Description Logics for Conceptual Design, Information Access, and Ontology Integration: Research Trends

Enrico Franconi

franconi@inf.unibz.it
http://www.inf.unibz.it/~franconi/
Faculty of Computer Science, Free University of Bozen-Bolzano, Italy

Abstract. In this short survey paper I want to briefly describe the role of Description Logics in the ontology world. I want to argue that good *Conceptual Modelling* and *Ontology Design* are required to support powerful *Query Management* and to allow for semantic based *Information Integration*.

Sommario. In questo breve articolo di rassegna descriverò il ruolo delle Logiche Descrittive nel mondo delle ontologie. Argomenterò che una buona *modellazione concettuale* ed una buona *progettazione di ontologie* sono necessarie per supportare la manipolazione di *interrogazioni* e per permettere *integrazione di informazione* basata sulla semantica.

1 Introduction

In recent years, data and knowledge base applications have progressively converged towards integrated technologies that try to overcome the limits of each single discipline. Research in Knowledge Representation (KR) originally concentrated around formalisms that are typically tuned to deal with relatively small knowledge bases, but provide powerful deduction services, and the language to structure information is highly expressive; research on formal languages for ontologies was originated from KR. In contrast, Information Systems and Database research mainly dealt with efficient storage and retrieval with powerful query languages, and with sharing and displaying large amounts of (multimedia) documents. However, data representations were relatively simple and flat, and reasoning over the structure and the content of the documents played only a minor role.

This distinction between the requirements in Knowledge Representation and Databases is vanishing rapidly. On the one hand, to be useful in realistic applications, such as the applications in the semantic web, a modern ontology KR system must be able to handle large data sets, and to provide expressive query languages. This suggests that techniques developed in the DB area could be useful for ontologies. On the other hand, the information stored on the web, in digital libraries, and in data warehouses is now very complex and with deep semantic structures, thus requiring more intelligent modelling languages and methodologies, and reasoning services on those complex representations to support design, management, retrieval, and integration. Therefore, a great call for an integrated view of Knowledge Representation and Database technologies is emerging.

***Networks* 2**: 25–32, 2003

Description Logics (DL) (Baader and Nutt 2002) are a very promising research area in KR with applications in 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 (Calvanese et al. 2002a; Donini 2002). Recently, basic progress has been made by establishing the theoretical foundations for the effective use of DL in information systems (Borgida 1995; Borgida et al. 2002). DL offer promising formalisms for solving several problems concerning Conceptual Data Modelling and Ontology Design (see, e.g., (Calvanese et al. 1998a; Borgida and Brachman 2002), or the DAML+OIL and OWL efforts (Fensel et al. 2000; Ian Horrocks 2002)), Intelligent Information Access and Query processing (see, e.g., (Borgida and Brachman 1993; Levy and Rousset 1998; Bresciani et al. 2000; Franconi 2000)), and Information Integration (see, e.g., (Calvanese et al. 1998b; Jarke et al. 2000; Mena et al. 2000; Goasdoue et al. 2000)).

This short survey will serve as a reference listing research trends, rather than being a strictly theoretical survey. Its aim is to let the audience understand why DL and DB technologies could be useful to semantic web research and applications; precise links to the important theoretical results and to the relevant references are given.

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*. 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 ontology integration will be presented.

2 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. Recently, a number of conceptual and ontology modelling languages has emerged as de-facto standard, in particular we mention Entity/Relationship (ER) for the relational data model, UML and ODMG for the object oriented data model, and XML, RDF(S), DAML+OIL and OWL for the web semi-structured data model. Still, many such languages do not have a formal semantics based on logic, or reasoners built upon them to support the designer. 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 techniques (Borgida and Brachman 2002). DL can be considered as an unifying formalism, since they allow the logical reconstruction and the extension of representational tools such as object-oriented data models (e.g., UML and ODMG), semantic data models (e.g., Entity/Relationship and ORM), frame-based ontology languages (e.g., DAML+OIL and OWL—the DL community has been heavily influential in the DAML+OIL and OWL web ontology languages proposals) (Calvanese et al. 1998a, 1999; Calì et al. 2001; Fensel et al. 2000). 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., (Artale and Franconi 2001; Artale et al. 2002; Artale and Franconi 2004; Franconi et al. 2000; Franconi and Sattler 1999; Franconi et al. 2003; Baader et al. 2002b)).

