

PARSING REGEX BACKREFERENCES WITH DERIVATIVES

For now, I'm just typesetting the derivative rules.

Some notations:

- Σ is the set of characters in the alphabet
- c is a metavariable for a single character from Σ
- Σ^* is a (possibly empty) string of characters
- ε is the empty string
- x is a metavariable for the name of a capturing group/backreference
- θ is a metavariable for a finite map from variables to strings
- right-biased merge of two finite maps is $A + B$

The basic idea is that we extend the accepts-empty function $\nu(r)$ from returning a simple boolean to returning possible substitutions. Then, ∂ is parameterized by a substitution. Annoyingly (but it's not that bad), ν must also be parameterized by a substitution.

Here are the core rules:

$$\begin{aligned}
\nu^\theta(\perp) &= \{\} \\
\nu^\theta(c) &= \{\} \\
\nu^\theta(\varepsilon) &= \{\theta\} \\
\nu^\theta(r \cdot r') &= \bigcup_{\theta' \in \nu^\theta(r)} \nu^{\theta'}(r') \\
\nu^\theta(r + r') &= \nu^\theta(r) \cup \nu^\theta(r') \\
\nu^\theta(r^*) &= \{\theta\} \\
\nu^\theta(x=\Sigma^*.r) &= \{\theta' + \{x \mapsto \Sigma^*\} \mid \theta' \in \nu^\theta(r)\} \\
\nu^\theta(x) &= \begin{cases} \{\theta\} & \text{if } \theta(x) = \varepsilon \\ \{\} & \text{otherwise} \end{cases} \\
\nu^\theta(\theta': r) &= \nu^{\theta+\theta'}(r) \\
\partial_c^\theta \perp &= \perp \\
\partial_c^\theta c' &= \begin{cases} \varepsilon & \text{if } c = c' \\ \perp & \text{otherwise} \end{cases} \\
\partial_c^\theta \varepsilon &= \perp \\
\partial_c^\theta(r \cdot r') &= \partial_c^\theta r \cdot r' + \sum_{\theta' \in \nu^\theta(r)} \partial_c^{\theta'} r' \\
\partial_c^\theta(r + r') &= \partial_c^\theta r + \partial_c^\theta r' \\
\partial_c^\theta r^* &= \partial_c^\theta r \cdot r^* \\
\partial_c^\theta x=\Sigma^*.r &= x=\Sigma^*c. \partial_c^\theta r \\
\partial_c^\theta x &= \begin{cases} \partial_c^\theta \Sigma^* & \text{where } \theta(x) = \Sigma^* \\ \perp & \text{if } x \notin \text{dom}(\theta) \end{cases} \\
\partial_c^\theta \theta': r &= \partial_c^{\theta+\theta'} r
\end{aligned}$$