

Informationsveranstaltung

BA Computerlinguistik

Korpus- und Computerlinguistik

Prof. Dr. Stephanie Evert

Lehrstuhl für Korpus- und Computerlinguistik

<https://www.linguistik.phil.fau.de/>

WS 2025/26



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

PHILOSOPHISCHE FAKULTÄT
UND FACHBEREICH THEOLOGIE

Was ist eigentlich ein Korpus?

Das Korpus
Die Korpora

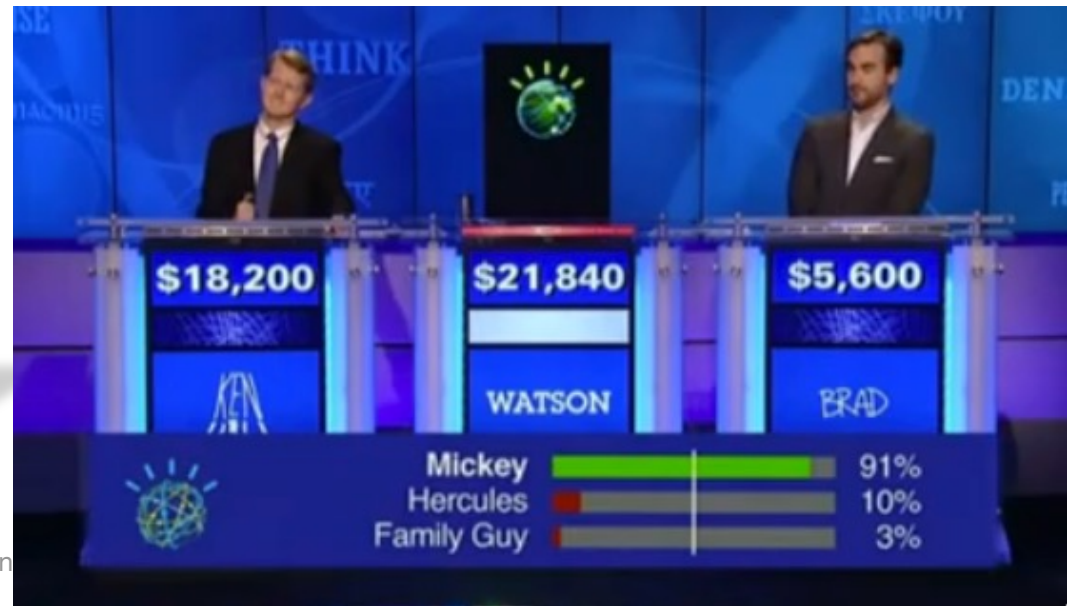
- Korpus¹ (im weiten Sinn)
= Sammlung von Sprachdaten / Texten in maschinenlesbarer Form
 - sehr große Korpora (≥ 100 M Wörter) sind besonders nützlich
 - Auswertung mit statistischen Methoden und maschinellen Lernverfahren
- Korpus² (im engen Sinn)
= Stichprobe authentischer Sprachdaten / Texte, die für eine bestimmte Sprache oder Sprachvarietät repräsentativ ist
 - z.B. literarische Korpora, Dialekte, gesprochene Sprache, IBK, ...
 - Basis für empirisch fundierte sprachwissenschaftliche Studien
 - zentral für Korpuslinguistik (im engen Sinn) und *Digital Humanities*
- **Korpus- und Computerlinguistik**
= Erstellung, maschinelle Verarbeitung und Auswertung von Korpora¹

Was macht eigentlich Computerlinguistik?

