

UNIVERSITÄT HEIDELBERG

Sommer 2012

Kommentiertes Verzeichnis der Lehrveranstaltungen

Computerlinguistik

(Stand: 24.2.2012)

Die ausführliche Darstellung der Veranstaltungen findet sich
unter <http://lsf.uni-heidelberg.de>



Computerlinguistik

Zusätzliche Lehrveranstaltungen und weitere Informationen werden durch Aushang am schwarzen Brett in INF 325, 1. Stock und auf unseren Internetseiten bekanntgegeben.

Vorbereitungskurse

Einführung in die Nutzung computerlinguistischer Ressourcen

Übung; SWS: 2

Block; 09:00 - 17:00; 10.04.2012 - 13.04.2012; INF 325 / PCPool; Simianer, P.

Inhalt Der Vorkurs gibt Grundlagen in der Nutzung von GNU/Linux-basierten (computerlinguistischen) Tools sowie der vorhandenen Korpora. Dabei geht es sowohl um allgemeine Grundlagen der Shell (wie z.B. Ein-/Ausgabeumleitung) als auch um einzelne Toolkits im Speziellen. Wir werden uns anschauen, wie bestimmte Tools zu benutzen sind, was man aus ihnen herausbekommt (und was nicht) und wie man solche Ausgaben automatisch weiterverarbeiten kann (und zum Beispiel an das nächste Tool weiterverfüttert). Zudem umfasst der Kurs eine Kurzeinführung in MapReduce und den am Institut vorhandenen Hadoop Cluster.

Leistungsnachweis Die Teilnahme am Blockkurs ist Voraussetzung für die Teilnahme am Softwareprojekt.

Voraussetzung Voranmeldung bitte per E-Mail an den Dozenten.

Bachelor (NBA; BA-2010)

Begleitveranstaltung zum Softwareprojekt - SP

Seminar; SWS: 2

Di; wöch; 14:15 - 15:45; ab 24.04.2012; INF 327 / SR 2; Gruppe 1;

Di; wöch; 16:15 - 17:45; ab 24.04.2012; INF 327 / SR 2; Gruppe 2;

Kommentar Leistungsbewertung:

BA-2010 (SP): 6 LP + 4 LP ÜK

NBA (SP): 6 LP + 4 LP ÜK

Inhalt Im Softwareprojekt soll eine computerlinguistische Aufgabenstellung weitgehend eigenverantwortlich und in Teamarbeit geplant, softwaretechnisch durchgeführt, dokumentiert und abschließend präsentiert werden. Neben der Vertiefung praktischer Programmierkenntnisse (Techniken und Werkzeuge für verteilte Programmerstellung, Testverfahren und Qualitätskontrolle, Dokumentation, etc.) sollen Teamfähigkeit und planerische Fähigkeiten geübt werden. Daneben werden grundlegende Techniken und Methoden wissenschaftlichen Arbeitens vermittelt.

Leistungsnachweis Teilnahme an allen Einführungsvorlesungen, Projekt-Spezifikationsvortrag, Projekt-Abschlussvortrag und Demo, Programmdokumentation, Archivierung und Poster

Voraussetzung Programmierprüfung, Einführung in die Benutzung computerlinguistischer Ressourcen

Voranmeldung: obligatorisch, an nastase@cl.uni-heidelberg.de, spätestens bis zum 1. April 2012

Literatur Abhängig vom Projekt; wird zu Beginn des Semesters bekannt gegeben

Abschlusskolloquium - Coll

Kolloquium; SWS: 3

Di; k.A.; 18:15 - 20:30; ab 24.04.2012; INF 325 / SR 24; Frank, A.;Nastase, V.;Riezler, S.

Kommentar	Leistungsbewertung: BA-2010 (Coll): 2 LP NBA (-): Freiwillig, keine LP Master (Coll): 2 LP
Inhalt	Präsentation laufender BA-, MA- und Magisterarbeiten Das Abschlusskolloquium bietet BA-, MA- und Magisterstudierenden ein Forum für die Vorstellung und Diskussion ihrer Abschlussarbeiten. Die Studierenden präsentieren den aktuellen Stand ihrer Arbeit und erhalten in der Diskussion Anregungen von Seiten der Studierenden und der Dozenten. Externe Vorträge Darüber hinaus bietet das Abschlusskolloquium allen Studierenden durch Vorträge geladener Gäste Einblicke in aktuelle Forschungsfragen der Computerlinguistik. Externe Vorträge finden im Rahmen des Computerlinguistischen Kolloquiums statt.
Leistungsnachweis	Vortrag (BA-2010, MA) und Ausarbeitung; regelmäßige Präsenz ist Voraussetzung für den Scheinerwerb. Ein Leistungserwerb ist nur für Examenskandidat/innen im Bachelorstudiengang (BA-2010) und Masterstudiengang (MA) vorgesehen. Jedoch sind alle Studierenden eingeladen, ihre Abschlussarbeiten vorzustellen, den Vorträgen zuzuhören und sich an den Diskussionen zu beteiligen.

Advanced Programming - AS-CL, SS-CL, SS-TAC

Vorlesung / Übung; SWS: 2

Mo; wöch; 11:15 - 12:45; ab 23.04.2012; INF 325 / SR 23; Simianer, P.

Kommentar	Leistungsbewertung: BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-TAC): 8 LP
Inhalt	We will discuss advanced programming techniques with a focus on parallelization, large data and general efficiency. The course contains practical work (in class and at home) where we will implement small projects, primarily applications of machine learning. This course also includes an introduction to the hadoop cluster of the ICL (and of course MapReduce).
Leistungsnachweis	Regular participation and completion of programming project(s).
Voraussetzung	At least 'Programmieren I'. Prior registration by Mail to lecturer.
Literatur	<ul style="list-style-type: none"> • Semisupervised Learning for Computational Linguistics • Hadoop: The Definitive Guide • http://www.ml-class.org/ • http://inst.eecs.berkeley.edu/~cs188/pacman/pacman.html • http://mahout.apache.org/

Computerlinguistisches Kolloquium

Kolloquium; SWS: 2

Do; wöch; 18:15 - 19:45; ab 26.04.2012; INF 328 / SR 25; Frank, A.;Riezler, S.;Strube, M.

