
Bachelor/Master Thesis Topic

A Grammar Based Process Mining Tool

Motivation and Background

Process mining techniques allow for extracting information from event logs. Process mining is closely related to data and workflow mining approaches. Unlike the classical data mining approaches, the main purpose of the algorithms are focusing on processes and provides queries for performance related requirements. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations [1]. Most of the existing approaches are focusing on the structure of the workflows and more information from event logs such as execution and response time, usage profile rates .. etc is needed.

Goals

The goal of this thesis is to create an efficient framework for process mining by using grammar inference techniques [2] with a focus on stochastic information to analyze more complex performance requirements.

Description of the Task

The specific tasks are:

- Understanding the background of process execution and workflows
- Implementing existing grammar inference techniques and extend the methods for process mining
- Performance evaluation and comparison with existing tools like ProM [3] with large data sets

Research Type

Theoretical Aspects: *****

Industrial Relevance: *****

Implementation: *****

Prerequisite

The student should be enrolled in the master of software engineering or master of computer science program, and has completed the required course modules to start a bachelor/master thesis.

Skills required

Mathematical foundations and programming skills in Java/C++ and possibly C are required. Background on Compiler design and Statistics is preferable.

References

[1] Process Mining – manifesto-

http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:process_mining_manifesto-small.pdf

[2] Probabilistic Pattern Matching and the Evolution of Stochastic Regular Expressions, Brian J. Ross, Kluwer Academic Publishers, 2000

[3] <http://www.processmining.org/prom/start>

Contacts

Sinem Getir, Humboldt-Universität zu Berlin, Institut für Informatik, Lehrstuhl Software Engineering, Unter den Linden 6, 10099 Berlin, Germany

Application

Please contact me during my office hours or send me an email with the title: "[ThesisProject]-ProcessMining" to se-career@informatik.hu-berlin.de