

Formally Defining the Behaviour of Anchored Expressions:

$L_p(r) = \{w: w \text{ is accepted by } r \text{ in programmer regex}\}$

$L_m(r) = \{w: w \text{ is accepted by } r \text{ in Mathematical RE}\}$

$f(r) \rightarrow$ the mathematical version of programmer expression r
that is, $L_p(r) = L_m(f(r))$

Define f , and prove $L_p(r) = L_m(f(r))$ for any valid programmer's regular expression r .

$f = a \circ p$ (first parse.js conversion, then handle anchor "expansion")

$p(\text{word boundary}) = \text{Error}$

$p(\text{backreference}) = \text{Error}$

$p(\text{[negative] look ahead/behind}) = \text{Error}$

$p(\dots)$	inside $[\dots]$	otherwise
$\backslash d$	$0-9$	$[0-9]$
$\backslash D$	Error	$[^0-9]$
$\backslash w$	$0-9A-Za-z_$	$[0-9A-Za-z_]$
$\backslash W$	Error	$[^0-9A-Za-z_]$
$\backslash s$	$\langle \text{space} \rangle$	$\langle \text{space} \rangle$
$\backslash S$	Error	$[^]$

$p(r+) = rr^*$

$p(r\{x\}) = r^x$

$p(r\{x, \infty\}) = r^x r^*$

$p(r\{i, x\}) = \bigwedge_{i=1}^{\lceil \log_2(x) \rceil} (\epsilon + p(r\{i\})) \cdot p(r\{, x - \lceil \log_2(x) \rceil \})$

$p(r\{x, y\}) = p(r\{x\}) p(r\{, y-x\})$

Anything unspecified remains unchanged $p(x) = x$