- Linguistische Analyse von Texten (Wortarten, Syntax, ...)
- Text Mining, Knowledge Mining, Argument Mining, ...
- Erkennung von Autoren, Meinungen, Sentiment, Emotionen, ...
- Erkennung von Fake News, Hate Speech, Desinformation, ...
- Maschinelle Übersetzung
- Spracherkennung & Sprachsynthese
- Dialogsysteme & virtuelle Assistenten
- Generierung und Überarbeitung von Texten
- Computerlexikographie & Terminologearbeit
- Rechtschreib- und Grammatikkorrektur, stilistische Verbesserungen
- Unterstützung für Korpuslinguistik, Digital Humanities und Computational Social Science

Künstliche Intelligenz (KI): virtuelle Assistenten

IN 1939's CARTOON
"THE POINTER", THIS
GUY GOT A NEW,
MORE PEAR-SHAPED
BODY & PUPILS WERE
ADDED TO HIS EYES



Künstliche Intelligenz (KI): virtuelle Assistenten

THIS CLAUSE IN A UNION CONTRACT SAYS THAT WAGES WILL RISE OR FALL DEPENDING ON A STANDARD SUCH AS COST OF LIVING



Methoden der CL: symbolische Verfahren



[Text](#) [Web Page](#) [RSS](#) [File](#) [Dictionary](#) [My Dictionary](#) [Register for free](#)

From: To: Options:

For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much — the wheel, New York, wars and so on — whilst all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man — for precisely the same reasons.

Zum Beispiel auf der Planet Erde, hatte Mann immer angenommen, dass er intelligenter war, als waren Delphine, weil er soviel erzielt hatte — das Rad, New York, Kriege und so weiter — während alle Delphine überhaupt getan hatten, Dreck ungefähr im Wasser, das eine gute Zeit hat. Aber andererseits, hatten die Delphine immer geglaubt, dass sie weit intelligenter als Mann — aus genau den gleichen Gründen waren.

Methoden der CL: statistische Verfahren

≡ Google Translate



Text

Documents

DETECT LANGUAGE

ENGLISH

SPANISH

FRENCH



GERMAN

ENGLISH

SPANISH



For instance, on the planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much—the wheel, New York, wars and so on— whilst all the dolphins had ever done was muck about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man—for precisely the same reasons. X

Zum Beispiel hatte der Mensch auf dem Planeten Erde immer angenommen, er sei intelligenter als Delfine, weil er so viel erreicht hatte - das Rad, New York, Kriege und so weiter -, während alles, was die Delfine jemals getan hatten, im Wasser herumgespielt hatte eine gute Zeit haben. Umgekehrt hatten die Delfine immer geglaubt, sie seien weitaus intelligenter als der Mensch - aus genau den gleichen Gründen. ☆



380 / 5000



Methoden der CL: Deep Learning



DeepL

Übersetzer ▾

DeepL Pro

Für Unternehmen ▾

Kostenloses Probeabo starten



Anmelden

Text übersetzen
31 SprachenDateien übersetzen
.pdf, .docx, .pptxDeepL Write
KI-Schreibassistent

Englisch (erkannt) ▾



Deutsch ▾

automatisch ▾

Glossar

For instance, on planet Earth, man had always assumed that he was more intelligent than dolphins because he had achieved so much—the wheel, New York, wars and so on—whilst all the dolphins had ever done was much about in the water having a good time. But conversely, the dolphins had always believed that they were far more intelligent than man—for precisely the same reasons.