Inhalt Das Kolloquium bietet Doktoranden des Instituts für Computerlinguistik sowie der Abteilung NLP des HITS (Heidelberg Institute for Theoretical Studies) ein Forum für die Vorstellung und Diskussion ihrer laufenden Doktorarbeiten, sowie gemeinsame Lektüre und Diskussion zu ausgewählten Themenbereichen der Computerlinguistik. Auch Bachelor- und Magisterabsolventen soll hier die Möglichkeit gegeben werden, ihre Abschlussarbeiten vorzustellen.

Im Rahmen des Kolloquiums finden externe Vorträge eingeladener Gastwissenschaftler des HITS und des Instituts für Computerlinguistik statt, zu denen interessierte Wissenschaftler und Studenten herzlich eingeladen sind.

Distributionelle Semantik - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Mi; wöch; 11:15 - 12:45; ab 25.04.2012; INF 325 / SR 24; Hartung, M.

Kommentar Leistungsbewertung:
BA-2010 (AS-CL): 8 LP
NBA (AS-CL): 8 LP
Master (SS-CL, SS-TAC): 8 LP

Inhalt Distributionelle Semantik ist eine aktuelle Forschungsrichtung innerhalb der Computerlinguistik und ihrer Nachbardisziplinen (Kognitionswissenschaft, Psychologie), die auf der "distributionellen Hypothese" (Harris 1951) als Grundannahme basiert. Demnach besteht eine Korrelation zwischen der Bedeutungsähnlichkeit linguistischer Einheiten (Wörter, Phrasen) und der Ähnlichkeit ihrer Verteilung über bestimmte sprachliche Kontexte, die aus empirischen Korpora gewonnen werden.

Im Unterschied zu symbolischen Bedeutungsrepräsentationen, wie sie aus der formalen Semantik bekannt sind, wird Wortbedeutung aus dem Blickwinkel der distributionellen Semantik mittels sog. Kontextvektoren repräsentiert, die die statistische Verteilung des Wortes über relevante sprachliche Kontexte erfassen. Mittels gängiger Verfahren aus der linearen Algebra können aus derartigen Kontextvektoren z.B. die semantische Ähnlichkeit zwischen einzelnen Wörtern oder die Bedeutung komplexer Phrasen berechnet werden.

Verfahren der distributionellen Semantik kommen beispielsweise in Suchmaschinen zum Einsatz, da sie im Umgang mit großen Datenmengen sehr effizient sind. Die Frage, wie viel und welche Art von linguistischem Wissen sich in derartigen Modellen erfassen lässt, ist Gegenstand einer Vielzahl von aktuellen Forschungsarbeiten, mit denen wir uns im Seminar auseinander setzen werden.

Leistungsnachweis

- Referat
- Hausarbeit
- Erwerb von "Activity Points" im Laufe des Semesters

Voraussetzung

- Formale Grundlagen
- Formale Semantik
- Statistik

Literatur

- Turney, Peter D. & Patrick Pantel (2010): From Frequency to Meaning. Vector Space Models for Semantics, in: Journal of Artificial Intelligence Research 37: 141 - 188

Weitere Literatur wird zu Beginn des Seminars bekannt gegeben.

Einführung in die lineare Algebra für Computerlinguistik - CS-CL, BS-CL, BS-AC

Vorlesung; SWS: 2

Mi; wöch; 14:15 - 15:45; ab 25.04.2012; INF 325 / SR 24; Fendrich, S.

Kommentar	Leistungsbewertung: BA-2010[100% 75%] (CS-CL): 6 LP BA-2010[50% 25%] (BS-CL, BS-AC): 4 LP NBA[100% 75%] (CS-CL): 6 LP NBA[50% 25%] (BS-CL, BS-AC): 4 LP
Inhalt	Die Vorlesung behandelt grundlegende Themen der linearen Algebra. Die Kenntnis dieser mathematischen Grundlagen ist wichtig, um die algebraischen Methoden zu verstehen, die z.B. in den Bereichen des maschinellen Lernens, des Information Retrieval oder der distributionellen Semantik eingesetzt werden. Mögliche Themenbereiche der Vorlesung sind: Theorie: <ul style="list-style-type: none"> • Vektorräume • Lineare Abbildungen • Skalarprodukt • Hyperebenen • Mehrdimensionale Ableitungen/Gradient • Determinanten • Eigenwerte und Eigenvektoren Anwendungen: <ul style="list-style-type: none"> • SVD • PCA • SVMs
Leistungsnachweis	regelmäßige Teilnahme, je nach Teilnehmerzahl Klausur oder mündliche Prüfung
Voraussetzung	<ul style="list-style-type: none"> • Formale Grundlagen • Einführung in die Statistik • Einführung in die Logik • Mathematische Grundlagen aus Schule oder Vorkurs
Literatur	Jede Einführung in die lineare Algebra. Weitere Literatur wird ggfs. im Kurs bekanntgegeben.

Einführung in die maschinelle Verarbeitung gesprochener Sprache - CS-CL, BS-CL, BS-AC

Vorlesung; SWS: 2

Mi; wöch; 09:15 - 10:45; ab 25.04.2012; INF 325 / SR 24; Jehl, L.

