

# Information Retrieval

## Computing Scores, Evaluation, Relevance Feedback

1. Consider the table of term frequencies for 3 documents Doc1, Doc2, Doc3. Compute the **tf-idf** weights for the terms car, auto, insurance, best, for each document. The size of the collection is 806,791 documents.

term	Doc1	Doc2	Doc3
car	27	4	24
auto	3	33	0
Insurance	0	33	29
best	14	0	17

term	df <sub>t</sub>
car	18,165
auto	6723
insurance	19,241
best	25,235

2. a) Compute the Euclidean normalized **document vectors** for each of the documents, where each vector has four components, one for each of the four terms  
b) With term weights as computed, rank the three documents by computed score for the query '**car insurance**', for each of the following cases of term weighting in the query:
  1. The weight of a term is 1 if present in the query, 0 otherwise.
  2. Euclidean normalized idf.
3. 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 N
  - a. What is the **MAP** of each system? Which has a higher MAP?
  - b. Does this result intuitively make sense? What does it say about what is important in getting a good MAP score?
  - c. What is the R-precision of each system? (Does it rank the systems the same as MAP?)
4. Suppose that a user's initial query is cheap CDs cheap DVDs extremely cheap CDs. The user examines two documents, d1 and d2. She judges d1, with the content CDs cheap software cheap CDs relevant and d2 with content cheap thrills DVDs non-relevant. Assume that we are using direct term frequency (with no scaling and no document frequency). There is no need to length-normalize vectors. Using **Rocchio relevance feedback**, what would the revised query vector be after relevance feedback? Assume  $\alpha = 1$ ,  $\beta = 0.75$ ,  $\gamma = 0.25$ .

## Solutions

1.

term	df <sub>t</sub>	idf <sub>t</sub>
car	18,165	1.65
auto	6723	2.08
insurance	19,241	1.62
best	25,235	1.5

terms	Doc1	Doc2	Doc3
Car	44.55	6.6	39.6
Auto	6.24	68.64	0
Insurance	0	53.46	46.98
Best	21	0	25.5

2. Doc1=(0.897,0.125,0,0.423), Doc2=(0.076,0.786,0.613,0) Doc3=(0.595,0,0.706,0.383)

(i) Term Weights

Term	Doc1	Doc2	Doc3
car	0.897	0.076	0.595
auto	0.125	0.786	0
insurance	0	0.613	0.706
best	0.423	0	0.383

Term	weight	W(t,q)	Doc1	Doc2	Doc3
car	1	1	0.897	0.076	0.595
auto	0	0	0	0	0
Insurance	1	1	0	0.613	0
best	0	0	0	0	0

Score(q,doc1)=0.897,Score(q,doc2)=0.689,Score(q,doc3)=0.595

Ranking=d1,d2, d3

(ii)

term	idf	W(t,q)
Car	1.65	0.478
Auto	2.08	0.602
Insurance	1.62	0.47
best	1.5	0.43

Score(q,doc1)=0.686, Score(q,doc2)=0.797, Score(q,doc3)=0.781

Ranking=d2,d3, d1

3. a.

MAP (System 1) =  $(1/4) * (1 + (2/3) + (3/9) + (4/10)) = 0.6$

MAP(System 2) =  $(1/4) * (1/2 + 2/5 + 3/6 + 4/7) = 0.493$

System1 has a higher average precision

b. MAP provides a single figure measure of quality across recall levels. For a good MAP score, it is essential to more relevant documents in the first few (3-5) retrieved ones.

c.  $R\text{-Precision}(\text{system 1}) = 1/2$   $R\text{-Precision}(\text{system 2}) = 1/4$  This ranks the system the same as MAP.

4.

word	q	d1	d2
CDs	2	2	0
cheap	3	2	1
DVDs	1	0	1
extremely	1	0	0
software	0	1	0
thrills	0	0	1

For  $1.0 \cdot q + 0.75 \cdot d_1 + 1 \cdot -0.25 \cdot d_2$ , we get:  $(3.5 \ 4.25 \ 0.75 \ 1 \ 0.75 \ -0.25)^T$  or  $(7/2 \ 17/4 \ 3/4 \ 1 \ 3/4 \ -1/4)^T$ . Negative weights are set to 0. The Rocchio vector thus is:  $(3.5 \ 4.25 \ 0.75 \ 1 \ 0.75 \ 0)^T$ .