
5. Execute - This executes and processes all files and then creates output file.
6. Quit - This exists the application

Options 2 to 4 are not required as they are provided for you.

How it works:

This program takes in an eBook and reads it line by line using virtual threads and then splits it up into individual words and puts it into a set, this is also done for the dictionary and common words which the program disregards when outputting the file. The program checks through the words from the eBook and dictionary and if they match it outputs its definition. Before this is done the most common words are removed from eBook words.