Kommentar	Leistungsbewertung: BA-2010[100% 75%] (CS-CL): 6 LP BA-2010[50% 25%] (BS-CL, BS-AC): 4 LP NBA[100% 75%] (CS-CL): 6 LP NBA[50% 25%] (BS-CL, BS-AC): 4 LP
Inhalt	In dieser Veranstaltung werden wir uns mit den Grundlagen der maschinellen Verarbeitung gesprochener Sprache beschäftigen. Zunächst werden wir uns einen Überblick verschaffen, wie Sprachlaute erzeugt werden, welche Eigenschaften sie haben und wie man sie mithilfe von Spektrogrammen analysieren kann. Anschließend werden wir uns mit den zwei Hauptanwendungen aus dem Bereich

der Sprachverarbeitung beschäftigen: Sprachsynthese (Speech Synthesis) und Spracherkennung (Speech Recognition). Ziel ist es sowohl die Theorie hinter diesen Anwendungen zu verstehen, als auch in praktischen Übungen das Gelernte direkt anzuwenden.

- Leistungsnachweis**
- Durchführung und Dokumentation einer der Praxisübungen
 - Bestehen der Klausur
- Voraussetzung** Erfolgreiche Teilnahme an den Veranstaltungen "Einführung in die Computerlinguistik" und "Formale Grundlagen"
- Literatur**
- P. Ladefoged (1996). Elements of Acoustic Phonetics. Second edition.
 - D. Jurafsky and J. Martin (2009). Speech and Language Processing. Second edition.
 - J. N. Holmes and W. J. Holmes (2001). Speech Synthesis and Recognition.

Einführung in die Nutzung computerlinguistischer Ressourcen

Übung; SWS: 2

Block; 09:00 - 17:00; 10.04.2012 - 13.04.2012; INF 325 / PCPool; Simianer, P.

Inhalt Der Vorkurs gibt Grundlagen in der Nutzung von GNU/Linux-basierten (computerlinguistischen) Tools sowie der vorhandenen Korpora. Dabei geht es sowohl um allgemeine Grundlagen der Shell (wie z.B. Ein-/Ausgabeumleitung) als auch um einzelne Toolkits im Speziellen. Wir werden uns anschauen, wie bestimmte Tools zu benutzen sind, was man aus ihnen herausbekommt (und was nicht) und wie man solche Ausgaben automatisch weiterverarbeiten kann (und zum Beispiel an das nächste Tool weiterverfüttert). Zudem umfasst der Kurs eine Kurzeinführung in MapReduce und den am Institut vorhandenen Hadoop Cluster.

Leistungsnachweis Die Teilnahme am Blockkurs ist Voraussetzung für die Teilnahme am Softwareprojekt.

Voraussetzung Voranmeldung bitte per E-Mail an den Dozenten.

Einführung in die Psycholinguistik für Computerlinguisten - CS-CL, BS-CL, BS-FL

Proseminar; SWS: 2

Di; wöch; 09:15 - 10:45; ab 24.04.2012; INF 327 / SR 6; Kremer, G.

Kommentar Leistungsbewertung:

BA-2010[100%|75%] (CS-CL): 6 LP
 BA-2010[50%|25%] (BS-CL, BS-FL): 4 LP
 NBA[100%|75%] (CS-CL): 6 LP
 NBA[50%|25%] (BS-CL, BS-FL): 4 LP

Inhalt Dieses Seminar gibt einen Überblick über Forschungsarbeiten in der Psycholinguistik. Dabei werden wir gemeinsam Artikel besprechen, die Themen aus den drei psycholinguistischen Teilbereichen Sprachverständnis, Sprachproduktion und Spracherwerb behandeln (z.B. Was ist das mentale Lexikon? Wie entsteht aus einem Gedanken ein gesprochener Satz? Warum beeinflussen Wörter der Zweitsprache die Sprachverarbeitung?)

Leistungsnachweis

- Angegebene Artikel kritisch lesen
- Aktive Teilnahme an Diskussionen
- Referat inklusive Handout

Voraussetzung Keine

Literatur

- Barbara Höhle (2010): Psycholinguistik
- Mark Gareth Gaskell (2007): The Oxford handbook of psycholinguistics
- Eva M. Fernández & Helen Smith Cairns (2011): Fundamentals of psycholinguistics

Einführung in psycholinguistische Methoden - AS-CL, SS-CL, SS-FAL

Hauptseminar; SWS: 2

Mi; wöch; 16:15 - 17:45; ab 25.04.2012; INF 325 / SR 24; Kremer, G.

Kommentar	Leistungsbewertung: BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-FAL): 8 LP
Inhalt	Um Ergebnisse aus computerlinguistischen Studien empirisch zu überprüfen oder durch experimentelle Daten zu ergänzen, bieten sich Methoden aus dem Bereich der Psycholinguistik an. Dieses Seminar stellt zunächst verhaltensbasierte, bildgebende und komputationelle Verfahren vor. Zusätzlich zu diesem Überblick der Methoden erörtern wir zusammen praktische Aspekte des Experimentdesigns (Festlegen der Aufgabe für den Untersuchungszweck, Auswahl der Teilnehmer, Daten sammeln, analysieren und interpretieren). In Gruppen erarbeiten die Kursteilnehmer größtenteils selbständig ein eigenes einfaches Experiment (von Grund auf ab der Idee), führen es durch und analysieren, interpretieren und präsentieren die Ergebnisse. Wöchentliche Sitzungen mit den Einzelgruppen unterstützen bei allen Schritten.
Leistungsnachweis	Durchführen eines eigenen kleinen Experiments in Gruppenarbeit, inklusive Präsentation und schriftlicher Ausarbeitung.
Voraussetzung	Beschränkt auf max. 20 Teilnehmer, verbindliche Voranmeldung beim Dozenten per E-mail.
Literatur	Wird im Hauptseminar bekanntgegeben.

Formale Grundlagen der Computerlinguistik: Mathematische Grundlagen - FF-FM

Vorlesung / Übung; SWS: 2

Do; wöch; 14:15 - 15:45; ab 26.04.2012; INF 306 / SR 13; Hartung, M.

