LING/C SC/PSYC 438/538

Lecture 6
Sandiway Fong
Homework 4

• Submit one PDF file
• Your submission should include code and sample runs
• Due date Monday 21\textsuperscript{st} September by midnight
Homework 4

• Question 1
  – write a program to compute and print a sorted word frequency table for a text file
  – allow both the option of case folding or not

• Example file:
  – Love for the the Bronx Bombers bubbled up in the absence of some local franchise.
  – On on on the table, we have some eggs.
  – I have Home Sharing turned on, on both my computer and my Apple tv.
  – I wish that that question had an answer.
  – Because he had had too many beers already, he skipped the Friday office happy hour.
  – Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.

no case folding:

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>buffalo</td>
<td>5</td>
</tr>
<tr>
<td>the</td>
<td>5</td>
</tr>
<tr>
<td>on</td>
<td>4</td>
</tr>
<tr>
<td>Buffalo</td>
<td>3</td>
</tr>
<tr>
<td>had</td>
<td>3</td>
</tr>
<tr>
<td>have</td>
<td>2</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
</tr>
<tr>
<td>some</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Homework 4

• Hint 1:
  – start with something like the following file I/O template:

```perl
open($fh, $ARGV[0]) or die "$ARGV[0] not found!
";
while ($line = <$fh>) {
  @words = split /\s+|[,]\.;\?/i, $line;
  print "@words"
}
```

/\s+|[,]\.;\?/ means a **regular expression** that's either
one or more spaces, or
one of , (comma), ; (semicolon), . (period), or ? (question mark)
Homework 4

• Hint 2:
  – use a hash to store the word frequencies
  – remember to sort the hash by value for print out
Homework 4

Question 2

• Using the same file as in Question 1.
• Write a Perl program that detects repeated words (many spell check/grammar programs can do this)
• Your program should read the input file and print a message stating the line number, the repeated word and its position if one exists.
• Example output:
  – Line 1: word 3, “the” is repeated 2 times
  – Line 2: word 1, “on” is repeated 3 times
  – etc.

Note: case
Homework 4

• Question 3: describe how a repeated word program could stop flagging legitimate examples of repeated words in a sentence
  – Examples:
    • I wish that that question had an answer
    • Because he had had too many beers already, he skipped the Friday office happy hour
Homework 4

- Microsoft Word:

- I saw the cat on the mat.
- The the cat sat

- I wish that that question had an answer.
- He had had had too many beers
- Buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo buffalo buffalo buffalo Buffalo

Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo buffalo.

is a grammatically correct sentence used as an example of how homonyms and homophones can be used to create complicated constructs. The sentence is unpunctuated and uses three different readings of the word “buffalo.” In order of their first use, these are:

- The city of Buffalo, New York.
- The animal "buffalo," in the plural (equivalent to "buffaloes"), in order to avoid articles.
- The verb "buffalo," meaning to confuse, deceive or intimidate.

Text excerpted from the Wikipedia articles Buffalo buffalo Buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo Buffalo buffalo Buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo Buffalo buffalo Buffalo buffalo buffalo Buffalo buffalo Buffalo buffalo, Homonym and Homophone. 30 March 2023.