

Paper Code	Examiner	Department	Office
INT309	Fangyu Wu	Department of Intelligence Science	SD555

Question 1

Suppose we have a collection of 20 documents, d_1, d_2, \dots, d_{20} , which have been judged for relevance to a query. A 3-point relevance scale was used, so relevant documents have been divided into Perfect, Good and just Relevant results. Weights for these levels are shown below:

Perfect	3
Good	2
Relevant	1
Non-relevant	0

Here are the documents and their judgments:

$$\begin{aligned}
 \text{Perfect} &= \{d_1, d_8\} \\
 \text{Good} &= \{d_4, d_9\} \\
 \text{Relevant} &= \{d_2, d_7, d_{10}\} \\
 \text{Non-relevant} &= \{d_3, d_5, d_6, d_{11} - d_{20}\}
 \end{aligned}$$

Consider now these two ordered result lists retrieved by different systems:

$$\begin{aligned}
 \text{Result}_1 &= \langle d_1, d_2, d_7, d_3, d_4, d_8, d_6, d_5 \rangle \\
 \text{Result}_2 &= \langle d_1, d_4, d_8, d_2, d_6, d_3, d_5, d_7 \rangle
 \end{aligned}$$

1. What are the precision and recall for result list Result_2 ?

Solution: (1 point each)
 $P = 5/8; R = 5/7$

2. What is the precision @4 of each result list?

Solution: (1 point each)
 $3/4$ and 1 .

3. What is the average precision of each result list?

Solution:

(2 points for each of the result list)

$$AP_1 = \frac{1+1+1+\frac{4}{5}+\frac{5}{6}}{5} = \frac{4.633}{5} = 0.9267$$

List	AP
R_1	0.9267
R_2	0.925

4. To measure/evaluate information retrieval (IR) effectiveness, what are the three elements required for a test collection, so the performance of the IR system could be compared?

Solution:

(2 points for each of the result list)

1) A benchmark document collection;

2) A benchmark suite of information needs, expressible as queries;

3) An assessment of either Relevant or Nonrelevant judgments for each query-document pair.

Question 2

Consider following documents with the stop word list: [when, in, the, and, I]

Doc 1: when walking in the rain

Doc 2: rain stopped walk, I ran, rain stop.

Doc 3: stop walking and run

Consider the query **rain stop** on a fictitious collection with $N = 1,000$ documents where the document frequencies of walk, rain, stop and run are respectively 50, 10, 100 and 100. What is the similarity score for this query with documents Doc 1 and Doc 2?

Use logarithmic term weighting for query and maximum tf (term frequency) formula normalization for documents, idf weighting for the query only. Length normalization is not required.

The maximum tf formula for normalization is listed as below:

$$0.25 + [0.75 \times tf_{t,d} / \max(tf_{t,d})]$$

Solution: 1) vector (tf-idf) of query

term	tf	wf	df	idf	tf-idf
rain	1	1	10	2	$wf \times idf : 1 \times 2 = 2$
stop	1	1	100	1	$wf \times idf : 1 \times 1 = 1$

2) vector (wf) for Doc 1

term	tf	w-tf
rain	1	$0.25 + [0.75 \times 1/1] = 1$
stop	0	$0.25 + [0.75 \times 0/1] = 0.25$

3) vector (wf) for Doc 2

term	tf	w-tf
rain	2	$0.25 + [0.75 \times 2/2] = 1$
stop	2	$0.25 + [0.75 \times 2/2] = 1$

The similarity score for query with Doc 1: $2 \times 1 = 2$

The similarity score for query with Doc 2: $2 \times 1 + 1 \times 1 = 3$

We focus only on terms that are present in both the document and the query.

———— *End of paper* ————