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 (see, e.g.,(Franconi and Ng 2000; Jarke et al. 2000)). Such an ontology language includes *taxonomic* relations to state containment assertions between entities and between relationships with the possibility to specify additional *covering* and *disjointness* constraints. The most interesting feature of this modelling language is the ability to completely *define* entities and relationships as *views* over other entities and relationships of the ontology (Calvanese et al. 1998a). The preferred adopted view language is DLR (Calvanese et al. 1998b), a Description Logic over unary and *n*-ary relationships. DLR is an interesting decidable fragment of first order logic: among others, inclusion dependencies with DLR views can express (a) unary inclusion dependencies, (b) typed inclusion dependencies without projection, (c) existence dependencies, (d) exclusion dependencies, and (e) full key dependencies. DLR is powerful enough to encode the full EER, the UML class diagrams and most of DAML+OIL and, of course, the complete OWL-DL.

Two additional extensions to the conceptual data model have also been considered. The first one is with multidimensional aggregations – that is, the conceptual data model is able to represent the structure of *aggregated entities* and of *multiply hierarchically organised dimensions*. The ability of representing aggregations at the conceptual level is crucial in modelling structured documents in data warehouses, in the semantic web and in digital libraries. The second one allows for the representation of standard temporal operators for temporal conceptual modelling and of a large class of temporal integrity constraints, useful to model the dynamics in the sematic web.

The i.com tool (Franconi and Ng 2000; Jarke et al. 2000) – which fully implements the above conceptual data model as UML class diagrams or EER schemas – 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.

3 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 (Ullman 1997; Calvanese et al. 2000b). 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 that the logical schema of the information system it is modelling. In some sense, quite often the conceptual schema plays the role of an 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 (Catarci et al. 2003). 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 (see, e.g., (Beeri et al. 1997)) and query answering (see, e.g., (Abiteboul and Duschka 1998; Calvanese et al. 2000a; Peim et al. 2002)). In the former approach, we are given a query Q, a set of view definitions characterising the actual data, and a set of (conceptual) constraints – all over the conceptual vocabulary – and the goal is to reformulate the query into an expression, the rewriting, that refers only to the views, and provides the answer to Q. Typically, the rewriting is formulated in the same language used for the query and the views. In the latter approach, besides Q, the view definitions and the constraints, we are also given the extensions of the (materialised) views. The goal is to compute the set of tuples that are implied by these extensions, i.e., the set of tuples that are in the answer set of Q in all the databases that are consistent with the views and the constraints.

This framework can be used to characterise several aspects of an information system. 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.

4 Information Integration

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 (Franconi et al. 2001; Catarci and Lenzerini 1993; Calvanese et al. 1998b; Jarke et al. 1999, 2000; Calvanese et al. 2002b; Peim et al. 2004).

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 *inter-model assertions*, and they are of the form of DLR inclusion dependencies. The union of the various schemas with the inter-model assertions 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 (Cal et al. 2004).

A particular but important case is the designing a Data Warehouse Conceptual Schema. In this case it is assumed to have a privileged schema – called the *Enterprise Model* – which is the conceptual representation of the global concepts and relationships reconciled and abstracted in the data warehouse, and it is not necessarily a complete model of all the source information. Such schema is integrated with the different source schemas. The crucial point is that not only the interrelationships between the source schemas and the Enterprise Model are modelled, but also the interdependencies between the source schemas themselves. Moreover, the global integrated schema – the Data Warehouse Conceptual Schema – is composed not only by the Enterprise Model, but also by the various source schemas and by the inter-model assertions. Global data warehouse queries are formally seen as views over the Enterprise Model.

In (Lenzerini 2002) 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.

Bibliografia

S. Abiteboul and O. Duschka. Complexity of answering queries using materialised views. In *Proc. of the 17th ACM Symp. on Principles of Database Systems (PODS'98)*, pages 254–265, 1998.

A. Artale and E. Franconi. A survey of temporal extensions of description logics. *Annals of Mathematics and Artificial Intelligence*, 30(1-4), 2001.

A. Artale and E. Franconi. Temporal description logics. In Dov Gabbay, Michael Fisher, and Lluis Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence*. Elsevier, "Foundations of Artificial Intelligence series, 2004.

A. Artale, E. Franconi, F. Wolter, and M. Zakharyaschev. 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.

F. Baader, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2002a.

F. Baader and W. Nutt. Basic description logics. In Baader et al. (2002a).

Franz Baader, Ralph Kuesters, and Frank Wolter. Extensions to description logics. In Baader et al. (2002a).

C. Beeri, A. Y. Levy, and M.-C. Rousset. Rewriting queries using views in description logics. In *Proc. of the 16th ACM Symp. on Principles of Database Systems (PODS'97)*, pages 99–108, 1997.

A. Borgida. Description logics in data management. *TKDE*, 7(5):671–682, 1995.

A. Borgida and R. J. Brachman. Conceptual modelling with description logics. In Baader et al. (2002a).

A. Borgida, M. Lenzerini, and R. Rosati. Description logics for databases. In Baader et al. (2002a).

Alexander Borgida and Ronald J. Brachman. Loading data into description reasoners. In *Proc. of 1993 ACM SIGMOD International Conference on Management of Data*, pages 217–226, 1993.

