

5. (10 points) Describe **two ways** to speed up the cosine algorithm on the attached page. For each of your suggested speedups, characterize the impact that your method has on the output of the algorithm.

6. (5 points) What aspect of ad hoc retrieval system performance is *relevance feedback* primarily intended to address?

7.

Lucene scoring equation

$$\sum_{t \in q} (tf(t \text{ in } d) \times idf(t)^2 \times boost(t, \text{field in } d) \times lengthNorm(t, \text{field in } d)) \times coord(q, d) \times queryNorm(q)$$

Cosine algorithm

COSINESCORE(q)

- 1 float $Scores[N] = 0$
- 2 Initialize $Length[N]$
- 3 for each query term t
- 4 do calculate $w_{t,q}$ and fetch postings list for t
- 5 for each pair($d, tf_{t,d}$) in postings list
- 6 do $Scores[d] += wf_{t,d} \times w_{t,q}$
- 7 Read the array $Length[d]$
- 8 for each d
- 9 do $Scores[d] = Scores[d] / Length[d]$
- 10 return Top K components of $Scores[]$