

# RegEx

Fahad Ausaf<sup>1</sup>, Roy Dyckhoff<sup>2</sup>, and Christian Urban<sup>1</sup>

<sup>1</sup> King's College London, United Kingdom  
<sup>2</sup> St Andrews

**Abstract.** BLA BLA Sulzmann and Lu [1]

**Keywords:**

## 1 Introduction

Regular expressions

$$r := \text{NULL} \mid \text{EMPTY} \mid \text{CHAR } c \mid \text{ALT } r_1 r_2 \mid \text{SEQ } r_1 r_2 \mid \text{STAR } r$$

Values

$$v := \text{Void} \mid \text{Char } c \mid \text{Left } v \mid \text{Right } v \mid \text{Seq } v_1 v_2 \mid \text{Stars } vs$$

The language of a regular expression

$$\begin{aligned} L \text{ NULL} &\stackrel{\text{def}}{=} \emptyset \\ L \text{ EMPTY} &\stackrel{\text{def}}{=} \{\emptyset\} \\ L (\text{CHAR } c) &\stackrel{\text{def}}{=} \{[c]\} \\ L (\text{SEQ } r_1 r_2) &\stackrel{\text{def}}{=} (L r_1) @ (L r_2) \\ L (\text{ALT } r_1 r_2) &\stackrel{\text{def}}{=} (L r_1) \cup (L r_2) \\ L (\text{STAR } r) &\stackrel{\text{def}}{=} (L r)^\star \end{aligned}$$

The nullable function

$$\begin{aligned} \text{nullable } \text{NULL} &\stackrel{\text{def}}{=} \text{False} \\ \text{nullable } \text{EMPTY} &\stackrel{\text{def}}{=} \text{True} \\ \text{nullable } (\text{CHAR } c) &\stackrel{\text{def}}{=} \text{False} \\ \text{nullable } (\text{ALT } r_1 r_2) &\stackrel{\text{def}}{=} \text{nullable } r_1 \vee \text{nullable } r_2 \\ \text{nullable } (\text{SEQ } r_1 r_2) &\stackrel{\text{def}}{=} \text{nullable } r_1 \wedge \text{nullable } r_2 \\ \text{nullable } (\text{STAR } r) &\stackrel{\text{def}}{=} \text{True} \end{aligned}$$

The derivative function for characters and strings

$$\begin{aligned}
\text{der } c \text{ } \mathit{NULL} &\stackrel{\text{def}}{=} \mathit{NULL} \\
\text{der } c \text{ } \mathit{EMPTY} &\stackrel{\text{def}}{=} \mathit{NULL} \\
\text{der } c \text{ } (\mathit{CHAR } c') &\stackrel{\text{def}}{=} \text{if } c = c' \text{ then } \mathit{EMPTY} \text{ else } \mathit{NULL} \\
\text{der } c \text{ } (\mathit{ALT } r_1 r_2) &\stackrel{\text{def}}{=} \mathit{ALT} (\text{der } c r_1) (\text{der } c r_2) \\
\text{der } c \text{ } (\mathit{SEQ } r_1 r_2) &\stackrel{\text{def}}{=} \text{if nullable } r_1 \text{ then } \mathit{ALT} (\mathit{SEQ} (\text{der } c r_1) r_2) (\text{der } c r_2) \\
&\quad \text{else } \mathit{SEQ} (\text{der } c r_1) r_2 \\
\text{der } c \text{ } (\mathit{STAR } r) &\stackrel{\text{def}}{=} \mathit{SEQ} (\text{der } c r) (\mathit{STAR } r) \\
\text{ders } [] r &\stackrel{\text{def}}{=} r \\
\text{ders } (c::s) r &\stackrel{\text{def}}{=} \text{ders } s (\text{der } c r)
\end{aligned}$$

The *flat* function for values

$$\begin{aligned}
|\mathit{Void}| &\stackrel{\text{def}}{=} [] \\
|\mathit{Char } c| &\stackrel{\text{def}}{=} [c] \\
|\mathit{Left } v| &\stackrel{\text{def}}{=} |v| \\
|\mathit{Right } v| &\stackrel{\text{def}}{=} |v| \\
|\mathit{Seq } v_1 v_2| &\stackrel{\text{def}}{=} |v_1| @ |v_2| \\
|\mathit{Stars } []| &\stackrel{\text{def}}{=} [] \\
|\mathit{Stars } (v::vs)| &\stackrel{\text{def}}{=} |v| @ |\mathit{Stars } vs|
\end{aligned}$$

