

Graduiertenkolloquium Angewandte Informatik

Semantic Similarity Measures and their Impact on Data-Driven Tasks over Knowledge Graphs

M. Sc. Ignacio Traverso Ribón

FZI Forschungszentrum Informatik am KIT

Precisely determining similarity values among real-world entities becomes a building block for data driven tasks, e.g., ranking, relation discovery, or integration. Semantic Web and Linked Data initiatives have promoted the publication of large semi-structured datasets in form of knowledge graphs. Knowledge graphs encode semantics that describes resources in terms of several aspects or resource characteristics, e.g., neighbors, class hierarchies or attributes. Existing similarity measures take into account these aspects in isolation, which may prevent them from determining accurate similarity values. In this talk, the relevant resource characteristics to determine accurately similarity values are identified and their impact in three data-driven tasks analyzed.

First, according to the identified characteristics, new similarity measures able to combine two or more of them are described. In total four similarity measures are presented in an evolutionary order. While the first three similarity measures combine the resource characteristics according to a human defined aggregation function, the last one makes use of the machine learning approach to determine the relevance of each resource characteristic during the computation of the similarity. Second, the suitability of each measure for real-time applications is studied means a theoretical and empirical comparison among the described similarity measures in terms of computational complexity.

Ultimately, the impact of the described similarity measures is shown in three data-driven tasks for the enhancement of knowledge graph quality: Relation discovery, dataset integration, and evolution analysis of annotation datasets. Empirical results show that accuracy of the tasks of relation discovery and dataset integration is enhanced when considering semantics encoded in semantic similarity measures. Furthermore, the annotation evolution task is also enhanced, and expressive metrics able to provide an aggregated overview of a set of annotated entities has been developed. All the improvements observed in these three tasks support the hypothesis that semantic similarity measures empower the performance of data-driven tasks.

Termin: Freitag, 19. Mai 2017, 14.00 Uhr

Ort: Kaiserstr. 89, 76133 Karlsruhe
Kollegiengebäude am Kronenplatz (Geb. 05.20), 1. OG, Raum 1C-04
(Hinweise für Besucher: www.aifb.kit.edu/web/Kontakt)

Veranstalter: Institut AIFB, Forschungsgruppe Web Science

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

A. Oberweis, H. Schmeck, Y. Sure-Vetter (Org.), J. M. Zöllner, H. Sack