Terrence Gausi
Hangbo Gu
Christian Lopez

Cmpe 130 project abstract

Web data mining and analytics

Abstract

- **Introduction:**

Data mining and analytics are by definition searching and generating new data based on systematic computational analysis. The team will gather content from web based sources and analyze it to generate basic deductions and will enable the team to predict what the future contents of the text will be. In order to implement the Data mining aspect of the project the team will utilize arrays with all the targeted webpages and to create filters to gather information about the data; furthermore the team will use a binary search tree to find the most common words in a text file and will sort the words based on the most recurring.

- **Proposed procedures:**

Extraction of raw data from HTML files:
The team will extract raw text data and parse it so that only the targeted text is left. Furthermore, the team will create a database to store the data to prepare for subsequent processes.

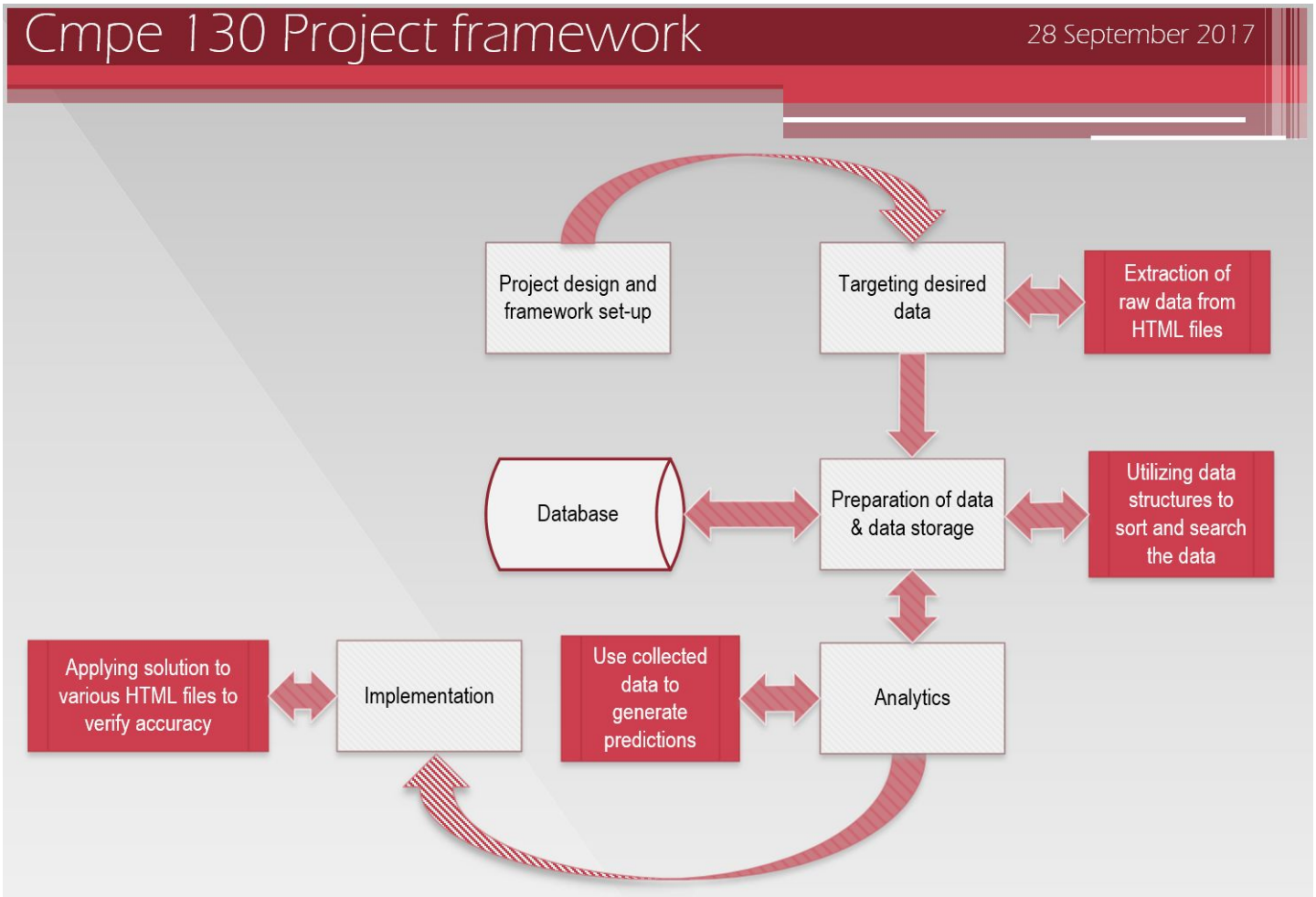
- **Data processing:**

During this stage the team will categorize the formatted text and give a quantifiable value to every word that was extracted.

- **Analytics:**

The most critical stage of the project, it will consist of analyzing the data and recognizing the patterns that emerge from the data processing stage. It is at this stage that the time will objectively evaluate the patterns found and make predictions based on statistical models.

Terrence Gausi
Hangbo Gu
Christian Lopez



Week of September 19

Finalizing project abstract and submitting

Week of September 26

All member have agreed on IDE and research for appropriate algorithms begins. Team member will meet to discuss specifics of the project (e.g. target data, data hosting, and tentative deadlines)

Week of October 10

Targeted data extracted from HTML files

Terrence Gausi
Hangbo Gu
Christian Lopez

Week of October 24

Preparation of data (Data storage in structures)

Begin analyzing data

Week of November 14

Results generated from analyzation stage

Week of November 30th

Project report finalized and turned in

December 5

Project presented and finalized