HOCHSCHULE
HANNOVER
UNIVERSITY OF
APPLIED SCIENCES
AND ARTS
–
Fakultät IV
Wirtschaft und
Informatik

Process Mining

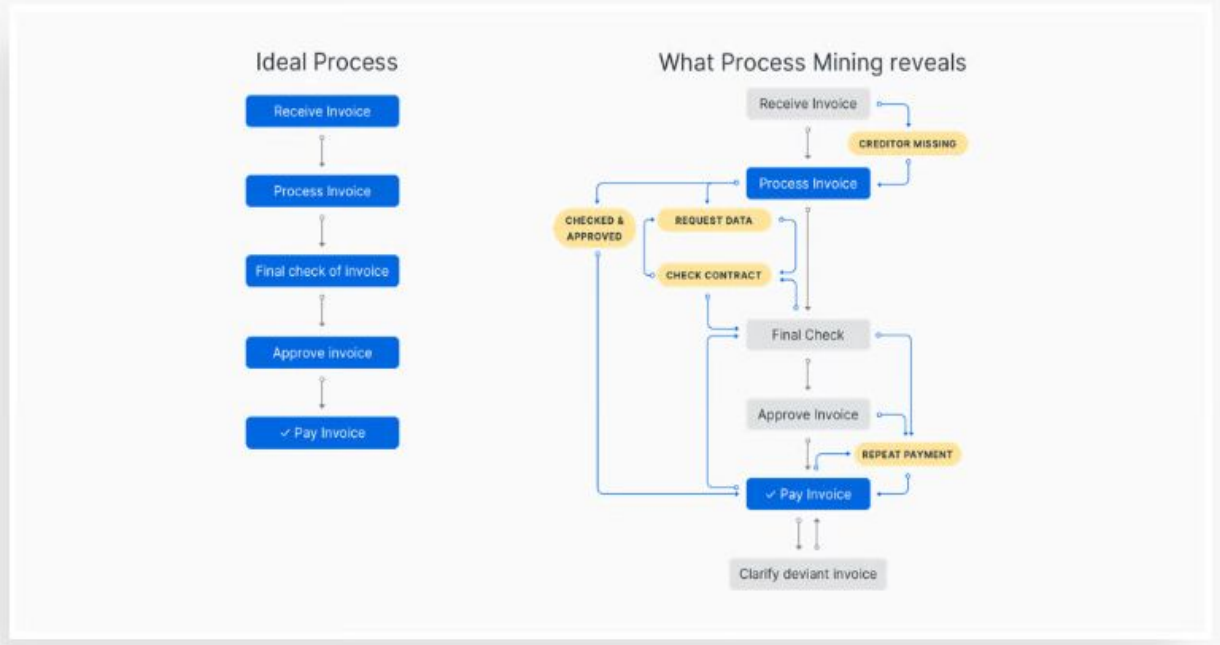
Mohammad Al Nasouh
Mario Vieilledent
Andreas Wolf Wolf
Laura María Camino Yuste
Usama El Moustamide

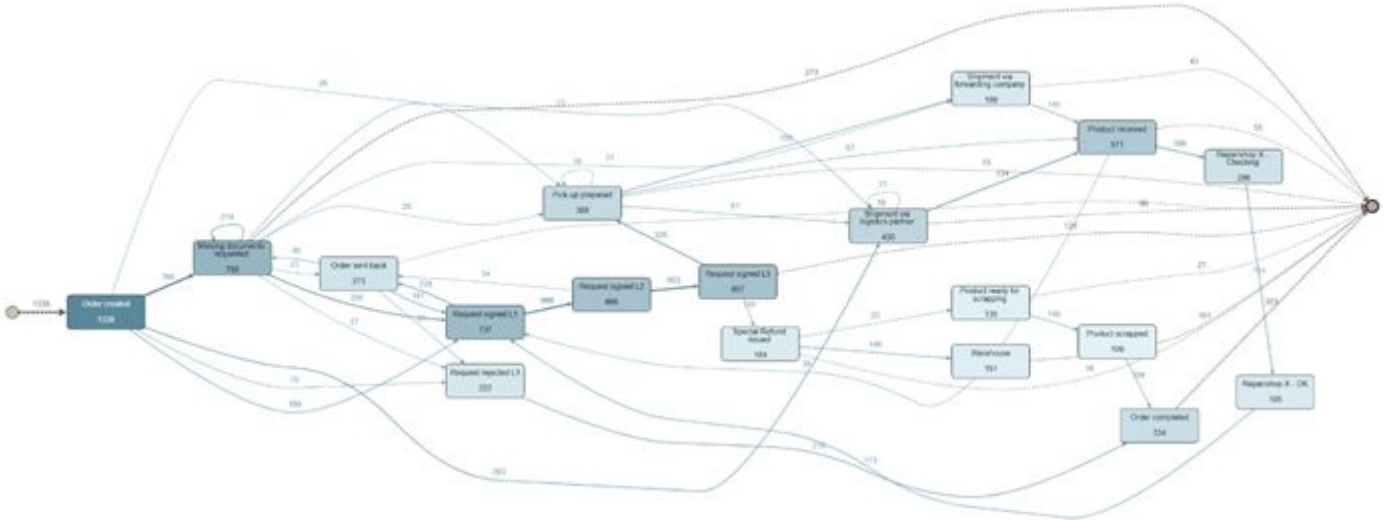
Outline of the presentation

1. Introduction
2. What Is Process Mining
3. History
4. Types Of Process Mining
5. Objectives and benefits
6. How does it work
7. Algorithms
8. Use Cases
9. Experiments
10. Conclusions
11. Literature

