Webdam Current Research Directions

Serge Abiteboul & Yannis Katsis
INRIA Saclay & ENS Cachan

Presented at LSV Yearly
meeting -- Barbizon 2010

Webdam

Webdam = Web Data Management

1. Distribution: Distributed knowledge
2. Uncertainty: Probabilistic trees
3. Verification and sequencing in Active XML
4. Optimization in Active XML systems
5. The Webdam textbook

http://webdam.inria.fr

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY – ÎLE-DE-FRANCE

Distribution: Distributed Knowledge

Thesis: The management of Web information should be seen as a holistic problem of distributed knowledge management

Model for the information that we manipulate

- XML: data, metadata including annotations
- Access control rights & keys (login, encryption)
- Localization (who provides what, replication)
- Beliefs & uncertain information

And for the knowledge

- Policies for storing/pushing/finding data
- Knowledge for inferring new information; Ontologies

PhD of Alban Galland – with Serge Abiteboul, Meghyn Bienvenue *(U. Paris Sud)* & Marie-Christine Rousset *(Grenoble U.)*

Uncertainty: Probabilistic Trees

Information on the Web is uncertain

Probabilistic trees: syntax for describing a set of trees with a probability distribution

- Attach probabilities to pieces of information

On-going works

- Queries with aggregation
- Richer models based on recursive Markov chains

PhD of Evgeny Kharlamov – with Serge Abiteboul & Pierre Senellart *(Telecom ParisTech)*

Verification & Sequencing in Active XML

Active XML: labeled, unranked trees with embedded function calls

- Evolve in time from results of calls

Verification of temporal properties

On-going work: comparison of various approaches to constrain the sequencing of tasks (workflow)

- Static constraints on the documents
- Guards to function calls (preconditions)
- Automata with data (state-transition)
- Temporal formulas (past-tree LTL)

PhD of Pierre Bourhis – with Serge Abiteboul, Luc Segoufin & Victor Vianu *(UC San Diego)*

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY – ÎLE-DE-FRANCE

5

Optimization in Active XML Systems

Active XML: captures systems of trees exchanging information (via function calls)

Question: equivalence of two systems
- This is at the foundation of optimization

We are studying the boundaries of decidability of equivalence

Yannis Katsis & Serge Abiteboul & Balder ten Cate *(UC Santa Cruz)*

The Textbook: Web Data Management and Distribution
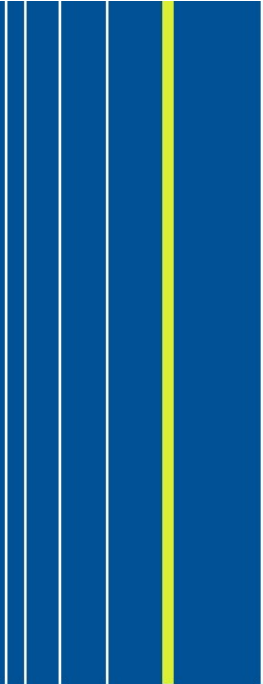
Book for a master's class; mostly system

3 main parts

- Semi-structured Data (tree model, query, typing, transformations)
- Semantic Web (description logics, query, data integration)
- Large-scale Data Computing (web search, P2P, map-reduce)

V0 is out – available on the Web – **http://webdam.inria.fr/textbook**

V1 end of the year

Serge Abiteboul, Ioana Manolescu (INRIA Saclay), Philippe Rigaux (CNAM Paris), Marie-Christine Rousset & Pierre Senellart

MERCI!
THANK YOU!

FRAPAR.