

CIS

Centrum für Informations- und Sprachverarbeitung

Ludwig-Maximilians-Universität München

**Kommentiertes Vorlesungsverzeichnis
Wintersemester 2004/2005**

Allgemeines

Das Centrum für Informations- und Sprachverarbeitung (CIS) befindet sich im Gebäudekomplex der Institute am Englischen Garten, Flügel B, 1. Stock. Adresse: Oettingenstr. 67, 80538 München. Tel: (089)2180-9721; FAX: (089) 2180-9701 Das Sekretariat ist für Studenten vormittags von 10-12 geöffnet.

Veranstaltungsbeginn und Räume

Die Veranstaltungen beginnen in aller Regel in der ersten Semesterwoche. Abweichungen davon werden rechtzeitig ausgehängt. Falls nicht anders angegeben, finden die Veranstaltungen im Gebäude an der Oettingenstraße statt. Veranstaltungen in Räumen des Hauptgebäudes werden mit „HGB“, in Räumen der Schellingstraße mit „SCH“ gekennzeichnet.

Vorlesungsverzeichnis im WWW

Das kommentierte Vorlesungsverzeichnis kann auch unter folgender Adresse im WWW eingesehen werden: <http://www.cis.uni-muenchen.de/kurse>

Nachbardisziplinen

Hingewiesen wird auch auf die Veranstaltungen des Instituts für Informatik, des Instituts für Philosophie, Logik und Wissenschaftstheorie (Ludwigstr.31), des Instituts für Phonetik und Sprachliche Kommunikation (Schellingstr.3) und schließlich auf die Veranstaltungen zur theoretischen Linguistik des Instituts für Deutsche Philologie (Schellingstr.3).

Hinweis

In diesem Semester findet erneut die Übung Technik des wissenschaftlichen Arbeitens statt. Hauptfachstudenten, die ihr Studium im Winter 2002 oder später

aufgenommen haben, sind verpflichtet, vor dem Beginn des Hauptstudiums diese Übung einmal erfolgreich zu absolvieren. Studienanfängern wird die Teilnahme an der anschließenden UNIX-Übung dringend nahegelegt.

Fachschaft

Die Fachschaft Computerlinguistik, die selbstverwaltete Interessenvertretung der Studenten am Institut, ist unter <http://www.cis.uni-muenchen.de/studenten/fachschaft> zu erreichen. Kommilitonen, die sich für die Belange der Computerlinguistikstudenten engagieren wollen, sind herzlich willkommen.

Studienberatung

Prof. Dr. Franz Guenther (Freisemester)
Prof. Dr. Klaus U. Schulz Di. 12-13
Dr. Max Hadersbeck Mo. 16-17
Dr. Stefan Langer nach Vereinbarung
Dr. Hans Leiß, Mi. 10-11
Dr. Petra Maier-Meyer Do. 11-12
Gerhard Rolletschek M.A. Di. 17-18
Dr. Daniel Schnorbusch nach Vereinbarung
Jörg Schuster M.A. Mi. 13-14
Dr. Petra Witschel, nach dem Seminar

Stand: Oct 11, 2004

Einführung in die Programmierung

M. Hadersbeck

Proseminarvorlesung

Zeit und Raum: Mo. 10-12, Raum 1.27
Übung I Mi. 10-12, Raum Z11 Takla Makan,
Übung II Fr. 14-16 Raum 1.14;

Inhalt und Ziel der Veranstaltung

In dieser Veranstaltung werden Sie einen Überblick über die Arbeitsweise von Computern bekommen. Wir lernen Strategien und Vorgehensweisen zur Entwicklung von Computerprogrammen. Außerdem lernen wir Datentypen, Standardkontrollstrukturen und den Aufbau regulärer Ausdrücke einer

Scriptsprache wie Python oder PERL kennen. In der Übung werden kleine Programme selbst geschrieben. Es gibt zwei Übungstermine (Mittwoch und Freitag), von denen einer wahrgenommen werden muss.

Voraussetzungen

keine

Teilnahmebedingungen

Aktive Mitarbeit

Scheinkriterien

Aktive Teilnahme, Bearbeitung von Übungsaufgaben, Klausur

Programmierpraktikum für Computerlinguisten

M. Hadersbeck

Proseminarvorlesung

Zeit und Raum: Mo. 14-16, Raum 0.15. und Mi. 12-14, Raum 1.27

Inhalt und Ziel der Veranstaltung

Ein zentrales Problem in der CL ist das Klassifizieren von Wörtern in geschriebenen Texten. In diesem Praktikum werden wir zuerst die Technik des Taggen kennenlernen: In der Vorlesung wird ein C++ Tagger vorgestellt und in der Übung um selbst geschriebene Programme verbessert und erweitert werden. Von den Teilnehmern wird erwartet, dass sie viel Zeit für die Übungen aufwenden.

