CMSC 724*Reading List

Sudarshan S. Chawathe†

Spring 2003

This document will change, so please check it often.

1 Schedule

This schedule is approximate and we may need to move topics around based on how fast we can cover material, feedback from the class, etc. It is almost guaranteed that we will not follow it exactly. *Do not plan major events (trips, weddings, proving P=NP, etc.) based on this schedule.* If you need to verify whether some date is OK, check with me.

#	Date	Material due (numbers refer to the sections below)
01	31 Jan	Introduction, 2.1
02	07 Feb	2.2
03	14 Feb	2.3
04	21 Feb	2.4
05	28 Feb	2.5
06	07 Mar	2.5 (continued) and review
07	14 Mar	5:00pm Take-home midterm assigned.
–	20 May	8:00am Deadline for submitting take-home midterm.
08	21 Mar	midterm discussion
–	28 Mar	Spring Break (no class)
09	04 Apr	2.6
10	11 Apr	2.9
11	18 Apr	Project Discussions
12	25 Apr	2.8
13	02 May	2.8 (continued)
14	09 May	Project Presentations
–	16 May	5:00pm Take-home final assigned.
–	20 May	8:00am Deadline for submitting take-home final.

*http://www.cs.umd.edu/class/spring2003/cmsc724/
†http://www.cs.umd.edu/users/chaw/

2 Topics

The following sections should be updated with details and summaries of the class discussions as we proceed. You should be able to find most of these papers very easily on the Web. (For more details, see Section 3.) You should read the material indicated below *before* the class meeting at which it is due so that you can actively participate in the discussion. You should read the papers critically, noting, for example, the advantages and limitations of the proposed methods. You should be prepared to both ask and answer questions intelligently. The class participation portion of your grade depends on such interactions. More importantly, if you do not do the readings before class, you will not benefit from the classroom discussions (which will assume you have read the material carefully).

2.1 Conjunctive and First-Order Queries

- Chapters 4 and 5 of [AHV95].

2.2 Standard query processing techniques.

- Chapters 15 and 16 of [GMUW02]
- Graefe's survey: [Gra93] Local copy
- The classic System R paper [SAC+79]

2.3 XML Query Languages

- Chapters 4-6 of [ABS99]
- The XPath language specification [CD99]
- The XQuery language specification [BCF+02]
- Two papers on Lore and the Lorel language [MAG+97, AQM+96]
- A paper on the UnQL language [BDHS96]

2.4 Datalog, Recursion, and Negation

- Chapters 14, 15, and 16 of [AHV95]

2.5 Data storage and indexing

- Chapters 12-14 of [GMUW02]

2.6 Describing Semistructured Data

- Chapter 7 of [ABS99]
- Representative Objects [NUWC97] Local copy
- Graph Schemas [BDFS96] Local copy
- DataGuides [GW97] Local copy
- Synopses
- DTD, RDF, XSchema, ...

2.7 Expressiveness and Complexity of Query Languages

- Chapters 16-18 of [AHV95]

2.8 Data Integration

- Chapter 20 of [GMUW02]
- Information Integration Using Logical Views [Ull97] Local copy
- Theory of Answering Queries using Views [Hal00] Local copy

2.9 Streaming Data

- Streaming XPath [PC03] Local copy

3 Resources

The ACM Digital Library[1]: Requires a subscription, but UMD has a site-wide subscription that gives access from all local machines.

The DBLP Bibiliography Server[2] has extremely good coverage of the Database and Logic Programming fields.

ACM SIGMOD[3].

VLDB Foundation[4].

SIGMOD Record[5]

[1] http://www.acm.org/dl/
[2] http://www.purl.org/net/dblp
[3] http://www.acm.org/sigmod/
[4] http://www.vldb.org
[5] http://www.acm.org/sigmod/record/

IEEE Data Engineering Bulletin[6]

Maryland Database Group[7] with pointers to other relevant DB resources.

