

Seminar/Proseminar im SoSe 2019

Natural Language Processing

Thomas Ruprecht

TU Dresden
Fakultät Informatik
Institut für Theoretische Informatik
Professur für Grundlagen der Programmierung

4. April 2019

(Pro-)Seminar?

Wir beschäftigen uns gemeinsam mit Literatur zu einem Themengebiet (*NLP*):

- ▶ jede/r Studierende
 - ▶ liest und versteht ein (eventuell auch zwei) Paper
 - ▶ schreibt eine Seminararbeit von 12-15 Seiten (Seminar)
 - ▶ bereitet ein Handout (1-2 Seiten) vor (Proseminar)
 - ▶ hält einen Vortrag von 30-35 Minuten am Semesterende
- ▶ Diskussion im Anschluss an jeden Vortrag

Zielgruppen

engagierte Studenten folgender Studiengänge

- ▶ Bachelor Informatik: Module INF-B-510 (Proseminar), INF-B-520 (Proseminar), INF-B-610 (Proseminar)
- ▶ Master Informatik: Modul INF-AQUA (Seminar)
- ▶ Diplom Informatik: Module INF-D-520 (Proseminar), INF-D-940 (Seminar)
- ▶ Diplom Informatik (Studienordnung 2004 oder älter): Hauptseminar, Fachgebiet Theorie der Programmierung

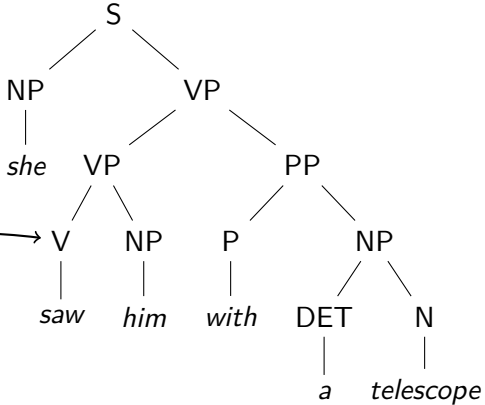
Voraussetzungen für die Teilnahme

- ▶ Bachelor Informatik: Modul INF-B-270 (Formale Systeme)
- ▶ Master Informatik: keine; Grundlagenwissen über die Themengebiete wird empfohlen
- ▶ Diplom Informatik: Modul INF-B-270 (Formale Systeme)
- ▶ Diplom Informatik (Studienordnung 2004 oder älter): Grundlagen der Theoretischen Informatik

Bewertungskriterien

- ▶ Verständnis des bearbeiteten Themas, Fähigkeit den Inhalt in eigenen Worten zu erklären und zusammenzufassen
- ▶ anschaulicher Vortrag: Auswahl und Präsentation der Inhalte, geeignete Medien, Rhetorik
- ▶ Qualität der schriftlichen Arbeit/des Handouts
- ▶ Anwesenheit bei allen Vorträgen, aktive Teilnahme an den Diskussionen
- ▶ fristgerecht und eigenständig Termine mit den Betreuern verabreden (mindestens 1 Woche im Voraus) sowie die geforderten Materialien abgeben

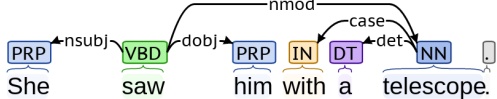
Syntactic Parsing



phrase structure
(constituent)
parsing

She saw him with a telescope.

dependency
parsing



Thema 1: The smallest grammar problem [Cha+05]

- ▶ Problemstellung:
 - ▶ gegebenes Wort
 - ▶ gesucht: kleinste CFG, die *nur das Wort* generiert
 - ▶ Größe: Symbole auf rechten Regelseiten
- ▶ Anwendung:
 - ▶ Kompression von Zeichenketten
 - ▶ Mustererkennung in DNA, natürlichsprachigen Texten
- ▶ Inhalt des Papers:
 - ▶ Schwere des Problems analysiert
 - ▶ Algorithmen zur Lösungsapproximation

a rose is a rose is a rose



$$\left. \begin{array}{l} S \rightarrow BBA \\ A \rightarrow a_{\sqcup} \text{rose} \\ B \rightarrow A_{\sqcup} \text{is}_{\sqcup} \end{array} \right\} \text{Größe: 14}$$