Paolo Bresciani, Michele Nori, and Nicola Pedot. A knowledge based paradigm for querying databases. In *Proc. of DEXA-00*, pages 794–804, 2000.

Andrea Cal`ı Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. A formal framework for reasoning on UML class diagrams. Technical report, Univ. of Rome "La Sapienza, Italy, 2001.

D. Calvanese, G. De Giacomo, M. Lenzerini, and D. Nardi. Reasoning in expressive description logics. In Baader et al. (2002a).

D. Calvanese, G. De Giacomo, and M. Lenzerini. Answering queries using views over description logics knowledge bases. In *Proc. of the 16th Nat. Conf. on Artificial Intelligence (AAAI 2000)*, 2000a.

D. Calvanese, G. De Giacomo, M. Lenzerini, and Moshe Y. Vardi. View-based query processing and constraint satisfaction. In *Proc. of the 15th IEEE Sym. on Logic in Computer Science (LICS 2000)*, 2000b.

D. Calvanese, M. Lenzerini, and D. Nardi. Description logics for conceptual data modeling. In J. Chomicki and G. Saake, editors, *Logics for Databases and Information Systems*. Kluwer, 1998a.

D. Calvanese, M. Lenzerini, and D. Nardi. Unifying class-based representation formalisms. *J. of Artificial Intelligence Research*, 11:199–240, 1999.

Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. Description logics for information integration. In A. Kakas and F. Sadri, editors, *Computational Logic: Logic Programming and Beyond, Essays in Honour of Robert A. Kowalski*, volume 2408 of *Lecture Notes in Computer Science*, pages 41–60. Springer, 2002b.

Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati. Information integration: Conceptual modeling and reasoning support. In *Proc. of the 6th Int. Conf. on Cooperative Information Systems (CoopIS'98)*, pages 280–291, 1998b.

Andrea Cal, Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. Data integration under integrity constraints. *Information Systems*, 29(2):147–163, 2004. Extended and revised version of [?].

T. Catarci and M. Lenzerini. Representing and Using Interschema Knowledge in Cooperative Information Systems. *Journal of Intelligent and Cooperative Systems*, 2(4): 375–398, 1993.

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 SEmantics*, November 2003.

F. Donini. Complexity of reasoning. In Baader et al. (2002a).

D. Fensel, I. Horrocks, F. van Harmelen, S. Decker, M. Erdmann, and M. Klein. OIL in a nutshell. In *Proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*, Lecture Notes In Artificial Intelligence. Springer-Verlag, 2000.

E. Franconi, F. Baader, U. Sattler, and P. Vassiliadis. Multidimensional data models and aggregation. In Jarke et al. (1999), chapter 5, pages 87–106.

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.

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.

E. Franconi and M. Kifer, editors. *Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99)*, July 1999. Linköping University Technical Report. Also electronically available as CEUR Publication, Vol. 21, RWTH Aachen, Germany.

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.

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.

Enrico Franconi. Knowledge representation meets digital libraries. In *Proc. of the 1st DELOS (Network of Excellence on Digital Libraries) workshop on "Information Seeking, Searching and Querying in Digital Libraries*, 2000.

Enrico Franconi, Ken Barker, and Diego Calvanese, editors. *Proceedings of the International Workshop on Foundations of Models for Information Integration (FMII-2001)*, 2001. Springer-Verlag.

Francois Goasdoue, Veronique Lattes, and Marie-Christine Rousset. The use of CARIN language and algorithms for information integration: the picsel system. *International Journal on Cooperative Information Systems*, 2000.

Christopher A. Welty Ian Horrocks, Deborah L. McGuinness. Digital libraries and web based information systems. In Baader et al. (2002a).

M. Jarke, M. Lenzerini, Y. Vassilious, and P. Vassiliadis, editors. *Fundamentals of Data Warehousing*. Springer-Verlag, 1999.

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.

Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. of PODS-2002*, pages 233–246, 2002.

Alon Y. Levy and Marie-Christine Rousset. Combining horn rules and description logics in CARIN. *Artificial Intelligence*, 104(1-2):165–209, 1998.

Eduardo Mena, Arantza Illarramendi, Vipul Kashyap, and Amit P. Sheth. OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distributed and Parallel Databases*, 8(2): 223–271, 2000.

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 Poulovassilis, editors, *The Functional Approach to Data Management*. Springer-Verlag, 2004.

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.

J. D. Ullman. Information integration using logical views. In *Proc. of the 6th Int. Conf on Database Theory (ICDT'97)*, pages 19–40, 1997.