

Webdam - Brainstorming on Foundations of Web Data Management

Serge Abiteboul, INRIA Saclay

Telecom Paris, 08/2009

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
SACLAY - ÎLE-DE-FRANCE

Webdam

| | |
|--------------|--------------------------------------------------------------------------------|
| Topic: | Foundations of Web Data Management European Research Council Advanced Grant |
| Founding: | 2.4 millions Euros |
| Timing: | 2009 – 2013 |
| Institution: | INRIA Saclay – Île-de-France |
| Hosts: | Univ. Paris Sud (Gemo) and ENS Cachan (Dahu) |

Organization

Technical content

The people

Meetings

Editing

Website

Discussion

Technical content

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
SACLAY - ÎLE-DE-FRANCE

Webdam thesis in brief

Management of local data

- Well-accepted model: relational model
- Well-accepted theory: FO & concurrency control & etc.
- Development of systems and theory in parallel

Management of distributed/Web data – not the case

- Rapid software development & no well-accepted model/theory

Illustration

- Relational database course: clean and neat course
- Distributed databases: a series of recipes

The Web is too complex to continue hacking systems with no unifying underlying model

A first shift: distribution in autonomous systems

Web/distributed/peer-to-peer data management

*Focus is on **information** residing on (possibly very many) **autonomous systems***

Information

- Structured data (relations)
- Semistructured (XML, graphs)
- Unstructured (text)
- Metadata
- Knowledge
- Physical data: indices, views, cache...

Focus is more on data management than on semantic Web but border is unclear

A model for Web data management

What should be expected for a such a model?

- Describe query processing, changes, monitoring, communication, service choreography, data integration/exchange, distributed applications
- Road map

What are the issues ?

First attempt – ActiveXML: XML with embedded service calls

- With “local” logic (tree pattern queries): a calculus for Web data
- An algebra
- Optimization
- Verification

A second shift: deduction

Relational system: FO on everyone's desk

Hardwired logic

- Concurrency control with 2PL
- Integrity constraint enforcement
- Access control enforcement

On the Web or in distributed systems: this is not available

Verification of distributed applications

- Work on business processes
- Verification of ActiveXML applications

What is the target?

Improve understanding of Web applications

Facilitate teaching Web data management technology

Improve Web applications: better performance, more reliable, better access control, etc.

Facilitate development of Web applications – programmers productivity

The people

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
SACLAY - ÎLE-DE-FRANCE

Webdam members

Already in: Abiteboul, Segoufin, Vianu, Senellart, Galland, Marinoiu, Bourhis, Kharlamov, ten Catte

Potentially: The members of Gemo and Dahu who are interested

Soon:

1. **Philippe Rigaux** – seconding “détachement” September 2009
2. **Marie-Christine Rousset** – ½ seconding “délégation” September 2009
3. **Yannis Katsis** – PhD UCSD
4. **Alin Tilea** – Engineer
5. **Amélie Marian** – short visit sept/dec 2009

Short visits – passed: Alkis Polizotis, Werner Nutt, Bruno Marnette

Assistant: Marie Domingues → Isabelle Biercewicz

Hiring season

Advertise there are positions

Do it primarily by contacting friends/good groups

Priority for Webdam:

researchers with some experience

Privileged contacts

France

- STREP Fox (Luc Segoufin): Foundations of XML
- ANR DataRing (Patrick Valduriez): Distributed data management
- ANR Docflow (Anca Muscholl): Verification of Active XML

Luc's
presentation

Top research groups

- Israel: Tel Aviv
- US: UCSD, Penn, Washington
- UK: Oxford, Edinburgh

Georg, Tova, Peter, Dan,
Val's presentations

Other ERC projects

- Stefano Ceri: Search computing

Stefano's
presentation

Editing

Foundations of databases

Text book

Target audience: PhD student with theory inclination & researchers

First Foundation

- Abiteboul Hull Vianu – *Alice's book*
- Soon available on the Web

Second Foundation

- New topics and new authors
- Same style – advanced textbook, proofs, homework
- 2 new parts are considered
 - Semistructured data and XML
 - Data integration

License: Probably creative commons on the Web

Web Distributed Management of Data

Text book

Used in courses at Orsay, Dauphine & Telecom ParisTech

Target audience: master students

Authors: Abiteboul, Rigaux, Rousset, Senellart + ...

License: Probably creative commons on the Web

Research in 2009

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
SACLAY - ÎLE-DE-FRANCE

Active XML, verification & workflows

Verification of Active XML systems

- SA, L. Segoufin, V. Vianu
- Continuation of PODS-2008
- Verification of unbounded systems



Victor's
presentation

Alternative ways of specifying data-centric work flows

- SA, **P. Bourhis**, V. Vianu

Active XML - others

Serge's
presentation

The Active XML Artifact Model

- SA, P. Bourhis, A. Galland, B. Marinoiu
- Model for describing distributed activities using business artifacts – data-centric workflow [Time09]

Equivalence and optimization of Active XML systems

- SA, B. ten Catte

Not in Balder's
presentation

Monitoring distributed systems

- SA, **B. Marinoiu, P. Bourhis**
- Satisfiability and relevance of queries for active docs [PODS09]
- P2P Monitoring system [EDBT09]

Distributed XML design

- SA, G. Gottlob, **M. Manna**

Not Active XML

Web data semantics

- M.-C. Rousset

Marie-Christine's
presentation

Probabilistic data

- SA, **E. Kharlamov**, W. Nutt, P. Senellart
- [BDA09]

Pierre's
presentation

Corroboration of imprecise/conflicting data

- SA, **A. Galland**, A. Marian, P. Senellart
- [BSA09]

Social networks and access control

- SA, **A. Galland**, N. Polyzotis

Website : <http://webdam.inria.fr>

