

From IE to ‘Deep’ Semantic Parsing

– Einführung und Organisation

Prof. Dr. Anette Frank
Institut für Computerlinguistik
Universität Heidelberg

24. Oktober 2013

Organisatorisches

- **Seminartermine**
 - Do: 11:15 – 12:45 Uhr, HS 5 INF 327
- **Studiengang:**
 - MA: SS-CL, SS-TAC (ggf. auch: SS-FAL)
 - NBA/BA-2010: AS-CL
 - 8 LP

Organisatorisches

- **Leistungsnachweis**
 - Regelmäßige, aktive Teilnahme
 - Lektüre der diskutierten Papiere
 - Referat
 - Ausarbeitung als Hausarbeit oder Projekt

Aktive Teilnahme

- Fragen
- Diskussionsbeiträge
- Freiwillige Beiträge und Stoffsammlungen für alle
 - Folienmaterial
 - Übersichten zu
 - Datensätze
 - Evaluationsmaße
 - Software/Tools
 - Verwandte/konkurrierende Ansätze

Austausch über Mailingliste

Referate

- Konzentration auf
 - **Problem** das angegangen wird / **Schwächen** bisheriger Ansätze
 - **Methode/Ansatz**
 - **Evaluierung & Performanz** (aussagekräftig?)
 - **Bewertung und Interpretation** der Resultate
 - **Impact:** Neuheit, Generalisierbarkeit, Perspektiven
 - Dabei: notwendige Hintergrundinformationen bereitstellen
- Abschließende **Würdigung**
 - Was gefällt Ihnen, was nicht?
 - Bezug/Vergleich zu anderen Ansätzen
 - Take-home message

Hausarbeit oder Projekt(skizze)

Hausarbeit

- Ausarbeitung des Referats
- Incl. Bewertung und Einordnung im Forschungskontext

Projektskizze

- Eine gut(!) ausgearbeitete Skizze für ein eigenes Projekt auf Basis etablierter Datensätze
- Kann als Basis für eine Abschlussarbeit dienen

Projekt

- Beschränkung auf einen fokussierten Aspekt

Themenüberblick

**From Information Extraction to
'Deep' Semantic Parsing**

Ziele und Fragestellungen

Informationsextraktion

- Derzeit dominiert durch unüberwachte **web-basierte 'OpenIE'** Ansätze
- Konzentriert auf *redundanzorientierte* Verfahren und hochfrequente Daten zu bekannten Entitäten und Relationen (DBpedia, Freebase)
- **Desiderat:**
IE auf Basis einzelner Texte oder kleiner Dokumentmengen
aka. (Deep) Semantic Parsing

Ziel des Seminars

- Erarbeitung des **Forschungsstands**, **Analyse** etablierter und neuer Techniken mit **Perspektive** für 'Deep' Semantic Parsing

Seminarplan

- State of the Art OpenIE
- Analyse: shallow vs. deep; Limitations of web-scale OpenIE
- Distant Supervision
- Semantic Methods (and more Semantic Methods) in RE
- Von Relationen zu Events und Parsing
- Events in Context
- Event Schema / Frame Induction
- Cross-document Event Alignment and Search
- IE in Biomedicine & Multi-document Summarization

State of the Art OpenIE

KnowItAll, TextRunner, Reverb, OLLIE ..

- ❑ **Fader et al. 2011:** Identifying Relations for Open Information Extraction. EMNLP.
- ❑ **Etzioni et al. 2011:** Open Information Extraction: The Second Generation.
- ❑ (Bkgr.) **Banko et al. 2007:** Open Information Extraction from the Web, IJCAI.

Analysis of State of the Art

Shallow vs. Deep Analysis in IE

- **Mesquita et al. 2013:** Effectiveness and Efficiency of Open Relation Extraction, EMNLP.

Limits of web-scale OpenIE: from Web-based IE to Textual Analysis

- **Weikum et al. 2012:** Big Data Methods for Computational Linguistics, IEEE.

Distant Supervision

- ❑ **Mintz et al. 2009:** Distant supervision for relation extraction without labeled data
- ❑ **Zhang et al. 2013:** Towards Accurate Distant Supervision for Relational Facts Extraction

Semantic Methods in Relation Extraction

- ❑ **Yao et al. 2012:** Unsupervised Relation Discovery with Sense Disambiguation, ACL 2012.
- ❑ **Chan & Roth 2011:** Exploiting Syntactico–Semantic Structures for Relation Extraction, ACL 2011.
- ❑ (alt.) **Plank & Moschitti (2013):** Embedding Semantic Similarity in Tree Kernels for Domain Adaptation of Relation Extraction, ACL 2013

More Semantic Methods in Relation Extraction

- ❑ **Lao et al. 2012:** Reading The Web with Learned Syntactic-Semantic Inference Rules, EMNLP 2012
- ❑ **Nakashole et al. 2012:** PATTY : A Taxonomy of Relational Patterns with Semantic Types, EMNLP 2012.

From Relations to Events and Parsing

- ❑ **McClosky et al. 2011:** Event Extraction as Dependency Parsing. ACL 2011
- ❑ **Goldberg & Orwant 2012:** A Dataset of Syntactic-Ngrams over Time from a Very Large Corpus of English Books.
*SEM 2013.
- ❑ (opt.) **Singh et al. 2013:** Joint Inference of Relations, Entities and Coreference.

Events in Context

- ❑ **Gerber & Chai 2010 / 2012**: Semantic Role Labeling of Implicit Arguments for Nominal Predicates. ACL/Comp.Ling.
- ❑ **Chambers & Jurafsky 2009**: Unsupervised Learning of Narrative Schemas and their Participants. ACL.
- ❑ (Bkgr.) **Silberer & Frank 2012**: Casting Implicit Role Linking as an Anaphora Resolution Task. *SEM.

Unsupervised Event Schema/Frame Induction

- ❑ **Cheung et al. 2013:** Probabilistic Frame Induction, NAACL 2013.
- ❑ **Chambers 2013:** Event Schema Induction with a Probabilistic Entity-Driven Model. EMNLP 2013.

Event Coreference Resolution

- ❑ **Bejan & Harabagiu 2010:** Unsupervised Event Coreference Resolution with Rich Linguistic Features. ACL 2010
- ❑ **Lee et al. 2012:** Joint entity and event coreference resolution across documents. EMNLP 2012

Cross-document Event Alignment and Search

- ❑ **Roth & Frank 2012:** Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering. EMNLP 2012.
- ❑ **Glavas & Snajder 2013:** Recognizing Identical Events with Graph Kernels. ACL 2013.

Sessions 11 and 12

23.01./30.1.

Applications I: Event Extraction in Biomedicine

- ❑ **McClosky et al. 2012:** Combining joint models for biomedical event extraction]]. BMC Bioinformatics.

Applications II: Multi-Document Summarization

- ❑ **Chi et al. 2013:** Towards Robust Abstractive Multi-Document Summarization : A Caseframe Analysis of Centrality and Domain, ACL.

Wrap-up und Diskussion

Wrap-up

- Diskussion der 'Take Home Messages'
- Stärken/Schwächen-Analyse einzelner Ansätze
- Perspektiven für weitere Entwicklungen
- Sichtung gesammelter Materialien

Gestaltung von Hausarbeiten und Projektentwürfe

- Abgabe: 4.4.2014