Management of Event Logs

1st KAOS Project Meeting
Process Mining

picture by Wil van der Aalst
Process Mining

picture by Wil van der Aalst
Actual Reality
Actual Reality

<table>
<thead>
<tr>
<th>ID</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alifah Syamsiyah</td>
</tr>
<tr>
<td>2</td>
<td>Marco Montali</td>
</tr>
<tr>
<td>3</td>
<td>Diego Calvanese</td>
</tr>
<tr>
<td>4</td>
<td>Wil van der Aalst</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Organizer</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>666</td>
<td>BPM 2015</td>
<td>2</td>
<td>2015-02-14 01:00:00</td>
</tr>
<tr>
<td>667</td>
<td>Caise 2015</td>
<td>4</td>
<td>2015-03-06 01:00:00</td>
</tr>
<tr>
<td>668</td>
<td>ER 2015</td>
<td>4</td>
<td>2015-03-26 01:00:00</td>
</tr>
<tr>
<td>669</td>
<td>EDOC 2015</td>
<td>2</td>
<td>2015-04-05 03:00:00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>CT</th>
<th>User</th>
<th>Conf</th>
<th>Type</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ontop at Work</td>
<td>2015-03-02 15:09:35</td>
<td>1</td>
<td>669</td>
<td>FP</td>
<td>RX</td>
</tr>
<tr>
<td>2</td>
<td>A Survey of Web Services</td>
<td>2015-03-02 12:36:01</td>
<td>3</td>
<td>668</td>
<td>SP</td>
<td>RX</td>
</tr>
<tr>
<td>3</td>
<td>The Definitive Guide for BPM</td>
<td>2015-03-04 13:36:20</td>
<td>1</td>
<td>666</td>
<td>FP</td>
<td>AB</td>
</tr>
</tbody>
</table>
IEEE XES standard for event logs

- Based on XML
- Minimalistic
- Data+metadata

Figure 2.1: The UML 2.0 class diagram for the complete meta-model for the XES standard

- One examination in which the x-ray machine is employed
- One visit of the website, by one specific user

Tag name for the trace object in the XML serialization of XES:

No XML attributes are defined for the trace tag.

2.1.3 Event

Every trace contains an arbitrary number (may be empty) of event objects. Events represent atomic granules of activity that have been observed during the execution of a process. As such, an event has no duration. Examples of an event are:

- Recording the client's personal information in the database has been completed

XES / Version 2.0
# Expected Reality

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Activity</th>
<th>Timestamp</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Submit paper</td>
<td>01-07-2015:10.05</td>
<td>Anna</td>
</tr>
<tr>
<td></td>
<td>Invite reviewer</td>
<td>03-07-2015:12.00</td>
<td>Budi</td>
</tr>
<tr>
<td></td>
<td>Get review</td>
<td>10-07-2015:16.10</td>
<td>Rudy</td>
</tr>
<tr>
<td></td>
<td>Acceptance</td>
<td>12-07-2015:15.00</td>
<td>Rudy</td>
</tr>
<tr>
<td></td>
<td>Submit final paper</td>
<td>19-07-2015:19.15</td>
<td>Anna</td>
</tr>
<tr>
<td>2</td>
<td>Submit paper</td>
<td>02-07-2015:17.19</td>
<td>John</td>
</tr>
<tr>
<td></td>
<td>Invite reviewer</td>
<td>03-07-2015:12.00</td>
<td>Tiara</td>
</tr>
<tr>
<td></td>
<td>Get review</td>
<td>11-07-2015:15.45</td>
<td>Clara</td>
</tr>
<tr>
<td></td>
<td>Rejection</td>
<td>12-07-2015:15.00</td>
<td>Clara</td>
</tr>
</tbody>
</table>
XES standard for event logs

```xml
<log xes.version="1.0" xes.features="nested-attributes">
  <trace>
    <string key="concept:name" value="1" />
    <event>
      <string key="concept:name" value="register request" />
      <date key="time:timestamp" value="2010-12-30T11:02:00.000+01:00" />
    </event>
  </trace>
</log>
```
Understanding Reality…
Impedance Mismatch

From here…
Impedance Mismatch

...to there!
Key Issues

- How to resolve the “impedance mismatch”?
- How to get a “view” of the data tailored to process mining?
Key Issues

• Need to resolve a second impedance mismatch problem!

