Introduction

HS Sprachtechnologie für eine bessere Welt (Winter semester 2022/23)

Nils Reiter,
nils.reiter@uni-koeln.de

October 11, 2022

INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

Section 1

Kennenlernen

Vier Fragen

- Wer sind Sie? (Name, Studiengänge, Lieblingsessen- oder videospiel, ...)
- Was erwarten Sie von dieser Veranstaltung?
  Gibt es etwas was sie gerne diskutieren/behandeln/lernen möchten?
- Wie würden Sie mit Sprachtechnologie die Welt verbessern?
- Wie geht's?

Section 2

Sprachtechnologie für eine Bessere Welt?

Language Technology and the World (1/2)

- NLP before 2010: Not good enough for serious applications
  - My professor to us: “Never use NLP for critical things!”

Language Technology and the World (1/2)

- NLP before 2010: Not good enough for serious applications
  - My professor to us: “Never use NLP for critical things!”
- 2011: IBM Watson beats two humans in Jeopardy

Video

Language Technology and the World (1/2)

- NLP before 2010: Not good enough for serious applications
  - My professor to us: "Never use NLP for critical things!"
- 2011: IBM Watson beats two humans in Jeopardy Video
- 2013: Word embeddings (word2vec etc.)
  - Enables: $\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \simeq \vec{\text{queen}}$

Language Technology and the World (2/2)

- 2013: Edward Snowden leaks how the NSA spies on everyone
  - Three factors make this possible: NoSQL, Machine learning/natural language processing, Apache Hadoop (WSJ, June 2013)

PDF in Ilias

Language Technology and the World (2/2)

- 2013: Edward Snowden leaks how the NSA spies on everyone
  - Three factors make this possible: NoSQL, Machine learning/natural language processing, Apache Hadoop (WSJ, June 2013)
- 2015: First release of deep learning library tensorflow
  - Since then: Adoption of DL in NLP, boosting performance numbers substantially

PDF in Ilias

Language Technology and the World (2/2)

- 2013: Edward Snowden leaks how the NSA spies on everyone
  - Three factors make this possible: NoSQL, Machine learning/natural language processing, Apache Hadoop (WSJ, June 2013) PDF in Ilias
- 2015: First release of deep learning library tensorflow
  - Since then: Adoption of DL in NLP, boosting performance numbers substantially
- 2018: Transformer models (BERT etc.)
  - Since then: Adoption of BERT in NLP, boosting performance numbers substantially (again!)

Language Technology and the World (2/2)

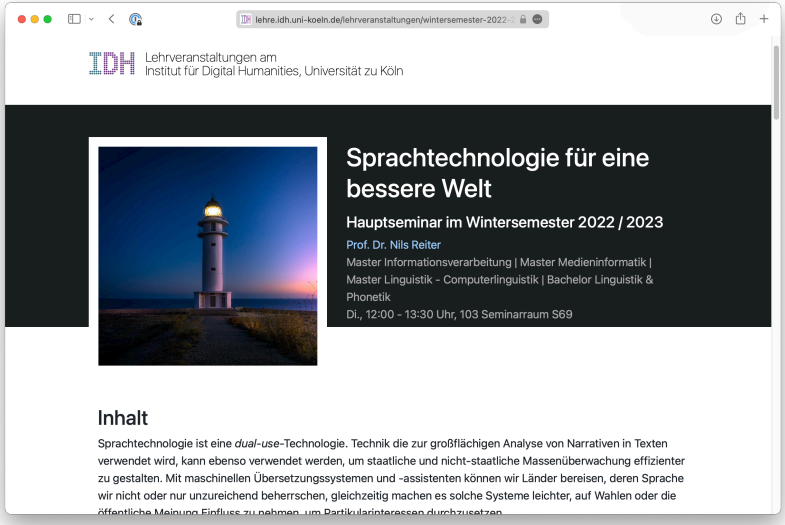
- 2013: Edward Snowden leaks how the NSA spies on everyone
  - Three factors make this possible: NoSQL, Machine learning/natural language processing, Apache Hadoop (WSJ, June 2013) PDF in Ilias
- 2015: First release of deep learning library tensorflow
  - Since then: Adoption of DL in NLP, boosting performance numbers substantially
- 2018: Transformer models (BERT etc.)
  - Since then: Adoption of BERT in NLP, boosting performance numbers substantially (again!)
- Today
  - Serious use of machine translation
  - Serious use of AI assistants
  - Automated decision making in more and more areas
  - 'Obvious' biases in applications
    - E.g., in back-and-forth translation with `https://www.deepl.com/translator`

Section 3

Ablauf und Organisatorisches

Lernziele

- Lesen und Verstehen von (technischer) Forschungsliteratur
- Aufbereitung eines Themas in Form eines Vortrages
- Überblick über aktuelle Forschungsthemen und -herausforderungen der Computerlinguistik
- Tieferen Einblick in ein Forschungsthema



https://uni.koeln/FCKSV

Modul: Verarbeitung von Textdaten

MA Informationsverarbeitung

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar		
Übung		
Kolloquium		
Modulprüfung		

Tabelle: Lehrveranstaltungen im Modul

Modul: Verarbeitung von Textdaten

MA Informationsverarbeitung

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	60
Übung	30	60
Kolloquium	30	60
Modulprüfung	–	270

Tabelle: Lehrveranstaltungen im Modul

- 18 Leistungspunkte
- $30\,\%$ der Fachnote

Modul: Profilmodul Computerlinguistik (1C)

MA Linguistik

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	60
Projektseminar	30	60
Projektseminar	30	60
Modulprüfung	–	180

