

Quiz 1 - Answers

1. tf-idf Weights: Which of the following is **true**:
 - a. the tf-idf weight is the ratio between tf and idf
 - b. the tf-idf weight is the sum of tf and idf
 - c. frequent terms obtain high tf score**
 - d. frequent terms obtain low tf score
2. Query Evaluation: In a ranked retrieval result, the result at position $k+1$ is not relevant. Which of the following is true ($P@k$ and $R@k$ are the precision and recall of the result set consisting of the k top ranked documents)
 - a. $P@k < P@k+1$
 - b. $R@k < R@k+1$
 - c. $R@k = R@k+1$**
 - d. $P@k = P@k+1$
3. Term-Document Matrix: If we sum up all values in a row of the term-document matrix, assuming all values are 1 and 1 indicates the presence of the term in the document, we obtain
 - a. The inverse document frequency of a term
 - b. The number of documents containing a term**
 - c. The maximal frequency of a term in the document collection
 - d. The number of terms in a document
4. Retrieval models: Which of the following statements is correct:
 - a. A Boolean retrieval model will never return any document that does not at least contain one term occurring in the query.
 - b. A Vector Space retrieval model will give non-zero similarity only if a document does at least contain one term occurring in the query. Corrected during the exercise session**
 - c. The Probabilistic retrieval model **without** smoothing will give non-zero similarity only if a document does at least contain one term occurring in the query.
 - d. In Boolean retrieval, adding a term to a document that is already contained in the document will change the result.
5. Inverted Files: Which components of an inverted index are growing as n^{β} ($0.4 < \beta < 0.6$) with the size of the document collection n .
 - a. The postings file and the document file.
 - b. The access structure and the index file.**
 - c. The index file and the document file.
 - d. The access structure and the postings file.

6. Inverted Files: Construction of partial indices and index merging for inverted file construction is used because
- a. The vocabulary cannot fit into the main memory
 - b. The whole document collection needs to be loaded into the main memory, but does not fit
 - c. The total aggregate size of posting lists exceeds the main memory capacity**
 - d. The method is always more efficient than constructing the complete inverted file in one pass
7. Query Feedback: The SMART algorithm for query relevance feedback modifies
- a. The original query weight vectors**
 - b. The document weight vectors
 - c. The similarity function computation
 - d. The keywords of the original user query
8. Probabilistic Retrieval: We are using a probabilistic retrieval model without smoothing ($\lambda = 0$). For document $d = \text{"information retrieval"}$ with language model M_d , which of the following is correct?
- a. $P(\text{information retrieval} \mid M_d) = 0$
 - b. $P(\text{information retrieval} \mid M_d) = 0.25$**
 - c. $P(\text{information retrieval} \mid M_d) = 0.5$
 - d. $P(\text{information retrieval} \mid M_d) = 1$