# Assignments course ‘Advanced Process Mining’

For the course ‘Advanced Process Mining’, the final mark will be based on an assignment (4 points) and a written exam (6 points, 3 hours). The assignment consists of three separate subassignments.

In general, the aim of the assignments is that students get hands-on experience with applying process mining techniques on event logs (both real-life and artificial). The application of these techniques will take place using the scientific workflow management tool RapidMiner 5 (see <http://rapid-i.com/> and <http://www.win.tue.nl/~rmans/RapidMiner/>) in which ProM 6 algorithms have been integrated. Furthermore, students will get hands-on experience with combining data mining techniqes and process mining techniques.

For the three assignments students are expected to spend 56 hours in total. Each assignment has to be performed individually by each student.

A description of each assignment is given below. Furthermore, before starting to work on each assignment please consider the slides which have been discussed during the lecture for each respective assignment!

## First Assignment

**Aim:** Given an event log, the student applies a wide variety of process mining techniques. In this way, the student gets hands-on experience with the three different kind of process mining techniques that are available (discovery, conformance, and enhancement) and the various algorithms that are part of each process mining type.

**Description:** An artificial log has been created for a healthcare process for diagnosing gynaecological oncology patients. So, this is the process starting from the first visit of the patient until a surgical examination. The student is expected to obtain results for each of the three different kinds of process mining techniques that are available (discovery, conformance, and enhancement). Next to that, results need to be obtained for the control-flow, organizational, and performance perspectives.

**Input:** The generated log for the patients following the gynaecological oncology diagnosis healthcare process. Note that the log is provided in XES format and is called ‘APM\_First\_assignment.xes’.

**Questions:** In order to get a sufficient grade for this assignment the following questions need to be *properly* addressed:

1. Is there a control-flow of the process discoverable in terms of a Petri net with a fitness of 1.0? If yes, provide the model, if not, explain why.
2. Discover information about the organizational perspective of the process. For each of the following networks, try to find a clear organizational structure and discuss the structure obtained. If no clear structure is to be found, explain why this is the case.
   * Similar-Task Social Network
   * Handover-of-Work Social Network
   * Subcontracting Social Network
   * Working-Together Social Network
   * Reassignment Social Network
3. Which bottlenecks within the process cause the biggest amount of delay within the entire process?
4. Identify the first decision point within the process. For this decision point, by which criterion is a certain path taken?
5. Based on the information that has been obtained for the previous questions, are there any opportunities for improving the care process? For each of the above questions can you mention any improvement opportunity? If yes, indicate how the process can be improved. If not, explain why.

**Deliverables:** The assignment should be submitted as a single zip file named **StudentNumber\_APM\_assignment\_1.zip** (StudentNumber is your student number) to the OASE folder ’APM\_Assignment\_1\_16-05-2014 23:59’ and it should consist of the following files:

1. **Report\_StudentNumber.docx**: A report describing the results that are obtained. Note that the report needs to be written using Microsoft Word and should not contain more than 20 pages.
2. **Process models for RapidMiner and any files you used as input there (except the assignment log)**.

## Second Assignment

**Aim:** Given an event log, the student defines several experiments. In each experiment, the robustness of various control-flow miners against different types of noise and against different severity is evaluated.

**Description:** For a log, the robustness of various control-flow miners (e.g. ILP miner, inductive miner) against noise (e.g. swapping or removal of events in a trace) is evaluated. Additionally, each noise type is inserted into the same log with various levels of severity (from 0% till 100%).

**Input:** The generated log for the patients following the gynaecological oncology diagnosis healthcare process. However, for the process we now only consider the ‘complete’ events. So, the log is provided in XES format and is called ‘APM\_Second\_assignment.xes’.

**Questions:** In order to get a sufficient grade for the third assignment the following questions need to be *properly* addressed.

To this end, Set-up experiments in which each control-flow discovery algorithm is tested against a specific type of noise (remove head, remove body, add event, swap tasks, remove task). Additionally, for each noise type various percentages of noise are assessed per control-flow discovery algorithm:

1. For which noise type, and with which percentage of noise, does the ‘alpha miner’ provide the ‘worst’ model in terms of fitness and understandability?
2. For which noise type, and with which percentage of noise, does the ‘ILP miner’ provide the ‘worst’ model in terms of fitness and understandability?
3. For which noise type, and with which percentage of noise, does the ‘passage miner’ provide the ‘worst’ model in terms of fitness and understandability?
4. For which noise type, and with which percentage of noise, does the ‘inductive miner’ provide the ‘worst’ model in terms of fitness and understandability?
5. Based on the results of the previous questions, can you indicate for each type of noise which is the most robust control-flow miner? Additionally, can you indicate which control-flow miner is the most robust against noise and which control-flow miner is the least robust against noise?

**Deliverables:** The assignment should be submitted as a single zip file named **StudentNumber\_APM\_assignment\_2.zip** (StudentNumber is your student number) to the OASE folder ’APM\_Assignment\_2\_23-05-2014 23:59’ and it should consist of the following files:

1. **Report\_StudentNumber.docx**: A report describing the results that are obtained. Note that the report needs to be written using Microsoft Word and should not contain more than 20 pages.
2. **Process models for RapidMiner and any files you used as input there (except the assignment log)**.

## Third Assignment

**Aim:** For this assignment, the student is provided with a real-life event log. The student will analyze the data using whatever (process mining) techniques are available within RapidMiner.

**Description:** For the real-life event log, several questions that need to be answered that are provided by the process owner. The student analyses the event log using whatever techniques are available within RapidMiner, focusing on the process owner's questions.

**Input:** A real-life event log for which several questions need to be answered. Note that the log is provided in XES format and is called ‘APM\_Third\_assignment’. Also, there are two additional logs provided in which the traces in the original log are splitted, namely ‘APM\_Third\_Assignment\_(Before).xes’ and ‘APM\_Third\_Assignment\_(After).xes’.

**Evaluation:** In order to get a sufficient grade for the third assignment the following criteria need to be satisfied:

* Each stakeholder question is translated into a concrete analysis question making clear which problem is going to be solved.
* For each analysis question an answer needs to be provided using the operators within RapidMiner. That is for each question, a detailed answer is given and it is indicated which process/data mining techniques have been used for answering the question.

**Deliverables:** The assignment should be submitted as a single zip file named **StudentNumber\_APM\_assignment\_3.zip** (StudentNumber is your student number) to the OASE folder ’APM\_Assignment\_3\_14-06-2014 23:59’ and it should consist of the following files:

1. **Report\_StudentNumber.docx**: A report describing the results that are obtained. Note that the report needs to be written using Microsoft Word and should not contain more than 20 pages.
2. **Process models for RapidMiner and any files you used as input there (except the assignment log)**.