Regular Expressions

- Regular Expression Functions
- Special Characters and Functions
- Creating character sets and repetitions
- Capture and non-Capture groups
- Backreferences

Regular Expression

- Stronger form of pattern matching
- **re** module available in python or regular expression

- *re.match(pattern, string)* : matches at start of string, None if no match
- *re.search(pattern, string)* : matches first occurrence anywhere in string
- *re.findall(pattern, string)/ finditer()* : returns all matches

Examples

- Write regex to search word **Python** in a string.
- String starts with Python
- String ends with Python

Regular Expression Special Characters – 1

- ^ : matches start of string
  - WAR to check whether string starts with word 'Python' or not.
- $: matches end of string
  - WAR to check that string should end with '!' symbol
- . (dot) : matches any single character except newline

Special Sequences

- \d : match any digit
- \D : match any non-digit
- \b : matches a word boundary
- \s : match any white space character
- \S : match any non-whitespace character
- \w : match any alphanumeric including _
- \W : match any non-alphanumeric

Regular Expression Special Characters – 1

- * : matches preceding RE 0 or more times
- + : matches preceding RE 1 or more times
- ? : matches preceding RE 0 or 1 times(non greedy)
- \ : is used as an escape sequence

Including and excluding specific characters

- [] : specify the characters to be included inside the []
- [x-y] : specify ranges. ex: [a-z], [0-9]
- [^] : specify inverse set ex: [^a-b], [^,.]

Repetitions

- {n} : matches exactly n repetitions of preceding re.
- {m, n} : match m up to n repetitions.

Groups and non-capture groups and alternative

- () : used to group as a single repetition unit
- (?:) : makes the group as non-capturing group
- | : works as or operator

Backreferences

- Backreference means to check for some pattern that had occurred previously in the expression
- Uses \n syntax, i.e. \1 matches the first capture group and so on.
- ([a-z]) .... \1