Fine Grain Provenance Using Temporal Databases
Outline of the talk

• Use case: Classic management of patient data
  • Data types, queries
  • History
  • Security and context information
• Fine grain provenance – I
• Smart management of patient data
  • Facts, knowledge, and information
  • The model
  • Classification and customization
• Fine grain provenance - II
• Implementation details
• Conclusion
Classic Management of Patient Data

Data Types

- Structured Data – SQL
- Semi structured data – XML
  - HL7 - Health Level-7
- DICOM - Digital Imaging and Communications in Medicine
- Text
- Any mix

Data Manipulation and (continuous) queries

- SQL 92 and 99
- XQuery
  - HL7 verbs
- DICOM verbs
  - Text processing verbs
  - Mixed use of languages

June 2011
TaPP 2011
History

• Data management for patient history
  • No extended data model
    • Simplifies programming significantly
  • Standard update, insert, delete

• Queries
  • The current values
  • The values/images at a specific time
  • The values/images as seen at a specific time
  • The evolution of values/images
Security and Context Information

- All queries and temporal queries support (fine grain) security
  - A doctor/nurse can only access data from patients s/he is currently treating
- Additional information recorded by the data base
  - The transactional context of any change or query
  - The transactional context includes host, database/OS user, program
Fine Grain Provenance - I

- The database is able to answer the following questions
  - What was a single or set of values at a given time – from the current perspective?
    - What was a single or set of values at a given time from an earlier perspective – imported to deal with corrections
  - What is the history of a single or a set of values
  - Was a value ever corrected?
    - What is the history of correction?
  - Who was responsible for providing/deleting a value?
  - Which program created the value?
  - Who looked at specific values?
Smart Management of Patient Data

• The issue:
  • Rapidly increasing amount and complexity of data
  • Rapidly increasing amount and complexity domain knowledge
  • Data and knowledge have grown way beyond the capacity of a human cognitive system

• A solution
  • Capture knowledge and personal preferences
    • Vocabularies, rules/models, classifications, customizations
  • Capture facts – as done in classic support
  • Transform data (facts) into information using captured knowledge
    • Alert medical personnel about time critical adverse conditions
The Model

Facts

Knowledge

Information

Patient Care
Applications**

Online Protocols

Raw data - indiscriminate

Quantitative

Near real time inference

Based on

Information - selective

Qualitative*

Online Protocols

Alerts

* Qualitative information is preferred by the human cognitive system

** The application is as declarative as possible
Use Case - Updated

- New functions
Information and Incidents

• Information is created as soon as new data/facts or new knowledge become available
  • The information is a compact and qualitative representation of important facts
    • The temperature is critical
    • The blood chemistry indicates a high probability of a cardiac arrest
    • The information has a high uncertainty, additional tests are recommended
  • Information is bundled as incidents
    • Alert is issued for time critical information
  • Doctors can review the status of the patient on a qualitative level
    • What is important; i.e., show incidents with certain characteristics
    • Show the history of selected incidents
      • Is the patient improving as expected?
    • If needed the doctor can also look at the quantitative data
Fine Grain Provenance - II

- Full auditing and tracking of facts
  - Implies full auditing and tracking of information
- Full Description and versioning of
  - Knowledge – rules, queries, model, programs,..
    - Who developed/tested/deployed/changed the knowledge elements and when
  - Classifications
    - Who developed/tested/deployed/changed the classification and when
  - Customizations
    - Who deployed/changed the customization
- The evolution of the information is now visible
  - What are the facts and knowledge behind information and incidents?
  - Do I accept the information?
  - Why did a colleague come to a (different) conclusion?
  - Why was the information (diagnosis) changed?
Conclusions

• Databases support management of and access to a wide variety of data
• Temporal databases provide full support for auditing and tracking – no user programming required
• Adding knowledge management to data management provides full support for provenance - no user programming is required
Questions
Comments
Suggestions
Read Consistency - Oracle’s Flashback

- One of the main features of Oracle is consistent read
  - No read locks are taken
  - Instead data is read as of a point in time in the past before all uncommitted changes (using undo)
- Flashback extends CR to be able to read data as of a point in time in the recent past (using undo)
- Total Recall extends flashback to go back far in the past
- Using flashback, it is possible to see data/information/knowledge as it was at any point in time, providing the main building block for provenance
Temporal Database Support – Oracle’s Total Recall

• Total recall provides a way to enable transaction time history on a table for a specified retention
• Using total recall it is possible to do flashback queries
• “As of” queries enable the user to read a row/table as of a point in time
• “Versions” enable the user to get all committed versions of a row/table between a range of time
  • Provides the transaction start/end time of version, transaction context of creator of version
• Audit used for tracking queries
• Valid time support can also be added in future
A Classification Model

Uniform classification of data
- Value: Normal, guarded, serious, critical
- Urgency: Stat, ASAP, none

Uniform classification of change
- Type: deteriorating, improving
- Rate: rapid, slow

Statistical temporal change model
- Patient is not improving as expected by model $M_1$

Uniform classification simplifies queries
- Find all patients with critical condition lasting more than 2 hrs in the last 5 years
- Identify important incidences/adverse conditions
Classification - Design Principles

Uniform classification
- Simplifies aggregating elementary quantitative information into highly compact representation
- Reduces the number of queries, rules, and models significantly
- A vital is deteriorating fast
- The patient does not improve as expected

Personalized classification rules
- Adjust to the preferences of a group, a doctor, or specific condition of a patient
- Adjusts to the specific situation of a patient

Classification Methods
- Decision tables, rules, models, manual
### Classification With a Decision Table

<table>
<thead>
<tr>
<th></th>
<th>Lower Range</th>
<th></th>
<th></th>
<th></th>
<th>Upper Range</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical</td>
<td>Serious</td>
<td>Guarded</td>
<td>Normal</td>
<td>Critical</td>
<td>Serious</td>
<td>Guarded</td>
<td>Normal</td>
</tr>
<tr>
<td>TEMPERATURE</td>
<td>34.5</td>
<td>36</td>
<td>37</td>
<td>37.0</td>
<td>38.4</td>
<td>38.4</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>HEART_RATE</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>60</td>
<td>100</td>
<td>100</td>
<td>125</td>
<td>150</td>
</tr>
<tr>
<td>SYSTOLIC_BP</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>90</td>
<td>140</td>
<td>140</td>
<td>160</td>
<td>190</td>
</tr>
<tr>
<td>DIASTOLIC_BP</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>60</td>
<td>90</td>
<td>90</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>MEAN_ARTERIAL_PRESSURE</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>70</td>
<td>105</td>
<td>105</td>
<td>110</td>
<td>115</td>
</tr>
<tr>
<td>RESPIRATORY_RATE</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>14</td>
<td>26</td>
<td>26</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>OXYGEN_SATURATION</td>
<td>80</td>
<td>85</td>
<td>90</td>
<td>90</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEIGHT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EKG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>3</td>
<td>4</td>
<td>4.0</td>
<td>6.0</td>
<td>6</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>CI</td>
<td>2.2</td>
<td>2.6</td>
<td>2.6</td>
<td>4.2</td>
<td>4.2</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>SVR</td>
<td>600</td>
<td>700</td>
<td>800</td>
<td>800</td>
<td>1200</td>
<td>1200</td>
<td>1400</td>
<td>1600</td>
</tr>
<tr>
<td>CWP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRA_ABD_PRESSURE</td>
<td>52</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Columns Guarded and Normal contain intentionally the same information.