Advanced Visualization for Process Mining
(Extended Abstract)

Tiago Lemos Ribeiro
(64066)

Instituto Superior Técnico – Taguspark Campus

May 2016

Abstract

Process mining techniques are used to analyze business processes from event logs. Typically, such techniques generate a model in the form of a graph representing the process behavior extracted from a sequence of events. Depending on this behavior, the graph can be very complex and difficult to analyze. The objective of this work is to create a process mining tool that takes advantage of Web technologies to create interactive models, which allow the user to rearrange the nodes and edges of the graph, filter elements, etc., in order to facilitate the understanding of these models.

Keywords: Process mining, process modeling, interactive visualization, dynamic content on the Web

1. Introduction

Process mining is a relatively new discipline that aims to process activity records from businesses and more, to extract models that reflect the operation of these activities from a global point of view[5]. The analyzed data and models generated have several applications such as: process discovery, compliance testing, forecasting, optimization and analysis of social networks within the organization or even technological systems such as websites[3].

Although the process mining techniques have evolved, the tools used to explore these techniques do not yet take advantage of Web technologies. These technologies offer advantages in terms of connectivity, distribution, ease of development and innovation of the libraries that make up this ecosystem.

This work offers a modern tool for data process mining, taking advantage of Web technologies and recent concepts of visual application development such as reactivity and responsiveness to provide a tool based on the existing ones but with interactive elements for graphic manipulation that enhance the user experience and facilitate understanding and analysis of the generated models.

2. Process mining

Process mining focuses on the analysis of events recorded by processes and building business models or the optimization of existing models using appropriate algorithms[2]. Process mining includes the discovery of processes from event logs, extraction and presentation of business models from data, compliance testing between generated models and the supposed models, and automation techniques to optimize the discovered models[5].

![Control-flow model generated with Disco](image)

Figure 1: Control-flow model generated with Disco

In process mining, event logs which can come in many formats such as input files or tables from a database etc. are used for model discovery[4], conformance testing between extracted models and existing models to verify if the supposed models correspond to reality and enhancement [5] of existing models such as extending existing models to include information coming from the records like the occurrence date or average process-time and increase the detail of existing models. Although there are other use cases, these are the most frequent and this work focuses mostly on process discovery and model generation.
2.1. Existing tools
There are several tools used in process mining for different tasks such as automatic generation of models or performing compliance testing. Two popular tools are ProM[6] and Disco and use these two as references for this project. Both are very useful tools to generate models from activity records.

The models generated by these applications are static and have limited interactive features, they don’t allow the user to change the arrangement of the elements of the created models to their own liking. The interactive manipulation of visual models is a modern application development trend which makes the process of analysis and exploration more intuitive and improves the user experience. It’s possible to add this interactive component using modern Web technologies and libraries for dynamic visualization, as well as creating a modern application that takes advantage of other benefits that result from the use of these technologies.

3. Dynamic visualization on the Web
There has been a constant evolution of the languages HTML, CSS and Javascript and the entire ecosystem of libraries, tools and browsers related to Web technologies. The browser is a central component of Web technologies which works as a display and communication engine for applications providing useful APIs, multiple features and bridging the gap between the Web applications or Web pages and operative system installed on the devices. Web applications are developed using scripts and markup languages that are interpreted by the browsers instead of compiled to machine code, this makes Web applications independent of the operating system and readily available from different devices, it also improves the development process speed and costs.

In addition to advances in Web technologies used to build applications, also paradigms have evolved with the introduction of concepts such as reactivity and MVVM design patterns [1]. Now applications are built increasingly dynamic and interactive providing a more intuitive and efficient user experience. Web technologies are becoming standard in application development and some of its innovations are still not accompanied by the process mining applications that are built as desktop applications using more traditional approaches.

For this study several libraries and tools were selected from a large ecosystem that contribute to the final quality of the project, allowing for interactivity, modularity and presentation that would be difficult to obtain without an increase of resources and time spent on application development.

4. Application development
The developed process mining tool analyses event logs and generates visual models according to different perspectives of process mining like most process mining tools. These features were improved using the web technologies and libraries that enable interactive visualization and manipulation of user-generated models.

To obtain a useful tool in the analysis of real event logs through process mining, other features such as filtering were added which allow for the user to nodes and arcs according to the occurrence counts or using a straightforward nodes filter list.

It was also given importance to the application architecture and production techniques to achieve a modular, scalable application which incorporates some of the most popular process mining techniques and algorithms. The generated models are presented to the user taking advantage of innovative libraries for vector graphics visualization and interaction which have been adapted and integrated in this application.

5. Tests with real-world event logs
To test the application in terms of compatibility and usability and to test its usefulness in real situations, tests were conducted using business event logs provided by the Business Business Processing Intelligence Challenge(BPIC). The BPIC holds an annual challenge of process mining which publishes business event logs so that participants can apply different process mining techniques and tools to understand the functioning of these companies and detect possible problems or sug-
The use of real event logs allowed to detect and correct problems in the application in terms of importing records with data containing badly formatted data such as unfilled column fields, they also pushed the application performance in when processing records with several dozen megabytes of data and hundreds of thousands of entries, which although is not unusual for real records, those are much larger than the event logs that were used during development.

6. Conclusion

The Web-based application developed has a scalable modular structure and with core mechanisms to import and process records in CSV format to present interactive models that facilitate the analysis and the usability of the tool. The use of Web technologies with access to the vast ecosystem of libraries and Javascript tools contributed to rapid development and prototyping.

In comparison with existing commercial tools, to some extent, the developed application offers similar results and features when it comes to importing records and generating visual models, plus the application can be downloaded and executed as a typical web page, requiring only a browser installed on the user the machine and internet access. The interactive component also helps the process of analysis of the generated models and contributes to an overall better user experience.

References