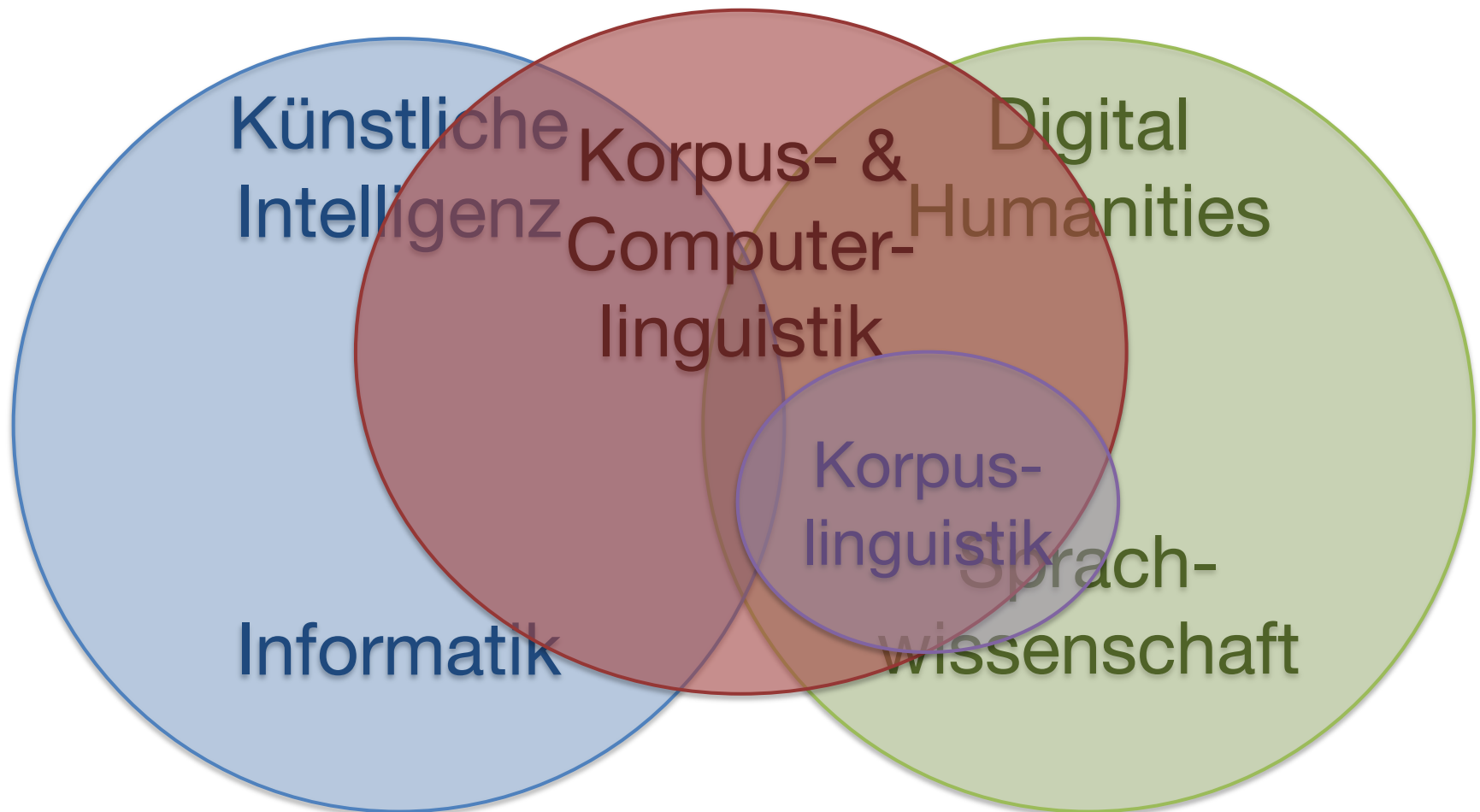
Auf dem Planeten Erde zum Beispiel hatte der Mensch immer angenommen, dass er intelligenter sei als die Delphine, weil er so viel erreicht hatte - das Rad, New York, Kriege und so weiter - während alles, was die Delphine je getan hatten, darin bestand, im Wasser herumzutollen und sich zu amüsieren. Umgekehrt hatten die Delfine immer geglaubt, sie seien viel intelligenter als der Mensch - aus genau denselben Gründen.



Methoden: generative KI



Orientierung: Korpus- und Computerlinguistik



BA Computerlinguistik

- Fokus auf maschinelle Lernverfahren (insb. **Deep Learning**)
 - folgt dem Trend der modernen Computerlinguistik (und KI)
 - Schwerpunkt des Studiums: Operationalisierung von computerlinguistischen Aufgaben als Lernprobleme
 - „Baukasten“ geeignet für zahlreiche Anwendungen in Sprachtechnologie, DH, Korpuslinguistik
- Programmierpraxis: **Python**
 - konsequenter Einsatz von Python und entsprechenden Frameworks im ganzen Studienverlauf
 - entsprechende Kenntnisse werden z.B. in praktischen Hauptseminaren vorausgesetzt
- Aber keine Reduktion auf *prompt engineering* für ChatGPT
 - formale und theoretische Grundlagen der Computerlinguistik
 - interdisziplinäre Anwendungen v.a. im Rahmen der Hauptseminare (politische Diskurse, Digital Humanities, Konstruktionsgrammatik, ...)

