



Probabilistic XML: A Data Model for the Web

Pierre Senellart

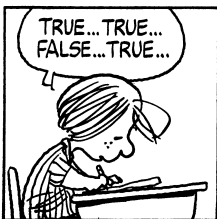
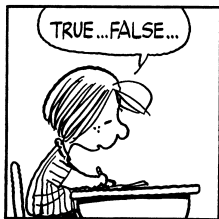
DBWeb

Webdam



Habilitation à diriger les recherches, 13 juin 2012

An Uncertain World





An Uncertain World

Uncertainty is Everywhere

Probabilistic Databases

Probabilistic XML Applications

Probabilistic XML: Models and Complexity

Reconciling Web Data Models with the Actual Web



Uncertain data

Numerous sources of **uncertain data**:

- Measurement errors
- Data integration from contradicting sources
- Imprecise mappings between heterogeneous schemata
- Imprecise automatic process (information extraction, natural language processing, graph mining, etc.)
- Imperfect human judgment
- Lies, opinions, rumors



Uncertain data

Numerous sources of **uncertain data**:

- Measurement errors
- Data integration from contradicting sources
- Imprecise mappings between heterogeneous schemata
- Imprecise automatic process (**information extraction**, natural language processing, graph mining, etc.)
- Imperfect human judgment
- Lies, opinions, rumors

Use case: Web information extraction

instance	iteration	date learned	confidence
<u>arabic, egypt</u>	406	08-sep-2011	(Seed) 100.0
<u>chinese, republic of china</u>	439	24-oct-2011	100.0
<u>chinese, singapore</u>	421	21-sep-2011	(Seed) 100.0
<u>english, britain</u>	439	24-oct-2011	100.0
<u>english, canada</u>	439	24-oct-2011	(Seed) 100.0
<u>english, england001</u>	439	24-oct-2011	100.0
<u>arabic, morocco</u>	422	23-sep-2011	100.0
<u>cantonese, hong kong</u>	406	08-sep-2011	100.0
<u>english, uk</u>	436	19-oct-2011	100.0
<u>english, south vietnam</u>	427	27-sep-2011	99.9
<u>french, morocco</u>	422	23-sep-2011	99.9
<u>greek, turkey</u>	430	07-oct-2011	99.9

Never-ending Language Learning (NELLM, CMU),

<http://rtw.ml.cmu.edu/rtw/kbbrowser/>

Use case: Web information extraction

The screenshot shows the Google Squared Labs interface. At the top, the search query 'comedy movies' is entered in a search box, with 'Square it' and 'Add' buttons to its right. Below the search bar, a table displays search results for 'comedy movies'. The table has columns for 'Item Name', 'Language', 'Director', and 'Release Date'. The first row shows 'The Mask' with language 'English' and director 'Chuck Russell', released on '29 July 1994'. Below the table, two dropdown menus are open, showing alternative values for 'Language' and 'Director' for the selected item. The 'Language' dropdown shows 'English' as the selected option, with other options including 'English Language', 'english, french', and 'Italian Language', all marked as 'Low confidence'. The 'Director' dropdown shows 'Chuck Russell' as the selected option, with other options including 'John R. Dilworth', 'Fiorella Infascelli', and 'Charles Russell', all marked as 'Low confidence'. Each dropdown menu includes a 'Search for more values' link.

Item Name	Language	Director	Release Date
<input type="checkbox"/> The Mask	English	Chuck Russell	29 July 1994
<input type="checkbox"/> Scary M	<input checked="" type="radio"/> English language for the mask www.infibeam.com - all 9 sources »	<input checked="" type="radio"/> Chuck Russell directed by for The Mask www.infibeam.com - all 9 sources »	
<input type="checkbox"/> Superba	Other possible values <input type="radio"/> English Language <i>Low confidence</i> language for Mask www.freebase.com	Other possible values <input type="radio"/> John R. Dilworth <i>Low confidence</i> director for The Mask www.freebase.com	
<input type="checkbox"/> Music	<input type="radio"/> english, french <i>Low confidence</i> languages for the mask www.dvdreview.com	<input type="radio"/> Fiorella Infascelli <i>Low confidence</i> directed by for The Mask www.freebase.com - all 2 sources »	
<input type="checkbox"/> Knocked	<input type="radio"/> Italian Language <i>Low confidence</i> language for The Mask www.freebase.com	<input type="radio"/> Charles Russell <i>Low confidence</i> directed by for The Mask www.freebase.com - all 2 sources »	

