Dataset:

It is provided in VU. Please download it and use it.

Assignment:

1. Use Lucene ranking algorithm for finding top 100 documents for each query
2. Implement TF-IDF ranking in Lucene and find top 100 documents.
3. Think about two more variations of TF\*IDF, try to implement them and test them

# How to do:

* For the qrel provided in VU
  + Create an appropriate format that is readable by trec\_eval
* Run queries over dataset using 4 algorithms (10 points each)
* *Output:* for each topic, return up to 100 documents, one result per line, tab-delimited, in following format  
         *topic\_id*  \t *Q0* \t *document\_id* \t *rank* \t *score* \t *your\_login*  
  e.g.,   
         *201*  \t *Q0* \t *FBIS-41571* \t *143* \t *101.24* \t *eugene*
* Use trec\_eval to find measures such as MAP and P@5 and the other measures (30 points)
* A report (3 pages), your steps of work, your variations of TF\*IDF and any justifications and the results of comparison (20 points)