

regEx:

- a regular expression is a string that represents a regular language.
- regular expression is a pattern that matches string or pieces of string.

- $[]$ એ સારું એ જ સારું, that means exactly one character
- $[]$ એ સારું first એ \wedge શરૂઆતે ઉપરના ચાર ચારિત્રોને હટાવે (negate)

દાખલ:

→ $[a, b, c]$ exactly a or b or c હશે

→ $[^a b c]$ any character except a, b, or c ; exactly એકો હશે

- a^* શરૂઆતે એ જ સારું એ જ સારું એ જ સારું એ જ સારું એ જ સારું
↳ $\epsilon, a, aa, aaa, \dots$

- $a \cdot b = ab$; • ફક્ત concatenation જ

- a^+ minimum એક સારું એ જ સારું એ જ સારું એ જ સારું એ જ સારું
↳ $a, aa, aaa, aaaa, \dots$

- $a \cup b$ or $a|b$ ફક્ત or operation જાય; either a or b.

- $[brn]^*$ as kleene closure જે બંધાયેલો છે એ છે choosing portion જે multiple time હશે; as kleene closure

સુધારા એ જ સારું એ જ સારું એ જ સારું એ જ સારું એ જ સારું

- $[b-chm-pp]$ at | ot

↳ $\{bat, cat, hat, mat, nat, oat, pat, fat, ot\}$

- $\backslash d$ ફક્ત digits જાય $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- $-$ ફરક range રૂબરૂ

- $a\{m\}$ ફરક a will be repeated exactly m times રૂબરૂ

- $a\{m, \}$ ફરક a will be repeated minimum m times રા ઠાર

હરફ રૂબરૂ

- $a\{m, n\}$ ફરક a will be repeated minimum m times and maximum n times રૂબરૂ

- $"a"$ ફરક exactly a character ઠેક રૂબરૂ

- $\backslash d\{5\} (-\backslash d\{4\})?$ \Rightarrow 11420 - 9812, 11417-8611

- Hello $\backslash n$ world \Rightarrow Hello $\backslash n$ ફરક newline રૂબરૂ, world

- mi...ft \Rightarrow contains a nine letter that starts with mi and ends with ft.

.... ફરક exactly any character at all રા any character except a newline રૂબરૂ

- $\backslash d^+ (1.\backslash d\backslash d)?$ \Rightarrow contains a positive integer or a floating point number with exactly two character after decimal points

- $\$$ ફરક end of line રૂબરૂ

• ? দিখায় বুলায় Zero or more occurrence.

compiler এর defined চিহ্নসমূহ হল - , . , " " , ' ' , () , { } , [] , ^ , \$, * , \ , ? , * , + যদি regex এ define করা লাগবে তখন literal character format এ use করে.

↳ for -

literal
character
format

{ " - " ,
" \ - " }

→ এতে সঠিক ভাবে ব্যাকরণ প্রকাশ করা
express করা হবে.

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 বা id এর জন্য token generate হয় আর একটি attribute value থাকবে.

Symbol table এ id এর,

lexeme type, location at which it is first found.

(for error tracking), ... আর info থাকবে.