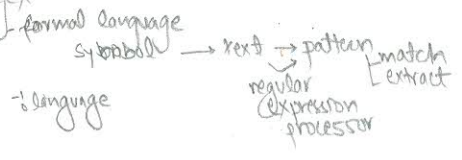


Regular Expression - Regex

Regular Expression



History

- 1943 - nervous system
- 1956 - model-algebra -> regular sets -> regular expression
- 1968 - Unix text editor -> global g/regular expression/p/print
- 1970 - Unix
- 1986 - standard POSIX Portable Operating System Interface extended grep
- 1986 - library in C
- 1987 - perl

Character

- Single
 - Literal character/strings
 - Case sensitive
 - Search tool
 - non-global/standard - leftmost is preferred
 - global
 - Metacharacter - character with special meaning

- Set [] match any one of any char
- ^ negation metacharacter not any one of any char

- any character except new line \t \r \n
- escape character \a \e \f \v

- each character one explicitly given
- Use defined range
 - alphabet, digit, int
 - Standard char sets \d digit, \w word -> alphanumeric, \s space

- POSIX Bracket Expression [[: :]]
 - alpha upper lower space blank
 - digit
 - alnum
 - x digit
 - punct print graph ctrl

- Repetition metacharacter preceding item
 - * >0
 - + >1
 - ? 0|1
 - {min,max}

- Expression
 - Greedy - regular expression process backtracking
 - Lazy - ? doesn't make efficiency

- Character
 - Single [] one of any
 - Set [] Set of character & metacharacter
 - Capturing Group

- Grouping Metacharacters - () Alternation
- Repeated & nested alternation precision, readability & efficiency

Start & End Anchors

- position
- width
- ^ start \A } String
- \$ end \Z } Line
- \b \B word boundary | space is not word boundary word character

Pattern

- Grouped Expression -> stored -> access -> Back reference
 - literal text will be matched
 - 11 to 19 1 to 9 pattern
 - with optional expression ? -> 0 width
- Captured Group Expression ()
- Non-capturing group expression (?:)
- different meaning non capturing group

Look around assertions

- check whether match exists
- 0 width
- stacked multiple assertion
- check for satisfying condition
- ? = positive lookahead
- ! = negative lookahead
- < = positive look behind
- <! = negative look behind