

CS 230
Programming Languages

10 / 06 / 2022

Instructor: Michael Eckmann

Today's Topics

- Questions? / Comments?
- More Regular Expressions (Regexs)
- Code examples in Java

Regex

- Reminder of the 4 principles (what takes precedence when attempting to match) that are followed:

<http://www.cs.rit.edu/~afb/20013/plc/perl5/doc/perlretut.html>

0. Match as early in the input text as possible
1. leftmost alternation that can match, matches
2. quantifiers are greedy
3. if multiple greedy quantifiers, leftmost quantifiers in the regex take precedence for greediness

Regex

- `^` `$` - beginning or end of a line (try w/ and w/o MULTILINE)
 - Note: `^` goes at beginning to match at beginning, `$` goes at end to match at end
- `\A` `\Z` – beginning or end of entire string (not lines within a string) --- useful if using MULTILINE flag and want to match the very beginning or very end of String.

- Let's try these

```
inputText = "Skidmore College\n815 N. Broadway\nSaratoga"
```

```
// note: 2 newlines within the String → three "lines" in the String
```

```
regex = "^[0-9]+";
```

```
// doesn't match in general, will match when MULTILINE flag is on
```

What if we want to store the number it matches 815 in a variable?

Regex

- `.` - matches any character except `\n`
- `.` - w/ DOTALL flag, matches any character including `\n`
- examples:
- `inputText = "Skidmore College\n815 N. Broadway\nSaratoga"`
- `regex = ".*";`
 - `// matches Skidmore College in general`
 - `// matches entire string if DOTALL flag is on`

Regex

- Let's try something like:
 - Regex: “\d+”
 - Text: “PO Box 41534”
 - How to force it to look for digits at the very beginning of the text?

Regex

- Let's write a few regular expressions.
- match any signed or unsigned integers of arbitrary length. e.g. it should match

-22

4567

1

+43

- but not things like:

+

4.56

abcd

etc.

Regex

- Let's try these:
- 1) match/find cs230 or cs106 or CS106 or CS230 but not Cs230, or cS106 etc. Find it anywhere in the text
- 2) find the first occurrence in a string of a date of format MM/DD/YY and save the month day and year in 3 separate variables if found. Do not worry if the month or day are invalid. Just find a date in that format.
- 3) ignore beginning whitespace if there is any, and match the word program OR an identifier (a letter followed by 0 or more letters/digits). Also, we want to store the program or the identifier in a variable as well as whatever comes after in a different variable. Note: it should fail if program or an identifier is NOT the first non-whitespace characters.

Regex

- Let's look at a larger parsing example using many of the features we just Learned: The “doing string selections” section of:

<http://www.troubleshooters.com/codecorn/littperl/perlreg.htm>

Regex

- Focus on principle 1:
- <http://www.cs.rit.edu/~afb/20013/plc/perl5/doc/perlretut.html>
- Consider this regex: “`^\s*(<|>|<=|>=)(.*)$`”
- Suppose input text contains “`<= 2)`”
- Will the regex match? If so, what will the first group match?

Regex

Let's write a program that repeatedly gets input from the user until the user enters the word "quit". We will inform the user that this is what will end the program.

For each line of text the user enters, tell the user whether it contains

- Skidmore
- a date in the format of 2 digits, followed by /, followed by 2 digits, followed by /, followed by 4 digits. e.g. 10/14/1994
- any lowercase letters
- any uppercase letters