

CMPUT 291-F2019 - Project 2 Evaluation

Demo: ____/80 Report: ____/15 Group Work: ____/5 Total Mark: ____/100
TA Name: _____ Date: ____/____/____

Name	CCID	Present?
1.		
2.		
3.		

Member CCID whose code is demoed: _____ On time? _____
If the code is changed, the number of edits (20 is the limit): _____ Penalty: _____

General Guidelines:

- In Phases 1 and 2, there is no partial mark if the expected output is not returned.
- For queries with many results, do a quick check in full-output mode to partially assess if the results are correct.
- Entries with gray backgrounds are valid for both 100k.txt and 20k.txt. If marking on 100k.txt file, look at all the results, if marking on 20k.txt file, only look at gray background results.
- Take off 20 marks if the 100k.txt file is not handled.
- If Phases 1 and 2 are done in the same step, measure the running time for both phases.
Give the mark for performance if the running time is within the sum of expected time for the two phases (i.e. 3 marks if under 100sec and 2 marks if under 2 min).
- In Phase 3, all queries should be instant (under 1 second to finish) except Sec VII. If not, deduct up to 50% of the mark. If a query takes more than 10 seconds, it should be considered as “fail” and no mark should be allocated.
- For Phase 3.VII, if a query takes more than 30 seconds, it should be considered as “fail” and no mark should be allocated.

Phase 1. Preparing Data Files ____/10

1. Performance of preparing data files (assuming correct files are produced) ____/2

Use “date; <running command>; date” to measure the elapsed time.

Total time: _____

____/1 Phase 1 is performed on 100k data in less than 25 seconds.

____/1 Phase 1 is performed on 100k data in less than 45 seconds.

2. Correctness of the data files ____/8

____/2 Correct terms.txt. The following command returns the given list.

grep marketbuzz terms.txt

```
b-marketbuzz:1
b-marketbuzz:1
b-marketbuzz:72038
b-marketbuzz:72038
b-marketbuzz:72155
b-marketbuzz:72155
b-marketbuzz:72155
b-marketbuzz:72155
b-marketbuzz:72197
b-marketbuzz:72197
b-marketbuzz:72197
```

b-marketbuzz:72197

_____/2 Correct dates.txt. The following command returns the given list.

```
grep -c 2000/08/09 dates.txt  
190 (28 for 20K file)
```

_____/2 Correct emails.txt. The following command returns the given list.

```
grep bryan.williams@enron.com emails.txt  
  
cc-bryan.williams@enron.com:722  
bcc-bryan.williams@enron.com:722  
to-bryan.williams@enron.com:7717  
to-bryan.williams@enron.com:57456  
to-bryan.williams@enron.com:67577  
to-bryan.williams@enron.com:67578  
to-bryan.williams@enron.com:67579  
to-bryan.williams@enron.com:83666  
to-bryan.williams@enron.com:83667  
to-bryan.williams@enron.com:83668  
to-bryan.williams@enron.com:84382  
to-bryan.williams@enron.com:95451
```

_____/2 Correct recs.txt. The following command returns the given.

```
grep 19987 recs.txt  
  
19987:<mail><row>19987</row><date>2000/12/13</date><from>rcavanagh@nrdc.org</from><to>jeff.d  
asovich@enron.com</to><subj>Re:Portfolio file for  
distribution</subj><cc></cc><bcc></bcc><body>Jeff, I&apos;d be grateful if you could  
retransmit this in more accessible code.</body></mail>
```

Phase 2. Indexing ____/5

1. Phase 2 is performed on 100k data in less than 75 seconds (and correct indexes are produced) ____/1

Use “date; <running command>; date” to measure the elapsed time.

Total time: ____

2. Correctness of the indexes ____/ 4

_____/1 A correct hash index re.idx. The following command returns the given output.

```
db_dump -p re.idx | grep 19987  
  
19987  
<mail><row>19987</row><date>2000/12/13</date><from>rcavanagh@nrdc.org</from><to>jeff.dasovic  
h@enron.com</to><subj>Re:Portfolio file for  
distribution</subj><cc></cc><bcc></bcc><body>Jeff, I&apos;d be grateful if you could  
retransmit this in more accessible code.</body></mail>
```

_____/1 A correct B+ tree index te.idx. The following command returns the given list.

```
db_dump -p te.idx | grep marketbuzz
```

Returns the word b-marketbuzz 4 times (once for 20K).