Grundstudium

- **Grundlagen der Computerlinguistik I – III**

- GdCL I = klassische symbolische Ansätze, Mengenlehre, Logik
- GdCL II = statistische Ansätze, maschinelle Lernverfahren, Wahrscheinlichkeitstheorie, lineare Algebra
- GdCL III = Deep Learning = neuronale Netze, Tensoranalysis, praktische Implementierung

- **Programmierung & Infrastrukturen I + II**

- Unix-Betriebssystem, Kommandozeile, Editor, reguläre Ausdrücke
- Einführung in Python
- Nutzung von Python-Bibliotheken und NLP-Standardwerkzeugen

- **Linguistische Grundkompetenzen**

- Grundlagen der germ. Linguistik 1 oder Basismodul II: Linguistics (A)
- bei Germanistik, Anglistik, Romanistik: DH-Modul 1 „Sprache & Text“

- **Proseminar** Computerlinguistik

- Rezeption von Originalarbeiten, Präsentation, konstruktive Diskussion
- Schreiben von Hausarbeiten, Literaturrecherche & Zitierung, LaTeX

Hauptstudium

- **3 Hauptseminare** zu wechselnden Themen
 - 1x mit klassischer Hausarbeit
 - 1x mit mündlicher Prüfung
 - 1x mit praktischem Projekt + Projektbericht
- **1 Praxisseminar**
 - angewandtes Gruppenprojekt (z.B. Teilnahme an Shared Task)
 - gemeinsamer Projektbericht in Form eines Konferenzbeitrags
 - Schwerpunkt auf Teamwork, Selbstorganisation, Arbeitsteilung, ...
- Angebot: mind. **2 Seminare / Semester**
- **Oberseminar** (verpflichtend)
 - Besuch der Vorträge über 2 Semester (→ Anmeldung in StudOn)
 - Essay über zwei ausgewählte Vorträge
- **Praktikum** (150h = 1 Monat Vollzeit)
 - bei externer Firma oder internes Forschungspraktikum

Erstfach ≠ Zweitfach

- **Zweitfach:** anwendungsorientierte Ergänzung zu geistes- oder sozialwissenschaftlichem Fach
 - komplett ohne Informatik-Importe
 - Vermittlung der mathematischen und informatischen Grundlagen ganz auf CL zugeschnitten
- **Erstfach:** Interesse an Informatik und NLP-Methoden
 - + 20 ECTS statt Schlüsselqualifikationen
 - 7,5 ECTS **Grundlagen der Informatik** (verpflichtend)
 - 7,5 ECTS + 5,0 ECTS **Wahlpflichtbereich**
 - zahlreiche Module aus dem Bachelor Informatik stehen zur Wahl
 - *Mathematik für Naturwissenschaftler* (7,5) und *Modellbildung & Statistik* (5)

Studienplan

Computerlinguistik (Zweifach)



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
PHILOSOPHISCHE FAKULTÄT
UND FACHBEREICH THEOLOGIE

Semester 1 (WiSe)	Semester 2 (SoSe)	Semester 3 (WiSe)	Semester 4 (SoSe)	Semester 5 (WiSe)	Semester 6 (SoSe)
VL Grundlagen der CL 1 2 SWS; 2 ECTS	VL Grundlagen der CL 2 2 SWS; 2 ECTS	VL Grundlagen der CL 3 2 SWS; 3 ECTS	Hauptseminar I 2 SWS; 5 ECTS; <i>Hausarbeit</i>	Hauptseminar III 2 SWS; 5 ECTS; <i>praktisches Projekt</i>	
Ü Grundlagen der CL 1 2 SWS; 3 ECTS	Ü Grundlagen der CL 2 2 SWS; 3 ECTS	Ü Grundlagen der CL 3 2 SWS; 7 ECTS	Hauptseminar II 2 SWS; 5 ECTS; <i>mündl. Prüfung</i>	Projektseminar 2 SWS; 5 ECTS; <i>Teamprojekt</i>	
S Grundkurs Programmierung 2 SWS; 5 ECTS	S Aufbaukurs Programmierung 2 SWS; 5 ECTS	Proseminar Computerling. 2 SWS; 5 ECTS	Oberseminar Computerlinguistik 2×1 SWS; 5 ECTS		
Einführung Linguistik 4 SWS; 5 ECTS; <i>Import Angl./Germ.</i>	DH-Modul 1: Sprache & Text 4 SWS; 5 ECTS; <i>Import DGuS</i>			Praktikum (extern / intern) 150 h = 1 Monat Vollzeit; 5 SWS; unbenotet	

