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.

<table>
<thead>
<tr>
<th>id</th>
<th>B</th>
<th>C</th>
<th>P</th>
<th>T</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>Ann</td>
<td>Zurich</td>
<td>0.0024</td>
<td>[4, 9]</td>
<td>p1</td>
</tr>
<tr>
<td>p2</td>
<td>John</td>
<td>Zurich</td>
<td>0.0778</td>
<td>[6, 11]</td>
<td>p2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>id</th>
<th>D</th>
<th>E</th>
<th>P</th>
<th>T</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>Sun</td>
<td>Zurich</td>
<td>0.0576</td>
<td>[7, 12]</td>
<td>w1</td>
</tr>
</tbody>
</table>

(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.
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


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