

17.09.2013

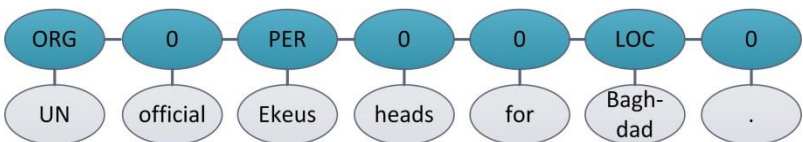
Bachelor-/Master-/Diplomarbeit

„Vergleich von Tools zur Erkennung von Eigennamen“ 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. In diesen Fällen ist es etwa schwierig, Erwähnungen von typischen Klassen wie „Person“ und „Ort“ zu erkennen oder auch neue Klassen wie „Technologie“ einzuführen.

Das Ziel der Arbeit besteht darin, einen systematischen Überblick über bestehende Tools zu gewinnen und die Tools anhand bestimmter Kriterien zu kategorisieren. Anhand einer Evaluation ist festzustellen, welche Tools sich in welchem Umfang für domänenspezifische Texte eignen. Hierzu werden Texte aus mehreren Domänen (z.B. industrielle Technologien, Medizin etc.) verwendet.

Was sollten Sie mitbringen?

- Interesse an semantischen Technologien und Text Mining (Vorkenntnisse von Vorteil, aber nicht notwendig).

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