

Homework 8

Student Number:

Name:

Problem 1. (40 points) An IR system returns 10 positive results, where 5 of them are relevant, and 5 are nonrelevant. There are a total of 1,000 documents in the collection and 20 of them are relevant.

- What is the precision of the system on this search, and what is its recall?
- What is the accuracy of this search?
- Let $F_\beta = \frac{(\beta^2+1)PR}{\beta^2P+R}$, What is the F_β of this search given $\beta = 1/2, 1, 2$? When we value more on precision than recall, which β should we use?

Problem 2. (30 points) Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result):

System 1	R	N	R	N	N	N	N	R	R
System 2	N	R	N	N	R	R	R	N	N

- What is the MAP of each system? Which has a higher MAP?
- Does this result intuitively make sense? What does it say about what is important in getting a good MAP score?
- What is the R-precision of each system? (Does it rank the systems the same as MAP?)

Problem 3. (30 points) Below is a table showing how two human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that you've written an IR system that for this query returns the set of documents [4, 5, 6, 7, 8].

docID	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

- Calculate the kappa measure between the two judges.
- Calculate precision, recall, and F1 of your system if a document is considered relevant only if the two judges agree.
- Calculate precision, recall, and F1 of your system if a document is considered relevant if either judge thinks it is relevant.