Steps:

**Preprocessing**

**Cleaning**

**Stopwords**

**Stemming** - Dictionary/ Linguistic Rules/

Lemmatization

Porter's Algorithm

Snowball Algorithm

**Exploring**

Document Term Matrix – Histogram

**Synonym Correction**

PyDictionary

NLTK

WordNet

ElasticSearch

**Model Tested**

K-Means clustering Using Tf-Idf Scoring

LDA

LSI/LSA

Word Cloud

Links

http://brandonrose.org/clustering

Resource Library

R

Text Mining - tm package

Web Scraping - rvest, twitteR

wordcloud

deeplearning - text2vec, syuzhet, topicmodels

Python

Text Mining - text mining

Nltk

Wordnet

Elasticsearch

Spelling Correction: <http://blog.notdot.net/2010/07/Damn-Cool-Algorithms-Levenshtein-Automata>

<http://blog.notdot.net/2007/4/Damn-Cool-Algorithms-Part-1-BK-Trees>

<http://norvig.com/spell-correct.html>

Twitter - tweepy

Web Scraping - scrapy, urllib2

word2vec, LDA - gensim, GloVe, theano

flipboard.com

Story Clustering by LDA

<http://engineering.flipboard.com/2017/02/storyclustering>

Bag of Words

<http://scikit-learn.org/stable/auto_examples/text/mlcomp_sparse_document_classification.html#sphx-glr-auto-examples-text-mlcomp-sparse-document-classification-py>

Sparse features bag of words

<http://scikit-learn.org/stable/auto_examples/text/document_classification_20newsgroups.html#sphx-glr-auto-examples-text-document-classification-20newsgroups-py>

Kmeans

<http://scikit-learn.org/stable/auto_examples/text/document_clustering.html#sphx-glr-auto-examples-text-document-clustering-py>

Spectral co clustering

<http://scikit-learn.org/stable/auto_examples/bicluster/bicluster_newsgroups.html#sphx-glr-auto-examples-bicluster-bicluster-newsgroups-py>

LDA

<http://scikit-learn.org/stable/auto_examples/applications/topics_extraction_with_nmf_lda.html#sphx-glr-auto-examples-applications-topics-extraction-with-nmf-lda-py>

<http://brandonrose.org/clustering>

Visualization

Python

Link Python and D3

matplotlib

NetworkX

plotly

R

Spelling Correction:

<http://www.sumsar.net/blog/2014/12/peter-norvigs-spell-checker-in-two-lines-of-r/>

ggplot2

igraph

plotly

networkD3