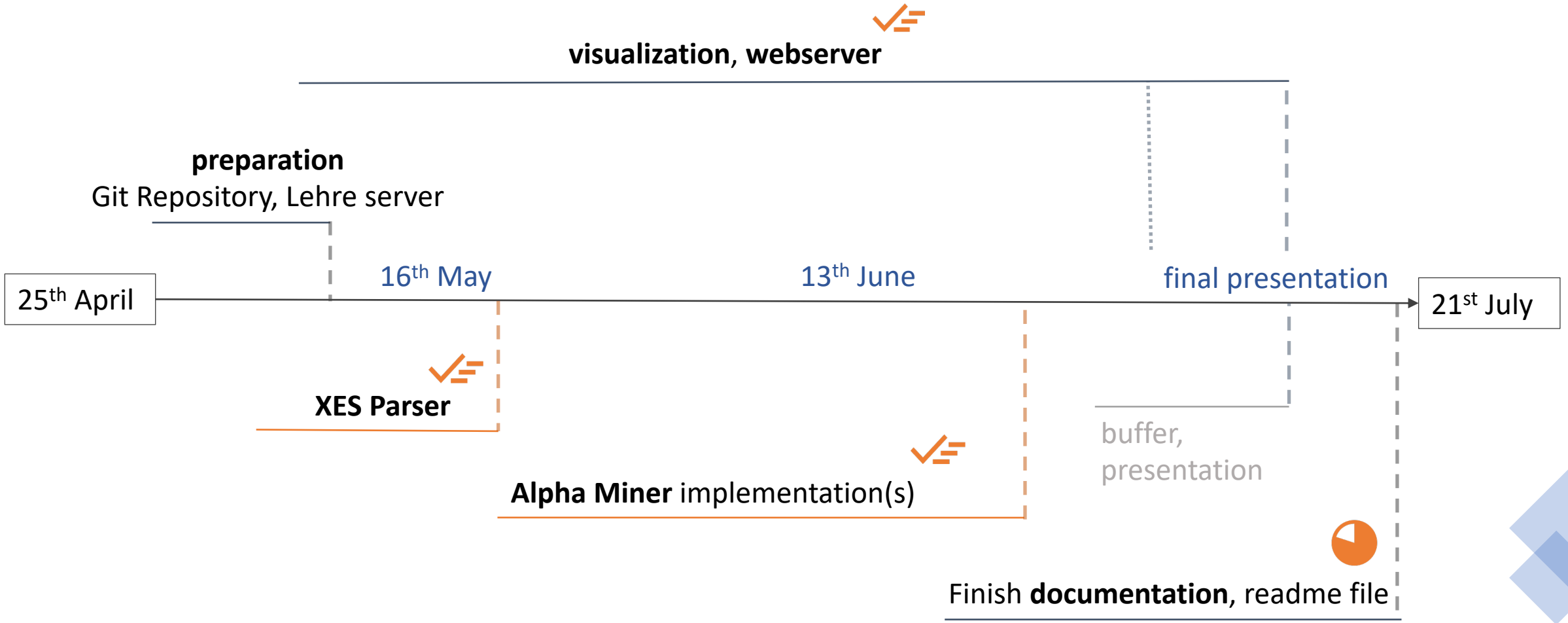
Practical Course: Process Mining – Final Presentation

Yidi Ma

18th July 2022

Original schedule



XES Parser

xml.etree.ElementTree module

- Separation of XES file into “header” and “body” and conversion to XML files
- Only needed <body.xml> for parsing

Parsing with data structure

- Generation of Event, Attribute, and Trace objects

Challenge: Lifecycle Transitions

- Cannot delete events with <lifecycle transition:complete> (necessary for statistics)
- Addition of a new attribute *<start>:* indicates whether an event is start of a lifecycle or not (ignored by Alpha Miner)

