

Name: Andre Freeman

Student ID: 610921313

INSTRUCTION on how to set Environment to execute my code.

NB. Most programs will need: extra_func.py

DATA pull

0. /data_pull/ - ftp_pull.py
 - Does FTP connect and request to FTP server that holds the weblogs for NDHU CSIE Server.
 - Need ../data_pull/data/ subdirectory

Need to copy contents from ../data_pull/data/ to ../data/ subdirectory

TSP – Timedelta Selection Process

1. TimeDelta Generation – ses_ana.py -T
 - Prerequisites:
 - i. Extra_func.py – Extra universal function (filtering)
 - Need ../tgen/ subdirectory for results
2. Timedelta analysis – td_cluster.py / td_cluster.py -S → (SHOW MODE)
 - Analyze the timedelta metadata
 - First chart: showing equal bin distribution of the timedelta before redistribution
 - Second chart: showing equal bin distribution of the timedelta after redistribution
 - The results shown on screen at end of this process.
 - Need ../tgen/ subdirectory
 - o Need files generated from #1

SATP – Session Aggregation and Transformation Process

3. SATP – fread.py / fread.py -T → (TEST MODE)
 - Give ID
 - Prerequisites:
 - i. url_tree.py – UTree Class
 - ii. url_session.py – SSession Class
 - iii. extra_func.py – Extra universal function
 - _tfs: 47.10 min
 - TEST MODE
 - o used default ID as 10

- Out of 342 files 0-2 files.
- _tfs: 60 min
- Need ../data/ subdirectory with files from #0
- Need ../gen/ subdirectory for results

AAP – Analysis Application Process

4. Apriori - apri.py
 - Select ID
 - Frequent itemset mining is applied to session-oriented data.
 - Prerequisites:
 - _APvalues.py – Global constants for min_sup and Min_conf
 - N.B. Demo may use relaxed min_sup: 0.015 and min_conf: 0.85
 - Parallel compute Apriori
 - Parallel load metafiles
 - Parallel Apriori compute
 - Need ../gen/ subdirectory and data from #3
 - Need ../a_res/ subdirectory for results
5. VLMC – vlmc.py
 - Select ID
 - N.B. DEMO version may have relaxed k is set as 0.
 - Parallel compute VLMC
 - Parallel load/read metafiles
 - Parallel VLMC compute and prune
 - Need ../gen/ subdirectory and data from #3
 - Need ../v_res/ subdirectory for results

AREA – Analysis Results Evaluation and Application

6. Result Evaluation – cros.py
 - Select ID
 - Press 2 for Computerized look up method
 - Need ../gen/ subdirectory and data from #3
 - Need ../a_res/ subdirectory and data from #4
 - Need ../comp/cros/ subdirectory for results
 - Prerequisites:
 - i. Extra_func.py
 - ii. _APvalues.py

The following don't write to file.

7. First and Last page – se_pages.py

- Select ID
- Select number for mode Targeted:
 - i. Enter target page number
- Show overall first and last pages to screen
 - i. Need ../gen/ subdirectory
 - ii. Need ../comp/cros/ subdirectory and data from #6

8. LDA – summary.py

- Select ID
- Show Topic probability on screen
- Show top 10 pages for English, Chinese and Both on screen
- Need ../gen/ subdirectory
- Need ../comp/cros/ subdirectory and data from #6
-