

## Probeklausur

1. Was sind die Aufgaben der Computerlinguistik?

Theoretische Computerlinguistik entwirft, implementiert und untersucht die Modelle mit dem Ziel, zum Verständnis, zur Verifikation und zur Verbesserung der zugrundeliegenden linguistischen und psychologischen Theorien beizutragen.

Angewandte Computerlinguistik entwirft, implementiert und untersucht die Modelle mit dem Ziel, Softwareanwendungen zu ermöglichen, die über eine (eingeschränkte) Beherrschung menschlicher Sprache verfügen.

2. Welche Ebenen der Sprachverarbeitung müssen bei einem Diktiersystem durchlaufen werden?

Das konstante Sprachsignal muss in Phoneme umgewandelt werden. Gemäß dieser Phoneme werden dann die Wörter in dem Lexikon nachgeschlagen und als Text angezeigt.

3. Welche Ebenen der Sprachverarbeitung müssen bei der maschinellen Textübersetzung durchlaufen werden?

Der Ursprungstext muss geparkt werden. Dann findet der Transfer der Wörter statt – diese werden dann mit den Regeln für die Zielsprache zu Sätzen zusammengefügt.

4. Konstruieren Sie einen Satz mit mindestens sechs Lesarten. Erläutern Sie die Lesarten.

Das Mädchen in dem Kleid, **das weiß war**, suchte den **Ball mit der Taschenlampe**.

Der Relativsatz kann sich entweder auf das Mädchen oder das Kleid beziehen. Ball ist zweideutig: zum einen kann es ein Spielzeug sein (der Fußball) oder ein

gesellschaftliches Ereignis (der Tanzball). Die Präpositionalphrase kann entweder an die NP „den Ball“ oder die VP „suchte den Ball“ angefügt werden.

weißes Mädchen + Spielzeug + das Mädchen hat die Taschenlampe

weißes Mädchen + Spielzeug + der Ball hat die Taschenlampe (als Abbildung)

weißes Mädchen + Ereignis + das Mädchen hat die Taschenlampe

weißes Mädchen + Ereignis + der Ball hat die Taschenlampe (als Abbildung)

weißes Kleid + Spielzeug + das Mädchen hat die Taschenlampe

weißes Kleid + Spielzeug + der Ball hat die Taschenlampe (als Abbildung)

weißes Kleid + Ereignis + das Mädchen hat die Taschenlampe

weißes Kleid + Ereignis + der Ball hat die Taschenlampe (als Abbildung)

5. Was sind typische Einsatzgebiete der Sprachtechnologie?

Informations- oder Transaktionsgebiete wie z.B. Bahnauskunft, ...

6. Nennen Sie drei Speech- und drei Texttechnologien.

Speech: Diktiersysteme, Text-To-Speech-Systeme, Dialogsysteme

Text: Maschinelle Übersetzung, Informationsextraktion, Textgenerierung

7. Sprachdialogsysteme können als endliche Automaten modelliert werden. Bitte geben Sie ein entsprechendes Modell für folgendes einfache "Kinoticketverkauf"-Szenario an: Der Benutzer soll ein Ticket für eine Kinovorstellung zu einer gewünschten Zeit (Abendvorstellung oder Spätvorstellung) erwerben können. Bitte halten Sie das Dialogmodell einfach.

B(enutzer), S(ystem)

B: (spricht System an)

S: Was wünschen Sie?

B: Ich möchte ein Ticket für Film XY kaufen.

S: Abend- oder Spätvorstellung?

B: Abendvorstellung.

S: Ein Ticket für die Abendvorstellung von XY kostet 5 Euro. Viel Spaß im Kino.

8. Nenne zwei typische Probleme einer Schlüsselwortsuche.

Schlüsselwörter können ambig sein und daher irrelevante Dokumente liefern.  
In einem relevanten Dokumenten können Synonyme des Schlüsselwortes, aber nicht das Schlüsselwort selbst auftauchen; diese werden dann nicht gefunden.

9. Nenne die Hauptkomponenten eines IR Systems und beschreibe kurz deren Aufgaben.

(Folien von Dr. Neumann: Folien 9-11)

10. Beschreibe kurz, wie im Vektorraummodell die Gewichtung von Termen und Dokumenten mittels des Maßes TF-IDF durchgeführt wird.

(Folien von Dr. Neumann: Folien 29-31)

13. Geben Sie beide Lesarten des Satzes "Every student presented a paper." in prädikatenlogischer Notation an.

A = Allquantor, E = Existenzquantor

$\forall x [ \text{student}(x) \Rightarrow \exists y [ \text{paper}(y) \ \& \ \text{present}(x,y) ] ]$

(Jeder Student stellt ein Paper vor, aber jeder ein unterschiedliches.)

$\exists y [ \text{paper}(y) \Rightarrow \forall x [ \text{student}(x) \ \& \ \text{present}(x,y) ] ]$

(Jeder Student stellt ein Paper vor, und zwar stellen alle dasselbe Paper vor.)

14. Was sind SynSets in WordNet?

SynSets in WordNet sind Mengen synonymer Worte.

15. Nennen Sie typische Anwendungen der statistischen Sprachverarbeitung.

Lesartendisambiguierung, POS-Tagging, Information Retrieval, Text-Klassifikation,  
...

16. Gegeben ist ein  $P(X)$ . Welche drei Eigenschaften müssen Sie überprüfen, um festzustellen ob  $P(X)$  eine Wahrscheinlichkeit ist.

Gemäß Kolmogorow:

- Wert von  $P(X)$  liegt zwischen 0 und 1.
- Das sichere Ereignis hat die Wahrscheinlichkeit 1:  $P(X) = 1$
- Die Wahrscheinlichkeit der Vereinigung von disjunkten Ereignissen ist die Summe der Einzelwahrscheinlichkeiten jedes Ereignisses.

17. Stellen Sie die folgende Äußerung in ihrer eigenen ad hoc Lautschrift dar, und alignieren Sie die Silben mit dem Sprachsignal:

Die Ernte kam sehr früh heim.