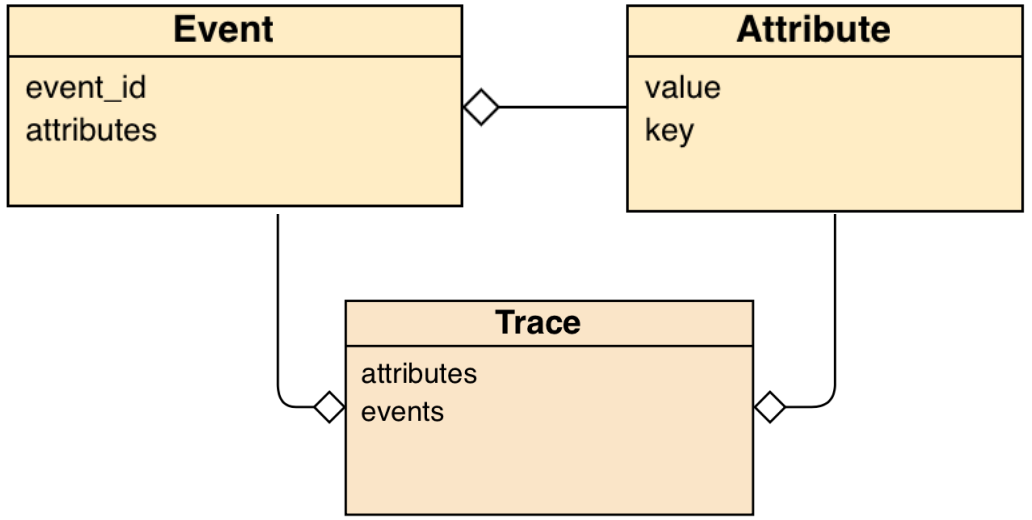


Fig. 1: class diagram for data structure used for parsing

Alpha Miner – Algorithm

Step 4

- **Following relation**: {event x}, {events that follow x}
  - $[(\{a\},\{b,e,c\}),(\{b\},\{c,d\}),(\{c\},\{d,b\})]$ → size of 1st element: 1
- **Fol. rel. reversed**: {events directly before x}, {event x}
  - $[(\{a,c\},\{b\}),(\{a,b\},\{c\}),(\{c,b,e\},\{d\})]$ → size of 2nd element: 1
- **Separate sets**, if events are unrelated with each other
  - $(\{a\},\{b,e,c\})$ → $(\{a\},\{b,e\}),(\{a\},\{e,c\})$
  - $(\{c,b,e\},\{d\})$ → $(\{c,e\},\{d\}),(\{e,b\},\{d\})$
- **Combine sets**, if possible

L1: <a, b, c, d>, <a, c, b, d>, <a, e, d>

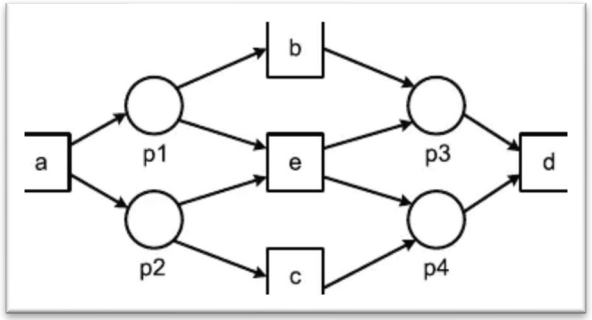


Fig. 2: Simplified petri net and traces for L1

Steps 1 – 3, 5, 6, 7: same as in paper

Alpha Miner - Statistics

Technologies: Matplotlib, Pandas

Type of statistics:

1. Total number of occurrences of each event
2. Duration of traces (7 shortest, 7 longest)
   - To emphasize the contrast between longest and shortest traces: y-axis might begin at a higher value than 00:00:00
   - Mean duration
   - Number of traces

- Difficulties with **duration as y-axis**
  - Matplotlib cannot sort *datetime-objects*

  → own **formatter:** HH:MM:SS
- Long computing times for XES files with many traces

