

Drehbuch FS25: Natural Language Processing (nlp)

Typ: Vorlesung
Studiengang: Informatik / iCompetence
Umfang: 3 ECTS

Verfasser: Stephan Heule
Datum: 17.02.2025

1 Funktion im Rahmen der Gesamtausbildung

1.1 Leitidee

Die Verarbeitung von Sprache (Natural Language Processing / NLP) ist ein sehr wichtiger Teil der Informationsbeschaffung und gehört deshalb zu den Data Science Grundkompetenzen. Sowohl unsere Gesellschaft als Ganzes als auch eine Grosszahl unsere Geschäftsprozesse sind einer fortschreitenden Digitalisierung unterworfen. Im Modul NLP thematisieren wir Herausforderungen, Techniken und Lösungsansätze rund um das Thema der maschinellen und algorithmischen Informationsgewinnung und -verarbeitung natürlicher Sprache.

1.2 Voraussetzungen

- Objektorientierte Programmierung (oop1 & oop2)
- Mathematische Grundlagen (mgli, lag & dist)

2 Unterricht

2.1 Lern- und Arbeitsformen

Die grundlegenden Ideen werden in 2 Kontaktlektionen pro Woche erarbeitet. Dazu kommt 1 Übungslektionen (mit Betreuung) sowie ca. 3 Stunden individuelle Arbeit (Selbststudium, Übungen und wahlweise einer benoteten Challenge) pro Woche.

2.2 Zeitaufwand

Kontaktunterricht + Übungen:	40h
Selbststudium:	50h
Gesamtaufwand:	90h

2.3 Leistungsbeurteilung

Die Modulnote ergibt sich aus:

- einer schriftlichen Modulschlussprüfung (90 Minuten) über den Vorlesungsstoff und die Übungen (ohne Hilfsmittel)
- auf freiwilliger Basis kann ein *individueller* Notenbonus von max. +0.4 erbracht werden durch Bearbeitung der wöchentlichen Übungen (wird zur schriftlichen Modulschlussprüfung addiert):
 - *schriftliche Abgabe*: pro schriftlich eingereichte Übungsaufgabe als Notebook oder Scan/Foto +0.05. Max. +0.2 Notenbonus (d.h. 4 eingereichte Aufgaben)
 - *Vorlösen* einer Aufgabe in der Übungsstunde: +0.2 Notenbonus
- freiwilliges, benotetes *NLP-Projekt* in Zweiergruppen. Bewertungskriterien:
 - Projekt-Vorschlag (→ One Pager ausfüllen und mein Feedback einbauen)

- Projektabgabe (Blog-Post, Paper oder Notebook; inkl. Git-Repo)
- Details werden mit Projekt-Vorschlag-One-Pager kommuniziert
- Für GPUs stehen CHF 60.- pro Person zur Verfügung

Bemerkungen:

- Pro Übungsserie kann nur ein Bonus erlangt werden (d.h. eine Aufgabe einreichen oder vorlösen). Für den vollen Bonus müssen somit 5 «Übungsblätter» bearbeitet werden
- Notenbonus: es wird keine perfekte Lösung erwartet (ausschlaggebend ist eine ernsthafte Auseinandersetzung mit dem Übungsstoff)
- Projektnote zählt 1/3, sofern sie den Notenschnitt der schriftlichen Prüfung erhöht (Bonus)
- Eine allfällige Nachholprüfung erfolgt mündlich und dauert 30 Minuten

2.4 Arbeitsmittel

Folien, Übungen, Tutorials und Literatur aus dem Internet:

- Dan Jurafsky and James H. Martin, **Speech and Language Processing**, 3rd edition draft, February 3rd, <https://web.stanford.edu/~jurafsky/slp3/>
- Allen B. Downey: Think Stats (ISBN: 978-1-491-90733-7) – Statistik und Wahrscheinlichkeitsrechnung an Praxisbeispielen mit Python (für Informatiker), gibt es kostenfrei als PDF: <http://greenteapress.com/wp/think-stats-2e/>
- L. Gonick, W. Smith: Statistik in Cartoons (ISBN: 978-3-8006-3598-6) – Gut verständliches Lehrbuch zur Statistik und Wahrscheinlichkeitsrechnung fürs Selbststudium geeignet.

Kommunikation via Teams: [E-NLP FS 2025 M365 | General | Microsoft Teams](#)

3 Drehbuch

Week	Date	Thema	Remarks
1	17.02.2025	Logistics, Introduction to NLP and Machine Learning	3h lecture
2	24.02.2025	Statistical Models	
3	03.03.2025	Vector Representation	
4	10.03.2025	Introduction to Neural Nets	
5	17.03.2025	Recurrent Neural Networks	
6	24.03.2025	Encoder-Decoder, Attention	
7	31.03.2025	Transformer	Tbd: Projektvorschlag einreichen
8	07.04.2025	Machine Translation	Feedback zum Projektvorschlag
9	14.04.2025	Chat Systems and GPT	
-	21.04.2025	Osterferien	
10	28.04.2025	Sampling and Prompting	
-	05.05.2025	Projektwoche	
11	12.05.2025	Safety and Privacy	
12	19.05.2025	Speech	
13	26.05.2025	Questions, Spare time	
14	02.06.2025	Exam (90 Min)	
15	09.06.2025	Exam discussion	Abgabe: Projekt