10 SWS 15 ECTS	6 SWS 10 ECTS	6 SWS 15 ECTS	5 SWS 12,5 ECTS	5 SWS 12,5 ECTS	0 SWS 5 ECTS
-------------------	------------------	------------------	--------------------	--------------------	-----------------

Studienplan

Computerlinguistik (Erstfach)



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
PHILOSOPHISCHE FAKULTÄT
UND FACHBEREICH THEOLOGIE

Semester 1 (WiSe)	Semester 2 (SoSe)	Semester 3 (WiSe)	Semester 4 (SoSe)	Semester 5 (WiSe)	Semester 6 (SoSe)
VL Grundlagen der CL 1 2 SWS; 2 ECTS	VL Grundlagen der CL 2 2 SWS; 2 ECTS	VL Grundlagen der CL 3 2 SWS; 3 ECTS	Hauptseminar I 2 SWS; 5 ECTS; <i>Hausarbeit</i>	Hauptseminar III 2 SWS; 5 ECTS; <i>praktisches Projekt</i>	
Ü Grundlagen der CL 1 2 SWS; 3 ECTS	Ü Grundlagen der CL 2 2 SWS; 3 ECTS	Ü Grundlagen der CL 3 2 SWS; 7 ECTS	Hauptseminar II 2 SWS; 5 ECTS; <i>mündl. Prüfung</i>	Projektseminar 2 SWS; 5 ECTS; <i>Teamprojekt</i>	
S Grundkurs Programmierung 2 SWS; 5 ECTS	S Aufbaukurs Programmierung 2 SWS; 5 ECTS	Proseminar Computerling. 2 SWS; 5 ECTS	Oberseminar Computerlinguistik 2×1 SWS; 5 ECTS		
Einführung Linguistik 4 SWS; 5 ECTS; <i>Import Angl./Germ.</i>	DH-Modul 1: Sprache & Text 4 SWS; 5 ECTS; <i>Import DGuS</i>			Praktikum (extern / intern) 150 h = 1 Monat Vollzeit; 5 SWS; unbenotet	
	Grundlagen der Informatik (GdI) 6 SWS; 7,5 ECTS; <i>Import Informatik</i>	Wahlpflichtbereich Informatik <i>ca. 6 SWS; 12,5 ECTS; Import Informatik (+ andere)</i>			
10 SWS 15 ECTS	12 SWS 17,5 ECTS	8 SWS 20 ECTS	9 SWS 20 ECTS	5 SWS 12,5 ECTS	0 SWS 5 ECTS

Nach dem Studium

- Masterstudium / Promotion
 - Computerlinguistik
 - Sprachwissenschaft / Linguistik / Lexikographie (EMLex)
 - Digital Humanities, Computational Social Science
 - Angewandte Informatik
- Tätigkeitsbereiche in der Wirtschaft (→ Sprachtechnologie & KI)
 - Google, Microsoft, Facebook, Amazon, Siemens, ...
 - Text Mining, Information Retrieval, Marktforschung, ...
 - Lexikographie und Terminologie, Legal Tech
 - Spracherkennung und Sprachsynthese, Dialogsysteme
 - Computergestützter Sprachunterricht (CALL)
 - Viele Start-Up-Unternehmen im IT-Bereich und der KI suchen Computerlinguist:innen!