_____/1 A correct B+ tree index da.idx. The following command returns the given list.

```
db_dump -p da.idx | grep -c 2000/08/09
```

190 (28 for 20K)

_____/1 A correct B+ tree index em.idx. The following command returns the given list *in the correct order*.

```
db_dump -p em.idx | grep bryan.williams@enron.com
```

bcc-bryan.williams@enron.com

cc-bryan.williams@enron.com

to-bryan.williams@enron.com (10 times for 100K and once for 20K)

Phase 3. Data Retrieval ____/65

Run the search program and perform the following queries. To get the full mark for each query, the system response must be quick and the results must be correct (check the general guideline on top for deductions). For accuracy, check the expected result of each query.

I. Single field search ____/12

_____/2 **subj: intel** (-0.5 for every extra/missing line)

```
10832 News from Intel(R)
11559 Re: Intel & Tamco High Plains Ethanol done deal
94366 FoolWatch: Duel Over Intel
94615 Breakfast With The Fool: IBM, Intel Announce Chip Advances
```

_____/1 **subj : intel**

[Same results as previous one. Deduct full mark if not match.]

_____/1 **subj:inTEL**

[Same results as previous one. Deduct full mark if not match.]

_____/2 **body:energy-reduction** (-0.5 for every extra/missing line)

```
19350 Re: Governor's Response to FERC
19352 Governor's Proposals Go Public
64045 Re: Governor's Response to FERC
65154 Energy Issues
```

_____/2 **to:charla.reese@enron.com** (-0.5 for every extra/missing line)

```
10153 Re: International Information - GIS
10284 Re: Charla Reese - New Position!
40766 CPS BEAR Volunteers
82830 Re: Executive Committee Invitation to 104 KRBE / Enron Earth Day Festival
```

_____/1 **from:wxm@cpuc.ca.gov** (-0.5 for every extra/missing line)

```
18565 FW: Edison bill on the role of the UDC
20756 FW: Strategic Planning Director
```

_____/2 cc : ccstaff@haas.berkeley.edu (-0.5 for every extra/missing line)

15579 This survey is only for students enrolled in courses using Infopier
16006 Training Workshops for Students
20433 Izio Infopier E-mail Messages and Accounts
21803 Who To See For Izio Infopier Support

_____/1 bcc : eric.merten@enron.com (-0.5 for every extra/missing line)

1511 BMC Update
60631 Correction: Risk and Purchasing Meeting -- August 8th

II. Multiple fields search ____/10

_____/3 subj: diversity body: there (-1 for every extra/missing line)

5710 Re: ENA Diversity Focus Groups
6591 Re: Mark Your Calendars: March 30 is Diversity/OLER Open House
96403 Nuclear diversity swap with Select Energy

_____/3 date: 2000/01/06 to:gretel.smith@enron.com (-1 for every extra/missing line)

6383 Houston Analyst North America Presentation
6386 Brenda Herod's new pager #
6388 Re: Alumni List

_____/2 Toronto-based from: ann.schmidt@enron.com (-0.5 for every extra/missing line)

17545 Enron Mentions - 07/19/01
63707 Enron Mentions
64163 Enron Mentions
66876 Enron Mentions - 07/19/01

_____/2 cc: ray.alvarez@enron.com bcc:vince.kaminski@enron.com (-0.5 for every extra/missing line)

16376 California 6/13
16694 California Update 6/25/01
17711 California Update 7/24/01
66910 California Update 7/16/01

III. Partial match search ____/10

_____/6 worldb% (-1 for every extra/missing line)

11835 Hotel California ISO
11896 Re: Hotel California ISO
16082 VentureWire, Tuesday, October 24, 2000
35437 FW: It's a tough game
35438 FW: It's a tough game
62436 Re: PRESENTATION ON CURRENT OIL AND GAS MARKETS

_____/4 subj:photograph% (-1 for every extra/missing line)

780 Re: Photograph for eCompany Now
10235 Re: Question re: photographs on MV/VP Biographic Profiles
44164 Re: Annual Report Photograph
69061 Denis Reggie Wedding Photography

