**INFORMATION RETRIEVAL \_ ASSIGNMENT 4**

**Content**: Evaluation of an Information Retrieval System.

**1. PREPARING**

1.1. Go to <https://archive.ics.uci.edu/ml/machine-learning-databases/20newsgroups-mld/> to download **20\_newsgroups.tar.gz** file and extract it.

1.2. Merge all articles into one collection.

**2. INDEXING**

2.1. Perform building index for the text collection using learned techniques.

**3. EVALUATION**

3.1. Build a set of queries of your own with the following conditions:

* Each query belongs to only one class of the original dataset
* There is not any class belonging to no query.

3.2. Perform retrieval using that set of query, output the performance by:

* Render the graph of Precision/Recall curve.
* Print out other measures: F-measure, MAP

The **Relevance** of each retrieved document is determined by this rule:

* 1 if that document contains at least one words of the query AND that document belongs to the same class/category as the query
* 0 otherwise