Regular Expression Grammar

Regex ::= StartOfStringAnchor? Expression

Expression ::= Subexpression("|" Expression)?

Subexpression ::= SubexpressionItem+

SubexpressionItem ::= Match | Group | Anchor | Backreference

Group ::= "(" GroupNonCapturingModifier? Expression ")" Quantifier?

GroupNonCapturingModifier ::= "?:"

Match ::= MatchItem Quantifier?

MatchItem ::= MatchAnyCharacter | MatchCharacterClass | MatchCharacter

MatchAnyCharacter ::= "."

MatchCharacterClass ::= CharacterGroup | CharacterClass | CharacterClassFromUnicodeCategory

MatchCharacter ::= Char

CharacterGroup ::= "[" CharacterGroupNegativeModifier? CharacterGroupItem+ "]"

CharacterGroupNegativeModifier ::= "^"

CharacterGroupItem ::= CharacterClass | CharacterClassFromUnicodeCategory | CharacterRange | Char /* excluding "]" */

CharacterClass ::= CharacterClassAnyWord | CharacterClassAnyWordInverted | CharacterClassAnyDecimalDigit | CharacterClassAnyDecimalDigitInverted

CharacterClassAnyWord ::= "\w"

CharacterClassAnyWordInverted ::= "\W"

CharacterClassAnyDecimalDigit ::= "\d"

CharacterClassAnyDecimalDigitInverted ::= "\D"

CharacterClassFromUnicodeCategory ::= "\p"{ UnicodeCategoryName }

UnicodeCategoryName ::= Letters

CharacterRange ::= Char ("-" Char)?

Quantifier ::= QuantifierType LazyModifier?

$$\text{QuantifierType} ::= \begin{cases} \text{ZeroOrMoreQuantifier} \\ \text{OneOrMoreQuantifier} \\ \text{ZeroOrOneQuantifier} \\ \text{RangeQuantifier} \end{cases}$$

$$\text{LazyModifier} ::= \text{"?"}$$

$$\text{ZeroOrMoreQuantifier} ::= \text{"*"}$$

$$\text{OneOrMoreQuantifier} ::= \text{"+"}$$

$$\text{ZeroOrOneQuantifier} ::= \text{"?"}$$

$$\text{RangeQuantifier} ::= \text{"\{" RangeQuantifierLowerBound ("," RangeQuantifierUpperBound?)? "\}"}$$

$$\text{RangeQuantifierLowerBound} ::= \text{Integer}$$

$$\text{RangeQuantifierUpperBound} ::= \text{Integer}$$

$$\text{Backreference} ::= \backslash\text{Integer}$$

$$\text{StartOfStringAnchor} ::= \text{"\^{}"}$$

$$\text{Anchor} ::= \begin{cases} \text{AnchorWordBoundary} \\ \text{AnchorNonWordBoundary} \\ \text{AnchorStartOfStringOnly} \\ \text{AnchorEndOfStringOnlyNotNewline} \\ \text{AnchorEndOfStringOnly} \\ \text{AnchorPreviousMatchEnd} \\ \text{AnchorEndOfString} \end{cases}$$

$$\text{AnchorWordBoundary} ::= \text{"}\backslash\text{b"}$$

$$\text{AnchorNonWordBoundary} ::= \text{"}\backslash\text{B"}$$

$$\text{AnchorStartOfStringOnly} ::= \text{"}\backslash\text{A"}$$

$$\text{AnchorEndOfStringOnlyNotNewline} ::= \text{"}\backslash\text{z"}$$

$$\text{AnchorEndOfStringOnly} ::= \text{"}\backslash\text{Z"}$$

$$\text{AnchorPreviousMatchEnd} ::= \text{"}\backslash\text{G"}$$

$$\text{AnchorEndOfString} ::= \text{"\$"}$$

$$\text{Integer} ::= [0\text{-}9]+$$

$$\text{Letters} ::= [a\text{-}zA\text{-}Z]+$$

$$\text{Char} ::= \begin{cases} \#x9 \\ \#xA \\ \#xD \\ [\#x20\text{-}\#xD7FF] \\ [\#xE000\text{-}\#xFFFD] \\ [\#x10000\text{-}\#x10FFFF] \end{cases}$$