Voraussetzungen

Gute Programmierkenntnisse in C++ .

Scheinkriterien

Für die erfolgreiche Programmierung der Aufgabe in einer Arbeitsgruppe gibt es einen Proseminarschein.

Computerlinguistik II

H. Leiß

Vorlesung mit Proseminar

Zeit und Raum: Di. 12-14, Raum SCH E06 und Fr. 14-16, Raum 1.14

Inhalt und Ziel der Veranstaltung

Die Veranstaltung behandelt die grundlegenden Aufgaben der Computerlinguistik anhand einer exemplarischen Implementierung der linguistischen Basismodule:

- Morphologie
- Syntax
- Semantik
- Pragmatik

Am Ende des Semesters werden wir ein natürlichsprachliches Generierungs- und Analyse-System für einen kleinen Sprachausschnitt in Prolog implementiert haben. Prologkenntnisse werden zu Anfang des Semesters bei Bedarf wiederholt.

Literatur

1. Pereira, F.C.N., S.M.Shieber (1987) *Prolog and Natural Language Analysis*, CSLI Lecture Notes Number 10, Stanford. (CIS-Bibliothek: PER I)
2. Gazdar, Gerald, Mellish, Chris (1989). *Natural Language Processing in Prolog. An Introduction to Computational Linguistics*. Wokingham u.a. : Addison-Wesley (CIS-Bibliothek: GAZ I)
3. Lehner, Christoph (1990). *Prolog und Linguistik*. München u.a. : Oldenbourg (CIS-Bibliothek: LEH II)
4. Wenger, Karsten, o.J. *Einführung in die Computerlinguistik*. CIS-Bericht.

Voraussetzungen

Kenntnisse in Prolog. Teilnehmer ohne Grundkenntnisse in Prolog bitte rechtzeitig beim Dozenten melden.

Scheinkriterien

Mitarbeit in den Übungsstunden, Lösung der Hausaufgaben und Klausur/Programm

Mathematik I

H. Leiß

Vorlesung mit Proseminar

Zeit und Raum: Di. 16 -18 , Raum 1.27 und Do. 8-10, Raum 1.14

Inhalt und Ziel der Veranstaltung

Es wird eine Einführung in Teilgebiete der diskreten Mathematik gegeben, die es erleichtert, die formalen Grundlagen anderer Veranstaltungen im Studiengang Computerlinguistik zu verstehen. Behandelt werden

1. mathematische Grundbegriffe wie Mengen, Relationen, Funktionen
2. Relationalstrukturen und Algebren
3. Syntax und Semantik der Aussagen- und Prädikatenlogik

Literatur

1. Mathematische Grundlagen I. Vorlesungsskriptum von Prof.K.Schulz.
2. B.H.Partee, A.Ter Meulen, R.E.Wall: Mathematical Methods in Linguistics

Voraussetzungen

Keine

Scheinkriterien

Mitarbeit in den Übungsstunden, Lösung der Hausaufgaben und Klausur

Einführung in formale Methoden und Arbeitsweisen der Computerlinguistik für Aufbaustudenten

P. Maier-Meyer

Proseminarvorlesung

Zeit und Raum: Mi. 14-16, Raum 0.13, Do. 16-18, Raum 0.11 / Z11 -Takla Makan

Inhalt und Ziel der Veranstaltung

Dieser Kurs richtet sich speziell an Studenten des Aufbaustudiengangs *Computerlinguistik*. Im Rahmen dieser Einführung werden die formalen Grundlagen und elementaren Arbeitstechniken der Computerlinguistik in Theorie und Praxis vermittelt. Im Theorieteil geht es vor allem um die mathematischen Grundlagen:

- Mengen, Relationen, Funktionen
- elementare Grundlagen der Aussagen- und Prädikatenlogik
- elementare Grundlagen von formalen Sprachen

Im Praxisteil lernen die Teilnehmer die Anwendung von verschiedenen Unix-Tools und der Skriptsprache Perl zur Lösung elementarer Aufgaben in der Computerlinguistik.

Literatur

1. Barbara H. Partee, Alice ter Meulen & Robert Wall: *Mathematical Methods in Linguistics*, Kluwer Academic Publishers, 1990.
2. Randal L. Schwartz, Tom Phoenix: *Learning Perl, 3rd Edition Making Easy Things Easy and Hard Things Possible*; 3rd Edition O'Reilly July 2001 0-596-00132-0.

