(Description) Logics for Information Modelling and Access

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

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.

Description Logics (DL) [BN02] 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 [CDLN02, Don02]. Recently, basic progress has been made by establishing the theoretical foundations for the effective use of DL in information systems [Bor95, BLR02]. DL offer promising formalisms for solving several problems concerning Conceptual Data Modelling and Ontology Design (see, e.g., [CLN98, BB02], or the OIL and DAML+OIL efforts [FHvH+00, IH02]), Intelligent Information Access and Query processing (see, e.g., [BB93, LR98, BNP00, Fra00, PFPG02]), and Information Integration (see, e.g., [CGL+98, JQC+00, MIKS00, GLR00]).

In the talk I argue that good *Conceptual Modelling* and *Ontology Design* is required to support powerful *Query Management* and to allow for semantic based *Information Integration*. Therefore, the talk has been structured into three parts. In the first part, an extended ontology language and a 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 (which will be just sketched), general issues about ontology integration will be presented.

Conceptual Modelling and Ontology Design. For the purpose of this talk, 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 and DAML+OIL 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 [BB02]. 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., OIL and DAML+OIL) [CLN98, CLN99, CCDGL01, FHvH$^+$00]. 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., [AF99, AFWZ02, FGM00, FS99, BKW02]).

At the end of this first part, a demo of the i.com tool [FN00, JQC$^+$00] – which implements the above conceptual data model as UML class diagrams or EER schemas – will be given. 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 [Ull97, CGLV00]. 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.

In this talk 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. 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.

I will introduce the two approaches to view-based query processing, namely query rewriting (see, e.g., [BLR97]) and query answering (see, e.g., [AD98, CGL00]). 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.

References

[AD98] 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.

[AF99] A. Artale and E. Franconi. Temporal ER modeling with description logics. In *Proc. of the International Conference on Conceptual Modeling (ER'99)*. Springer-Verlag, November 1999.

[AFWZ02] 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)*. Springer-Verlag, 2002.

[BB93] 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.

[BB02] A. Borgida and R. J. Brachman. Conceptual modelling with description logics. In Baader et al. [BMNPS02].

[BKW02] Franz Baader, Ralph Kuesters, and Frank Wolter. Extensions to description logics. In Baader et al. [BMNPS02].

[BLR97] 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.

[BLR02] A. Borgida, M. Lenzerini, and R. Rosati. Description logics for databases. In Baader et al. [BMNPS02].

[BMNPS02] F. Baader, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2002.

[BN02] F. Baader and W. Nutt. Basic description logics. In Baader et al. [BMNPS02].

[BNP00] Paolo Bresciani, Michele Nori, and Nicola Pedot. A knowledge based paradigm for querying databases. In *Proc. of DEXA-00*, pages 794–804, 2000.

[Bor95] A. Borgida. Description logics in data management. *TKDE*, 7(5):671–682, 1995.

[CCDGL01] Andrea Calì 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.

[CDLN02] D. Calvanese, G. De Giacomo, M. Lenzerini, and D. Nardi. Reasoning in expressive description logics. In Baader et al. [BMNPS02].

[CGL+98] 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, 1998.

[CGL00] 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)*, 2000.

[CGLV00] 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)*, 2000.

[CL93] 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.

[CLN98] 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, 1998.

[CLN99] D. Calvanese, M. Lenzerini, and D. Nardi. Unifying class-based representation formalisms. *J. of Artificial Intelligence Research*, 11:199–240, 1999.

[Don02] F. Donini. Complexity of reasoning. In Baader et al. [BMNPS02].

[FBSV99] E. Franconi, F. Baader, U. Sattler, and P. Vassiliadis. Multidimensional data models and aggregation. In Jarke et al. [JLVV99], chapter 5, pages 87–106.

[FGM00] 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.

[FHvH+00] 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.

[FK99] 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.

[FN00] 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.

[Fra00] 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.

[FS99] 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.

[GLR00] 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.

[IH02] Christopher A. Welty Ian Horrocks, Deborah L. McGuinness. Digital libraries and web based information systems. In Baader et al. [BMNPS02].

[JLVV99] M. Jarke, M. Lenzerini, Y. Vassilious, and P. Vassiliadis, editors. *Fundamentals of Data Warehousing*. Springer-Verlag, 1999.

[JQC+00] 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.

[LR98] Alon Y. Levy and Marie-Christine Rousset. Combining horn rules and description logics in CARIN. *Artificial Intelligence*, 104(1-2):165–209, 1998.

[MIKS00] 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.

[PFPG02] M. Peim, E. Franconi, N. W. Paton, and C. A. 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-2002)*. IEEE Press, 2002.

[Ull97] J. D. Ullman. Information integration using logical views. In *Proc. of the 6th Int. Conf on Database Theory (ICDT'97)*, pages 19–40, 1997.