Stopword indonesia

List stopword indonesia https://github.com/masdevid/ID-Stopwords/blob/master/id.stopwords.02.01.2016.txt

Stemmer indonesia

Bisa pek package python sastrawi di sini <https://pysastrawi-demo.appspot.com/>

data hasil crawler --> filtering (regex) --> bikin tiap dok  
jadi vektor,  
kemudian datanya di clustering (k means)  
simpen centroid tiap cluster  
  
  
Spesifikasi IR  
-VSM, pembobotan selain tf idf, rangkingnya pake jarak cosine  
sort descending  
  
sistem kita (web)  
localhost, pake angular.  
  
sistematik hubungan IR dengan web nya pake rest json