



## **Modulkatalog**

---

### **Computerlinguistik und Sprachtechnologie**

---

Programmformat: Minor 60

---

Studienstufe: Bachelor

---

Erstellt am 04.06.2024

---

### **Modulgruppen des Programms**

---

Einführung in die Computerlinguistik

Wissenschaftliche Vertiefung

Kernbereich Computerlinguistik und Sprachtechnologie

Informatik

Praxis der Computerlinguistik und Sprachtechnologie

---

---

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.



**06B-7521e01 Einführung in die Computerlinguistik**

---

06SM521-001	Einführung in die Computerlinguistik 1 (P 6)	7
06SM521-002	Einführung in die Computerlinguistik 2 (P 3)	8
06SM521-003	Mathematische Grundlagen der Computerlinguistik (P 6)	9
06SM521-004	Programmiertechniken der Computerlinguistik 1 (P 9)	10
06SM521-005	Programmiertechniken der Computerlinguistik 2 (P 6)	11

---



**06B-7521i01 Wissenschaftliche Vertiefung**

---

06SM521-s04 [Seminar] (W 6)

12

---



**06B-7521i02 Kernbereich Computerlinguistik und Sprachtechnologie**

---

06SM521-009	Maschinelles Lernen in der Computerlinguistik (P 6)	13
06SM521-006	Erstellung und Annotation von linguistischen Ressourcen (WP 6)	14
06SM521-008	Maschinelle Übersetzung (WP 6)	15
06SM521-010	Phonetic Data Analysis (WP 6)	16
06SM521-012	Digital Accessibility (WP 6)	17
06SM521-013	Sprachtechnologische Webapplikationen (WP 9)	18
06SM521-014	Text Analytics in the Digital Humanities (WP 6)	19
06SM521-015	Text Mining (WP 6)	20
06SM521-023	Introduction to Speech Processing (WP 6)	21
06SM521-s10	[Sommerschule] (W 6)	22

---



**06B-7521i03 Informatik**

---

Diese Modulgruppe enthält ausschliesslich Wahlmodule. Informieren Sie sich im Vorlesungsverzeichnis über das aktuelle Angebot.

---



**06B-7521i04 Praxis der Computerlinguistik und Sprachtechnologie**

---

06SM521-016	Praktikum extern (WP 6)	23
06SM521-017	Praktikum intern (WP 6)	24
06SM521-018	Programmierprojekt 1 (WP 6)	25
06SM521-019	Übungsleitung/Tutorat 1 (WP 6)	26
06SM521-021	Übungsleitung/Tutorat 2 (WP 6)	27
06SM521-022	Programmierprojekt 2 (WP 6)	28

---



## Einführung in die Computerlinguistik 1

---

06SM521-001

---

**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 Studienleistung)

---

**Unterrichtssprache** Englisch

---

**Lehrformen** Vorlesung mit integrierter Übung, Tutorat

---

### Lernziel

Die Studierenden (1) kennen die grundlegende Thematik von Computerlinguistik, Sprachtechnologie und Linguistik sowie ihre interdisziplinäre Verknüpfung (2) verstehen die wichtigen linguistischen und formalen Begriffe und können sie abgrenzen und anwenden (3) ermitteln detaillierte Syntaxanalysen für deutsche oder englische Sätze, verstehen Grammatikregeln und entwickeln diese weiter (4) kennen Zweck, Probleme und Lösungsansätze der textuellen Vorverarbeitung, automatischen Wortformenanalyse, Wortartenbestimmung, Syntaxanalyse, Eigennamenerkennung, Informationsextraktion, Semantikanalyse, Verarbeitung gesprochener Sprache, Maschineller Übersetzung (5) erkennen die für sprachtechnologische Anwendungen benötigten computerlinguistischen Techniken (6) können quantitative Evaluierungen interpretieren und anwenden.

---

### Allgemeine Beschreibung

Dieser Kurs vermittelt die praktischen und theoretischen Grundlagen der Computerlinguistik. Vorgestellt werden beispielhafte Systeme und Anwendungen wie maschinelle Übersetzung, Sprachsynthese sowie Sprach- und Sprechererkennung, Semantikanalyse, Informationsextraktion, sowie Methoden und Probleme der automatischen Syntaxanalyse von Sprachen. Im Kurs wird das notwendige linguistische Wissen erarbeitet.

