

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. - Nicht Systeme, sondern unterschiedliche Anwendungen: Also z.B. nicht "LIFT", sondern "Dialogschnittstelle zur Geräte- bzw. Fahrstuhlsteuerung".
2. Juravsky/Martin beschreiben das fünffach mehrdeutige Beispiel "I made her duck". Welche Arten von Mehrdeutigkeit spielen hier mit? Bitte nicht die Lesarten aufzählen, sondern (mindestens 3) Quellen für die Mehrdeutigkeit angeben (mit ganz kurzer Beschreibung).
3. 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?
4. 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).
5. Was sind die Grundprobleme bei der automatischen Erkennung gesprochener Sprache (im Unterschied zur Verarbeitung von Texteingaben)? Bitte nennen sie, möglichst jeweils in einem Satz, drei Aspekte.
6. Welche grundsätzlich verschiedenen Möglichkeiten gibt es, um die Sprachausgabe von Dialogsystemen zu realisieren? Bitte zwei Fälle angeben! Was sind die grundsätzlichen Vorzüge/Nachteile?
7. Welche Struktureigenschaften von Zustandsdiagrammen sind für den Nicht-Determinismus von NEAs verantwortlich?

8. 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?
9. Ein Lemmatisierer kann zur Wortartidentifikation verwendet werden: Ein Morphologie-Automat identifiziert mögliche Substantiv-, Verb-, Adjektivendungen. Wenn der verbleibende Stamm als Substantiv, Verb bzw. Adjektiv im (Stamm-) Lexikon vorkommt, handelt es sich um eine Substantiv-, Verb- bzw. Adjektivform. Statistische Wortart-Tagger dienen ebenfalls der Zuweisung von Wortarten. Welche Vorteile besitzen sie gegenüber Lemmatisierern?
10. Beschreiben Sie kurz (ggf. mit Beispielen), in welcher Hinsicht die syntaktische Struktur natürlicher Sprachen komplexer ist als die morphologische Struktur.
11. Geben Sie einen endlichen Automaten an, der die Sprache L über dem Alphabet $\{a,b\}$ mit $L = \{w \mid w \text{ enthält mindestens ein } a \text{ und höchstens ein } b\}$ beschreibt !
12. Geben Sie einen endlichen Automaten an, der die starken Substantiv-Endungen im Deutschen akzeptiert (z.B. für "Schaf" oder "Tisch").
13. 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!

14. Gegeben sei eine kontextfreie Grammatik mit den Produktionsregeln

$S \rightarrow S PP$	$V \rightarrow \text{verfolgte, schlief}$
$S \rightarrow NP V NP$	$N \rightarrow \text{Detektiv, Gangster, Motorrad, Sonnenbrille}$
$S \rightarrow NP V$	$Prp \rightarrow \text{auf, mit}$
$NP \rightarrow NP PP$	$Det \rightarrow \text{der, dem, den}$
$PP \rightarrow Prp NP$	
$NP \rightarrow Det N$	

Leiten Sie den Satz *Der Detektiv verfolgte den Gangster auf dem Motorrad*

auf zwei unterschiedliche Weisen ab. (Geben Sie die unterschiedlichen Strukturbäume an, die aus der Ableitung resultieren.)

15. Leiten Sie mit den Regeln aus 15. zwei Wortfolgen ab, die keine korrekten Sätze des Deutschen sind. Kommentieren Sie, gegen welche grammatischen Regeln dabei verstoßen wird.
16. Geben Sie eine kontextfreie Grammatik an, die arithmetische Gleichungen der Form $x=y$, $(x+y)=(z:x)$, $(x*(y-(-z)))=(((-x)*(-y)))$ 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 (zweistelliger Operator).