

## NLP

*Bleier, Roman; roman.bleier@uni-graz.at*

*Natural Language Processing* (NLP) oder auch maschinelle Sprachverarbeitung beschäftigt sich mit der algorithmengestützten Verarbeitung von natürlicher Sprache. Teilaufgaben von NLP beschäftigen sich unter anderem mit Spracherkennung, Tokenisierung von Texten, *Part-Of-Speech-Tagging/PoS*, *Named Entity Recognition/NER* und dem automatisierten Erkennen und der Extraktion der Bedeutung von Wörtern im Satzgefüge und den Beziehungen zu anderen Wörtern sowie von Sätzen zueinander. In sogenannten NLP-Pipelines werden mehrere dieser Teilbereiche aufeinanderfolgend ausgeführt und ein Text (*plain text*) schrittweise mit Information angereichert. Ein Beispiel dafür ist das Onlineservice *WebLicht*. Es erlaubt der Nutzerin bzw. dem Nutzer, eine NLP-Pipeline mit austauschbaren Teilaufgaben, die von Drittanbietern zur Verfügung gestellt werden, zusammenzustellen und auf Texte anzuwenden.

Im Kontext von Digitalen Editionen kann NLP bei der Aufbereitung und Normalisierung von Texten und der automatisierten Anreicherung mit semantischen Informationen zum Einsatz kommen, zum Beispiel beim automatisierten Erkennen von Sätzen und dem Taggen von Personen und Orten. Die Daten von Digitalen Editionen können aber auch als Grundlage für NLP-Analysen dienen. In diesem Falle ist es der fertig edierte, elektronische Text, der mit NLP-Methoden für die Expertenanalyse aufbereitet wird. Ein weiterer Anwendungsfall wäre, dass die Editionsdaten als Trainingsdaten für *Machine Learning* verwendet werden (z. B. Personen- und Ortsdaten für NER).

### Literatur:

- Bird, Steven; Klein, Ewan; Loper, Edward: *Natural Language Processing with Python*. Beijing u.a.: 2009, URL: <https://www.nltk.org/book/>.
- Jurafsky, Daniel; Martin, James H.: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, New Jersey, United States of America: 2009.
- Piotrowski, Michael: *Natural Language Processing for Historical Texts*. Hrsg. von und Graeme Hirst: 2012, URL: <http://doi.org/10.2200/S00436ED1V01Y201207HLT017>.
- *WebLicht*. Main Page Main page. URL: [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page)

### Verweise:

Part-of-Speech-Tagging, spaCy, Named Entity Recognition / NER, Weblicht, xTokenizer, Tagger

**Software:**

CLARIN-mediated NLP-services, enrich/stanbol (ACDH-OeAW), Manuscript Desk, SaltNPepper, topic-modelling-tool, weblicht, TEITOK, Apache OPENNLP, CLAWS POS-Tagger for English, Chaos, CiceroLight, DBpedia Spotlight, FreeLing, IMPACT Tools and Data, Natural Language Toolkit (nltk), Classical Language Toolkit (cltk), spacy , flair, German NER, GermaNER, TreeTagger, RNNTagger, SoMeWeTa, Surrogate Generation, NeuroNER, Digilib, LemmaGen, MontyLingua, Morfette, CST's Lemmatiser, acdh-spacytei, xsl-tokenizer, ANNIS

**Themen:**

Einführung, Natural Language Processing

**Zitiervorschlag:**

Bleier, Roman. 2021. NLP. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.145>