1a-regex_set

Ben Schmidt

February 26, 2015

Exercises

Download the dictionary from benschmidt.org/words.txt and load it into Rstudio.

Try to think about these. Give it a shot. Some of them are hard–if you're simply banging your head against the wall, take a break, or simply don't finish.

Searches

1. The word "picalilli" contains five consecutive "l" or "i" letters. What word contains 6 consecutive "i" or "ls"?
2. What is the longest substring of your name for which a word contains all the matches? For example, my name is "Ben Schmidt" and I can match the first five letters wtih the capitalized letters in the word `BirkENStoCk`. What is the regex for it?
3. What dictionary words contain the same letter, three times in a row? [hard]
4. Besides the word found in question 1, are there any other words in the dictionary that contain two identical letters 6 times in a row? 7 times in a row? [harder]

Replacements

5. Design a regex that replaces the text strings `"NU"` and `"NEU"` with the word "Northeastern." **For example,** it would transform

   ```
   The NU huskies are competing in Thursday's game: email
   m.meehan@husky.neu.edu for more information.
   ```

   into

   ```
   The Northeastern huskies are competing in Thursday's game: email
   m.meehan@husky.Northeastern.edu for more information.
   ```

6. Improve your regex so that it doesn't replace strings that are part of longer words; for example, it should not replace "entrepreneur" with "entrepreNortheaternr".
7. Sometimes documents have excessive spaces in them. (For instance, if you copy and paste from the Internet). Write a regex that reduces any string of spaces down to just one. **For example,** this text:

   ```
   Good day, everyone.
   1     4      6

   3   4       10
   Good night, ladies
   ```

   Would be reduced to:

   ```
   Good day, everyone. 1 4 6 3 4 10 Good night, ladies
   ```

8. Write a regex that changes the spelling of all words in a document so that they conform to the rule “I before e, except after c.” [hard]

Concordances

An online version of the bible is at benschmidt.org/bible. It allows you to filter and replace at once on the bible. This may take some time to run, so it initially will only show values for the book of Matthew.

9. Pick a biblical figure–Jesus, for instance. Assume that the Bible is in the past tense, and that all verbs end with ‘ed’. What sort of stuff does your figure do the most? What are his most common verbs? (You don’t have to use the replacements, but there’s a way to copy and paste the top line and use a single `\1` grouping that will make the output much easier to read.)
10. Create a regex that reduces the bible to a concordance for the word “love” that shows 3 words before and 3 words after. (very hard)
11. Edit that regex so that it includes the book/line/verse number as the beginning. (very hard)