Google Squared (terminated), screenshot from [Fink et al., 2011]



Use case: Web information extraction

Subject	Predicate	Object	Confidence
Elvis Presley	diedOnDate	1977-08-16	97.91%
Elvis Presley	isMarriedTo	Priscilla Presley	97.29%
Elvis Presley	influences	Carlo Wolff	96.25%

YAGO, <http://www.mpi-inf.mpg.de/yago-naga/yago> [Suchanek et al., 2007]



Uncertainty in Web information extraction

- The information extraction system is **imprecise**
- The system has some **confidence** in the information extracted, which can be:
 - a **probability** of the information being true (e.g., conditional random fields)
 - an **ad-hoc** numeric confidence score
 - a rough **discrete** level of confidence (low, medium, high)
- What if this uncertain information is not seen as something final, but is used as a source of, e.g., a query answering system?



An Uncertain World

Uncertainty is Everywhere

Probabilistic Databases

Probabilistic XML Applications

Probabilistic XML: Models and Complexity

Reconciling Web Data Models with the Actual Web



Managing uncertainty

Objective

Not to pretend this imprecision does not exist, and manage it as rigorously as possible throughout a long, automatic and human, potentially complex, process

Especially:

- Represent **different forms** of uncertainty
- **Probabilities** are used to measure uncertainty in the data
- Query data and retrieve **uncertain** results
- Allow adding, deleting, modifying data in an **uncertain** way



Managing uncertainty

Objective

Not to pretend this imprecision does not exist, and manage it as rigorously as possible throughout a long, automatic and human, potentially complex, process

Especially:

- Represent **different forms** of uncertainty
- **Probabilities** are used to measure uncertainty in the data
- Query data and retrieve **uncertain** results
- Allow adding, deleting, modifying data in an **uncertain** way



Why probabilities?

- Not the only option: **fuzzy set** theory [Galindo et al., 2005], **Dempster-Shafer** theory [Zadeh, 1986]
- **Mathematically rich** theory, nice semantics with respect to traditional database operations (e.g., joins)
- Some applications already **generate probabilities** (e.g., statistical information extraction or natural language probabilities)
- Naturally arising in case of **conflicting information**, based on the trust in the sources
- In other cases, we “cheat” and pretend that (normalized) **confidence scores** are probabilities: see this as a first-order approximation



Probabilistic relational DBMSs

Numerous studies on the modeling and querying of probabilistic relations:

block-independent disjoint (BIDs) databases [Dalvi et al., 2009] tuples either mutually exclusive or independent, depending on key values

probabilistic c-tables [Imieliński and Lipski, 1984, Green and Tannen, 2006, Senellart, 2007], tuples annotated with arbitrary probabilistic formulas

Two main probabilistic relational DBMS:

Trio [Widom, 2005] Various **uncertainty operators**: unknown value, uncertain tuple, choice between different possible values, with probabilistic annotations

MayBMS [Koch, 2009] Uncertain tables can be constructed using a REPAIR-KEY operator, similar to BIDs, and more elaborate correlations can result from querying



Probabilistic relational DBMSs

Numerous studies on the modeling and querying of probabilistic relations:

block-independent disjoint (BIDs) databases [Dalvi et al., 2009] tuples either mutually exclusive or independent, depending on key values

probabilistic c-tables [Imieliński and Lipski, 1984, Green and Tannen, 2006, Senellart, 2007], tuples annotated with arbitrary probabilistic formulas

