Ontology Alignment: An annotated Bibliography

Natasha Noy[1], Heiner Stuckenschmidt[2]

[1]Stanford University
[2]Vrije Universiteit Amsterdam

Ontology mapping, alignment, and translation has been an active research component of the general research on semantic integration and interoperability. In our talk, we gave our own classification of different topics in this research. We talked about types of heterogeneity between ontologies, various mapping representations, classified methods for discovering methods both between ontology concepts and data, and talked about various tasks where mappings are used. In this extended abstract of our talk, we provide an annotated bibliography for this area of research, giving readers brief pointers on representative papers in each of the topics mentioned above. We did not attempt to compile a comprehensive bibliography and hence the list in this abstract is necessarily incomplete. Rather, we tried to sketch a map of the field, with some specific reference to help interested readers in their exploration of the work to-date.

1 Survey Articles

For more detailed descriptions and bibliography of the field we refer the readers to several recently published surveys:

- The survey of methods for automatic schema matching by Rahm and Bernstein [Rahm and Bernstein, 2001] is probably the most detailed map of different approaches to automatic mapping discovery up to 2001.
- A new (and much more brief survey) of the semantic-integration work in the database community by Doan and Halevy [Doan and Halevy, 2005] will be published in one of the upcoming issues of the AI Magazine.[1]
- On the ontology side, Kalfoglou and Schorlemmer [Kalfoglou and Schorlemmer, 2003b] provide an extremely comprehensive review of the state-of-the-art of ontology mapping.
- If you are looking for a brief review of ontology-based approaches to semantic integration, a survey by Noy [Noy, 2004] may provide just that.
- One of the deliverables for the European KnowledgeWeb project [Bouquet et al., 2004] also provides a detailed overview of mostly European projects on semantic integration.

The rest of this annotated bibliography is structured around the topics that we have outlined in the beginning of this section.

[1] At the time of this writing, the paper is not publicly available yet. Please contact us or the authors directly for a copy.

Dagstuhl Seminar Proceedings 04391
Semantic Interoperability and Integration
http://drops.dagstuhl.de/opus/volltexte/2005/48

2 Classifications of Heterogeneity

There has been some work on analyzing and classifying different types of mismatch between representation. Most of this work is targeted at database heterogeneity: [Kim and Seo, 1991,Kashyan and Sheth, 1996,Goh, 1997]. In his PhD thesis Wache compares and integrates these classifications into a common model [Wache, 2004]. Only newer approaches also look at heterogeneities occurring in different ontologies [Visser et al., 1998,Klein, 2001]. Most of these classifications distinguish between different levels of heterogeneity, often between syntactic, structural and semantic heterogeneities. For the case of ontologies, the aspect of having to deal with logical language of different expressiveness adds complications whereas work in the data area mostly assumes the entity relationship model to be the common frame of reference.

3 Representing Mappings

Different representations of mappings have been proposed recently. We can distinguish between approaches that have a different purpose in use:

Frameworks: Several researchers have proposed frameworks for describing mappings on an abstract level. These frameworks try to capture general aspects of mappings, often independent of a particular encoding language or an intended use.

- Calvanese and others [Calvanese et al., 2001] describe a framework for mapping ontologies that is motivated by their previous work on database integration. The framework is based on the traditional database integration architecture with a global and several local models and re-applies common database notions like the GAV and LAV approach to integration.
- The work of Madhavan and others[Madhavan et al., 2002] is also inspired by the database integration problem but allows more flexible architectures than the Calvanese paper. The general framework consists of some core definitions and a number of reasoning problems that are illustrated in the relational framework.
- In the context of the European Network KwowledgeWeb a general framework for the representation of mappings between semantic models has been developed [Bouquet et al., 2004]. The framework in intentionally independent of a particular representation language and only defines different types and elements of mappings.

Terminological Reasoning The ability to reason about the content of an ontology has been identified as an important tool for ontological engineering. Consequently, mapping representations that allow to perform reasoning across the mapped ontologies have been proposed.

