

Regular Expression Language

M. Sahebbonar, Ph.D.

- A regular expression provide a powerful, flexible & efficient method for processing text
- The extensive pattern notation of regular expression
 - To find specific character patterns
 - To extract, edit, replace, or delete text substring

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside \Rightarrow `[hc]at`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`
- `?` The preceding item can occur zero or one times \Rightarrow `colou?r`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`
- `?` The preceding item can occur zero or one times \Rightarrow `colou?r`
- `{n}` The preceding item is matched exactly n times \Rightarrow `Moham{2}ad`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`
- `?` The preceding item can occur zero or one times \Rightarrow `colou?r`
- `{n}` The preceding item is matched exactly n times \Rightarrow `Moham{2}ad`
- `{n,}` The preceding item is matched n or more times \Rightarrow `[a-z]{4,}`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`
- `?` The preceding item can occur zero or one times \Rightarrow `colou?r`
- `{n}` The preceding item is matched exactly n times \Rightarrow `Moham{2}ad`
- `{n,}` The preceding item is matched n or more times \Rightarrow `[a-z]{4,}`
- `{n,m}` The preceding item is matched at least n times, but not more than m times \Rightarrow `[0-9]{4,6}`

Regular Expression Basics

- `^` Start of line anchor \Rightarrow `^Hello`
- `$` End of line anchor \Rightarrow `World$`
- `.` Any single character \Rightarrow `a.e`
- `|` Alternation operator \Rightarrow `the|this`
- `[]` Character set, any one character listed inside `[]` \Rightarrow `[hc]at`
- `[^]` Negative character set, any one character not listed inside `[]` \Rightarrow `[^aeiou]`
- `-` Range indicator \Rightarrow `[A-Z]`
- `*` The preceding item can occur zero or more times \Rightarrow `file1*`
- `+` The preceding item can occur one or more times \Rightarrow `gro+ve`
- `?` The preceding item can occur zero or one times \Rightarrow `colou?r`
- `{n}` The preceding item is matched exactly n times \Rightarrow `Moham{2}ad`
- `{n,}` The preceding item is matched n or more times \Rightarrow `[a-z]{4,}`
- `{n,m}` The preceding item is matched at least n times, but not more than m times \Rightarrow `[0-9]{4,6}`
- `\` Escape the next character \Rightarrow `*\.txt`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`
- `\D` Anything not a digit : `[^ 0-9]`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`
- `\D` Anything not a digit : `[^ 0-9]`
- `\w` Any word character : `[A-Za-z_0-9]`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`
- `\D` Anything not a digit : `[^ 0-9]`
- `\w` Any word character : `[A-Za-z_0-9]`
- `\W` Anything not a word character : `[^\w]`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`
- `\D` Anything not a digit : `[^ 0-9]`
- `\w` Any word character : `[A-Za-z_0-9]`
- `\W` Anything not a word character : `[^\w]`
- `\s` White space : `[space , \t, \n, \v, \f, \r]`

Regular Expression Predefined Character Classes

- `\d` Any digit : `[0-9]`
- `\D` Anything not a digit : `[^ 0-9]`
- `\w` Any word character : `[A-Za-z_0-9]`
- `\W` Anything not a word character : `[^\w]`
- `\s` White space : `[space , \t, \n, \v, \f, \r]`
- `\S` Anything not whitespace : `[^\s]`

Regular Expression Predefined Character Classes

- `[: alnum :]` Alphanumeric characters
- `[: alpha :]` Alphabetic characters
- `[: blank :]` Whitespace characters
- `[: cntrl :]` Control characters in ASCII
- `[: digit :]` Numeric characters
- `[: graph :]` Graphical characters
- `[: lower :]` Lower-case alphabetic characters
- `[: print :]` Printable characters
- `[: punct :]` Punctuation characters
- `[: space :]` Space characters
- `[: upper :]` Upper-case alphabetic characters
- `[: xdigit :]` Hexadecimal digits