

Project #2: PM4Py, ProM and Disco

Make a hospital audit report concerned about the process of handling patients with suspected sepsis based on the analysis of the [Sepsis.xes.gz](https://doi.org/10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460)¹ using PM4Py, ProM, and Disco tools. The event log description is given below:

This real-life event log contains events of sepsis cases from a hospital. Sepsis is a life threatening condition typically caused by an infection. One case represents the pathway through the hospital.

The events were recorded by the ERP (Enterprise Resource Planning) system of the hospital. There are about 1000 cases with in total 15,000 events that were recorded for 16 different activities. Moreover, 39 data attributes are recorded, e.g., the group responsible for the activity, the results of tests and information from checklists.

Events and attribute values have been anonymized. The time stamps of events have been randomized, but the time between events within a trace has not been altered.

Attributes

=====	
Name	/ Description

Age	/ Age in 5-year groups
Diagnostic*	/ Various checkboxes on the triage document
DisfuncOrg	/ Checkbox: Disfunctional organ
Hypotensie	/ Checkbox: Hypotension
Hypoxie	/ Checkbox: Hypoxia
InfectionSuspected	/ Checkbox: Suspected infection
Infusion	/ Checkbox: Intravenous infusion required
Oligurie	/ Checkbox: Oliguria
SIRSCritHeartRate	/ Checkbox: One of the SIRS criteria
SIRSCritLeucos	/ Checkbox: One of the SIRS criteria
SIRSCritTachypnea	/ Checkbox: One of the SIRS criteria
SIRSCritTemperature	/ Checkbox: One of the SIRS criteria
SIRSCriteria2OrMore	/ Checkbox: Two or more of the SIRS criteria
Leucocytes	/ Leucocytes measurement
CRP	/ CRP measurement
LacticAcid	/ Lactic-acid measurement

¹ Event log source: <https://doi.org/10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460>

Include the following points in the report:

- Description of the problem and description of the origin of the event log.
- Analysis of event log content, description of a typical process flow, distribution of business case frequency, occurrence of loops, distribution of process duration.
- Analysis of the completeness of business cases recorded in the event log. If there are incomplete cases, please handle the problems arising from them.
- The results of two analyses – the first in the PM4Py or ProM tools, the second in Disco.
- PM4Py and ProM: Build a process model in at least two representations chosen out of C-net, process tree, Petri net, and do the following:
 - Filter the log and/or fine-tune the parameters of the algorithm to reject noisy dependencies and incomplete cases, but at the same time do not lose the full picture on the process. Test different variants of the training algorithms and check if the obtained models are sound, i.e. they are free of deadlock and livelock, do not have dead parts, the final state is reachable from each reachable state and after obtaining the final state there are no tokens left inside the model. Analyze the conformance of models with the log. Choose a model that is sound (or close to), has a high fitness and precision and do the next steps.
 - Based on the model, describe the typical process flow. Is this observation consistent with previous manual log analysis?
 - Describe on the basis of the model and/or simulation what the most common deviations from the typical process look like and what they may be caused by.
 - Make temporal analysis of the process using the model and log to determine which activities are the main causes of slowdown and inefficiencies and try to assess why.
 - Build and evaluate decision classifiers for XOR-split/OR-split nodes, which explain the decisions made using case and/or event attributes.²
 - Analyze the compatibility of the discovered model with domain knowledge, e.g.: [\[1\]](#), [\[2\]](#).
- Disco: Build a process map and do the following analyses:
 - Filter log and tune algorithm parameters.

² Extension scored extra, max 1p. In PM4Py replay the log on the model using Token-Based Replay (TBR) and create classifiers for decision points. In ProM use the models with "Data" in the name, e.g.: "Data Causal net", „Data Petri net”.

- Describe a typical process flow according to the process map. Is it consistent with the observations in the log?
- Analyze the frequency of activities: the longest activities, the most repeated. Try to explain reasons for long service and waiting times and repetitions.
- Are there loops of length 1 (activity A -> activity A)? If so, justify what could be the cause?
- Analyze periodicity of work. If it occurs, include it in the analysis.
- What is the key factor affecting the total active time, what affects the waiting time, and what affects the increase in the number of events/activities in the process?
- Compare the results in both tools. Which tool creates better models and why? Do one of the tools explain the process better than the other? Describe your findings from working with these tools.
- It is possible to expand with more analyses and observations than mentioned above.³
- Save partial analysis results, models, etc. to files, so that they can be run in PM4Py, ProM, or Disco to verify the results presented in the report.

The report should be a text document submitted in accordance with good [typesetting and report drafting rules](#). The text should be accompanied by tables, e.g.: presenting data and/or statistics; illustrations, e.g. models, charts and bibliography – references to the algorithms used and/or facts from domain knowledge. The use of LaTeX or [the LyX](#) overlay on it is welcome, but it is not required.

The report should be delivered in electronic form (PDF) to the lecturer. No need to print. The length of the document does not affect the rating, so please avoid overblowing the text with formatting and/or oversized illustrations. What counts is the completeness of the content and the brevity of the presentation.

Discussion/explanation of terms from domain knowledge:

- Triage – rapid classification procedure for patients, see: [\[1\]](#).
- ER – Emergency Room [\[3\]](#).
- IV – Intravenous therapy [\[4\]](#).
- CRP – C-Reactive Protein [\[5\]](#).
- NC – Normal Care ward,
- IC – Intensive Care ward [\[6\]](#).

³ Extensions scored additionally, max 1p.

- Release A-E – ways of discharging the patient (the assignment of letters to types is not known, but in some cases, it can be deduced from the process model):
 - Discharge without admission,
 - Admission to normal ward,
 - Discharge to home,
 - Admission to intensive care (Admission ICU),
 - Admission to a normal ward and transition of the patient to the intensive care within 72 hours.

For details, see document [\[7\]](#).