- In the context of the i•com tool for conceptual modelling Franconi and Ng described the use of subsumption axioms in the $\mathcal{DLR}$ description logic to specify mapping between different conceptual database schemata [Franconi and Ng, 2000]. The tool automatically verifies mappings introduced by the user using satisfiability checking in $\mathcal{DLR}$.

- Stuckenschmidt and Klein introduce the notion of modular ontologies in terms of description logic models connected by conjunctive queries [Stuckenschmidt and Klein, 2003]. The semantics of the mappings is based on distributed description logics and allows to infer subsumption relations between concepts based on their mappings to other modules.
- Bouquet and others extend the Web Ontology Language with explicit notions of mappings [Bouquet et al., 2003]. As the modular ontology approach, the resulting language C-OWL is based on distributed description logics and allows to connect complex OWL expressions in different ontologies in a loose way. A completely distributed reasoning procedure can be used to perform typical description logic reasoning.
- In the OntoMerge system [Dou et al., 2002], authors use a general-purpose inference engine to enable translation between mapped ontologies. In OntoMerge the correspondence between two ontologies is expressed as a set of *bridging axioms* relating classes and properties of the two source ontologies. The two source ontologies, together with the bridging axioms are then treated as a single theory by a theorem prover optimized for ontology-translation task.

Data Transformation In many domains, ontologies are used to structure information. Exchanging information between sources that apply this kind of structuring requires a transformation of the data to fit the structures in the other source.

- Omelayenko introduces a model for specifying relations between heterogeneous RDF schema models for the purpose of data transformation in e-Commerce [Omelayenko, 2002]. The idea is to construct a separate RDF model that defines the relations in terms of so-called bridges. These bridges are accompanied by transformations that execute the translation.
- Maedche and others [Maedche et al., 2002] describe an approach that is very similar to the one of Omelayenko. They also define 'bridges' between elements of the different models and add transformation descriptions. As in the work of Omelayenko, the semantics of the bridges is only specified in terms of an RDf schema.
- The *mapping ontology* by Crubézy and colleagues [Crubézy and Musen, 2003] defines the structure of specific mappings and the transformation functions to transfer instances from one ontology to another. This ontology can then be used by tools to perform the transformations. The ontology provides different ways of linking concepts from the source ontology to the target ontology, transformation rules to specify how values should be changed, and conditions and effects of such rules.
- A more formal approach to the data translation problem is reported by [Dou et al., 2002]. Here integration is achieved by creating a merged model that consists of both ontologies plus a set of type mapping rules

and first order logic axioms that define the relation between the two models.

Query Processing Answering structured queries is one of the most intensively investigated tasks. The use of an ontology to structure and describe data can support this task. With respect to mappings, the problem of querying across sources with different ontologies has been addressed recently.

- A early approach that addresses the problem querying multiple sources with different ontologies is summarized in [Mena and Illarramendi, 2001]. The authors propose the use of linguistic relations between class names from different description logic ontologies. These relations are used to translate query terms and estimate the amount of information lost during translation.
- In [Serafini et al., 2003] the authors argue that the global schema assumptions of traditional database integration approaches in inadequate for decentralized information systems. They propose the local relational model for connecting different databases for the purpose of query processing using local model semantics and limited first order axioms for specifying semantic relations between models.
- The approach described in [Franconi et al., 2004] is very similar to the local relational model. It uses many of its ideas like the local model semantics and the use of coordination axioms, but restrict them to datalog. Using this restriction they discuss a number of reasoning tasks and formal properties.

4 Mapping Discovery

One of the most active areas of research in ontology alignment is the automatic and semi-automatic mapping discovery. The KnowledgeWeb deliverable mentioned earlier [Bouquet et al., 2004] provides details about many of these tools. In this annotated bibliography we categorize the tools based on the type of information they employ in their algorithms.

