COMP20003 Algorithms and Data Structures

程序代[Assignment 45编程辅导

Taxi & For-Hire Vehicle Trip Dataset Information Binary Search Trees

> t: Friday, 23 of August , Monday, 9 of September

Purpose

The purpose of this assi

- Increase your proficiency in C programming, your dexterity with dynamic memory allocation and your understanding of linked data structures, through programming a dictionary.
- Increase your understanding of how computational complexity can affect the performance of an algorithm by conducting orderly experiments with your program and comparing the results of your experimentation with theory ment Project Exam Help
- Increase your proficiency in using UNIX utilities.

Background Email: tutorcs@163.com

A dictionary is an abstract data type that stores and supports lookup of key, value pairs. For example, in a telephone directory the outring key is a person of company name, and the value is the phone number. In a student record tookup, the key would be a student ID number and the value would be a complex structure containing all the other information about the student.

A dictionary can be implented in Catholic of the pair into a dictionary; insert a new item (key, value pair) into a dictionary; search for a key in the dictionary, and return the associated value. Most dictionaries will also support the operation delete an item.

Your task

In this assignment, you will create a simple instance of a dictionary, and we'll use it to look up information about for-hire vehicle trips in New York City.

There are two stages in this project. In the first stage you will code a dictionary in the C programming language, using a binary search tree as the underlying data structure. You will insert records into this dictionary from a file, and look up records by key. In the second stage, you will code additional functions to retrieve information from this dictionary. You will use a Makefile to direct the compilation of two separate executable programs, one for Stage 1 and one for Stage 2.

In both stages of the assignment, you will report on the number of key comparisons used for search and analyse what would have been expected theoretically. The report should cover each file used to initialize the dictionary.

You are *not* required to implement the delete functionality.

Stage 1 (7 marks)

In Stage 1 of this assignment, war Mke I evil for the South at on the program called dict1. The program dict1 takes two command line arguments: the first argument is the name of the data file used to build the dictionary; the second argument is the name of the output file, containing the data file to be consisted number of records, one per line, where the second argument is the name of the output file.

```
VendorID - Code t
                                                                                                                          rendor which produced the record
passenger_count
trip_distance -
                                                                                                                              in miles
RatecodeID - Code
                                                                                                                                 fare rate for the trip
store_and_fwd_flag
                                                                                                                    ther trip records were stored locally
PULocationID - TLC Taxi Zone where passengers were picked up
DOLocationID - TLC Taxi Zone where passengers were dropped off
payment_type - Cod to indigate payment type - Cod to indigate payment - Cod t
fare_amount - Fare for the trip
extra - Extra charges in USD
mta_tax - MTA tax in USD.
tip_amount - Tip iAuSSignment Project Exam Help
tolls_amount - Tolls in USD
improvement_surcharge - Improvement surcharge in USD
total_amount - Total
PUdatetime - Date time passengers
DOdatetime - Date/time passengers were dropped off
trip_duration - Duration of the trip in minutes
```

This data comes from a publicly-available dataset released by the New York City Taxi & Limousine Commission. More information about the dataset can be found at:

```
 \begin{array}{c} \text{https://wwwl.nyc.qov/site/t/about/tlc-trip-record-data.page} \\ \textbf{NUPS://tUtorcs.com} \end{array}
```

The field <PUdatetime> is an alphabetic string representing the date and time of the taxi trip in the format YYYY-MM-DD HH:mm:ss (year-month-day hour:minute:second). *The other columns* can be treated simply as the associated <data> field. Build a data structure of strings to save the associated data collected about each taxi trip. The maximum size that any string can be is 128 characters. Each string is separated by a comma ",". This is a standard csv format where the delimiter used is a comma.

The <PUdatetime> field will serve as the dictionary key, so the records will be sorted in temporal order. Note that because the datetime information is stored in lexicographical order, the values can be compared as strings (e.g., with strcmp()) to determine which trip is earlier/later. The <data> is the information sought during lookup.

In this assignment the search keys are not guaranteed to be unique – there are instances where multiple taxis pick up passengers at exactly the same day and time. You should handle duplicates by implementing a linked list for items with same key.

