

Übungsklausur

1. In der Vorlesung wurden "Sprachassistentz", "Dialog" und "Informationsmanagement" als drei wichtige Gruppen sprachtechnologischer Anwendungssysteme genannt. Geben Sie für jede der drei Gruppen zwei möglichst unterschiedliche Beispiele an, und zwar nicht Systeme, sondern unterschiedliche Anwendungen: Also z.B. nicht "LIFT", sondern "Dialogschnittstelle zur Geräte- bzw. Fahrstuhlsteuerung".
2. Lexikalische Information in Sprachverarbeitungssystemen kann alternativ in Wortformenlexika und in Stammlexika bereitgestellt werden. Was sind Vorteile und Nachteile? Bitte kurz und stichwortartig beantworten! Wieso ist es für das Deutsche wichtiger als für das Englische, mit einem Stammlexikon zu arbeiten?
3. 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 (System-Benutzer-Dialogsequenz) zu den beiden Systemen (Sie können die Beispiele frei, und falls Ihnen das lieber ist, auch in deutscher Sprache, wählen).
4. Eine naive Methode, um im Internet Dokumente zu identifizieren, die für eine Anfrage relevant sind, ist zu testen, ob Worte aus der Anfrage im Dokumente vorkommen. Beschreiben Sie kurz, wie Information Retrieval (IR) diese Methode verfeinert und mindestens zwei Vorteile der Verfeinerung.
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. Beschreiben Sie kurz (ggf. mit Beispielen), in welcher Hinsicht die syntaktische Struktur natürlicher Sprachen komplexer ist als die morphologische Struktur.
7. Geben Sie einen endlichen Automaten an, der die Sprache L über dem Alphabet $\{a,b\}$ mit $L = \{w|w \text{ enthält höchstens ein } a \text{ und mindestens ein } b\}$ beschreibt !
8. Gegeben sei der folgende NEA:

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

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

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

9. Gegeben sei eine kontextfreie Grammatik mit den Produktionsregeln

S	-> S PP	V	-> verfolgte, schlief
S	-> NP V NP	N	-> Detektiv, Gangster, Motorrad, Sonnenbrille
S	-> NP V	Prp	-> auf, mit
NP	-> NP PP	Det	-> der, dem, den
PP	-> Prp NP		
NP	-> Det N		

Leiten Sie den Satz *Der Detektiv verfolgte den Gangster mit der Sonnenbrille.* auf zwei unterschiedliche Weisen ab, indem Sie die beiden Strukturbäume angeben, die aus der Ableitung resultieren.

10. Leiten Sie mit der Grammatik aus 9. zwei Wortfolgen ab, die keine korrekten Sätze des Deutschen sind. Kommentieren Sie, gegen welche grammatischen Regeln dabei verstoßen wird.
11. Erklären Sie kurz, was Rauschen (Noise) in einem Korpus ist, und weshalb es ein Problem darstellt. Geben Sie ein konkretes Beispiel.
12. Was heißt es, eine probabilistische Grammatik zu lexikalisieren? Welche Vorteile hat man dadurch? Sie können zur Illustration z.B. die Verben *husten* (intransitiv) und *vergessen* (transitiv) verwenden. Was für Probleme von Baumbank-Grammatiken werden durch Lexikalisierung verschärft?
13. Wie heißen die zwei Maße, die zur Evaluation von automatischen Klassifikationen verwendet werden? Wieso genügt es nicht, nur eines der beiden Maße zu betrachten?
14. Geben Sie eine kontextfreie Grammatik an, die arithmetische Gleichungen der Form

$$\begin{aligned}x &= y, \\ (x + y) &= (z : x), \\ (x * (y - (-z))) &= ((-x) * (-y))\end{aligned}$$

erzeugt. Das heißt, Terminalsymbole sind $x, y, z, =, +, -, *, :, ($ und $)$, und alle komplexen Terme werden in Klammern geschrieben. Verwenden Sie als nicht-terminale Symbole S (Gleichung), T (Term), O (Operator).