Kommentar	Leistungsbewertung: BA-2010 (FF-FM): 6 LP NBA (FF-FM): 6 LP
Inhalt	Die Veranstaltung vermittelt eine Einführung in die Typen formaler Sprachen sowie die mathematischen Grundlagen der symbolischen und probabilistischen Sprachmodellierung. Zentrale Themen der Vorlesung sind im einzelnen: <ul style="list-style-type: none"> • Wahrscheinlichkeitsmodelle und Estimierungsverfahren • allgemeine Regelsprachen und Regelgrammatiken • Chomsky-Hierarchie der formalen Sprachen • endliche Automaten und Markov-Ketten • Hidden Markov Models • probabilistische kontextfreie Grammatiken
Leistungsnachweis	<ul style="list-style-type: none"> • Klausur • Übungsaufgaben
Voraussetzung	In der Vorlesung werden grundlegende Konzepte der Wahrscheinlichkeitstheorie vorausgesetzt, die in der Veranstaltung "Mathematischer Vorkurs" eingeführt werden. Der Besuch dieses Vorkurses wird daher allen Teilnehmern dringend empfohlen.

Achtung: Der Mathematische Vorkurs findet bereits am Ende des vorhergehenden Wintersemesters statt: <http://www.cl.uni-heidelberg.de/courses/ws11/mathevorkurs/>

- Literatur**
- Charniak, E.: Statistical Language Learning, MIT Press, 1996
 - Hopcroft, J.E. and Ullmann, J.D.: Introduction to Automata Theory, Languages and Computation, Addison Wesley, 1979
 - Klabunde, R.: Formale Grundlagen der Linguistik, Narr, 1998
 - Partee, B. et al.: Mathematical Methods in Linguistics, Kluwer, 1990
 - Vossen, G. und Witt, K.-U.: Grundlagen der Theoretischen Informatik mit Anwendungen, Vieweg, 2001

Formale Syntax - FSyn

Vorlesung / Übung; SWS: 4

Di; wöch; 11:15 - 12:45; ab 24.04.2012; INF 327 / SR 5; Frank, A.

Do; wöch; 11:15 - 12:45; ab 26.04.2012; INF 327 / SR 5; Frank, A.

Kommentar Leistungsbewertung:

NBA (FSyn): 6 LP
BA-2010 (FSyn): 6LP

Inhalt

Die Vorlesung vermittelt vertiefte Kenntnisse der syntaktischen Sprachbeschreibung am Beispiel eines theoretisch fundierten Grammatikformalismus.
Wir konzentrieren uns insbesondere auf die formalen Grundlagen der Lexikalisch-funktionalen Grammatik (LFG), als einen der bedeutenden lexikalischen constraint-basierten Grammatikformalismen.
Die Vorlesung gibt einen Überblick über die Kernbereiche und einige spezielle Phänomene der syntaktischen Analyse und ihrer formalen Beschreibung im Rahmen der Lexikalisch-funktionalen Grammatik.
Den Abschluss bildet ein Überblick über verwandte Grammatikformalismen: HPSG (Head-driven Phrase Structure Grammar), CCG (Combinatory Categorical Grammar), TAG (Tree Adjoining Grammar), Dependenzgrammatik).

Leistungsnachweis Klausur
Voraussetzung für den Scheinerwerb sind regelmäßige Präsenz, aktive Teilnahme und die erfolgreiche Bearbeitung der Übungsaufgaben (60%).

- Literatur**
- Yehuda Falk (2001): Lexical-Functional Grammar. An Introduction to Parallel Constraint-Based Syntax. University of Chicago Press.
 - Joan Bresnan (2001). Lexical-Functional Syntax. Oxford: Blackwell.
 - Mary Dalrymple (2001). Lexical Functional Grammar. Volume 34, Syntax and Semantics. Academic Press.
 - Peter Sells (1985). Lectures on Contemporary Syntactic Theories. An Introduction to Government-Binding Theory, Generalized Phrase Structure Grammar, and Lexical-Functional Grammar. CSLI Lecture Notes No. 3, CSLI Publications, Stanford, CA. (Chapter 3)
 - Tracy Holloway King, Stefanie Dipper, Anette Frank, Jonas Kuhn, John Maxwell (2004): "Ambiguity Management in Grammar Writing". In Research on Language and Computation, 2 (2): 259-280, Kluwer Academic Publications.
 - Christian Fortmann (2005): Lexikalisch-Funktionale Grammatik. Vorlesungsskript, Univ. Zürich.

Weitere Literatur wird im Laufe des Semesters bekannt gegeben.

Information Extraction - CS-CL, BS-CL, BS-AC

Proseminar; SWS: 2

Mo; wöch; 14:15 - 15:45; ab 23.04.2012; INF 325 / SR 23; Nastase, V.

Kommentar Leistungsbewertung:

BA-2010[100%|75%] (CS-CL): 6 LP
 BA-2010[50%|25%] (BS-CL, BS-AC): 4 LP
 NBA[100%|75%] (CS-CL): 6 LP
 NBA[50%|25%] (BS-CL, BS-AC): 4 LP

Inhalt In this course we will study methods to deal with large amounts of textual data, and how to extract relevant information for various tasks.

The first part of the semester will consist of lectures, the second part will consist of student presentations based on papers on the topic that I will assign to them. Throughout the semester students will have to implement an IR system, and give a demo and short presentation at the end of the course.

Leistungsnachweis

- implement a seminar project
- pass a written exam

Programmieren II: Fortgeschrittene - PII

Vorlesung / Übung; SWS: 4

Di; wöch; 14:15 - 15:45; ab 24.04.2012; INF 327 / SR 3; Spreyer, K.

Mi; wöch; 11:15 - 12:45; ab 25.04.2012; INF 327 / SR 4; Spreyer, K.

Kommentar Leistungsbewertung:

BA-2010 (PII): 6 LP
 NBA (PII): 6 LP

Inhalt In der Vorlesung werden die in "Programmieren I" erworbenen Grundkenntnisse anhand einer zweiten Programmiersprache und der Entwicklung von elementaren Kenntnissen der Softwareentwicklung erweitert und vertieft. Neben einer Einführung in objektorientierte Programmierung in der Programmiersprache Java wird dabei auch in allgemeine Themen der Softwaretechnik eingeführt; u.a. zentrale Methoden und Vorgehensweisen der Softwareentwicklung wie z.B. (objektorientiertes) Design, Software-Dokumentation, Modultests und Deployment.