Voraussetzungen

keine

Scheinkriterien

Übungsaufgaben und Klausur

Morphologie, Lexikographie & Computerlexikographie

G. Rolletschek

Proseminarvorlesung

Zeit und Raum: Di. 18-20, Raum 1.27 und Do. 12-14, Raum 1.14

Inhalt und Ziel der Veranstaltung

In diesem Seminar geht es zunächst um die Vermittlung vertiefter Kenntnisse der Flexions- und Wortbildungsmorphologie. Nach einem knappen Überblick über die verschiedenen Wortbildungsverfahren im Deutschen und anderer Sprachen werden wir uns mit den Problemen einer theoretisch angemessenen Erfassung der Daten befassen. Neben den klassischen Konzeptionen (I&A, I&P, W&P) werden Modelle, die in der CL relevant sind (Finite-State-Morphology, Two-Level-Morphology) behandelt. Der zweite Schwerpunkt des Seminars ist die maschinelle Lexikographie, wobei es ein zentrales Thema sein wird, wie morphologische, syntaktische und semantische Eigenschaften von Wörtern in einem elektronischen Lexikon repräsentiert werden können. Themen werden hier unter anderem sein:

- Gemeinsamkeiten und Unterschiede zwischen papierenen und elektronischen Wörterbüchern
- Struktur lexikalischer Einträge in elektronischen Lexika
- Die Semantik lexikalischer Einheiten und ihre Repräsentation im Lexikon
- Struktur des institutseigenen Lexikonsystems CISLEX im Vergleich zu anderen Systemen (z.B. WordNet, COMLEX, Celex...)
- automatische Extraktion lexikalischer Information aus Korpora

Literatur

Haspelmath, Martin (2002). *Understanding Morphology*. London: Arnold

Teilnahmebedingungen

Aktive Mitarbeit, Übungen

Scheinkriterien

Aktive Mitarbeit, Übungen und Klausur

Formale Sprachen und Automaten

Proseminarvorlesung

Zeit und Raum: Mo. 12-14, Raum 0.13, Do. 14-16, Raum 1.27

Inhalt und Ziel der Veranstaltung

Die Vorlesung gibt eine Einführung in die Theorie der formalen Sprachen und der zu ihrer Erkennung nötigen Automatenmodelle. Der Zusammenhang zur Semantik von sequentiellen Programmiersprachen wird erläutert. Damit werden die Grundlagen für eine Beschäftigung mit maschineller Analyse natürlicher Sprachen gelegt. Behandelt werden:

- Reguläre Ausdrücke, als Sprachdefinitionen und als iterative Programme
- Erkennung regulärer Sprachen durch endliche Automaten
- Kontextfreie Grammatiken, als Sprachdefinitionen und rekursive Programme
- Erkennung kontextfreier Sprachen durch Kellerautomaten
- Chomskys Hierarchie formaler Sprachklassen
- Wiederholungs- und andere Struktureigenschaften kontextfreier Sprachen
- Grundprinzipien der Syntaxanalyse für kontextfreie Grammatiken

Literatur

1. Hopcroft / Ullman (1979). *Introduction to Automata Theory, Languages and Computation*. Addison Wesley
2. N.Moll / M.Arbib / A.J.Kfoury (1988). *An Introduction to Formal Language Theory*. New York/Berlin: Springer
3. H.Leiß (1997). *Formale Sprachen und Automaten. Vorlesungsskript*. München: CIS

Voraussetzungen

Mathematischer Grundkurs oder entsprechende Kenntnisse

Scheinkriterien

Schein bei aktiver Teilnahme und Lösung von Übungsaufgaben.

Computerlinguistik I

J. Schuster

Proseminarvorlesung

Zeit und Raum: Mi. 8 - 10 , Raum 1.14 und Fr. 8-10, Raum 1.14

Inhalt und Ziel der Veranstaltung

Die Veranstaltung zerfällt in einen Vorlesungs- und einen Übungsteil. Der Vorlesungsteil gibt einen Überblick über grundlegende Begriffe der theoretischen Linguistik. Es werden nacheinander die klassischen Teilgebiete der Linguistik betrachtet:

- Phonologie
- Morphologie
- Syntax
- Semantik
- Pragmatik.

In den Übungen versuchen wir, kleine Programme und/oder Algorithmen zu den im Vorlesungsteil betrachteten Problemen zu schreiben. Da manche linguistischen Probleme zu komplex sind, um in sinnvoller Zeit ein Programm zu ihrer (teilweisen) Lösung zu schreiben, werden praktischer und theoretischer Teil der Veranstaltung streckenweise unabhängig voneinander sein.