For the purposes of this assignment, you may assume that the input data is well-formatted, that the input file is not empty, and that the maximum length of an input record (a single full line of the csv file) is 256 characters. This number could help you to determine the buffer size to use when reading the file.

In this first stage of the assignment, you will:

• Construct a binary search tree to store the information contained in the file specified in the

command line argument. Each record should be stored in a separate Node.

• Search the binary starch records, i.e. from the screen.

ate a file of keys to be searched, one per line, and redirect For testing, it is of the input from the operator < to redirect input from a file.

- Examples of use:
 - dict1 dat then type in keys; or
 - dict1 dat < keyfile
- Your program will look up each key and output the information (the data found) to the output file specified by the second command line parameter. If the key is not found in the tree, you must output the word NOTEQUND.

 We Chat: cstutorcs

The number of key comparisons performed during both successful and unsuccessful lookups should be written to stdout.

- Remember that the entries of the file of not necessarily have unique keys. Your search must locate and output all the data found for a matching key.
- Email: tutorcs@163.com • Example output:
 - output file (information):

```
2018-12-15 01:49:13 --> VendorID: 1 || passenger_count: 1 || trip_distance: 1.9
|| RatecodeID: 1.0 || store_and_fwd_flag: 0 || PULocationID: 79 || DOLocationID: 234 ||
payment_type: http://payourtutology.com/mta_tax: 0.5 || tip_amount: 2.15
|| tolls_amount: 0.0 || improvement_surcharge: 0.3 || total_amount: 12.95 || DOdatetime:
2018-12-15 02:00:00 || trip_duration: 10 ||
2018-12-15 01:49:13 --> VendorID: 1 || passenger_count: 1 || trip_distance: 0.6
|| RatecodeID: 1.0 || store_and_fwd_flag: 0 || PULocationID: 79 || DOLocationID: 114 ||
payment_type: 1 || fare_amount: 5.0 || extra: 0.5 || mta_tax: 0.5 || tip_amount: 1.00
|| tolls_amount: 0.0 || improvement_surcharge: 0.3 || total_amount: 7.35 || DOdatetime:
2018-12-02 01:53:38 || trip_duration: 4 ||
1901-11-06 12:03:14 --> NOTFOUND
```

- stdout (comparisons):

```
2018-12-15 01:49:13 --> 423
1901-11-06 12:03:14 --> 401
```

Note that the key is output to both the file and to stdout, for identification purposes. Also note that the number of comparisons is only output at the end of the search, so there is only one number for key comparisons per key, even when multiple records have been located for that key.

The format need not be exactly as above. Variations in whitespace/tabs are permitted. The number of comparisons shown above was made up; do not take it as an example of a correct result.

Stage 2 (2 marks)

In Stage 2, you will code a function which takes a tantoone ID support is as front and returns to the output file all of the PUdatetime> keys from records which match the CPUlocationID>, using in-order tree traversal. The keys should be output in sorted temporal order (that is, earlier records should be printed first) the requested CPUlocationID> exist in the database, this function should wr
UND to the output file. As in Stage 1, the the number of comparisons made duri

The <PulocationID> to the specific point of the zones at the dataset website linked above, but you do not need these map store the <PulocationID> to the control of the zone as simply an integer. You may store the <PulocationID> inside the <data> field. As in Stage 1, you should handle duplicate keys by implementing a linked list for items with same key. Note that this means there may be more than one matching <PulocationID> for a single key. If this is the case, the key should be output multiple times to reflect the number of matches.

You should compile your code using a Makefile to produce an executable program called dict2. The program dict2 takes to Schalard the Insuments Office targum and the rank of the data file used to build the dictionary; the second argument is the name of the output file, containing the data located in the searches. You may reuse your record insertion code from Stage 1 to build the dictionary from the data lie in Stage 2. tutores @ 163.com

- Examples of use:
 - dict2 data cout out the type in ocation IDs; or
 - dict2 datafile outputfile < idsfile
- $\bullet \ \ \textbf{Example output:} \ \ https://tutorcs.com$
 - output file (information):

```
79 --> 2018-12-08 19:36:57
79 --> 2018-12-08 21:22:08
79 --> 2018-12-15 01:49:13
79 --> 2018-12-15 01:49:13
79 --> 2018-12-23 17:26:42
```

- stdout (comparisons):

79 --> 1528

The number of comparisons shown above was made up; do not take it as an example of a correct result.