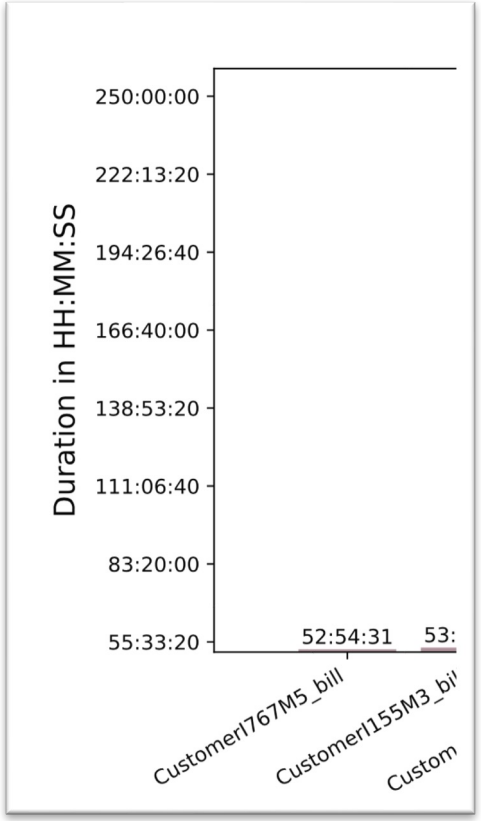


Fig. 3: Y-axis from diagram that shows the duration of traces

Alpha Miner – Remarks

- Incorrect results for **limitation-XES files**
  - **Lifecycle transitions** were handled incorrectly → events with attribute `start = False` are ignored
  - Algorithm could handle **loops of length 1**, but Alpha Miner should not be able to do so
- Testing with Python **Unittest framework**
  - Creating *Edge, Transition, Place* objects manually and compare with results from Alpha Miner
  - One test for each given XES file
- Difficulties with **visualization** (Graphviz)
  - Missing token in start-place → add note/remark?
  - Edges could overlap with places/transitions
  - Different petri-nets when reload (non-deterministic)
  - Places, Transitions: name mapping

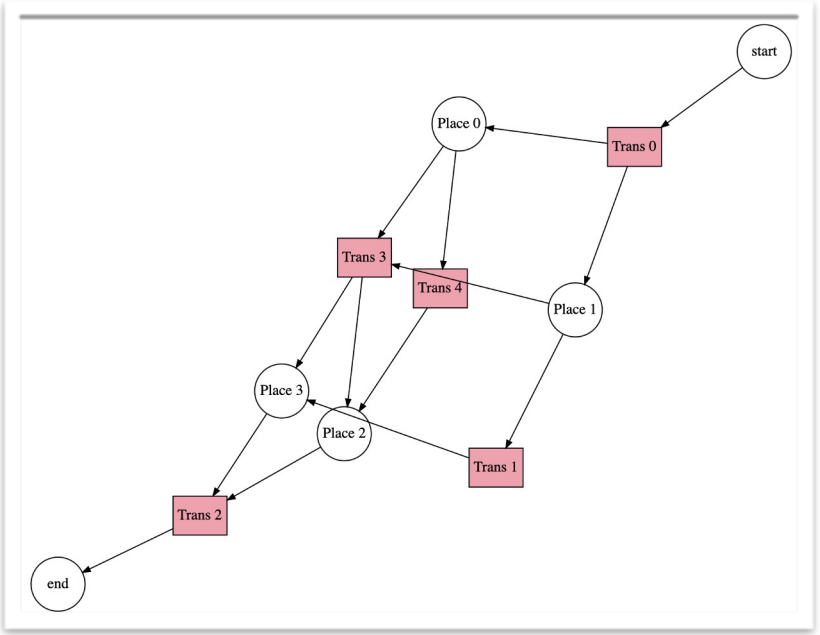


Fig. 4: Petri-net for L1.xes generated from algorithm

Webserver

- **Technologies**: Flask, HTML, CSS

- Added functionalities:
  - **Download** as PDF/SVG
  - **Upload XES files** and generate respective petri-net and statistics

- **Challenge: Deployment on server**
  - the automatically added prefix “/ports/<port number>/” was unknown to Flask
  - Workaround solution with Blueprint: hardcoded the prefix for deployment on server

- DEMO