

# Study Guide 2007/08 Process Mining (1BM45)

Version 2.0

*Note: The Studyweb (course 1BM45) will be used for posting up-to-date information about this course! This includes the material necessary to perform the assignments, as well as the answers for the assignments and the final grades.*

## **Teachers:**

dr. A.J.M.M. (Ton) Weijters (TM-IS)

dr. Ana Karla Alves de Medeiros (TM-IS)

## **Lectures:**

Dates: Jan 28th; Feb 11th, 18th, 25th; and Mar 3rd, 2008 (Mondays).

Time: 08:45 to 10:30

Location: Paviljoen building, B-2.

## **Instructionss:**

Dates: Feb 1st; Feb 15th, 22th, 29th; and Mar 7th, 2008 (Fridays).

Time: 10:45 to 12:30

Location: Paviljoen building, B-2.

## **1 General information**

### **1.1 Goal**

As stated in the OWinfo webpage, the objective of this course are:

“Unlike classical data mining techniques, process mining focuses on causal dependencies between activities in a process. The goal is to use process mining techniques and tools (in particular the ones used in the ProM framework and knowledge discovery tools) to investigate the behavior of processes and

organizations. After the course students should be able to analyze processes based on their event logs.”

In other words, the aim of this course is to teach the students how to use process mining techniques to analyze different aspects of (business) processes in organizations. The hands-on experience comes from the use of the ProM tool to mine (artificially created) event logs.

## 1.2 Content

The course has two main parts: one focusing on the data mining techniques that are commonly used in practice and the other focusing on process mining. The main reason to also let the students experiment with traditional data mining techniques is to get them familiar with the concept of mining in general. Furthermore, since some process mining techniques are based on conventional data mining techniques, the lectures about data mining will provide the background necessary to understand these process mining techniques.

## 1.3 Planning

In general terms, the subjects to be addressed during the lectures and instructions are:

### Lecture 1 - 28/01/2008

**Content:** (1) Motivation of the course (Why data mining? Why process mining? Presentation of some use cases using real-life logs); (2) Explanation of course outline, structure, assignments and assessment; and (3) Data mining - Part 1 (Main concepts and techniques; Decision Trees; Tool Alpha Miner).

**Material:** Slides, Tutorial about data mining.

### Instruction 1 - 01/02/2008

**Content:** Assignment using decision trees.

**Material:** Tool *Alpha Miner*; Files with data to be mined.

### Lecture 2 - 11/02/2008

**Content:** Data Mining - Part 2 (explain the two other techniques commonly used in practice: clustering and genetic algorithms)

**Material:** Slides; Tutorial about data mining; Articles:

- Introduction to Data Mining and Knowledge Discovery. Third Edition. Two Crows Corporation. Strong focus on the whole data mining process.

Link: <http://www.twocrows.com/intro-dm.pdf>

### Instruction 2 - 15/02/2008

**Content:** Discuss answers for assignment previous instruction; Assignment using clustering and genetic algorithms.

**Material:** Tool *Alpha Miner*; Files with data to be mined; Articles:

- Ton Weijters and Jan Paredis (2002). Genetic Rule Induction at an Intermediate Level. Knowledge-Based Systems, Vol 15/1-2, pp 85-94.

Link: [http://dx.doi.org/10.1016/S0950-7051\(01\)00124-1](http://dx.doi.org/10.1016/S0950-7051(01)00124-1)

### Lecture 3 - 18/02/2008

**Content:** Brief recap motivation process mining; Explanation about three perspectives: discovery, conformance and extension; Explanation some discovery techniques: control-flow mining algorithms ( $\alpha$ -miner, Heuristics net miner, Genetic Miner and Fuzzy miner).

**Material:** Slides; Research Articles/Chapters:

- (Alpha Miner + Heuristics Miner) W.M.P. van der Aalst and A.J.M.M. Weijters. Process Mining. In M. Dumas, W.M.P. van der Aalst, and A.H.M. ter Hofstede, editors, Process-Aware Information Systems: Bridging People and Software through Process Technology, pages 235-255. Wiley & Sons, 2005.

Link: <http://tabu.tm.tue.nl/wiki/publications/aalst2005e>

- (Genetic Algorithms) Ana Karla Alves de Medeiros Genetic Process Mining PhD Thesis. Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2006. (see **Chapters 4 and 5**)

Link: <http://tabu.tm.tue.nl/wiki/publications/thesisgeneticprocessmining>

- (Fuzzy Miner) C.W. Günther and W.M.P. van der Aalst. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics.

In G. Alonso, P. Dadam, and M. Rosemann, editors, International Conference on Business Process Management (BPM 2007), volume 4714 of Lecture Notes in Computer Science, pages 328-343. Springer-Verlag, Berlin, 2007.

