Assignment 10: Tables

CS230
Fall 2000

Allen B. Downey
Computer Science Department

Due: Wednesday 22 November

The purpose of this assignment is to implement a `Hashtable` and some of the algorithms that use one.

Count them words

Goal: Find the 20 most common words in our wondrous textbook.

1. Type in the code from the book that uses the built-in `Hashtable` to count word frequencies.
2. Download the text of the book from

   `http://rocky.wellesley.edu/cs230/code/hw10/text.txt`

3. Think about how to handle punctuation and any other wierdness that appears in the file.
4. How many different words are there?
5. Think about how to find the 20 most common. Hint: is there another data structure that might help?

Check those lists

Write a method for the `List` class that checks whether the list contains a loop. Your method should be linear in the number of nodes.

Implement that hashtable

Goal: write an implementation of a hashtable as described in the book. Test it by using it for the previous two programs.

1. Start with the simple `List` implementation.
2. Use the `hashCode` method provided.
3. Modify the implementation to use either an array of Lists or a Vector of Lists, whichever you prefer.
4. To implement `keys`, put all the keys in a vector and then invoke `elements` on the vector.

Random words

Goal: Write a method named `randomWord` in the `WordCount` class. It should choose and return a random word from the hashtable, such that the probability that each word appears is in proportion to its frequency in the original document.

1. It might help to keep track of n, the total number of words in the document. Then you can choose a random integer in the range $(0, n)$.
2. As you traverse the keys in the hashtable, keep a running sum of the frequencies you have seen. When this total exceeds your random number, you've found your word.