

Week 4

Readings

Chapter 8

Chapter 9

Look over chapter 8 - "Named Captures" but keep in mind that unless you are using Perl 5.10 you will not be able to use named captures and named back references (page 126-127).

Key notes to keep in mind (AKA: how Perl differs)

- To match a pattern to a string not stored in `$_` using the binding operator `=~` (page 121 – "The Binding Operator, `=~`")
 - Note that the pattern matches to `$_` by default as many other operators will work on `$_` by default unless otherwise specified (Chapter 3, page 48 – "Perl's Favorite Default: `$_`")
- When storing your regular expression in a variable use single quotes to preserve the escaping of metacharacters, these are lost with double quotes and will cause unexpected results (page 122-123 – "Interpolating into Patterns")
- Perl's quantifiers are by default greedy, to make them act as lazy append a question mark ? (page 141-142 – "Nongreedy Quantifiers")
- Take a look at the `perldoc perlre` page by typing `perldoc perlre` at the command prompt or visiting <http://www.perl.com/doc/manual/html/pod/perlre.html> for a quick reference to Perl's regular expressions. Much of this may still seem foreign but don't worry you won't have to know what isn't covered in the readings, just use the beginning as a convenient reference.

Typos in the Reading

- Page 142
 - `s#(.*)#$1#g;` should be `s#<BOLD>(.*</BOLD>#$1#g;`
 - the bold tags are missing

Exercises

Book exercises from Chapter 8 (except number 4) and Chapter 9.

Write a program to read in a given file specified on the command line, add each unique word to a hash and count each occurrence (after converting it to lower case). Then print all the words and the number of times they appear, sorted alphabetically in a formatted list using `printf`.

Some caveats:

- punctuation marks and quotes should not be counted as part of the word, but hyphens and contractions should