• From here…
Key Issues

• …To there!

Figure 2.1: The UML 2.0 class diagram for the complete meta-model for the XES standard

3

• One examination in which the x-ray machine is employed
• One visit of the website, by one specific user

Tag name for the trace object in the XML serialization of XES: <trace>
No XML attributes are defined for the <trace> tag.

2.1.3 Event

Every trace contains an arbitrary number (may be empty) of event objects. Events represent atomic granules of activity that have been observed during the execution of a process. As such, an event has no duration. Examples of an event are:
• Recording the client’s personal information in the database has been completed

3

XES / Version 2.0
Key Issues

• From here…

<table>
<thead>
<tr>
<th>ID</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alifah Syamsiyah</td>
</tr>
<tr>
<td>2</td>
<td>Marco Montali</td>
</tr>
<tr>
<td>3</td>
<td>Diego Calvanese</td>
</tr>
<tr>
<td>4</td>
<td>Wil van der Aalst</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Organizer</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>666</td>
<td>BPM 2015</td>
<td>2</td>
<td>2015-02-14 01:00:00</td>
</tr>
<tr>
<td>667</td>
<td>Caise 2015</td>
<td>4</td>
<td>2015-03-06 01:00:00</td>
</tr>
<tr>
<td>668</td>
<td>ER 2015</td>
<td>4</td>
<td>2015-03-26 01:00:00</td>
</tr>
<tr>
<td>669</td>
<td>EDOC 2015</td>
<td>2</td>
<td>2015-04-05 03:00:00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>CT</th>
<th>User</th>
<th>Conf</th>
<th>Type</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ontop at Work</td>
<td>2015-03-02 15:09:35</td>
<td>1</td>
<td>669</td>
<td>FP</td>
<td>RX</td>
</tr>
<tr>
<td>2</td>
<td>A Survey of Web Services</td>
<td>2015-03-02 12:36:01</td>
<td>3</td>
<td>668</td>
<td>SP</td>
<td>RX</td>
</tr>
<tr>
<td>3</td>
<td>The Definitive Guide for BPM</td>
<td>2015-03-04 13:36:20</td>
<td>1</td>
<td>666</td>
<td>FP</td>
<td>AB</td>
</tr>
</tbody>
</table>
Key Issues

• …To there!

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Activity</th>
<th>Timestamp</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Submit paper</td>
<td>01-07-2015:10.05</td>
<td>Anna</td>
</tr>
<tr>
<td></td>
<td>Invite reviewer</td>
<td>03-07-2015:12.00</td>
<td>Budi</td>
</tr>
<tr>
<td></td>
<td>Get review</td>
<td>10-07-2015:16.10</td>
<td>Rudy</td>
</tr>
<tr>
<td></td>
<td>Acceptance</td>
<td>12-07-2015:15.00</td>
<td>Rudy</td>
</tr>
<tr>
<td></td>
<td>Submit final paper</td>
<td>19-07-2015:19.15</td>
<td>Anna</td>
</tr>
<tr>
<td>2</td>
<td>Submit paper</td>
<td>02-07-2015:17.19</td>
<td>John</td>
</tr>
<tr>
<td></td>
<td>Invite reviewer</td>
<td>03-07-2015:12.00</td>
<td>Tiara</td>
</tr>
<tr>
<td></td>
<td>Get review</td>
<td>11-07-2015:15.45</td>
<td>Clara</td>
</tr>
<tr>
<td></td>
<td>Rejection</td>
<td>12-07-2015:15.00</td>
<td>Clara</td>
</tr>
</tbody>
</table>
Ontology-Based Data Access

Ontology-based data integration framework

Query answering

Ontology languages

Mappings

Identity

Conclusions

Ontologies for Data Integration

Diego Calvanese (FUB)
Resolving the Impedance Mismatch

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>CT</th>
<th>User</th>
<th>Conf</th>
<th>Type</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ontop at Work</td>
<td>2015-03-02 15:09:35</td>
<td>1</td>
<td>669</td>
<td>FP</td>
<td>RX</td>
</tr>
<tr>
<td>2</td>
<td>A Survey of Web Services</td>
<td>2015-03-02 12:36:01</td>
<td>3</td>
<td>668</td>
<td>SP</td>
<td>RX</td>
</tr>
<tr>
<td>3</td>
<td>The Definitive Guide for BPM</td>
<td>2015-03-04 13:36:20</td>
<td>1</td>
<td>666</td>
<td>FP</td>
<td>AB</td>
</tr>
</tbody>
</table>

