CMSC 476/676 Sample-test-1

February 15, 2021

Choose the best answer.

1. How many distinct characters can be encoded in 16 bits?
   1. 2^7
   2. 2^8
   3. 2^32
   4. None of the above
2. Choose the answer that correctly completes this sentence. An inverted index
   1. uses the primary and secondary keys in a relational database table.
   2. uses a hash table for quick lookup of words called tokens.
   3. uses the characters in a word to compute the hash value into an array.
   4. b. and c. but not a.
3. The best weighting scheme that can be used to create a word cloud is
   1. term frequency (tf)
   2. inverse document frequency (idf)
   3. term frequency \* inverse document frequency (tf\*idf)
   4. none of the above.
4. When scraping the WWW for documents, what protocol is used?
   1. RSS
   2. SFTP
   3. HTTP
   4. All of the above.
5. Suppose there are 20 documents in a collection. dog occurs 4 times, cat occurs 5 times, and rat occurs 2 times.
   1. How many documents are estimated to contain both dog and rat?
      1. 2 (.1 of the 20)
      2. 4 (.2 of the 20)
      3. 8 (.4 of the 20)
      4. None of the above
   2. How many of documents can be estimated to contain dog, cat, and rat?
      1. 2 (.1 of the 20)
      2. 4 (.2 of the 20)
      3. 8 (.4 of the 20)
      4. None of the above
   3. How many of the documents can be estimated to contain cat?
      1. 2 (.1 of the 20)
      2. 4 (.2 of the 20)
      3. 8 (.4 of the 20)
      4. None of the above
6. Given that the total number of word occurrences in an English document collection is 50,000,000, that the constant c=.1 for English, and that the frequency of word A is 5000. (Remember: rank\*probability is approximately c, and frequency\*rank = constant k)
   1. What is the probability of A occurring?
      1. .01
      2. .1
      3. 1
      4. None of the above
   2. What is the rank of word A?
      1. 10
      2. 100
      3. 1000
      4. None of the above

For the questions below, use the following:

= Relevant documents for Query 1

= Relevant documents for Query 2

Query 1

Ranking

Query 2

Ranking

1. What is the Average Precision for Query 1 Ranking?
   1. .62
   2. .72
   3. .52
   4. None of the above.
2. What is the Average Precision for Query 2 Ranking?
   1. .5
   2. .3
   3. .4
   4. None of the above
3. What is the Mean Average Precision for Query 1 and 2?
   1. .48
   2. .51
   3. .67
   4. None of the above