Introduction

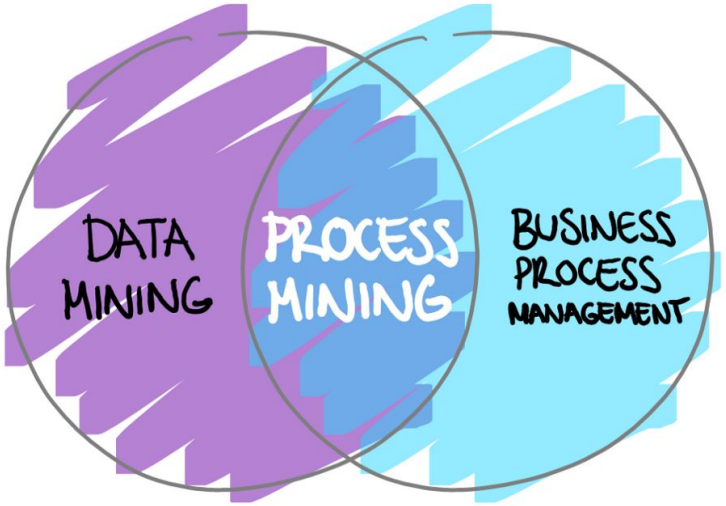
- Companies have try to describe and understand their business process for centuries.
- Business logic is getting more complex as years pass by.
- We need a new way of discovering, understanding and improve our business processes.

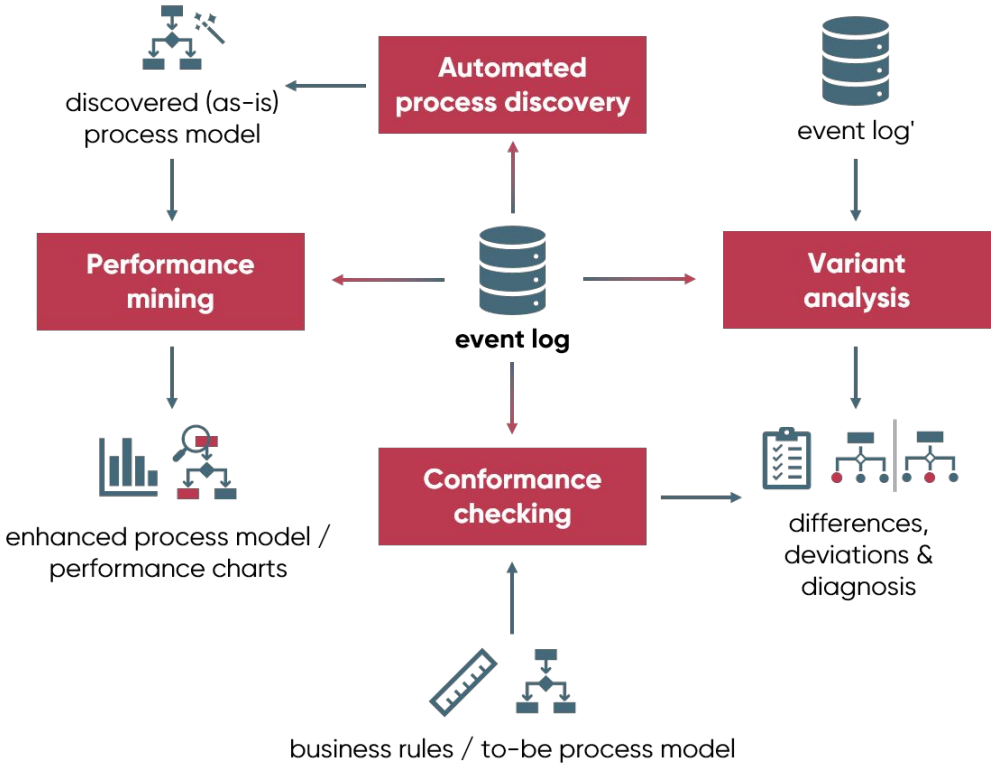




What is Process Mining?

- Is a technique used to analyse and track processes on a business model
- It can be apply on different contexts and businesses.
- It has some common aspects with Data Mining and Business Process Management.



