CSE 91 HW2 Due January 21st

In this assignment, we explore the Google search box to get a hint at what algorithms Google uses "under the hood". Words contained in square brackets [like this] are search queries that should be entered as is with all characters and punctuation that fall between the brackets. Queries can be submitted at http://www.google.com/

Make sure you are logged out of google if you have a google account before attempting the assignment.

1) Word order

In the English language, the order that words appear in can greatly change the meaning of a phrase. For example, "climbing rope" and "rope climbing" have different meanings. Does google use word order to change the results of their search queries? Give an example to support your answer.

2) Context sensitivity

a)Search for the phrase ["* fish"], with the quotation marks included. How does Google interpret the "*" character in this query?

b)Now search for the phrase [6 * 7]. Has the meaning of the "*" sign changed? How so?

3) Stemming

In language processing, "stemming" refers to treating different forms of a word as equivalent by chopping off the suffixes. For example, a program that uses stemming would treat "quick", "quickly", and "quicker" as equivalents. Does Google use stemming to determine their results? Give an example that supports your answer.

4)Stop words

A common practice in language models is to ignore a list of "stop words". Stop words are commonly used words that add little meaning to a phrase, such as "the", "and", "of", etc. However, search for [who] and [the who] and you'll see that the results are very different. Can you find two queries that differ only by stop words and have the same top result?

5)Adwords

Google uses terms in search queries to place context sensitive ads on the result screen.

a)Give an example of a query that displays an ad for Rolex with its results but does not contain the word "Rolex".

b)Can you find a query that returns no ads? Give an example.