TTugle Search

Project 2

Due: Wednesday, November 1, 2006 @ 11:59:59 PM

1 Description

A good search engine can quickly identify all pages that contain a given word. Hashtables provide a past method to search for words. Use a hashtable to build your search engine. For a given word, build an html file that contains links to all matching webpages.

2 Assumptions

- Number of unique words: 100,000
- Number of unique pages: 1,000

3 Building the Dictionary

Create wordlist, a vector of word info. Each element of wordlist contains

- word
- key
- the index position in the vector
- common or uncommon boolean flag
- list of all pages where it occurs

Create a hashtable. The hashtable should record the index position in the wordlist vector. The hash table should use

- open addressing
- good hash function
- appropriate number of spaces to insure only a 20% chance of collisions

Fill up the wordlist with the common words so they can be easily ignored. Then, add the given pair (word, pageid) to the hashtable and wordlist as long as word is not common.

4 Searching the Dictionary

For a given word, search the hashtable for it. If the word is common, state it in a simple webpage. Otherwise, generate a webpage that contains links to the appropriate webpages where the word occurs. The webpages should correspond to filenames with "http://www.cs.ttu.edu" appended.