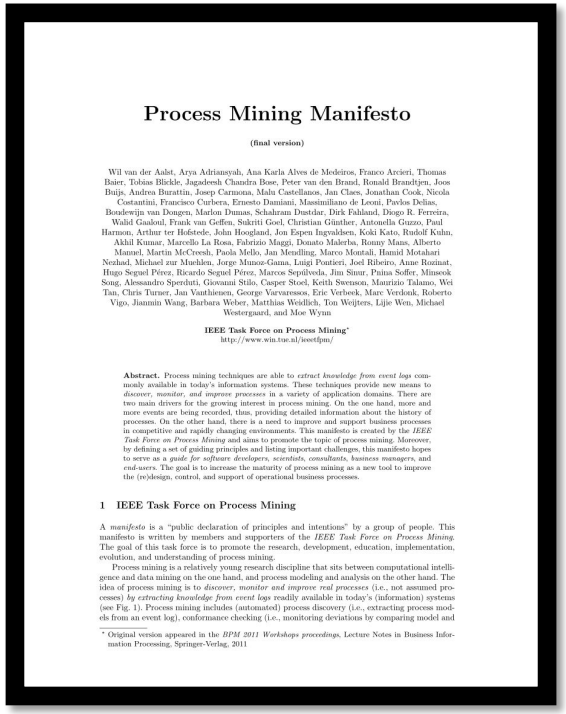


History

At the beginning of the 21st century a teacher a computer scientist from the Eindhoven university, Wil Van De Aalst started to work on the concept of Process Mining, a way of making Process Mining more transparent, automatic and efficient.

At first, this new concept did not rise too much interest in the community, but after the IEEE published an article back on 2011 more and more people started to show interest in the topic.

Nowadays, Process Mining is a fundamental part of business organization and management.



Process Mining Manifesto

(final version)

Wil van der Aalst, Arya Adriansyah, Ana Karla Alves de Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter van den Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malu Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano de Leoni, Pavlos Delias, Boudewijn van Dongen, Marlon Dumas, Schahram Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank van Geffen, Sukriti Goel, Christian Günther, Antonella Guzzo, Paul Harmon, Arthur ter Hofstede, John Hoogland, Jon Espen Ingvaldsen, Koki Kato, Rudolf Kuhn, Akhil Kumar, Marcello La Rosa, Fabrizio Maggi, Donato Malerba, Ronny Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid Motahari Nezhad, Michael zur Muehlen, Jorge Munoz-Gama, Luigi Pontieri, Joel Ribeiro, Anne Rozinat, Hugo Seguel Pérez, Ricardo Seguel Pérez, Marcos Sepúlveda, Jim Sinur, Paina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, Casper Stoel, Keith Swenson, Maurizio Talamo, Wei Tan, Chris Turner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, Barbara Weber, Matthias Weidlich, Ton Weijters, Lijie Wen, Michael Westergaard, and Moe Wynn

IEEE Task Force on Process Mining*
http://www.win.tue.nl/ieeetfpm/

Abstract. Process mining techniques are able to *extract knowledge from event logs* commonly available in today's information systems. These techniques provide new means to *discover, monitor, and improve processes* in a variety of application domains. There are two main drivers for the growing interest in process mining. On the one hand, more and more events are being recorded, thus, providing detailed information about the history of processes. On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments. This manifesto is created by the *IEEE Task Force on Process Mining* and aims to promote the topic of process mining. Moreover, by defining a set of guiding principles and listing important challenges, this manifesto hopes to serve as a *guide for software developers, scientists, consultants, business managers*, and *end-users*. The goal is to increase the maturity of process mining as a new tool to improve the (re)design, control, and support of operational business processes.

1 IEEE Task Force on Process Mining

A *manifesto* is a "public declaration of principles and intentions" by a group of people. This manifesto is written by members and supporters of the *IEEE Task Force on Process Mining*. The goal of this task force is to promote the research, development, education, implementation, evolution, and understanding of process mining.

Process mining is a relatively young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. The idea of process mining is to *discover, monitor and improve real processes* (i.e., not assumed processes) *by extracting knowledge from event logs* readily available in today's (information) systems (see Fig. 1). 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

