university of groningen

Structured Prediction for Named Entity Recognition

Joachim Daiber & Carmen Klaussner

Information Science
University of Groningen

24th. October, 2012

university of groningen

Motivation

university of groningen

Outline

university of groningen

Named Entity Recognition (NER)

The task of **Named Entity Recognition** divides into two subtasks:

- Named Entity Detection
- Named Entity Classification

university of groningen

Entity Classes in NER

- Person Names: `John Bateman`
- Organisations: `Lavazza`
- Locations: `France, Bristol`
- Time Expressions: `26 May, 2009`
- Monetary Values: `$ 500`
- Percentages and Addresses: `69%/www.uni-bremen.de`

university of groningen

Approaches to NER

1. linguistic grammar-based techniques
2. statistical models

1. ⇒ hand-crafted rules may obtain slightly better precision, at cost of low recall and extensive work by computational linguists
2. ⇒ Statistical NER systems require large amount of manually annotated training data

university of groningen

university of groningen

Issues for NER

Ambiguity

university of groningen

Training and Test Data

- CoNLL 2002/2003
- Languages: English, German, Spanish & Dutch

university of groningen

university of groningen

References I


Xavier Carreras.
Learning Structured Predictors.
Lecture at Lisbon Machine Learning School 2012, http://lxmls.it.pt/strlearn.pdf, 2012.


Michael Collins.
Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms.
In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 1–8, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

university of groningen

References II


David Nadeau and Satoshi Sekine.
A survey of named entity recognition and classification.
Linguisticae Investigationes, 30(1):3–26, January 2007.
Publisher: John Benjamins Publishing Company.


Erik F. Tjong Kim Sang and Fien De Meulder.
Introduction to the CoNLL-2003 shared task: language-independent named entity recognition.
In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4*, CONLL '03, pages 142–147, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.