Two main probabilistic relational DBMSs:

Trio [Widom, 2005] Various **uncertainty operators**: unknown value, uncertain tuple, choice between different possible values, with probabilistic annotations

MayBMS [Koch, 2009] Uncertain tables can be constructed using a REPAIR-KEY operator, similar to BIDs, and more elaborate correlations can result from querying

Probabilistic relational DBMSs

```
test=# select * from R;
dummy | weather | ground | p
-----+-----+-----+-----
dummy | rain    | wet    | 0.35
dummy | rain    | dry    | 0.05
dummy | no rain | wet    | 0.1
dummy | no rain | dry    | 0.5
(4 rows)
```

Numero

block-in

probabi

```
test=# create table S as
repair key Dummy in R weight by P;
SELECT
```

Two ma

```
test=# select Ground, conf() from S group by Ground;
ground | conf
-----+-----
dry    | 0.55
wet    | 0.45
(2 rows)
```

ons:

either

lues

Senellart,

ulas

lue,

, with

MayBMS [Koch, 2009] Uncertain tables can be constructed using a REPAIR-KEY operator, similar to BIDs, and more elaborate correlations can result from querying



An Uncertain World

Uncertainty is Everywhere

Probabilistic Databases

Probabilistic XML Applications

Probabilistic XML: Models and Complexity

Reconciling Web Data Models with the Actual Web

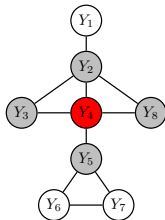
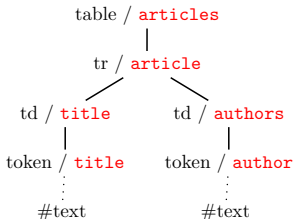


Why probabilistic XML?

- Different typical querying languages: SQL and conjunctive queries vs XPath and tree-pattern queries (possibly with joins)
- Cases where a tree-like model might be appropriate:
 - No schema or few constraints on the schema
 - Independent modules **annotating** freely a content warehouse
 - Inherently tree-like data (e.g., mailing lists, parse trees) with naturally occurring queries involving the descendant axis

Remark

Some results can be transferred from one model to the other. In other cases, connection much trickier!

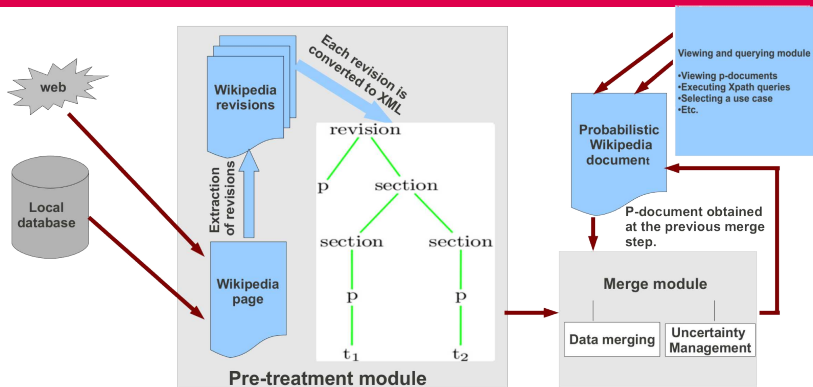


- Annotate HTML Web pages with possible **labels**
- Labels can be learned from a **corpus of annotated documents**
- **Conditional random fields for XML**: estimate **probabilities of annotations** given annotations of neighboring nodes
- Provides **probabilistic labeling** of Web pages



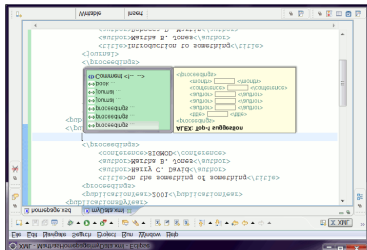
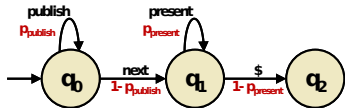
Uncertain version control

