

Source file

The source file that is used to generate inputFileOfWords file was came from <https://raw.githubusercontent.com/dwyl/english-words/master/words.txt>

With command

```
wget -q -O inputFileOfWords
https://raw.githubusercontent.com/dwyl/engli
sh-words/master/words.txt
```

It is downloaded on August 16, 2018

Prepare the input

And then run command

```
cat inputFileOfWords | sed 's/[^a-zA-Z ]//g' |
tr [:upper:] [:lower:] | tr -s ' ' '\n' | sort
| uniq > temp
mv temp inputFileOfWords
```

First use sed to remove all none alphabet character, and then use tr to translate all uppercase letters to lowercase letters. After that, use tr to translate all whitespace to newline character in order to place different words in different lines. Finally, use sort to sort all words in order to place all duplicated words in adjacent position, and uniq command can be used to eliminate the duplication now.

Number of words

After download the file from original URL. Command 'wc -w words.txt' output is 466544. And after the preparing input, command 'wc -w inputFileOfWords' output is 455247

Output

```
number of Latin root: 517329
number of Greek root: 284028
```