ANHANGC: DIE SYNTAX UND SEMANTIK *REGULÄRER AUSDRÜCKE* IN (F)LEX

Dieser Anhang gibt informell einen Überblick über *Reguläre Ausdrücke* in **(f)lex**. Für weitere Details sei der interessierten Leser auf Herold (1995:49-67) verwiesen.

C.1 Syntax

C.1.1. Zeichensatz

Reguläre Ausdrücke bestehen aus einfachen (ASCII-) Zeichen und Meta-Zeichen. Meta-Zeichen sind:

\ ^ $. [ ] * + { }

C.1.2 *Einfache Reguläre Ausdrücke*

Einfache Reguläre Ausdrücke sind :

(1) einfache Zeichen
(2) die Sequenz des Metazeichen "\"und eines weiteren Metazeichens (z.B. "*")
(3) bestimmte ESCAPE-Sequenzen, wie "\n" für "Newline".
(4) das Metazeichen "." ("ein beliebiges Zeichen ausser "\n").
(5) Zeichenklassen mit Abkürzungen, wie [abc] ("a oder b oder c") oder [A-Z] ("eins der Zeichen aus A bis Z")
(6) Komplementklassen von Zeichen, z.B. [^a] ("ein beliebiges Zeichen ausser a")

C.1.3 *Reguläre Ausdrücke*

Einfache Reguläre Ausdrücke sind *Reguläre Ausdrücke*. Wenn R_1, ..., R_n *Reguläre Ausdrücke* sind, sind es auch:

(1)	$(R_1 \mid ... \mid R_n)$	(Disjunktion, "R_1 oder...oder R_n")
(2)	R_1R_2	(Konkatenation, die Verkettung von R_1 und R_2)
(3)	R*	(Kleene-Abschluß, "0 oder beliebig viele Male R")
(4)	R+	(+-Abschluß, "ein oder mehrmals R")
(5)	R?	("ein oder keinmal R")
(6)	R{m, n}	(für m,n, ganze Zahlen: "m bis n-mal R")

C.2 Semantik

C.2.1 *Einfache Reguläre Ausdrücke*

(1) Jedes einfache Zeichen denotiert einen String, der nur dieses Zeichen enthält. So denotiert a den String a.
(2) Eine Sequenz von "\" und einem Metazeichen denotiert das Metazeichen. Z.B. denotiert "*" "*".
(3) ESCAPE-Sequenzen denotieren das jeweilige ASCII-Symbol, z.B. \n "Newline".
(4) Zeichenklassen denotieren die Vereinigungsmenge der in ihnen enthaltenen Strings, Abkürzungen wie "a-c" oder "5-8" stehen dabei für die Folge aller Zeichen, deren ASCII- Wert größer/gleich als das linke und kleiner/gleich als das rechte Begrenzungszeichen ist. [a-c] steht also für [abc] und [5-8] für [5678].
(5) "." denotiert die Menge aller Strings der Länge 1 ausser "\n".
(6) Komplementklassen denotieren die Zeichenklasse für alle Zeichen, die nicht angeführt sind. Z.B. steht [^a] für $[x_1...x_n]$, wobei $x_1...x_n$ alle ASCII-Zeichen ausser a sind.

C.2.2 *Reguläre Ausdrücke*

(1)-(4) s. **2.1.3** in **§1/2**.
(5) $(RA_1)(RA_2)?(RA_3)$ ist eine Abkürzung für $((RA_1) (RA_2) (RA_3) | (RA_1) (RA_3))$
(6) $R\{m,n\}$ ist eine Abkürzung für $(R_1...R_m | R_1...R_{m+1} | .... | R_1...R_{1n-1}|R_1...R_n)$[1]

C.2.3. Sonderzeichen

verlieren ihren Sonderstatus durch das bereits besprochene Voranstellen von "\", oder aber in Zeichen und Komplementklassen, sowie in *Regulären Ausdrücken*, die von Anführungszeichen begrenzt sind. Die folgenden RAs sind äquivalent:

*bc\?de? [*]bc[?]de? "*bc?"de?

[1]Die Indizes stehen hier für verschiedene Instanzen von R.