The first class of tools deals with the case where the two ontologies to be mapped share a common reference ontology. Several upper ontologies, such as SUMO [Niles and Pease, 2001] and DOLCE [Gangemi et al., 2003] are developed specifically for the goal of facilitating knowledge sharing. Grüninger and Kopena propose an approach to ontology integration that is based specifically on the idea of a shared interlingua [Grüninger and Kopena, 2003].

When a shared ontology is not available, mapping-discovery tools use other types of information: lexical and structural information, user input, external resources, or prior matches.

The tools developed by Hovy and colleagues [Hovy, 1998] are probably the most representative of the tools using lexical information, such as concept names and definitions, their lexical structure, distance between strings, and so on.

The majority of tools for ontology mapping use some sort of structural or definitional information to discover new mappings. This information includes such elements as subclass–superclass relationships, domains and ranges of properties, analysis of the graph structure of the ontology, and so on. Some of the tools in this category include IF-Map [Kalfoglou and Schorlemmer, 2003a], QOM [Ehrig and Staab, 2004], Similarity Flooding [Melnik et al., 2002], and the Prompt tools [Noy and Musen, 2003,Noy and Musen, 2004].

User input is another important source of information. Most researchers believe that completely automatic ontology mapping is beyond our reach and therefore some user interaction is required. This interaction may include seeding the mapping algorithm with initial set of matching pairs, verifying the matches that an algorithm produces, or configuring the specific matchers used [Noy and Musen, 2003,McGuinness et al., 2000,Mitra et al., 2000].

Many external sources available in electronic form, provide useful information for mapping discovery. The S-match algorithm [Giunchiglia et al., 2004], for instance, uses annotations from WordNet to help in finding mappings.

Other type of information outside the ontologies themselves are prior matches. Matching is laborious and error-prone process, and once ontology mappings are discovered, it may be a good idea to use these verified mappings in order to find new ones. One architecture for using prior mappings is peer-to-peer: given mappings between ontologies A and B and ontologies B and C, we can combine these mappings to find the mapping between A and C. Several researchers studies mapping composition in this setting [Kementsietsidis et al., 2003,Madhavan and Halevy, 2003]. In the next stage, one can imagine a network of ontologies and mappings available between some of them. Again, given two ontologies in this network, one could find different paths of existing mappings and combine them to produce the required one. The Semantic gossiping framework [Aberer et al., 2003] takes this approach.

Another architecture for using prior mappings is to use them as a corpus for machine-learning approaches. In this setting, given a large number of schemas and mappings, one could collect statistics on commonly matching terms. distribution of mappings, and so on and use this information to create a mapping between two previously unseen ontologies [Madhavan et al., 2005].

Matching ontologies, however, is only one thrust of mapping-discovery methods. Recently, researchers gave considerable attention to efforts in matching data elements: The task of data matching is the task of discovering whether two data elements refer to the same real-life object. Some of the representative examples of work in this area include using probabilistic models to match citations [Pasula et al., 2003] and using external sources to match objects [Michalowski et al., 2004]

References

[Aberer et al., 2003] Aberer, K., Cudré-Mauroux, P., and Hauswirth, M. (2003). Start making sense: The chatty web approach for global semantic agreements. *Journal of Web Semantics*, 1(1).

[Bouquet et al., 2004] Bouquet, P., Euzenat, J., Franconi, E., Serafini, L., Stamou, G., and Tessaris, S. (2004). Specification of a common framework for characterizing alignment. Deliverable 2.2.4, KnowledgeWeb.

[Bouquet et al., 2003] Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., and Stuckenschmidt, H. (2003). C-OWL: Contextualizing ontologies. In *Second International Semantic Web Conference ISWC'03*, volume 2870 of *LNCS*, pages 164–179. Springer.

[Calvanese et al., 2001] Calvanese, D., De Giacomo, G., and Lenzerini, M. (2001). A framework for ontology integration. In *Proceedings of the Semantic Web Working Symposium*, pages 303–316, Stanford, CA.

[Crubézy and Musen, 2003] Crubézy, M. and Musen, M. A. (2003). Ontologies in support of problem solving. In Staab, S. and Studer, R., editors, *Handbook on Ontologies*, pages 321–342. Sringer.

