RWTH AACHEN UNIVERSITY

Customizable interface for LTL conformance checking

Project Initiation Document

Group 2

Teodora Staneva
Mohamed Amine Kooli
Ugo Detaille
Vishisht Choudhary
Daniel Henel

Aachen, October 2022

Contents

1. Introduction

In this project, we aim to implement a customizable interface for LTL checking and would like to introduce a new system covering the following features:

- Composition of filtering system in accordance with the rules of LTL checking
- Representation of event logs deviating from the rules
- Identification of point of deviation

In chapter 2 we will touch upon the relevant business cases which LTL model checking can be applied in. After listing x examples, we will present the benefits of our product for Internship.

In chapter 3 we present three examples of the usage of LTL Checking. Three different platforms (bank, online shop, online encyclopaedia) are introduced and for each they are three different cases where LTL Checking is proved useful.

In chapter 4 we will discuss the feasibility of our product by taking theoretical and technical aspects into account. We will support the analysis by presenting similar use cases.

In chapter 5 we provide a GANTT chart outlining and explaining the time frame of our project.

In chapter 6 the tools used during our project will be listed along with an explanation for the usage of the specific tools.

Chapter 7 of this document is dedicated to revising the development of our team. We will also introduce ourselves and offer an impression of our personal profiles.

In chapter 8 we give a brief explanation of the roles needed for the project and designate a member of our team to each role.

Chapter 9 concludes with a Phase review, where each member describes the process of creating this document and evaluates the work done.

2. Business Case

2.1. Restructuring of IT Landscape

Company X is a large bank with many different practices ranging from customer relations to investment banking. During day-to-day business operations the bank uses various IT-Applications. Since the bank has been existing for many years and applications

are rather added than gotten rid of, the bank has found many applications that are not being used anymore. And since the bank does not want to unnecessarily pay for licences they do not need, company X would like to use process mining to identify which applications are actually being used during the business processes. Following this practice company X will be able to unsubscribe from applications they do not use and will avoid paying for those licences. But because they want to be sure they are not making any mistakes with crucial bank data; they would also like to verify their models using conformance checking. Using our application company X will be able to verify their business process models and will be able to delete applications from their IT landscape with a high certainty.

2.2. Police Department

The Fiction City Police Department has recently been accused of not handling incidents and arrests the right way. Neglect, Violence and arguably unfounded arrests are amongst the accusations from the citizens against their city's police department. Hence, the Fiction City Police Department needs to identify eventual neglect in the handling of cases to regain the trust of its citizens. In fact, this task can neither be solved based on raw data of arrests or witnesses' statements, nor on abstract improvements of the code of conduct. To answer this question, precise understanding of the real behaviour of police agents is required, i.e. understanding how the agents really work. That is why we, a process mining business, are best suited to answer the Fiction City Police Department's needs. By recording all interventions in and outside of the office through body cameras and neutral security staff documenting the police department's work and behaviour, we can obtain enough data on how incidents are being handled. The surveillance of the agents from neutral security staff ensures the **"four-eyes"-principle**. Next, our powerful process mining algorithms will be key to create a process model, which will then be compared to the formal code of conduct to understand eventual deviations and breaches of the code of conduct from the Police Department. Formally, we will provide our client with the following services:

1. Precise visualization of case handling in and outside the office through powerful process mining algorithms
2. Comparison of the results from 1) to the code of conduct, outlining strong deviations as well as strong similarities. This will help our client identify the strengths and weaknesses of the police staff.
3. Standardized data storing instructions to facilitate future work, so that our company will be able to reanalyse new sample data more efficiently.
4. Absolute discretion and privacy of personal data of all staff members across the total organization in order to minimize the risk of exposure.