[Abdessalem et al., 2011, Ba et al., 2011]



Use trees with probabilistic annotations to represent the **uncertainty in the correctness** of a document under open version control (e.g., Wikipedia articles)

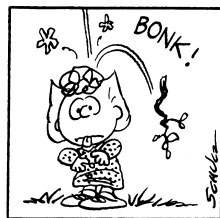
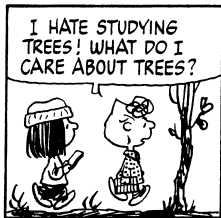




- Transform an XML schema (deterministic top-down tree automaton) into a **probabilistic generator** (probabilistic tree automaton) of XML documents
- Probability distribution **optimal** with respect to a given corpus
- **Application:** Optimal **auto-completions** in an XML editor



Probabilistic XML: Models and Complexity





An Uncertain World

Probabilistic XML: Models and Complexity

Modeling

Querying

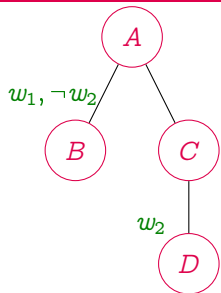
Updating

Reconciling Web Data Models with the Actual Web



A general discrete Probabilistic XML Model

[Abiteboul and Senellart, 2006, Senellart, 2007, Abiteboul et al., 2009]



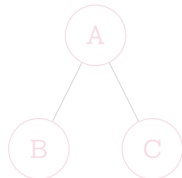
semantics →



$p_1 = 0.06$



$p_2 = 0.70$



$p_3 = 0.24$

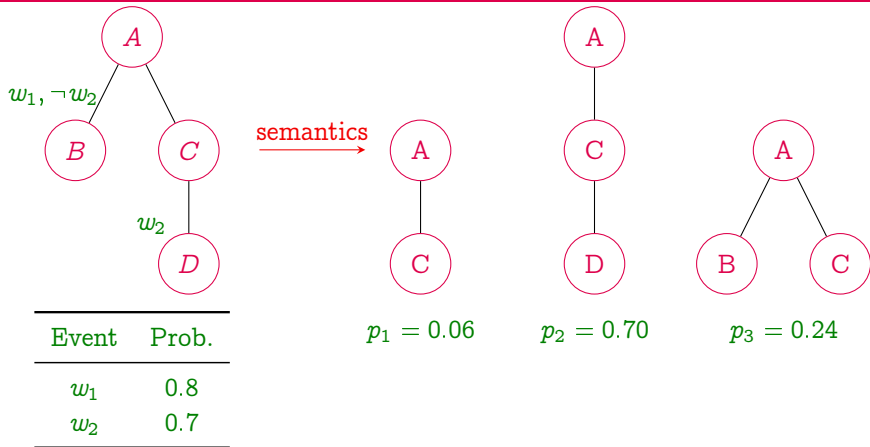
Event	Prob.
w_1	0.8
w_2	0.7

- Expresses arbitrarily complex dependencies
- Analogous to probabilistic c-tables



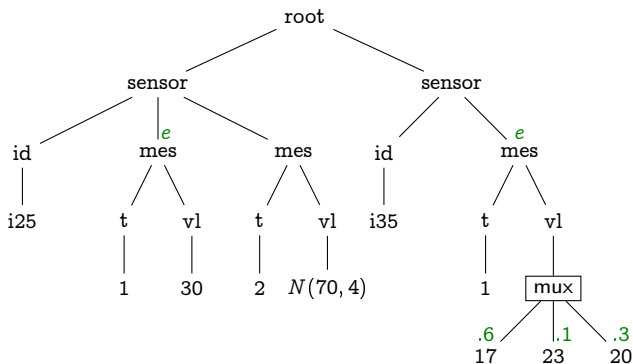
A general discrete Probabilistic XML Model