Literatur

Adrian Akmajian et alii (1984). *Linguistics. An Introduction to Language and Communication*. Cambridge MA: MIT Press.

Voraussetzungen

Gleichzeitiger Besuch der Veranstaltungen *Einführung in die Programmierung* und *UNIX-Grundlagen* oder entsprechende Kenntnisse. Bereitschaft, von der ersten Stunde an Aufgaben am Rechner zu lösen.

Scheinkriterien

Referat, Übungen und Klausur. Unter „Referat“ wird hier die Vorstellung einer Musterlösung zu einer Übung verstanden.

Datenbanken

C. Draxler

Proseminar

Zeit und Raum: Do. 10-12, Raum 0.43

Inhalt und Ziel der Veranstaltung

Datenbanken erlauben eine anwendungsunabhängige Speicherung von und einen kontrollierten Zugriff auf sehr große Datenmengen - z.B. Sprachsignale, Transkriptionen, Lexika. Dieses Seminar führt anhand praktischer Beispiele aus der Phonetik und Computerlinguistik in das Gebiet ein:

- Datenmodell, Datenbankentwurf
- Hierarchisches Datenmodell und Netzwerk-Datenmodell
- Relationenmodell
 - Relationenalgebra, -kalkül
 - Normalisierung
 - SQL und QBE
 - Anwendung I: Aussprachelexikon
- Deduktive Datenbanken
 - Anwendung II: Phonetische Datenbank
- Objekt-orientiertes Datenmodell
- World Wide Web

Literatur

Zum Seminar erscheint ein Skript im WWW. Literatur wird im Seminar bekanntgegeben.

Voraussetzungen

Bereitschaft zur Arbeit am Computer

Scheinkriterien

Vortrag und Hausarbeit

Einführung in Java

C. Draxler

Proseminar

Zeit und Raum: Do. 8-10, Raum Z11 - Takla Makan

Inhalt und Ziel der Veranstaltung

Java ist eine moderne objektorientierte und plattformunabhängige Programmiersprache, die sowohl zur Programmierung von eigenständigen Applikationen als auch von Applets, das sind Programme, die innerhalb eines WWW-Browsers laufen, verwendet wird.

Der Javakurs besteht aus zwei Teilen: eine Einführung ins Programmieren in Java im Wintersemester, und fortgeschrittenes Programmieren in Java im Sommersemester.

Themen im Wintersemester sind: objektorientiertes Programmieren, elementare Programmierkonstrukte wie Schleifen, Fallunterscheidungen und Rekursion sowie der Aufbau eigener Datenstrukturen und elementare Ein- und Ausgabe. Themen des Sommersemesters sind grafische Oberflächen, Multimedia, das WWW sowie Datenbankzugriffe von Java aus.

Die Programmierbeispiele stammen größtenteils aus dem Bereich der Phonetik bzw. Linguistik. Im Kurs stelle ich Programmfragmente vor, die in eigener Arbeit dann vervollständigt werden müssen.

Literatur

1. Java Tutorial, <http://java.sun.com/docs/books/tutorial/>,
2. <http://fishmac.phonetic.uni-muenchen.de/JavaKurs/>,
3. David Flanagan (1997). *Java in a Nutshell* Sebastopol: O'Reilly Press

Voraussetzungen

Programmiererfahrung in einer anderen Programmiersprache wünschenswert, selbstständiges Arbeiten während der Woche.

Scheinkriterien

auf Wunsch stelle ich einen Teilnahmechein aus. Einen Schein mit Leistungsnachweis gibt es für ein selbstständig ausprogrammiertes Anwendungsprogramm.

Perl für Fortgeschrittene

M. Heller

Proseminar

Zeit und Raum: Di. 18-20, Raum Z11- Takla Makan, Do. 18-20 Raum Z11-Takla Makan

Inhalt und Ziel der Veranstaltung

Die Details gibt der Dozent in der ersten Sitzung bekannt.

Voraussetzungen

Grundkenntnisse in Perl

Scheinkriterien

siehe Inhalt

Sprachakustik

T. Horn

Proseminar

Zeit und Raum: Do. 16-18, Raum 0.41

Inhalt und Ziel der Veranstaltung

Sprachakustik befasst sich mit der Wahrnehmung, Erzeugung, Übertragung und Verarbeitung von Sprachschall aus psychologischer, physiologischer, physikalischer und technischer Sicht.

