CAB431 Tutorial (Week 7): Evaluation

****

A is set of relevant documents (e.g., a benchmark)**,**

B is set of retrieved documents (e.g., the outputs of an IR model)

	Relevant	Non-Relevant
Retrieved	$A \cap B$	$\overline{A} \cap B$
Not Retrieved	$A \cap \overline{B}$	$\overline{A} \cap \overline{B}$

$$Recall = \frac{|A \cap B|}{|A|}$$

$$Precision = \frac{|A \cap B|}{|B|}$$

F-Measure (F1),

$$F = \frac{1}{\frac{1}{2}(\frac{1}{R}+\frac{1}{P})} = \frac{2RP}{(R+P)}$$

where *R* is Recall, *P* is Precision.

For a given ranking, at each position *n*,

$$Recall@n = \frac{|A \cap B_n|}{|A|}$$

$$Precision@n = \frac{|A \cap B_n|}{n}$$

where B_n is a set of ***top-n*** retrieved documents.

TASK 1: read a topic-doc-assignments file (e.g., topicdocassign.txt, the benchmark) and a retrieved topic-doc-assignments file (e.g., topicdocassigntest.txt, the output of an IR model); calculate three evaluation measures of Recall, Precision, and F-Measure.

- Please download two topic-doc-assignment files, open and read through. The given topic-assignment files are in format of "topic documentID Relevance judgment" using 1 to indicate relevant and 0 for non-relevant.

- Create a new class evaluator, read in two given topic-assignment files. Build up ***a set of relevant documents*** containing all relevant documents based on topic-doc-assignment file, ***a set of retrieved documents*** containing all retrieved documents based on a rank file.
- Create a method to calculate Recall, a method to calculate Precision, and a method to calculate F-measure. Display three evaluation measures.

Example of output

```
The number of relevant documents = 10;
The number of retrieved relevant documents 7
recall = 0.700000
precision =
F-Measure =
```

TASK 2: read two top-10 ranking result files (e.g., rank1.txt, and rank2.txt), calculate average Precision.

- Please download two top-10 ranking result files, open and read through. The given top-10 ranking result files are in format of "rankingNo documentId".
- Create a method to parse two ranking result file and generate a HashMap (dictionary in Python) of rankingNo:documentId pair.
- Create a method to calculate Recall and Precision at fixed rank positions i.e. 1-10, taking a set of relevant documents built in task 1 and HashMap of rankingNo:documentId as two parameters. Then calculate the average precision. Print out the results.

Example of output

```
At position 1, precision=1.000000 recall=0.062500
At position 2, precision=1.000000 recall=0.125000
At position 3, precision=1.000000 recall=0.187500
At position 4, precision=1.000000 recall=0.250000
```

```
At position 5, precision=1.000000 recall=0.312500
At position 6, precision=1.000000 recall=0.312500
At position 7, precision=0.857143 recall=0.375000
…

In top-10, average precision=0.941518


At position 1, precision=0.000000 recall=0.000000
At position 2, precision=0.500000 recall=0.062500

…

In top-10, average precision=0.714286
```