



NLP: Praktische Einführung mit Python

Vorbesprechung

2019-10-17, Timm Lichte



- 1 Vorstellung
- 2 Thema der Vorlesung
- 3 Ziel der Vorlesung
- 4 Ablauf (vorläufig)
- 5 Nächste Sitzung



- 1 Vorstellung
- 2 Thema der Vorlesung
- 3 Ziel der Vorlesung
- 4 Ablauf (vorläufig)
- 5 Nächste Sitzung



Meine Person:

- Interessen: Computerlinguistik/NLP, insbesondere
 - Präzisionsgrammatiken & Grammar Engineering
 - Semantisches/Syntaktisches Parsen
 - Computational Lexicon (lexikalische Ambiguität, Mehrwortheiten)
 - Metaphern/Lesarten-Erkennung
 - Korpuslinguistik
 - Komplexitätstheorie bei natürlicher Sprache
- Studium (auch Informatik) & Promotion in Tübingen
- PostDoc in Düsseldorf
- seit 2018 am FB Informatik als Koordinator + Lehre



Meine Person:

- Interessen: Computerlinguistik/NLP, insbesondere
 - Präzisionsgrammatiken & Grammar Engineering
 - Semantisches/Syntaktisches Parsen
 - Computational Lexicon (lexikalische Ambiguität, Mehrwortheiten)
 - Metaphern/Lesarten-Erkennung
 - Korpuslinguistik
 - Komplexitätstheorie bei natürlicher Sprache
- Studium (auch Informatik) & Promotion in Tübingen
- PostDoc in Düsseldorf
- seit 2018 am FB Informatik als Koordinator + Lehre
- relevante Hobbys: Emacs & org-mode, \LaTeX



- 1 Vorstellung
- 2 Thema der Vorlesung**
- 3 Ziel der Vorlesung
- 4 Ablauf (vorläufig)
- 5 Nächste Sitzung



“NLP: Praktische Einführung mit Python”

Was ist NLP?



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗
- Neuro-Linguistisches Programmieren ✗



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗
- Neuro-Linguistisches Programmieren ✗
- nicht-lineares Programm ✗



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗
- Neuro-Linguistisches Programmieren ✗
- nicht-lineares Programm ✗
- Natural Language Processing ✓



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗
- Neuro-Linguistisches Programmieren ✗
- nicht-lineares Programm ✗
- Natural Language Processing ✓



“NLP: Praktische Einführung mit Python”

Was ist NLP?

- Nationalpark ✗
- Neuro-Linguistisches Programmieren ✗
- nicht-lineares Programm ✗
- Natural Language Processing ✓

Natural Language Processing / Maschinelle Sprachverarbeitung

Unter dem Begriff der maschinellen Sprachverarbeitung (Natural Language Processing, NLP) fasst man die maschinelle Analyse und Generierung von “natürlichen” Sprachdaten.



- Spracherkennung
- Rechtschreib- und Grammatikkorrektur
- Spam-Erkennung
- Suchmaschinen (in Texten)
- maschinelle Übersetzung
- Chatbots und Sprachassistenzsysteme
- usw.



mostly solved

Spam detection

OK, let's meet by the big ...

Dick too small? Buy VIAGRA ...

Text categorization

Phillies shut down Rangers 2-0 **SPORTS**

Jobless rate hits two-year low **BUSINESS**

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Obama met with UAW leaders in Detroit ...

Information extraction (IE)

You're invited to our bunga bunga party, Friday May 27 at 8:30pm in Cordura Hall

Party May 27 [add](#)

making good progress

Sentiment analysis

The pho was authentic and yummy.

Waiter ignored us for 20 minutes.

Coreference resolution

Obama told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my mouse.

Syntactic parsing

I can see Russia from my house!

Machine translation (MT)

Our specialty is panda fried rice.

我们的专长是熊猫炒饭

still really hard

Semantic search

people protesting globalization Search

...demonstrators stormed IMF offices...

Question answering (QA)

Q. What currency is used in China?

A. The yuan

Textual inference & paraphrase

T. Thirteen soldiers lost their lives ...

H. Several troops were killed in the ... **YES**

Summarization

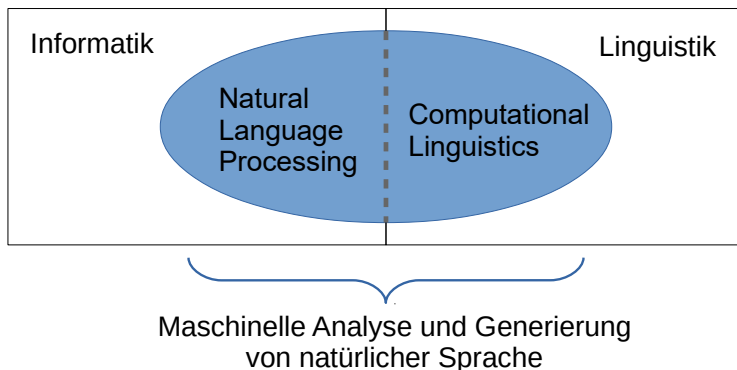
Sheen continues rant against ... Sheen is nuts

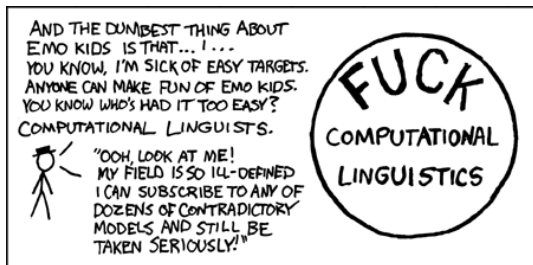
Discourse & dialog

Where is Thor playing in SF?

