

Probeklausur

Besprechung der Musterlösung in den Übungen am 6.2.

Diese Probeklausur dient zur Klausurvorbereitung und soll einen Eindruck von möglichen Aufgabentypen vermitteln. Bitte beachten Sie dass alle Inhalte aus den Vorlesungen, Übungsaufgaben und Lese-Aufgaben (siehe Übungsblätter) zum Prüfungsstoff gehören. Die Klausur kann also durchaus Aufgaben zu Themen enthalten, die in der Musterlösung nicht angesprochen werden.

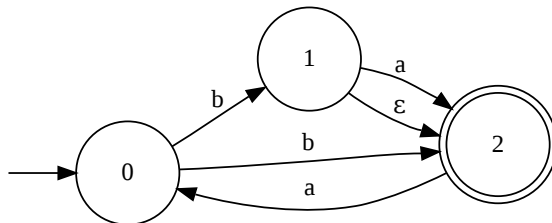
Aufgabe 13 kann (**optional!**) bis zum 6.2. als Übungsaufgabe eingereicht werden (wie üblich entweder in gedruckter Form oder per e-Mail) um zusätzliche Übungspunkte zu erwerben.

1. 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).
2. Beantworten sie kurz die folgenden Fragen zu Korrekturprogrammen:
 - Warum ist eine lexikonbasierte Rechtschreibkorrektur (d.h. es wird angenommen dass jedes Wort das nicht genau so im Lexikon steht ein Nicht-Wort ist) für Sprachen wie z.B. Deutsch nicht ausreichend?
 - Nachdem in einem Korrekturprogramm ein Nicht-Wort erkannt wurde, soll dieses durch das korrigierte Wort ersetzt werden. Welche Methode kennen sie um das richtige Wort auszuwählen oder dem Benutzer mögliche Ersetzungen vorzuschlagen.
 - Warum ist im allgemeinen eine kontextabhängige Korrektur wünschenswert?
3. In folgenden Beispielen sind abgetrennte Verbzusätze (VZ) hervorgehoben:

Det N V VZ	Der Zug fährt ab
Det N V Det N VZ	Der Kellner sammelt die Teller ein
Det N V Det N Det N VZ	Der Trainer teilt den Spielern die Positionen zu

- a) Entwerfen Sie ein Zustandsdiagramm, das einfache Sätze mit abgetrennten Verbzusätzen erkennt. Akzeptiert werden sollen Folgen der Art:
Nominalausdruck - Verb - beliebig viele Nominalausdrücke - Verbzusatz
Nominalausdrücke sollen immer aus genau einem Artikel gefolgt von einem Nomen bestehen.
- b) Geben Sie ein zweites Zustandsdiagramm an, das zwischen Verb und Verbzusatz maximal zwei Nominalausdrücke akzeptiert, ansonsten aber wie das erste Diagramm ist.

4. Betrachten Sie den folgenden NEA A_1 über dem Alphabet $\Sigma = \{a, b\}$:



- a) Konstruieren Sie mit den in der Vorlesung vorgestellten Verfahren:
 - i. einen zu A_1 äquivalenten buchstabierenden Automaten.
 - ii. durch Potenzautomatenkonstruktion einen DEA, der die gleiche Sprache wie A_1 erkennt.
 - b) Beschreiben Sie in Worten die Sprache, die der Automat erkennt.
5. 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?
6. Geben Sie zwei Zerlegungen des Wortes *“Abteilungen”* an, die ein automatenbasiertes Morphologiesystem erkennen könnte. Beschreiben sie die morphologischen Prozesse denen die Zerlegungen jeweils entsprechen (Flexions-Morphologie, Ableitungs-/Derivations-Morphologie oder Kompositazerlegung).
7. Gegeben sei eine kontextfreie Grammatik mit den Produktionsregeln
- | | | | | | |
|----|---|--------|-----|---|----------------------|
| S | → | NP VP | V | → | trifft |
| VP | → | V NP | PN | → | Maria Peter Hans |
| VP | → | VP PP | PRP | → | mit |
| NP | → | PN | | | |
| NP | → | NP PP | | | |
| PP | → | PRP NP | | | |
- a) Leiten Sie den Satz *“Peter trifft Hans mit Maria”* auf zwei unterschiedliche Weisen ab. Geben Sie die beiden Strukturbäume an, die aus der Ableitung resultieren und charakterisieren Sie beide Lesarten umgangssprachlich.
 - b) Zeigen Sie wie der Earley-Algorithmus den Satz verarbeitet. Geben Sie die Chart an, sowie die Schritte, die angewendet wurden.
8. a) Was ist der Unterschied zwischen syntaktischer Kategorie und Funktion?
 b) Wählen Sie aus dem folgenden Satz zwei Konstituenten und geben Sie jeweils deren Kategorie und Funktion an.
- “Der Forscher erfindet ein neues Mittel gegen Müdigkeit”*
9. Betrachten Sie alle möglichen Paarungen der folgenden Merkmalsstrukturen:

- a) $\left[\begin{array}{c} A \left[\begin{array}{c} B \ a \\ C \ \boxed{1} \ b \end{array} \right] \\ B \ \boxed{1} \end{array} \right]$ b) $\left[\begin{array}{c} B \ b \\ A \ \left[\begin{array}{c} C \ a \end{array} \right] \end{array} \right]$ c) $\left[\begin{array}{c} A \ \left[\begin{array}{c} B \ \boxed{1} \end{array} \right] \\ C \ \left[\begin{array}{c} \boxed{1} \ a \end{array} \right] \end{array} \right]$ d) $\left[\begin{array}{c} A \ \left[\begin{array}{c} B \ a \end{array} \right] \end{array} \right]$

a) Welche Paare sind konsistent/unifizierbar, welche Paare nicht (✓ oder ✗)?

	a)	b)	c)	d)
a)				
b)				
c)				
d)				

b) Zwischen welchen Paaren besteht die Subsumptionsrelation \sqsubseteq und in welcher Richtung?

10. Gegeben sei eine Grammatik mit Merkmalsconstraints, sowie ein paar lexikalische Einträge.

- S → NP VP
 < S HEAD = VP HEAD >
 < S HEAD SUBJ = NP HEAD >
 VP → V NP
 < VP HEAD = V HEAD >
 < V SUBCAT = NP HEAD >

- Gänse: $\left[\begin{array}{c} \text{CAT} \quad \text{NP} \\ \text{HEAD} \left[\begin{array}{c} \text{AGR} \left[\begin{array}{c} \text{NUM} \quad \text{pl} \end{array} \right] \\ \text{SEM} \quad \text{gänse} \end{array} \right] \end{array} \right]$ Maria: $\left[\begin{array}{c} \text{CAT} \quad \text{NP} \\ \text{HEAD} \left[\begin{array}{c} \text{AGR} \left[\begin{array}{c} \text{NUM} \quad \text{sg} \end{array} \right] \\ \text{SEM} \quad \text{maria} \end{array} \right] \end{array} \right]$
- sieht: $\left[\begin{array}{c} \text{CAT} \quad \text{V} \\ \text{HEAD} \left[\begin{array}{c} \text{SUBJ} \left[\begin{array}{c} \text{AGR} \left[\begin{array}{c} \text{NUM} \quad \text{sg} \end{array} \right] \\ \text{SEM} \ \boxed{1} \end{array} \right] \\ \text{SEM} \left[\begin{array}{c} \text{PRED} \quad \text{sehen} \\ \text{ARG1} \ \boxed{1} \\ \text{ARG2} \ \boxed{2} \end{array} \right] \end{array} \right] \end{array} \right]$
 SUBCAT $\left[\begin{array}{c} \text{SEM} \ \boxed{2} \end{array} \right]$

Leiten Sie mit dieser Grammatik den Satz "Maria sieht Gänse" ab, geben Sie insbesondere die Merkmalsstruktur für den gesamten Satz an.

11. Erklären Sie den Unterschied zwischen Transfer- und Interlingua-Modell in der maschinellen Übersetzung. Nennen Sie jeweils Vor- und Nachteile.
12. Geben Sie **jeweils zwei** Beispiele für folgende von WordNet erfasste Relationen. Machen Sie (wo sinnvoll) deutlich, in welche Richtung die Relation besteht (Sie können dafür deutsche Bezeichnungen verwenden).

- a) Hypernymie-Hyponymie
- b) Holonymie-Meronymie
- c) Synonymie

13. Wörter können in einem Vektorraummodell durch die Verteilung benachbarter Wörter repräsentiert werden. Unter <http://clg.wlv.ac.uk/demos/similarity/> ist ein solches Modell online zu testen.

Lassen Sie die nächsten Nachbarn zu *fünf* Wörtern ausgeben (wählen Sie den Cosinus als Ähnlichkeitsmaß). Versuchen Sie, für *jeweils drei*¹ interessante ausgegebene Wörter die Relation zu finden, in der sie zum eingegebenen Wort stehen (z.B. *ist-ein*, *Oberbegriff-von*, *Teil-von*,...). In welchen Fällen ist es schwierig, eine passende Relation zu finden?

¹insgesamt werden also 15 Paare betrachtet