IV. Range search ____/7

____/2 **date: 1999/06/11** (-0.5 for every extra/missing line)

```
9082 Topock diesel pigging strategy
9093 Cleaning Postponement of the pipeline between TW PG&E M/S and the Topock C/S
87242 Nacogdoches PSA
99015 Re: EOTT v. Eunice Patricia Wales, et al.
```

____/2 **date > 2002/07/12** (0 if not empty)

[No results]

____/3 **date> 2001/07/31 date <=2001/08/01** (-0.5 for every extra/missing line)

```
9269 Re: EOTT Crude Oil Tanks-Rev 1
9270 EOTT PL Tks 8-01.xls
9271 Phone Conversation with TNC
```

V. Complex search ____/12

____/3 **subj:shortage body: excess% to: alan.comnes@enron.com** (-1 for every extra/missing line)

```
13945 Power shortage shifts attitudes --Davis Plummets in Polls
22751 CA Leg Analyst Forecasts Significant Power Shortage in CA
65535 CA Leg Analyst Forecasts Significant Power Shortage in CA
```

____/4 **to: jdasovic@enron.com date > 2000/11/08 ecompan%** (-1 for every extra/missing line)

```
18510 VentureWire, Monday, November 13, 2000
18548 VentureWire, Tuesday, November 14, 2000
19257 VentureWire, Thursday, November 30, 2000
19338 VentureWire, Friday, December 1, 2000
```

____/2 **body: 12647 cc: jeff.dasovich@enron.com bcc: dirk.vanulden@enron.com**
(-1 for every extra/missing line)

```
18135 Re: CSU Fresno, PG&E leter
20010 Re: Minutes from call with City of Fresno, Attorney
```

____/3 **date<= 2000/01/10 from: george.robinson@enron.com to: larry.campbell@enron.com**
(-1 for every extra/missing line)

```
8751 AFUDC on Remediation Projects
9493 Re: NNG Eunice Station
9711 Re: Environmental Costs for PG&E and KN
```

VI. Full output view ____/8

____/2 Ability to switch to full output by typing **output=full** and back to brief mode by typing **output=brief** (zero mark if the default output mode is not brief or the output mode changes without an explicit command)

____/2 All columns are shown in the full mode. (Deduct 0.5 marks for each missing/redundant columns)

Change the output mode to full. Type **output=full**

For each query, assign 0 if full output mode not available or if any incorrect rows are returned.

____/2 **date:2001/10/27**

[One result with **id=9384**]

_____/2 **bcc: margaret.huson@enron.com**

[Two results with id=14058 and id=41417]

Change back the output mode to brief. Type **output=brief**

VII. Search with large result sets _____/6
(use a clock to time the queries)

_____/3 **oil%**

___/2 no crash and many results

___/0.5 less than 10 seconds

___/0.5 less than 5 seconds

_____/3 **th% date>2001/08/01**

___/2 no crash and many results

___/0.5 less than 10 seconds

___/0.5 less than 5 seconds

VIII. Others _____ (up to 4 deductions)

_____ The system terminates only if the user chooses to do so (-1 mark for each inappropriate termination up to -2 marks).

_____ Record ID or subject is missing in brief mode (-1 mark for each time up to -2 marks)

I. Documentation and source code quality ____/15

____/4 General overview of the system and sample usage of the system. The general overview of the system gives a high level introduction and may include a diagram showing the flow of data between different components; this can be useful for both users and developers of your application. The user guide should have instructions for running your code for phases 1, 2 and 3.

____/3 A description of the test strategy, test cases and their coverage

____/4 A description of your algorithm for efficiently evaluating queries, in particular evaluating queries with multiple conditions and wild cards and range searches and an analysis of the efficiency of your algorithm.

____/4 Source code quality. Easy to read and properly commented code, i.e., use of header comments and comments for non-trivial statements.

II. Group Cooperation ____/5

____/2 Task break-down among partners

____/2 Time spent and the progress made by each member

____/1 Method of coordination

OTHER COMMENTS (This is important in case the instructor, who did not see the demos, needs to make a decision regarding a possible appeal.)