

Theoretische Informatik

Prof. Dr. B. Nebel, Prof. Dr. G. Lausen
M. Ragni, K. Simon und C.-N. Ziegler
WS 2004/2005

Universität Freiburg
Institut für Informatik

Übungsblatt 12

Abgabe: 28. Januar 2004

Aufgabe 12.1 (Reguläre Ausdrücke mit GREP – 6 Punkte (4×1.5))

Reguläre Ausdrücke finden in der Praxis vielfältige Anwendung. Hierzu zählen beispielsweise Automaten zur lexikalischen Analyse von Token Streams im Compilerbau, sowie Werkzeuge zur Vereinfachung alltäglicher Tätigkeiten, darunter vornehmlich die Textsuche in Dokumenten.

Zum Beispiel dient das vornehmlich unter UNIX weithin verbreitete Tool GREP der Extraktion von Text aus Dateien mittels regulärer Ausdrücke. Im Folgenden soll dieses Werkzeug zur Auffindung bestimmter Zeichenfolgen dienen. Den zugrunde liegenden Text sowie Hinweise zur Benutzung von GREP und eine unter Windows lauffähige Version können aus dem Netz geladen werden. Siehe hierzu die Informationen auf der Vorlesungsseite.

Beispiel: Der Shell-Aufruf `grep -E -o -e "[a-zA-Z]+" regex.txt` gibt jedes im Text vorkommende Wort aus. Die Option `-E` ist hierbei unter UNIX nötig, um erweiterte Patterns erkennen zu können, die Syntax unter Windows kann unter Umständen abweichen. Im Beispiel dürfen jene Worte weder Ziffern noch Leerzeichen oder andere Sonderzeichen enthalten.

Finden Sie folgende Zeichenketten mittels GREP:

- (a) **Alle in \LaTeX -Formeln auftretenden Sub- und Superskripte.**
Sub- und Superskripte in \LaTeX (z.B. ist i in a_i ein Subskript und in a^i ein Superskript) werden durch “`_`” bzw. “`^`” eingeleitet und durch “`}`” geschlossen. So drückt in \LaTeX die Zeichenfolge “ `$\$a_{\{i\}}\$$` ” beispielsweise a_i aus. Für den gesuchten regulären Ausdruck nehme man an, dass keine Schachtelung von Sub- oder Subskripten und keine weitere Klammerung “`{`” und “`}`” im Sub-/Superskript vorliegt.
- (b) **Durch `\textit`, `\textbf` und `\textsf` hervorgehobene Wörter.**
Finden Sie mit Hilfe von GREP alle durch die \LaTeX -Makros “`\textit{...}`”, “`\textbf{...}`” und “`\textsf{...}`” hervorgehobenen Wörter und geben Sie diese mitsamt den klammernden Makros aus.
- (c) **Alle Dezimalzahlen, sowohl mit als auch ohne Nachkommastelle.**
Gemäß dem gängigen IEEE Standard definieren unter anderem die Zeichenfolgen “`-27`”, “`1.043e-17`”, “`+14.120E12`” und “`0.5`” legitime Dezimalzahlen. Beachte, dass Vorkammateil und Exponent keine führenden Nullen enthalten dürfen. Außerdem sind Vorzeichen “`+`” und “`-`” optional.
- (d) **Ausgabe der Inhalte von \LaTeX -Umgebungen.**
 \LaTeX -Umgebungen werden mit “`\begin{x}`” eingeleitet und “`\end{x}`” beendet, wobei x der Umgebungsname ist, z.B. “`example`” oder “`tabular`”. Sämtliche Umgebungen samt Inhalt sollen nun ausgegeben werden,

jedoch nur, wenn diese im Textfile über keine Zeilengrenze hinweg gehen. Es ist darauf zu achten, dass stets bis zum *entsprechenden* Umgebungs-ende gelesen wird. Erkennt GREP also “`\begin{center}`”, muss folglich bis “`\end{center}`” eingelesen werden. Verwenden Sie hierzu Gruppierung durch “(” und “)”.

Hinweis: Um die Aufgaben zu erleichtern, sind in der ebenfalls auf der Webseite verfügbaren Datei `regex.out.txt` die geforderten Ausgaben für jede einzelne Teilaufgabe explizit aufgelistet. Verwenden Sie diese, um Ihre eigenen Ergebnisse auf ihre Richtigkeit zu überprüfen.

Aufgabe 12.2 (Pumping Lemma, KFG – 6 Punkte (3 × 2))

Sei L wie folgt definiert:

$$L = \{w \in \{a, b\}^+ \mid w \neq vv \text{ für } v \in \{a, b\}^+\}$$

- (a) Zeigen Sie mittels Pumping-Lemma, dass L nicht regulär ist.
- (b) Zeigen Sie, dass die Sprache L jedoch kontextfrei ist, indem Sie eine kontextfreie Grammatik G angeben.
- (c) Beweisen Sie, dass die von G erzeugte Sprache $L(G)$ genau L ist.

Aufgabe 12.3 (Chomsky-Hierarchie – 4 Punkte (2 + 2))

Eine generative Chomsky-Grammatik wird in folgender Form angegeben:

(Nichtterminale, Terminale, Regelmenge, Startsymbol).

Ordnen Sie jede der folgenden Grammatiken in die Chomsky-Hierarchie ein und geben Sie die erzeugte Sprache an. Ist die erzeugte Sprache regulär oder nur kontextfrei? Beweisen Sie ihre Antwort.

- (a) $G_1 := (\{X, Y, Z\}, \{a, b, c\}, \{X \rightarrow aYZ, Y \rightarrow bZ, Y \rightarrow \epsilon, Z \rightarrow cY\}, X)$
- (b) $G_2 := (\{X, Y, Z\}, \{a, b, c\}, \{X \rightarrow bYZ, Y \rightarrow cX, Z \rightarrow a, Y \rightarrow c\}, X)$

Achtung: Die Voraussetzungen für die Zulassung zur Klausur umfassen die Teilnahme an den Übungen sowie die sinnvolle Bearbeitung der Übungsblätter. Die Übungsblätter sollen in Gruppen von zwei Studenten bearbeitet werden. Bitte schreiben Sie beide Namen, die Nummer Ihrer Übungsgruppe und den Namen Ihres Tutors auf den Lösungszettel.

Abgabe bis 11.15 Uhr in der Vorlesung oder Einwurf in die entsprechenden Briefkästen im Erdgeschoss von Gebäude 51.