

Homework 9: Tables

cs230
Spring 2002

Allen B. Downey
Computer Science Department

Due: Tuesday 16 April

The purpose of this assignment is to use the built-in `Hashtable`. Before starting, read Chapter 19 and the documentation of the `Hashtable` class.

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/hw9/text.txt`

3. Think about how to handle punctuation and any other wierdness that appears in the file.
4. Write programs that answer the following questions and then include the answers in your program as comments.
 - How many words are there in the book?
 - How many times does the word “encapsulation” appear?
 - How many *different* words are there in the book? For purposes of comparison, there are (very roughly) 500,000 words in English, of which (very roughly) 50,000 are in current use. What percentage of this vast lexicon did I find fit to use?
 - What are the 20 most common words in the book? Hint: there is another data structure that might help with this part of the homework.