Tabelle: Lehrveranstaltungen im Modul

- 15 Leistungspunkte
- $40\,\%$ der Fachnote

Modul: Verarbeitung von Textdaten

M.Sc. Informatik

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	240
Übung	30	60
Kolloquium	30	60
Modulprüfung	–	–

Tabelle: Lehrveranstaltungen im Modul

- 15 Leistungspunkte
- $13.1\,\%$ der Fachnote

Modul: Professionalisierung: Forschung (EM 1a)

MA Deutsche Sprache und Literatur

Veranstaltung	Kontaktzeit	Selbststudium
Hauptseminar	30	60
Kolloquium/Oberseminar	30	60
Modulprüfung		180

Tabelle: Lehrveranstaltungen im Modul

- 12 Leistungspunkte
- Das Modul geht nicht in die Fachnote ein.

Ablauf des Seminars

Zeitraum	Ref.	Inhalt
bis 15.11.	Dozent	Grundlagen
22.11. – 13.12.	Stud.	Anwendungen
ab 20.12.	Dozent	Weiterführende Themen

Studienleistung

Die Studienleistung besteht im (betreuten) Gestalten einer Seminarsitzung zu einem Thema in einer Zweiergruppe. Innerhalb dieses Rahmens können und sollen Schwerpunkte gesetzt werden. Es ist nicht erforderlich, eine vollumfängliche Darstellung der Papiere zu liefern. Stattdessen soll das jeweilige Thema anhand der Papiere vorgestellt werden.

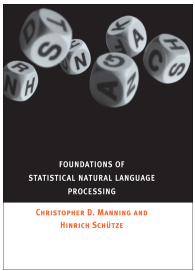
Lehre.IDH

Studienleistung

Ablauf

- Mehr als zwei Wochen vor Termin: In Thema einarbeiten, Literatur durcharbeiten, Forschungsstand aufarbeiten.
  - ⚠ Es ist zwingend erforderlich, sich dafür Zeit zu nehmen und es muss ggf. Sekundärliteratur hinzugezogen werden
- Spätestens zwei Wochen vorher: Sprechstunde zu inhaltlichen Fragen zur Literatur
- Spätestens eine Woche vorher: Konzept zum Referat abgeben. Das Konzept beinhaltet
  - die Gliederung/Struktur,
  - Aufgaben/Fragestellungen für Kleingruppen sowie erwartete Ergebnisse und
  - selbst erstellte Beispiele, die das gesagte verdeutlichen und konkretisieren
- Sitzung leiten. Jede Sitzung muss eine Gruppenaktivität beinhalten
- Im Anschluss: Feedbackgespräch

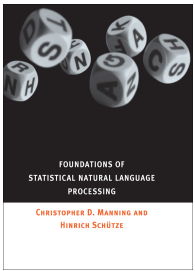
Wichtige Sekundärliteratur



Christopher D. Manning/Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts und London, England: MIT Press

Ilias: Chapters 1-4

Wichtige Sekundärliteratur



Christopher D. Manning/Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts und London, England: MIT Press

Ilias: Chapters 1-4

Dan Jurafsky/James H. Martin (2021). *Speech and Language Processing*. 3. Aufl. Draft of December 29, 2021. Prentice Hall. URL: https://web.stanford.edu/~jurafsky/slp3/

Link

Modulprüfungen

- Thema
  - Findung und Wahl: Ihre Aufgabe
  - Kann, muss aber nicht, etwas mit dem Seminar zu tun haben
  - Mit mir absprechen

Modulprüfungen

- Thema
  - Findung und Wahl: Ihre Aufgabe
  - Kann, muss aber nicht, etwas mit dem Seminar zu tun haben
  - Mit mir absprechen
- Praktischer Anteil: Offen.
  Beispiele: Experiment zur automatischen Identifikation eines Textphänomens, Annotationsexperiment, quantitativer Vergleich verschiedener Korpora, ...
- Am Ende: Hausarbeit von max. 8 S Länge
- Übung vor der Master-Arbeit

Modulprüfungen

- Thema
  - Findung und Wahl: Ihre Aufgabe
  - Kann, muss aber nicht, etwas mit dem Seminar zu tun haben
  - Mit mir absprechen
- Praktischer Anteil: Offen.
  Beispiele: Experiment zur automatischen Identifikation eines Textphänomens, Annotationsexperiment, quantitativer Vergleich verschiedener Korpora, ...
- Am Ende: Hausarbeit von max. 8 S Länge
- Übung vor der Master-Arbeit
- Ilias-Gruppe mit Details (Aufnahme nach Anmeldung)

To Do

- Ab sofort
  - Mit Themen beschäftigen
  - Nachdenken/Erinnern: Was waren besonders gute Referate die Sie erlebt haben? Warum waren die gut?
- Ab 13.10., 9 Uhr: In Ilias ein Thema (und Termin) 'buchen'

To Do

- Ab sofort
  - Mit Themen beschäftigen
  - Nachdenken/Erinnern: Was waren besonders gute Referate die Sie erlebt haben? Warum waren die gut?
- Ab 13.10., 9 Uhr: In Ilias ein Thema (und Termin) 'buchen'
- Nächste Woche
  - Wissenschaftliche Literatur lesen
  - Wissenschaftliche Themen aufbereiten und präsentieren

References I

- Jurafsky, Dan/James H. Martin (2021). *Speech and Language Processing*. 3rd ed. Draft of December 29, 2021. Prentice Hall. URL: https://web.stanford.edu/~jurafsky/slp3/.
- Manning, Christopher D./Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts and London, England: MIT Press.