The *mkeps* function

$$\begin{aligned}
\text{mkeps } \mathit{EMPTY} &\stackrel{\text{def}}{=} \mathit{Void} \\
\text{mkeps } (\mathit{SEQ } r_1 r_2) &\stackrel{\text{def}}{=} \mathit{Seq} (\text{mkeps } r_1) (\text{mkeps } r_2) \\
\text{mkeps } (\mathit{ALT } r_1 r_2) &\stackrel{\text{def}}{=} \text{if nullable } r_1 \text{ then } \mathit{Left} (\text{mkeps } r_1) \text{ else } \mathit{Right} (\text{mkeps } r_2) \\
\text{mkeps } (\mathit{STAR } r) &\stackrel{\text{def}}{=} \mathit{Stars } []
\end{aligned}$$

The *inj* function

$$\begin{aligned}
\text{inj } \mathit{EMPTY } c \mathit{Void} &\stackrel{\text{def}}{=} \mathit{Char } c \\
\text{inj } (\mathit{CHAR } d) c \mathit{Void} &\stackrel{\text{def}}{=} \mathit{Char } d \\
\text{inj } (\mathit{CHAR } d) c (\mathit{Char } c') &\stackrel{\text{def}}{=} \mathit{Seq} (\mathit{Char } d) (\mathit{Char } c') \\
\text{inj } (\mathit{ALT } r_1 r_2) c (\mathit{Left } v_1) &\stackrel{\text{def}}{=} \mathit{Left} (\text{inj } r_1 c v_1) \\
\text{inj } (\mathit{ALT } r_1 r_2) c (\mathit{Right } v_2) &\stackrel{\text{def}}{=} \mathit{Right} (\text{inj } r_2 c v_2) \\
\text{inj } (\mathit{SEQ } r_1 r_2) c (\mathit{Seq } v_1 v_2) &\stackrel{\text{def}}{=} \mathit{Seq} (\text{inj } r_1 v_1 v_2) v_2.0 \\
\text{inj } (\mathit{SEQ } r_1 r_2) c (\mathit{Left} (\mathit{Seq } v_1 v_2)) &\stackrel{\text{def}}{=} \mathit{Seq} (\text{inj } r_1 c v_1) v_2 \\
\text{inj } (\mathit{SEQ } r_1 r_2) c (\mathit{Right } v_2) &\stackrel{\text{def}}{=} \mathit{Seq} (\text{mkeps } r_1) (\text{inj } r_2 c v_2) \\
\text{inj } (\mathit{STAR } r) c (\mathit{Seq } v (\mathit{Stars } vs)) &\stackrel{\text{def}}{=} \mathit{Stars} ((\text{inj } r c v)::vs)
\end{aligned}$$

The inhabitation relation:

$$\begin{array}{c}
\frac{\frac{\frac{\vdash v_1 : r_1}{\vdash \text{Seq } v_1 v_2 : \text{SEQ } r_1 r_2}}{\vdash v_1 : r_1} \quad \frac{\frac{\vdash v_2 : r_2}{\vdash \text{Seq } v_1 v_2 : \text{SEQ } r_1 r_2}}{\vdash v_2 : r_2}}{\vdash \text{Seq } v_1 v_2 : \text{SEQ } r_1 r_2}} \\
\frac{\frac{\vdash v_1 : r_1}{\vdash (\text{Left } v_1) : \text{ALT } r_1 r_2}}{\vdash v_1 : r_1} \quad \frac{\frac{\vdash v_2 : r_1}{\vdash (\text{Right } v_2) : \text{ALT } r_2 r_1}}{\vdash v_2 : r_1}}{\vdash (\text{Left } v_1) : \text{ALT } r_1 r_2} \quad \vdash (\text{Right } v_2) : \text{ALT } r_2 r_1} \\
\frac{}{\vdash \text{Void} : \text{EMPTY}} \quad \frac{}{\vdash (\text{Char } c) : \text{CHAR } c} \\
\frac{}{\vdash \text{Stars } [] : \text{STAR } r} \quad \frac{\frac{\vdash v : r \quad \vdash \text{Stars } vs : \text{STAR } r}{\vdash \text{Stars } (v::vs) : \text{STAR } r}}{\vdash v : r}
\end{array}$$

We have also introduced a slightly restricted version of this relation where the last rule is restricted so that  $|v| \neq []$ . This relation for *non-problematic* is written  $\models v : r$ .

