

Objective: work with binary search trees.

Write a C++ program that reads words from a text file and store them in lowercase with their file location in a binary search tree. For each word, store the line numbers and number of occurrences within each line. A word may exist in multiple lines and multiple times per line. Remove the possessive ('s) from all words. If a word begins or ends in a punctuation (period, comma, question mark, exclamation point, brackets, etc.), remove it before storing it in the index tree.

Your program should ignore the words stored in "ignore.txt". Some of these words include:

The, a, that, at, an, on, in, over, under, for, of, from, to, into, he, she, they, them, him, his, her, hers, their, ...

Each node in the tree should include a word and a list of locations (line number, number of occurrences).

Write a program that shows a menu of options to do the following:

1. Display Index in order (See sample output)
2. Search for a word
Ask the user for a word and display the node information where the word was found.
Found ethel: [<6, 3>, <8, 1>]
3. Most common word (return the first most common word and how many times it occurred).
The most common word is turing, it occurred 15 times.
4. Quit

Submit the two files

- indexing.cc: your main program
- index-tree.h: the tree class and all the functions related to the tree.

Option 1 Sample "Display Index":

```
12th: [<8, 1>]
1873-1947: [<6, 1>]
1881-1976: [<6, 1>]
2012: [<10, 1>]
active: [<10, 1>]
alan: [<1, 1>, <8, 1>]
anglo-irish: [<6, 1>]
army: [<8, 1>, <10, 1>]
baronet: [<6, 1>, <8, 1>]
baston: [<10, 1>]
batholomew's: [<6, 1>]
bengal: [<8, 1>]
between: [<10, 1>]
```

```
birth: [<8, 1>, <10, 1>]
ethel: [<6, 3>, <8, 1>]
family: [<5, 1>, <6, 2>, <8, 1>]
turing: [<1, 1>, [<6, 6>, <8, 4>, <10, 4>]
```

Grading:

Programs that contain syntax errors will earn zero points.

Programs that use global variables, other than constants, will earn zero points.

(70 points):

- (25 points) Creating the tree with all the entries in lowercase, includes loading the file and the insert function
- (10 points) all the ignore words are excluded
- (10 points) punctuations removed
- (10 points) Search function
- (10 points) most common word
- (5 points) Style and documentations

Follow the coding style outline on GitHub:

<https://github.com/nasseef/cs2400/blob/master/docs/coding-style.md>