Project Name: Clusty (Ultimate News Search Engine)

Team Members: Swati Shah, Karan Sachan

Project Proposal: News Clustering and representation Web App.

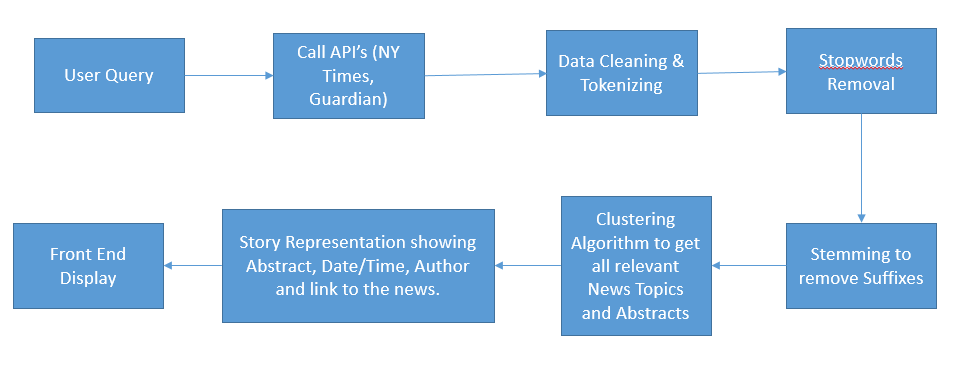
In current scenario, News Absorption is inefficient and time consuming. Its hard to keep up with the latest relevant news in our day to day busy schedule. We are providing a solution which will cluster the news on-demand for the topics that matters the most to you. It is basically a refined search engine which would present the news which is easy to comprehend and digest.

Now you need not to worry about keeping updates for the topics you would to follow on go.

Detail Description:

Our website will be used as a Search Engine for gathering relevant news as per user request. It plans to fetch the news articles from NY Times and Guardian and will represent the news in chronological order. The news will further be categorized and will be assigned a color coding. The category of the news articles is inferred from its abstract and each separate category is assigned a different color encoding to differentiate the genre of the news. Both Clustering and Classification algorithms used are unsupervised. Our News search Engine results in Story Representation and Topic Summarization given a user query.

Below is the architecture of our Project:



Modules:

1. Data Visulatization
2. Data Mining
3. Machine Learning – Clustering Algorithm
4. Crawling-Search Engine.

Technology Stack:

1. HTML
2. NODEJS
3. JAVASCRIPT
4. CSS

Risks:

1. Creating interactive user interface which shows the all the news coherently.
2. Getting News Abstract from the API’s needs some algorithm to fine tune it.
3. Accuracy of the Search is still not known and there is no way we could measure it.
4. Lot of dependency on NYT and Guardian API’s.

References:

<http://xpo6.com/list-of-english-stop-words/>

https://tartarus.org/martin/PorterStemmer/js.txt

https://github.com/awaisathar/lda.js