Thema 2: An efficient best-trees algorithm for weighted tree automata over the tropical semiring [BDZ15]

- ▶ Problemstellung:
 - ▶ Die Bäume mit dem höchsten Gewichte einer gewichteten Baumsprache sollen bestimmt werden.
- ▶ Anwendung:
 - ▶ Decoding bei Machine Translation, Parsing
- ▶ Inhalt des Papers:
 - ▶ Algorithmus der die n -besten Bäume im Fall von WTA über dem tropischen Semirings berechnet
 - ▶ Komplexitätsanalyse

Thema 3: Guided parsing of range concatenation languages [Bar+01]

- ▶ Problemstellung:
 - ▶ Parsing von *range concatenation languages* ist polynomiell, aber aufwendig
- ▶ Anwendung:
 - ▶ diskontinuierliches Konstituentenparsing
- ▶ Inhalt des Papers:
 - ▶ Parse-Guides helfen bei nicht-deterministischen Entscheidungen
 - ▶ Konstruktion von Guides aus Parses mit weniger-komplexen Grammatik
 - ▶ experimentelle Untersuchung für Konstituentenparsing

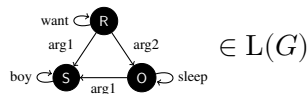
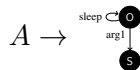
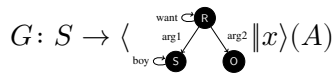
Thema 4: Coarse-to-fine parsing for expressive grammar formalisms [TKG17]

- ▶ Problemstellung:
 - ▶ langsames Parsing mit großen statistischen Grammatiken
 - ▶ weniger-komplexe Modelle durch Zusammenfassen von Nichtterminalen
- ▶ Anwendung:
 - ▶ (diskontinuierliches) Konstituentenparsing
- ▶ Inhalt des Papers:
 - ▶ *coarse-to-fine Parsing* mittels groben Klassen von Nichtterminalen
 - ▶ experimentelle Evaluation für verschiedene Grammatikformalismen

$$G: S \rightarrow NP VP$$
$$VP \rightarrow VP NP + to VP NP + loves$$
$$NP \rightarrow john + mary$$
$$\downarrow C(x) \begin{cases} S & \text{if } x = S \\ HP & \text{if } x \in \{NP, VP\} \\ \dots \end{cases}$$
$$G_C: S \rightarrow HP HP$$
$$HP \rightarrow HP HP + to HP HP + loves$$
$$+ john + mary$$

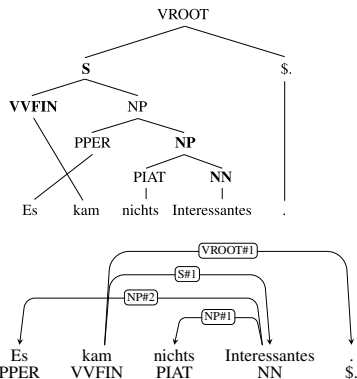
Thema 5: Graph parsing with s-grammars [GKT15]

- ▶ Problemstellung:
 - ▶ Synchroner Graph-String-Grammatiken zur Überführung von Sätzen in AMR
 - ▶ Generierung von Sätzen aus AMR erfordert Parsing von Graphen
- ▶ Anwendung
 - ▶ Machine Translation (Generierung in Zielsprache)
 - ▶ Training von synchronen Grammatiken
- ▶ Inhalt des Papers:
 - ▶ Parsing mit dieser Grammatik über Graphen
 - ▶ experimentelle Evaluation



Thema 6: Parsing as reduction [FM15]

