Français (https://moodle.epfl.ch/mod/quiz/review.php?attempt=23641&lang=fr)
English (https://moodle.epfl.ch/mod/quiz/review.php?attempt=23641&lang=en)

Started on	Thursday, 12 April 2018, 10:47
State	Finished
Completed on	Thursday, 12 April 2018, 11:03
Time taken	16 mins 21 secs
Grade	**6.00** out of 8.00 (**75**%)

Question 1

Incorrect

Mark 0.00 out of 1.00

Flag question

Given the 2-itemsets {1,2}, {1,5}, {2,5}, {1,4}, {1,3}, when generating the 3-itemsets we will:

Select one:

- a. Generate **4** 3-itemsets after the join and **2** 3-itemsets after the prune
- b. Generate **5** 3-itemsets after the join and **2** 3-itemsets after the prune ✗
- c. Generate **6** 3-itemsets after the join and **1** 3-itemsets after the prune
- d. Generate **4** 3-itemsets after the join and **1** 3-itemsets after the prune

The correct answer is: Generate **6** 3-itemsets after the join and **1** 3-itemsets after the prune

Question 2

Incorrect

Mark 0.00 out of 1.00

Flag question

When representing the adjacency list of a Web page in a connectivity server by using a reference list from another Web page, the reference list is searched only in a neighboring window of the Web page's URL, because:

Select one:

- a. Typically many URLs in a web page are similar to each other
- b. Subsequent URLs in an adjacency list have typically small differences ✗
- c. Most extra nodes are found in the neighboring window
- d. Often many URLs among two pages with similar URL are similar

The correct answer is: Often many URLs among two pages with similar URL are similar

Question 3

Correct

When computing PageRank iteratively, the computation ends when:

Mark 1.00 out of 1.00

Flag question

Select one:

- a. The norm of the difference of rank vectors of two subsequent iterations falls below a predefined threshold ✓
- b. The probability of visiting an unseen node falls below a predefined threshold
- c. All nodes of the graph have been visited at least once
- d. The difference among the eigenvalues of two subsequent iterations falls below a predefined threshold

The correct answer is: The norm of the difference of rank vectors of two subsequent iterations falls below a predefined threshold

Question 4

Correct

Mark 1.00 out of 1.00

Flag question

When constructing a word embedding, negative samples are:

Select one:

- a. All less frequent words that do not occur in the context of a given word
- b. Word - context word combinations that are not occurring in the document collection ✓
- c. Context words that are not part of the vocabulary of the document collection
- d. Only words that never appear as context word

The correct answer is: Word - context word combinations that are not occurring in the document collection

Question 5

Correct

Mark 1.00 out of 1.00

Flag question

Which of the following statements on Latent Semantic Indexing (LSI) and Word Embeddings (WE) is **correct**?

Select one:

- a. LSI does take into account the frequency of words in the documents, whereas WE does not
- b. The dimensions of LSI can be interpreted as concepts, whereas those of WE cannot
- c. LSI does not take into account the order of words in the document, whereas WE does ✓
- d. LSI is deterministic (given the dimension), whereas WE is not

The correct answer is: LSI is deterministic (given the dimension), whereas WE is not

Question 6

Correct

Mark 1.00 out of 1.00

Flag question

For his awesome research, Tugrulcan is going to use the PageRank with teleportation and HITS algorithm, not on a network of webpages but on the retweet network of Twitter! The retweet network is a directed graph, where nodes are users and an edge going out from a user A and to a user B means that "User A retweeted User B". Which one is **false** about a Twitter bot that retweeted other users frequently but got never retweeted by other users or by itself?

Select one:

- a. It will have a non-zero hub value
- b. It will have a PageRank of zero ✓
- c. Its authority value will be equal to the hub value of a user who never retweets other users
- d. It will have an authority value of zero

The correct answer is: It will have a PageRank of zero

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

Given the following list of transactions: {apple,milk}, {milk, bread}, {apple, bread, milk}, {bread}, which of the following is **correct**?

Select one:

- a. apple → milk has support ½ and confidence 1 ✓
- b. milk → bread has support ½ and confidence 1
- c. milk → apple has support ½ and confidence 1
- d. bread → milk has support ½ and confidence 1

The correct answer is: apple → milk has support ½ and confidence 1

Question 8

Correct

Mark 1.00 out of 1.00

Given the following teleporting matrix (E) for nodes A, B and C:

[0 ½ 0]
[0 0 0]
[0 ½ 1]

Flag question

and making no assumptions about the link matrix (R), which of the following is **correct**:

(Reminder: columns are the probabilities to leave the respective node.)

Select one:

- a. A random walker can never reach node A
- b. A random walker can never leave node A
- c. A random walker can always leave node C
- d. A random walker can always leave node B ✓

The correct answer is: A random walker can always leave node B

Finish review (https://moodle.epfl.ch/mod/quiz/view.php?id=981144)