“verification and databases”

ENS CACHAN

informatics mathematics
Inria

cnrs

Research interests

Database theory:

- Modelization
- Query language
  - expressive power
  - evaluation

Verification:

- Specification
- Model checking

Emphasis on the intersection: database driven verification

Research interests

Database theory:

- Modelization
- Query language
  - expressive power
  - evaluation

Verification:

- Specification
- Model checking

Emphasis on the intersection: database driven verification

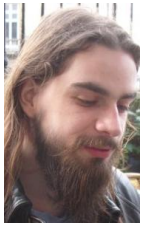
Distributed data management, heterogeneity, redundancy, inconsistency, lack of reliability, automated verification

Current members

Serge Abiteboul
DR INRIA

Luc Segoufin
DR INRIA

Cristina Sirangelo
MCF ENS Cachan



Emilien Antoine
WebDaM

Nadime Francis
ENS-Cachan

Johann Brault-Baron
PDoc INRIA 2013/2014

Arnaud Durand
délégation CNRS 2013/2014

Sylvain Schmitz
Délégation INRIA 2013/2014

Victor Vianu
chaire INRIA 2013/2018

Members 2009-2013

Stéphane Demri
moved to INFINI in 2012

Florent Jacquemard
INRIA @ IRCAM since 2011

Former Ph.Ds

Wojtek Kazana, PDoc, UCSD

Pierre Bourhis, CNRS, Lille

Alban Galland, Ministry of finance

Diego Figueira, PDoc in Edinburgh

Thomas Place, MCF Bordeaux

Camille Vacher

Former PDocs/visitors

Yukiko Kawai, Kyoto, Japan

Gerome Miklau, UMASS

Szymon Toruńczyk, Warsaw

Balder ten Cate, UCSC/logicBox

Daniel Deutch, Beersheba

Yannis Katsis, UCSD

Bruno Marnette, Winton Capital Management

Results

Theme 1: Specification and verification of database driven systems.

Theme 2: Distributed data management.

Theme 3: Tree automata theory.

Theme 1: Verification and databases

1. **Modelization and verification of data centric systems.**
   AXML model, extension of LTL, automata with registers
   XML: ordered, labelled, unranked trees
   standard for Web data
   AXML: XML with embedded function calls

2. **Static analysis of query languages.**
   Decidability, XPath with joins

3. **Static analysis of updates and access control policies.**
   XQuery Update facility, inference algorithms

Zoom: XPath Satisfiability

Path expressions: $\epsilon, \uparrow, \downarrow, \rightarrow, \leftarrow \mid [\phi] \mid \alpha \cup \beta \mid \alpha \cdot \beta \mid \alpha^*$

Node expressions: $a \mid \wedge, \vee, \neg \mid \alpha = \beta \mid \alpha \neq \beta$

Known result: XPath($\uparrow$,$\downarrow$,$\rightarrow$,$\leftarrow$) is undecidable.

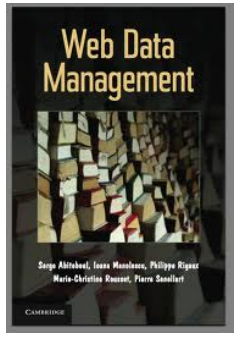
New results: XPath($\uparrow$,$\downarrow$) is decidable.
XPath($\downarrow$,$\rightarrow$) is decidable.

Proof technique: construction of a wqo, non primitive recursive complexity.

PH.D thesis of Diego Figueira, 2010
Also PODS'09, MFCS'09, ICDT'10, STACS'11.

Theme 2: Distributed data management

1. **Distributed knowledge base.**
   Webdamlog, Datalog rules
2. **Probabilistic XML.**
   Probabilistic models, tractability, query evaluation, updates...
3. **Data exchange and Web incomplete information.**
   Restructuring data, handling incompleteness, query evaluation
4. **Enumeration of queries**
   Linear preprocessing, constant delay
5. **http://webdam.inria.fr/Jorge**
   Creative commons and Cambridge University Press



Zoom: WebDamlog

Goal: a language for distributed knowledge management
Approach: an extension of datalog with distribution, update, delegation...

```
$message@$peer($name, "Happy birthday!") :-
    today@my-iphone($d/$m/$y),
    birthday@my-iphone($name, $message, $peer, $d/$m)
```

Novelty: declarative language for distributed computation supporting the exchange of data and knowledge
Contributions:

- Design and theory of Webdamlog PODS11
- Webdamlog system (open source) demo at SIDMOD13
  PhD thesis of A. Galland (2011) and E. Antoine (2013)
  https://github.com/Emilien-Antoine/webdamlog-engine
- Access control DBPL13
- Probabilistic Webdamlog ICDT13

Zoom: Data exchange

Source schema σ, Target schema τ,
Dependencies Σ

Σ: Employee(Name, Phone) $\rightarrow$ Emp(Name, Id), PhoneDir(Id, Phone)

Problems:

- given D over σ, find a E consistent with τ and Σ
- answer queries stated directly over τ

Existing semantics: OWA, CWA

Contributions:

- mixed CWA-OWA semantics
- structural restriction for deciding FO query answering

Cristina Sirangelo (with L. Libkin), 2011

Theme 3: Tree automata

In the other themes, we use extensively tree automata techniques, e.g., in works on XML queries.
In this theme, we study more abstractly tree automata problems, some encountered in the other themes.

1. **Register Automata.**
   Alternation, expressive power, emptiness

2. **Automata with counters.**
   Emptiness, coverability, LTL properties

3. **Subclasses/superclasses of regular tree languages.**
   Expressive power, emptiness decidability

4. **Transformations of tree languages.**
   Tree transducers, Post*, Pre*

Main Grants

ERC WebDaM (Web Data Management): 2009-2013.
PI: Serge Abiteboul.

STREP FoX (Foundations of XML): 2009 - 2012.
With Bojańczyk, Gottlob, Libkin, Marx, Neven and Schwentick
PI: Luc Segoufin.

Others: ARC INRIA, ANR $\cdots$ see document

Objectives

Theme 1: Specification and verification of database driven systems.

- Better complexities
- More powerful models

Theme 2: Distributed data management.

- Management of access control and data localization
- Reasoning with imprecision and inconsistencies

Theme 3: Tree automata theory. Stopped.

Theme 3 bis: Query processing for the Web.

- Query languages for graph databases
- Enumeration of query answers

Query languages for graph databases

thesis of Nadime Francis

Graph database: social networks
web semantic (RDF)
...

Queries: reachability
paths
pattern matching
...

Problems: mix recursion (regular expression) with data.
$\longrightarrow$ simple query languages have hard query evaluation complexities.

Goals: find better query languages, design suitable algorithms, query using views...

Enumeration with constant delay

project with Arnaud Durand and Johann Brault-Baron

Query evaluation: M Database, $\phi(\bar{x})$ query

$$\phi(M) = \{\bar{a} : M \models \phi(\bar{a})\}$$

$\phi(M)$ may be **huge**

We are looking for:

- $O(n)$ precomputation phase $\longrightarrow$ index structure
- enumeration phase: all answers with constant delay between two consecutive outputs

Goal: find when this is (not) possible

$\rightsquigarrow$ restrictions on the class of databases

$\rightsquigarrow$ restrictions on the class of queries

n = size of the database

Dahu during the evaluation period

- 6 Ph.D thesis
- 2 major EU grants: FoX and WebDam
- Top class visitors
- Top quality papers: JACM, TODS, PODS, ICDT, LICS, ICALP, STACS ...