[Doan and Halevy, 2005] Doan, A. and Halevy, A. (2005). Semantic integration research in the database community: A brief survey. *AI Magazine*, to appear.

[Dou et al., 2002] Dou, D., McDermott, D., and Qi, P. (2002). Ontology translation by ontology merging and automated reasoning. In *Proceedings of the EKAW Workshop on Ontologies for Multi-Agent Systems*, Siguenza, Spain.

[Ehrig and Staab, 2004] Ehrig, M. and Staab, S. (2004). QOM - Quick Ontology Mapping. In *3rd International Semantic Web Conference (ISWC2004)*, Hiroshima, Japan.

[Franconi et al., 2004] Franconi, E., Kuper, G., Lopatenko, A., and Zaihrayeu, I. (2004). Queries and updates in the codb peer to peer database system. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB'04)*.

[Franconi and Ng, 2000] Franconi, E. and Ng, G. (2000). The i•com tool for intelligent conceptual modelling. In *Proceedings of the 7th Intl. Workshop on Knowledge Representation meets Databases (KRDB'00)*, Berlin, Germany.

[Gangemi et al., 2003] Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. (2003). Sweetening wordnet with DOLCE. *AI Magazine*, 24(3):13–24.

[Giunchiglia et al., 2004] Giunchiglia, F., Shvaiko, P., and Yatskevich, M. (2004). Semantic matching. In *1st European semantic web symposium (ESWS'04)*, pages 61–75, Heraklion, Greece.

[Goh, 1997] Goh, C. (1997). *Representing and reasoning about semantic conflicts in heterogeneous Information Sources*. PhD thesis, MIT.

[Grüninger and Kopena, 2003] Grüninger, M. and Kopena, J. (2003). Semantic integration through invariants. In Doan, A., Halevy, A., and Noy, N., editors, *Workshop on Semantic Integration at ISWC-2003*, Sanibel Island, FL.

[Hovy, 1998] Hovy, E. (1998). Combining and standardizing largescale, practical ontologies for machine translation and other uses. In *The First International Conference on Language Resources and Evaluation (LREC)*, pages 535–542, Granada, Spain.

[Kalfoglou and Schorlemmer, 2003a] Kalfoglou, Y. and Schorlemmer, M. (2003a). IF-Map: an ontology mapping method based on information flow theory. *Journal on Data Semantics*, 1(1):98–127.

[Kalfoglou and Schorlemmer, 2003b] Kalfoglou, Y. and Schorlemmer, M. (2003b). Ontology mapping: the state of the art. *The Knowledge Engineering Review*, 18(1):1–31.

[Kashyan and Sheth, 1996] Kashyan, V. and Sheth, A. (1996). Semantic and schematic similarities between database objects. *VLDB Journal*, 5(4):276–304.

[Kementsietsidis et al., 2003] Kementsietsidis, A., Arenas, M., and Miller, R. J. (2003). Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *ACM SIGMOD International Conference on Management of Data*, pages 325–336.

[Kim and Seo, 1991] Kim, W. and Seo, J. (1991). Classifying schematic and data heterogeneity in multidtabase systems. *IEEE Computer*, 24(12):145–169.

[Klein, 2001] Klein, M. (2001). Combining and relating ontologies: an analysis of problems and solutions. In Gomez-Perez, A., Gruninger, M., Stuckenschmidt, H., and Uschold, M., editors, *Workshop on Ontologies and Information Sharing, IJCAI'01*, Seattle, WA.

[Madhavan et al., 2002] Madhavan, J., Bernstein, P., Domingos, P., and Halevy, A. (2002). Representing and reasoning about mappings between domain models. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI'2002)*, Edmonton, Canada.

[Madhavan et al., 2005] Madhavan, J., Bernstein, P. A., Doan, A., and Halevy, A. (2005). Corpus-based schema matching. In *International Conference on Data Engineering (ICDE)*, Tokyo, Japan.

