CSE102 - Introduction to Programming Spring 2022

Final Assignment, Due date: 10/06/2022 Friday, 23:59

- A MINI DESKTOP SEARCH ENGINE -

Note: Final assignment affects overall 30% of your overall grade. The assignment must be prepared in C programming language.

You are given a set of txt documents and an "ignoreList.txt" file in documents folder attached.

The file "ignoreList.txt" contains list of words (should be ignored during this analysis, such as common short words (a, the, ...), SGML/ HTML tags (lines starting with <), and punctuations (".", ",", ...).

Instructions- Part I

- (5 pts) Get a file name as input.
- (5 pts) Read all words in the input file document.
- (10 pts) Store all words in an array (it may be a struct array optionally) (ignore the words in which are on the ignore list).
- (25) Find the frequency count of all words
- (5 pts) Print all the words and their frequencies as in Example 1.
- (10 pts) Also print then output to "words.txt" file.

Example 1:

Content of the input file: *the first case is that of the wide flanged beam with convective heat transfer into the outer faces of the flanges. The second case considered is that of convective heat transfer into one side of a flat*

Content of the output file:
first,1
case,2
beam,1
convective,2
faces,1
wide,1
flanged,1
heat,2
flanges,1
flat,1
transfer,2
into,2
outer,1
one,1
second,1
side,1

Instructions- Part II

In the remaining part, user will search if a word exists in the file or not. A sample run is illustrated in Figure I.

- (10 pts) Sort the word list alphabetically applying a sorting algorithm as you wish and construct a new sorted array.
- (5 pts) Ask user "Enter a search string:".
- (20 pts) Search the input word by a **recursive search algorithm**.
- (5 pts) Report the existence status of searched word as in Figure I.

Figure 1- A sample run

Submission Details

Your program and documents folder must be in the same directory.
Submitted folder must contain all materials in a single folder as shown in Figure 2.

Ad	Değiştirme tarihi	Tür	Boyut
documents	24.05.2022 14:18	Dosya klasörü	
ignoreList	20.04.2015 15:33	Metin Belgesi	2 KB
miniSearchEngine	24.05.2022 14:33	C Source File	5 KB
screenshot	24.05.2022 14:33	JPEG resmi	110 KB
words	20.04.2015 15:33	Metin Belgesi	2 KB

Figure 2- Submitted folder must contain documents inputs, ignoreList.txt, your minisearch engine .c code, a screenshot of the running program and the output words.txt file in the same directory.

Please send your projects on time. If you submit your project late, you will lose 5 points for each late days. Please keep this in mind and promptly start working on your projects.

You are going to submit the compressed folder in a single zip file via aduzem. Zip file must be named as yournameyoursurname_studentnumber (example→ gozdealp_2007900011.rar or gozdealp_2007900011.zip)

You are required to exhibit an individual effort on this homework. In other word, everyone will send a separate homework for final assignment. Note that, **similarity test (jplag)** will be applied to all homeworks. Copying and cheating will cause getting zero grade.

Good Luck!