---

### Voraussetzungen

-

---



## Einführung in die Computerlinguistik 2

---

06SM521-002

---

**Anbietende Organisationseinheit**

PhF: Institut für Computerlinguistik

---

**ECTS Credits**

3

---

**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, 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 1

---

06SM521-004

---

**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 (75% Schriftliche Prüfung und 25% Nachweis von im Selbststudium erbrachten Studienleistungen)

---

**Unterrichtssprache** Englisch

---

**Lehrformen** Vorlesung mit integrierter Übung, Tutorat

---

### Lernziel

Die Studierenden können mit Unix-basierten Systemen umgehen und sind fähig, deren Werkzeuge zur Textverarbeitung effizient einzusetzen; sie kennen die grundlegenden Datentypen, Kontrollstrukturen und Funktionen von Python; sie konzipieren und implementieren einfache Problemlösungen in Python und wenden Textanalyse-Werkzeuge für korpuslinguistische Fragestellungen sowie einfache Verfahren zur Sprachsignalverarbeitung an.

---

### Allgemeine Beschreibung

Das Modul beginnt mit einer Einführung in Unix und seine Kommandozeilenschnittstelle, speziell in dessen Werkzeuge zur Textverarbeitung. Danach folgt eine Einführung in die Programmiersprache Python. Vermittelt werden Kontrollstrukturen, Operatoren und Funktionen, der Umgang mit Listen und Dictionaries sowie ein grundlegendes Verständnis für objektorientierte Programmierung. Die Studierenden sollen die Fähigkeit erwerben, Texte für die weitere Verarbeitung aufzubereiten und erste Erfahrung mit der Verarbeitung gesprochener Sprache sammeln. In regelmässigen Übungen soll das im Modul vermittelte Wissen zur praktischen Programmierfähigkeit der Teilnehmenden werden.

---

### Voraussetzungen

Paralleler Besuch der Vorlesung «Einführung in die Computerlinguistik 1» empfohlen.

---



## 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.

---



## [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

-



## 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% final exam und 25% exercises)

---

**Unterrichtssprache**

Deutsch und/oder Englisch

---

**Lehrformen**

Vorlesung mit integrierter Übung, Tutorat

---

### Lernziel

Students understand the mathematical details of the basic approaches in the field of Deep Learning. They know the most common network architectures and are able to independently implement models for processing typical language technology problems in a Python-based programming environment.

---

### Allgemeine Beschreibung

The course is an introduction to the field of Deep Learning: the learning of neural networks. The mathematical basics are discussed and different approaches like RNN, LSTM etc. are presented. Besides the theoretical part, there is also a practical part: We will use a current program library (framework) such as Pytorch to investigate prototypical tasks in human language processing.

---

### Voraussetzungen

Mathematical Foundations of Computational Linguistics

---



## Erstellung und Annotation von linguistischen Ressourcen

---

06SM521-006

---

**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

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

---



## Maschinelle Übersetzung

---

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 (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.

---



## Phonetic Data Analysis

---

06SM521-010

---

**Anbietende Organisationseinheit** PhF: Institut für Computerlinguistik

---

**ECTS Credits** 6

---

**Angebotsmuster** 2-semesterig, Beginn im Herbstsemester

---

**Bewertung/Benotung** 1-6, in Halbschritten

---

**Repetierbarkeit** einmal wiederholbar, erneut buchen

---

**Leistungsnachweis** Portfolio (Referat 1/3, schriftliche Arbeit 2/3)

---

**Unterrichtssprache** Deutsch und/oder Englisch

---

**Lehrformen** Seminar

---

### Lernziel

Students know how to make high-quality audio recordings for phonetic research purposes and how to organize them efficiently for analysis. They can also annotate them manually and use automatic tools for segmentation. They perform reliable measurements (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

Research on language increasingly relies on the analysis of large databases and a large number of such corpora already exist, in particular for English. In spoken language, data analysis is more complicated compared to written language as the speech signal is a highly variable continuous stream of events which first needs to be processed to arrive at the underlying linguistic units. For this reason, phoneticians regularly resort to the aid of technical devices in order to record, file, edit and analyze large speech corpora. In this module we look at the technical side of phonetic database compilation, processing and analysis. The students acquire and develop skills and techniques necessary to carry out recordings of speech, organize and edit them efficiently and learn how to analyze them. We will look at sound recording equipment and software solutions geared specifically towards the need of phonetic processing and analysis.

