

G00397072 – Eoghan Barker

The user is given a menu of option. The first 5 relate to building the index and the other 3 are used for creating an output file.

A SkipListSet is used to hold the 1000 most common words. HashMap is used to store the Words and Definitions from the dictionary and for storing the index which is a key value pair of each word and its details (definition and pages). Only valid words that appear in the dictionary and not in the most common words list are stored in the index.

Each line of the input text file is handled by its own virtual thread. When adding a page number to the WordDetail of that word it uses a synchronized method to avoid data inconsistencies.

There are two options for text to output to a file:

- 1) A count of the unique words in the text document, this is equal to the size of the index as only unique words are added to it.
- 2) Each word in alphabetical order with its definition and the page numbers it appears on.