CHAPTER 8

Complex Data Types

One of the key requirements of the relational model is that data values be atomic: multivalued, composite, and other complex data types are disallowed by the core relational model. However, there are many applications where the constraints on data types imposed by the relational model cause more problems than they solve. Several *non-atomic data types* are now widely used, including semi-structured data, object-based data, textual data, and spatial data.

Bibliographical Notes

Several object-oriented extensions to SQL have been proposed. POSTGRES ([Stonebraker and Rowe (1986)] and [Stonebraker (1986)]) was an early implementation of an object-relational system. Other early object-relational systems include the SQL extensions of O_2 ([Bancilhon et al. (1989)]) and UniSQL ([UniSQL (1991)]).

SQL:1999 was the product of an extensive (and long-delayed) standardization effort, which originally started off as adding object-oriented features to SQL and ended up adding many more features, such as procedural constructs, which we saw earlier. Support for multiset types was added as part of SQL:2003.

[Melton (2002)] concentrates on the object-relational features of SQL:1999. [Eisenberg et al. (2004)] provides an overview of SQL:2003, including its support for multisets.

A number of object-oriented database systems were developed in the late 1980s and early 1990s. Among the notable commercial ones were ObjectStore ([Lamb et al. (1991)]), O_2 ([Lecluse et al. (1988)]), and Versant. The object database standard ODMG is described in detail in [Cattell (2000)].

Oracle provides a fairly complete implementation of the object-relational features of SQL, while PostgreSQL provides a smaller subset of those features. More information on support for these features may be found in their respective manuals.

[Salton (1989)] is an early textbook on information-retrieval systems. [Manning et al. (2008), Chakrabarti (2002), Grossman and Frieder (2004), Witten et al. (1999)], and [Baeza-Yates and Ribeiro-Neto (2011)] provide textbook descriptions of information retrieval and data mining relevant to information retrieval.

[Agrawal et al. (2002), Bhalotia et al. (2002)] and [Hristidis and Papakonstantinou (2002)] cover keyword querying of relational data. Keyword querying of XML data are addressed by [Florescu et al. (2000)] and [Guo et al. (2003)], among others.

A tutorial on JSON can be found at www.w3schools.com/js/js_json_intro.asp. More information about XML can be found in Chapter 30, available online. More information about RDF can be found at www.w3.org/RDF/. Apache Jena provides an RDF implementation, with support for SPARQL; a tutorial on SPARQL can be found at jena.apache.org/tutorials/sparql.html

Bibliography

[Agrawal et al. (2002)] S. Agrawal, S. Chaudhuri, and G. Das, "DBXplorer: A System for Keyword-Based Search over Relational Databases", In *Proc. of the International Conf. on Data Engineering* (2002), pages 5–16.

[Baeza-Yates and Ribeiro-Neto (2011)] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, 2nd edition, ACM Press (2011).

[Bancilhon et al. (1989)] F. Bancilhon, S. Cluet, and C. Delobel, "A Query Language for the O_2 Object-Oriented Database", In *Proc. of the Second Workshop on Database Programming Languages* (1989), pages 122–138.

[Bhalotia et al. (2002)] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, "Keyword Searching and Browsing in Databases using BANKS", In *Proc. of the International Conf. on Data Engineering* (2002), pages 431–440.

[Cattell (2000)] R. Cattell, editor, *The Object Database Standard: ODMG 3.0*, Morgan Kaufmann (2000).

[Chakrabarti (2002)] S. Chakrabarti, *Mining the Web: Discovering Knowledge from HyperText Data*, Morgan Kaufmann (2002).

[Eisenberg et al. (2004)] A. Eisenberg, J. Melton, K. G. Kulkarni, J.-E. Michels, and F. Zemke, "SQL:2003 Has Been Published", *ACM SIGMOD Record*, Volume 33, Number 1 (2004), pages 119–126.

[Florescu et al. (2000)] D. Florescu, D. Kossmann, and I. Monalescu, "Integrating Keyword Search into XML Query Processing", In *Proc. of the International World Wide Web Conf.* (2000), pages 119–135. Also appears in *Computer Networks*, Vol. 33, pages 119-135.

[Grossman and Frieder (2004)] D. A. Grossman and O. Frieder, *Information Retrieval: Algorithms and Heuristics*, 2nd edition, Springer Verlag (2004).

[Guo et al. (2003)] L. Guo, F. Shao, C. Botev, and J. Shanmugasundaram, "XRANK: Ranked Keyword Search over XML Documents", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (2003), pages 16–27.

[Hristidis and Papakonstantinou (2002)] V. Hristidis and Y. Papakonstantinou, "DISCOVER: Keyword Search in Relational Databases", In *Proc. of the International Conf. on Very Large Databases* (2002), pages 670–681.

[Lamb et al. (1991)] C. Lamb, G. Landis, J. Orenstein, and D. Weinreb, "The ObjectStore Database System", *Communications of the ACM*, Volume 34, Number 10 (1991), pages 50–63.

[Lecluse et al. (1988)] C. Lecluse, P. Richard, and F. Velez, "O2: An Object-Oriented Data Model", In *Proc. of the International Conf. on Very Large Databases* (1988), pages 424–433.

[Manning et al. (2008)] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press (2008).

[Melton (2002)] J. Melton, *Advanced SQL:1999 – Understanding Object-Relational and Other Advanced Features*, Morgan Kaufmann (2002).

[Salton (1989)] G. Salton, *Automatic Text Processing*, Addison Wesley (1989).

[Stonebraker (1986)] M. Stonebraker, "Inclusion of New Types in Relational Database Systems", In *Proc. of the International Conf. on Data Engineering* (1986), pages 262–269.

[Stonebraker and Rowe (1986)] M. Stonebraker and L. Rowe, "The Design of POSTGRES", In *Proc. of the ACM SIGMOD Conf. on Management of Data* (1986), pages 340–355.

[UniSQL (1991)] *UniSQL/X Database Management System User's Manual: Release 1.2*. UniSQL, Inc. (1991).

[Witten et al. (1999)] I. H. Witten, A. Moffat, and T. C. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd edition, Morgan Kaufmann (1999).