3. Examples of using LTL Checking

3.1 Bank

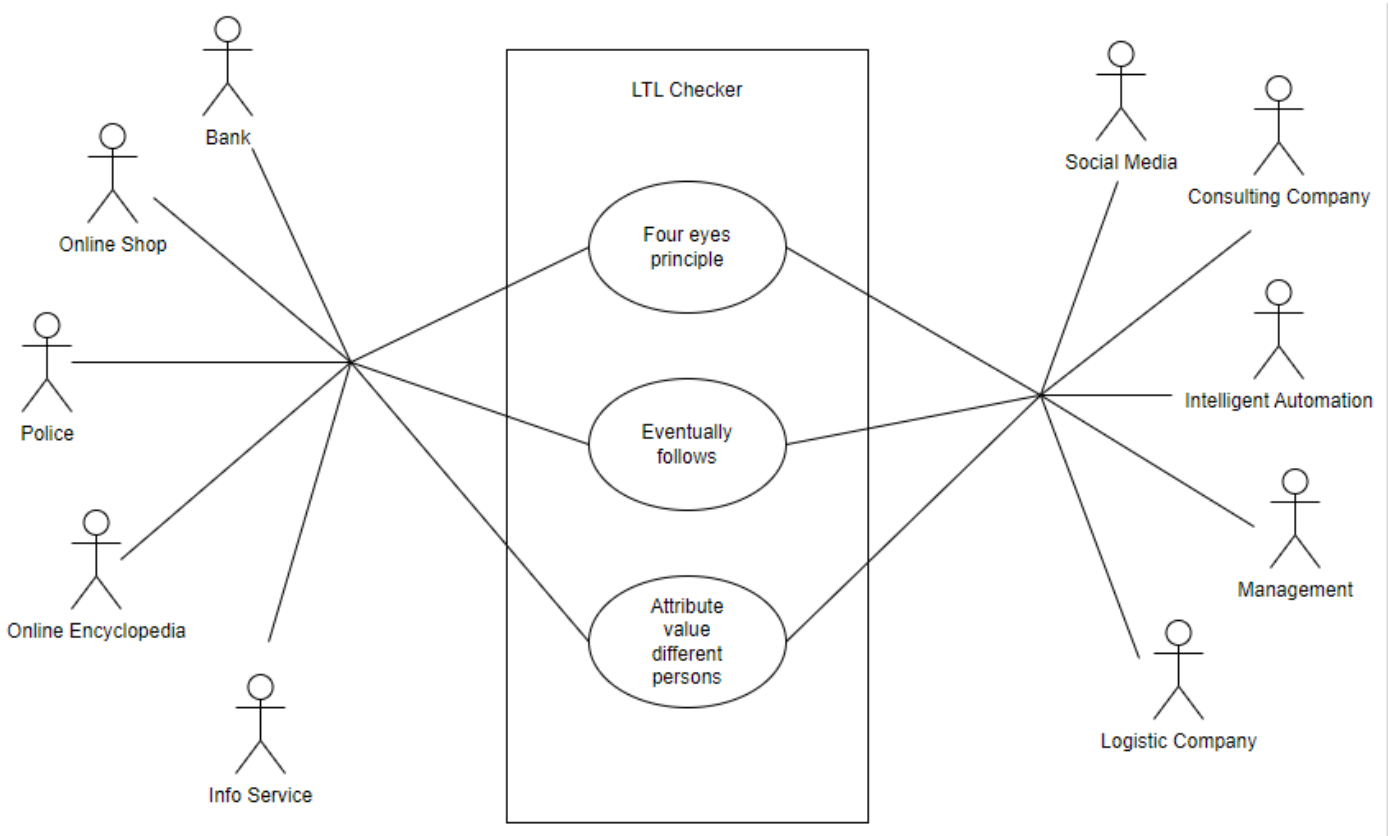
Case 1	When someone wants to get a loan, first the agreement must be signed by a customer and then approved by a bank employee, who must check if the customer has the permission to get the loan. We can provide **"Four eyes principle"** filtering to analyse log data in order to avoid such situations that employees approve their own credits.
Case 2	The bank's clients need to know statuses of their transactions, i.e. whether they are approved or rejected. We can use the "**Eventually follows"** filter to divide the transactions into these two groups. When a transaction is approved, the "Approved" activity must follow the "Start" activity. When a transaction is rejected, the "Rejected" activity must follow the "Start" activity.
Case 3	A client can be logged into his bank account on multiple devices simultaneously. Thanks to the **"Attribute value different persons"** filter we can detect insecure situations when the "Logged in" activity was performed on more than one device.

3.2 Online shop

Case 1	In a propriate way, when a client buys something on the internet, the seller sent him ordered product. By assumption, both the client and the seller should be two different people. Otherwise, a person could buy something from himself to generate artificial income for example. Using the **"Four eyes principle"** filter we can detect such situations and then block the account of said person.
Case 2	Users can bookmark the products that they like. Of course, not all subscribed items will be sold. Using the **"Eventually follows"** filter, we can check what action followed the "Bookmark" activity. That way, we can divide the action into two distinct groups: successfully sold and unsubscribed. Subsequently, we can analyze the data in order to create more individual advertisements for each customer.
Case 3	One of the forms of sale is a bidding. The most popular biddings should be shown on the top of the website. The **"Attribute value different persons"** filter could help us to find such biddings which have more than one auctioneer.

3.3 Online Encyclopaedia Service

Case 1	To avoid such situations where the information in the encyclopaedia isn't correct, we could provide a verification flow that each article after the changes must be approved by a person other than the author of the changes. The **"Four eyes principle"** filter would be helpful in verifying that the approver is really a different user.
Case 2	There are many links in each article. The user can fast check what the difficult word in the text means just by clicking on the word. We could use the **"Eventually follows"** filter to find out if the user checked a given word.
Case 3	In the online encyclopaedia service, each article can be edited multiple times. The developers would create a new button called "Show author" or "Show authors" depending on the number of authors. Using the **"Attribute value different persons"** filter, we can find out very fast which page has more than one author, because in the log data each change on the article's page is saved as a "Change" activity in the same case assigned to the given article.



