

## Übungsklausur

1.
  - a) Die Potenzautomatenkonstruktion erzeugt aus einem buchstabierenden NEA mit  $n$  Zuständen einen DEA mit  $2^n$  Zuständen. Der Potenzautomat für einen NEA mit 10 Zuständen hat also 1024 Zustände. Trotzdem kommt man beim Aufschreiben des DEA meist mit einer weit kleineren Zustandsmenge aus. 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) Was ist der Unterschied zwischen einem Referenzkorpus und einem Monitorkorpus?
  - f) Cosinus und Euklidische Distanz sind zwei Maße für die Ähnlichkeit zweier Dokumente in einem Vektorraum-Modell. Welches der Maße wird im Information Retrieval bevorzugt verwendet und warum?
  - g) Sollte ein statistisches Sprachverarbeitungssystem auf hohe Präzision oder hohen Recall optimiert werden?
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).
3. In folgenden Beispielen sind abgetrennte Verbzusätze (VZ) hervorgehoben:

Det N V VZ	<i>Der Zug fährt ab</i>
PN V Det N VZ	<i>Peter sammelt die Teller ein</i>
Det N V Det N Det N VZ	<i>Der Trainer teilt den Spielern die Positionen zu</i>

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.

4. Gegeben sei der folgende NEA:

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

$\Delta = \{ \langle 1, ba, 1 \rangle, \langle 1, b, 2 \rangle, \langle 1, b, 3 \rangle, \langle 2, bab, 3 \rangle, \langle 3, \varepsilon, 2 \rangle \}$

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

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).

6. Gegeben sei eine kontextfreie Grammatik mit den 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		

- a) 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.
- b) Beschreiben Sie alle grammatischen Merkmalsabhängigkeiten, die in diesem Satz vorkommen, informell (explizite Merkmalsstrukturen brauchen nicht angegeben werden).
7. Zeigen Sie die Folge von Inhalten des Kellers an, wenn ein Top-Down-Parser den Satz *Der Musiker spielte* in der kürzestmöglichen Weise abarbeitet (das heißt, jeweils zielführende Regeln der Grammatik in 6 anwendet).
8. Beschreiben Sie die Funktionen von Predictor und Scanner im Earley-Algorithmus kurz und informell. Geben Sie jeweils ein Beispiel für die Funktionsweise.
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.*
10. Welche grundsätzlich unterschiedlichen Schichten von Bedeutungsinformation können beim Wort unterschieden werden? Nennen Sie 3 solcher Bedeutungskomponenten und erläutern Sie sie an einem Beispiel, nämlich der Bezeichnung für ein Musikinstrument Ihrer Wahl.

**Hinweis: Bonus-Punkte für fehlende Punkte bei den Übungsaufgaben werden nur für die Übungen 3-7 vergeben.**