Metreon at 4:30 and 7:30

(McCartney-Slides, Stanford)





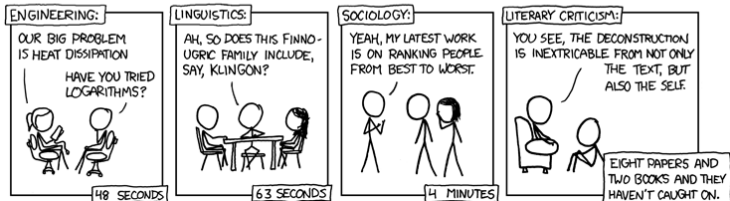
Cursor-Text: “Chomskyists, generative linguists, and Ryan North, your days are numbered.”

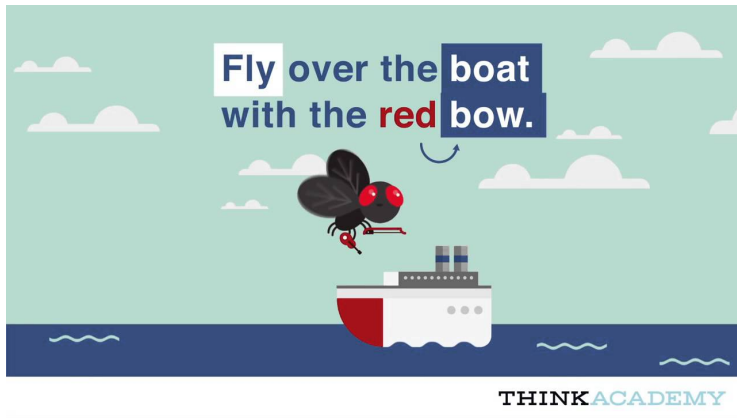
(<https://xkcd.com/114/>)



MY HOBBY:

SITTING DOWN WITH GRAD STUDENTS AND TIMING
HOW LONG IT TAKES THEM TO FIGURE OUT THAT
I'M NOT ACTUALLY AN EXPERT IN THEIR FIELD.






(Quelle: https://www.youtube.com/watch?v=tu5v-gu_5pY)



(Quelle: https://www.youtube.com/watch?v=tu5v-gu_5pY)



- ambig
- fuzzy & vage
 - *rot* 
- komplex & rekursiv
 - *a woman whom another woman whom another woman hired hired hired another woman*
- in ständigem Wandel
 - Neologismen, Code Switching, . . .



- 1 Vorstellung
- 2 Thema der Vorlesung
- 3 Ziel der Vorlesung**
- 4 Ablauf (vorläufig)
- 5 Nächste Sitzung



Wissen:

- grundlegende Konzepte der Linguistik
- grundlegende Konzepte, Ressourcen & Verfahren der NLP

Fähigkeiten:

- Programmierung von einfachen NLP-Pipelines mit Python
- Achtung: Dies ist kein Programmierkurs!

Zielgruppe: Fortgeschrittene Bachelorstudierende



- etabliert
- relativ einfach
- Jupyter-Notebooks
- gute NLP-Bibliotheken
 - NLTK



(<https://xkcd.com/1312/>)



- Inverted Classroom:
 - Sie erhalten die Folien und Notebooks schon vor der Sitzung.
⇒ Ilias
 - Während der Sitzung wird beides besprochen.
- Prüfungsleistungen:
 - Hausaufgaben in den Notebooks (je ca. vier)
 - kleines Abschlussprojekt in Form eines Notebooks
- Benotungsgrundlage:
 - Anteil der richtig bearbeiteten, lauffähigen Hausaufgaben
 - Qualität des dokumentierten und lauffähigen Abschlussprojekts



- 1 Vorstellung
- 2 Thema der Vorlesung
- 3 Ziel der Vorlesung
- 4 Ablauf (vorläufig)**
- 5 Nächste Sitzung



13 Sitzungen:

- Einführung NLTK
- Worte: Normalisierung
- Worte: Distanzmaße und Spell Checking
- Worte: semantische Wortnetze
- Sätze: Tokenisierung
- Sätze: Tagging
- Sätze: Chunking & Chinking
- Sätze: Parsing & Generierung
- Sätze: NER & WSD
- Zusammenfassung



- 1 Vorstellung
- 2 Thema der Vorlesung
- 3 Ziel der Vorlesung
- 4 Ablauf (vorläufig)
- 5 Nächste Sitzung**



Thema: Einführung in NLTK

Voraussetzungen: Installation von Python & Jupyter

Abgabe der Hausaufgaben: spätestens Donnerstag, 24.10., 16:00,
per Email (timmlichte@uni-tuebingen.de)

