

Merkblatt Parsing

Prof. Dr. Udo Hahn Sven Büchel

Lehrstuhl für Computerlinguistik
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

28. März 2017

Inhaltsverzeichnis

1	Eine Beispielgrammatik	1
2	Top-Down-Parsing	2
3	Bottom-Up-Parsing	4
4	Left-Corner-Parsing	6

1 Eine Beispielgrammatik

Regeln:

$S \rightarrow NP VP$

$NP \rightarrow det n$

$VP \rightarrow v$

$VP \rightarrow v NP$

$det \rightarrow an \mid the$

$n \rightarrow elephant \mid man \mid shot$

$v \rightarrow shot$

Beispielsatz: *the man shot an elephant*

2 Top-Down-Parsing

Hier wird der Spezialfall des Links-Rechts-Parsings mit Tiefensuche vorgestellt.

Algorithmus

1. Erzeuge das Startsymbol der Grammatik
2. Wiederhole solange der Satz noch nicht fertig geparkt ist und Parsing nicht gescheitert ist:
 - (a) Wähle den Knoten zur Betrachtung der am weitestens links und unten im aktuellen Teil-Parsebaum liegt und kein Wortknoten ist.
 - (b) Wenn aus dem aktuellen Knoten unmittelbar nur Wortknoten folgen können, führe die Regel aus, die das nächste Wort der Eingabekette erzeugt, wenn es so eine Regel gibt. Wenn es diese Regel nicht gibt, backtracke bis zur letzten Entscheidung für eine Regel.
 - (c) Wenn aus dem aktuellen Knoten nicht direkt Wortknoten folgen, betrachte alle Regeln der Grammatik, die das Symbol des aktuellen Knotens auf der linken Seite haben. Führe davon, die oberste aus, die an dieser Stelle noch nicht ausgeführt worden ist.

Notation

- Zeichne für jeden hinzugefügten Knoten einen eigenen Teilparsebaum.
- Mache Backtracking explizit deutlich ('Backtracking', Blitzsymbol, o.ä.).
- Nummeriere die Teilbäume übersichtlich und konsistent durch. Achte besonders darauf, dass durch die Nummerierung beim Backtracking deutlich wird, zu welcher Stelle zurückgesprungen wurde und wie häufig schon Backtracking gemacht wurde (die Entscheidungspfade sollen klar nachvollziehbar sein). Empfehlung z.B.: 1, 2, 3, Backtracking, 2a, 3a, 4a, Backtracking, 3b,...
- Mache bei einem erfolgreichen Durchlauf des Algorithmus dieses explizit deutlich ("Erfolg!" oder Häkchen).

Beispielparse

<p>1 S</p>	<p>2 NP VP</p>	<p>S NP VP det N</p>
<p>S NP VP det N the</p>	<p>S NP VP det N the man</p>	<p>S NP VP det N v the man</p>
<p>S NP VP det N v the man shot ⚡</p>	<p>S NP VP det N v NP the man</p>	<p>S NP VP det N v NP the man shot</p>
<p>S NP VP det N v NP the man shot det n</p>	<p>S NP VP det N v NP the man shot det n an</p>	<p>S NP VP det N v NP the man shot det n an elephant ✓</p>