Modern Information Retrieval [BYRN99]. Use this book for an overview of Information Retrieval. The huge list of references is a big plus.

Readings in Database Systems [SH98]. This collection of papers is typically covered in CMSC 624 and similar courses. It includes many famous papers, such as "the System R paper," "the ARIES paper," and Gray et al.'s locking paper.

Principles of Distributed Database Systems [OV99]. Look here for distributed query optimization, distributed transaction processing, etc.

References

[ABS99] Serge Abiteboul, Peter Buneman, and Dan Suciu. *Data on the Web: From Relations to Semistructured Data and XML.* Morgan Kaufmann, first edition, October 1999.

[AHV95] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases.* Addison-Wesley, 1995.

[AQM+96] S. Abiteboul, D. Quass, J. McHugh, J. Widom, and J. Wiener. The Lorel query language for semistructured data. *Journal of Digital Libraries*, 1(1):68–88, November 1996.

[BCF+02] Scott Boag, Don Chamberlin, Mary F. Fernandez, Daniela Florescu, Jonathan Robie, and Jerome Simeon. XQuery 1.0: An XML query language. W3C Working Draft. `http://www.w3.org/TR/2002/WD-xquery-20021115/`, November 2002.

[BDFS96] P. Buneman, S. Davidson, M. Fernandez, and D. Suciu. Adding structure to unstructured data. Technical Report MS-CIS-96-21, University of Pennsylvania, Computer and Information Science Department, 1996.

[BDHS96] P. Buneman, S. Davidson, G. Hillebrand, and D. Suciu. A query language and optimization techniques for unstructured data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 505–516, Montréal, Québec, June 1996.

[BYRN99] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval.* Addison-Wesley, first edition, May 1999.

[6]http://www.research.microsoft.com/research/db/debull

[7]http://www.cs.umd.edu/areas/db/

[CD99] James Clark and Steve DeRose. XML path language (XPath) version 1.0. W3C Recommendation `http://www.w3.org/`, November 1999.

[GMUW02] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. Prentice-Hall, 2002.

[Gra93] Goetz Graefe. Query evaluation techniques for large databases. *ACM Computing Surveys*, 25(2):73–169, 1993.

[GW97] R. Goldman and J. Widom. DataGuides: Enabling query formulation and optimization in semistructured databases. In *Proceedings of the Twenty-third International Conference on Very Large Data Bases*, Athens, Greece, 1997.

[Hal00] Alon Y. Halevy. Theory of answering queries using views. *SIGMOD Record*, 29(4), December 2000.

[MAG+97] J. McHugh, S. Abiteboul, R. Goldman, D. Quass, and J. Widom. Lore: A database management system for semistructured data. *SIGMOD Record*, 26(3):54–66, September 1997.

[NUWC97] S. Nestorov, J. Ullman, J. Wiener, and S. Chawathe. Representative objects: Concise representations of semistructured, hierarchial data. In *Proceedings of the International Conference on Data Engineering*, pages 79–90, 1997.

[OV99] M. Tamer Ozsu and Patrick Valduriez. *Principles of Distributed Database Systems*. Prentice-Hall, Upper Saddle River, New Jersey, second edition, 1999.

[PC03] Feng Peng and Sudarshan S. Chawathe. XPath queries on streaming data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, San Diego, California, June 2003. To appear. Available at `http://www.cs.umd.edu/~chaw/`.

[SAC+79] P. G. Selinger, M. M. Astrahan, D. D. Chamberlin, R. A. Lorie, and T. G. Price. Access path selection in a relational database management system. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 23–34, 1979.

[SH98] M. Stonebraker and J. Hellerstein, editors. *Readings in Database Systems*. Morgan Kaufmaann, San Francisco, California, third edition, 1998.

[Ull97] Jeffrey D. Ullman. Information integration using logical views. In *Proceedings of the International Conference on Database Theory*, 1997.