

Graduiertenkolloquium Angewandte Informatik

Instance Matching for Heterogeneous Structured Data

M.Sc. Yongtao Ma

AIFB

Abstract:

Structured data is abundantly available in enterprises and also largely increasing in the Web setting. Generally speaking, it can be conceived as structured descriptions of real-world entities. One main problem towards the effective usage of structured data is instance matching, where the goal is to find instance representations referring to the same real-world thing. However, the structured data on the Web is heterogeneous, e.g. type information of instances is missing or too general to be useful. Besides, the challenges that lie ahead for typical instance matching approaches also include dealing with the low-quality data and high computation complexity.

We tackle these challenges in different steps of the instance matching process. The first step is typification, in which the type semantics is derived by an unsupervised approach. The second step, blocking, aims to reduce the quadratic complexity of the instance matching process through the efficient and effective generation of match candidates. We propose an unsupervised approach to learn the most representative attributes of instances called keys, based on which two instances are considered as a match candidate if they share the same value of the key. The third step classification aims to deal with the low quality of data, for which we propose an almost-parameter-free approach for learning instance-matching rules to classify candidate instance pairs into matches and non-matches. In the last filtering step, we propose a parameter-free solution that leverages only simple Boolean functions and exploits fine-grained word-level dissimilarity evidences to further filter out the non-matches. We evaluate our approaches against the latest baselines. The results show advances beyond the state-of-the-art.

Termin: Freitag, 28. Februar 2014, 14.00 Uhr

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

Veranstalter: Institut AIFB, Forschungsgruppe Wissensmanagement

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