

Einführung in die Computerlinguistik

Chart-Parsing

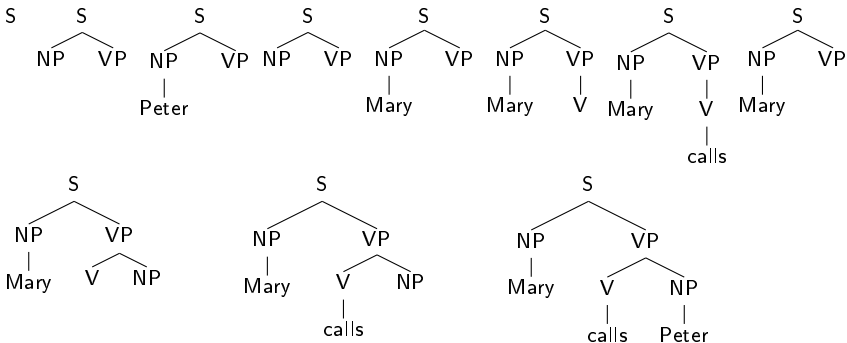
Dozentin: Wiebke Petersen

21.12.2009

Naives Parsen \Rightarrow redundante Berechnungen

$$P = \left\{ \begin{array}{l} S \rightarrow NP VP \quad VP \rightarrow V \quad VP \rightarrow V NP \\ V \rightarrow \text{calls} \\ NP \rightarrow \text{Peter} \quad NP \rightarrow \text{Mary} \end{array} \right\}$$

Mary calls Peter



Dynamische Programmierung

Ziel: Parsing-Methode,

- die einmal berechnete Teilbäume speichert und nicht erneut berechnet
- kein Problem mit Linksrekursion hat

⇒ Chart-Parser speichern Zwischenergebnisse in einer Chart

Chart-Parser: CYK (Cocke-Younger-Kasami)

- setzt voraus, dass die Grammatik in Chomsky-Normalform ist
- wird auch CKY-Parser genannt

Chomsky-Normalform

Definition 1

Eine Grammatik ist in **Chomsky-Normalform** (CNF), wenn alle Regeln die Gestalt

- 1 $A \rightarrow a$
- 2 $A \rightarrow BC$

mit $A, B, C \in T$ und $a \in \Sigma$ haben (und gegebenenfalls $S \rightarrow \epsilon$, dann aber ohne S in den rechten Regelseiten).

Theorem 2

Zu jeder kontextfreien Sprache gibt es eine Grammatik in Chomsky-Normalform, die sie generiert.

Was sind die charakteristischen Eigenschaften einer Grammatik in CNF?

Jede kontextfreie Sprache kann durch eine Grammatik in Chomsky-Normalform generiert werden

3 Umwandlungsschritte

- 1 Eliminierung der Regeln der Form $A \rightarrow B$.
Ersetze diese Regel durch alle Regeln $A \rightarrow \beta$, für die es eine Regel $B \rightarrow \beta$ gibt.
- 2 Anpassen der Grammatik, so dass Terminale nur in Regeln der Form $A \rightarrow a$ vorkommen.
Einfügen eines Dummy-Nichtterminalen für jeden Terminalen, der in einer falschen Regel steht.
- 3 Eliminierung der Regeln der Form $A \rightarrow B_1 B_2 \dots B_n$ mit $n \geq 3$.
Wiederholtes Einfügen eines Dummy-Nichtterminalen für jedes Cluster $B_2 \dots B_n$.

CYK-Parser

- Demo von Roy Mennicke
<http://www.informatik.uni-leipzig.de/alg/lehre/ss08/AUTO-SPRACHEN/Java-Applets/CYK-Algorithmus.html>
Wandelt Grammatiken in CNF um, beschreibt detailliert die Umwandlung und den CYK-Algorithmus.
- Demo von Martin Lazarov
<http://homepages.uni-tuebingen.de/student/martin.lazarov/demos/cky.html>
Erlaubt die Eingabe von Grammatiken mit komplexen Terminal- und Nichtterminalsymbolen (z.B., "Maria", "NP")
- Exorciser
Zeigt die implizit erzeugten Derivationsbäume an.