

Description Logics for Interoperability

Enrico Franconi

<http://www.inf.unibz.it/~franconi/>

Faculty of Computer Science, Free University of Bozen-Bolzano, Italy

Description Logics (DL) [5] are a very promising research area in Knowledge Representation (KR) with applications in databases (DBs). The main effort of the research in DL is in providing both theories and systems for expressing structured knowledge and for accessing and reasoning with it in a principled way. Recently, basic progress has been made by establishing the theoretical foundations for the effective use of DL in information systems [8]. DL offer promising formalisms for solving several problems concerning Conceptual Data Modelling and Ontology Design (see, e.g., [7], or the DAML+OIL and OWL efforts [20]), Intelligent Information Access and Query processing, and Information Integration and Interoperability.

I want to argue that good *Conceptual Modelling* and *Ontology Design* is required to support powerful *Query Management* and to allow for semantic based *Information Integration and Interoperability*. Therefore, this short survey has been structured into three parts. In the first part, the notions of ontology language and of methodology for conceptual and ontology design will be introduced. In the second part, the query management problem in the presence of the previously devised conceptual model will be considered: a global framework will be introduced, together with various basic tasks involved in information access. In the last part, general issues about integration and interoperability will be presented. The most relevant research work carried out in our group will be cited in the paper, while for all the other relevant citations we refer to [4]. This work has been partially supported by the EU projects Sewasie, KnowledgeWeb, and Interop.

Conceptual Modelling and Ontology Design

For the purpose of this short survey, an Ontology will be considered as a Conceptual Schema expressed in a suitable conceptual data model (i.e., an Ontology Language). Good *conceptual data models* put their emphasis on the correct and semantically rich representation of *complex* properties and relations that may exist between documents. They should allow for an abstract representation of data which resembles the way they are actually perceived and used in the real world, thus shortening (with respect to the more traditional data models) the semantic gap between the domain and its representation.

Conceptual (or Ontology) modelling deals with the question on how to describe in a declarative and reusable way the domain information of an application, its relevant vocabulary, and how to constrain the use the data, by understanding what can be drawn from it. Not surprisingly, conceptual modelling tasks have always been in the mainstream of KR research – see for example the research on Ontology representation and design – and can be considered now one of the main applications of KR languages and reasoning technique. In addition, given the high complexity of the modelling task when complex data is involved, in the semantic web field there is the demand of more sophisticated and expressive languages than for normal information systems. Again, DL research is very active in providing expressive ontology languages to capture various aspects of the information (see, e.g., [1, 3, 2, 11, 15, 12, 6]).

A big part of the DL community likes to see a generic ontology language as the generalisation of both the object-oriented data model based on UML class

diagrams and the extended Entity-Relationship (EER) semantic data model, strictly related to the ontology web languages such as DAML+OIL and OWL. Our work in this direction includes the i.com tool [14, 22] – which fully implements an extended conceptual data model generalising the UML class diagrams and the EER schemas – and which is available online for the evaluation of the principles just exposed at the public web address <http://www.inf.unibz.it/~franconi/icom/>. i.com allows for the specification of multiple EER (or UML) diagrams and inter- and intra-schema constraints. Complete logical reasoning is employed by the tool using an underlying DL inference engine to verify the specification, infer implicit facts and stricter constraints, and manifest any inconsistencies during the conceptual modelling phase.

Information Access

Only recently has KR research started to have an interest in query processing and information access. Recent work has come up with advanced reasoning techniques for query evaluation and rewriting using views under the constraints given by the ontology – also called view-based query processing. This means that the notion of accessing information through the navigation of an Ontology modelling the document’s domain – which can be seen as a conceptual schema – has its formal foundations.