[Abiteboul and Senellart, 2006, Senellart, 2007, Abiteboul et al., 2009]



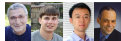
- Expresses **arbitrarily complex** dependencies
- Analogous to probabilistic c-tables





- e : event “it did not rain” at time 1
- mux : mutually exclusive options
- $N(70, 4)$: normal distribution

- Two kinds of dependencies in discrete probabilistic XML: global (e) and local (mux)
- Add **continuous leaves**; non-obvious semantics issues



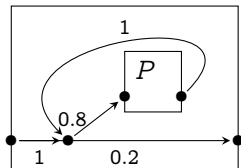


Recursive Markov chains [Benedikt et al., 2010]

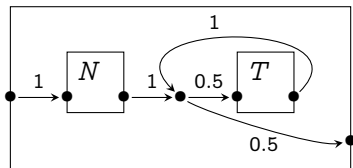
```
<!ELEMENT directory (person*)>
```

```
<!ELEMENT person (name,phone*)>
```

D : directory



P : person



N : name

T : phone



- Probabilistic model that **subsumes** local discrete models
- Allows generating documents of **unbounded** width or depth





An Uncertain World

Probabilistic XML: Models and Complexity

Modeling

Querying

Updating

Reconciling Web Data Models with the Actual Web



Querying probabilistic XML

Semantics of a (Boolean) query = **probability** the query is true:

1. Generate **all possible worlds** of a given probabilistic document
2. In each world, **evaluate the query**
3. **Add up** the probabilities of the worlds that make the query true

EXPTIME algorithm! Can we do better, i.e., can we apply directly the algorithm on the probabilistic document?

We shall talk about **data complexity** of query answering.



Querying probabilistic XML

Semantics of a (Boolean) query = **probability** the query is true:

1. Generate **all possible worlds** of a given probabilistic document (possibly exponentially many)
2. In each world, **evaluate the query**
3. **Add up** the probabilities of the worlds that make the query true

EXPTIME algorithm! Can we do better, i.e., can we apply directly the algorithm on the probabilistic document?

We shall talk about **data complexity** of query answering.



Previously known results

[Kimelfeld et al., 2009, Cohen et al., 2009]

- Tree-pattern queries are evaluable in linear time over local dependencies with a bottom-up, **dynamic programming** algorithm
- Indeed, any **monadic second-order** query is **tractable** over **local dependencies**
- Even trivial queries are **#P-hard** over **global dependencies**
- **Monte-Carlo sampling** works
- **Multiplicative approximation** is also tractable (existence of a FPRAS)

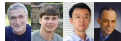


Aggregation queries [Abiteboul et al., 2010]

Aggregate Queries: sum, count, avg, countd, min, max, etc.
Distributions? Possible values? Expected value?

- Computing expected values of sum and count **tractable** with global dependencies; everything else **intractable**
- Computing expected values of every of these aggregate functions **tractable** with local dependencies
- Computing distributions and possible values **tractable** for count, min, max, **intractable** for the others

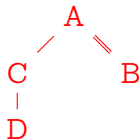
Continuous distributions do not add any more complexity!





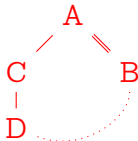
Queries with joins are hard [Kharlamov et al., 2011]

Tree-pattern queries (TP) `/A[C/D]//B`



Tree-pattern queries with joins (TPJ) for `$x` in `$doc/A/C/D`

`return $doc/A//B[.= $x]`



Over local dependencies, **dichotomy** for queries with a single join:

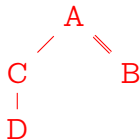
- If equivalent to a join-free query, **linear-time**
- Otherwise, **#P-hard**





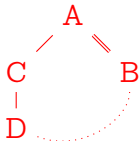
Queries with joins are hard [Kharlamov et al., 2011]