Der Kurs gibt eine allgemeinverständliche Einführung, indem einige hundert grundlegende Fachbegriffe möglichst anschaulich und beispielreich erläutert werden, im Sinne einer didaktisch geordneten Enzyklopädie.

Teilnahmebedingungen

keine

Scheinkriterien

kurze mündliche Prüfung

Sprachspezifische Aspekte der automatischen Sprachverarbeitung

S. Langer

Proseminar

Zeit und Raum: Do. 16-18, Raum HGB B09

Inhalt und Ziel der Veranstaltung

In diesem Kurs geht es um grundlegende Fragen der Verarbeitung natürlicher Sprachen in multilingualen Anwendungen. Der Kurs beginnt mit einer Einführung in Zeichensätze und Zeichensatzkodierungen, wobei sowohl auf sprachspezifische Kodierungen (etwa die ISO-8859-X Zeichensätze und Multibytekodierungen für die asiatischen Sprachen) eingegangen wird, als auch auf den sprachübergreifenden Zeichensatz UNICODE und sein Kodierungen (u.a. UTF-8). In diesem Rahmen werden auch Programme, Programmmodule und Programmbibliotheken zur Zeichensatzkonversion vorgestellt. Im weiteren Verlauf des Seminars sollen Fragen der automatischen Sprachenidentifizierung, der Normalisierung, Tokenisierung und der Lemmatisierung unterschiedlicher Sprachen behandelt werden.

Voraussetzungen

Grundkenntnisse in Linguistik, insbesondere der Morphologie, sowie Fertigkeiten in mindestens einer Programmiersprache (Perl, Python oder C(++)) und im Umgang mit Unix/Linux. Englischkenntnisse. Anmeldung in der ersten Sitzung. Eine Voranmeldung ist nicht nötig.

Scheinkriterien

Regelmäßige Teilnahme, Bereitschaft, ein Referat von ca. 30 min zu übernehmen und die Inhalte des Referats für die anderen Teilnehmer schriftlich aufzuarbeiten.

(Möglicherweise kann auch ein Hauptseminarschein erworben werden; steht aber noch nicht fest)

Probleme der Konnexion

D. Schnorbusch

Proseminar

Zeit und Raum: Mo. 18-20, Raum 0.37

Beginn am 25.10.2004

Inhalt und Ziel der Veranstaltung

Voraussetzung für alle höheren Anwendungen der CL ist das korrekte Zerlegen von beliebigem Text und den darin enthaltenen Satzgefügen in einfache Sätze. Gegenstand des Seminars sind daher die Konnektoren des Deutschen, Ausdrücke also, die i.d.R. syntaktisch zwei einfache Sätze und semantisch zwei einfache Prädikationen miteinander verbinden. Neben den „klassischen“ und in ihrer Anzahl überschaubaren Konnektoren, den Konjunktionen und Konjunkionaladverbialen, werden wir uns vor allem mit komplexen Konnektoren vom Typ *unter der Voraussetzung, dass...* befassen. Diese stellen - und entgegen der Darstellung in der Standardliteratur - aufgrund ihrer Vielzahl und ihrer syntaktischen und semantischen Eigenschaften ein erhebliches Problem für die CL dar.

Literatur

Pasch, R. et al. (2003). *Handbuch der Deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfen*. Berlin /New York: de Gruyter

Teilnahmebedingungen

Anmeldung beim Dozenten während der vorlesungsfreien Zeit. Regelmäßige Mitarbeit. Übernahme eines Referats.

Scheinkriterien

Erfüllung der Teilnahmebedingungen und Abfassung einer schriftlichen Hausarbeit.

Methoden stochastischer Modellierung von Sprache

P. Witschel

Proseminar

Zeit und Raum: Mo. 16-18, Raum 1.15

Inhalt und Ziel der Veranstaltung

Beginnend mit den statistischen Grundlagen zur Sprachmodellierung werden im Wesentlichen Sprachmodelle basierend auf N-Grammen vorgestellt, das so genannte Training von Sprachmodellen näher betrachtet und Bewertungsgrößen für den Test erarbeitet. Speziell behandelte Themen sind unter Anderem

- Sprachmodelle basierend auf linguistischen Klassen
- Annotieren von Trainingstexten
- Adaption von Sprachmodellen
- praktischer Einsatz von Sprachmodellen in der Automatischen Spracherkennung

