

Probeklausur

Aufgabe 1

- (a) Nennen Sie ein Beispiel für ein Phänomen der menschlichen Sprache, das mit endlichen Automaten nicht modelliert werden kann und erläutern Sie, wieso.
- (b) Lexikalische Information in Sprachverarbeitungssystemen kann alternativ in Wortformenlexika und in Stammlexika bereitgestellt werden. Wieso ist es für das Deutsche wichtiger als für das Englische, mit einem Stammlexikon zu arbeiten?
- (c) Geben Sie anhand von einfachen Merkmalsstrukturen (abstrakt oder mit konkreten grammatischen Merkmalen) je ein Beispiel für eine erfolgreiche Unifikation (mit Resultat) und einen Fehlschlag.
- (d) Die Kontinuität und die Varianz des Schallsignals sind Probleme für die Spracherkennung. Geben Sie für Kontinuität und Varianz jeweils zwei unterschiedliche Aspekte an.
- (e) Spracherkenner, die in Dialogsystemen eingesetzt werden, sind immer noch ungenau. Nennen Sie zwei Strategien, wie ein Dialogsystem damit umgeht.
- (f) Wann ist Kontextwissen bei der Rechtschreibkorrektur nötig? Erläutern Sie an zwei Beispielen.
- (g) Nennen Sie drei potentielle Probleme, die ein maschinelles Übersetzungsprogramm (z.B. ins Englische) mit dem folgenden Satz haben kann: *Das Mädchen las seiner Mutter auf der Bank vor dem Haus aus dem Buch vor.*
- (h) Setzen Sie die folgenden Begriffe durch die passenden semantischen Relationen miteinander in Beziehung: *Musikinstrument, Geige, Saite, Streichinstrument, Bratsche, Trompete.*
- (i) Geben Sie die Abhängigkeitsstruktur für die NP *der große dicke Baum hinter dem Haus* an. Denken Sie an die Kantenbeschriftungen.

Aufgabe 2

Sie haben zu Beginn der Vorlesung die beiden „klassischen“ Sprachverarbeitungssysteme ELIZA und SHRDLU kennengelernt. Der wesentliche Unterschied zwischen beiden Systemen ist, dass SHRDLU im Gegensatz zu ELIZA wissensbasiert arbeitet. Erläutern Sie diese Feststellung an je einem Beispiel (d.h. einer Dialogsequenz mit System- und Benutzeräußerung) zu den beiden Systemen (Sie können die Beispiele frei, und falls Ihnen das lieber ist, auch in deutscher Sprache, wählen).

Aufgabe 3

In folgenden Beispielen sind abgetrennte Verbzusätze (VZ) hervorgehoben:

Det N V VZ

Der Zug fährt **ab**

PN V Det N VZ

Peter sammelt die Teller **ein**

Det N V Det N Det N VZ

Der Trainer teilt den Spielern die Positionen **zu**

Entwerfen Sie einen endlichen Automaten, der einfache Sätze mit abgetrennten Verbzusätzen erkennt. Akzeptiert werden sollen Folgen der Art: Nominalausdruck - Verb - beliebig viele Nominalausdrücke - Verbzusatz. Nominalausdrücke sollen immer aus einem Eigennamen oder einem Artikel gefolgt von einem Nomen bestehen. Geben Sie einen zweiten Automaten an, der zwischen Verb und Verbzusatz maximal zwei Nominalausdrücke akzeptiert, sich ansonsten aber wie der erste Automat verhält.

Aufgabe 4

Gegeben sei der folgende NEA:

$A = \langle \{1, 2, 3\}, a, b, \Delta, 1, \{2, 3\} \rangle$ mit

$\Delta = \langle 1, ba, 1 \rangle, \langle 1, b, 2 \rangle, \langle 1, b, 3 \rangle, \langle 2, bab, 3 \rangle, \langle 3, \epsilon, 2 \rangle$

- Beschreiben Sie informell die Sprache, die der Automat akzeptiert.
- Konstruieren Sie nach dem in der Vorlesung eingeführten dreistufigen Verfahren einen zu A äquivalenten DEA. Geben Sie dabei insbesondere auch die Übergangstabelle und den resultierenden DEA an.

Aufgabe 5

Gegeben ist das Wort „Abteilungen“.