Tree-pattern queries (TP) `/A[C/D]//B`



Tree-pattern queries with joins (TPJ) for `$x` in `$doc/A/C/D`

`return $doc/A//B[.= $x]`



Over local dependencies, **dichotomy** for queries with a single join:

- If equivalent to a join-free query, **linear-time**
- Otherwise, **#P-hard**





Rely on approximations... when needed

[Senellart and Souihli, 2011, Souihli, 2011, Souihli and Senellart, 2012]

- Monte-Carlo is very good at approximating high probabilities
- Sometimes the structure of a query makes the probability of a query easy to evaluate
- For small formulas, naïve evaluation techniques good enough
- Refined approximation methods best when everything else fails

⇒ Use a **cost model** to decide which evaluation algorithm to use!



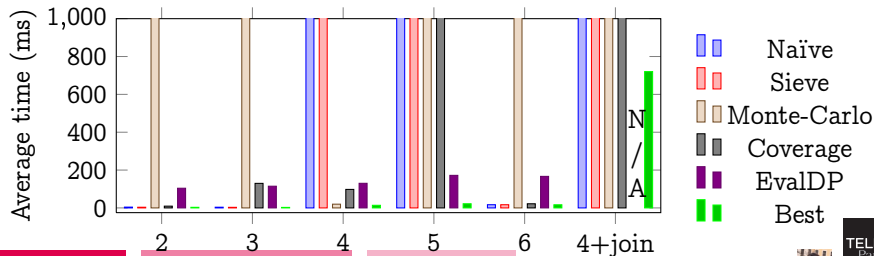


Rely on approximations... when needed

[Senellart and Souihli, 2011, Souihli, 2011, Souihli and Senellart, 2012]

- Monte-Carlo is very good at approximating high probabilities
- Sometimes the structure of a query makes the probability of a query easy to evaluate
- For small formulas, naïve evaluation techniques good enough
- Refined approximation methods best when everything else fails

⇒ Use a **cost model** to decide which evaluation algorithm to use!





An Uncertain World

Probabilistic XML: Models and Complexity

Modeling

Querying

Updating

Reconciling Web Data Models with the Actual Web



The complexity of updates

[Senellart and Abiteboul, 2007, Kharlamov et al., 2010]

Updates defined by a query (cf. XUpdate, XQuery Update).

Semantics: for all matches of a query, insert or delete a node in the tree at a place located by the query.

Results

- Most updates are **intractable** with local dependencies: the result of an update can require an exponentially larger representation size
- Insertions with a for-each-match semantics are **tractable** with arbitrary dependencies; deletions are **intractable**
- Some insert-if-there-is-a-match operations **tractable** for local dependencies but not for arbitrary dependencies



Reconciling Web Data Models with the Actual Web





An Uncertain World

Probabilistic XML: Models and Complexity

Reconciling Web Data Models with the Actual Web

Applying Probabilistic XML to Web data

Formal Models for Web Content Acquisition



Applying Probabilistic XML to Web data

- Current **mismatch** (in my research, but also in general) between:
 - Applications that produce and have to deal with uncertainty:
 - Information extraction
 - Graph mining
 - Natural language parsing
 - Data integration
 - etc.
 - Probabilistic database management systems
- Analogy to **data management in the 1960's**:
 - Each data management application (accounting, cataloging, etc.) was developing its own way of managing data
 - Each application dealing with imprecise data is developing its own way of managing imprecision
- Probabilistic DBMSs are becoming more and more mature – time to put them to use for **real applications**?

Applying Probabilistic XML to Web data

- Current **mismatch** (in my research, but also in general) between:
 - Applications that produce and have to deal with uncertainty:
 - Information extraction
 - Graph mining
 - Natural language parsing
 - Data integration
 - etc.
 - Probabilistic database management systems
- Analogy to **data management in the 1960's**:
 - Each data management application (accounting, cataloging, etc.) was developing its own way of managing data
 - Each application dealing with imprecise data is developing its own way of managing imprecision
- Probabilistic DBMSs are becoming more and more mature – time to put them to use for **real applications**?