3 Bottom-Up-Parsing

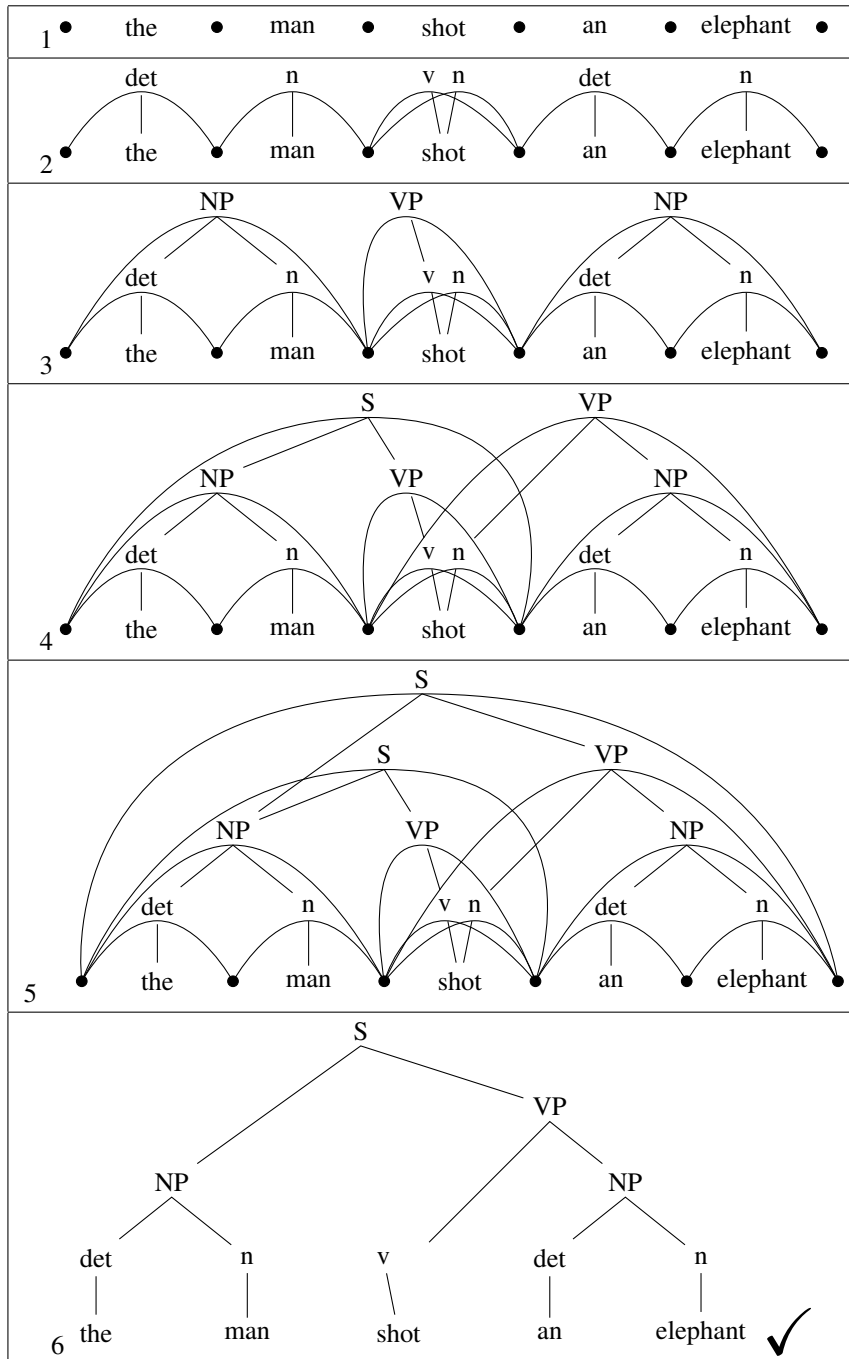
Algorithmus

1. Füge für jedes Wort der Eingabekette ein Knoten mit dem entsprechenden Symbol hinzu.
2. Finde für jedes Wort der Eingabeketten alle lexikalischen Regeln, bei denen das jeweilige Wort auf der rechten Seite steht. Erzeuge alle entsprechenden Nicht-Terminal-Symbole und verbinde sie mit den entsprechenden Terminalsymbolen. Der Abdeckungsbereich jeder dieser NTs ist jeweils das Wort, das es als direkten Tochterknoten hat.
3. Wiederhole:
 - (a) Finde alle Regeln, deren rechtsseitige Symbole so in den aktuellen Teilparseebäumen zu finden ist, dass ihre Abdeckungsbereiche unmittelbar aneinander anschließen und die Reihenfolge der Abdeckungsbereiche der Reihenfolge der Symbole in der Regel entspricht (dabei sind nicht nur die aktuell obersten Symbole in den Teilparseebäumen zu beachten!).
 - (b) Erzeuge für jede so gefundenen Regel das entsprechende linkseitige Symbol und verbinde alle dazugehörigen rechtsseitigen Symbole damit.
 - (c) Der Abdeckungsbereich der so erzeugten Symbole ist der Bereich der von irgendeinem der Töchterknoten abgedeckt wird.
 - (d) Wenn keine Regeln mehr gefunden werden, beende die Wiederholung.