Experimentation (4 marks)

You will run various files through your program to test its accuracy and also to examine the number of key comparisons used when searching different files. You will report on the key comparisons used by your Stage 1 dictionary dict1 for various data inputs and the key comparisons used by your Stage 2 dictionary dict2 for various data inputs too. You will compare these results with what you expected based on theory (big-O) for these algorithms and data structure.

Your experimentation should be systematic, varying the size and characteristics of the dataset files you use (e.g. sorted or random), and observing how the number of key comparisons varies. Repeating a test case with different key and taking the average database useful.

Some useful UNIX commands for creating test files with different characteristics include sort, sort -R (man sort for mor -R option), and shuf. You can randomize your input data and pick the first x eywords.

You will write up your findings and submit your results separately through the Turniting system. You will describe your results from each traditional tool on pare the Compared to the Compare

Tables and graphs are useful presentation methods. Selectionly informative data; more is not always better.

You should present your findings clearly, in light of what you know about the data structures used in your programs and in light of their known computational complexity. You may find that your results are what you expected, based on theory. Alternatively, you may find your results do not agree with theory. In either case, you should state what you expected from the theory, and if there is a discrepancy you should suggest possible reasons. You might want to discuss space-time trade-offs, if this is appropriate to your code and late.

You are not constrained to any particular structure in this report, but a useful way to present your findings might be:

- Introduction: Summary of data structures and inputs.
- Stage 1:
 - Data (number of key comparisons)
 - Comparison with theory
- Stage 2:
 - Data (number of key comparisons)
 - Comparison with theory
- Discussion

Implementation Requirements

The following implementation requirements must be adhered to:

• You *must* code your dictionary in the C programming language.

- You *must* code your dictionary in a modular way, so that your dictionary implementation could be used in another program without extensive rewriting or copying. This means that the dictionary operations are keptingether in a separate confidence in the main program.
- Your code should be assily extensible to allow for multiple dictionaries. This means that the functions for insertion

 take as arguments not only the item being inserted or a key for searching a pointer to a particular dictionary, e.g. insert (dict, item).
- Your program shows the space of the space for the space
- A Makefile is not provided for you. The Makefile should direct the compilation of two separate programs: dict1 and dict2. To use the Makefile, make sure it is in the same directory of your code, and type make dict1 to make the dictionary for Stage 1 and make dict2 to make the dictionary for Stage 2 you must submit your makefile with your assignment. Hint: If you havent used make before, try it on simple programs first. If it doesn't work, read the error messages carefully. A common problem in compiling multifile executables is in the included header files. Note also that the whitespace before the command is a tab, and not multiple spaces. It is not a goodine as had your program as a single file and then try to break it down into multiple files. Start by using multiple files, with minimal content, and make sure they are communicating with each other before starting more serious coding.

Email: tutorcs@163.com

Data

The data files are provided to home 1988 170 stafiles / on JupyterHub. The data format is as specified above in Suge 1.

No attempt has been made to remove or prevent duplicate keys in the original files, so you should expect duplicate keys. Our strip Sily of that of the Sata County making sure it complies with a csv standard specification.

Resources: Programming Style (2 Marks)

