Prof. Franz Schweiggert — Abteilung Angewandte Informationsverarbeitung — 13. April 2005
Christian Ehrhardt — Blatt 1

Unix-basierte Implementierung kleiner Datenbanken (SS 2005)

Abgabetermin 20.04.2005

Formale Sprachen (10 Punkte)

Grammatik (4 Punkte)

Gegeben sei die folgende Grammatik (“␣” steht für ein einzelnes Leerzeichen, Großbuchstaben sind Non-Terminale, Startsymbol ist A):

$$
\begin{array}{rcl}
A & \longrightarrow & C\ B\ into\ B\ D\ B\ values\ B\ F \\
B & \longrightarrow & \text{␣} \mid B\ \text{␣} \\
C & \longrightarrow & insert \\
D & \longrightarrow & E\ \{E\} \\
E & \longrightarrow & a \mid b \mid c \mid \ldots \mid z \\
F & \longrightarrow & (\ G\ ) \\
G & \longrightarrow & H\ \{\ ,\ H\ \} \\
H & \longrightarrow & '\ D\ '
\end{array}
$$

Erklärt Eurem Tutor welche Sätze zu der durch diese Grammatik definierten Sprache gehören und gebt einige Sätze an. Formuliert anschließend (falls möglich) einen regulären Ausdruck, der die selbe Sprache definiert oder erklärt wieso es keinen passenden regulären Ausdruck gibt.

Reguläre Ausdrücke (6 Punkte)

- Beschreibt Eurem Tutor, welche Worte zu der Sprache gehören, die durch den regulären Ausdruck `\<[a-z]+ *[<=>]=? *[a-z]+\>` definiert wird.
- Gebt einen regulären Ausdruck an, der eine Sprache definiert, die genau die (syntaktisch) gültigen Web-Adressen (URLs) enthält. Ihr könnt dabei davon ausgehen, daß eine solche Adresse neben Punkt (“.”, Doppelpunkt (“:”) und Slash (“/”) nur Kleinbuchstaben, Ziffern, das Minuszeichen (“-”) und den Unterstrich (“_”) enthalten darf. Syntaktisch gültig sind also z.B.:
  - `http://dbk.de/vatican/`
  - `http://de`
  - `http://www.xx`
  - `http://www.mathematik.uni-ulm.de/sai/ss05/unixdb/index.html`

  Nicht gültig sind dagegen z.B.:
  - `www.ulm.org` (http fehlt)
  - `http://www..uni.de` (“`..`” im Hostteil ist nicht erlaubt)
- Gebt einen von egrep (siehe Manualseite) unterstützten erweiterten regulären Ausdruck an, der alle Zeilen in einer Datei findet, die zwei gleiche *Worte* enthalten. Die beiden gleichen Worte müssen dabei nicht notwendigerweise hintereinander vorkommen.