



NLP

Eine kurze Einführung



Anna Weißhaar

“Was mit Sprache” gelernt (Linguistik @unihd)

“Was mit Computern” gelernt (Informatik @hska)

Seit 2015 Search Engineer @inovex



Tilman Wittl

“Was mit Sprache und Computern” gelernt
(Computerlinguistik @ Uni Heidelberg)

Seit 2016 Search Engineer @inovex

NLP vs. NLP

Wir beschäftigen uns im Labor mit dem coolen NLP (Natural Language Processing), nicht dem esoterischen (neurolinguistische Programmierung).



Buzzword NLP

Natürliche Sprache dient als Interface

(Amazon Alexa, Google Home, Chatbots,...)

große Mengen an wertvollen Sprachdaten

(Social Media, Kundenbewertungen,...)

Herausforderungen

Natürliche Sprache ...

... ist unendlich.

... ist nicht normiert.

... referenziert auf außersprachliche Dinge.

... verfügt über Subtext.

...

Beispiele

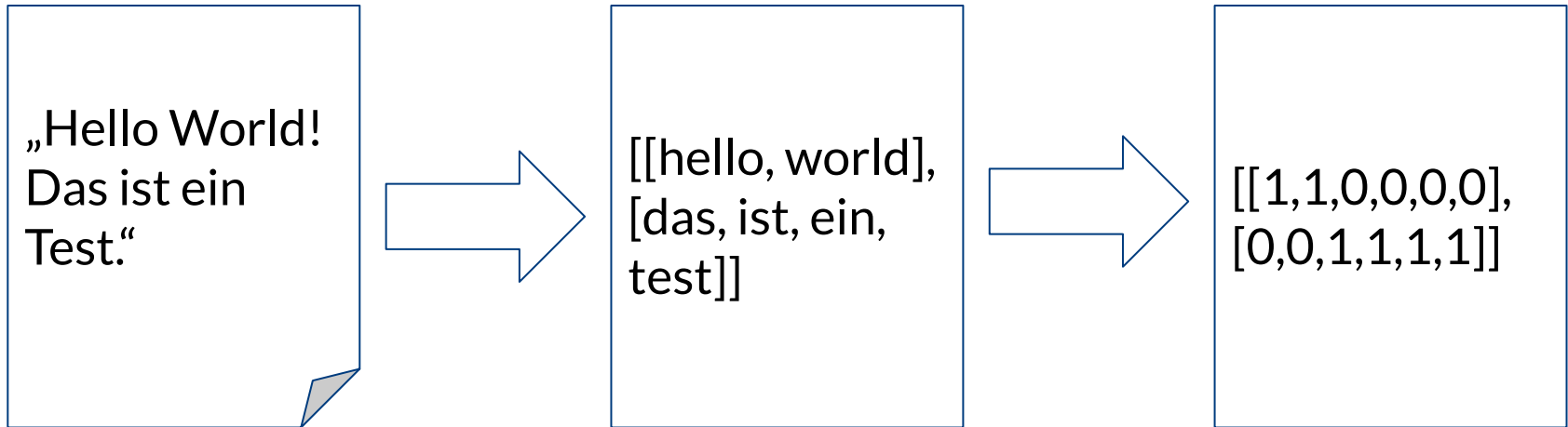
„Alexa, stell den Timer.“, „Ok Google, Timer auf 3 Minuten.“,
...

„Top Ebayer, 5 Sterne, gerne wieder. “

„Ich möchte ein Ticket von Karlsruhe nach Hamburg kaufen.“

Wie kommt die Sprache in den Computer?

NLP: die klassische Verarbeitungskette



NLP: Stichworte

- Vorverarbeitung:
 - Tokenisierung/Normalisierung:
 - Satzgrenzen/Wortgrenzen erkennen
 - Stemming/Lemmatisierung
 - Featurisierung: Token => Feature
 - POS
 - Embedding (semantisch)
- Anwendungen:
 - Spracherkennung
 - NER
 - Sprachgenerierung
 - Textklassifikation
 - Sentiment Analysis
 - ...

<https://github.com/hskaailabnlp/sentiment-analysis>

Vielen Dank

Anna Weißhaar

aweisshaar@inovex.de

Tilman Wittl

twittl@inovex.de

