



Modulkatalog

Digitale Linguistik

Programmformat: Minor 30

Studienstufe: Master

Gültig ab: Herbstsemester 2019

[Erstellt am 10.09.2020]

Modulgruppen des Programms

Wissenschaftliche Vertiefung

Kernbereich Digitale Linguistik

Praxis der Digitalen Linguistik

Der Modulkatalog enthält alle beständigen Angaben zu den Pflicht- und Wahlpflichtmodulen des Programms, die semesterbezogenen Informationen dazu entnehmen Sie dem Vorlesungsverzeichnis.

Ebenfalls im Vorlesungsverzeichnis finden Sie das aktuelle Angebot an Wahlmodulen sowie weiterführende Informationen zu Modulen anderer Fakultäten.



Wissenschaftliche Vertiefung

06SM521-s04 [Seminar]

5



Kernbereich Digitale Linguistik

| | | |
|-------------|---|----|
| 06SM271-518 | Language Data Processing | 6 |
| 06SM271-516 | Quantitative Methods | 7 |
| 06SM271-517 | Language Data Acquisition | 8 |
| 06SM521-002 | Einführung in die Computerlinguistik 2 | 9 |
| 06SM521-003 | Mathematische Grundlagen der Computerlinguistik | 10 |
| 06SM521-005 | Programmiertechniken der Computerlinguistik 2 | 11 |
| 06SM521-006 | Erstellung und Annotation von linguistischen Ressourcen | 12 |
| 06SM521-007 | Grammatikformalismen und Parsing | 13 |
| 06SM521-008 | Maschinelle Übersetzung | 14 |
| 06SM521-009 | Maschinelles Lernen in der Computerlinguistik | 15 |
| 06SM521-011 | Semantic Web & Linked Data | 16 |
| 06SM521-012 | Sprachtechnologie als Beitrag zur Barrierefreiheit | 17 |
| 06SM521-013 | Sprachtechnologische Webapplikationen | 18 |
| 06SM521-014 | Text Analytics in the Digital Humanities | 19 |
| 06SM521-015 | Text Mining | 20 |
| 06SM521-519 | Fundamentals of speech sciences and signal processing | 21 |
| 06SM521-520 | Instrumental techniques of phonetic research | 22 |
| 06SM521-521 | Phonetic Transcription | 23 |
| 06SM521-529 | Intermediate Methods and Programming in Digital Linguistics | 24 |



Praxis der Digitalen Linguistik

| | | |
|-------------|--------------------|----|
| 06SM521-017 | Praktikum intern | 25 |
| 06SM521-018 | Programmierprojekt | 26 |



[Seminar]

06SM521-s04

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, einmalig

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit keine Wiederholungsmöglichkeit

Leistungsnachweis Portfolio (50% schriftliche Arbeit und 50% Referat)

Unterrichtssprache

Lehrformen Seminar

Lernziel

Die Studierenden: - vertiefen ihr Wissen in einem spezifischen Gebiet; - erlernen die methodischen Grundkompetenzen wissenschaftlicher Arbeit; - üben das verständliche Präsentieren komplexer Sachverhalte; - können eine wissenschaftliche Arbeit verfassen.

Allgemeine Beschreibung

Das Seminar dient der wissenschaftlichen Bearbeitung eines spezifischen Themenbereichs. Dabei werden die Methoden wissenschaftlichen Arbeitens erlernt, z.B. das Rezipieren von Fachliteratur, die Informationssuche, das Strukturieren von Fakten und Theorien, sowie das eigenständige Beurteilen empirischer Befunde. Daneben lernen die Studierenden einen Vortrag zu konzipieren und zu halten. Sie lernen mit anderen über das Gehörte zu diskutieren und es zu bewerten. Schliesslich erwerben sie die Fähigkeit zur schriftlichen Ausarbeitung ihres Vortrags.

Voraussetzungen

-



Language Data Processing

06SM271-518

Anbietende Organisationseinheit

PhF: Linguistik Zentrum Zürich

ECTS Credits

15

Angebotsmuster

1-semesterig, jedes Herbstsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, Wiederholungsprüfung oder erneut buchen

Leistungsnachweis

Portfolio (80% written exam and 20% proof of self-study achievements)

Unterrichtssprache

Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Students get to know the core methods and tools for automatic corpus analysis, annotation and evaluation. They learn about cross-language alignment and gain insights into the advantages of parallel corpora. Students learn how to use Unix language processing tools and obtain programming knowledge in a modern scripting language (e.g. R or Python) with a focus on the processing of linguistic data.