Über uns

- Prof. Dr. Stephanie Evert
stephanie.evert@fau.de
- Dr. Besim Kabashi
- Philipp Heinrich, M.Sc.
philipp.heinrich@fau.de
- Nathan Dykes, M.A.
- Bao Minh Doan Dang, M.A.
baominh.d.dang@fau.de
- Dr. Alexander Piperski
- Mahdi Mantash, M.Sc.
- Naveed Unjum, M.Sc.
- Steffen Bothe, M.Sc.
- Timm Weber, B.A.
- Anke Lutz
ccl-sekretariat@fau.de

www.linguistik.phil.fau.de/team/

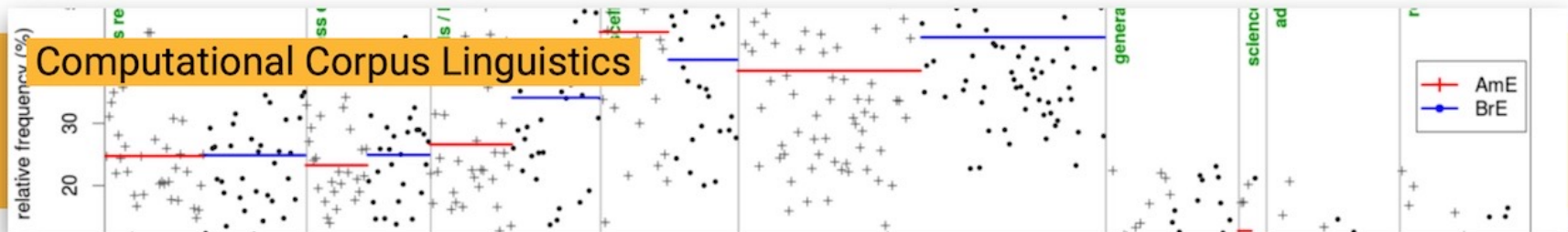
Bismarckstr. 6
Raum 4.000
(Lehrstuhlbüro)

Philologien



(CIP-Pool)
Bismarckstr. 12
Raum 0.320

Ganz wichtig: <https://www.linguistik.fau.de/>



Welcome!

The computational corpus linguistics group carries out foundational methodological research on the quantitative analysis of large text corpora. Our research lies at the intersection of corpus linguistics, computational linguistics, and the digital humanities.

We offer a Bachelor's degree program in computational linguistics ("Linguistische Informatik"). Teaching usually takes place in German. Students can find information about the program and currently offered courses in the teaching portal.

Computational Corpus Linguistics

Lehrstuhl für Korpus- und Computerlinguistik
Department Germanistik und Komparatistik

Bismarckstr. 6
91054 Erlangen

+49913185-22426

info@linguistik.uni-erlangen.de

http://linguistik.fau.de

CL Events

OCT 17 4:15 PM - 6:00 PM
Einführungsveranstaltung BA

- Sehr kleiner Studiengang (bisher ca. 30 Studierende)
- Fachschaftsinitiative (bitte dort engagieren!):
<https://www.linguistik.phil.fau.de/teaching/fsi-computerlinguistik/>
- FSI-Stammtisch

Einführungsveranstaltung: wichtigen Informationen zum neuen Studiengang BA Computerlinguistik am Montag, 17.10.2023 (14:15-16:00 Uhr, Raum 2), Bismarckstr. 1a.

Stundenplan

siehe <https://www.linguistik.fau.de/teaching/lehrveranstaltungen/>

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag
08–10				09:30–10:15 Fragestunde zu GdCL 1 (P. Heinrich) Bismarckstr. 12, 0.320	
10–12			Sprechstunde Prof. Evert Bismarckstr. 6, 4.006 / Zoom (Anmeldung über StudOn)	Übung GdCL 1: Traditionelle Verfahren (P. Heinrich) Bismarckstr. 12, 0.320	Proseminar Computerlinguistik (S. Evert) Bismarckstr. 12, 0.320
12–14					
14–16		Vorlesung GdCL 3: Deep Learning (S. Evert) Bismarckstr. 12, 0.320			HS Erkennung und Bewertung KI-generierter Texte (S. Evert) Bismarckstr. 12, 0.320 + Sondertermine gem. StudOn
16–18	Programmierung & Infrastrukturen I (B. M. Doan Dang) Bismarckstr. 12, 0.320	Übung GdCL 3: Deep Learning (S. Evert) Bismarckstr. 12, 0.320	Oberseminar Computerlinguistik (S. Evert) Bismarckstr. 12, 0.320		