Voraussetzungen

Mathematische und linguistische Grundkenntnisse

Scheinkriterien

Aktive Mitarbeit, Referat

Exemplarische Satzanalysen

F. Guenthner

Hauptseminar

Zeit und Raum: Mi. 10-12, Raum 0.11

Inhalt und Ziel der Veranstaltung

Anhand einer Reihe von syntaktischen Phänomenen des Deutschen (und anderer Sprachen) wollen wir einen neuen Ansatz zur systematischen Beschreibung von „einfachen“ Sätzen vorstellen, der auf vorhandene Vorschläge von Z. Harris, M. Gross und I. Mel'cuk basiert. Die Grundidee dabei wird sein, dass man ein taxonomisches Lexikon von „Prädikat-Argument-Strukturen“ braucht, welches alle elementaren Sätze einer Sprache und deren Varianten umfasst. Abwechselnd mit Detailanalysen verschiedener Satzbeispiele werden wir die Grundbegriffe der Gross'schen „Lexikon-Grammatik“ sowie der Arbeiten von Harris und Mel'cuk unter die Lupe nehmen.

Scheinkriterien

Ausarbeitung eines Teilsbereichs eines Prädikat-Argument-Strukturen Lexikons des Deutschen.

Erste Sitzung: 27. Oktober, 2004

Voraussetzungen

- Erfolgreiche Teilnahme am Proseminar SYNTAX
- Anmeldung beim Dozenten vor der ersten Sitzung (auch per email/Telefon) wegen Bereitstellung von Kursunterlagen

Computerlinguistische Analyse von Bioscience- Texten

F. Guenthner / J. Goller

Hauptseminar

Zeit und Raum: Mi. 14-16, Raum 0.11

Inhalt und Ziel der Veranstaltung

Ziel des Seminars ist die eingehende linguistische Analyse von wissenschaftlichen

Texten aus dem Bereich der Zell- und Proteinforschung, sowie die Bereitstellung von Mitteln zu ihrer automatischen Verarbeitung. Es werden keine Kenntnisse aus den Bio-Wissenschaften vorausgesetzt, sie stellen für das Seminar nur einen exemplarischen, ausreichend kleinen Ausschnitt der wissenschaftlichen Sprache dar, der es erlaubt, den folgenden Zielen mit vertretbarem Aufwand gerecht zu werden:

- Lexikographische Analyse des Vokabulars
- Betrachtung typischer Diskurs- und Argumentationsstrukturen
- Studie bereichsspezifischer Prädikat-Argumentstrukturen
- Informationsextraktion und -normalisierung

Die Betrachtung und Diskussion dieser Techniken im theoretischen Teil des Seminars (2-stündig) wird stattfinden mit Blick auf ihre Anwendung in einer neuartigen Suchmaschine, deren Bau sich der praktische Teil widmen wird (1-stündig). Hierzu werden wir eine ausreichend große Kollektion wissenschaftlicher Abstracts den genannten Technologien unterziehen und die Ergebnisse in Form von Übungsaufgaben und gemeinsamen Diskussionen erörtern.

Erste Sitzung: 27. Okt. 2004

Scheinkriterien

Übernahme eines Referats und Mitarbeit an der Suchmaschine; Referatsthemen sollten vor der ersten Sitzung mit Prof. Guenther vereinbart werden.

Semantik II: Interpretation und Auswertung

H. Leiß

Hauptseminar

Zeit und Raum: Mo. 16-18, Raum 0.13

Im Seminar soll die Semantik verschiedener Konstruktionen natürlicher Sprachen, z.B. Plural, koordinierte Nominalphrasen, Reflexiv-, Reziprok- und Diversitäts-Pronomina, distributive Verblesarten, Adverbien, Gradausdrücke, behandelt werden. Dazu sollen Grammatikregeln mit passenden semantischen Auswertungsregeln geschrieben und in Beispielinterpretationen getestet werden.

Für die linguistische Analyse und ihre theoretische Aufarbeitung werden die beiden angegebenen Bücher benutzt, für die Implementierung soll ML-Yacc benutzt werden.

Literatur

1. M. Böttner (1999). *Relationale Semantik*. Tübingen: Niemeyer.
2. H.Kamp, U.Reyle (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Voraussetzungen

Grundkenntnisse der Montague-Semantik.

Scheinkriterien

Seminarvortrag mit Ausarbeitung bzw. Implementierung.

Kompaktkurs Information Retrieval

K. Schulz

Hauptseminar

Zeit und Raum: 20.09. - 01.10, Raum n.n.