Allgemeine Beschreibung

This 2-course-module introduces automatic corpus annotation ("Introduction to Language Data Processing") and programming ("Programming for Linguists") and consists of the following parts: Annotation, manipulation and extraction of linguistic data, basic skills in Natural Language Processing on various linguistic levels (morphology, syntax, semantics), basic Unix commands for text handling, regular expressions for pattern matching, file formats and markup languages, encoding and compression, programming in a modern scripting language (e.g. R or Python), introduction to linguistic databases.

Voraussetzungen

none

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Quantitative Methods

06SM271-516

Anbietende Organisationseinheit PhF: Linguistik Zentrum Zürich

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Herbstsemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, Wiederholungsprüfung oder erneut buchen

Leistungsnachweis Portfolio (80% written exam and 20% written exercises)

Unterrichtssprache Englisch

Lehrformen Vorlesung mit integrierter Übung, Tutorat

Lernziel

Students are familiar with the basic concepts and methods of statistical analyses of linguistic data and are able to perform such analyses themselves.

Allgemeine Beschreibung

This course introduces the basic concepts of statistical analysis as used in modern linguistics, covering data description and visualization as well as basic techniques of machine learning and modelling. The course also introduces the basic concepts of frequentist vs Bayesian approaches and the use of simulations and baseline models.

Voraussetzungen

Notice: the following knowledge of high school mathematics is required and has to be solid: - Concept of spaces / number systems (e.g., natural, rational and real numbers); - Basic Functions: Linear, polynomial, exponential, logarithmic function; - Differential Calculus: Extreme values, derivatives, integrals; - Linear Algebra: Vectors, vector spaces, linear transformations, matrices, dot (scalar) product; - Probability Theory: Random variables, probability distributions (uniform, binomial, normal, exponential), joint and marginal distributions.

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Language Data Acquisition

06SM271-517

Anbietende Organisationseinheit

PhF: Linguistik Zentrum Zürich

ECTS Credits

9

Angebotsmuster

1-semesterig, jedes Herbstsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, Wiederholungsprüfung oder erneut buchen

Leistungsnachweis

Portfolio (80% written exam and 20% proof of self-study achievements)

Unterrichtssprache

Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Students are familiar with the core methods and techniques of data collection both in experimental and naturalistic settings: They know how to design and carry out experiments, are familiar with the use of video and audio recording devices and editing tools, and know how to design and build up a corpus. They have a basic understanding of the constraints on data acquisition for both qualitative and quantitative purposes and are familiar with ethical and legal issue of data.

Allgemeine Beschreibung

This course introduces methods and techniques of linguistic data acquisition and consists of the following parts: Audiovisual techniques of data acquisition, metadata management, transcription, corpus design and corpus building, conversions (e.g. OCR), experiments (including field experiments), basic qualitative methods as well as ethical and legal issues.

Voraussetzungen

none

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Einführung in die Computerlinguistik 2

Introduction to Computational Linguistics 2

06SM521-002

| | |
|--|---|
| Anbietende Organisationseinheit | PhF: Institut für Computerlinguistik |
| ECTS Credits | 3 |
| Angebotsmuster | 1-semesterig, jedes Frühlingssemester |
| Bewertung/Benotung | 1-6, in Halbschritten |
| Repetierbarkeit | einmal wiederholbar, erneut buchen |
| Leistungsnachweis | Portfolio (75% schriftliche Prüfung und 25% Nachweis von im Selbststudium erbrachten Studienleistungen) |
| Unterrichtssprache | Deutsch und/oder Englisch |
| Lehrformen | Vorlesung mit integrierter Übung, Tutorat |

Lernziel

Die Studierenden: - kennen die Anwendungs- und Problembereiche der Semantik innerhalb der Computerlinguistik: lexikalische Semantik, insbesondere Wortnetze, Framesemantik und Satzsemantik; - kennen Logik als Basis der Satzsemantik; - gewinnen Einblick in weitere Probleme beim automatischen Verarbeiten der natürlichen Sprache; - kennen computerlinguistische Verfahren zur Berechnung von semantischen Strukturen und Bedeutungsdesambiguierung; - erarbeiten praktische Fähigkeiten für die Modellierung von semantischen Beziehungen mit Wortnetzen, für die einfache semantische Annotation von Sätzen in FrameNet und für die Übersetzung von Sätzen in logische Formen; - erlangen Kenntnisse über die Modellierung von Wissen, die Grundlagen des Semantic Web und Basiskenntnisse in XML.

Allgemeine Beschreibung

Die Vorlesung beschäftigt sich mit der Semantik der natürlichen Sprache, d.i. die lexikalische Semantik und die Satzsemantik. Wir beginnen mit dem Problem der Ambiguität, diskutieren die Grundlagen der lexikalischen Semantik und widmen uns dann den sog. Sinnrelationen (lexikalische Funktionen), die in den Wortnetzen zur Strukturierung des Wortschatzes eingesetzt werden. Wir werden ein probabilistisches Verfahren zur Bedeutungsdesambiguierung kennenlernen. Danach behandeln wir die Aussagen- und Prädikatenlogik und untersuchen, wie man Sätze mittels Prädikatenlogik semantisch repräsentieren kann. Schliesslich befassen wir uns mit dem Semantic Web, also dem Versuch, Internetseiten mit semantischer Information anzureichern, um die Informationssuche flexibler zu gestalten.

Voraussetzungen

Keine, jedoch Modul «Einführung in die Computerlinguistik 1» empfohlen.



Mathematische Grundlagen der Computerlinguistik

Mathematical Foundations of Computational Linguistics

06SM521-003

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühlingsemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis Portfolio (90% schriftliche Prüfung und 10% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung, Tutorat

Lernziel

Die Studierenden: - kennen die mathematischen Grundlagen für die quantitative Sprachverarbeitung; dazu gehören Wahrscheinlichkeitstheorie, Kombinatorik, deskriptive Statistik, Inferenzstatistik und wichtige Konzepte der linearen Algebra.

Allgemeine Beschreibung

Dieser Kurs behandelt die mathematischen Grundlagen der Computerlinguistik. Es werden die zentralen Konzepte der Wahrscheinlichkeitstheorie und Kombinatorik eingeführt. Ausserdem behandeln wir Fragestellungen der Statistik (Deskriptive Statistik, Inferenzstatistik) und der linearen Algebra. Der Kurs vermittelt die theoretischen Grundlagen und dient zudem der praktischen Einübung und Anwendung der numerischen Verfahren.

Voraussetzungen

Programmierkenntnisse im Umfang von «Programmiertechniken in der Computerlinguistik 1».



Programmiertechniken der Computerlinguistik 2

Programming Techniques in Computational Linguistics 2

06SM521-005

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühlingssemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis Portfolio (75% schriftliche Prüfung und 25% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung, Tutorat

Lernziel

Die Studierenden: - beherrschen die benötigten Programmiertechniken und -kenntnisse zur Lösung einfacher sprachtechnologischer Aufgaben, insbesondere objektorientiertes Programmieren, Speicher-/Laufzeitoptimierung im Umgang mit grossen Datenmengen und einfaches maschinelles Lernen.

Allgemeine Beschreibung

Ziel dieses Kurses ist die Vertiefung von Programmierkenntnissen in Python. Relevante Themen und Techniken (s. Lernziele) werden anhand von ausgewählten Anwendungen aus dem Bereich der Computerlinguistik vermittelt.

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang; Kenntnisse im Umfang von «Einführung in die Computerlinguistik 1».



Erstellung und Annotation von linguistischen Ressourcen

Creation and Annotation of Linguistic Resources

06SM521-006

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung bestanden/nicht bestanden

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis Nachweis von im Selbststudium erbrachten Studienleistungen

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung

Lernziel

Die Studierenden: - kennen und beherrschen alle Teilschritte, welche für die nachhaltige Erstellung von sprachtechnologischen Ressourcen notwendig sind; - kennen gängige Verfahren, um manuelle Annotation effizient und in guter Qualität zu erzeugen; - produzieren gemeinsam eine annotierte sprachtechnologische Ressource, welche wissenschaftliche Standards erfüllt.

Allgemeine Beschreibung

Manuell oder automatisch annotierte Ressourcen sind mittlerweile die Grundlage für viele computerlinguistische Systeme. Dieses Modul vermittelt die Grundlagen, Probleme und Best-Practice-Verfahren der Erstellung von Korpora und (ihrer) Annotation anhand eines praktischen Projekts zur Ressourcenerstellung, welche alle Teilschritte umfasst. Bezüglich Ressourcen werden Themen angesprochen wie Rohdatenbeschaffung (Crawling), automatische Datenaufbereitung, Verwendung von nachhaltigen XML-Standards als Austauschformat, Lizenzierung. Das Thema Annotation umfasst neben den klassischen automatischen computerlinguistischen Verfahren die Erstellung von Guidelines und Dokumentation, manuelle Annotation (Crowdsourcing) und Inter-Annotator-Übereinstimmung sowie effiziente halbautomatische Verfahren.

Voraussetzungen

Programmierkenntnisse



Grammatikformalismen und Parsing

Grammar Formalisms and Parsing

06SM521-007

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Herbstsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (90% schriftliche Prüfung und 10% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung

Lernziel

Die Studierenden: - kennen die Fragestellungen, Probleme und Lösungsansätze der Syntaxanalyse; - verstehen ausgewählte Grammatikformalismen und können sie anwenden, sowie ausgewählte Grammatikphänomene des Deutschen modellieren; - arbeiten sich ein in verschiedene Ansätze zum (probabilistischen) Parsing und ihrem Training anhand von Baumbanken.

Allgemeine Beschreibung

Der Kurs behandelt die automatische Syntaxanalyse. Dazu werden die zentralen Begriffe der syntaktischen Analyse eingeführt, es werden verschiedene Grammatikformalismen behandelt und anhand einer ausgewählten Theorie wird Grammatikmodellierung betrieben. Daneben widmen wir uns den probabilistischen Verfahren der Syntaxanalyse (Parsing), ihren Grundlagen, ihrem Training anhand von Baumbanken und ihrer Evaluierung.

Voraussetzungen

Programmierkenntnisse im Umfang von «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang.



Maschinelle Übersetzung

Machine Translation

06SM521-008

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (75% schriftliche Prüfung und 25% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung

Lernziel

Die Studierenden gewinnen praktische Erfahrungen und ein vertieftes Verständnis im Bereich der Maschinellen Übersetzung.

Allgemeine Beschreibung

Der Kurs besteht aus Vorlesungen und praktischen Übungen in Maschinellem Übersetzung. Die Teilnehmer bauen und evaluieren MÜ-Systeme über unterschiedlichen Korpora. Themenbeispiele: - Einführung in die maschinelle Übersetzung - Parallele Korpora und Alignierung - Aktuelle Methoden der Maschinellen Übersetzung - Hybride maschinelle Übersetzungssysteme - Gesellschaftliche Auswirkungen der Maschinellen Übersetzung

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang; Modul «Einführung in die Computerlinguistik 1» empfohlen.



Maschinelles Lernen in der Computerlinguistik

Machine Learning in Computational Linguistics

06SM521-009

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Herbstsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (90% schriftliche Prüfung und 10% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Die Studierenden: - verstehen die Methodik des supervisierten und unsupervisierten maschinellen Lernens; - kennen die Grundlagen, Techniken und Anwendungen der wichtigsten maschinellen Lernverfahren, die in der Computerlinguistik und Sprachtechnologie relevant sind, insbesondere Regression, Klassifikation und Sequenzklassifikation; - erwerben die praktischen Fähigkeiten, um mit einem modernen Machine-Learning-Framework in Python selbständig konkrete sprachtechnologische Probleme zu lösen.

Allgemeine Beschreibung

Die Methodik des supervisierten und unsupervisierten maschinellen Lernens beinhaltet das Verstehen der Funktion von Trainings- Entwicklungs- und Testdaten, die Vorverarbeitung von Daten, die Merkmalsextraktion und -auswahl, Modellauswahl, Optimierung von Hyperparametern, Evaluationsmetriken und -techniken (Kreuzvalidierung), Visualisierungsmöglichkeiten. Die grundlegenden, für die Sprachtechnologie relevanten, maschinellen Lernverfahren werden theoretisch und praktisch vermittelt.

Voraussetzungen

Erfolgreich bestandene Module «Programmiertechniken in der Computerlinguistik 1», «Programmiertechniken in der Computerlinguistik 2», «Mathematische Grundlagen der Computerlinguistik» oder nachgewiesene Kenntnisse im gleichen Umfang.



Semantic Web & Linked Data

Semantic Web & Linked Data

06SM521-011

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes 2. Frühlingssemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis Portfolio (60% Schriftliche Prüfung und 40% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung

Lernziel

The students: - understand the fundamental ideas of knowledge representation (entities, concepts, instances, ontologies, etc) and are aware of the historical approaches (e.g. Conceptual Graphs); - are very familiar with the Semantic Web standards (RDF, SPARQL, OWL) and are able to apply them for the construction and querying of Knowledge Bases; - understand what Linked Data is, and are able to make use of the concepts and goals of the Linked Open Data movement.

Allgemeine Beschreibung

A formalized computable representation of human knowledge has long been a goal of data science. Several approaches have been proposed since the dawn of computer science and artificial intelligence. In this course we will revise some of the historical approaches (First Order Logic, Semantic Networks, Frames, Conceptual Graphs, etc.) and then introduce the most recent approaches based on the semantic web standards (RDF, SPARQL, OWL). We will then describe how these standards can be used to achieve the goals of a distributed shared knowledge infrastructure (Linked Open Data).

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang.



Sprachtechnologie als Beitrag zur Barrierefreiheit

Language Technology for Accessibility

06SM521-012

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis schriftliche Prüfung

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung

Lernziel

Die Studierenden: - kennen die verschiedenen Zielgruppen von Sprachtechnologiesystemen im Kontext von Barrierefreiheit; - können verschiedene Anwendungen von Sprachtechnologiesystemen im Kontext von Barrierefreiheit beschreiben; - können den sprachtechnologischen Gehalt dieser Systeme genauer erläutern.

Allgemeine Beschreibung

Gehörlose und hörbehinderte, blinde und sehbehinderte Menschen und Menschen mit kognitiven Beeinträchtigungen stehen im Alltag immer wieder vor Hindernissen. Um diese Hindernisse abzubauen, werden laufend neue Werkzeuge entwickelt. Die Vorlesung beschäftigt sich mit Sprachtechnologie-Werkzeugen, die einen Beitrag zur Barrierefreiheit leisten: Wir untersuchen unter anderem die Funktionsweise von Screen-Readern, Systemen zur Untertitelung für Hörbehinderte (Respeaking mit Spracherkennung), Systemen zur Audiodeskription, Gebärdensprach-Avataren, Systemen zur automatischen Übersetzung in Gebärdensprache, Systemen zur automatischen Textvereinfachung (Herstellung von leichter Sprache) und Systemen zur Konvertierung in Braille.

Voraussetzungen

Kenntnisse im Umfang der Lehrveranstaltung "Einführung in die Computerlinguistik 1".



Sprachtechnologische Webapplikationen

Language Technology and Web Applications

06SM521-013

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

9

Angebotsmuster

1-semesterig, jedes Herbstsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (60% schriftliche Prüfung und 40% Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Die Studierenden: - können verschiedene Arten der Daten- und Softwaremodellierung interpretieren, erläutern und selbst anwenden; - sind in der Lage, Datenstrukturen in Datenbankschemata zu übertragen, diese mit Daten zu füllen und einfache Anfragen an die Datenbank zu formulieren, sowie die Resultate mittels Webservice-Schnittstelle in einer Webapplikation darzustellen; - kennen grundlegende Konzepte der Datenvisualisierung und können sie anwenden; - erwerben die Fähigkeit im Rahmen eines Programmierprojektes die anfallenden Arbeitsschritte zu planen, die Aufgaben untereinander aufzuteilen und die Ergebnisse zu integrieren; - sind imstande, ein komplexes System verständlich zu präsentieren und dokumentieren:

Allgemeine Beschreibung

Die Veranstaltung widmet sich Datenstrukturen und Benutzerschnittstellen für computerlinguistische Anwendungen. Einen Schwerpunkt bildet die Datenmodellierung und die Abbildung dieser Modelle in Datenbankschemata, einen anderen die Konzipierung und Implementierung von interaktiven Webanwendungen und die Visualisierung von Daten. Die Teilnehmer erstellen während des Kurses ein Programmierprojekt in Gruppenarbeit und präsentieren dieses am Ende des Semesters.

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang; ausserdem empfohlen: «Programmiertechniken in der Computerlinguistik 2».



Text Analytics in the Digital Humanities

Text Analytics in the Digital Humanities

06SM521-014

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung 1-6, in Halbschritten

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis dokumentierte praktische Arbeit

Unterrichtssprache Englisch

Lehrformen Seminar

Lernziel

Participants will: - know the fundamental questions, problems, challenges and methods in Digital Humanities; - be able to apply the methods to their own scientific questions and datasets; - acquire an overview of existing resources.

Allgemeine Beschreibung

After an introduction to Text Analytics and Digital Humanities, we learn to program key methods ourselves and to use existing tools. The methods comprise concordancing, collocations, machine learning, topic modelling, keyword detection, named entity recognition, network analysis, stylometry, sentiment detection, visualization.

Voraussetzungen

Keine, jedoch Modul «Einführung in die Computerlinguistik 1» empfohlen.



Text Mining

Text Mining

06SM521-015

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (75% schriftliche Prüfung und 25% Nachweis von im Selbststudium erbrachten Studienleistungen). Die genaue notenrelevante Gewichtung der Einzelteile wird Anfang des Semesters durch die Dozierenden bekannt gegeben und kann von den oben genannten Werten abweichen.

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung

Lernziel

Die Studierenden: - kennen die wichtigen Fragestellungen der Erschliessung von strukturierter Information aus natürlichsprachlichen Texten, die damit verbundenen Probleme und die computerlinguistischen Ansätze zu ihrer Lösung; - kennen insbesondere die Teilgebiete der Sentimentanalyse und des automatischen Media-Monitoring und haben eine Übersicht, welche sprachtechnologischen Komponenten und Ressourcen benötigt werden; - kennen verschiedene aktuelle Forschungsansätze für die Teilprobleme und gewinnen praktische Erfahrung mit eigener Programmierarbeit.

Allgemeine Beschreibung

Das Modul führt ins Text-Mining ein. Es vermittelt dabei die relevanten sprachtechnologischen Ansätze für Sentimentanalyse und Media-Monitoring, zwei wichtige Anwendungsbereiche der Computerlinguistik. Sentimentanalyse, d.h. die Erkennung von positiven/negativen Bewertungen in klassischen oder sozialen Medien (insbesondere nutzergenerierte Inhalte wie Kommentare, Blogs, Kurznachrichten), ist zu einem zentralen Bereich der Informationsextraktion geworden. Sie ist eine zentrale Komponente von Media-Monitoring-Systemen, welche sich damit beschäftigen, welche Themen und Entitäten (Personen, Orte, Produkte) prominent sind. Das Modul beginnt mit den Fragestellungen, Methoden und Problemen. Anschliessend werden ausgewählte aktuelle Forschungsansätze vertiefter behandelt. Relevant sind dabei Themen wie lexikalische Ressourcen, Multilingualität, informelle Sprache, Named-Entity-Erkennung und Relationserkennung. Die Teilnehmenden implementieren in Übungen einzelne Aspekte.

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 1» oder nachgewiesene Kenntnisse im gleichen Umfang.



Fundamentals of speech sciences and signal processing

Fundamentals of speech sciences and signal processing

06SM521-519

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Frühjahrssemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio: (a) 6 written assignments throughout the term (60%) (b) written exam (40%)

Unterrichtssprache

Englisch

Lehrformen

Vorlesung

Lernziel

(a) Fundamental skills in speech signal processing (b) Understanding of speech acoustics like signal types, signal transformations, acoustic systems and signal and system analysis (c) Application of the signal processing techniques in research and industrial products.

Allgemeine Beschreibung

In research and in numerous practical applications, the processing of speech signals (e.g. recoding, manipulating, replaying) is fundamental. In this lecture series we will acquire a fundamental understanding of signal and system theory that is necessary for understanding speech communication in humans and the processing of speech in numerous technical applications.

Voraussetzungen

-

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Instrumental techniques of phonetic research

Instrumental techniques of phonetic research

06SM521-520

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Frühlingsemester

Bewertung/Benotung

bestanden/nicht bestanden

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio: At the beginning of the semester students create a pool of their own recordings. During the semester they run guided analyses on their material both as part of the course but also as personal homework. In addition students are required to hand in a small-scale empiric study (7-10 pages) using their recordings and/or additional recordings from other sources to be handed in at the last meeting of the semester. Both their analyses during the semester and their final report form their portfolio and thus the basis for the evaluation of their performance.

Unterrichtssprache

Englisch

Lehrformen

Übung

Lernziel

Students know how to make high-quality audio recordings for phonetic research purposes. They can annotate sound files, make reliable measurements in them (formants, pitch, intensity, etc.) and produce meaningful visualizations (wave forms, spectra, spectrograms, etc.) with suitable software. They also understand how to read spectrograms so as to draw informed conclusions about the temporal and spectral characteristics of speech events. Moreover, students understand the most important key notions and concepts in automatizing measurements and in making them replicable (scripting).

Allgemeine Beschreibung

Since speech is a transient event, phoneticians regularly resort to the aid of technical devices in order to record, describe and analyse the production, the acoustics and the perception of speech sounds. Hence, in this module we look at the technical side of phonetic research and the students acquire and develop skills and techniques necessary for the successful deployment of such devices, ranging from sound recording equipment (especially recorders and microphones) to more specialized phonetic equipment (such as the laryngograph) to software solutions geared specifically towards the need of phoneticians (such as Praat or the R-package 'vowels').

Voraussetzungen

-

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Phonetic Transcription

06SM521-521

| | |
|--|--|
| Anbietende Organisationseinheit | PhF: Institut für Computerlinguistik |
| ECTS Credits | 6 |
| Angebotsmuster | 1-semesterig, jedes Herbstsemester |
| Bewertung/Benotung | bestanden/nicht bestanden |
| Repetierbarkeit | einmal wiederholbar, erneut buchen |
| Leistungsnachweis | Portfolio (Students upload solutions of different exercises of phonetic transcription to the platform OLAT). |
| Unterrichtssprache | Englisch |
| Lehrformen | Übung |

Lernziel

Students are familiar with the different alphabets used for phonetic transcription (ranging from Boehmer-Ascoli to X-SAMPA), gaining a thorough knowledge of the International Phonetic Alphabet (IPA). They are aware of the theoretical and methodological aspects involved in phonetic transcription and are able to use the necessary practical tools (e.g., fonts, computer software, etc.) in an adequate manner.

Allgemeine Beschreibung

This tutorial offers a historical overview of phonetic transcription from the foundations in the nineteenth century (i.e. the creation of the dialectological transcription systems and the introduction of the IPA in 1886) to the novel tools used in the field of speech technology (e.g. the X-SAMPA-Alphabet). The practical exercises focus on different types of transcription (broad/narrow, phonemic/phonetic, systematic/impressionistic), on the basis of both written texts and audio recordings and illustrating better and lesser known languages. The technical tools consist in particular of phonetic fonts (including shortcuts and the "IPA-Palette") and the use of acoustic software for manual and automatic segmentation/annotation of audio recordings. An outlook on some non-phonetic procedures of annotating spoken language (e.g. in the field of conversation analysis or working with video recordings) will round off the tutorial.

Voraussetzungen

The participation in "The Sounds of the World's Languages" is highly recommended.

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Intermediate Methods and Programming in Digital Linguistics

Intermediate Methods and Programming in Digital Linguistics

06SM521-529

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Frühlingsemester

Bewertung/Benotung

1-6, in Halbschritten

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

Portfolio (80% written exam and 20% proof of self-study achievements)

Unterrichtssprache

Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Students will be able to use Unix-based systems and Unix language processing tools efficiently. Students will know the basic data types, control structures and functions of Python. Students will be able to design problem solutions and to implement them in Python. Students will know how to use basic tools and programming libraries for corpus linguistics.

Allgemeine Beschreibung

This course is designed to refresh and to deepen programming skills in Unix and Python. We teach basic operators and functions, the handling of lists and dictionaries as well as the basics of object-oriented programming. It is particularly important that the students acquire the ability to prepare text and speech data for further processing. Through practical tasks and exercises, we train the algorithmic and programming skills of the participants.

Voraussetzungen

Successful completion of the introductory course "Language Data Processing" of the Master Linguistics



Praktikum intern

Practical Training In-House

06SM521-017

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Semester

Bewertung/Benotung bestanden/nicht bestanden

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis dokumentierte praktische Arbeit

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Praktikum

Lernziel

Die Studierenden: - erhalten Einblick in die wissenschaftliche Projektarbeit; - rezipieren Fachliteratur; - sind beteiligt an Evaluations- und Revisionszyklen; - übernehmen Teilarbeiten im Projekt - sind an Publikationsvorhaben beteiligt - gewinnen insgesamt Erfahrung in praktischer Arbeit - und vertiefen ihr Wissen in einem Thema;

Allgemeine Beschreibung

Das Modul ist gedacht als praktisches Kennenlernen von wissenschaftlicher Projektarbeit. Die Studierenden werden an wissenschaftliches Arbeiten herangeführt. Sie recherchieren, lesen Fachliteratur, bereiten Daten vor, annotieren, wenden statistische Methoden und Verfahren des Maschinellen Lernens an, sind beteiligt an Publikationsarbeiten und diskutieren Lösungsansätze. Die Studierenden arbeiten selbständig an einer Teilaufgabe im Kontext eines wissenschaftlichen Themas oder Projektes. Dieses Modul kann gebucht werden, um Mitarbeit bei wissenschaftlichen Projekten der UZH anrechnen zu lassen. Dieses Modul kann mit einer Punktzahl von 6 oder 9 gebucht werden. Die Punktzahl wird in Absprache mit dem Modulverantwortlichen festgelegt.

Voraussetzungen

Dieses Modul kann nicht selbst gebucht werden, die Buchung des Moduls muss vom Modulverantwortlichen autorisiert werden. Es besteht kein Anspruch auf ein internes Praktikum, sondern das Modul wird nur angeboten, wenn in einem Projekt eine entsprechende Möglichkeit besteht. Die Voraussetzungen werden dem Thema entsprechend festgelegt. In der Regel werden Programmierkenntnisse erwartet und/oder der erfolgreiche Besuch bestimmter Lehrveranstaltungen.



Programmierprojekt

Programming Project

06SM521-018

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

6

Angebotsmuster

1-semesterig, jedes Semester

Bewertung/Benotung

bestanden/nicht bestanden

Repetierbarkeit

einmal wiederholbar, erneut buchen

Leistungsnachweis

dokumentierte praktische Arbeit

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Selbststudium

Lernziel

Die Studierenden: - planen ein Projekt; - setzen einen Projektplan um; - realisieren Software; - halten Dokumentationsstandards ein; - evaluieren; - nutzen Softwarerepositories.

Allgemeine Beschreibung

Programmierprojekte dienen der Festigung der Programmierkenntnisse und dem Erwerb von Fertigkeiten, wie sie im Kontext des Softwareengineering erforderlich sind. Ausgehend von einer computerlinguistischen Fragestellung und ggfs. relevanter Fachliteratur wird ein Lösungskonzept erarbeitet, es werden die Meilensteine fixiert, es werden Daten beschafft und/oder eigens annotiert, es wird ein Programm implementiert und schliesslich anhand der Daten evaluiert. Dieses Modul kann gebucht werden, um ein Programmierprojekt anrechnen zu lassen. Dieses Modul kann mit einer Punktzahl von 3, 6 oder 9 gebucht werden. Die Punktzahl wird in Absprache mit dem Modulverantwortlichen festgelegt.

Voraussetzungen

Dieses Modul kann nicht selbst gebucht werden, die Buchung des Moduls muss vom Modulverantwortlichen autorisiert werden. Bevor ein Programmierprojekt erarbeitet wird, muss zwingend Rücksprache mit dem Modulverantwortlichen genommen werden (per E-Mail). Die Voraussetzungen werden dem Thema entsprechend festgelegt.