CS 445 Homework

Structed Data on the Web
http://dl.acm.org/citation.cfm?id=1897816.1897839

MapReduce
http://dl.acm.org/citation.cfm?doid=1327452.1327492

MapReduceExamples
http://code.google.com/edu/parallel/mapreduce-tutorial.html#MRExamples

Write up answers to the following sets of numbered questions.
Share your answers in a GoogleDoc with profchadd@gmail.com by 1pm on Nov 28.

Be prepared in class on Nov 28 to discuss the questions below.

0.   Write definitions for any terms you run across and are not familiar with.

1.   What are key/value pairs?

2.   What does Map do?

3.   What does Reduce do?

4.   Explain, in your own words, how a MapReduce solution to "counting the number of occurrences of each word in a large collection of documents" works.

5.   Why is the problem in question #4 difficult for relational databases?

6.   What other examples of MapReduce problems can you find?

7.   What is "Web-embedded structured data"?

8.   What is the "Deep Web" from the article? Give an example of data in the Deep Web. Note that "Deep Web" has two distinct meanings on the Internet.

9.   Why do we, or Google, care about the data discussed in this paper?