Two locally-written papers containing useful guidelines on coding style and structure can be found on the LMS Resources \rightarrow Project Coding Guidelines, by Peter Schachte, and below and adapted version of the LMS Resources \rightarrow C Programming Style, written for Engineering Computation COMP20005 by Aidan Nagorcka-Smith. Be aware that your programming style will be judged with 2 marks.

```
1  /** ************************
2 * C Programming Style for Engineering Computation
3 * Created by Aidan Nagorcka—Smith (aidann@student.unimelb.edu.au) 13/03/2011
4 * Definitions and includes
5 * Definitions are in UPPER_CASE
6 * Includes go before definitions
7 * Space between includes, definitions and the main function.
8 * Use definitions for any constants in your program, do not just write them
9 * in.
10 *
11 * Tabs may be set to 4—spaces or 8—spaces, depending on your editor. The code
12 * Below is ``gnu'' style. If your editor has ``bsd'' it will follow the 8—space
13 * style. Both are very standard.
14 */
15
```

```
16 /**
17 * GOOD:
                        程序代写代做 CS编程辅导
18 */
20 #include <stdio.h>
21 #include < stdlib .h>
22 #define MAX_STRING_SI
23 #define DEBUG 0
24 int main(int argc,
25
26
27 /**
  * BAD:
28
29
31 /* Definitions and includes
32 #include <stdlib.h>
33 #define MAX_STING_SIZE 1000
/* Definitions are give hames like variables to torcs
36 #include <stdio.h>
_{
m 37} /* No spacing between includes, definitions and main function */
38 int main(int argc, char **argw) {
                        Assignment Project Exam Help
39
41 /** ****
  * Variables
* Give them useful lover case rames or camel case for her 3 incom

44 * as long as you are consisted and apply the same 3 lecom
  * Initialise them to something that makes sense.
46
47
                        QQ: 749389476
48 /**
  * GOOD: lower_case
49
50
51
  int main(int argc, chatters://tutorcs.com
52
53
    int i = 0;
54
    int num_fifties = 0;
55
    int num_twenties = 0;
56
    int num_tens = 0;
57
58
59
60 /**
  * GOOD: camelCase
  int main(int argc, char **argv) {
65
    int i = 0;
66
    int numFifties = 0;
67
    int numTwenties = 0;
68
    int numTens = 0;
69
70
71
   . . .
72 /**
  * BAD:
73
74
75
  int main(int argc, char **argv) {
76
77
    /* Variable not initialised - causes a bug because we didn't remember to
78
   * set it before the loop */
79
    int i;
80
   /* Variable in all caps - we'll get confused between this and constants
```

```
82
     int NUM_FIFTIES = 0;
83
     /* Overly abbreviate程a序代的的的S编程辅导int nt = 0
84
     int nt = 0
85
86
     while (i < 10) {
87
88
89
       i++;
90
91
92
93
94
     Spacing:
95
    Space intelligently
                                          oup blocks of code that are doing a
     specific operation, or to separate
                                          variable declarations from other code.
   * One tab of indentation within either a function or a loop.
   * Spaces after commas.
   * Space between ) and Wenchatt i Cstuetorics the main
   * function.
  * When declaring a pointer variable or argument, you may place the asterisk
* adjacent to either the type or to the variable name.

* Lines at most 80 changes grament Project Exam Help

* Closing brace goes on its owering.
107
                        Email: tutorcs@163.com
109 /**
   * GOOD:
110
111
   */
112
   int main(int argc, cha
                              ): 749389476
113
114
    int i = 0;
115
116
    for(i = 100; i >= 0; i--) {
117
118
      if (i > 0) {
        printf("%d bottlhttps;//atutorcs.dcom around,"
119
         % d bottles of beer. \sqrt{n}, i, i - 1);
120
121
      } else
        printf("%d bottles of beer, take one down and pass it around."
122
         We're empty.\n", i);
123
124
    }
125
126
127
    return 0;
128
129
   /**
130
131
   * BAD:
132
133
  /* No space after commas
134
   * Space between the ** and argv in the main function definition
135
   st No space between the ) and \{ at the start of a function st/
136
   int main(int argc, char ** argv){
137
    int i = 0;
138
    /* No space between variable declarations and the rest of the function.
139