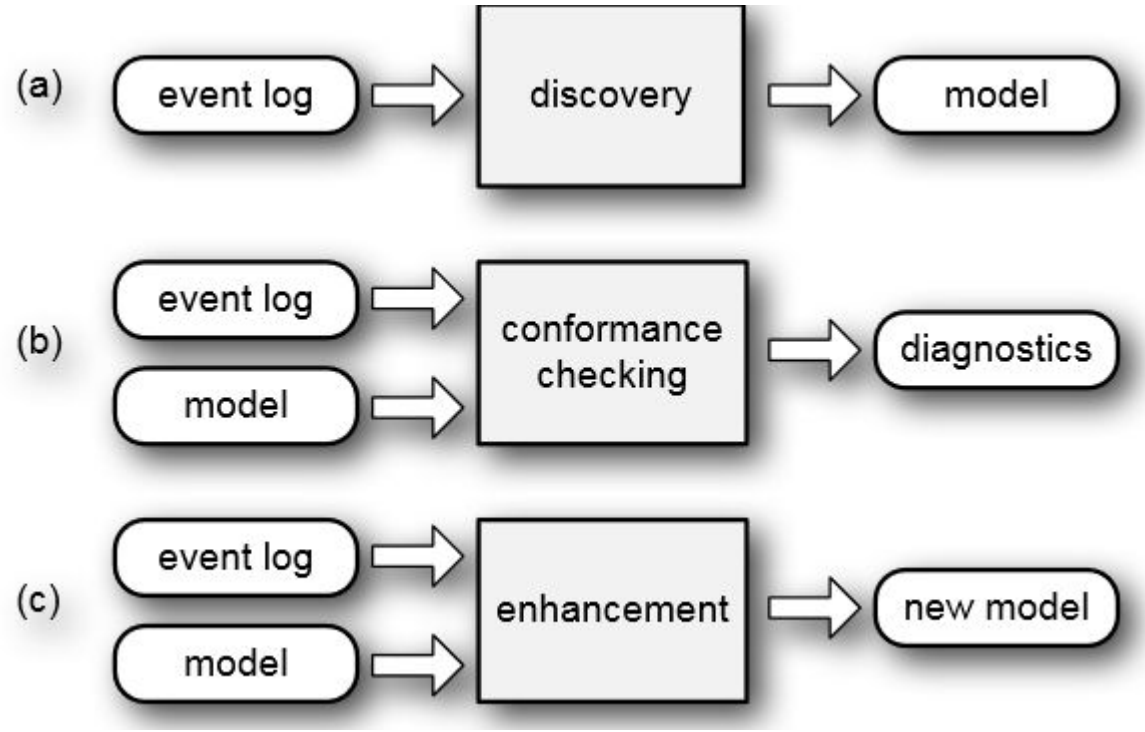
* Original version appeared in the *BPM 2011 Workshops proceedings*, Lecture Notes in Business Information Processing, Springer-Verlag, 2011

Types of Process Mining:

a) **Discovery:** Consists of going over the logs to automatically construct a model based on observable events, without outside influence. Mainly use for discussing problems among stakeholders.

b) **Conformance:** Uses a model and an event log as input. It checks whether the actual model is in fact reflected in practice. It is mainly used to check the accuracy of a model.

c) **Enhancement:** It consists of extending or improving an existing model with the event log and additional information.



