CSC2538   TOPICS IN DATABASES: FOUNDATIONS OF XML

Fall 2005

http://www.cs.toronto.edu/~libkin/csc2538

General:

Class meets on Fridays
BA 1200, 3–5pm

Instructor:

Leonid Libkin
libkin@cs.toronto.edu
PT388a
Office hours: by appointment

Prerequisites:

basic knowledges of databases, logic, complexity, theory of computation

Requirements:

Presentation
class notes
small-scale research project
(only for A+)
exercises???

Topics   (both the list and the order will probably change)

The basics of XML: data model, DTDs, XPath, XQuery, etc.

The basics of automata theory: string and tree automata. Logic background: FO and MSO.

Formal model of XML documents: unranked trees.

Automata for unranked trees. Logics for unranked trees; connections between logic and automata; complexity of decision problems.

DTDs: extended CFGs, connections with automata, extended DTDs, automata and MSO.

XPath: a fragment of FO, connections with temporal logics, conditional XPath, containment.

Logics for querying XML, complexity of query evaluation (examples may include ETL, monadic datalog, temporal logics).

XML transformations and tree transducers.

Streaming XML: models, expressiveness, complexity.

XML constraints: keys, foreign keys, consistency.

Student presentations (list of topics TBD but will likely include: typechecking of XML transformations, complexity of XQuery evaluation, automata with counting, tree-walking automata, compressed trees, conjunctive queries for XML, data extraction on the web)

Sources   (links will be added on the webpage, or use google)

- Thomas Schwentick, *Trees, Automata, and XML*, Invited Tutorial at PODS'04.
- Frank Neven, *Automata, Logic, and XML*. Invited talk at CSL 2002.
- Nils Klarlund, Thomas Schwentick and Dan Suciu. *XML: Model, Schemas, Types, Logics, and Queries.* In "Logics for Emerging Applications of Databases", 2003, Springer.
- Leonid Libkin, *Logics for Unranked Trees: An Overview*, Invited talk at ICALP 2005.
- This list will be expanded and links will be added on the webpage.