[Madhavan and Halevy, 2003] Madhavan, J. and Halevy, A. (2003). Composing mappings among data sources. In *International Conference on Very Large Databases (VLDB) 2003*, Berlin, Germany.

[Maedche et al., 2002] Maedche, A., Motik, B., Silva, N., and Volz, R. (2002). MAFRA - a mapping framework for distributed ontologie. In *Proceedings of 13th European Conference on Knowledge Engineering and Knowledge Management (EKAW)*, Siguenza, Spain.

[McGuinness et al., 2000] McGuinness, D. L., Fikes, R., Rice, J., and Wilder, S. (2000). An environment for merging and testing large ontologies. In Cohn, A. G., Giunchiglia, F., and Selman, B., editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR2000)*. Morgan Kaufmann Publishers, San Francisco, CA.

[Melnik et al., 2002] Melnik, S., Garcia-Molina, H., and Rahm, E. (2002). Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *18th International Conference on Data Engineering (ICDE-2002)*, San Jose, California. IEEE Computing Society.

[Mena and Illarramendi, 2001] Mena, E. and Illarramendi, A. (2001). *Ontology-Based Query Processing for Global Information Systems*. Kluwer.

[Michalowski et al., 2004] Michalowski, M., Thakkar, S., and Knoblock, C. A. (2004). Exploiting secondary sources for unsupervised record linkage. In *2004 VLDB Workshop on Information Integration on the Web*.

[Mitra et al., 2000] Mitra, P., Wiederhold, G., and Kersten, M. (2000). A graph-oriented model for articulation of ontology interdependencies. In *Proceedings Conference on Extending Database Technology 2000 (EDBT'2000)*, Konstanz, Germany.

[Niles and Pease, 2001] Niles, I. and Pease, A. (2001). Towards a standard upper ontology. In *The 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Ogunquit, Maine.

[Noy, 2004] Noy, N. (2004). Semantic integration: A survey of ontology-based approaches. *SIGMOD Record*, to appear.

[Noy and Musen, 2003] Noy, N. F. and Musen, M. A. (2003). The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024.

[Noy and Musen, 2004] Noy, N. F. and Musen, M. A. (2004). Ontology versioning in an ontology-management framework. *IEEE Intelligent Systems*, page in press.

[Omelayenko, 2002] Omelayenko, B. (2002). Rdft: A mapping meta-ontology for business integration. In *Proceedings of the Workshop on Knowledge Transformation for the Semantic Web (KTSW 2002) at the 15-th European Conference on Artificial Intelligence*, pages 76–83, Lyon, France.

[Pasula et al., 2003] Pasula, H., Marthi, B., Milch, B., Russell, S., and Shpitser, I. (2003). Identity uncertainty and citation matching. *Advances in Neural Information Processing Systems*, 15.

[Rahm and Bernstein, 2001] Rahm, E. and Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *VLDB Journal*, 10(4).

[Serafini et al., 2003] Serafini, L., Giunchiglia, F., Mylopoulos, J., and Bernstein, P. (2003). Local relational model: a logical formalization of database coordination. In Blackburn, P., Ghidini, C., Turner, R., and Giunchiglia, F., editors, *Proceedings of the 4th international and interdisciplinary conference on modeling and using context (CONTEXT 2003)*, pages 286–299, Stanford, CA.

[Stuckenschmidt and Klein, 2003] Stuckenschmidt, H. and Klein, M. (2003). Integrity and change in modular ontologies. In *Proceedings of the International Joint Conference on Artificial Intelligence - IJCAI'03*, Aapulco, Mexico.

[Visser et al., 1998] Visser, P., Jones, D., Bench-Capon, T., and Shave, M. (1998). Assessing hetrogeneity by classifying ontology mismatches. In *Proceedings International Conference on Formal Ontology in Information System (FOIS'98)*. IOS Press.

[Wache, 2004] Wache, H. (2004). *Semantic Mediation for heterogeneous Information Sources*. PhD thesis, Univeristy of Bremen. in German.