

Probeklausur - Abgabe: 07.02.2011

Aufgabe 1

- a) Nennen Sie ein Beispiel für ein Phänomen der menschlichen Sprache, das mit endlichen Automaten nicht modelliert werden kann und erläutern Sie es.
- 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) Welchen Zweck erfüllt das Grounding in einem Dialogsystem? Nennen Sie zwei verschiedene Groundingtypen mit Beispiel.
- f) Sollte ein statistisches Sprachverarbeitungssystem auf hohe Präzision oder hohen Recall optimiert werden?
- g) Was ist der Unterschied zwischen dem Transfer- und dem Interlingua-Modell in der maschinellen Übersetzung? Nennen Sie Vor- und Nachteile.
- h) Nennen Sie drei potentielle Probleme, die ein maschinelles Übersetzungsprogramm (z.B. ins Englische) mit dem folgenden Satz haben kann: *Das Mädchen nahm vor dem Tor von seinem Vater Abschied.*
- i) Setzen Sie die folgenden Begriffe durch die passenden semantischen Relationen miteinander in Beziehung: *Musikinstrument, Geige, Saite, Streichinstrument, Bratsche, Trompete.*

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:

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

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

Konstruieren Sie nach dem in der Vorlesung eingeführten dreistufigen Verfahren einen zu A äquivalenten DEA.

Aufgabe 5

Geben Sie zwei Zerlegungen des Wortes „Abteilungen“ an, die ein Morphologiesystem erkennen könnte. 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 → NP VP	V → spielte, gähnte
VP → V NP	N → Musiker, Sinfonie, Paukenschlag, Klavier
VP → V	Prp → auf, mit
VP → VP PP	Det → der, dem, den, die, der
NP → NP PP	
PP → Prp NP	
NP → Det N	

- Leiten Sie den Satz *Der Musiker spielte die Sinfonie mit dem Paukenschlag.* auf zwei unterschiedliche Weisen ab. Geben Sie die beiden Strukturbäume an, die aus der Ableitung resultieren.
- Beschreiben Sie alle grammatischen Merkmalsabhängigkeiten, die in diesem Satz vorkommen, informell (explizite Merkmalsstrukturen brauchen nicht angegeben werden).

Aufgabe 7

Stellen Sie (durch Angabe der einzelnen Bäume) 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:

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

Aufgabe 10

Kontextvektoren sind eine Möglichkeit, die semantische Ähnlichkeit von Wörtern zu ermitteln. In einem Corpus wurden die folgenden Werte für das gemeinsame Auftreten der Wörter *Scheck*, *Bank* und *Tisch* mit jedem von zwei Kontextwörtern ermittelt.

	Geld	Garten
Scheck	10	150
Bank	180	280
Tisch	80	20

(a) Berechnen Sie die paarweise Cosinus-Ähnlichkeit der drei Wörter.

$$\text{(Formel: } sim_{cos}(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \text{)}$$

(b) Veranschaulichen Sie die semantische Ähnlichkeit der drei Begriffe in einem Koordinatensystem.

(c) Welches generelle Problem dieses semantischen Ähnlichkeitsmaßes wird hier deutlich?

Abgabe (verbindlich nur für diejenigen, die noch Punkte brauchen) bis **07.02.2011** um 18 Uhr entweder als Email im pdf-Format an i2cl@coli.uni-sb.de oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.