Our Posix relation  $s \in r \rightarrow v$

$$\begin{array}{c}
\frac{}{[] \in \text{EMPTY} \rightarrow \text{Void}} \quad \frac{}{[c] \in \text{CHAR } c \rightarrow (\text{Char } c)} \\
\frac{s \in r_1 \rightarrow v}{s \in \text{ALT } r_1 r_2 \rightarrow (\text{Left } v)} \quad \frac{s \in r_2 \rightarrow v \quad s \notin (L r_1)}{s \in \text{ALT } r_1 r_2 \rightarrow (\text{Right } v)} \\
\frac{\frac{s_1 \in r_1 \rightarrow v_1 \quad s_2 \in r_2 \rightarrow v_2}{\nexists s_3 s_4. s_3 \neq [] \wedge s_3 @ s_4 = s_2 \wedge s_1 @ s_3 \in (L r_1) \wedge s_4 \in (L r_2)}}{(s_1 @ s_2) \in \text{SEQ } r_1 r_2 \rightarrow \text{Seq } v_1 v_2} \\
\frac{s_1 \in r \rightarrow v \quad s_2 \in \text{STAR } r \rightarrow \text{Stars } vs \quad |v| \neq []}{(s_1 @ s_2) \in \text{STAR } r \rightarrow \text{Stars } (v::vs)} \quad \frac{}{[] \in \text{STAR } r \rightarrow \text{Stars } []}
\end{array}$$

Our version of Sulzmann's ordering relation

$$\begin{array}{c}
\frac{v_1 \succeq_{r_1} v_1' \quad v_1 \neq v_1'}{Seq\ v_1\ v_2 \succeq_{SEQ\ r_1\ r_2} Seq\ v_1'\ v_2'} \quad \frac{v_2 \succeq_{r_2} v_2'}{Seq\ v_1\ v_2 \succeq_{SEQ\ r_1\ r_2} Seq\ v_1\ v_2'} \\
\frac{len(|v_1|) \leq len(|v_2|)}{Left\ v_2 \succeq_{ALT\ r_1\ r_2} Right\ v_1} \quad \frac{len(|v_2|) < len(|v_1|)}{Right\ v_1 \succeq_{ALT\ r_1\ r_2} Left\ v_2} \\
\frac{v_2 \succeq_{r_2} v_2'}{Right\ v_2 \succeq_{ALT\ r_1\ r_2} Right\ v_2'} \quad \frac{v_1 \succeq_{r_1} v_1'}{Left\ v_1 \succeq_{ALT\ r_1\ r_2} Left\ v_1'} \\
\\
\frac{}{Void \succeq_{EMPTY} Void} \quad \frac{}{Char\ c \succeq_{CHAR} c\ Char\ c} \\
\frac{|Stars\ (v::vs)| = []}{Stars\ [] \succeq_{STAR\ r} Stars\ (v::vs)} \quad \frac{|Stars\ (v::vs)| \neq []}{Stars\ (v::vs) \succeq_{STAR\ r} Stars\ []} \\
\\
\frac{v_1 \succeq_r v_2}{Stars\ (v_1::vs_1) \succeq_{STAR\ r} Stars\ (v_2::vs_2)} \\
\\
\frac{Stars\ vs_1 \succeq_{STAR\ r} Stars\ vs_2}{Stars\ (v::vs_1) \succeq_{STAR\ r} Stars\ (v::vs_2)} \quad \frac{}{Stars\ [] \succeq_{STAR\ r} Stars\ []}
\end{array}$$

Some lemmas

*If nullable  $r$  then  $\vdash mkeps\ r : r$ .*

*If nullable  $r$  then  $|mkeps\ r| = []$ .*

...

## References

1. M. Sulzmann and K. Lu. POSIX Regular Expression Parsing with Derivatives. In *Proc. of the 12th International Conference on Functional and Logic Programming (FLOPS)*, volume 8475 of LNCS, pages 203–220, 2014.