FullPaper

creationTime: DateTime
title: String

<table>
<thead>
<tr>
<th>mappingId</th>
<th>fp-mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>paper{ID} a :FullPaper; :title {Title}; :creationTime{CT}</td>
</tr>
<tr>
<td>source</td>
<td>select I.ID, I.Title, I.CT from PaperInfo I where I.Type = “FP”</td>
</tr>
</tbody>
</table>
Multiple Log Views

- **Upload Submitted**: 
  - upload time: DateTime

- **Accepted Paper**: 
  - <<no time>>
  - 0.1 corresponds to 1

- **Upload Accepted**: 
  - upload time: DateTime

- **Conference**: 
  - creation time: DateTime
  - conf name: String

- **Paper**: 
  - creation time: DateTime
  - title: String

- **Decision**: 
  - decision time: DateTime
  - outcome: Bool

- **Review Request**: 
  - invitation time: DateTime

- **Review**: 
  - submission time: DateTime

- **User**: 
  - creation time: DateTime
  - username: String

- **Event** for:
  - trace: follow upload by activity name: “upload submitted”
    - timestamp: upload time
    - resource: follow UhasP
    - type: complete

- **Event** for:
  - trace: follow has author activity name: “decision author”
    - timestamp: decision time
    - resource: follow PhasD
    - type: complete

- **Event** for:
  - trace: follow by activity name: “decision chair”
    - timestamp: decision time
    - resource: follow PhasD
    - type: complete attributes: outcome

- **Event** for:
  - trace: follow has & reviewer activity name: “review”
    - timestamp: submission time
    - resource: follow RhasR & for type: complete
And Now?

Domain Ontology → Annotation → Event Ontology

Mapping → Data base → Data base → Data base

? → XES Log Extraction
Mapping Synthesis

1. Annotation transformed into an ontology-to-ontology mapping $M'$
2. $M'$ is “rewritten” using the data-to-domain ontology mapping
3. The result is a mapping connecting the XES ontology directly to the data
Mapping Synthesis

Two Approaches
1. Materialization into XES
2. Virtual, on-demand access
Log Materialization

SELECT DISTINCT ?t ?v ?e
       ?t :TcontainsE ?e.}

SELECT DISTINCT ?e ?t

SELECT DISTINCT ?e ?t
WHERE {?e :EcontainsA ?a . ?a :keyA ?t.}

SELECT DISTINCT ?e ?t
Log Virtualization

xlog.get(7).get(90) to retrieve the event in index 7th inside the 90th trace in a log
Questions

• How to optimize and test the scalability of the approach? Fine-tuning is a must!

• Is the “virtual” approach useful? How do process mining algorithms access the data?

• High-level approach vs “hands-on” language (a general question for process mining)

• Case studies!
Figure 2.1: The UML 2.0 class diagram for the complete meta-model for the XES standard

- One examination in which the x-ray machine is employed
- One visit of the website, by one specific user

Tag name for the trace object in the XML serialization of XES: `<trace>`

No XML attributes are defined for the `<trace>` tag.

2.1.3 Event

Every trace contains an arbitrary number (may be empty) of event objects. Events represent atomic granules of activity that have been observed during the execution of a process. As such, an event has no duration. Examples of an event are:

- Recording the client's personal information in the database has been completed

---

XES / Version 2.0
Acknowledgments

All coauthors of this research,
in particular

Diego Calvanese (UNIBZ)
Giuseppe De Giacomo (UNIROMA)
Riccardo De Masellis (FBK-Trento)
Alin Deutsch (UCSD)
Chiara Difrancescomarino (FBK-Trento)
Chiara Ghidini (FBK-Trento)
Fabio Patrizi (UNIBZ)
Sergio Tessaris (UNIBZ)
Alifah Syamsiyah (TU/e)
Wil van der Aalst (TU/e)