Leistungsnachweis

- erfolgreiche Bearbeitung von Übungsaufgaben
- Bestehen der Abschlussklausur

Zur Klausur wird nur zugelassen, wer in den Übungsaufgaben eine Mindestanzahl von Punkten erreicht hat.

Voraussetzung Erfolgreicher Abschluss des Kurses "Programmieren I".

Literatur

- <http://docs.oracle.com/javase/6/docs/api/>
- D. Flanagan (2005). "Java in a Nutshell". O'Reilly.
- K. Sierra & B. Bates (2005). "Head First Java". O'Reilly.
- J. Bloch (2008). "Effective Java". Addison-Wesley.
- E. Freeman et al. (2004). "Head First Design Patterns". O'Reilly
- S. Oaks & H. Wong (2004). "Java Threads". O'Reilly

Statistical Learning Theory - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Di; wöch; 16:15 - 17:45; ab 24.04.2012; INF 306 / SR 21; Riezler, S.

Kommentar Leistungsbewertung:

BA-2010 (AS-CL): 8 LP
 NBA (AS-CL): 8 LP
 Master (SS-CL, SS-TAC): 8 LP

Inhalt	<p>"Statistical Learning Theory" (SLT) beschäftigt sich mit Fragestellungen, die allen Machine Learning Systemen zugrundeliegen bzw. zu deren Entwicklung geführt haben: Wie kann von begrenzten Datenmengen so gelernt werden, dass nicht nur das Rauschen der Trainingsdaten modelliert wird ("overfitting"), sondern auf ungesesehenen Testdaten gute Voraussagen gemacht werden können ("generalization performance")? Zentrale Konzepte der SLT sind "generalization bounds", also Abschätzungen des Fehlers auf Testdaten, die von der Kapazität der Machine Learning Systeme und von der Verfügbarkeit von Trainingsdaten abhängen. Ziel des Hauptseminars ist das Kennenlernen dieser mathematischen Werkzeuge, besonders in Hinblick auf konkrete Anwendungen zur "model selection". Wir werden folgende Kapazitätskonzepte kennenlernen:</p> <ul style="list-style-type: none"> • Vapnik-Chervonenkis (VC) Dimension • Covering Numbers • Rademacher Penalization • Bootstrap Penalization • Resampling Penalization
Voraussetzung	<p>Grundbegriffe von Statistik und Machine Learning aus einschlägigen Veranstaltungen, z.B. aus Statistical Methods for Computational Linguistics (http://www.cl.uni-heidelberg.de/courses/ws11/statistics/)</p>
Literatur	<p>Eine Liste möglicher Referatsthemen wird in der ersten Seminarsitzung bekanntgegeben.</p>

Statistical Methods for Information Retrieval - AS-CL, SS-CL, SS-TAC

Vorlesung / Übung; SWS: 4

Di; wöch; 11:15 - 12:45; ab 24.04.2012; INF 327 / SR 1; Vorlesung; Riezler, S.

Do; wöch; 11:15 - 12:45; ab 26.04.2012; INF 327 / SR 4; Übung; Riezler, S.

Kommentar	<p>Leistungsbewertung:</p> <p>BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-TAC): 8 LP</p>
Inhalt	<p>Die Vorlesung führt in statistische Methoden des Information Retrieval ein, mit einem Fokus auf mathematische und algorithmische Grundlagen. Mögliche Themenbereiche sind:</p> <ul style="list-style-type: none"> • Random Walks for Ranking: Page Rank and HITS • Randomization for Duplicate Detection • Statistical Language Models for Ranking • Statistical Machine Translation for Ranking and Query Expansion • Supervised Learning-to-Rank for Information Retrieval • Latent Semantic Indexing • Sentiment Classification • Social Network Analysis
Leistungsnachweis	<ul style="list-style-type: none"> • Regelmäßige Teilnahme an Vorlesung und Übung • Bearbeitung von Übungsaufgaben • Klausur
Voraussetzung	<p>Grundbegriffe des Information Retrieval, z.B. aus dem Hauptseminar Information Retrieval WS11/12 (http://kontext.fraunhofer.de/haenelt/kurs/InfoRet/titel.php)</p>
Literatur	<p>Grundlage der Vorlesung ist:</p>

- Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press. (<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>)
- Bing Liu (2007). Web Data Mining. Springer. (<http://www.cs.uic.edu/~liub/WebMiningBook.html>)

Topic Models - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Do; wöch; 14:15 - 15:45; ab 26.04.2012; INF 325 / SR 24; Nastase, V.

Kommentar	Leistungsbewertung: BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-TAC): 8 LP
Inhalt	In this seminar we will understand what topic models are, and how they are useful for the processing of texts. We will study first the basic topic model (LDA), and then extensions of this in various dimensions. Course organization: In the first part of the semester I will give lectures, in the second part, the students will present and discuss papers that I will assign to them on the topic. Throughout the semester the students will have to implement a topic model, and give a demo and short presentation at the end of the course.
Leistungsnachweis	<ul style="list-style-type: none"> • implement a seminar project • pass a written exam
Literatur	<ul style="list-style-type: none"> • http://cocosci.berkeley.edu/tom/papers/SteinGriffiths.pdf • http://www.cs.princeton.edu/~blei/papers/Blei2011.pdf • http://videlectures.net/mlss09uk_blei_tm/

Master

Abschlusskolloquium - Coll

Kolloquium; SWS: 3

Di; k.A.; 18:15 - 20:30; ab 24.04.2012; INF 325 / SR 24; Frank, A.; Nastase, V.; Riezler, S.

