CSE/CPEG Final Year Project/Thesis  
  
Monthly-Report Submission Guideline

**Objective:** To track the progress of the Final Year Project/Thesis (FYP/FYT).

* There are **5** monthly reports which account for **5%** of the final grade.
* There are 4 monthly reports due in the Fall, i.e. Sept., Oct., Nov. and Dec. We recommend the meetings take place before the proposal report is due in Sept. and middle of the month in Oct., Nov. and Dec.
* The last monthly report is due in Feb., before the progress report is due.
* It is the responsibility of the student to submit the completed report to the FYPMS.

**Monthly Report for CSE FYP/FYT**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Project Code: | DL2 | Supervisor(s): | | Professor Dik LEE | | |
| Project Title: | Predicting trends of Interests in Twitter | | | | | |
| Group Member(s) and Student ID(s): | Luk Wing San (20193803)  Tsang Chiu Sing (20133126)  Kwok Chun Ho Andy (20272350) | | | | | |
| Reporting Period:   * Scan report and submit via the FYPMS | Report #1  Sept.  Report #2  Oct.  Report #3  Nov.  Report #4  Dec.  Report #5  Feb | | | | | |
| Progress:   * List the work completed in this reporting period. * Identify the major difficulties encountered. * Comment on the overall progress. | * Included a new data quality analysis before any data preprocessing/ analysis takes place to eliminate data that do not contribute much to general trend prediction (e.g. due to spelling mistake, not-commonly-used vocabulary) * Updated our data preprocessing method to include some extra information that would be useful for prediction (e.g. follower count, like count, retweet count) * Filtered out only the 538 most frequently used hashtags for our algorithms training input. Since the counts of hashtags drop exponentially, we have to select hashtags that have enough counts to be analyzed * One difficulty we are facing is that the data is extracted as a sample of the database, not the whole information contained in the database. Therefore, some of the hashtags may not be that representative. The trend of those hashtags may not give much insights into future trend. We need a better methodology to remove these data. | | | | | |
| Future Plan:   * Write down the working plan | * We are going to develop a better data filtering rule to remove less representative data, making sure the input data to training algorithms would be general enough to represent a large set of behavior. * We observed that there are hashtags that share a similar meaning and even behaviors (e.g. Football, WorldCup). We are planning to develop a method to group similar hashtags to improve the accuracy of prediction. | | | | | |
| Supervisor’s  Comments: |  | | | | | |
| Supervisor’s Overall Evaluation:  F  D  C-  C  C+  B-  B  B+  A-  A  A+ | (please circle) | | | | | (letter grade) |
| Meeting  Date & Time: | 4th November, 2017 | | | | | |
| Group Representative’s Signature: |  | | Supervisor’s Signature: | |  | |

(Version 2015-09-17)