Link: <http://wwwis.win.tue.nl/~wvdaalst/publications/p400.pdf>

- (Multi-Phase Miner) B. van Dongen and W.M.P. van der Aalst. Multi-Phase Mining: Aggregating Instances Graphs into EPCs and Petri Nets. In D. Marinescu, editor, Proceedings of the Second International Workshop on Applications of Petri Nets to Coordination, Workflow and Business Process Management, pages 35-58. Florida International University, Miami, Florida, USA, 2005.

Link: <http://wwwis.win.tue.nl/~wvdaalst/publications/p262.pdf>

### **Instruction 3 - 22/02/2008**

**Content:** Discuss answers for assignment previous instruction; Assignment using control-flow mining techniques.

**Material:** ProM Tutorial; Tool *ProM 4.2*; Material of Lecture 3; Assignment 3.

### **Lecture 4 - 25/02/2008**

**Content:** Brief recap previous lecture; Discovery techniques (continuation): social network (Handover of work, SNA Analysis, Organizational Miner); Conformance techniques: algorithms Conformance Checker and LTL Checker.

**Material:** Slides, Research Articles:

- (Social network) M. Song and W.M.P. van der Aalst. Towards Comprehensive Support for Organizational Mining, BETA Working Paper Series, WP 211, Eindhoven University of Technology, Eindhoven, 2007. Link: <http://tabu.tm.tue.nl/wiki/publications/beta.211>
- (Conformance Checker) A. Rozinat and W.M.P. van der Aalst. Conformance Checking of Processes Based on Monitoring Real Behavior. Information Systems, 2007. Link: <http://tabu.tm.tue.nl/wiki/publications/confchecking>
- (LTL Checker) W.M.P. van der Aalst, H.T. de Beer, and B.F. van Dongen. Process Mining and Verification of Properties: An Approach

based on Temporal Logic. In R. Meersman and Z. Tari et al., editors, On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, volume 3760 of Lecture Notes in Computer Science, pages 130-147. Springer-Verlag, Berlin, 2005.

Link: <http://tabu.tm.tue.nl/wiki/publications/aalst2005f>

#### **Instruction 4 - 29/02/2008**

**Content:** Discuss answers for assignment previous instruction; Assignment using social network algorithms and conformance ones.

**Material:** ProM Tutorial; Tool *ProM 4.2*; Material of Lecture 4; Assignment 4.

#### **Lecture 5 - 03/03/2008**

**Content:** Brief recap previous lecture; Extension techniques: Performance Analysis with Petri Nets, Decision Miner; **Invited Talk: Futura Technology** (Start-up company in the area of process mining. They will show their commercial process mining tool and share some of their experiences so far).

**Material:** Slides, Research Articles:

- (Performance Analysis with Petri Nets) Peter Hornix Master Thesis. Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2007. Link: <http://tabu.tm.tue.nl/wiki/publications/hornix2007>
- (Decision Miner) A. Rozinat and W.M.P. van der Aalst In S. Dustdar, J.L. Fiadeiro, and A. Sheth, editors, BPM 2006, volume 4102 of Lecture Notes in Computer Science, pages 420-425. Springer-Verlag, Berlin, 2006  
Link: <http://tabu.tm.tue.nl/wiki/publications/bpm2006a>

#### **Instruction 5 - 07/03/2008**

**Content:** Discuss answers for assignment previous instruction; Assignment in the process mining area. This assignment will mimic a real-life situation as close as possible. The aim is to assess how good the students are at solving problems in the process mining area.

**Material:** ProM Tutorial; Tool *ProM 4.2*; Material of *all* lectures; Assignment 5.

Assignment Number	Available by	Must be posted by	Mark range	Percentage final score
1	01/02/2008	11/02/2008 - 08:45 a.m.	[0,1]	10%
2	15/02/2008	20/02/2008 - 11:00 a.m.	[0,1]	10%
3	22/02/2008	27/02/2008 - 11:00 a.m.	[0,1]	10%
4	29/02/2008	05/03/2008 - 11:00 a.m.	[0,1]	10%
5	07/03/2008	14/03/2008 - 06:00 p.m.	[0,6]	60%

Tab. 1: Dates for posting the assignments at the Studyweb and their mark ranges.

## 1.4 Assessment

As the main aim of the course is to *prepare the students to apply process mining techniques in practice*, the grade will be given based on assignments distributed during the instructions. Thus, after every instruction the students have to hand in the answers for an assignment. The answers should be posted in respective folders in Studyweb. In total, five assignments will be graded. The first four assignments are individual ones and they correspond to 40% of the final grade. Depending on the number of students, the last assignment may be an individual or a group one. This assignment is more complex than the other four ones because it will closely reproduce typical real-life situations. The mark for this last assignment corresponds to 60% of the final grade of a student in this course. Table 1 contains the relevant dates for assignments and mark range.