An Uncertain World

Probabilistic XML: Models and Complexity

Reconciling Web Data Models with the Actual Web

Applying Probabilistic XML to Web data

Formal Models for Web Content Acquisition



Formal Models for Web Content Acquisition

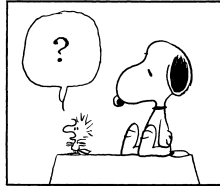
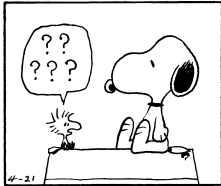
- Current **mismatch** between:
 - Actual Web mining and Web data management applications:
 - Web graph mining
 - wrapper induction
 - Web crawling
 - Research on the foundations of Web data management
- Plan: **bridge the two**
 - Use **formal models for data exchange** [Senellart and Gottlob, 2008, Gottlob and Senellart, 2010] to perform wrapper induction or ontology matching
 - Use **static analysis techniques for JavaScript** to perform deep Web data extraction [Benedikt et al., 2012b]
 - Use results of formal studies of the **containment of recursive query languages** [Benedikt et al., 2011, 2012a] to optimize query answering over the deep Web



Formal Models for Web Content Acquisition

- Current **mismatch** between:
 - Actual Web mining and Web data management applications:
 - Web graph mining
 - wrapper induction
 - Web crawling
 - Research on the foundations of Web data management
- Plan: **bridge the two**
 - Use **formal models for data exchange** [Senellart and Gottlob, 2008, Gottlob and Senellart, 2010] to perform wrapper induction or ontology matching
 - Use **static analysis techniques for JavaScript** to perform deep Web data extraction [Benedikt et al., 2012b]
 - Use results of formal studies of the **containment of recursive query languages** [Benedikt et al., 2011, 2012a] to optimize query answering over the deep Web

Questions?



Talel Abdessalem, M. Lamine Ba, and Pierre Senellart. A probabilistic XML merging tool. In *Proc. EDBT*, pages 538–541, Uppsala, Sweden, March 2011. Demonstration.

Serge Abiteboul and Pierre Senellart. Querying and updating probabilistic information in XML. In *Proc. EDBT*, pages 1059–1068, Munich, Germany, March 2006.

Serge Abiteboul, Benny Kimelfeld, Yehoshua Sagiv, and Pierre Senellart. On the expressiveness of probabilistic XML models. *VLDB Journal*, 18(5):1041–1064, October 2009.

Serge Abiteboul, T-H. Hubert Chan, Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Aggregate queries for discrete and continuous probabilistic XML. In *Proc. ICDT*, pages 50–61, Lausanne, Switzerland, March 2010.

Serge Abiteboul, T-H. Hubert Chan, Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Capturing continuous data and answering aggregate queries in probabilistic XML. *ACM Transactions on Database Systems*, 36(4), 2011.

Serge Abiteboul, Yael Amsterdamer, Daniel Deutch, Tova Milo, and Pierre Senellart. Finding optimal probabilistic generators for XML collections. In *Proc. ICDT*, Berlin, Germany, March 2012a.

Serge Abiteboul, Yael Amsterdamer, Tova Milo, and Pierre Senellart. Auto-completion learning for XML. In *Proc. SIGMOD*, pages 669–672, Scottsdale, USA, May 2012b. Demonstration.

M. Lamine Ba, Talel Abdessalem, and Pierre Senellart. Towards a version control model with uncertain data. In *Proc. PIKM*, Glasgow, United Kingdom, October 2011.

Michael Benedikt, Evgeny Kharlamov, Dan Olteanu, and Pierre Senellart. Probabilistic XML via Markov chains. *Proceedings of the VLDB Endowment*, 3(1):770–781, September 2010. Presented at the VLDB 2010 conference, Singapore.

Michael Benedikt, Georg Gottlob, and Pierre Senellart. Determining relevance of accesses at runtime. In *Proc. PODS*, pages 211–222, Athens, Greece, June 2011.

