



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

Ekaterina Kuleshova

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

Zürich, April 1, 2011

Facharbeit in Informatik (9KP)
Datenbanktechnologie

Topic: Provenance in Temporal Databases

Interval timestamped temporal databases are databases that store the history of the data. With each record of the data a time-interval is associated during which the data is considered to be valid. The presence of this timestamp attribute represents a major challenge when the data needs to be manipulated, since the data is not in 1st normal form.

For this reason temporal operators, i.e., operators that manipulate temporal data have been defined based on snapshot reducibility (SR) using non-temporal operators. In particular the result of a temporal operator at each point in time needs to be equivalent to the result of the corresponding non-temporal operator applied to the data valid at that specific point in time.

Such an approach only defines the output on single time-points, the data model however is based on time-intervals, so time-points need to be merged into intervals. To do so the concept provenance and lineage of the data can be used. More specifically single time-points in the result are merged into an interval, when they have value-equivalent data and the same provenance and lineage information.

The aim of the work is to work out provenance and lineage for temporal data as follows:

1. Study of literature [5, 1] about interval-timestamped temporal databases and their operation.
2. Study of literature [6, 2, 3] about provenance and lineage of non-temporal data.
3. Work out examples of the 6 basic temporal relational operators with their corresponding provenance and lineage.



4. Work out examples of provenance and lineage on composite temporal queries, e.g., aggregation followed by selection.
5. Work out and describe the differences between provenance of non-temporal data and of temporal data.

The expected outcome is a report where the various aspects and challenges of the work are described. The report should focus on the result and outcomes of the work and be structured as a scientific paper. At the end of the work the student should give a 10 min. presentation at one of the group meetings.

The student is encouraged to develop the report in an incremental fashion during the working period. There will be weekly or biweekly meetings with the supervisor, where the student sends the current state of the report one day before the meeting to the supervisor.

References

- [1] M. H. Böhlen, C. S. Jensen, and R. T. Snodgrass. Temporal statement modifiers. *ACM Trans. Database Syst.*, 25(4):407–456, 2000.
- [2] Y. Cui, J. Widom, and J. L. Wiener. Tracing the lineage of view data in a warehousing environment. *ACM Trans. Database Syst.*, 25(2):179–227, 2000.
- [3] T. J. Green, G. Karvounarakis, and V. Tannen. Provenance semirings. In L. Libkin, editor, *PODS*, pages 31–40. ACM, 2007.
- [4] L. Liu and M. T. Özsu, editors. *Encyclopedia of Database Systems*. Springer US, 2009.
- [5] N. A. Lorentzos. Period-stamped temporal models. In Liu and Özsu [4], pages 2094–2098.
- [6] W. C. Tan. Provenance. In Liu and Özsu [4], page 2202.

Supervisor: Anton Dignös

Start date: April 15, 2011

End date: October 15, 2011

University of Zürich
Department of Informatics

Prof. Dr. Michael Böhlen