    * No spaces around the boolean operators */
    for (i=100; i>=0; i--) {
141
    /* No indentation */
142
   if (i > 0) {
143
    /* Line too long */
144
  printf("%d bottles of beer, take one down and pass it around, %d
bottles of beer.\n", i, i - 1);
147 } else {
```

```
/* Spacing for no good reason. */
148
149
   printf("%d bottles o 程er序性响响代的 CS编程辅导" We're empty.\n", i 程序性响响性的 CS编程辅导
150
152
153
154
   /* Closing brace not
155
   return 0;}
156
157
158 /** *****
  * Braces:
159
160 * Opening braces go o
                                         the loop or function name
161 * Closing braces go o
                                        🗖 tion level as the thing they are
162 * Closing braces go
163 * closing
164 */
165
166 /**
                       WeChat: cstutorcs
167 * GOOD:
168 */
169
  int main(int argc, char **argw) {
170
                        Assignment Project Exam Help
171
172
173
    for (...) {
                       Email: tutorcs@163.com
175
176
177
    return 0;
178
179
                       QQ: 749389476
180
181 /**
   * BAD:
182
183
   int main(int argc, chattags://tutorcs.com
184
185
186
187
188
    /* Opening brace on a different line to the for loop open */
189
    for (...)
190
191
192
      /* Closing brace at a different indentation to the thing it's
193
      closing
196
197
    /* Closing brace not on its own line. */
198
        return 0;}
199
200
201 /** ********
202 * Commenting:
203 * Each program should have a comment explaining what it does and who created
205 * Also comment how to run the program, including optional command line
206 * parameters.
207 * Any interesting code should have a comment to explain itself.
  * We should not comment obvious things — write code that documents itself
209 */
210
211 /**
212 * GOOD:
213 */
```

```
214
  /* change.c
215
  * Created by Aidan Na 程序。代语句《Strick In CS.编程辅导
216
218 13/03/2011
219
                                          uld be needed to make up some
220
  * Print the number of
221 change
   * that is input by th.
222
223
   * To run the program
224
   * ./coins —num_coins
                                         Lapezoid —output blabla.txt
225
226
   * To see all the inpu
227
   * ./coins —help
   * Options::
                             Show help message
230
      --help
      --num_coins arg
                             Input number of coins
231
                          Vacabahan x Stutoras 1
      -shape_coins arg
232
      —bound arg (=1)
233
      -output arg
                             Output solution file
234
235
236
                          Assignment Project Exam Help
237
   int main(int argc, char
     int input_change = 0;
240
    printf("Please inpuEmail: tutores@c163.com
241
242
     inclusive):\n");
243
    scanf("%d", &input_change);
244
    printf("\n");
245
246
     // Valid change values are 0-99 inclusive.
247
     if (input_change < 0 \mid \mid input_change > 99) {
248
      printf("Input not in the range 0-99.\n")
249
250
                        https://tutorcs.com
251
252
253
  /**
254
   * BAD:
255
256
257
   /* No explanation of what the program is doing */
258
   int main(int argc, char **argv) {
260
     /* Commenting obvious things */
261
     /* Create a int variable called input_change to store the input from
262
263
     the
264
     * user. */
265
     int input_change;
266
267
268
269
   * Code structure:
   * Fail fast - input checks should happen first, then do the computation.
272 * Structure the code so that all error handling happens in an easy to read
273 * location
274 */
275
276 /**
277 * GOOD:
278 */
279 if (input_is_bad) {
```

```
printf("Error: Input was not valid. Exiting.\n");
280
    exit(EXIT_FAILURE);
281
                              予代写代做 CS编程辅导
282
283
  /* Do computations here
284
285
286
287 /**
  * BAD:
288
289
290
  if (input_is_good) {
291
                                          ne else part off the screen.
    /* lots of computa
292
293
294
    else {
295
    fprintf(stderr, "Error:
                           Input was not valid. Exiting.\n");
296
    exit(EXIT_FAILURE);
297
                        WeChat: cstutorcs
298
```