Kommentar	Leistungsbewertung: BA-2010 (Coll): 2 LP NBA (-): Freiwillig, keine LP Master (Coll): 2 LP
Inhalt	Präsentation laufender BA-, MA- und Magisterarbeiten Das Abschlusskolloquium bietet BA-, MA- und Magisterstudierenden ein Forum für die Vorstellung und Diskussion ihrer Abschlussarbeiten. Die Studierenden präsentieren den aktuellen Stand ihrer Arbeit und erhalten in der Diskussion Anregungen von Seiten der Studierenden und der Dozenten. Externe Vorträge Darüber hinaus bietet das Abschlusskolloquium allen Studierenden durch Vorträge geladener Gäste Einblicke in aktuelle Forschungsfragen der Computerlinguistik. Externe Vorträge finden im Rahmen des Computerlinguistischen Kolloquiums statt.

Leistungsnachweis Vortrag (BA-2010, MA) und Ausarbeitung; regelmäßige Präsenz ist Voraussetzung für den Scheinerwerb.
Ein Leistungserwerb ist nur für Examenskandidat/innen im Bachelorstudiengang (BA-2010) und Masterstudiengang (MA) vorgesehen. Jedoch sind alle Studierenden eingeladen, ihre Abschlussarbeiten vorzustellen, den Vorträgen zuzuhören und sich an den Diskussionen zu beteiligen.

Advanced Programming - AS-CL, SS-CL, SS-TAC

Vorlesung / Übung; SWS: 2

Mo; wöch; 11:15 - 12:45; ab 23.04.2012; INF 325 / SR 23; Simianer, P.

Kommentar Leistungsbewertung:

BA-2010 (AS-CL): 8 LP

NBA (AS-CL): 8 LP

Master (SS-CL, SS-TAC): 8 LP

Inhalt We will discuss advanced programming techniques with a focus on parallelization, large data and general efficiency. The course contains practical work (in class and at home) where we will implement small projects, primarily applications of machine learning. This course also includes an introduction to the hadoop cluster of the ICL (and of course MapReduce).

Leistungsnachweis Regular participation and completion of programming project(s).

Voraussetzung At least 'Programmieren I'.

Prior registration by Mail to lecturer.

Literatur

- Semisupervised Learning for Computational Linguistics
- Hadoop: The Definitive Guide
- <http://www.ml-class.org/>
- <http://inst.eecs.berkeley.edu/~cs188/pacman/pacman.html>
- <http://mahout.apache.org/>

Computerlinguistisches Kolloquium

Kolloquium; SWS: 2

Do; wöch; 18:15 - 19:45; ab 26.04.2012; INF 328 / SR 25; Frank, A.;Riezler, S.;Strube, M.

Inhalt Das Kolloquium bietet Doktoranden des Instituts für Computerlinguistik sowie der Abteilung NLP des HITS (Heidelberg Institute for Theoretical Studies) ein Forum für die Vorstellung und Diskussion ihrer laufenden Doktorarbeiten, sowie gemeinsame Lektüre und Diskussion zu ausgewählten Themenbereichen der Computerlinguistik. Auch Bachelor- und Magisterabsolventen soll hier die Möglichkeit gegeben werden, ihre Abschlussarbeiten vorzustellen.

Im Rahmen des Kolloquiums finden externe Vorträge eingeladener Gastwissenschaftler des HITS und des Instituts für Computerlinguistik statt, zu denen interessierte Wissenschaftler und Studenten herzlich eingeladen sind.

Distributionelle Semantik - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Mi; wöch; 11:15 - 12:45; ab 25.04.2012; INF 325 / SR 24; Hartung, M.

Kommentar Leistungsbewertung:

BA-2010 (AS-CL): 8 LP

NBA (AS-CL): 8 LP
 Master (SS-CL, SS-TAC): 8 LP

Inhalt	<p>Distributionelle Semantik ist eine aktuelle Forschungsrichtung innerhalb der Computerlinguistik und ihrer Nachbardisziplinen (Kognitionswissenschaft, Psychologie), die auf der "distributionellen Hypothese" (Harris 1951) als Grundannahme basiert. Demnach besteht eine Korrelation zwischen der Bedeutungsähnlichkeit linguistischer Einheiten (Wörter, Phrasen) und der Ähnlichkeit ihrer Verteilung über bestimmte sprachliche Kontexte, die aus empirischen Korpora gewonnen werden.</p> <p>Im Unterschied zu symbolischen Bedeutungsrepräsentationen, wie sie aus der formalen Semantik bekannt sind, wird Wortbedeutung aus dem Blickwinkel der distributionellen Semantik mittels sog. Kontextvektoren repräsentiert, die die statistische Verteilung des Wortes über relevante sprachliche Kontexte erfassen. Mittels gängiger Verfahren aus der linearen Algebra können aus derartigen Kontextvektoren z.B. die semantische Ähnlichkeit zwischen einzelnen Wörtern oder die Bedeutung komplexer Phrasen berechnet werden.</p> <p>Verfahren der distributionellen Semantik kommen beispielsweise in Suchmaschinen zum Einsatz, da sie im Umgang mit großen Datenmengen sehr effizient sind. Die Frage, wie viel und welche Art von linguistischem Wissen sich in derartigen Modellen erfassen lässt, ist Gegenstand einer Vielzahl von aktuellen Forschungsarbeiten, mit denen wir uns im Seminar auseinander setzen werden.</p>
Leistungsnachweis	<ul style="list-style-type: none"> • Referat • Hausarbeit • Erwerb von "Activity Points" im Laufe des Semesters
Voraussetzung	<ul style="list-style-type: none"> • Formale Grundlagen • Formale Semantik • Statistik
Literatur	<ul style="list-style-type: none"> • Turney, Peter D. & Patrick Pantel (2010): From Frequency to Meaning. Vector Space Models for Semantics, in: Journal of Artificial Intelligence Research 37: 141 - 188 <p>Weitere Literatur wird zu Beginn des Seminars bekannt gegeben.</p>

Einführung in psycholinguistische Methoden - AS-CL, SS-CL, SS-FAL

Hauptseminar; SWS: 2

Mi; wöch; 16:15 - 17:45; ab 25.04.2012; INF 325 / SR 24; Kremer, G.

