

Natural Language Processing

Magdalena Wolska Matti Wiegmann Henning Wachsmuth
Benno Stein Martin Potthast Matthias Hagen

Contents

- I. Introduction
- II. Corpora
- III. Text Models
- IV. Words
- V. Syntax
- VI. Semantics
- VII. Discourse
- VIII. Speech
- IX. Ethics, Fairness, and Bias
- X. NLP Applications

Objectives

- ❑ understand natural language phenomena
- ❑ understand and explain basic concepts of natural language processing
- ❑ be able to implement and apply algorithms to process natural language
- ❑ be able to compare and combine approaches to solve language problems
- ❑ be able to self-educate

Related Fields

1. Linguistics [paradigms, models]
2. Statistics
3. Psychology
4. Machine Learning [methods, algorithms]
5. Data Mining
6. Information Retrieval
7. Knowledge Processing
8. Text-to-Speech Systems [applications]
9. Dialog Systems
10. Conversational Systems

Literature

Natural Language Processing:

- ❑ D. Jurafsky, J. H. Martin.
Speech and Language Processing
Prentice Hall 2008. (3rd edition draft)
- ❑ C. D. Manning, H. Schütze.
Foundations of Statistical Natural Language Processing
MIT Press 1999.
- ❑ J. Eisenstein
Natural Language Processing
MIT Press 2018. (november draft)

Literature

Top-tier natural language processing conferences:

- ❑ AAACL.
Conference of the Asian Chapter of the ACL.
- ❑ ACL. (the only A* NLP conference)
Annual Meeting of the Association for Computational Linguistics.
- ❑ COLING. (linguistic focus)
International Conference on Computational Linguistics.
- ❑ CoNLL.
International Conference on Natural Language Learning.
- ❑ EACL.
Conference of the European Chapter of the Association for Computational Linguistics.
- ❑ EMNLP. (empirical focus, probably the second best NLP conference)
Conference on Empirical Methods in Natural Language Processing.
- ❑ NAACL. Conference of the North-American Chapter of the ACL.

Literature

Other relevant natural language processing conferences:

- ❑ CICLing.
International Conference on Computational Linguistics and Intelligent Text Processing.
- ❑ IJCNLP.
International Joint Conference on Natural Language Processing.
- ❑ INLG. (generation focus)
International Conference on Natural Language Generation.

Conferences from related fields:

- ❑ SIGIR, ECIR, ICTIR, AIRS, CHIIR, TREC, CLEF
- ❑ CIKM, WSDM, WWW, SPIRE

Top-tier Journals:

- ❑ ACM TOIT. toit.acm.org
- ❑ Computational Linguistics. www.mitpressjournals.org/loi/coli
- ❑ TACL. Transactions of the ACL. transacl.org

Software

Annotation Software:

- ❑ *Prodigy by ExplosionAI*
Closed Source, focus on active learning while annotating
prodi.gy
- ❑ *Label Studio by Heartex*
Open Source, very flexible
labelstud.io
- ❑ *Doccano by Hiroki Nakayama*
Open Source, limited functionality but easy to use
doccano.herokuapp.com

NLP Toolkits:

- ❑ *spaCy by ExplosionAI*
Open Source, fast, flexible, good performance, go-to toolkit
spacy.io
- ❑ *Stanza by Stanford NLP Group*
Open Source, focus on SotA performance
stanfordnlp.github.io/stanza

Software

Algorithm Collections:

- ❑ Natural Language Toolkit *by NLTK Project*
Open Source, large collection of basic algorithms
nltk.org
- ❑ Stanford NLP Software *by Stanford NLP Group*
nlp.stanford.edu/software/

Machine Learning for NLP:

- ❑ scikit-learn *by the sklearn community*
Open Source, many utilities for text modeling and transformation
scikit-learn.org/
- ❑ HuggingFace *by Huggingface*
Open Source, SotA models for many advanced NLP tasks
huggingface.co
- ❑ Flair *by Humboldt University Machine Learning Group*
Open Source, focus on embeddings and sequence tagging tasks
github.com/flairNLP/flair