Additional Support Assignment Project Exam Help

Your tutors will be available to help with your assignment during the scheduled workshop times. Questions related to the assignment may be posted on the Piazza forum, using the folder tag assignment for new posts. You should feel free to answer other students questions if you are confident of your skills.

A tutor will check the Discussion Forum regularly, and answer some questions, but be aware that for some questions you will just need to use your independ document your thinking. For example, a question like, How much data should I use for the experiments?, will not be answered; you must try out different data and see what makes sense.

In this subject, we'll be stopleting the started figure Hub Office Hub Office

Submission

You will need to make *two* submissions for this assignment:

- Your C code files (including your Makefile) will be submitted through the LMS page for this subject: Assignment 1 → Assignment 1: Code.
- Your experiments report file will be submitted through the LMS page for this subject: Assignments
 → Assignment 1 → Assignment 1: Experimentation. This file can be of any format, e.g. .pdf, text
 or other.

Program files submitted (Code)

Submit the program files for your assignment and your Makefile.

If you wish to submit any scripts or code used to generate input data, you may, although this is not required. Just be sure to submit all your files at the same time.

Your programs *must* compile and run correctly on the shared JupyterHub instance. You may have developed your program in another environment, but it still *must* run on the JupyterHub at submission

time. For this reason, and because there are often small, but significant, differences between compilers, it is suggested that if you are working in a different environment, you upload and test your code on the shared JupyterHut instance at reasonably frequent intervals.

A common reason for programs not to compile is that a file has been inadvertently omitted from the submission. Please checally a large and resubmit all files if necessary.

Experiment file subr

As noted above, your explicit f and f be submitted through the LMS, via the Turnitin system. Go to the LMS page for f and follow the prompts f and follow the prompts f and f are f are f and f are f are f and f are f are f are f and f are f are f and f are f are f are f and f are f are f and f are f are

Your file can be in any format. Plain text or .pdf are recommended, but other formats will be accepted. It is expected that your experimental work will be in a single file, but multiple files can be accepted. Add your uselvance to the top of your experiments file.

Please do not submit large data files. There is no need to query every key on the dictionary.

Assessment

Assignment Project Exam Help

There are a total of 15 marks given for this assignment, 7 marks for Stage 1, 2 marks for Stage 2, and 4 marks for the separate of the separat

Your C program will be marked on the possible for contract the possible formatted on the possible formatted on the possible formatted on the possible formatted so their job, etc. The documentation should explain all major design decisions, and should be formatted so that it does not interfere with reading the code. As much as possible, try to make possible formatted so that it does not interfere with reading the code. As much as possible, try to make possible formatted so that it does not interfere with reading the code.

Your experimentation will be marked on the basis of orderliness and thoroughness of experimentation, comparison of your results with theory, and thoughtful discussion.

Plagiarism

This is an individual assignment. The work must be your own.

While you may discuss your program development, coding problems and experimentation with your classmates, you must not share files, as this is considered plagiarism.

If you refer to published work in the discussion of your experiments, be sure to include a citation to the publication or the web link.

Borrowing of someone elses code without acknowledgment is plagiarism. Plagiarism is considered a serious offense at the University of Melbourne. You should read the University code on Academic honesty and details on plagiarism. Make sure you are not plagiarizing, intentionally or unintentionally.

You are also advised that there will be a C programming component (on paper, not on a computer) in the final examination. Students who do not program their own assignments will be at a disadvantage for this part of the examination.

Administrative issues

When is late? What do in an late: The ductate and the arrive of the front of this document. The lateness policy is on the handout provided at the first lecture and also available on the subject LMS page. If you decide to make a late submission, you should send an email directly to the lecturer as soon as poss; in the lateness policy is on the handout provided at the first lecture and also available on the subject LMS page. If you decide to make a late submission, you should send an email directly to the lecturer as soon as poss; in the lateness policy is on the handout provided at the first lecture and also available on the subject LMS page. If you decide to make a late submission, you should send an email directly to the lecturer as soon as poss; in the lateness policy is on the handout provided at the first lecture and also available on the subject LMS page. If you decide to make a late submission, you should send an email directly to the lecturer as soon as poss; in the lateness policy is on the handout provided at the first lecture and also available on the subject LMS page.

What are the marks a transfer teria Recall that this project is worth 15% of your final score. There is also a high transfer to use the projects to pass this transfer to the projects to pass this transfer terial recall that this project is worth 15% of your final score. There is also a high transfer terial recall that this project is worth 15% of your final score.

Finally Despite all these here to help! There is information about getting help in this subject on the LMS project will be answered in the LMS discussion group.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com