2004

Early Work in Database Research on Schema Mapping/Merging/Transformation, Semantic Heterogeneity, and Use of Ontology and Description Logics for Schematic and Semantic Integration

Amit P. Sheth
Wright State University - Main Campus, amit.sheth@wright.edu

A. Sheth, *Early work in database research on schema mapping/merging/transformation, semantic heterogeneity, and use of ontology and description logics for schematic and semantic integration, Dagstugh Seminar on* Semantic Interoperability and Integration, September 2004, *http://www.dagstuhl.de/04391/Materials/*

Official copy at:
http://www.dagstuhl.de/files/Proceedings/04/04391/04391.SWM2.Other.htm

Here are the DB publications specifically related to matching and types of mappings, or use of ontologies and DL in DB/IS research, that were mentioned in the evening session on September 21, 2004. These are by no means complete or representative but historical and biased sample, focusing on works that were primarily performed 10 to 15 years ago. There are quite a few new relevant efforts that are more recent that have again looked at the mapping/merging issues (most notably, the model management work Phil Bernstein has spearheaded) that are easy to find and are not listed here. .Also, the multidatabase (loosely coupled) and federated database (tightly coupled) information integration system architectures were replaced by LAV/GLAV, and these are not discussed here either. If anyone wants to submit additions, I would be happy to add.

1981: "semantic relativism" discussed by Michael Brodie—I think it appeared in (please verify) Michael Brodie: On Modelling Behavioural Semantics of Databases VLDB 1981: 32-42. Felix Saltor talked about concept of three "worlds," I think he called them model world, representation world and real world (can't find documentary proof right now). In my view this was a way to recognize existence of (and refer to, selectively) "open world" while living in "closed world".

1984: Litwin introduces multidatabases in ealy 1980s with his MRDSM system and this paper: Witold Litwin: MALPHA: A Relational Multidatabase Manipulation Language. ICDE 1984: 86-93

1985: After a report in 1981, McLeod's federated database architecture takes off with this: Dennis Heimbigner, Dennis McLeod: A Federated Architecture for Information Management. ACM Trans. Inf. Syst. 3(3): 253-278 (1985)

1986: Schema Integration survey that discusses a five step methodology, that encompasses the more simplified map/merge view of recent days. This methodology was follwed by most subsequent schema integration approached in the subsequent 5 or more years. Batini, Navathe, Lenzerini, "A comparative analysis of methodologies for database schema integration. http://portal.acm.org/citation.cfm?id=27634

1986: Concept of semantic and dynamic attribute capturing types of mapping involing four types of mapping/translation techniques ranging from syntactic, table, functional and program based mappings: Witold Litwin, Abdelaziz Abdellatif: Multidatabase Interoperability. IEEE Computer 19(12): 10-18 (1986).

1987: Probably the first tutorial on heterogeneous database integration by Sheth at ICDE (IEEE Data Engg)—the Sheth/Larson 1990 paper document this later. Further added to semantic heterogeneity issues Litwin introduced earlier.

1988: Second reported Schema Integration tool (first in US); first reported tool was probably SIS in Spain; semi-automated matching and automated merging: Amit P.

Sheth, James A. Larson, Aloysius Cornelio, Shamkant B. Navathe: A Tool for Integrating Conceptual Schemas and User Views. ICDE 1988: 176-183. http://portal.acm.org/citation.cfm?id=653395 or Digital copy: http://lsdis.cs.uga.edu/lib/download/SLCN88.pdf

1989: urge to use"real world semantics" in schema integration so as to use semantic in matching/integration that is not considered by syntactic approach of Larson et al 1987: A. Sheth and S. Gala, "Attribute Relationships: An Impediment in Automating Schema Integration," Proceedings of the Workshop on Heterogeneous Database System, December 1989.

1989: An early view on matching, but was later recognized as focusing only on syntactic and some structural matching, and was found to miss out on semantic considerations. James A. Larson, Shamkant B. Navathe, Ramez Elmasri: A Theory of Attribute Equivalence in Databases with Application to Schema Integration. IEEE Trans. Software Eng. 15(4): 449-463 http://citeseer.ist.psu.edu/context/69486/0

1989: Discussion of more than first order representation to define schematic heterogeneity: R. Krishnamurthy,W.Litwin, and W. Kent. *Language features for interoperability of databases with schematic discrepancies*. In J. Clifford and R. King, editors, Proceedings of ACM SIGMOD Conference, pages 40--49. ACM, May1991. http://citeseer.ist.psu.edu/context/69486/0

1990: Comprehensive taxonomy of issues/requirements, architectures, methodology, analysis and survey of federated databases (encompassing Litwin and McLeod's related visions of loosely coupled and tightly coupled federated architectures, respectively) with key dimensions of distribution, heterogeneity and autonomy and their impact; classification of system to semantic heterogeneity, various key issues including schema and data translation/mapping/integration (merging), including new (at that time) look at methodology. A. P. Sheth and J. A. Larson. *Federated Database Systems for Managing Distributed*, Heterogeneous, and Autonomous Databases. ACM Computing Surveys, 22(3):183--236, 1990. http://citeseer.ist.psu.edu/context/4393/0 Digital Copy in ACM Protal r LSDIS library.

