## **CSCI 6370 IR and Web Search**

## Homework Assignment Three

Total points: 100. Due: Monday, June 22.

### Problem 1. [40 points]

Assume that a document corpus follows Zipf’s Law. Show that in this corpus, the number of words with frequency 1 is N/2 statistically, where N is the total number of words in the corpus.

### Problem 2. [40 points]

Suppose that you own a web search service. You would like to allow advertisers to bid for “keywords”. That is, ads for highest bidders will be displayed when user query contains a purchased keyword. Since you the popularity of search keywords meets Zipf’s Law, can you explain a meaningful way to set the base prices for keywords to help you make lots of profits?

### Problem 3. [20 points]

Suppose that you need to implement a web search service – Fishing Guide for the Gulf of Mexico. As a part of this project, you need to collect and index web pages pertaining to fishing in the Gulf of Mexico. Explain how you may use the topic directed crawling to help your implementation.