

Graduiertenkolloquium Angewandte Informatik

Ontologie-basierte Monosemierung

Dipl.-Inform. (FH) Joachim Kleb
FZI

Von Beginn an sind wir mit dem Problem der Namensgebung konfrontiert. Zunächst sind es unsere Eltern, die unseren Namen bestimmen. Wie Eltern wissen ist dies nicht einfach und viele versuchen hierbei einen einzigartigen Namen zu wählen. Der Hintergrund ist hierbei, dass Namen nicht nur einmal vergeben werden, sondern derselbe Namen eine Vielzahl von unterschiedlichen Objekten identifizieren kann. Dieses Phänomen beschäftigt nicht nur Eltern sondern erschwert jedem von uns bereits das Erlernen der Sprache. Die einzige Hilfe liegt hierbei in der Beachtung zusätzlicher Merkmale, die es ermöglichen die verschiedenen Bedeutungen auseinanderzuhalten. Beispielsweise hilft für "Blatt" mit den Bedeutungen "Blatt Papier" und "Blatt einer Pflanze" oftmals bereits die Umgebung in der dieses Wort genannt wird, z.B. in der Nähe des Schreibtisches oder in der freien Natur.

Betrachtet man den Bereich der maschinellen Bearbeitung, sprach bereits Tim Berners-Lee vom Zwang zur Eindeutigkeit. Diese Anforderung ist offensichtlich, wenn man bedenkt, dass die maschinelle Bearbeitung auf Regeln basiert und deren eindeutige Identifikation eine grundlegende Anforderung ist. Ontologien, die in modernen Programmen als Hintergrundwissen zum Einsatz kommen, setzen eine eindeutige Identifikation der Ontologieelemente über URI's voraus und umgehen dadurch das Problem der Mehrdeutigkeit.

Wird jedoch eine Bearbeitung anhand natürlich-sprachlicher Bezeichner, z.B. aufgrund der Analyse von Nachrichtentexten etc. notwendig, so wird eine Identifikation von Einträgen der ontologischen Wissensbasis durch natürlich-sprachliche Begriffe notwendig. Allerdings, sobald verschiedene Einträge solche Identifikatoren teilen, z.B. mehrere Personen mit demselben Namen enthalten sind, tritt auch hier das Phänomen der Mehrdeutigkeit auf.

Dieser Vortrag beschäftigt sich mit der Bestimmung von Objekten in Ontologien, die mit ihrem natürlich-sprachlichen Bezeichner in einem vorliegenden Text genannt wurden und daher anhand dieses Bezeichners identifiziert werden müssen. Hierbei wird zunächst die Problematik der Mehrdeutigkeit näher betrachtet. Im Vortrag wird ein grundlegendes Verfahren zur Monosemierung vorgestellt, das eine Auflösung der Mehrdeutigkeit und somit eine eindeutige Bestimmung der korrekten Referenz für einen gegebenen Bezeichner ermöglicht. Weiterhin werden Modifikationen dieses Verfahrens vorgestellt.

Termin: Mittwoch, 14. September 2011, 15:45 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe
Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 231
(Hinweise für Besucher: www.aifb.kit.edu/Allgemeines/Besucher)

Veranstalter: Institut AIFB, Forschungsgruppe Wissenmanagement

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

Andreas Oberweis, Hartmut Schmeck, Detlef Seese, Wolffried Stucky, Rudi Studer (Org.), Stefan Tai