- Geben Sie zwei Zerlegungen des Wortes an, die ein Morphologiesystem erkennen könnte.
- Benennen Sie die einzelnen Morpheme (Stamm, Flexionsendung,...).
- Beschreiben Sie die morphologischen Prozesse, denen die Zerlegungen jeweils entsprechen (Flexions-Morphologie, Ableitungs-/Derivations-Morphologie oder Kompositazerlegung).

Aufgabe 6

Gegeben sei eine kontextfreie Grammatik mit den folgenden Produktionsregeln:

$S \rightarrow NP VP$ $V \rightarrow$ spielte, gähnte
 $VP \rightarrow V NP$ $N \rightarrow$ Musiker, Sinfonie, Paukenschlag, Klavier
 $VP \rightarrow V$ $Prp \rightarrow$ auf, mit
 $VP \rightarrow VP PP$ $Det \rightarrow$ der, dem, den, die, der
 $NP \rightarrow NP PP$
 $PP \rightarrow Prp NP$
 $NP \rightarrow Det N$

- (a) Leiten Sie den Satz *Der Musiker spielte die Sinfonie mit dem Paukenschlag.* auf zwei unterschiedliche Weisen ab. Geben Sie dazu die beiden Strukturbäume an, die aus der Ableitung resultieren.
- (b) Beschreiben Sie alle grammatischen Merkmalsabhängigkeiten, die in diesem Satz vorkommen, informell (explizite Merkmalsstrukturen brauchen nicht angegeben werden).
- (c) Kann ein Top-Down Parser, der die Regeln in der angegebenen Reihenfolge ausprobiert, Probleme mit Linksrekursion bekommen? Begründen Sie.

Aufgabe 7

Stellen Sie durch Angabe der einzelnen (Teil-)Bäume und des zugehörigen Stacks dar, wie ein Top-Down-Parser den Satz *Der Musiker spielte* in der kürzestmöglichen Weise abarbeitet (das heißt, jeweils zielführende Regeln der Grammatik in Aufgabe 6 anwendet).

Aufgabe 8

Beschreiben Sie die Funktionen von Predictor und Scanner im Earley-Algorithmus kurz und informell. Geben Sie jeweils ein Beispiel für die Funktionsweise.

Aufgabe 9

Kommentieren Sie, welche besonderen Schwierigkeiten der folgende Satz für ein TTS-System macht. Nennen Sie 4 Punkte.

Bei der Bundestagswahl am 27.9.2009 erhielt die CDU im Wahlkreis Coesfeld ca. 48% der Zweitstimmen.

Aufgabe 10

Die folgende Tabelle zeigt die Evaluation eines Klassifikators für die Erkennung von Fehlern in einem Rechtschreibkorrektursystem.

	echter Fehler	echtes korrektes Wort
klass. als Fehler	40	60
klass. als korrektes Wort	10	390

- (a) Der F-Score für die Klasse „korrektes Wort“ beträgt ca. 0,92. Bedeutet das, dass der Klassifikator besonders gut ist? Begründen Sie.
- (b) Berechnen Sie den F-Score für die Klasse Fehler.
- (c) Halten Sie es bei diesem System für wichtiger für die Klasse *Fehler* auf Präzision oder auf Recall zu optimieren? Begründen Sie.

Aufgabe 11

Mit Hilfe von Kontextwörtern soll ein Klassifikator zur Word-Sense-Disambiguierung für die beiden Lesarten von *Ball* gelernt werden. Das Trainingskorpus besteht aus 50 Dokumenten für $Ball_1$ und 50 Dokumenten für $Ball_2$.

Dabei ermittelt man die folgenden Kontextwortfrequenzen und beobachtet danach ein Auftreten von *Ball* in dem Satz „Nach dem ersten Tanz warf sie Ihre Schuhe in die Ecke und verließ den Ball“.

	Ball ₁	Ball ₂	???
Tor	33	2	
werfen	25	5	
Tanz	7	46	
Schuh	12	10	

Außerdem gelten a-priori-Wahrscheinlichkeiten von $P(Ball_1) = 0,9$ und $P(Ball_2) = 0,1$. Ergänzen Sie das Merkmalsmuster für das beobachtete Auftreten von *Ball* und berechnen Sie die wahrscheinlichere Lesart. Nehmen Sie an, dass ein Stemmer verwendet wird.

Sie müssen die Probeklausur nicht abgeben. Sie wird am nächsten Dienstag in der Vorlesung besprochen.