Kommentar	<p>Leistungsbewertung:</p> <p>BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-FAL): 8 LP</p>
Inhalt	<p>Um Ergebnisse aus computerlinguistischen Studien empirisch zu überprüfen oder durch experimentelle Daten zu ergänzen, bieten sich Methoden aus dem Bereich der Psycholinguistik an. Dieses Seminar stellt zunächst verhaltensbasierte, bildgebende und komputationelle Verfahren vor. Zusätzlich zu diesem Überblick der Methoden erörtern wir zusammen praktische Aspekte des Experimentdesigns (Festlegen der Aufgabe für den Untersuchungszweck, Auswahl der Teilnehmer, Daten sammeln, analysieren und interpretieren). In Gruppen erarbeiten die Kursteilnehmer größtenteils selbständig ein eigenes einfaches Experiment (von Grund auf ab der Idee), führen es durch und analysieren, interpretieren und präsentieren die Ergebnisse. Wöchentliche Sitzungen mit den Einzelgruppen unterstützen bei allen Schritten.</p>

Leistungsnachweis	Durchführen eines eigenen kleinen Experiments in Gruppenarbeit, inklusive Präsentation und schriftlicher Ausarbeitung.
Voraussetzung	Beschränkt auf max. 20 Teilnehmer, verbindliche Voranmeldung beim Dozenten per E-mail.
Literatur	Wird im Hauptseminar bekanntgegeben.

Forschungsmodul für MA-Studenten - SS-FM

Hauptseminar; SWS: 2; LP: 20

Di; wöch; 14:15 - 15:45; ab 24.04.2012; INF 325 / SR 24; Frank, A.

Kommentar	Leistungsbewertung: Master (SS-FM): 20 LP
Inhalt	Die Veranstaltung dient zur Einübung selbstständigen wissenschaftlichen Arbeitens und dient damit auch als Vorbereitung für die Master-Arbeit. In Absprache mit den Teilnehmern werden Themen der aktuellen Forschung festgelegt, in deren Rahmen Forschungsprojekte definiert und durchgeführt werden. Die Projektarbeit umfasst Literaturrecherche, Entwurf eines Forschungsprojekts, Projektdurchführung und Implementation, sowie wissenschaftliche Aufarbeitung und Präsentation der Ergebnisse. Themen des Forschungsmoduls im SS 2012 sind Vertiefungsgebiete zu diskursorientierter semantischer Verarbeitung. Kenntnisse aus dem HS "Semantic NLP - from a discourse perspective" (WS 2011/12) werden vorausgesetzt. Teilnehmer dieses Hauptseminars werden bevorzugt berücksichtigt.
Leistungsnachweis	Erfolgreiche Bearbeitung aller Teilaspekte des Forschungsprojekts
Voraussetzung	Anmeldung bis 15.02.2012; Vorbesprechung (obligatorische Teilnahme): 20.02.2012
Literatur	Siehe Hauptseminar Frank 2011/12

Statistical Learning Theory - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Di; wöch; 16:15 - 17:45; ab 24.04.2012; INF 306 / SR 21; Riezler, S.

Kommentar	Leistungsbewertung: BA-2010 (AS-CL): 8 LP NBA (AS-CL): 8 LP Master (SS-CL, SS-TAC): 8 LP
Inhalt	"Statistical Learning Theory" (SLT) beschäftigt sich mit Fragestellungen, die allen Machine Learning Systemen zugrundeliegen bzw. zu deren Entwicklung geführt haben: Wie kann von begrenzten Datenmengen so gelernt werden, dass nicht nur das Rauschen der Trainingsdaten modelliert wird ("overfitting"), sondern auf ungesehenen Testdaten gute Voraussagen gemacht werden können ("generalization performance")? Zentrale Konzepte der SLT sind "generalization bounds", also Abschätzungen des Fehlers auf Testdaten, die von der Kapazität der Machine Learning Systeme und von der Verfügbarkeit von Trainingsdaten abhängen. Ziel des Hauptseminars ist das Kennenlernen dieser mathematischen Werkzeuge, besonders in Hinblick auf konkrete Anwendungen zur "model selection". Wir werden folgende Kapazitätskonzepte kennenlernen: <ul style="list-style-type: none"> • Vapnik-Chervonenkis (VC) Dimension • Covering Numbers

- Rademacher Penalization
- Bootstrap Penalization
- Resampling Penalization

Voraussetzung Grundbegriffe von Statistik und Machine Learning aus einschlägigen Veranstaltungen, z.B. aus Statistical Methods for Computational Linguistics (<http://www.cl.uni-heidelberg.de/courses/ws11/statistics/>)

Literatur Eine Liste möglicher Referatsthemen wird in der ersten Seminarsitzung bekanntgegeben.

Statistical Methods for Information Retrieval - AS-CL, SS-CL, SS-TAC

Vorlesung / Übung; SWS: 4

Di; wöch; 11:15 - 12:45; ab 24.04.2012; INF 327 / SR 1; Vorlesung; Riezler, S.

Do; wöch; 11:15 - 12:45; ab 26.04.2012; INF 327 / SR 4; Übung; Riezler, S.