---

### Voraussetzungen

Students are required to have passed an introductory module in linguistics at bachelor level.

---





## Digital 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** 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, 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 Frühjahrssemester

**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 Frühjahrssemester

---

**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".

---



## **[Sommerschule]**

06SM521-s10

<b>Anbietende Organisationseinheit</b>	PhF: Institut für Computerlinguistik
<b>ECTS Credits</b>	6
<b>Angebotsmuster</b>	1-semesterig, einmalig
<b>Bewertung/Benotung</b>	bestanden/nicht bestanden
<b>Repetierbarkeit</b>	keine Wiederholungsmöglichkeit
<b>Leistungsnachweis</b>	Nachweis von im Selbststudium erbrachten Studienleistungen
<b>Unterrichtssprache</b>	
<b>Lehrformen</b>	Selbststudium

### **Lernziel**

Die Studierenden (1) repetieren und verfestigen Gelerntes (2) erarbeiten neue Inhalte/Themenbereiche in kompakter Form (3) lernen aktuelle Trends kennen (4) tauschen Erfahrungen aus mit Studierenden anderer Universitäten (5) machen Networking auf internationaler Ebene

### **Allgemeine Beschreibung**

Sommerschulen dienen dazu, den Studierenden vertiefenden Einblick in spezifische computerlinguistische Themenbereiche zu geben. Sie festigen das bisher im Studium Gelernte, erweitern ihre Kenntnis der zentralen Inhalte und bieten ihnen in kompakter Weise Zugang zu neuen Ansätzen. Sie werden auf aktuelle Trends aufmerksam, tauschen Erfahrungen und Einschätzungen mit Studierenden anderer Universitäten aus, und bekommen auf diese Weise die Möglichkeit, das internationale Niveau kennenzulernen und gleichzeitig Beziehungen zu knüpfen, die über das Studium hinaus hilfreich sein können. Dieses Modul kann gebucht werden, um den Besuch von Sommerschulen im Bereich der Sprachverarbeitung anrechnen zu lassen. Dieses Modul kann mit einer Punktzahl von 3 oder 6 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 die Sommerschule besucht wird, muss die Anrechnung mit dem Modulverantwortlichen abgesprochen werden.



## Praktikum extern

06SM521-016

**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 Arbeitswelt (2) arbeiten an der Verbindung von Theorie und Praxis (3) lernen die Abläufe in Firmen kennen (4) wenden Gelerntes an (5) erweitern ihr Wissen um praxisrelevante Fragestellungen

### Allgemeine Beschreibung

Die Studierenden sammeln Erfahrung in der Praxis. Sie erhalten Einblick in die Abläufe in Betrieben, sind beteiligt an der Umsetzung computerlinguistischer Methoden zur Lösung firmenspezifischer Probleme. Sie wenden das im Studium Gelernte auf kommerzielle Fragestellungen an und passen es an die Anforderungen der Arbeitswelt an. Externe Praktika sind in der Regel Aufenthalte in Firmen oder öffentlichen Organisationen, welche auf dem Gebiet der Sprachtechnologie tätig sind. Die Praktika müssen fachbezogen sein und müssen selber organisiert werden. Dieses Modul kann mit einer Punktzahl von 3 oder 6 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. Damit externe Praktika angerechnet werden können, bedarf es der Bewilligung eines vorgängigen Gesuchs durch den Modulverantwortlichen. Externe Praktika sind für fortgeschrittene Studierende gedacht.



## 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.





## Programmierprojekt 1

---

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 (1) planen ein Projekt (2) setzen einen Projektplan um (3) realisieren Software (4) halten Dokumentationsstandards ein (5) evaluieren (6) 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

Im Laufe einer Studienstufe können max. zwei Programmierprojekte gebucht werden. 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.

---



## Übungsleitung/Tutorat 1

---

06SM521-019

---

**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

Die Studierenden (1) verstehen fachspezifische Inhalte aus einer Lehrperspektive (2) lernen Vorlesungsinhalte zielgruppengerecht aufzubereiten und als Übungsmaterial zu konzipieren (3) lernen zu korrigieren und Rückmeldung zu geben

---

### Allgemeine Beschreibung

Ein Tutorat/Übungsleitung dient der Erlernung von ersten didaktischen Fertigkeiten. Dies setzt ein reflektiertes Verständnis fachspezifischer Inhalte voraus und erfordert die Aufbereitung von Lernmaterialien für das Repetieren der Inhalte aus der Vorlesung, sowie das Konzipieren und Korrigieren von Übungsblättern. Dieses Modul kann mit einer Punktzahl von 3 oder 6 gebucht werden. Die Punktzahl wird in Absprache mit dem Modulverantwortlichen festgelegt.