Inhalt und Ziel der Veranstaltung

Der Kurs gibt eine Einführung in folgende Teilgebiete des Information Retrieval: Geschichte, Grundlagen, Daten-Dateien-Texte, IR-Modelle, Evaluierung von IR-Systemen, Einzelheiten des Vektorraummodells, computerlinguistische Hilfsmittel, Abfrage strukturierter Dokumente und Indexieren von Multimedia-Datenbanken. Der Kurs findet parallel zu einem virtuellen Seminar der Universitäten Tübingen und Tilburg statt, das in Zusammenarbeit mit dem CIS entwickelt wird. In Teilen werden die in diesem Rahmen entwickelten webbasierten Unterrichtsmaterialien und die Internet-Lernplattform dieses Kurses verwendet.

[Skript-PS](#)

Voraussetzungen

Computerlinguistische und mathematische Grundkenntnisse

Scheinkriterien

Lösen von Übungsaufgaben und/oder einfachen Programmieraufgaben

Indexieren, Abfragen und Ranking strukturierter Dokumente

K.U. Schulz & F. Weigel

Hauptseminar

Zeit und Raum: Di. 16-18, Raum 0.13

Inhalt und Ziel der Veranstaltung

XML ist eine inzwischen weit verbreitete Sprache zur Auszeichnung von Dokumenten („Markup“) und zur Datenrepräsentation im allgemeinen. Sie wird sowohl im Internet als auch außerhalb verwendet. Für die Verarbeitung von XML-Dokumenten werden Anfragesprachen benutzt, mit denen man in den Dokumenten Elemente selektiert und daraus gegebenenfalls neue Dokumente konstruiert. Diesen Anfragesprachen liegen meist Datenbank-Paradigmen zugrunde; es gibt aber auch vereinzelte Ansätze, die sich am Information Retrieval orientieren und z. B. nach Relevanz sortierte Antworten erlauben („Ranking“). In der Veranstaltung werden sowohl standardisierte Anfragesprachen für XML (wie z. B. XQuery) als auch neue Forschungsansätze vorgestellt. Die Grundlagen von XML werden dabei vorausgesetzt. Wenn genügend Zeit bleibt, sollen auch Indexverfahren für die Suche in XML-Daten vorgestellt werden.

Literatur

1. Robert Eckstein (2000). *XML kurz und gut*. O'Reilly/VVA
2. Serge Abiteboul, Peter Buneman, Dan Suciu (1999). *Data on the Web: From Relations to Semistructured Data and XML (Kapitel 1-5)*. Morgan-Kaufmann.
3. David Hunter et al.(2004). *Beginning XML*. 3rd Edition, Wrox
4. World Wide Web Consortium: www.w3.org
5. Robin Cover's XML Pages: xml.coverpages.org

Voraussetzungen

XML-Grundkenntnisse und evtl. Datenbank-Grundkenntnisse

Teilnahmebedingungen

Keine

Scheinkriterien

Übernahme eines Vortrages und/oder einer Programmieraufgabe

Dialogsysteme und Voice-XML

K.U. Schulz & Chr. Ringlstetter

Hauptseminar

Zeit und Raum: Di. 14-16, Raum 1.27

Inhalt und Ziel der Veranstaltung

Dialogsysteme haben inzwischen eine Vielzahl praktischer Anwendungen wie etwa Telebanking, automatische Auskunftssysteme, automatische Klientenverteilung im Call Center Bereich, Car Telematics oder Computer Aided Language Learning (CALL). In diesem Seminar soll ein Überblick zu aktuellen Entwicklungen im Bereich Dialogsysteme gegeben werden. Teilthemen die im Rahmen von Vorträgen behandelt werden, sind Arten von Dialogsystemen, Design, Entwicklung und Evaluierung solcher Systeme aber auch deren Grenzen. Daneben sollen auch praktische Fertigkeiten zur Erstellung von Dialogsystemen erarbeitet werden. Hierzu wird einerseits eine Einführung in die Dokumentbeschreibungssprache VoiceXML gegeben, andererseits wird die GALAXY Architektur des MIT vorgestellt.

Literatur

1. M.F. McTear (2002). *Spoken Dialogue Technology: Enabling the Conversational User Interface*. ACM Computing Surveys 43 (1).
2. D. Jurafsky, J. H. Martin (2000). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Kapitel 18-20*. Prentice Hall.
3. Voice Extensible Markup Language (VoiceXML) Version 2.0.
<http://www.w3.org/TR/voicexml20/>, W3C Recommendation, 2004.

