Natural Language Processing

András Simonyi

2021

Lectures

Introduction

1. Introduction: what is NLP, related fields, typical applications, central themes
2. Overview of Computational Linguistics and the components of the traditional NLP pipeline: tokenization and sentence segmentation, POS tagging and morphological analysis, syntax, semantics, pragmatics

Basic methods

1. Tokenization: regexes, normalization and edit distance, subword tokenization methods
2. N-gram based language modeling
3. Classification and sequence tagging: Sentiment analysis, POS-tagging, morphology, NER
4. Dependency parsing
5. Lexical semantics and LSA

Neural methods

1. Word2vec and neural embeddings
2. Recurrent Neural Networks and RNN-based Language Models
3. Machine Translation, Seq2Seq and Attention
4. Contextual word representations and fine tuning: BERT and co.