

Regular Expressions



Regular Expressions

CHARACTER SPECIFICATIONS	
<code>[a-m]</code>	<i>Range. A characters between a and m, inclusive</i>
<code>[a-m[A-M]]</code>	<i>Union. a through m or A through M</i>
<code>[abc]</code>	<i>Set. The character a, b, or c</i>
<code>[^abc]</code>	<i>Negation. Any character except a, b, or c</i>
<code>[a-m&&[^ck]]</code>	<i>Intersection. a though m but neither c nor k</i>
PREDEFINED SPECIFICATIONS	
<code>.</code>	<i>Any character</i>
<code>\d</code>	<i>A digit, [0-9]</i>
<code>\s</code>	<i>A whitespace character, [\t\n\x0B\f\r]</i>
<code>\w</code>	<i>A word character, [a-zA-Z_0-9]</i>
<code>\p{Punct}</code>	<i>A punctuation, [! " # \$ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` { } ~]</i>
QUANTIFIERS	
<code>x?</code>	<i>x, once or not at all</i>
<code>x*</code>	<i>x, zero or more times</i>
<code>x+</code>	<i>x, one or more times</i>
<code>x{n,m}</code>	<i>x, at least n but no more than m times</i>

Java Escape Characters

- Whitespace characters
 - `\t` for a tab
 - `\n` for a newline
- Special characters not on the keyboard (hello)
 - `\u4f60\u597d` (Chinese, 你好)
 - `\u0633\u0644\u0627\u0645` (Farsi, سلام)
 - `\u0393\u03B5\u03B9\u03AC \u03C3\u03BF\u03C5` (Greek, Γειά σου)
- Special meaning in Java
 - `\"` for a quote
 - `\\` for a backslash (very important for Reg. Ex.)

Exercise: PostalCodeChecker

- Task:
 - Output whether or not a String is a valid postal code
 - E.g., “M3J 1P3” returns `true`
- Code:
 - Presented in lecture