Notation

- Beginne das erste Bild mit der Eingabekette.
- Zeichne ein neues Bild, jedes Mal, wenn Symbole (bzw. die dazugehörigen Kanten und Abdeckungsbereiche) zu den Teilparseebäumen hinzugefügt werden.
- Zeichne dabei alle diese Symbole "gleichzeitig" ein. Das heißt, anders als bei Top-Down- und Left-Corner-Parsing wird nicht mit jedem Bild bloß ein Knoten hinzugefügt, sondern ggf. mehrere.
- Zeichen die Abdeckungsbereiche als Bögen, die den abgedeckten Baum umspannen. Das Symbol der Wurzel dieses Baumes wird direkt oberhalb dieses Bogens gezeichnet.
- Füge nach Ende des Algorithmus noch ein letztes Bild ein, in dem du nur die Knoten und Kanten zeigst, die letztlich einen korrekten Parse dieses Satzes erzeugen (auch die Abdeckungsbereiche werden dafür nicht gezeichnet). Bei synkatischer Ambiguität zeichnen für alle korrekten, gefunden Parseebäume ein eigenes Bild.
- Markiere diese finalen Bäume mit "Erfolg!", einem Häkchen o.ä.

Beispielparse



4 Left-Corner-Parsing

Algorithmus

1. Die zu parsende Konstituente wird mit C bezeichnet (anfangs meistens das Startsymbol S).
2. Lies das erste Wort der Eingabekette ein, das noch nicht eingelesen worden ist. Dieses ist jetzt die aktuell betrachtete Konstituente K .
3. Wiederhole bis $C = K$ (Erfolg) oder das Parsing fehlschlägt (keine Regeln mehr gefunden werden können). Bei Fehlschlag Backtracking.
 - (a) Finde alle Regeln, die K ganz links auf der rechten Seite stehen haben (*left corner*). Wähle davon die oberste aus, die an dieser Stelle noch nicht angewandt worden ist.
 - (b) Erzeuge das linksseitige Symbol dieser Regel, dieses wird mit K' bezeichnet. Verbinde K mit K' .
 - (c) Erzeuge alle anderen rechtseitigen Symbolen der Regel, verbinde diese mit K' . Für jedes dieser Symbole, wende Left-Corner-Parsing rekursiv an, um die entsprechende Konstituente zu parsen.
 - (d) K' wird fortan mit K bezeichnet.

Notation

- Zeichne ein eigenes Bild für jedes neue Nicht-Terminal-Symbol, das erzeugt wird (für die Terminalsymbole müsse keine eigenen Bilder gezeichnet werden).
- Die rekursiv erzeugten Teilparsebäume sind zunächst unverbunden vom "Haupt"-Parsebaum daneben oder darunter zu schreiben. Erst wenn in diesem Nebenparsebaum das benötigte Symbol erzeugt wurde, werden beide Parsebäume in einem separaten Bild zusammengefügt.
- Genau wie bei Top-Down-Parsing soll Backtracking und Erfolg explizit deutlich gemacht werden. Auch die Nummerierung der Bilder bei Backtracking unterscheidet sich nicht vom Top-Down-Parsing.

Beispielparse

<p style="text-align: center;">det 1 the</p>	<p style="text-align: center;">NP / \ det n 2 the</p>
<p style="text-align: center;">NP / \ / \ det n n 3 the man</p>	<p style="text-align: center;">NP / \ det n 4 the man</p>
<p style="text-align: center;">S / \ NP VP / \ / \ det n 5 the man</p>	<p style="text-align: center;">S / \ NP VP / \ / \ det n n 6 the man shot ⚡</p>
<p style="text-align: center;">S / \ NP VP / \ / \ det n v 6a the man shot</p>	<p style="text-align: center;">S / \ / \ NP VP VP / \ / \ det n v 7a the man shot</p>
<p style="text-align: center;">S / \ NP VP / \ / \ det n v 8a the man shot ⚡</p>	<p style="text-align: center;">S / \ / \ NP VP VP / \ / \ det n v NP 7b the man shot</p>

(Fortsetzung auf der nächsten Seite.)