- ▶ Problemstellung:
 - ▶ Konstituenten-Parsing durch große statistische Modelle ist langsam
 - ▶ Dependency-Parsing i.d.R. weniger aufwendig
- ▶ Anwendung
 - ▶ (diskontinuierliches) Konstituenten-Parsing
- ▶ Inhalt des Papers:
 - ▶ Isomorphismus zw. speziellen Dependency-Parses und Konstituenten-Parses
 - ▶ experimentelle Evaluation mit natürlichsprachigen Korpora



Thema 7: Weighting finite state transducer with neural context [RCE16]

- ▶ Problemstellung
 - ▶ Zeichenreihen sollen in andere Zeichenreihen übersetzt werden
 - ▶ FST erlauben Modellierung struktureller Einschränkungen bei lokalem Kontext
 - ▶ Neuronale Modelle erlauben automatische Features mit globalem Kontext
- ▶ Anwendung
 - ▶ Beugung von Worten, Wortstammbestimmung
- ▶ Inhalt des Papers
 - ▶ Verbindung von WFST mit neuronalen Netzen
 - ▶ Modell, Trainings- und Dekodieralgorithmen und Experimente

Themenübersicht

- ▶ The smallest grammar problem [Cha+05]
- ▶ An efficient best-trees algorithm for weighted tree automata over the tropical semiring [BDZ15]
- ▶ Guided parsing of range concatenation languages [Bar+01]
- ▶ Coarse-to-fine parsing for expressive grammar formalisms [TKG17]
- ▶ Graph parsing with s-grammars [GKT15]
- ▶ Parsing as reduction [FM15]
- ▶ Weighting finite state transducer with neural context [RCE16]

Zeitplan

- 04. April erstes Treffen und Themenvergabe im Raum APB/3027 (Fakultät Informatik)
- bis zum 26. April Termine mit Betreuern (rechtzeitig verabreden!); Ziel: in der Lage sein, dem Betreuer das Problem und die beschriebenen Ansätze zu skizzieren und auf Nachfragen zu reagieren; Fragen an den Betreuer müssen ganz konkret formuliert werden.
- bis zum 17. Mai Vorabversion der Seminararbeit abgeben, Termine zur Besprechung machen
- bis zum 31. Mai fertige Seminararbeit abgeben
- bis zum 14. Juni Vorabversion der Vortragsmaterialien abgeben, Termine zur Besprechung machen
- bis zum 28. Juni fertige Vortragsmaterialien abgeben
- 05. Juli Vorträge im Raum APB/3027

Informationen

Den Link zu jedem Paper sowie weitere Hilfestellungen findet ihr unter:
<https://www.orchid.inf.tu-dresden.de/teaching/2019ss/seminar/>

References I



François Barthélemy u. a. “Guided parsing of range concatenation languages”. In: *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics. 2001, S. 42–49.



Johanna Björklund, Frank Drewes und Niklas Zechner. “An efficient best-trees algorithm for weighted tree automata over the tropical semiring”. In: *International Conference on Language and Automata Theory and Applications*. Springer. 2015, S. 97–108.




Moses Charikar u. a. “The smallest grammar problem”. In: *IEEE Transactions on Information Theory* 51.7 (2005), S. 2554–2576.




Daniel Fernández-González und André FT Martins. “Parsing as reduction”. In: *arXiv preprint arXiv:1503.00030* (2015).

References II



Jonas Groschwitz, Alexander Koller und Christoph Teichmann. “Graph parsing with s-graph grammars”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Bd. 1. 2015, S. 1481–1490.



Pushpendre Rastogi, Ryan Cotterell und Jason Eisner. “Weighting finite-state transductions with neural context”. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016, S. 623–633.



Christoph Teichmann, Alexander Koller und Jonas Groschwitz. “Coarse-to-fine parsing for expressive grammar formalisms”. In: *Proceedings of the 15th International Conference on Parsing Technologies*. 2017, S. 122–127.