(Modal) Logics for Semistructured Data

Natasha Alechina
School of Computer Science and IT
University of Nottingham
Nottingham, NG7 2RD, UK
email: nza@cs.nott.ac.uk

The area of semistructured data includes collections of data items which have in some ways similar but not identical structure. Examples of semistructured data range from heterogeneous databases to the World Wide Web [Abi97]. The area is obviously quite heterogeneous itself. However there are some important features common to all kinds of semistructured data, namely:

- data is represented as an edge labelled graph;
- querying data involves traversing paths in a graph.

Both features suggest that modal logic techniques can be successfully used for analysing logical properties of semistructured data. The reason for believing this is that modal logics were successfully applied to express similar properties over similar graphs, for example transition systems and feature structures.

Logical issues involved in working with semistructured data (the list below contains mutually dependent items!) are:

- which operations should a query language have;
- expressive power and complexity of query evaluation;
- how precisely do we want to describe structures (what is the right notion of equivalence);
- how do we express information available about the format of data (what is the right description language).

The talk will contain an overview of work on semistructured data and advocate the use of 'modal fragments' of transitive closure logic [Imm87] as a formal basis for query and description languages for semistructured data. By a modal fragment of some language containing first order quantifiers $\forall$ and $\exists$ we mean a fragment where the range of quantifiers is restricted to elements which are 'accessible' from the parameters of the quantified formula [Ale95]. In the context of first order logic extended with transitive closure operator, accessibility means existence of a path. We believe that this restriction reflects the spirit of navigational query languages such as Lorel [AQM97].

References

[Abi97] S. Abiteboul. Querying semi-structured data. In *Proceedings ICDT'97*, 1997.

[AQM97] S. Abiteboul, D. Quass, J. McHugh, J. Widom, and J.L. Wiener. The Lorel query language for semistructured data. *Journal of Digital Libraries*, 1(1):68-88, 1997.

[Ale95] N. Alechina. *Modal Quantifiers.* PhD thesis, ILLC, University of Amsterdam, 1995.

[Imm87] N. Immerman. Languages that capture complexity classes. *SIAM Journal of Computing*, 16(4):760-778, 1987.