Query Enrichment Approach- Pseudo relevance feedback

Pseudo relevance feedback (PRF) is a type of technique where the user is not prompted to identify the relevant documents; the system simply assumes that the top ranked documents are relevant. The expansion terms generated by pseudo-relevance feedback will depend on the whole query, since they are extracted from documents ranked highly for that query, but the quality of the expansion will be determined by how many of the top-ranked documents in the initial ranking are in fact relevant.

Reasons for the choice of query enrichment:

* It doesn’t require user intervention.
* The words that occur more frequently in the relevant documents are used to expand the query which will result in getting better results.
* It is found to increase the recall and precision values.

Algorithm/Approach used:

* Pseudo relevance model was run on the BM25 scores for the queries. This retrieval model was giving the best recall and precision values and hence this was considered.
* Initially a query Q was considered and the top 5 documents obtained.
* The most commonly occurring words in those documents were considered after removing the stop words provided in the common\_words file.
* Different number of top k documents and top n frequently occurring combinations were considered and evaluated.

According to ‘Magdy W. and G. J. F. Jones. A Study on Query Expansion Methods for Patent Retrieval PAIR 2011 - CIKM 2011’ paper. It was stated that the best evaluations scores were found when 5 most frequently considered words were considered from top 5 documents

Trial 1:

Criteria- Three most frequently occurring words in top 5 documents were considered.

* MAP: 0.358
* MRR: 0.572

Trial 2:

Criteria: Five most frequently occurring words in top 5 documents were considered.

* MAP: 0.424
* MRR: 0.673

The evaluation clearly backed the claim made in the paper and hence we considered the second trial’s criteria in the project

<http://doras.dcu.ie/16517/1/A_Study_on_Query_Expansion_Methods_for_Patent_Retrieval.pdf>

Snippet Generation Approach