EDITH COWAN UNIVERSITY

CSI3105: Software Testing

Module 2d: Strings/ Languages/ Reg Ex

Creative thinkers made here.

String, Languages and Regular Expressions

1. Read the material from the slides (from the textbook) on BB as well as pages 98 and 99 of the textbook.

String – Slides from the textbook

Strings play an important role in testing. A string serves as a test input.

- Examples: 1011; AaBc; “Hello world”.

Alphabet

- A collection of symbols is known as an alphabet. We use an upper case letter such as X and Y to denote alphabets.
- Though alphabets can be infinite, we are concerned only with finite alphabets. For example, X={0, 1} is an alphabet consisting of two symbols 0 and 1. Another alphabet is Y={dog, cat, horse, lion}that consists of four symbols ``dog", ``cat", ``horse", and ``lion".

A string over an alphabet X is any sequence of zero or more symbols that belong to X. For example, 0110 is a string over the alphabet {0, 1}. Also, dog cat dog dog lion is a string over the alphabet {dog, cat, horse, lion}.

We will use lower case letters such as p, q, r to denote strings. The length of a string is the number of symbols in that string.

- Given a string s, we denote its length by |s|. Thus, |1011|=4 and |dog cat dog|=3. A string of length 0, also known as an empty string, is denoted by ε.

String – Slides from the textbook

EDITH COWAN UNIVERSITY

String Concatenation

- Let s1 and s2 be two strings over alphabet X. We write s1.s2 to denote the concatenation of strings s1 and s2.
- For example, given the alphabet X={0, 1}, and two strings 011 and 101 over X, we obtain 011.101=011101. It is easy to see that |s1.s2|=|s1|+|s2|. Also, for any string s, we have s. ε =s and ε.s=s.

Languages - Slides from the textbook

A set L of strings over an alphabet X is known as a language. A language can be finite or infinite.

For example, a set of all strings consisting of zeros and ones is the language of binary numbers.

- The following sets are finite languages over the binary alphabet {0, 1}:
  - ∅: The empty set
  - {ε}: A language consisting only of one string of length zero
  - {00, 11, 0101}: A language containing three strings

Regular Expressions - Slides from the textbook

A pattern of special characters (called metacharacters) which is used to match strings in a search. Any non-meta character matches itself

Examples:

- /a|x / -- matches a, x, aa, aax, aaxx, aaa, xx, etc
  (Note that | -- metacharacter for logical OR)
- abc* --- ab, abc, abcc, abccc ……
- (abc)* -- “ ”, abc, abcabc, abcabcabc, ….
- (a|b)(c|d) -- ac, ad, bc, bd
- (0|1)* --- all binary strings

Metacharacter	Matches…
.	**Any one character, except new line**
[a-z]	**Any one of the enclosed characters (e.g. a-z)**
*	**Zero or more of preceding character**
?	**Zero or one of the preceding characters**
+	**One or more of the preceding characters**

Regular Expressions - Metacharacter

EDITH COWAN UNIVERSITY

RE Metacharacter	Matches…
^	**beginning of line**
$	**end of line**
char	**Escape the meaning of *char* following it**
[^]	**One character not in the set**
\<	**Beginning of word anchor**
\>	**End of word anchor**
( ) or \(\)	**Tags matched characters to be used later (max = 9)**
| or |	**Or grouping**
x\{m\}	**Repetition of character x, m times (x,m = integer)**
x\{m,\}	**Repetition of character x, at least m times**
x\{m,n\}	**Repetition of character x between m and m times**