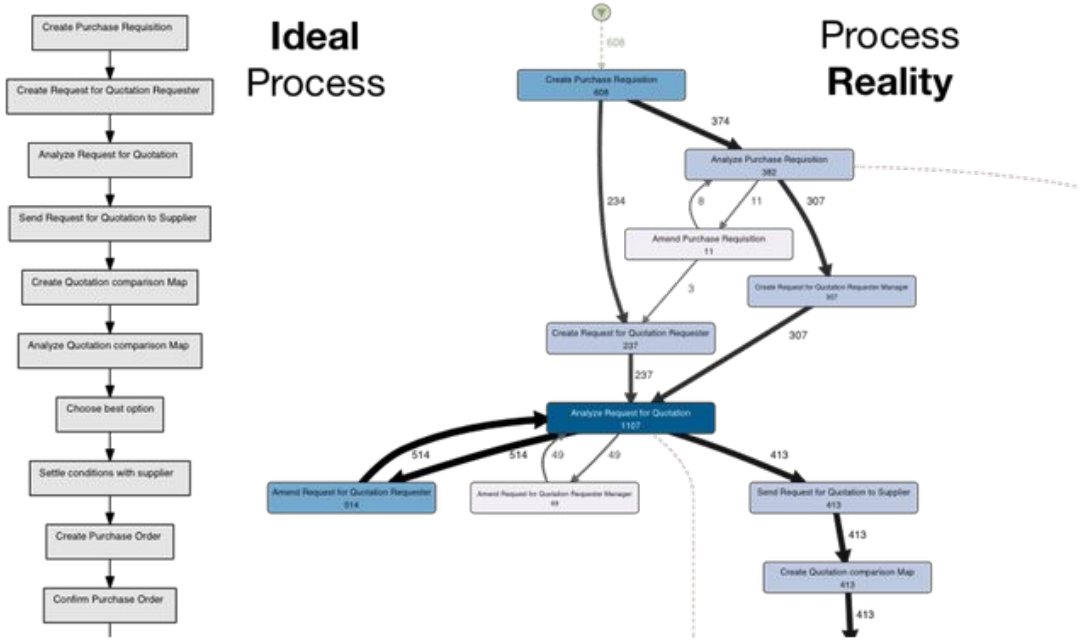
Objectives

GP1	**Event Data Should Be Treated as First-Class Citizens** Events should be *trustworthy*, that is, it should be safe to assume that the recorded events actually happened and that the attributes of events are correct. Event logs should be *complete*, that is, given a particular scope, no events may be missing. Any recorded event should have well-defined *semantics*. Moreover, the event data should be *safe* in the sense that privacy and security concerns are addressed when recording the event log.
GP2	**Log Extraction Should Be Driven by Questions** Without concrete questions it is very difficult to extract meaningful event data. Consider, for example, the thousands of tables in the database of an ERP system like SAP. Without questions one does not know where to start.
GP3	**Concurrency, Choice and Other Basic Control-Flow Constructs Should Be Supported** Basic workflow *patterns* supported by all mainstream languages (e.g., BPMN, EPCs, Petri nets, BPEL, and UML activity diagrams) are *sequence*, *parallel routing* (AND-splits/joins), *choice* (XOR-splits/joins), and *loops*. Obviously, these patterns should be supported by process mining techniques.
GP4	**Events Should Be Related to Model Elements** Conformance checking and enhancement heavily rely on the relationship between *elements in the model* and *events in the log*. This relationship may be used to "replay" the event log on the model. Replay can be used to reveal discrepancies between event log and model (e.g., some events in the log are not possible according to the model) and can be used to enrich the model with additional information extracted from the event log (e.g., bottlenecks are identified by using the timestamps in the event log).
GP5	**Models Should Be Treated as Purposeful Abstractions of Reality** A model derived from event data provides a *view on reality*. Such a view should serve as a purposeful abstraction of the behavior captured in the event log. Given an event log, there may be multiple views that are useful.
GP6	**Process Mining Should Be a Continuous Process** Given the dynamical nature of processes, it is not advisable to see process mining as a one-time activity. The goal should not be to create a fixed model, but to breathe life into process models such that users and analysts are encouraged to look at them on a daily basis.

How does it work?

Process Mining can be divided into the following steps:

1. **Reading digital footprint:** Process Mining reads and transforms data read from the system into logs to work with.
2. **Model Generation:** Using the logs, Process Mining creates a model in which you can see all the process which are being executed.
3. **Working over the model:** Process Mining allows you to focus on a specific area of business and you can start to prioritize and fix concrete problems from them.
4. **Monitoring:** For a Process Mining to be successful the key is to keep monitoring the business in case new inefficiencies appear.



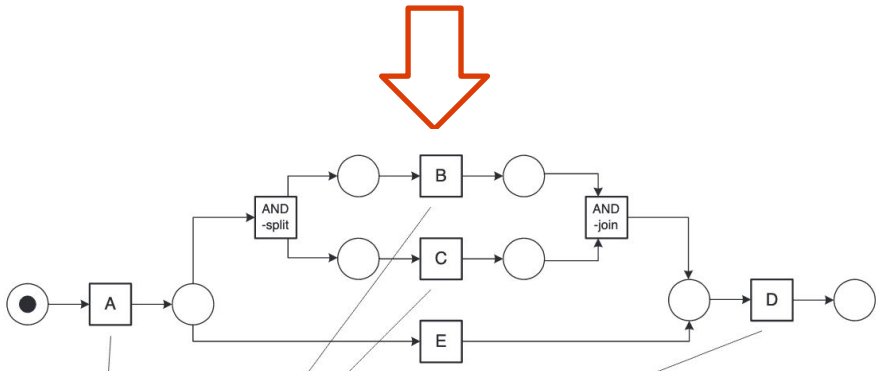
Algorithms

The main algorithms we use for Process Mining are:

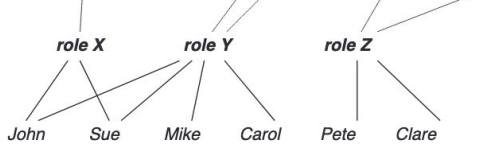
- **Alpha Miner:** Connects observed data and event logs in the Discovery phase. Returns a model that show the sequence of events.
- **Heuristic Miner:** Evolution of Alpha Miner. It copes well with noise and takes frequencies in account. It is use to show the main behaviour.
- **Generic Miner:** Use when we have incomplete data or too much noise. It generates a Petri net considering the missing data.
- **Inductive Miner:** Provides a soundless model with Good values of fitness.
- **Fuzzy Miner:** Specially useful for processes which are poorly structured.

case id	activity id	originator	time stamp
case 1	activity A	John	9-3-2004:15.01
case 2	activity A	John	9-3-2004:15.12
case 3	activity A	Sue	9-3-2004:16.03
case 3	activity B	Carol	9-3-2004:16.07
case 1	activity B	Mike	9-3-2004:18.25
case 1	activity C	John	10-3-2004:9.23
case 2	activity C	Mike	10-3-2004:10.34
case 4	activity A	Sue	10-3-2004:10.35
case 2	activity B	John	10-3-2004:12.34
case 2	activity D	Pete	10-3-2004:12.50
case 5	activity A	Sue	10-3-2004:13.05
case 4	activity C	Carol	11-3-2004:10.12
case 1	activity D	Pete	11-3-2004:10.14
case 3	activity C	Sue	11-3-2004:10.44
case 3	activity D	Pete	11-3-2004:11.03
case 4	activity B	Sue	11-3-2004:11.18
case 5	activity E	Clare	11-3-2004:12.22
case 5	activity D	Clare	11-3-2004:14.34
case 4	activity D	Pete	11-3-2004:15.56

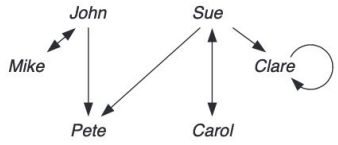
Heuristic *Miner*



(a) The control-flow structure expressed in terms of a Petri net.