4. S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue (1998). *GALAXY-II: A Reference Architecture for Conversational System Development*. Proc. ICSLP 98, Sydney, Australia, November 1998

Voraussetzungen

Abgeschlossenes Grundstudium, Anmeldung

Teilnahmebedingungen

Keine

Scheinkriterien

Vortrag, Programmieraufgaben, Klausur

Technik des wissenschaftlichen Arbeitens

C. Draxler, S. Langer, H. Leiß & D. Schnorbusch

Übung

Zeit und Raum: Fr. 12-13, Raum 1.14

Inhalt und Ziel der Veranstaltung

Ziel der Übung ist es, die Grundfertigkeiten eines (Computerlinguistik-) Studiums zu vermitteln. Neben allgemeinen Hinweisen zum Studium werden folgende Fragen behandelt:

- Welche Literatur spielt in der CL eine Rolle und wie kann man diese finden?
 - einführende Literatur, Handbücher, zentrale Aufsätze
 - Zeitschriften der CL und ihrer Nachbarwissenschaften
 - Literaturrecherche in Bibliotheken und via Internet
 - etc.
- Wie wird ein Referat im Seminar vorbereitet und gehalten?
 - Zeitplanung
 - Medien: Begleitblätter, Folien, Notebook+Beamer
 - Vortragsstil

- etc.
- Wie wird eine Seminararbeit geschrieben?
 - Form und Struktur einer Seminararbeit
 - Zitierkonventionen
 - Anmerkungen
 - etc.

HINWEIS: Studenten, die bisher noch nicht an dieser Übung teilgenommen haben, wird dies dringend nahegelegt.

Voraussetzungen

keine

Scheinkriterien

regelmäßige Teilnahme

UNIX-Grundlagen

W. Mederle & Th. Schäfer

Übung

Zeit und Raum: Fr. 13-14, Raum 1.14 / Z11 Takla Makan

Inhalt und Ziel der Veranstaltung

Die Übung bietet eine Einführung in das am CIS und auch sonst an den Universitäten vornehmlich eingesetzte Betriebssystem Unix bzw. Linux. Themen werden sein:

- Einführung in GNU/Linux/Unix fuer Computerlinguisten
- Prinzipien unixartiger Betriebssysteme
- Shell-Bedienung, Kommandozeilentools, einfache Bash-Skripte
- sicheres Arbeiten im Netzwerk
- Grundlagen der wichtigsten Netzwerkprotokolle
- Emacs-Einführung

- Bearbeiten multisprachlicher Dokumente/Programmieren mit Emacs

HINWEIS: Studienanfängern wird die Teilnahme an dieser Übung dringend nahegelegt. Hier werden die Grundlagen im Umgang mit dem Rechner vermittelt, dem wichtigsten Arbeitsgerät der Computerlinguisten.

Voraussetzungen

keine

Scheinkriterien

regelmäßige Teilnahme

Kolloquium Computerlinguistik

K. Schulz

Kolloquium

Zeit und Raum: Fr.10.00 - 12.00, Raum 0.05

Inhalt und Ziel der Veranstaltung

Das Kolloquium findet jeweils nach Ankündigung statt. Hier berichten Mitarbeiter und Gäste des Instituts über ihre laufenden Arbeiten.

Teilnahmebedingungen

Die Teilnahme am Kolloquium steht jedem offen.

Logik in der Informatik

Dozenten der Informatik & K.Schulz

Kolloquium

Zeit und Raum: Fr. 8.00 - 10.00, Raum Theresienstr. E27

Inhalt und Ziel der Veranstaltung

Kolloquium des Graduiertenkollegs „Logik in der Informatik“. Die Kollegiaten und Gäste halten Vorträge. Siehe Ankündigungen.

Linguistisches Kolloquium

Dozenten der Linguistik

Zeit und Raum: unregelmäßig mittwochs, zumeist in Raum E06, Schellingstr. 3

Inhalt und Ziel der Veranstaltung

Gäste des Instituts für Deutsche Philologie halten Vorträge zu sprachwissenschaftlichen Themen.

Einführung in die Programmierung, vector is unstable.

Amtliche Bekanntmachungen der TU Bergakademie Freiberg, the binomial of Newton, despite external influences, laterally reflects the Mediterranean shrub. A methodological approach to key competencies in informatics, by which scales philosophical fear.

Kognitionen, the maximum deviation of the substrate causes the chorale.

Schanda, Franz: Computer-Lernprogramme. Wie damit gelernt wird. Wie sie entwickelt werden. Was sie im Unternehmen leisten, contemplation firmly discredited empirical alluvium.

Statistical Literacy: Eine Basiskompetenz in der Informationsgesellschaft, a priori bisexuality monotonically scales the integral of a variable.

Luhmanns Systemtheorie aus der Sicht der Verteilten Künstlichen Intelligenz, style, obviously, accumulates interplanetary Muscovite.