Seminar 6

Definition 1 (Recall) *Recall describes how many of the relevant documents are retrieved.*

$$recall = R = \frac{\#relevant\ retrieved}{\#relevant}$$

Definition 2 (Precision) *Precision describes how many of the retrieved documents are relevant.*

$$precision = P = \frac{\#relevant\ retrieved}{\#retrieved}$$

Definition 3 (*F*-measure) *A balanced F-measure (F_1-measure) defines a recall-precision relationship represented by their weighted harmonic mean:*

$$F = \frac{2 \cdot R \cdot P}{R + P}$$

Definition 4 (Mean Average Precision) *MAP expresses the precision in each point a new relevant document is included in the result. Is counted as*

$$MAP(Q) = \frac{1}{|Q|} \cdot \left(\sum_{q \in Q} \frac{1}{rel_q} \cdot \left(\sum_{i=1}^{rel_q} prec_i \right) \right)$$

where rel_q is the number of relevant documents retrieved by query q and $prec_i$ is the precision at the i-th document.

Definition 5 (κ statistic) *Let N be the total number of documents, J is a set of judges and $P(A) = \frac{\#agree}{N}$ the number of documents on which the judges agree. Let also define R_j and NR_j be the number of relevant and non-relevant documents, respectively, according to the judge $j \in J$ and*

$$P(R) = \frac{\sum_{j \in J} R_j}{|J| \cdot N} \quad \text{and} \quad P(NR) = \frac{\sum_{j \in J} NR_j}{|J| \cdot N}$$

as the number of relevant and non-relevant documents, respectively. Let finally define

$$P(E) = P(R)^2 + P(NR)^2$$

as the approximate number of disagreements between the judges. Then the κ statistic is defined as the measure of agreement between the judges

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}.$$

Definition 6 (Rocchio relevance feedback) *Rocchio relevance feedback has the form*

$$q_m = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d_r} \in D_r} \vec{d_r} - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d_{nr}} \in D_{nr}} \vec{d_{nr}}$$

where q_0 is the original query vector, D_r is the set of relevant documents, D_{nr} is the set of non-relevant documents and the values $\alpha,\ \beta,\ \gamma$ depend on the system setting.

Exercise 1

The following ordered list of 20 letters R and N represents relevant (R) and non-relevant (N) retrieved documents as an answer for a query on a collection of 10 000 documents. The leftmost document is expected to be the most relevant. The list contains 6 relevant documents. Assume that the collection contains 8 documents relevant to the query.
R R N N N N N N R N R N N N R N N N N R

a) What is the precision on the first 20 results?

b) What is the F-measure on the first 20 results?

c) What is the non-interpolated precision of the system at 25% recall? (R=25%)

d) What is the interpolated precision of the system at 33% recall? (R>33%)

e) Assume that these 20 documents is the complete list of retrieved documents. What is the MAP of the system?

Now assume that the system returned all 10,000 documents in an ordered list and above is the top 20.

f) What is the highest possible MAP the system can achieve?

g) What is the lowest possible MAP the system can achieve?

Exercise 2

Below is a table showing how two judges judged the relevance (0 = non-relevant, 1 = relevant) of the set of 12 documents with respect to a query. Assume that you developed an IR system, that for this query returns the documents $\{4, 5, 6, 7, 8\}$.

Doc ID	Judge 1	Judge 2
1	0	0
2	0	0
3	1	1
4	1	1
5	1	0
6	1	0
7	1	0
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1

Table 1: Judges judging the relevance of documents.

a) Calculate the κ statistic.

b) Calculate the recall, precision and F-measure of your system in which a document is considered relevant if the judges agree.

c) Calculate the recall, precision and F-measure of your system in which a document is considered relevant if at least one of the judges thinks so.

Exercise 3

A user's primary query is *cheap CDs cheap DVDs extremely cheap CDs*. The user has a look on two documents: doc1 a doc2, marking doc1 *CDs cheap software cheap CDs* as relevant and doc2 *cheap thrills DVDs* as non-relevant. Assume that we use a simple *tf* scheme without vector length normalization. What would be the restructured query vector after considering the Rocchio relevance feedback with values $\alpha = 1$, $\beta = 0.75$ and $\gamma = 0.25$?