Chapter 1.20
Theories of Meaning in Schema Matching:
A Review

Joerg Evermann
Memorial University of Newfoundland - St. John's, Canada

ABSTRACT

Schema matching is the identification of database elements with similar meaning as preparation for subsequent database integration. Over the past 20 years, different schema-matching methods have been proposed and have been shown to be successful to various degrees. However, schema matching is an ongoing research area and the problem is not yet considered to be solved. This article reviews existing schema-matching methods from the perspective of theories of meanings drawn from philosophy and psychology. It positions existing methods, raises questions for future research based on these theories, and shows how these theories can form a firm theoretical basis as well as guide future schema-matching research.

INTRODUCTION

Database integration is an increasingly important activity. Information systems development often requires integrating existing systems and their databases. Data warehousing requires the integration of data from multiple transaction-processing systems for decision support. Electronic business between organizations requires the integration of trade partners' business data for efficient business processes. Interagency information sharing in e-government requires integration of data about citizens (Bajaj & Ram, 2003).

Database integration is a process with multiple steps, leading from the identification of the databases to the testing of the integrated system. The central step of the database integration process is

the identification of those elements in the schemata of the databases that match each other. This step is termed schema matching. Schema matching can be applied to different data representations, whether they are relational, object oriented, or XML (extensible markup language) based. Existing work on schema matching has focused primarily on relational databases. Schema matching can be applied on the conceptual level, where schema elements are entity types or attributes and relationship types, or it can be applied on the physical level, where schema elements are tables, fields, and foreign-key relationships.

Many different methods for matching database schema elements have been proposed (Batini, Lenzerini, & Navathe, 1986; Rahm & Bernstein, 2001). The partial success of the proposed methods shows that the problem is not yet completely solved.

The purpose of a review is to provide a new conceptualization of a field and raise unresolved research issues (Webster & Watson, 2002). Consequently, this research review article has three aims. First, it proposes a new theoretical foundation for schema matching. Second, it uses this theoretical foundation to examine and position existing schema-matching methods. Third, based on this theoretical foundation and the positioning of existing schema-matching methods, it points out possible avenues for future research in schema matching.

A solid theoretical basis that informs the research in a discipline is seen as essential to the discipline's maturation (Weber, 1997, 2003). In this sense, the proposed theoretical foundation for schema matching is similar to the role that ontology plays in the wider discipline of information systems development (Wand & Weber, 1990). Wand and Weber's ontological foundation provided a framework with which to understand existing IS development research and point out future research directions (Wand & Weber, 2002). We hope this article will make a similar contribution to schema-matching research.

The remainder of the article is structured as follows. The second section proposes a new conceptualization of schema-matching research based on theories of meaning. This conceptualization is followed by an introduction to existing schema-matching methods. Then, the article introduces specific theories of meaning and analyzes their use in schema matching. It also presents research questions raised in connection with individual theories of meaning. Next, we position schema-matching methods into the theoretical foundation and offer suggestions for further empirical studies, before concluding the article.

SCHEMA MATCHING AND MEANING

Consider a situation that requires the integration of production management data with marketing data for a decision support system. The marketing database contains information about products and articles, while the production database contains information about parts and components (Figure 1). How does a database integrator decide which of these elements match each other? What criteria does the database integrator apply? We propose that the database integrator makes use of an, often implicitly held, theory of meaning. In the example data model in Figure 1, the schema elements Product and Article are matched if the database integrator decides they have the same meaning. Similarly, the elements ProductID and SerialNum are matched if they have the same meaning to the database integrator. The theory of meaning held by the database integrator determines whether they have the same meaning. Hence, the theories of meaning held by database integrators, or humans in general, are relevant to schema matching.

Knowing that database integrators hold a theory of meaning is not yet helpful to the concrete schema-matching problem at hand. How does the database integrator decide whether Product and

Related Content

Techniques for Sampling Online Text-Based Data Sets
Lynne M. Webband Yuanxin Wang (2014). *Big Data Management, Technologies, and Applications (pp. 95-114).*
www.irma-international.org/chapter/techniques-for-sampling-online-text-based-data-sets/85452

Business-to-Business (B2B) Integration
Christoph Bussler (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends (pp. 828-836).*
www.irma-international.org/chapter/business-business-b2b-integration/20768

Evaluating XML-Extended OLAP Queries Based on Physical Algebra
Xuepeng Yinand Torben Bach Pedersen (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 2510-2542).*
www.irma-international.org/chapter/evaluating-xml-extended-olap-queries/8049

Towards Autonomic Workload Management in DBMSs
Baoning Niu, Patrick Martinand Wendy Powley (2009). *Journal of Database Management (pp. 1-17).*
www.irma-international.org/article/towards-autonomic-workload-management-dbmss/4122

Cognitive Neuroscience in Information Systems Research
Yeli Zhaoand Keng Siau (2016). *Journal of Database Management (pp. 58-73).*
www.irma-international.org/article/cognitive-neuroscience-in-information-systems-research/160351