

MTAT.03.231 – Business Process Management

Homework 5 (HW5) – Mining (5 points)

Due on 31.05 at 23:59

Complete the homework individually or in groups of 2 students.

The goal of this assignment is to analyze the logs provided on the course website. **Describe in detail the procedure you follow (using Disco and/or Apromore) to answer the questions. Direct answers without explanations will not be taken into consideration.**

Part 1.

Your task is to analyze the following synthetic log using process mining techniques:

<https://courses.cs.ut.ee/2019/bpm/spring/uploads/Main/HW5.zip>

Use the alpha algorithm in Apromore to discover a BPMN model from the log.

1. Provide the model returned by Apromore
2. Explain step by step how you extract that model using the footprint matrix of the log.

Given the logs

CC1= {<A, B, G>²⁰⁰, <A, C, D, E, F, G>³²⁰, <A, C, E, D, G>¹², <A, B>³}

CC2= {<A, B, G>²⁵⁰, <A, C, D, E, F, G>⁴²⁰, <A, C, E, D, G>²², <A, B>⁹, <A, C, E, D, F, G>¹}

1. Compute the fitness with the trace-based approach for both logs (justify the answer)
2. Compute the fitness with the event-based approach for both logs (justify the answer)
3. Compute the fitness with the alignment-based approach (using the standard cost function) for both logs (justify the answer)
4. Compute the precision for both logs (justify the answer)

Part 2.

Your task is to analyze the following real-life log using process mining techniques:

<https://courses.cs.ut.ee/2019/bpm/spring/uploads/Main/Sepsis.zip>

1. Give the global statistics (number of events, number of cases, number of activities, Median and Mean case duration, initial and final timestamp in the log) and the map of the most

frequent variant of the entire log. How many cases follow the most frequent variant? What is the highest and the lowest duration among these cases?

2. Give the global statistics (number of events, number of cases, number of activities, Median and Mean case duration, initial and final timestamp in the log) and the map of the most frequent variant of cases containing org:group = A and with a duration of max 10 days. How many cases follow the most frequent variant? What is the highest and the lowest duration among these cases?
3. Give the global statistics (number of events, number of cases, number of activities, Median and Mean case duration, initial and final timestamp in the log) and the map of the most frequent variant of cases containing patients of age = 80, 85 and 90 and start with "ER Registration" and end with "CRP". How many cases follow the most frequent variant? What is the highest and the lowest duration among these cases?