



UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Andrin Betschart, Martin Leimer

**Prof. Dr. Michael Böhlen**  
Professor  
Phone +41 44 635 43 33  
Fax +41 44 635 68 09  
boehlen@ifi.uzh.ch

Zürich, February 1, 2013

**Master-Project (18 ECTS)**

**Topic: Lineage in Query Evaluation on Temporal Probabilistic Databases**

The most important issue to deal with in query evaluation in temporal probabilistic databases is the correct computation of the probabilities of the result tuples. The approach adopted in this assignment involves exploiting trio-lineage for confidence computation [3]. Trio-Lineage is a formula indicating the way the result tuples have been produced from the argument tuples. It is recorded as a supplementary attribute in a temporal probabilistic relation and it demands that each tuple is associated with a unique identifier. It is produced through the appropriate combination of tuple identifiers using disjunction, conjunction and negation operators.

**p**

id	B	C	P	T	$\lambda$
$p_1$	Ann	Zurich	0.0024	[4,9)	$p_1$
$p_2$	John	Zurich	0.0778	[6,11)	$p_2$

**w**

id	D	E	P	T	$\lambda$
$w_1$	Sun	Zurich	0.0576	[7,12)	$w_1$

(a) A Temporal Probabilistic Database  $D^{TP}$

The goal of this master project is the understanding of the query evaluation process in temporal probabilistic databases and its implementation in the kernel of PostgreSQL so that it offers native support. This process demands:

- The adaptation of the temporal primitives [1, 2], already known in temporal databases as a means to adjust the time intervals, so that the trio-lineage of each tuple is also considered.
- The extension of the conventional, non-temporal and non-probabilistic, operators so that they are able to correctly compute the trio-lineage of the result tuples.
- The definition of a function appropriate for the computation of the probability of each result tuple using the corresponding lineage.



$$p^{TP} \Phi^L w^{TP}$$

B	C	P	T	$\lambda$
Ann	Zurich		[7,9)	$p_1$
John	Zurich		[6,7)	$p_2$
John	Zurich		[7,11)	$p_2$

$$w^{TP} \Phi^L p^{TP}$$

D	E	P	T	$\lambda$
Sun	Zurich		[7,9)	$w_1$
Sun	Zurich		[7,11)	$w_1$
Sun	Zurich		[11,12)	$w_1$

(b) Adaptation of the temporal aligner

$$p \bowtie^{TP} w$$

	B	C	D	P	T	$\lambda$
$r_1$	Ann	Zurich	Sun		[7,9)	$p_1 \wedge w_1$
$r_2$	John	Zurich	Sun		[7,11)	$p_2 \wedge w_1$

(c) The result of a temporal probabilistic join

### Tasks

1. Implementation of the temporal primitives [1, 2]
2. Definition and representation of the lineage formula [3]
3. Computation of confidence values using lineage [3]
4. Experimental evaluation of the query evaluation process
5. Technical Report (10-20 pages)

### References

- [1] Anton Dignös, Michael H. Böhlen, and Johann Gamper. Temporal alignment. In *SIGMOD Conference*, pages 433–444, 2012.
- [2] Anton Dignös, Michael H. Böhlen, and Johann Gamper. Query time scaling of attribute values in interval timestamped databases. In *ICDE*, 2013.
- [3] Anish Das Sarma, Martin Theobald, and Jennifer Widom. Exploiting lineage for confidence computation in uncertain and probabilistic databases. In *ICDE*, 2008.

**Supervisor:** Katerina Papaioannou (papaioannou@ifi.uzh.ch)

**Start date:** 15-04-2013

**End date:** 12-10-2013

University of Zürich  
Department of Informatics

Prof. Dr. Michael Böhlen