

20.07.2013

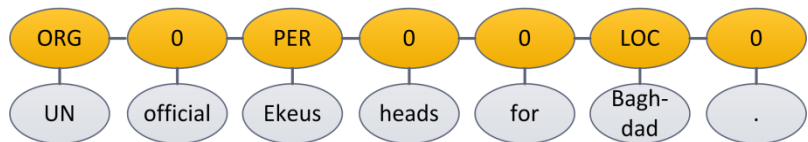
Master-/Diplomarbeit

„*Relational Markov Network* als grafisches Modell zur Eigennamenerkennung und -disambiguierung“

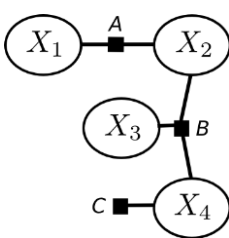
zu vergeben!

Um was geht es?

Bei der Eigennamenerkennung (*named entity recognition and classification (NERC)*) geht es darum, in



einem Text Erwähnungen von Personen, Organisationen, Orten usw. automatisch zu erkennen und zu klassifizieren. Mit dieser Art von Informationsextraktion können wichtige Aussagen über den Text getroffen werden. Dabei ist es wichtig, Mehrdeutigkeiten richtig aufzulösen (Disambiguierung).



In der Vergangenheit sind sehr viele Software-Tools zur Eigennamenerkennung entwickelt worden. In der Regel sind diese jedoch nur auf alltägliche Texte anwendbar, nicht jedoch auf domänenspezifische, wenn es etwa um die Beschreibung von Technologien geht. Mathematisch-statistische grafische Modelle (*graphical models*) [1] zur Eigennamenerkennung zu verwenden, ist ein relativ neuer und vielversprechender Ansatz.

Die Abschlussarbeit stützt sich auf den Ansatz *Relational Markov Networks (RMN)*. Das Ziel der Arbeit besteht darin, diese probabilistisch-statistische Methode (ggf. anhand existierender Frameworks wie *factorie*) für o.g. Aufgabe zu implementieren. Anhand einer Evaluation ist ferner festzustellen, ob sich der Ansatz für domänenspezifische Texte eignet. Schließlich ist zu klären, wie strukturiert abgelegte Informationen in das Modell mit integriert werden können.

Was sollten Sie mitbringen?

- Gutes mathematisches Grundverständnis, insbesondere im Bereich Statistik.
- Bereitschaft, sich in komplexe, mitunter sehr mathematisch-statistische Themengebiete einzuarbeiten.
- Bereitschaft zum Programmieren.

[1] Literatur hierzu wird bereitgestellt.

Kontaktperson:
Michael Färber
michael.farber@kit.edu
Tel.: 0721/608 479 46