**INFORMATION RETRIEVAL \_ ASSIGNMENT 5**

**Content**: Evaluation of an Information Retrieval System (cont.)

**1. PREPARING**

Go to <http://trec.nist.gov/data/t9_filtering.html> to download **filtering.tar.gz** file and extract it.

*This dataset is a set of 348,566 references from MEDLINE, the on-line medical information database, consisting of titles and/or abstracts from 270 medical journals over a five-year period (1987-1991).*

**2. INDEXING**

2.1. Read README to understand the meaning and format of extracted files

2.2. Go to directory ./ trec9-train/

2.3. ohsumed.87 is the file containing all documents. Perform indexing of documents in this file

**3. EVALUATION**

3.1. Go to directory ./ trec9-train/. We have:

* *query.\* files contain queries*
* *qrels.\* files contain groundtruth for the provided queries*

3.2. Perform querying and use provided groundtruth to evaluate the system by the following measures:

* Precision/Recall Curve (draw this curve)
* R-Precision
* F-Measure
* MAP