(b) The organizational structure expressed in terms of an activity-role-performer diagram.

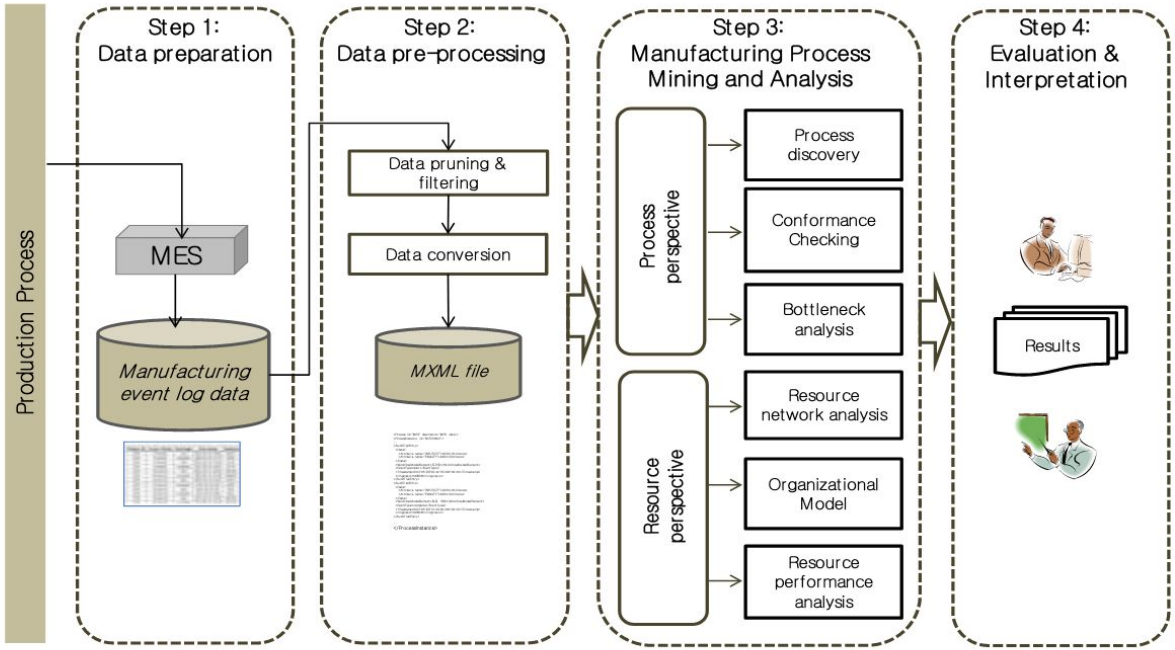


(c) A sociogram based on transfer of work.