I will thus consider DL for formalising not only the ontology but also the query processing as well. The (DL-based) conceptual schema as defined in the previous section can be seen as a set of constraints over a vocabulary which is usually richer than the logical schema of the information system it is modelling. In some sense, quite often the conceptual schema plays the role of a general ontology of the domain, very close to the user’s rich vocabulary, rather than of a set of constraints over the poor logical vocabulary structuring the data. With this perspective in mind, the user would prefer to query the information system using the richer vocabulary of the ontology [9]. The vocabulary of the basic data (i.e., the logical schema) could be seen in turn either as a subset of the conceptual vocabulary – this is the simplistic view – or more generally as a set of (materialised) views over the vocabulary of the ontology. However, in this case we have to solve the problem of view-based query processing. The problem requires to answer a query posed to a database – the one defined by the ontology – only on the basis of the information in a set of (materialised) views, which are again queries over the same database. In the process, the information contained in the conceptual schema of the database should be of course taken into account.

Two approaches to view-based query processing exist, namely query rewriting and query answering (see, e.g., [25]). This framework can be used to characterise several aspects of an information system to support interoperability. In query optimisation, view-based query processing is relevant because using the views may speed up query processing. In data integration, the views represent the only information sources accessible to answer a query. A data warehouse can be seen as a set of materialised views, and, therefore, query processing reduces to view-based query answering. Finally, since the views provide partial knowledge on the database, view-based query processing can be seen as a special case query answering with incomplete information.

Information Integration and Interoperability

In this last part I will hint how the technologies introduced in the first two parts, namely a very expressive ontology language and view-based query processing over it, can be used in the framework of Information Integration [16, 21, 22, 24].

Let us suppose to have multiple databases to be integrated. Each database will have its own conceptual schema and logical schema, where, as seen in the previous part, the logical schema is just a set of views over the conceptual schema (local-as-view approach). We assume that each symbol of each schema is identified by a unique global symbol, i.e., the various databases have disjoint signatures. Interdependencies between entities and relationships in different schemas are represented by means of integrity constraints involving symbols of the schemas. Such interdependencies are called *coordination formulas*, and they are of the form of inclusion dependencies expressed in a suitable view language (e.g., a DL itself, or a SPJ query language. The union of the various schemas with the coordination formulas and the local views forms the global integrated schema, or the *mediator*. It is worth noting that the integration process is incremental – since the integrated schema can be monotonically refined as soon as there is new understanding of the different component schemas – and that the resulting unified schema is strongly dependent from (actually, it includes) the schemas of the single information sources.

This approach gives both a clear semantics to the integration process of ontologies, and a calculus for deriving inconsistencies and checking the validity of integrity constraints in the integrated schema. Most importantly, in this framework global queries can be defined as views over single ontologies, or they can be generalised to span over multiple ontologies. The view-based query processing mechanism will guarantee the correct answer to the global query from the local sources.

In [23] a comparison is given between the above local-as-view approach to processing global queries and the global-as-view approach, which is more common in current information integration architectures.

More recent research work is dealing with the application of the above ideas in a peer-to-peer interoperability framework [17, 19, 18]. Here, the difference with the classical framework are the following: (a) the role of the coordination formulas between nodes is for data migration (as opposed to the role of logical constraints in classical data integration systems); (b) computation is delegated to single nodes (distributed local computation); (c) the topology of the network may dynamically change; (d) local inconsistency does not propagate; (e) computational complexity can be low.

References

1. A. Artale and E. Franconi. A survey of temporal extensions of description logics. *Annals of Mathematics and Artificial Intelligence*, 30(1-4), 2001.
2. A. Artale and E. Franconi. Temporal description logics. In Dov Gabbay, Michael Fisher, and Lluís Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence*. Elsevier, “Foundations of Artificial Intelligence” series, 2004.
3. A. Artale, E. Franconi, F. Wolter, and M. Zakharyashev. A temporal description logic for reasoning over conceptual schemas and queries. In *Proc. of the 8th European Conference on Logics in Artificial Intelligence (JELIA-2002)*, 2002.
4. F. Baader, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2002.
5. F. Baader and W. Nutt. Basic description logics. In Baader et al. [4].
6. Franz Baader, Ralph Kuesters, and Frank Wolter. Extensions to description logics. In Baader et al. [4].
7. A. Borgida and R. J. Brachman. Conceptual modelling with description logics. In Baader et al. [4].
8. A. Borgida, M. Lenzerini, and R. Rosati. Description logics for databases. In Baader et al. [4].

