

Lecture 04-05

self note

RegEx:

- a regular expression is a string that represents a regular language.
- regular expression is a pattern that matches string or pieces of string.
- [] ഏ ഒരു നി മാറ്റം, that means exactly one character
- [] ഏ ഒരു first അ അക്കൻ അല്ലെങ്കിൽ ഒരു വാക്കുക്കുണ്ട് (negative)
- { } തമാജ്: exactly a or b or c
- [a,b,c] exactly a or b or c
- [^abc] any character except a, b, or c; exactly എല്ലാ രംഗം
- a^* മാറ്റണ്ട് സിരാറ്റ് പഠണ്ട് എന്തും മാറ്റണ്ട് അല്ലെങ്കിലും ഒരു രംഗം
- $a \cdot b = ab$; concatenation
- a^+ minimum ഏകവർ ഇന്തലിന് സിരാറ്റ്
- $a \cup b$ or $a|b$ ഇംഗ്ലീഷ് operation ഇല്ലായ്ക്കും either a or b.
- [brn]^{*}. as kleene closure ഓരോ പോളിംഗ് എം ചോസി-portion
D multiple time ഇല്ലായ്ക്കും; as kleene closure
- [b-chm-pp] at 1st ot
- \d ഇംഗ്ലീഷ് digits തുകാമ്പ് {0,1,2,3,4,5,6,7,8,9}

- - હારા range રૂપાય
- $a\{m\}$ હારા a will be repeated exactly m times રૂપાય
- $a\{m,n\}$ હારા a will be repeated minimum m times or up to n times
- ~~at least~~ રૂપાય
- $a\{m,n\}$ હારા a will be repeated minimum m times and maximum n times રૂપાય
- "a" હારા exactly a character ટોકાં રૂપાય
- $\backslash d\{5\}(-\backslash d\{4\})?$ \Rightarrow 11420 - 9812, 11417 - 8611.
- Hello\n world \Rightarrow Hello \n હારા newline રૂપાય, world
- mi....ft \Rightarrow contains a nine letter that starts with mi and ends with ft.
- હારા exactly any character at all
- or any character except a newline રૂપાય,
- $\backslash d^+(\backslash 0\backslash d\backslash d)?$ \Rightarrow it contains a positive integer or a floating point number with exactly two character after decimal point
- \$ [છાર] end of line રૂપાય

- ? നിന്ന് മുകളിൽ zero or more occurrence.
 - # compiler ഏ ഡീഫീൻഡി ചെയ്ത - , . , " " , . , () , { } , [] , ^ , \$, . , \ , ? , * , + ഫോർമേറ്റ് റെജിൾ എക്സ്പ്രസ്സ് അഥവാ literal character format ഉപയോഗിച്ച് കൊണ്ട് വരുത്താം.
 - ↳ for -
- literal character format

$$\left\{ \begin{array}{l} "-" \\ "\\" \end{array} \right\} \rightarrow \text{എംഗേജ് എന്റെ ഭാഷാഗാ ഫോർമേറ്റ്}$$

 express കുറഞ്ഞ രീതി.
- # A token is a pair consisting of a token name and an optional attribute value.
 - # A pattern is a description of the form that the lexemes of a token may take
 - # A lexeme is a sequence of characters in the source program that matches the pattern for a token and is identified by the lexical analyzer as an instance of that token
 - # identifier ഓ id എ തന്നെ തരം token generate ചെയ്യുന്നതു ഫോർമേറ്റ് attribute value ശിക്കാം.
 - # Symbol table ഓ id ഓ, lexeme type, location at which it is first found. (for error tracking), ... എംഗേജ് info ശിക്കാം,