A Predictive Query Engine for Probabilistic XML

Asma Souihli
Pierre Senellart

Probabilistic Data

Integration

Example: Probabilistic XML documents

A model that uses conjunctive propositional formulas to express the uncertainty related to the data

Probabilistic Data

Extraction

The probability of the *gmail* address is:

\[
\Pr(gmail.com) = \Pr(C_1 \lor C_2 \lor C_3 \lor C_4)
\]

Probabilities of the satisfying assignments for the DNF (lineage formula):

\#P-Hard problem

No polynomial time algorithm for the exact solution if \(P \neq NP\)

\#P problems ask "how many" rather than "are there any"

How many graph coloring using \(k\) colors are there for a particular graph \(G\)?

Context Uncertainty

Example

Repository

Employee

Details

Distribution

Asma Souihli

Name

t_1 e_8 Contact

e-mail souihli@gmail.com

t_2 e_9 Contact

e-mail souihli@gmail.com

 XPath query Q1:

\[/Employee[Name="Asma Souihli"] // e-mail / text()\]

Matches in the database:

gmail.com: \(e_2 \land e_8 \land e_1\) \(C_1\)

gmail.com: \(e_2 \land e_8 \land e_6\) \(C_2\)

gmail.com: \(e_2 \land e_9 \land e_{10}\) \(C_3\)

gmail.com: \(e_2 \land e_9 \land e_6\) \(C_4\)

Approach

ProApproX reveals its originality through the following major features:

- A broader range of XPath queries
- A more general data model
- A cost model for a variety of probability evaluation algorithms (including Monte Carlo and the Self-Adjusting Coverage Algorithm)
- Lineage Decomposition into independent computational units
- Custom-made error bound \(\varepsilon\) and confidence \(\delta\) for the desired probabilistic approximation
- Well-grounded propagation mechanisms of \(\varepsilon\) and \(\delta\) between computational units
- An exploration of the space of evaluation plans based on the proposed cost model

ProApproX 2.0.

(Processing)

Lineage Extraction

User Interface

Query Translation

Result \(\Pr(Q)\)

PrXML database

User input:

XPath Query

Q

ANSWER

Compilation

Exploration (best execution plan)

Computation

Lineage preprocessing

System Architecture