Use cases

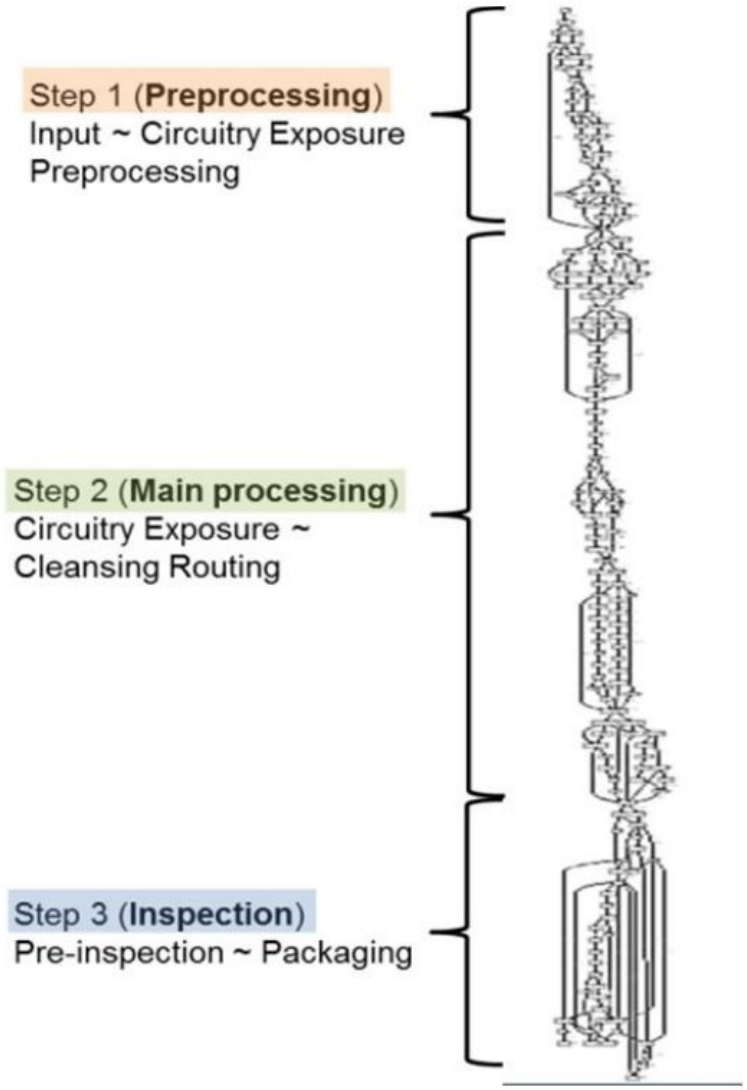
Samsung Electro-Mechanics

- Samsung Electro-Mechanics is one of the largest manufacturers in the whole are of East Asia.
- There were 11,226 cases, almost 1 million events (related to the year 2012) using a MES system.



Use cases

Samsung Electro-Mechanics

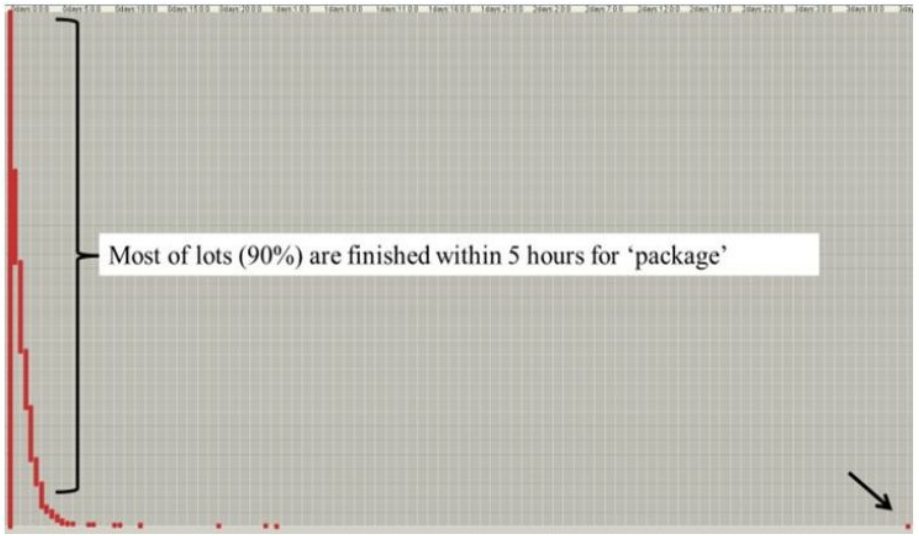


- Using process perspective we extract a manufacturing process and analysizes it (using the Heuristic Algorithm).
- The model starts with the input tasks and outputs the Packaging task.
- The model is then discovered and a conformance check is applied.
- This model found abnormalities on the delivery inspection and final test.

Use cases

Samsung Electro-Mechanics

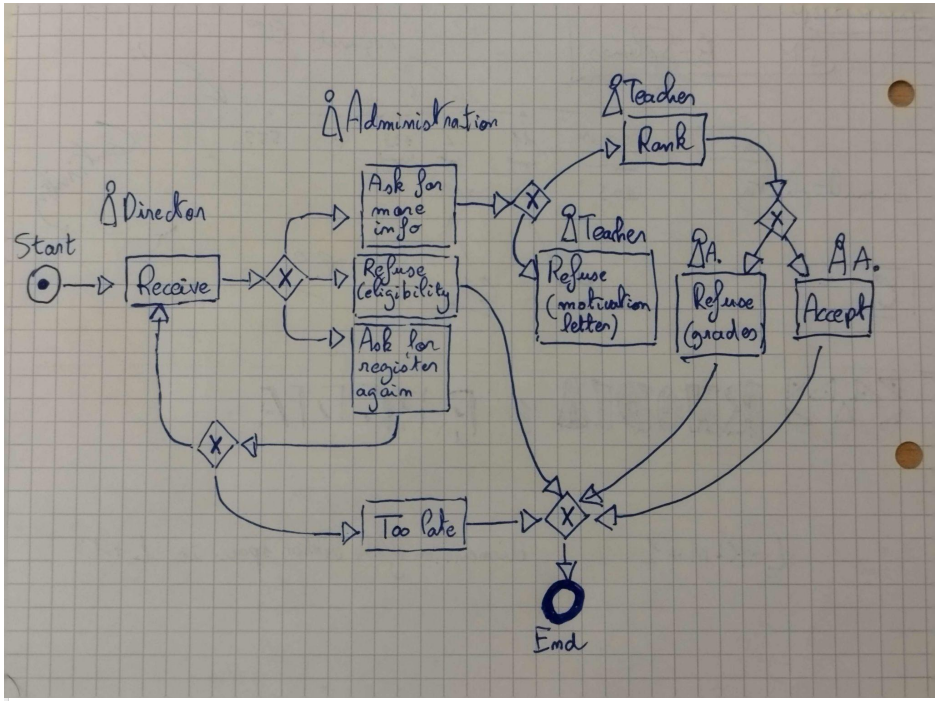
Machine	Package
N28026	5,612
N28027	2,965
N20031	2,143
N20030	500
N14074	2
N28025	2
N28146	2



