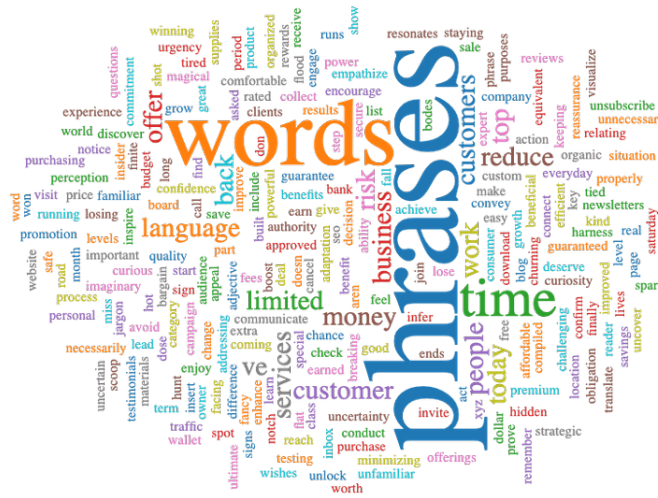


Assignment 5: The Spell Checker



Description/Contents of Main (Part A)

Spell checkers. Mankind's worst invention. The very reason why we cannot spell words properly and we why we cannot use their, there, and they're correctly. But it exists, and there is nothing we can do about it. Thus we will create our own spell checking program. In order to enhance the program's runtime, we will use an `std::unordered_map` to check the spelling of each word. Below is the outline of the program:

1. Load a dictionary file that contains around 60,000 words in the american english dictionary into an `std::unordered_map`, the dictionary file is called `"usa.txt"` and you can hard code this in your program
2. Read in a file and spell check the file: read the file and parse each word and verify if it's in the dictionary (filename is read in through standard input)
3. If the word is not in the dictionary (misspelled word) and the word has not been seen yet, you need to list out a set of suggested words (in alphabetic order)
4. You continue until you have reached the end of the file, you can use more than one `std::unordered_map` to implement your code

Suggested Words

Let's go over what I mean by suggested words. Suppose the word **"DOG"** is a misspelled word, then you need to try all words that are one letter off, here is a list of possible suggested words for **"DOG"**

- Sub list of words with different first characters
- "AOG"
- "BOG"
- "COG"
- "DOG" (Not a suggested word)

- "EOG"
- ...
- ...
- ...
- Sub list of all words with different second characters
- "DAG"
- "DBG"
- "DCG"
- "DDG"
- ...
- ...
- ...
- Sub list of words with different third characters
- "DOA"
- "DOB"
- "DOC"
- "DOD"
- ...
- ...
- ...
- Sub list of words with different letters added in front
- "ADOG"
- "BDOG"
- "CDOG"
- ...
- ...
- ...
- Sub list if words with different letters added between the first and second letter
- "DAOG"
- "DBOG"
- "DCOG"
- ...
- ...
- ...
- Sub list of words with a letter between the second and third letter
- "DOAG"
- "DOBG"
- "DOCG"
- ...
- ...
- ...
- Sub list of words with different letter added to the end

- "DOGA"
- "DOGB"
- "DOGC"
- ...
- ...
- ...

You may use any function in the string library to generate these words, you would probably need to insert them into a `std::vector<std::string>` object, and then you want to check each word if it's in the dictionary (if one of these words is in the dictionary then it's a valid suggested word), if not then you need to remove this word from the list (or you can just maintain another list of valid suggested words)

You will need to sort the list that you will display, you can use the `std::sort` function, you would need to include `<algorithm>` into your program, if your vector of strings is called words, then the following statement will sort the list

```
std::sort(words.begin(), words.end());
```

Specifications (Part A)

- Document your code
- No global variables
- All searches must use `std::unordered_map` find function or `[]` operator (no linear searching allowed in part a)
- Do not process the same misspelled word multiple times if it appears multiple times in the text file
- You must name the file `mainMap.cpp`

Part B

You will implement the same algorithm in part A except you will only use vectors (so you will only be allowed to do linear searching), thus your dictionary would be a `std::vector<std::string>` and you can have an additional vectors but not `std::unordered_map`

Specifications (Part B)

- Document your code
- No global variables
- All searches must a linear searches on vectors
- Do not process the same misspelled word multiple times if it appears multiple times in the text file
- You must name the file `mainVec.cpp`

Example Output

Part A and Part B code should give the same output

```
% g++ mainMap.cpp
% ./a.out
```

Enter filename: resident_evil.txt

CAPCOM is misspelled

No suggestions

STORYLINE is misspelled

No suggestions

GAMEPLAY is misspelled

No suggestions

% ./a.out

Enter filename: syphon_filter.txt

SCE is misspelled

Suggested words

ACE

ICE

SEE

SHE

SSE

SUE

EIDETIC is misspelled

No suggestions

LOGAN is misspelled

Suggested words

LORAN

SLOGAN

BIOTERRORISM is misspelled

No suggestions

PLAYSTATION is misspelled

No suggestions

HANDHELD is misspelled

Suggested words

HANDHOLD

SPINOFFS is misspelled

No suggestions

% ./a.out

Enter filename: gears.txt

ICONIC is misspelled

Suggested words
IRONIC

SCI is misspelled

Suggested words
SKI

FI is misspelled

Suggested words
BI
FA
FIB
FIG
FIN
FIR
FIT
FIX
FT
HI
II
MI
PI
TI
VI
XI

FENIX is misspelled

No suggestions

DOMINIC is misspelled

Suggested words
DOMINICA

ANYA is misspelled

Suggested words
ANNA

STROUD is misspelled

Suggested words
SHROUD

GROUNDBREAKING is misspelled

No suggestions

% ./a.out

Enter filename: patriots.txt

FOXBOROUGH is misspelled

No suggestions

NFL is misspelled

Suggested words

NIL

NUL

AFC is misspelled

Suggested words

ABC

AFT

ARC

DFC

BRADY is misspelled

Suggested words

BEADY

BRADS

BRANDY

BELICHICK is misspelled

No suggestions

% ./a.out

Enter filename: rickenbacker.txt

RICKENBACKER is misspelled

No suggestions

ICONIC is misspelled

Suggested words

IRONIC

BASSISTS is misspelled

No suggestions

PICKUPS is misspelled

No suggestions

FINGERBOARD is misspelled

No suggestions

EVERYONES is misspelled

No suggestions

Written Portion

You will need to submit a write up for the assignment along with your source code, the write up needs to discuss the following: If you are given n amount of words in the dictionary and m amount of words in the file you need to spell check and the length of every word is $O(1)$

1. What is the runtime if an unsorted vector of words is used to maintain the dictionary? Please explain with detail
2. What is the runtime if a sorted vector of words is used to maintain the dictionary? Please explain with detail
3. What is the runtime if an unordered_map of words is used to maintain the dictionary? Please explain with detail
4. What is the runtime if an AVL tree of words is used to maintain the dictionary? Please explain with detail

Submission

Submit the source code to code and the write up to canvas by the deadline

References

- Link to the top image can be found at <https://www.wordstream.com/blog/ws/2021/01/13/best-words-and-phrases-for-marketing>
- Supplemental Video <https://youtu.be/9BzVBPuTZxA>
- The list of words used in the dictionary can be found at <http://www.gwicks.net/dictionaries.htm>
- The text file were generated from the following link <https://deepai.org/machine-learning-model/text-generator>