Kommentar Leistungsbewertung:

BA-2010 (AS-CL): 8 LP
 NBA (AS-CL): 8 LP
 Master (SS-CL, SS-TAC): 8 LP

Inhalt Die Vorlesung führt in statistische Methoden des Information Retrieval ein, mit einem Fokus auf mathematische und algorithmische Grundlagen. Mögliche Themenbereiche sind:

- Random Walks for Ranking: Page Rank and HITS
- Randomization for Duplicate Detection
- Statistical Language Models for Ranking
- Statistical Machine Translation for Ranking and Query Expansion
- Supervised Learning-to-Rank for Information Retrieval
- Latent Semantic Indexing
- Sentiment Classification
- Social Network Analysis

Leistungsnachweis

- Regelmäßige Teilnahme an Vorlesung und Übung
- Bearbeitung von Übungsaufgaben
- Klausur

Voraussetzung Grundbegriffe des Information Retrieval, z.B. aus dem Hauptseminar Information Retrieval WS11/12 (<http://kontext.fraunhofer.de/haenelt/kurs/InfoRet/titel.php>)

Literatur Grundlage der Vorlesung ist:

- Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze (2008). Introduction to Information Retrieval. Cambridge University Press. (<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>)
- Bing Liu (2007). Web Data Mining. Springer. (<http://www.cs.uic.edu/~liub/WebMiningBook.html>)

Topic Models - AS-CL, SS-CL, SS-TAC

Hauptseminar; SWS: 2

Do; wöch; 14:15 - 15:45; ab 26.04.2012; INF 325 / SR 24; Nastase, V.

Kommentar Leistungsbewertung:

BA-2010 (AS-CL): 8 LP
 NBA (AS-CL): 8 LP
 Master (SS-CL, SS-TAC): 8 LP

Inhalt	<p>In this seminar we will understand what topic models are, and how they are useful for the processing of texts. We will study first the basic topic model (LDA), and then extensions of this in various dimensions.</p> <p>Course organization: In the first part of the semester I will give lectures, in the second part, the students will present and discuss papers that I will assign to them on the topic. Throughout the semester the students will have to implement a topic model, and give a demo and short presentation at the end of the course.</p>
Leistungsnachweis	<ul style="list-style-type: none"> • implement a seminar project • pass a written exam
Literatur	<ul style="list-style-type: none"> • http://cocosci.berkeley.edu/tom/papers/SteiyversGriffiths.pdf • http://www.cs.princeton.edu/~blei/papers/Blei2011.pdf • http://videlectures.net/mlss09uk_blei_tm/

Magister

Abschlusskolloquium - Coll

Kolloquium; SWS: 3

Di; k.A.; 18:15 - 20:30; ab 24.04.2012; INF 325 / SR 24; Frank, A.;Nastase, V.;Riezler, S.

Kommentar	<p>Leistungsbewertung:</p> <p>BA-2010 (Coll): 2 LP NBA (-): Freiwillig, keine LP Master (Coll): 2 LP</p>
Inhalt	<p>Präsentation laufender BA-, MA- und Magisterarbeiten</p> <p>Das Abschlusskolloquium bietet BA-, MA- und Magisterstudierenden ein Forum für die Vorstellung und Diskussion ihrer Abschlussarbeiten. Die Studierenden präsentieren den aktuellen Stand ihrer Arbeit und erhalten in der Diskussion Anregungen von Seiten der Studierenden und der Dozenten.</p> <p>Externe Vorträge</p> <p>Darüber hinaus bietet das Abschlusskolloquium allen Studierenden durch Vorträge geladener Gäste Einblicke in aktuelle Forschungsfragen der Computerlinguistik.</p> <p>Externe Vorträge finden im Rahmen des Computerlinguistischen Kolloquiums statt.</p>
Leistungsnachweis	<p>Vortrag (BA-2010, MA) und Ausarbeitung; regelmäßige Präsenz ist Voraussetzung für den Scheinerwerb.</p> <p>Ein Leistungserwerb ist nur für Examenskandidat/innen im Bachelorstudiengang (BA-2010) und Masterstudiengang (MA) vorgesehen. Jedoch sind alle Studierenden eingeladen, ihre Abschlussarbeiten vorzustellen, den Vorträgen zuzuhören und sich an den Diskussionen zu beteiligen.</p>

Computerlinguistisches Kolloquium

Kolloquium; SWS: 2

Do; wöch; 18:15 - 19:45; ab 26.04.2012; INF 328 / SR 25; Frank, A.;Riezler, S.;Strube, M.

Inhalt	<p>Das Kolloquium bietet Doktoranden des Instituts für Computerlinguistik sowie der Abteilung NLP des HITS (Heidelberg Institute for Theoretical Studies) ein Forum für die</p>
---------------	---

Vorstellung und Diskussion ihrer laufenden Doktorarbeiten, sowie gemeinsame Lektüre und Diskussion zu ausgewählten Themenbereichen der Computerlinguistik. Auch Bachelor- und Magisterabsolventen soll hier die Möglichkeit gegeben werden, ihre Abschlussarbeiten vorzustellen.

Im Rahmen des Kolloquiums finden externe Vorträge eingeladener Gastwissenschaftler des HITS und des Instituts für Computerlinguistik statt, zu denen interessierte Wissenschaftler und Studenten herzlich eingeladen sind.

Einführung in die Nutzung computerlinguistischer Ressourcen

Übung; SWS: 2

Block; 09:00 - 17:00; 10.04.2012 - 13.04.2012; INF 325 / PCPool; Simianer, P.

Inhalt Der Vorkurs gibt Grundlagen in der Nutzung von GNU/Linux-basierten (computerlinguistischen) Tools sowie der vorhandenen Korpora. Dabei geht es sowohl um allgemeine Grundlagen der Shell (wie z.B. Ein-/Ausgabeumleitung) als auch um einzelne Toolkits im Speziellen. Wir werden uns anschauen, wie bestimmte Tools zu benutzen sind, was man aus ihnen herausbekommt (und was nicht) und wie man solche Ausgaben automatisch weiterverarbeiten kann (und zum Beispiel an das nächste Tool weiterverfütert). Zudem umfasst der Kurs eine Kurzeinführung in MapReduce und den am Institut vorhandenen Hadoop Cluster.

Leistungsnachweis Die Teilnahme am Blockkurs ist Voraussetzung für die Teilnahme am Softwareprojekt.

Voraussetzung Voranmeldung bitte per E-Mail an den Dozenten.