

For Do query expansion based on Indri Pseudo Relevant Documents:

ExtractFromTrec: Extract the pseudo relevant documents from microblogTrack2011.

input file: microblogTrack2011 7 folders

result_file which is the indri result of the original queries

output file: rel/ folder which is top 10 pseudo relevant documents from indri extracted from microblogTrack2011.

PRF: get the expansion terms based on the tf-idf score.

input file: rel/ folder which is the output from ExtractFromTrec.

queries.txt

relMap/ folder which is the inverted index 2.

output file: expandedQueries.txt

GetQueryParam: convert expandedQueries.txt to param.xml format.

input file: expandedQueries.txt from PRF.

output file: expandedQuery_param.xml

FilreqOnXML: add filter to filter those tweet date after the query date.

input file: expandedQuery_param.xml

queries.txt

output file: expandedQuery_param_before10.xml

(Because they are the separate projects name on my Eclipse, they are not in the package. But it could be viewed as in different package.)

The flow of programs is as below:

ExtractFromTrec => PRF => GetQueryParam => FilreqOnXML

The main function of PRF is in the Test2.java.

The structure of classes in PRF is as below:

