



Modulkatalog

Digitale Linguistik

Programmformat: Minor 30

Studienstufe: Master

Gültig ab: Herbstsemester 2019

[Erstellt am 22.01.2024]

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.



06M-7526i01 Wissenschaftliche Vertiefung

06SM521-s04 [Seminar] (W 6)

5



06M-7526i02 Kernbereich Digitale Linguistik

06SM271-518	Language Data Processing (P 15)	6
06SM271-516	Quantitative Methods (WP 6)	7
06SM271-517	Language Data Acquisition (WP 9)	8
06SM521-002	Einführung in die Computerlinguistik 2 (WP 3)	9
06SM521-003	Mathematische Grundlagen der Computerlinguistik (WP 6)	10
06SM521-005	Programmiertechniken der Computerlinguistik 2 (WP 6)	11
06SM521-006	Erstellung und Annotation von linguistischen Ressourcen (WP 6)	12
06SM521-007	Grammatikformalismen und Parsing (WP 6)	13
06SM521-008	Maschinelle Übersetzung (WP 6)	14
06SM521-009	Maschinelles Lernen in der Computerlinguistik (WP 6)	15
06SM521-012	Digital Accessibility (WP 6)	16
06SM521-013	Sprachtechnologische Webapplikationen (WP 9)	17
06SM521-014	Text Analytics in the Digital Humanities (WP 6)	18
06SM521-015	Text Mining (WP 6)	19
06SM521-023	Introduction to Speech Processing (WP 6)	20
06SM521-519	Fundamentals of speech sciences and signal processing (WP 6)	21
06SM521-520	Instrumental techniques of phonetic research (WP 6)	22
06SM521-521	Phonetic Transcription (WP 6)	23
06SM521-529	Intermediate Methods and Programming in Digital Linguistics (WP 6)	24
06SM521-530	Eye tracking: Experiment design and machine learning methods (WP 6)	25



06M-7526i03 Praxis der Digitalen Linguistik

06SM521-017	Praktikum intern (WP 6)	26
06SM521-512	Programming Project 1 (WP 6)	27
06SM521-513	Student Teaching Assistant 1 (WP 6)	28
06SM521-517	Programming Project 2 (WP 6)	29



[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 (1) vertiefen ihr Wissen in einem spezifischen Gebiet (2) erlernen die methodischen Grundkompetenzen wissenschaftlicher Arbeit (3) üben das verständliche Präsentieren komplexer Sachverhalte (4) 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, erneut buchen

Leistungsnachweis

Portfolio (40% written exam for Introduction to Language Data Processing, 40% written exam for Programming for Linguists, 20% proof of self-study achievements in both courses). All elements of this portfolio must be completed. If an element is not completed, the module is considered as «failed».

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, erneut buchen
Leistungsnachweis	Portfolio (80% written exam and 20% written exercises). All elements of this portfolio must be completed. If an element is not completed, the module is considered as «failed».
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, erneut buchen

Leistungsnachweis Portfolio (80% written exam and 20% proof of self-study achievements). All elements of this portfolio must be completed. If an element is not completed, the module is considered as «failed».

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

06SM521-002

Anbietende Organisationseinheit

PhF: Institut für Computerlinguistik

ECTS Credits

3

Angebotsmuster

1-semesterig, jedes Frühlingsemester

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

Students (1) know the application and problem areas of semantics within computational linguistics: lexical semantics, especially word nets, frame semantics, and sentence semantics; they (2) learn logics (as the basis of sentence semantics); they (3) gain insight into computational linguistic approaches for computing semantic structures and meaning desambiguation; they (5) practise the modeling of semantic relations with word nets.

Experimental and Computational Psycholinguistics: Students are familiar with different theories of sentence processing and know how these can be tested empirically. Students have gained insights into how psycholinguistic theories can be implemented as computational cognitive models.

Allgemeine Beschreibung

The lecture deals with the semantics of natural language, i.e. lexical semantics and sentence semantics. We will start with the problem of ambiguity, discuss the basics of lexical semantics, and then turn to the so-called sense relations (lexical functions) used in word nets to structure vocabulary. We will cover propositional and predicate logic and examine how to semantically represent sentences using predicate logic. This course also gives an introduction into Experimental and Computational Psycholinguistics. We will discuss current theories of human language processing and how these theories can be evaluated experimentally and by means of computational simulations.

Voraussetzungen

keine



Mathematische Grundlagen der Computerlinguistik

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

The students know the mathematical basics for quantitative language processing; this includes probability theory, combinatorics, hypothesis testing and important concepts of analysis and linear algebra. In addition, they know the basic procedures of machine learning and can apply them to selected problems in language processing. The basis for this is a program library such as sklearn.

Allgemeine Beschreibung

This course covers the mathematical foundations of computational linguistics. The central concepts of probability theory and combinatorics are introduced. Furthermore, we deal with problems of statistics (descriptive statistics, inferential statistics) and linear algebra. The course not only provides the theoretical foundations but also looks at the application of numerical methods. Additionally, an introduction to traditional machine learning methods such as Logistic Regression or the Perceptron and their mathematical background is provided.

Voraussetzungen

keine



Programmiertechniken der Computerlinguistik 2

06SM521-005

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 (25% exercises, 25% mid-term exam, 50% final exam)

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.

Allgemeine Beschreibung

Ziel dieses Moduls ist die Vertiefung von Programmierkenntnissen in Python und das Kennenlernen der Bausteine von moderner Softwareentwicklung (Versionskontrolle, Projektverwaltung, Testen). Relevante Themen und Techniken werden anhand von ausgewählten Anwendungen aus dem Bereich der Computerlinguistik und Sprachverarbeitung vermittelt.

Voraussetzungen

Modul «Programmiertechniken in der Computerlinguistik 1» (PCL1) oder nachgewiesene Kenntnisse im gleichen Umfang; Kenntnisse im Umfang von «Einführung in die Computerlinguistik 1» (ECL1). Interessierte Studierende, die PCL1 oder ECL1 nicht absolviert/bestanden haben, müssen sich vor der Buchung bei der Studienberatung melden.



Erstellung und Annotation von linguistischen Ressourcen

06SM521-006

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

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung

Lernziel

Die Studierenden (1) kennen und beherrschen alle Teilschritte, welche für die nachhaltige Erstellung von sprachtechnologischen Ressourcen notwendig sind (2) kennen gängige Verfahren, um manuelle Annotation effizient und in guter Qualität zu erzeugen und für maschinelles Lernen zu nutzen (3) produzieren gemeinsam eine annotierte sprachtechnologische Ressource, welche wissenschaftliche Standards erfüllt.

Allgemeine Beschreibung

Manuell oder automatisch annotierte Ressourcen (z.B. Texte, Aufnahmen gesprochener Sprache) sind mittlerweile die Grundlage für viele Forschungsfragen und Anwendungssysteme. Dieses Modul vermittelt die Grundlagen, Probleme und Best-Practice-Verfahren der Erstellung von Korpora und (ihrer) Annotation anhand von Vorträgen und eines praktischen Projekts zur Ressourcenerstellung, welches alle Teilschritte umfasst. Bezüglich Ressourcen werden Themen angesprochen wie Rohdatenbeschaffung (Crawling, Experimentelle Verfahren), automatische Datenaufbereitung, Verwendung von nachhaltigen XML-Standards als Austauschformat, Lizenzierung. Das Thema Annotation umfasst neben den automatischen computerlinguistischen Verfahren die Erstellung von Richtlinien und Dokumentationen, manuelle Annotation (Crowdsourcing) und Inter-Annotator-Übereinstimmung sowie effiziente halbautomatische Verfahren.

Voraussetzungen

Programmierkenntnisse



Grammatikformalismen und 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 (1) kennen die Fragestellungen, Probleme und Lösungsansätze der Syntaxanalyse (2) verstehen ausgewählte Grammatikformalismen und können sie anwenden, sowie ausgewählte Grammatikphänomene des Deutschen modellieren (3) 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

06SM521-008

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 (25% Übungen, 75% schriftliche Prüfung)

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung, Tutorat

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: (1) Einführung in die maschinelle Übersetzung (2) Parallele Korpora und Alignierung (3) Aktuelle Methoden der Maschinellen Übersetzung (4) Hybride maschinelle Übersetzungssysteme (5) Gesellschaftliche Auswirkungen der Maschinellen Übersetzung

Voraussetzungen

Erfolgreich bestandene Module «Mathematische Grundlagen der Computerlinguistik» und «Maschinelles Lernen in der Computerlinguistik» oder nachgewiesene Kenntnisse im gleichen Umfang; 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

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 (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 verstehen die mathematischen Details der grundlegenden Ansätze im Bereich Deep Learning. Sie kennen die gängigsten Netzwerkarchitekturen und sind in der Lage, eigenständig Modelle zur Bearbeitung typischer sprachtechnologischer Probleme in einer auf Python basierenden Programmierumgebung zu realisieren.

Allgemeine Beschreibung

Der Kurs ist eine Einführung in das sogenannte Deep Learning: das Lernen von neuronalen Netzen. Es werden die mathematischen Grundlagen behandelt und verschiedene Ansätze wie RNN, LSTM etc. vorgestellt. Neben der theoretischen gibt es auch eine praktische Herangehensweise. Wir werden eine aktuelle Programmbibliothek (Framework) wie z.B. Pytorch verwenden, um prototypische Aufgabenstellungen bei der Verarbeitung menschlicher Sprache zu untersuchen.

Voraussetzungen

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



Digital Accessibility

06SM521-012

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 Written exam

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung

Lernziel

Students (1) are aware of different target groups in the context of accessibility; (2) are aware of barriers that these target groups face in accessing information and communication as well as information and communication technologies; (2) know about tools and measures for reducing some of these barriers; (3) know how to operate a selection of these tools.

Allgemeine Beschreibung

Blind and visually impaired, deaf and hearing-impaired, cognitively, motor-impaired, and persons with speech and language disorders face many barriers in their everyday lives, often related to access to information and communication. This course provides an overview of common barriers and introduces language-based assistive technologies and e-accessibility measures. Specifically, the course deals with guidelines for creating accessible electronic documents (e.g., Web pages), with screen readers, Braille, operation of mobile devices, sign language, tactile signing, Augmentative and Alternative Communication (AAC), and ethics. Students will gain hands-on practice using various tools as part of the exercises accompanying the course. This course is complemented by a course on Master's level with a focus on deep learning approaches.

Voraussetzungen

Knowledge to the extent of the course «Einführung in die Computerlinguistik 1»



Sprachtechnologische Webapplikationen

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 (1/3 schriftliche Prüfung [Take-Home-Exam], 2/3 Nachweis von im Selbststudium erbrachten Studienleistungen)

Unterrichtssprache

Deutsch und/oder Englisch

Lehrformen

Vorlesung mit integrierter Übung, Tutorat

Lernziel

Die Studierenden (1) können verschiedene Arten der Daten- und Softwaremodellierung interpretieren, erläutern und selbst anwenden (2) 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 (3) kennen grundlegende Konzepte der Datenvisualisierung und können sie anwenden (4) erwerben die Fähigkeit im Rahmen eines Programmierprojektes die anfallenden Arbeitsschritte zu planen, die Aufgaben untereinander aufzuteilen und die Ergebnisse zu integrieren (5) 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

06SM521-014

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 dokumentierte praktische Arbeit

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Seminar

Lernziel

Participants will (1) know the fundamental questions, problems, challenges and methods in Digital Humanities; (2) be able to apply the methods to their own scientific questions and datasets; (3) 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, visualisation and transformers (LLMs). They will be demonstrated by case studies and hands-on experiments. The aim of the seminar is to enable students to conduct and interpret experiments themselves. Central concepts of statistics, namely overuse, significance, classification, regression, and distributional semantics such as topic models and BERT models are shown by example, using real-world corpora and giving tangible results and quantitative support. Simple data pre-processing steps are demonstrated. The statistical programming language R is used to create wordlists, conduct statistical tests and for simple visualisations and BERT zero-shot. For tasks like topic modelling, document classification and conceptual maps, recipes for using the tools `mallet`, `LightSide`, `textplot`, `gephi`, and others are given.

Voraussetzungen

1) Programmierkenntnisse sind hilfreich, aber nicht Voraussetzung. Der Kurs richtet sich explizit auch an Studierende ohne Programmierkenntnisse. 2) Grundkenntnisse in Linguistik, Literatur oder Geschichte. 3) Grundkenntnisse in Statistik



Text Mining

06SM521-015

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 (25% Übungen, 75% schriftliche Prüfung)

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung, Tutorat

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. Sie kennen insbesondere die Teilgebiete der Sentimentanalyse und des automatischen Media-Monitoring und haben eine Übersicht, welche sprachtechnologischen Komponenten und Ressourcen benötigt werden. Sie kennen sich mit Linked Data, ihrer Verwendung und Rolle im Text-Mining aus. Sie kennen verschiedene aktuelle Forschungsansätze für die Teilprobleme und gewinnen praktische Erfahrung mit eigener Programmierarbeit.

Allgemeine Beschreibung

Das Modul führt ins Text-Mining ein, das heisst mit Verfahren, um aus unstrukturierten Textdaten strukturierte Information zu extrahieren. Es vermittelt dabei die relevanten sprachtechnologischen Ansätze für Sentimentanalyse und Media-Monitoring, zwei wichtige Anwendungsbereiche der Computerlinguistik. Sentimentanalyse, d.h. die Erkennung von positiven oder negativen Bewertungen in sozialen Medien (insbesondere nutzergenerierte Inhalte wie Kommentare, Blogs, Kurznachrichten), ist zu einem wichtigen Bereich der Informationsextraktion geworden. Sie ist eine Basis-Komponente von Media-Monitoring-Systemen, die sich damit beschäftigen, welche Themen und Entitäten (Personen, Orte, Produkte) prominent sind. Weiter wird im Modul die Rolle von Linked Data als strukturierte Wissensrepräsentation eingeführt sowie ihre Verwendung in Text-Mining-Ansätzen betrachtet. Das Modul beginnt mit den Fragestellungen, Methoden und Problemen. Anschliessend werden ausgewählte aktuelle Forschungsansätze vertiefter behandelt. Relevant sind dabei Themen wie Sprachmodelle, Multilingualität, informelle Sprache, Named-Entity-

Voraussetzungen

Erfolgreich bestandenes Modul «Programmiertechniken in der Computerlinguistik 2» oder nachgewiesene Kenntnisse im gleichen Umfang. Grundlegende Kenntnisse in Maschinellem Lernen.



Introduction to Speech Processing

06SM521-023

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

Unterrichtssprache Deutsch und/oder Englisch

Lehrformen Vorlesung mit integrierter Übung

Lernziel

Students will gain practical experience and a deeper understanding of techniques in the field of automatic speech processing.

Allgemeine Beschreibung

The module consists of lectures and practical exercises in the field of automatic speech processing. Students will learn the fundamentals of speech processing systems such as speech recognition and speech generation (e.g., text-to-speech synthesis), along with their technical implementations, and apply them in practical exercises.

Voraussetzungen

Required programming knowledge equal or equivalent to that taught in "Programmiertechniken in der Computerlinguistik 1".

Recommended background knowledge to the following courses: "Maschinelles Lernen in der Computerlinguistik" and "Einführung in die Computerlinguistik 1".



Fundamentals of speech sciences and signal processing

06SM521-519

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: (a) weekly assignments, 40% (b) end of term exam, 60%

Unterrichtssprache Englisch

Lehrformen Vorlesung

Lernziel

(1) Fundamental skills in speech signal processing (2) Understanding of speech acoustics like signal types, signal transformations, acoustic systems, signal and system theory (3) Application of the signal processing techniques in research and industrial products.

Allgemeine Beschreibung

Experience the captivating world of speech signal processing. Discover the essential techniques that enable us to decode, manipulate, and reproduce the human communication with speech. Learn about signal and system theory necessary for speech processing in both human interaction and cutting-edge technological applications. This lecture series will equip you with the fundamental knowledge needed to unravel the intricacies of speech communication and embrace the possibilities it holds.

Voraussetzungen

An interest in speech signal processing is with computers required.

Dieses Modul ist als vorgezogenes Mastermodul geeignet.



Instrumental techniques of phonetic research

06SM521-520

Anbietende Organisationseinheit PhF: Institut für Computerlinguistik

ECTS Credits 6

Angebotsmuster 1-semesterig, jedes Frühlingssemester

Bewertung/Benotung bestanden/nicht bestanden

Repetierbarkeit einmal wiederholbar, erneut buchen

Leistungsnachweis During the semester students run guided analyses on spoken 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) to be handed in a fortnight after 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

Students are required to have attended an introductory module in phonetics at bachelor or master level.

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

06SM521-529

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

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



Eye tracking: Experiment design and machine learning methods

06SM521-530

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

Unterrichtssprache Englisch

Lehrformen Übung

Lernziel

Students will learn how to develop an experiment design for a given research question. Students will further learn how to implement and conduct an eye tracking experiment, choose and implement adequate preprocessing algorithms and implement state-of-the-art statistical and machine learning methods for the analysis of eye tracking data.

Allgemeine Beschreibung

In this course, we will study eye tracking methodology for (psycho-)linguistic basic research as well as technological applications including cognitively enhanced/interpretable NLP. The course covers an introduction to eye tracking hardware, the design and implementation of experiments and the preprocessing of the data. The main focus will be the analysis and implementation of state-of-the-art machine learning methods for the analysis of eye tracking data.

Voraussetzungen

Python programming skills at least on the level of the Module «Programmiertechniken in der Computerlinguistik 1».



Praktikum intern

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 (1) erhalten Einblick in die wissenschaftliche Projektarbeit (2) rezipieren Fachliteratur (3) sind beteiligt an Evaluations- und Revisionszyklen (4) übernehmen Teilarbeiten im Projekt (5) sind an Publikationsvorhaben beteiligt (6) gewinnen insgesamt Erfahrung in praktischer Arbeit (7) 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.



Programming Project 1

06SM521-512

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 Sonstiges

Lernziel

The students (1) autonomously design a project (2) realize the project plan (3) use existing tools (4) do software engineering (5) document their work according to standards (6) evaluate the results (7) use software repositories

Allgemeine Beschreibung

Programming projects aim at the consolidation of programming skills and the acquisition of software engineering skills. Starting with a particular research question and relevant literature, they work on a solution, define milestones, acquire and/or annotate data, implement a program and evaluate it using appropriate data. This module can be booked to credit work done in a first programming project. This module can be booked with 3, 6 or 9 ECTS points. The amount of points will be decided in consultation with the module coordinator.

Voraussetzungen

In the duration of a study level a maximum of two programming projects can be booked. This module can be booked to credit work done in a programming project. It cannot be booked by the students themselves, the booking has to be authorized by the module coordinator. Before a programming project is started, it is essential to get the permission of the module coordinator (per Email). The prerequisites will be set according to the topic.



Student Teaching Assistant 1

06SM521-513

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 Sonstiges

Lernziel

The students (1) cope with computational linguistics content from a teaching perspective (2) learn to prepare computational linguistics content in a way tailored to a student's audience (3) learn to correct exercises and give appropriate feedback

Allgemeine Beschreibung

A student teaching assistance serves the acquisition of basic teaching skills. This requires a deeper insight of the contents of the associated lecture and the ability to prepare teaching material in order to help the students to better understand it. The task also involves the preparation and correction of exercises. This module can be booked to credit the conducting of exercises/tutorials. This module can be booked with 3 or 6 ECTS points. The amount of points will be decided in consultation with the module coordinator.

Voraussetzungen

In the duration of a study level a maximum of two modules «Student Teaching Assistant» can be booked, whereby the two modules must differ in content (also to any previously completed student teaching assistant modules). This module is booked in order to receive credit for a first job as a student teaching assistant at master's level.

This module is an application module, the application has to be authorized by the module coordinator (per Email). The lecturers have to be included in the communication. The open positions for student teaching assistants are usually posted on the mailing list of the Institute of Computational Linguistics (clist@lists.ifi.uzh.ch) a few weeks before the semester starts. Students interested in conducting exercises/tutorials of a specific course can apply anytime for the position directly with the lecturer and the module coordinator. The module in question must have been passed successfully beforehand.



Programming Project 2

06SM521-517

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 Sonstiges

Lernziel

The students (1) autonomously design a project (2) realize the project plan (3) use existing tools (4) do software engineering (5) document their work according to standards (6) evaluate the results (7) use software repositories

Allgemeine Beschreibung

Programming projects aim at the consolidation of programming skills and the acquisition of software engineering skills. Starting with a particular research question and relevant literature, they work on a solution, define milestones, acquire and/or annotate data, implement a program and evaluate it using appropriate data. This module can be booked with 3, 6 or 9 ECTS points. The amount of points will be decided in consultation with the module coordinator.

Voraussetzungen

In the duration of a study level a maximum of two programming projects can be booked. This module can be booked to credit work done in a second programming project. It cannot be booked by the students themselves, the booking has to be authorized by the module coordinator. Before a programming project is started, it is essential to get the permission of the module coordinator (per Email). The prerequisites will be set according to the topic.
