Started on	Thursday, 22 March 2018, 10:46 AM
State	Finished
Completed on	Thursday, 22 March 2018, 11:00 AM
Time taken	13 mins 46 secs
Grade	6.00 out of 8.00 (**75**%)

Question 1
Correct
Mark 1.00 out of 1.00
Flag question

Tugrulcan wanted to plan his next summer vacation so he wrote “best beaches” to his favourite search engine. Little did he know, his favourite search engine was using pseudo-relevance feedback and the top-k documents that are considered relevant were about the beaches only in Turkey. What is this phenomenon called?

Select one:
- a. Query Drift ✓
- b. Query Confounding
- c. Query Malfunction
- d. Query Bias

The correct answer is: Query Drift

Question 2
Incorrect
Mark 0.00 out of 1.00
Flag question

Which of the following is **wrong** about inverted files?

Select one:
- a. The index file has space requirement of $O(n^{\beta})$, where β is about ½
- b. Storing differences among word addresses reduces the size of the postings file
- c. Variable length compression is used to reduce the size of the index file ✓
- d. The space requirement for the postings file is O(n)

The correct answer is: Storing differences among word addresses reduces the size of the postings file

Question 3
Incorrect
Mark 0.00 out of 1.00
Flag question

The SMART algorithm for query relevance feedback modifies...

Select one:
- a. The original document weight vectors
- b. The keywords of the original user query ✗
- c. The result document weight vectors
- d. The original query weight vectors

The correct answer is: The original query weight vectors

Question 4
Correct
Mark 1.00 out of 1.00
Flag question

Which of the following statements about index merging (when constructing inverted files) is **correct**?

Select one:
- a. While merging two partial indices on disk, the inverted lists of a term are concatenated without sorting ✓because they're already sorted
- b. While merging two partial indices on disk, the vocabularies are concatenated without sorting
- c. The size of the final merged index file is $O(n\log_2(n)*M)$, where M is the size of the available memory
- d. Index merging is used when the vocabulary does no longer fit into the main memory

The correct answer is: While merging two partial indices on disk, the inverted lists of a term are concatenated without sorting

Question 5
Correct
Mark 1.00 out of 1.00
Flag question

What is the benefit of LDA over LSI?

Select one:
- a. LSI is based on a model of how documents are generated, whereas LDA is not
- b. LDA represents semantic dimensions (topics, concepts) as weighted combinations of terms, whereas LSI does not
- c. LDA has better theoretical explanation, and its empirical results are in general better than LSI’s ✓
- d. LSI is sensitive to the ordering of the words in a document, whereas LDA is not

The correct answer is: LDA has better theoretical explanation, and its empirical results are in general better than LSI’s

Question 6

Correct

Mark 1.00 out of 1.00

Flag question

In general, what is **true** regarding Fagin's algorithm?

Select one:

- a. It provably returns the k documents with the largest aggregate scores ✓
- b. It performs a complete scan over the posting files
- c. It never reads more than $(kn)^{½}$ entries from a posting list
- d. Posting files need to be indexed by the TF-IDF weights

The correct answer is: It provably returns the k documents with the largest aggregate scores

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

How does LSI querying work?

Select one:

- a. The query vector is transformed by Matrix S; then cosine similarity is computed
- b. The query vector is treated as an additional document; then cosine similarity is computed ✓
- c. The query vector is multiplied with an orthonormal matrix; then cosine similarity is computed
- d. The query vector is treated as an additional term; then cosine similarity is computed

The correct answer is: The query vector is treated as an additional document; then cosine similarity is computed

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

Which of the following is **true** when comparing Vector Space Model (VSM) and Probabilistic Language Model (PLM)?

Select one:

- a. Both VSM and PLM are based on a generative language model
- b. Both VSM and PLM use collection frequency in the model
- c. Both VSM and PLM require parameter tuning
- d. Both VSM and PLM take into account multiple term occurrences ✓

The correct answer is: Both VSM and PLM take into account multiple term occurrences

Finish review (https://moodle.epfl.ch/mod/quiz/view.php?id=978942)