9. Tiziana Catarci, Tania Di Mascio, Enrico Franconi, Giuseppe Santucci, and Sergio Tessaris. An ontology based visual tool for query formulation support. In *Proc. of ODBASE'03, International Conference on Ontologies, Databases and Applications of SEmanatics*, November 2003.
10. E. Franconi, F. Baader, U. Sattler, and P. Vassiliadis. Multidimensional data models and aggregation. In Jarke et al. [21], chapter 5, pages 87–106.
11. E. Franconi, F. Grandi, and F. Mandreoli. A semantic approach for schema evolution and versioning in object-oriented databases. In *Proc. of the 1st International Conf. on Computational Logic (CL'2000), DOOD stream*. Springer-Verlag, July 2000.
12. E. Franconi, F. Grandi, and F. Mandreoli. Description logics for modelling dynamic information. In Jan Chomicki, Ron van der Meyden, and Gunter Saake, editors, *Logics for Emerging Applications of Databases*. Springer-Verlag, 2003.
13. E. Franconi and M. Kifer, editors. *Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99)*. Linköping University Technical Report, July 1999. Also electronically available as CEUR Publication, Vol. 21, RWTH Aachen, Germany.
14. E. Franconi and G. Ng. The ICOM tool for intelligent conceptual modelling. In *Proc. of the 7th International Workshop on Knowledge Representation meets Databases (KRDB'2000)*, 2000.
15. E. Franconi and U. Sattler. A data warehouse conceptual data model for multidimensional aggregation. In *Proceedings of the Workshop on Design and Management of Data Warehouses (DMDW'99)*, 1999.
16. Enrico Franconi, Ken Barker, and Diego Calvanese, editors. *Proceedings of the International Workshop on Foundations of Models for Information Integration (FMII-2001)*. Springer-Verlag, 2001.
17. Enrico Franconi, Gabriel Kuper, Andrei Lopatenko, and Luciano Serafini. A robust logical and computational characterisation of peer-to-peer database systems. In *Proc. of the VLDB International Workshop On Databases, Information Systems and Peer-to-Peer Computing (DBISP2P-2003)*, 2003.
18. Enrico Franconi, Gabriel Kuper, Andrei Lopatenko, and Ilya Zaihraeu. The coDB robust peer-to-peer database system. In *Proc. of the 2nd Workshop on Semantics in Peer-to-Peer and Grid Computing*, 2004.
19. Enrico Franconi, Gabriel Kuper, Andrei Lopatenko, and Ilya Zaihraeu. A distributed algorithm for robust data sharing and updates in p2p database networks. In *Proc. of the EDBT International Workshop on Peer-to-Peer Computing and Databases*, 2004.
20. Christopher A. Welty Ian Horrocks, Deborah L. McGuinness. Digital libraries and web based information systems. In Baader et al. [4].
21. M. Jarke, M. Lenzerini, Y. Vassilios, and P. Vassiliadis, editors. *Fundamentals of Data Warehousing*. Springer-Verlag, 1999.
22. Mathias Jarke, V. Quix, D. Calvanese, Maurizio Lenzerini, Enrico Franconi, S. Ligoudistiano, P. Vassiliadis, and Yannis Vassiliou. Concept based design of data warehouses: The DWQ demonstrators. In *2000 ACM SIGMOD International Conference on Management of Data*, May 2000.
23. Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. of PODS-2002*, pages 233–246, 2002.
24. Martin Peim, Enrico Franconi, and Norman Paton. Applying functional languages in knowledge-based information integration systems. In Peter Gray, Larry Kerschberg, Peter King, and Alex Poulouvasilis, editors, *The Functional Approach to Data Management*. Springer-Verlag, 2004.
25. Martin Peim, Enrico Franconi, Norman Paton, and Carole Goble. Query processing with description logic ontologies over object-wrapped databases. In *Proc. of the 14th International Conference on Scientific and Statistical Database Management (SSDBM'02)*, July 2002.