Code Artifacts for the Project Clusty

1. Stopword Removal Module: We used below reference for our choosing our stopwords and created a file stopwords.js which will be used to remove the stopwords

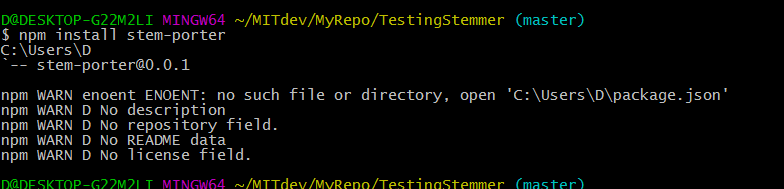
Reference Link: <http://xpo6.com/list-of-english-stop-words/>

File: stopwords.js

Attachment: 

Snapshot: 

1. Stemmer Removal : We thought of using the npm package for stemming algorithm but it gave error while installing. Below is the screenshot



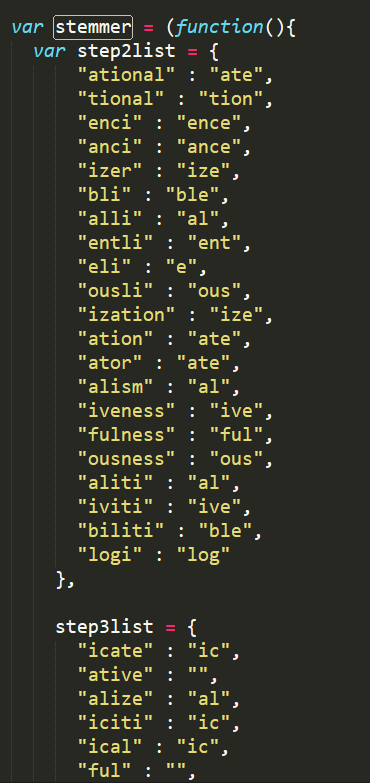
Finally we choose to use the Porters Stemming Algorithm written in Javascript.

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

File: PorterStemmer.js

Attachment: 

Snapshot:



1. APIs: We are using below links for the APIs of NY Times and Guardian

[http://api.nytimes.com/svc/search/v2/articlesearch.json?q=<Search](http://api.nytimes.com/svc/search/v2/articlesearch.json?q=%3cSearch) String>&sort=newest&api-key=3a9948f7410b6bd9138716c33ff4ed75:1:75166459

[http://content.guardianapis.com/search?q=<Search](http://content.guardianapis.com/search?q=%3cSearch) String>&order-by=newest&api-key=neukrcw8u9xm4ks5zejvx3uj&show-tags=contributor

1. Open Source Implementation of Clustering Algorithm:

Looking at numerous clustering implementation algorithm suiting our requirements

* Dirichlet Process Clustering.
* K-Means Clustering.
* Canopy Clustering.
* Fuzzy K-Means.
* Expectation Maximization.
* Mean Shift Clustering.
* Hierarchical Clustering.
* Latent Dirichlet Allocation.