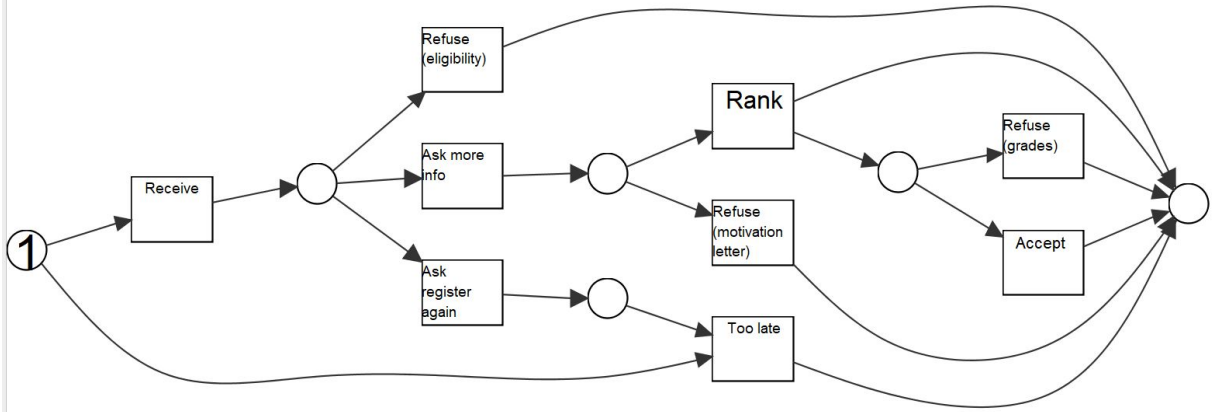
Experiment

- Idea: get a better understanding of PM using a PM tool and a Discovery Algorithm
- We developed rules for a business end-to-end process (students applying for a university)
- We developed an application generating random event logs following the rules
- We used ProM Lite (open-source tool) with Alpha Miner to generate a BPM only based on the generated event log

Experiment



- The BPM generated is surprisingly close to the rules we specified
- High (almost perfect) fitness, simplicity, precision and generalisation
- This experiment helped us putting in practice our knowledge
- Demonstration



Conclusion

- **Process Mining**
  - Focuses on processes, not on data
  - Use real facts to develop BPM, not human opinions
- What are the **objectives** and **benefits** of Process Mining?
  - High scale analysis of a Business Model and Accuracy based on fact
  - No bottlenecks or deviation
  - Continuously monitoring processes and can be apply to ANY industry
  - Applicable in ANY Area
- What is the **potential** of Process Mining?
  - Save costs, increase productivity, improve performance

Conclusion

- Types of Process Mining:
  - **Discovery:** automatically construct a model based on observable events, without outside influence.
  - **Conformance:** checks whether the actual model is in fact reflected in practice. It identifies any deviation of the model from real data and looks for overfitting or underfitting cases. It is mainly used to check the accuracy of a model.
  - **Enhancement:** extends the previous model and could make it useful for predictions and recommendations.
- Many tools and algorithms supporting Process Mining, even open-source softwares.

Conclusion

- **For the future:**
  - There are many things to research in the future within the framework of Process Mining.
  - Very interesting and useful for many companies to get a better understanding of their business to improve quality and productivity.

Literature

Weijters, A. & Aalst, Wil & Medeiros, Alves. (2006). Process Mining with the Heuristics Miner-algorithm.

Ghawi, R., & Pfeffer, J. (2016). Mining social networks from linked open data

Gupta, E., & Gupta, A. E. (2014). Process mining algorithms process mining algorithms

Discovering knowledge in data : An introduction to data mining / daniel T. larose, chantal D. larose

Reinkemeyer Editor, L. (2020). Process mining in action : Principles, use cases and outlook.

van der Aalst, W. (2012). Process mining. ACM Transactions on Management Information Systems.

van der Aalst, W., Adriansyah, A., de Medeiros, A.,Karla Alves, Arcieri, F., Baier, T., Blickle, T., . . . Wynn, M. (2012). Process mining manifesto.

Thank you for your attention!

Vielen Dank für Ihre Aufmerksamkeit!