

6. Übungsblatt - Abgabe: 8.12.2015

Aufgabe 6.1

Gegeben ist die Sprache $L_1 = \{ww^R \mid w \in \{a,b\}^+\}$ über dem Alphabet a, b , wobei w^R die Spiegelung von w ist, d.h. es enthält die Zeichen von w in umgekehrter Reihenfolge. L_1 enthält also alle (nichtleeren) Wörter, die vorwärts wie rückwärts gelesen gleich sind, z.B. *abaaaaba*.

Außerdem ist die Sprache $L_2 = \{w cw^R \mid w \in \{a,b\}^*\}$ gegeben. Wörter aus L_2 sind also z.B. *c*, *abcba*, *bbbaabacabaabbb*.

- Geben Sie für L_1 und L_2 jeweils einen möglichst intuitiven PDA, der ohne Kategoriensymbole auf dem Speicher auskommt.
- Geben Sie außerdem für L_1 und L_2 jeweils eine kontextfreie Grammatik an.
- Erzeugen Sie aus jeder dieser Grammatiken nach dem Schema auf den aktuellen Vorlesungsfolien jeweils einen PDA als Top-Down- und Bottom-Up-Parser.
- Geben Sie für jeden Ihrer erzeugten PDAs (also insgesamt 6) an, ob sie deterministisch sind, oder nicht.

Aufgabe 6.2

Gegeben ist die folgende Grammatik:

$S \rightarrow NP VP$
 $VP \rightarrow VP ADV$
 $VP \rightarrow geht$
 $NP \rightarrow Hans$
 $ADV \rightarrow heute \mid schnell$

Betrachten Sie den Satz „Hans geht schnell“.

- Warum kann ein Top-down-Parser (Tiefensuche mit Backtracking, wir probieren passende Regeln in der Reihenfolge aus, wie sie in der Grammatik stehen) mit dieser Grammatik den Satz nicht parsen? Wie lässt sich dieses Problem durch Anpassen der Grammatik beheben?
- Zeigen Sie, wie der Bottom-up-Parser (=Shift-Reduce-Parser) diesen Satz verarbeiten würde. Geben Sie für jeden Schritt den Stack und die verwendete Grammatikregel an. Geben Sie auch den dabei entstehenden Parsebaum an. (Nur einmal, Sie brauchen keine Zwischenergebnisse als Bäume anzugeben)

- (c) Stellen Sie sich einen Top-down-Parser vor, der zuerst immer das offene Nicht-Terminal ganz rechts expandiert, also von rechts nach links vorgeht und ansonsten wie in a) vorgeht, also auch die Regeln der Reihenfolge nach durchprobiert und Tiefensuche mit Backtracking betreibt. Parsen Sie den Satz mit dem so veränderten Algorithmus und der nicht-abgeänderten Grammatik. Geben Sie auch hier für jeden Schritt die benutzte Regel und den Stack an.

Aufgabe 6.3

Betrachten Sie den Earley-Algorithmus, wie er in der Vorlesung besprochen wurde (Er ist nicht ganz identisch mit dem Algorithmus im Jurafsky & Martin).

- (a) Benutzen Sie den Earley-Algorithmus, um beide Analysen für den folgenden Satz zu finden:

„Der Student sieht den Mann mit der Brille.“

Benutzen Sie folgende Grammatik:

$S \rightarrow NP VP$
 $VP \rightarrow V NP$
 $VP \rightarrow VP PP$
 $NP \rightarrow Det N$
 $NP \rightarrow NP PP$
 $PP \rightarrow P NP$
 $V \rightarrow \text{sieht}$
 $N \rightarrow \text{Mann}$
 $N \rightarrow \text{Student}$
 $N \rightarrow \text{Brille}$
 $Det \rightarrow \text{der}$
 $Det \rightarrow \text{den}$
 $P \rightarrow \text{mit}$

Geben Sie die Chart als Tabelle an (Achtung, es sind an die 60 einzelne Chart-einträge!).

Hinweis: Der selbe Eintrag braucht nicht mehrfach auf einer Chartposition abgelegt zu werden.

- (b) Geben Sie durch Verlinkung (Pfeile) der vollständigen Charteinträge die beiden Lesarten des Satzes an (vgl. die vorletzte Vorlesungsfolie). Benutzen Sie verschiedene Farben für die beiden Lesarten.
- (c) Stellen Sie für jede Lesart die Teilchart, die nur die vollständigen Einträge enthält, graphisch dar (d.h. mit „Bögen“ für jede Regel, die vollständig abgearbeitet wurde, vgl. letzte Vorlesungsfolie).

Aufgabe 6.4

Beim *Probabilistischen Parsing* haben Regeln in der Grammatik eine Wahrscheinlichkeit, mit der sie angewendet werden, gegeben ein bestimmtes nicht-Terminal auf der linken Seite der Regel. Die Wahrscheinlichkeiten schätzt man mithilfe eines großen Korpus ab. Wenn im Korpus beispielsweise 80% aller NPs zu Det N expandieren und 20% zu Det NE, dann nehmen wir an, dass auch die Wahrscheinlichkeit der Regel $NP \rightarrow Det N$ 80% sein soll und die für $NP \rightarrow Det NE$ 20%.

- (a) Benutzen Sie nun wie auf dem letzten Übungsblatt `tgrep2` und das `negra`-Corpus, um die Häufigkeiten für diese beiden Regeln abzuschätzen. Denken Sie daran, dass mit STTS annotiert wurde und Sie deswegen Kategoriensymbole entsprechend anpassen müssen. Geben Sie Ihre Befehle und die Ergebnisse an.
Tipp: Schauen Sie sich insbesondere Teile 4.3 und 4.6 aus dem `TGrep2` Manual an. In Abschnitt 4.6 wird beschrieben, wie Sie Knoten Labels geben können, um für eine Knoten mehrere Bedingungen angeben zu können.
- (b) Ermitteln Sie die Wahrscheinlichkeiten für die beiden Regeln aus a) und beschreiben Sie Ihr Vorgehen.

Abgabe in Gruppen von bis zu drei Studierenden am **8.12.2015** vor der Vorlesung.