Homework Set 4

This problem set is due in parts.

- The hash table investigation proposal is due by Friday, **February 24**

- The hash table implementation/assessment are due Friday, **March 2**

Solutions should be submitted in PDF form using LaTeX.

Hash Table Investigation

Proposal

In this part of Homework 4, I want you to engage and think creatively. I want you to work in groups of 2 or 3 on a project investigating both hash functions and hash table implementation design choices. Our goal is to select a universe of keys that is "interesting" to your group. As suggested in class, that key-universe could be Movie titles from IMDB (or actor names, etc.), or it could be words in a corpus of books from the Gutenberg project, or it could be DNA sub strands, or it could be some other key-universe that your group is just dying to investigate. Your key universe should, however, be large, and it should not be something like simple integers.

Given a key universe, your group should identify a (small) set of possible hash functions, a set of at least two hash table implementations (chaining plus an open-addressing design), and finally, evaluation criterion for comparing the different hash functions and hash-table design choices. These should all be rolled together into a hash table investigation proposal.

Come talk to me if you want to brainstorm on any of the above elements, from key universe selection to implementation choices to input data sets and strategies for driving your hash table implementation.

Implementation/Assessment

The second phase of the assignment is to build the hash table implementation as proposed, with "testability" features so that the various options and evaluation criterion may be exercised. The final deliverable should be an assessment and analysis reporting your findings.