Week 1 Practice Quiz

TOTAL POINTS 10

1. Consider the instantiation of the vector space model where documents and queries are represented as **term frequency vectors**. Assume we have the following query and two documents:

   1 point

   Q = “future of online education”

   D1 = “Coursera is shaping the future of online education; online education is affordable.”

   D2 = “In the future, online education will dominate.”

   Let V(X) = [c1 c2 c3 c4] represent a part of the term frequency vector for document or query X, where c1, c2, c3, and c4 are the term weights corresponding to “future,” “of,” “online,” and “education,” respectively.

   Which of the following is true?

   ● V(Q) = [1 1 1 1]   V(D1) = [1 1 2 2]   V(D2) = [1 0 1 1]

   ○ V(Q) = [1 1 1 1]   V(D1) = [1 1 1 1]   V(D2) = [1 0 1 1]

   ○ V(Q) = [1 1 1 1]   V(D1) = [1 1 2 2]   V(D2) = [1 1 1 1]

2. Consider the same scenario as in Question 1, with the dot product as the similarity measure. Which of the following is true?

   1 point

   ○ Sim(Q,D1) = 6   Sim(Q,D2) = 4

   ● Sim(Q,D1) = 6   Sim(Q,D2) = 3

   ○ Sim(Q,D1) = 4   Sim(Q,D2) = 3

3. Which is NOT the reason that ranking is preferred over document selection?

   1 point

   ○ The boundary between relevant vs. not relevant documents is hard to define.

   ○ Not all relevant documents are equally relevant.

● Ranking is computationally easier than selection.

○ Users prefer to browse results in a sequential manner.

4. Which of the following application scenarios in text retrieval relies LESS on NLP?

1 point

○ Use query in English to retrieve documents in French

● Compare homework submissions from students to check if plagiarism occurs

5. What information about text is lost using "bag of words" representation?

1 point

□ Word occurrence counts

☑ Word ordering

□ Number of unique words in document

☑ Phrases formed by multiple words

□ Document length

6. Select the following applications that provide "push" information access.

1 point

☑ Netflix (netflix.com)

□ Google Maps (maps.google.com)

□ Spotify mobile (iOS/Android app)

□ Bing (bing.com)

7. Which of the following factors make text retrieval more difficult than database retrieval?

1 point

☑ Ambiguity in text

☑ Unstructured/free text query and data

☑ Subjective judgement of relevance and empirical evaluation involving humans

8. Stop words, such as "*the,*" "*is,*" "*at,*" "*which,*" and "*on*" can be identified with TF-IDF: 1 point

○ Low TF, low IDF

● High TF, low IDF

○ Low TF, high IDF

○ High TF, high IDF

9. How many syntactic structures can you identified in sentence “A man saw a boy with a telescope”? 1 point

○ 1

● 2

○ 3

10. In VSM model, which of the following similarity/distance measure would be affected by document length? 1 point

● L2 distance: $||v_1 - v_2||_2$

○ Cosine similarity: $cos(v_1, v_2)$

☑ I, **BAL KRISHNA NYAUPANE**, understand that submitting work that isn’t my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera’s Honor Code