---

### Voraussetzungen

Im Laufe einer Studienstufe können max. zwei Module «Übungsleitung/Tutorat» gebucht werden, wobei sich die beiden Module inhaltlich unterscheiden müssen. Dieses Modul wird gebucht, um sich eine erste Übungsleitung anrechnen zu lassen.

Für dieses Modul muss man sich bewerben. Zuvor muss zwingend Rücksprache mit dem Modulverantwortlichen genommen werden (per E-Mail). Die Dozierenden der Lehrveranstaltung müssen in den E-Mail-Verkehr einbezogen sein. Übungsleitungen werden normalerweise einige Wochen vor Semesterbeginn auf der Mailingliste der Computerlinguistik ([cl-list@lists.ifi.uzh.ch](mailto:cl-list@lists.ifi.uzh.ch)) ausgeschrieben. Die Studierenden können sich auch gerne jederzeit spontan bei den Dozierenden bewerben, für deren Lehrveranstaltung sie eine Übungsleitung übernehmen möchten. Für Übungsleitungen/Tutorate muss das entsprechende Modul selbst erfolgreich bestanden sein.

---



## Übungsleitung/Tutorat 2

---

06SM521-021

---

**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

Die Studierenden (1) verstehen fachspezifische Inhalte aus einer Lehrperspektive (2) lernen Vorlesungsinhalte zielgruppengerecht aufzubereiten und als Übungsmaterial zu konzipieren (3) lernen zu korrigieren und Rückmeldung zu geben

---

### Allgemeine Beschreibung

Ein «Tutorat/Übungsleitung» dient der Erlernung von ersten didaktischen Fertigkeiten. Dies setzt ein reflektiertes Verständnis fachspezifischer Inhalte voraus und erfordert die Aufbereitung von Lernmaterialien für das Repetieren der Inhalte aus der Vorlesung, sowie das Konzipieren und Korrigieren von Übungsblättern. Dieses Modul kann mit einer Punktzahl von 3 oder 6 gebucht werden. Die Punktzahl wird in Absprache mit dem Modulverantwortlichen festgelegt.

---

### Voraussetzungen

Im Laufe einer Studienstufe können max. zwei Module "Übungsleitung/Tutorat" gebucht werden, wobei sich die beiden Module inhaltlich unterscheiden müssen. Dieses Modul wird gebucht, um sich eine zweite Übungsleitung anrechnen zu lassen.

Für dieses Modul muss man sich bewerben. Zuvor muss zwingend Rücksprache mit dem Modulverantwortlichen genommen werden (per E-Mail). Die Dozierenden der Lehrveranstaltung müssen in den E-Mail-Verkehr einbezogen sein. Übungsleitungen werden normalerweise einige Wochen vor Semesterbeginn auf der Mailingliste der Computerlinguistik ([cl-list@lists.ifi.uzh.ch](mailto:cl-list@lists.ifi.uzh.ch)) ausgeschrieben. Die Studierenden können sich auch gerne jederzeit spontan bei den Dozierenden bewerben, für deren Lehrveranstaltung sie eine Übungsleitung übernehmen möchten. Für Übungsleitungen/Tutorate muss das entsprechende Modul selbst erfolgreich bestanden sein.

---



## Programmierprojekt 2

---

06SM521-022

---

**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 (1) planen ein Projekt (2) setzen einen Projektplan um (3) realisieren Software (4) halten Dokumentationsstandards ein (5) evaluieren (6) nutzen Softwarerepositories

---

### Allgemeine Beschreibung

Dieses Modul kann gebucht werden, um ein zweites Programmierprojekt anrechnen zu lassen. 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 mit einer Punktzahl von 3, 6 oder 9 gebucht werden. Die Punktzahl wird in Absprache mit dem Modulverantwortlichen festgelegt.

---

### Voraussetzungen

Im Laufe einer Studienstufe können max. zwei Programmierprojekte gebucht werden. 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.

---