Lehrangebot im WS 2024/25

- VL + Ü Grundlagen der Computerlinguistik 1
 - PS Programmierung & Infrastrukturen 1
 - Linguistische Grundkompetenzen (teilweise)
 - Modul Grundlagen der Informatik (Erstfach, optional)
- } **1. Sem.**
- VL + Ü Grundlagen der Computerlinguistik 3
 - PS Computerlinguistik
- } **3. Sem.**
- HS: Erkennung & Bewertung KI-generierter Texte
 - PraxisS: Erkennung & Bewertung KI-generierter Texte
 - ggf. weitere HS als Importe (→ selbst anfragen)
- } **5. Sem.**
- Oberseminar Computerlinguistik
 - Wahlpflichtbereich Informatik
 - Schlüsselqualifikationen
- } **allgemein**

Informationen für Erstsemester

Grundlagen der Computerlinguistik I

- **Vorlesung:** **asynchron** (Zeitfenster **Mo 14:00 – Do 9:00**)
 - Bereitstellung als Screencasts von 45–60 Minuten Dauer
 - Dozent: Prof. Dr. Stefan Evert 😊
 - **Fragestunde:** **Do 9:30–10:15** (CIP-Pool)
 - Fragen und Diskussion zur Vorlesung
 - **Übung:** **Do 10:15–11:45** (CIP-Pool)
 - Veranschaulichung mit Beispielen und interaktiven Übungen
 - wöchentliche Übungsaufgaben zur schriftlichen Bearbeitung
 - Dozentin: Prof. Dr. Stephanie Evert
 - **Tutorium:** **n.V.** (Tutor/in: Etienne Krembs, Laura Tatlow)
- } **Beginn: Do 16.10.**

Informationen für Erstsemester

Programmierung & Infrastrukturen I

- **Grundkurs Programmierung:** Mo 16:15–17:45 (CIP-Pool)
 - erste Sitzung am Montag 20.10. (→ weitere Organisation)
 - Dozent: Bao Minh Doan Dang, M.Sc.
- **Tutorium:** n.V. (Tutor/in: Etienne Krembs / Laura Tatlow)

Informationen für Erstsemester

Importmodule

- **Grundlagen der germanistischen Linguistik 1 oder Basismodul II: Linguistics (A)**
 - falls zweites Fach nicht Germanistik, Anglistik oder Romanistik
 - jeweils mehrere Gruppen in verschiedenen Zeitfenstern
- **Grundlagen der Informatik (nur Erstfach)**
 - Vorlesung: Do 8:30–12:00 / „asynchron“ als Videoaufzeichnung
 - Fragestunde zu GdI: Di 10:15–11:45 (online)
 - Sprechstunden zu GdI: verschiedene Termine (Chat / virtuell)
 - alternativ: im 2. Semester belegen

Nächste Schritte

- Alle LVen werden durch **StudOn-Kurse** begleitet
 - separater Kurs für jede Lehrveranstaltung in Campo
 - dort aktuelle Informationen, Materialien, Zugang Videos + Zoom
 - gilt unseres Wissens als offizielle Anmeldung
- Lehrveranstaltungen für 1. Semester in dieser Woche
 - Vorlesung GdCL1: ersten Screencast schauen (bis Do 16.10.)
 - Fragestunde + Übung GdCL 1: beginnt **am Do 16.10.**
 - Linguistische Grundkompetenzen / GdI nicht vergessen!
- Prüfungsanmeldung über **Campo** nicht vergessen!
 - Anmeldezeitraum: **17.11.–07.12.2025**
- **Accounts** für CIP-Pool = IdM-Kennung + Passwort
- **Sprechstunde** Prof. Dr. Stephanie Evert: **hybrid**
 - Mi 10:00–12:00 (nur mit Voranmeldung per StudOn)
 - Termine ab Mi 05.11. verfügbar