(CS704) Process Mining Seminar – Kick-Off Meeting

Prof. Dr. Heiner Stuckenschmidt

Timo Sztyler
Goals

In this seminar, you will ...

• Read, understand, and explore scientific literature

• Summarize a current research topic in a concise report (10-15 pages)

• Give a presentation about your topic (before the submission of the report)

• Moderate a scientific discussion about a topic of one of your fellow students

• Provide feedback to a report and a presentation of a fellow student (verbal)
Requirements

• The report has to be written with Latex
  • If you need help or an introduction – just let me know

• No programming skills are required but they might be helpful

• Lectures such as Process Management (Master) and Algorithms and Data Structure (Bachelor) are recommended.

• Report and presentation have to be in English

If you have any kind of question – just ask or contact me!
## Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.09.</td>
<td>Deadline for the registration</td>
</tr>
<tr>
<td>21.09.</td>
<td>Kick-Off meeting / Topic assignment</td>
</tr>
<tr>
<td>04.-07.10.</td>
<td>Tutor meeting (not mandatory)</td>
</tr>
<tr>
<td>24.10.</td>
<td>Slides submission (by e-mail)</td>
</tr>
<tr>
<td>26.10.</td>
<td>Presentation (10:00 a.m., B6 A2.07)</td>
</tr>
<tr>
<td>31.-04.11.</td>
<td>Tutor meeting (not mandatory)</td>
</tr>
<tr>
<td>23.12.</td>
<td>Final report submission (by e-mail)</td>
</tr>
</tbody>
</table>
Grading

1. Presentation 20%
2. Moderate the discussion 10%
3. Seminar Report 70%

Again - the report has to be written in Latex

• Template: [http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/Lehre/Thesis/vorlage_abschlussarbeiten.zip](http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/Lehre/Thesis/vorlage_abschlussarbeiten.zip)
• Change the Format from “Book” to “Article”
Presentation

• Everyone has to introduce his topic in one month
  • You do not have to submit any kind of report beforehand but the slides
  • I will assign a “moderator” to each presentation
  • This should help you to get feedback

• The moderator has to lead the discussion and has to ask questions (I will send him the slides beforehand)

• Duration: 15 minutes presentation / 5 minutes discussion
Topic Assignment

1. Wenwen, Liao - Process Diagnostics
2. Siefert, Mike - The ProM Framework
3. Komorek, Clemens - Process Mining (?)
4. Nguyen Ngoc Nam, Trung - Trace Clustering/Alignment

There are no conflicts =D
Next

• You should be familiar with your topic
• You should start a literature research (scholar.google.com)
  ▪ to get an overview of the related work
  ▪ W.M.P. Van der Aalst is a famous guy in this community
  ▪ most of the paper are only available if you are using the university network!

• I will tell you during the tutorial session what I expect and you can ask me any kind of questions

If you have any kind of question – just ask or contact me!
Final submission

• You have to submit your final seminar report via e-mail
  • It is not necessary to print it

• The deadline is 23.12. (hard, no extension!, by e-mail)

• If you want some feedback you can send me a pre-version (at least three weeks before the deadline)

• Please submit your tex-files as well as the compiled version (as pdf)
Questions

Are there any questions?
Organizational Issues

Organizers

• Prof. Dr. Heiner Stuckenschmidt
  heiner@informatik.uni-mannheim.de

• Timo Sztyler (contact)
  timo@informatik.uni-mannheim.de
Process Mining – Event Logs

Process mining is impossible without proper event logs

How to redesign a process to improve its performance?
What really happened in the past?
Why did it happen?
What is likely to happen in the future?

<table>
<thead>
<tr>
<th>Case id</th>
<th>Event id</th>
<th>Properties</th>
<th>Timestamp</th>
<th>Activity</th>
<th>Resource</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35654423</td>
<td>30-12-2010:11.02</td>
<td>Register request</td>
<td>Pete</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654424</td>
<td>31-12-2010:10.06</td>
<td>Examine thoroughly</td>
<td>Sue</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654425</td>
<td>05-01-2011:15.12</td>
<td>Check ticket</td>
<td>Mike</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654426</td>
<td>06-01-2011:11.18</td>
<td>Decide</td>
<td>Sara</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654427</td>
<td>07-01-2011:14.24</td>
<td>Reject request</td>
<td>Pete</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>35654483</td>
<td>30-12-2010:11.32</td>
<td>Register request</td>
<td>Mike</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654485</td>
<td>30-12-2010:12.12</td>
<td>Check ticket</td>
<td>Mike</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654487</td>
<td>30-12-2010:14.16</td>
<td>Examine casually</td>
<td>Pete</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654488</td>
<td>05-01-2011:11.22</td>
<td>Decide</td>
<td>Sara</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35654489</td>
<td>08-01-2011:12.05</td>
<td>Pay compensation</td>
<td>Ellen</td>
<td>200</td>
<td></td>
</tr>
</tbody>
</table>
Process Mining – Process Discovery (1)

Based on the an event-log, a process model is constructed thus capturing the behavior seen in the log

\[ L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle] \]

A process discovery algorithm is a function that maps an (simple) event log onto a process model (e.g., Petri net)
Process Mining – Process Discovery (2)

The $\alpha$-algorithm is the most known algorithm to discover workflow nets (e.g., Petri Nets) from an event log

\[(1) \quad TL = \{t \in T \mid \exists \sigma \in L \ t \in \sigma\}\]
\[(2) \quad TI = \{t \in T \mid \exists \sigma \in L \ t = \text{first}(\sigma)\}\]
\[(3) \quad TO = \{t \in T \mid \exists \sigma \in L \ t = \text{last}(\sigma)\}\]
\[(4) \quad XL = \{(A,B) \mid A \subseteq TL \land A = \emptyset \land B \subseteq TL \land B = \emptyset \land \forall a \in A \forall b \in B \ a \rightarrow L b \land \forall a_1, a_2 \in A \ a_1 \# L a_2 \land \forall b_1, b_2 \in B \ b_1 \# L b_2\}\]
\[(5) \quad YL = \{(A,B) \in XL \mid \forall (A',B') \in XL \ A \subseteq A' \land B \subseteq B' \Rightarrow (A,B) = (A',B')\}\]
\[(6) \quad PL = \{p(A,B) \mid (A,B) \in YL\} \cup \{iL, oL\}\]
\[(7) \quad FL = \{(a,p(A,B)) \mid (A,B) \in YL \land a \in A\} \cup \{(p(A,B), b) \mid (A,B) \in YL \land b \in B\} \cup \{(iL, t) \mid t \in TI\} \cup \{(t, oL) \mid t \in TO\}\]
\[(8) \quad \alpha(L) = (PL, TL, FL)\]