Regular Expressions

Presented by : Yogesh Nikam

Agenda

- What is Regular Expression?

- Regular Expression Syntax

- Special Characters

- Examples

- Regular Expressions in PHP

What is Regular Expression?

- A regular expression is special text string for describing a search pattern

- Used to search/edit or manipulate text and data

- Used by many text editors to manipulate the text

- Many programming languages support regular expressions for string manipulation

- Combination of alpha-numeric characters and match characters

- The patterns described by regular expressions are used to search strings, extract desired parts of strings, and to do search and replace operations

What is Regular Expression?

- Example:

  "I am John, my age is 26. I work with Cybage and my mobile number is 9800123456. I live in Pune."

  Select phone number from the above text?

www.cybage.com

Cybage

Regular Expression Syntax

- Start and End characters

- Escape Characters

- Special Characters

- Valid Patterns:
    /<\/\w+>/
    |(\d{3})-\d+|Sm
    /^(?i)php[34]/
    {^\s+(\s+)?$}

- Invalid Patterns:
    /href='(.*)' - missing ending delimiter
    /\w+\s*\w+/J - unknown modifier 'J'
    1-\d3-\d3-\d4| - missing starting delimiter

5

Special Characters (Match characters)

Char	Meaning	Example
^	Match the beginning of the line	^cyb = cybage, cyberage, cybercity etc. (netcyber)
$	Match the end of the line (or before newline at the end)	hood$ = childhood, neighborhood, brotherhood etc... (mounthoodArea)
.	Matches any character except newline	c.t = cat, cot, cut
*	match 0 or more times	go*gle = matches ggle, gogle, google, gooogle etc...
+	match 1 or more times	go+gle = gogle, google, gooogle, etc. (but not ggle).
?	match 0 or 1 times	colou?r = matches both 'color' and 'colour'
|	Alternative	a|b = matches either 'a' or 'b'
( )	Grouping	(ab) = ab
[ ]	Character class	[abc] = a , b or c
{ }	Repetition modifier	ab{4} = abbbb (b should be 4 times) a{3,5} = aaa, aaaa or aaaaa
\	Quote next character (Escape)	\d, \s, a\+b = matches "a+b"
-	Range Operator	0-9, a-z, A-Z, k-p etc...

6

Special Characters (predefined character class)

Char	Meaning	Example
\d	[0-9]	matches any digit character
\D	[^0-9]	matches any non-digit character
\s	[\t\n\x0B\f\r]	matches any whitespace character (space, tab, newline)
\S	[^\s]	A non-whitespace character
\w	[a-zA-Z_0-9]	A word character
\W	[^\w]	A non-word character

- Range

[a-z]   -   matches any character between a to z

[A-Z]   -   matches any character between A to Z

[0-9]   -   matches any character between 0 to 9

- Repeatation

a*   -   zero or more a's

a+   -   one or more a's

a?   -   zero or one a's (i.e., optional a)

a{m}   -   exactly m number of a's

a{m,}   -   at least m number of a's

a{m,n}   -   at least m but at most n a's

Example:

"My Phone Number is 9890012345."

     [0-9]+   or   \d+

     [0-9]{10}   or   \d{10}

"My Phone Number is 9890012345. customer care 121."

7

Examples

- How to match IP Addresses?

  ex. 192.168.6.82, 255.255.255.0, 10.47.92.1

  \d+

  \d+\.   (escape '.' , Since '.' is a special character)

  \d+\.\d+\.\d+\.\d+

  The above Regular expression can match 12345.345.987.12345
  This is not a valid IP address.

  Improved Regular Expression:

  \d{1,3}

  \d{1,3}\.

  \d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}

- How to match a date?

  ex. 11-05-2007

  \d{1,2}\-\d{1,2}\-\d{4}

Examples

- How to match a floating point number?
  ex. 0.172, 3.41, 100, -34.67, +45, .86
  [-+]
  [-+]?
  [-+]?[0-9]*
  [-+]?[0-9]*\.?
  [-+]?[0-9]*\.?[0-9]*

- How to match an email address?
  ex. yogeshni@cybage.com, abc@yahoo.com
  \w+
  \w+@
  \w+@\w+\.
  \w+@\w+\.[a-z]{3}

Examples

"Visit W3Schools"
/w3schools/

"Visit W3Schools"
/w3schools/i

"Is this all there is?"
/is/g

"Is this all there is?"
/is/gi

Regular Expressions in PHP (PCRE)

- **preg_grep**: Return array entries that match the pattern
  array preg_grep(string pattern, array input)

```
// return all array elements containing floating point numbers
$fl_array = preg_grep("/^(\d+)?\.\d+$/", $array);
```

- **preg_match**: Perform a regular expression match
  int preg_match(string pattern, string subject [, array &matches])

```
// The "i" after the pattern delimiter indicates a case-insensitive search
if (preg_match("/php/i", "PHP is the web scripting language of choice.")) {
   echo "A match was found.";
} else {
   echo "A match was not found.";
}
```

- **preg_match_all**: Perform a global regular expression match
  int preg_match_all(string pattern, string subject [, array &matches])

Regular Expressions in PHP (PCRE)

- **preg_quote**: Quote regular expression characters

    string preg_quote(string str [, string delimiter])

```
$keywords = '$40 for a g3/400';
$keywords = preg_quote($keywords, '/');
echo $keywords; // returns \$40 for a g3\/400
```

- **preg_replace**: Perform a regular expression search and replace

    mixed preg_replace (mixed pattern, mixed replacement, mixed subject [, int limit [, int &count]])

```
$string = 'April 15, 2003';
$pattern = '/(\w+) (\d+), (\d+)/i';
$replacement = '${1}1,$3';
echo preg_replace($pattern, $replacement, $string);
```

- **preg_split**: Split string by a regular expression

    array preg_split(string pattern, string subject)

```
//split the phrase by any number of commas or space charas, which include " ", \r, \t, \n
$keywords = preg_split("/[\s,]+/", "hypertext language, programming");
```

Any Questions?

Thank you!