Week 5 Quiz

1
point

1.
Which of the following is **not** true about GFS?

○ The GFS keeps multiple replicas of the same file chunk.

● The file data transfer happens directly between the GFS client and the GFS master.

○ The file data transfer happens directly between the GFS client and the GFS chunkservers.

1
point

2.
MapReduce allows parallelizing the creation of the inverted index.

○ False

● True

1
point

3.
In MapReduce, the Reduce function is called for each unique key of the output key-value pairs from the Map function.

○ False

◉ True

4.

Which of the following would cause a web page P to have a higher PageRank score?

◉ Add to another page Q a link that points to page P

○ Add to page P a link that points to another page Q

5.

Imagine if the web is fully connected with N pages such that for any pair of pages, P and Q, there exists a link from P to Q, then which of the following is true?

○ At least one page will have a PageRank score larger than 1/N.

◉ All the pages will have a PageRank score of 1/N.

○ At least one page will have a PageRank score smaller than 1/N.

6.

Which of the following is/are the difference(s) between pseudo and implicit feedback?

☑ Pseudo feedback assumes top ranked documents are relevant.

- [x] Implicit feedback assumes user-clicked documents are relevant.

- [ ] Both methods do not involve user activity.

1
point

7.

What is true about Rocchio feedback?

- (•) It can be used for relevance feedback and also pseudo feedback.

- ( ) Negative examples are as important as positive examples.

- ( ) It works the best if you discard original query weights during feedback.

- ( ) All words are used (no truncation) so that performance and efficiency can be guaranteed.

1
point

8.

The advantages of incremental crawling are:

- [x] It learns from past experience.

- [ ] It adds overhead.

- [x] It targets frequently accessed pages.

- [x] It targets frequently updated pages.

9.

What is NOT the reason that PageRank works better than "citation counting"?

- ( ) PageRank considers "indirect citations."

- ( ) PageRank has smoothing.

- (●) It utilizes text information other than link information.

10.

What is WRONG about HITS algorithm?

- ( ) It has two sets of parameters: a (authority) and h (hub).

- (●) The adjacency matrix is symmetric.

- ( ) It's an iterative algorithm.