Michael Benedikt, Pierre Bourhis, and Pierre Senellart. Monadic datalog containment. In *Proc. ICALP*, pages 79–91, Warwick, United Kingdom, July 2012a.

Michael Benedikt, Tim Furché, Andreas Savvides, and Pierre Senellart. ProFoUnd: Program-analysis-based form understanding. In *Proc. WWW*, pages 313–316, Lyon, France, April 2012b. Demonstration.

Sara Cohen, Benny Kimelfeld, and Yehoshua Sagiv. Running tree automata on probabilistic XML. In *PODS*, 2009.

- Nilesh Dalvi, Christopher Ré, and Dan Suciu. Probabilistic databases: Diamonds in the dirt. *Communications of the ACM*, 52(7), 2009.
- Robert Fink, Andrew Hogue, Dan Olteanu, and Swaroop Rath. SPROUT²: a squared query engine for uncertain web data. In *SIGMOD*, 2011.
- José Galindo, Angelica Urrutia, and Mario Piattini. *Fuzzy Databases: Modeling, Design And Implementation*. IGI Global, 2005.
- Georg Gottlob and Pierre Senellart. Schema mapping discovery from data instances. *Journal of the ACM*, 57(2), January 2010.
- Todd J. Green and Val Tannen. Models for incomplete and probabilistic information. In *Proc. EDBT Workshops, IIDB*, Munich, Germany, March 2006.
- Tomasz Imieliński and Witold Lipski. Incomplete information in relational databases. *Journal of the ACM*, 31(4):761–791, 1984.

Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Updating probabilistic XML. In *Proc. Updates in XML*, Lausanne, Switzerland, March 2010.

Evgeny Kharlamov, Werner Nutt, and Pierre Senellart. Value joins are expensive over (probabilistic) XML. In *Proc. LID*, pages 41–48, Uppsala, Sweden, March 2011.

Benny Kimelfeld, Yuri Kosharovsky, and Yehoshua Sagiv. Query evaluation over probabilistic XML. *VLDB Journal*, 18(5), 2009.

Christoph Koch. MayBMS: A system for managing large uncertain and probabilistic databases. In Charu Aggarwal, editor, *Managing and Mining Uncertain Data*. Springer, 2009.

Pierre Senellart. *Comprendre le Web caché. Understanding the Hidden Web*. PhD thesis, Université Paris XI, Orsay, France, December 2007.

Pierre Senellart and Serge Abiteboul. On the complexity of managing probabilistic XML data. In *Proc. PODS*, pages 283–292, Beijing, China, June 2007.

Pierre Senellart and Georg Gottlob. On the complexity of deriving schema mappings from database instances. In *Proc. PODS*, pages 23–32, Vancouver, Canada, June 2008.

Pierre Senellart and Asma Souihli. ProApproX: A lightweight approximation query processor over probabilistic trees. In *Proc. SIGMOD*, pages 1295–1298, Athens, Greece, June 2011.
Demonstration.

Pierre Senellart, Avin Mittal, Daniel Muschick, Rémi Gilleron, and Marc Tommasi. Automatic wrapper induction from hidden-Web sources with domain knowledge. In *Proc. WIDM*, pages 9–16, Napa, USA, October 2008.

- Asma Souihli. Efficient query evaluation over probabilistic XML with long-distance dependencies. In *EDBT/ICDT PhD Workshop*, 2011.
- Asma Souihli and Pierre Senellart. Optimizing approximations of dnf query lineage in probabilistic xml, April 2012. Preprint.
- Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: A core of semantic knowledge. Unifying WordNet and Wikipedia. In *WWW*, pages 697–706, 2007. ISBN 978-1-59593-654-7.
- Jennifer Widom. Trio: A system for integrated management of data, accuracy, and lineage. In *CIDR*, 2005.
- Lotfi A. Zadeh. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, 7(2), 1986.