1991: A classification of heterogeneity especially for relational DB schemas: Kim & Seo,"Classifying Schematic and Data Heterogeneity in Multidatabase Systems" http://portal.acm.org/citation.cfm?id=125845

1991: Early attempt to use Description for Schema Integration further refined in 1993 paper. Ashoka Savasere, Amit P. Sheth, Sunit K. Gala, Shamkant B. Navathe, H. Markus: On Applying Classification to Schema Integration. RIDE-IMS 1991: 258-261 1991: Discussion of an experiment to integrate two large real world schemas using BERDI tool at Bellcore: http://lsdis.cs.uga.edu/lib/download/Issues_in_Schema_Integration.pdf

1992: Highly influential mediator architecture was introduced by Gio Wiederhold which was used for second wave of information integration systems (after the earlier wave of multidatabase/federated database systems). This is also the time when ontology was reintroduced to information systems/AI research by Gruber, so number of mediator-based systems used an ontology. G. Wiederhold, Mediators in the Architecture of Future Information Systems IEEE Computer 25(3) March, 1992. Digital copy at ACM Portal.

1992: A fairly comprehensive list of heterogeneity not limited to relational model; and recognizes schematic vs semantic differences/heterogeneity. Introduces Semantic Proximity/Feature set based representation of types of more complex/inexact mappings (consisting of domain (later replaced by domain ontology), model/abstraction, context and state).
Amit P. Sheth, Vipul Kashyap: So Far (Schematically) yet So Near (Semantically). DS-5 1992: 283-312. Digital Copy in LSDIS library.

1993: Support for structural/epistemological relationships/mappings (equality, inclusion/subtype, disjoint, and inconsistent relationships); use the notion of model semantics and real-world semantics to "reason" about semantic relationships; Amit P. Sheth, Sunit K. Gala, Shamkant B. Navathe: On Automatic Reasoning for Schema Integration. Int. J. Cooperative Inf. Syst. 2(1): 23-50 (1993). Digital Copy at: http://lsdis.cs.uga.edu/library/download/IJCIS-On_Automatic_Reasoning.pdf

1993: BERDI: A practically is usable Schema Integration tool that was internally evaluated and used in Bellcore. http://lsdis.cs.uga.edu/lib/download/SM94.pdf

1993/1994: Use of common ontology as a representation of the "real world" for integration of database of schemas in the "model world.". Use of semantic proximity to represent schema correspondences (mapping) with structural and semantic mapping components. Need to deal with multiple ontologies and initial approaches for that, including represntation of query context and information resource concept, and intermediation. Vipul Kashyap, Amit P. Sheth: Semantics-Based Information Brokering. CIKM 1994: 363-370 [The 1993 TR version has additional discussion about respurce discovery on Web, etc. "Semantics_based Information Brokering A step towards realizing the Infocosm"]. For CIKM version, click. (TR available by email).

1993: Around the same time, the ISI's SIM's system (Arens & Knoblock) also worked on use of ontology for information integration, which might be the earliest prototype of that type. See: http://portal.acm.org/citation.cfm?id=171566

1996: Significant refinement of So Far yet So Near: esp. relevant to more comprehensive type of mapping called semantic proximity/distance/similarity, taxonomy of schematic conflicts, modeling and reasoning of context with DL, and refinement to types of heterogeneity. Kashyap and Sheth, Semantic and schematic similarities between database objects: a context-based approach, VLDB Journal. See http://citeseer.ist.psu.edu/48524.html:

1996:Multi-ontology query processing, inter-ontological relationships (selected mappings/relationships between concepts in different ontologies, not merging of ontologies): Mena et al., OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies. http://citeseer.ist.psu.edu/mena96observer.html (extended version: http://portal.acm.org/citation.cfm?id=346215).

1999: Detailed example of semantic heterogeneity from financial domain: Goh, C.H., et al., *Context Interchange: New Features and Formalisms for the Intelligent Integration of Information*. ACM Transactions on Information Systems, 1999. 17(3): p. 270-293.

2000: On query reformulation between query specified using one ontology and information source described using another; more interesting an attempt to define “information loss” (definitions based on intensional and extensional aspects): Mena et al. Imprecise Answers In Distributed Environments: Estimation Of Information Loss For Multi-Ontology Based Query Processing.
http://www.worldscinet.com/journals/ijcis/09/0904/S0218843000000193.html

2001: Meersman’s lucid discussion on ontologies and DB research: Ontologies and Databases: More than a Fleeting Resemblance.
http://citeseer.ist.psu.edu/meersman01ontologies.html:.