Discussion of query refinement/relevance feedback techniques used

1. Query Refinement – Wordnet

* Nltk’s wordnet returns a list of Sysnets for a given word. We used wordnet to expand a given query before doing a VSM on the new query vector.
* So given a query, wordnet will return us some synonyms of the query. We then add these synonyms to the query vector and do VSM. For example, if the query given is ‘motorcar’, wordnet will return us ['car', 'auto', 'automobile', 'machine', 'motorcar']. This list then becomes our new query vector.
* On immediate effect of doing this is more documents being returned as our output.
  + With more documents, our precision drops because many documents returned does not contain any of the original queries
* Sometimes wordnet gives us synonyms that are irrelevant of what we are trying to find. For example, ‘good’ can mean both ‘superb’ or ‘commodity’, and these are words with very different meaning. Hence without being clear of what the user is trying to find, wordnet tends to result in too many irrelevant documents.
* There is a slight increase in the runtime taken when using wordnet
* We do not use wordnet to increase the constraint when trying to find a strict ‘AND’ query (step 1.b.2 in our README). E.g. ‘motorcar AND ocean’ will not become ‘motorcar AND car AND ocean’ because we will miss out on potential documents.
* Wordnet is turned off because of the many irrelevant documents that it returns.