Informatik-Kolloquium
Fachgruppe Informatik
Lehrstuhl für Informatik 5

RWTH
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN

EINLADUNG

Zeit:	Freitag, 9. Juli, 14.00 Uhr
Ort:	5056, Ahornstr. 55
Referent:	Prof. Dr. Avigdor Gal Technion, Haifa, Israel
Titel:	Tuning the ensemble selection process of schema matchers

Abstract:

Schema matching is the task of providing correspondences between concepts describing the meaning of data in various heterogeneous, distributed data sources. It is recognized to be one of the basic operations required by the process of data and schema integration and its outcome serves in many tasks such as targeted content delivery and view integration. Schema matching research has been going on for more than 25 years now. An interesting research topic, which was largely left untouched involves the automatic selection of schema matchers to an ensemble, a set of schema matchers. To the best of our knowledge, none of the existing algorithmic solutions offer such a selection feature. In this paper we provide a thorough investigation of this research topic. We introduce a new heuristic, Schema Matcher Boosting SMB. We show that SMB has the ability to choose among schema matchers and to tune their importance. As such, SMB introduces a new promise for schema matcher designers. Instead of trying to design a perfect schema matcher, a designer can instead focus on finding better than random schema matchers. For the effective utilization of SMB, we propose a complementary approach to the design of new schema matchers. We separate schema matchers into first-line and second-line matchers. First-line schema matchers were designed by-and-large as applications of existing works in other areas (e.g., machine learning and information retrieval) to schemata. Second-line schema matchers operate on the outcome of other schema matchers to improve their original outcome. SMB selects matcher pairs, where each pair contains a first-line matcher and a second-line matcher. We run a thorough set of experiments to analyze SMB ability to effectively choose schema matchers and show that SMB performs better than other, state-of-the-art ensemble matchers.

Es laden ein: Die Dozenten der Informatik