LTL Checking has many different uses.

4. Feasibility Study

4.1. Theoretical and Technical Feasibility

This project focuses on creating an interface for conformance checking. In process mining, conformance checking involves verifying process models which have been constructed using real-life event log data. By looking at discrepancies between the process model and the real-life data, we can inspect the quality of our model and tweak it to further represent the real sequence of events during a process. Our approach during this project is to apply model-checking to Linear Temporal Logic (LTL) models. LTL is a logic allowing to depict processes using transition diagrams and mathematical formulae.

The translation of such diagrams into code is possible by using the ProM framework [1]. ProM is a so-called pluggable environment, which means it can be modified with packages which are relevant to the tool we want to implement. The packages to be used are:

1) Mining Plugin

   This Plugin is used to load and filter our MXML File (containing our event data)

2) Import Plugin

   This Plugin is used to load LTL files into the program

In Combination with ProM we will also use the frameworks Declare and P4MPy which we will be able to visualize our processes in form of a colorized finite automata with.

4.2. Use Cases

4.2.1. Python

Feasibility Study: Python [1] [2]

Python is a high-level, general-purpose programming language. The programming language has gained great popularity in recent years and is amongst the most-used programming languages today, especially in business and web application environments. In fact, Python is also our preferred programming language, and here is why:

1) Design philosophy: Python aims to make its code very readable. As we are working in a team, it is crucial that we can all read each other's code. Pythons easy access for all coders makes it the language of choice. Additionally, Python is simply a language our group likes to use. We all know how important a good work environment is, so using tools and languages one is comfortable with is imperative for a good result.
2) Standard library: Python is probably most known for its very broad, standard library. As of June 2022, PyPI contains over 380.000 packages, equipped with a wide range of functionalities. Not only are Django and Flask two of the most famous web frameworks for web applications, but libraries like NumPy, SageMath and MatplotLib allow also for powerful mathematical and numerical tools. Finally, another vastly used application of Python lies in the fields of artificial intelligence and machine learning. Again, libraries such as TensorFlow, Scikit-learn, Pandas and others are very powerful and one of the main reasons Python is nowadays the go-to programming language when it comes to Data Science. Plus, Python is open source.
3) Scalability and security: The web frameworks provided in the standard library guarantee high security, convenience and scalability to its applications.
4) Global representation: For future work, we may have to work together with other teams or outsource our code for others. With over 8 million programmers worldwide, Python can be considered a language many programmers will have in their arsenal. This makes our code more accessible

to third parties. Big Tech Companies that use Python code are e.g., Netflix, Spotify, Facebook, Microsoft or Pinterest, just to name a few.

5) Swift integrations: Python enables smooth integration of and transition from other (older) programming languages such as Java, PHP or others. Therefore, we will be able to provide new additions to our client's already existing information system, preserving valuable, established code and combining it with new additions in Python.

In particular, we intend on using the Flask framework for this application. But why Flask, one may ask. The following aspects are just some of the reasons that make Flask the right choice: [3] [4]

1) Popularity: Flask is very popular and there is great documentation of it in many forums: Along with Django, it was the best rated Python web-development framework as of 2018 according to GitHub.
2) Simplicity: Containing only two dependencies, „Werkzeug" and „Jinja2" makes Flask also very accessible for all programmers. Simplicity is one of the key motives of Flask, allowing more freedom for the developers. Still, a great number of external libraries and plug-ins are available to extend the minimalistic framework. In fact, that is one of the reasons why Flask is considered more beginner-friendly than Django.
3) Low-level Dependencies: Low-level dependencies lead to lower levels of abstractions. Flask's lightweight modular design enables higher performance since there will not be any redundancies taking up a considerable amount of time by using many packages.
4) Individual framework: Due to its simplicity, Flask does not really have any form of conventions or limitations on how it is supposed to be used or how code should be implemented. Again, this gives us developers the freedom to create results perfectly suited to our clients' needs.

4.2.2. Process Mining and Conformance Checking

Granada, located in Southern Spain, is one of the most vivid cities of an already very touristic country. In fact, Granada is known for its many touristic and agriculture businesses. Consequently, every year, the municipality of the city must deal with many appeals against its collection of taxes and/or revenues. Over time, the population started complaining about heavy delays in the handling of dossiers. Even though the responsible party managed to identify the time gaps where dossiers handling would come to a hold, they were not able to identify the cause of those delays. So, together with the tax collection department of the Granada Council, the Department of Computer Science of the University of Granada applied a process

mining analysis to check which steps of the regulatory business model were creating said delays. Applying process mining and conformance checking techniques outlined flaws not only in the real processes, but also in the business model itself. [5]

The technical setup was as follows: [6] [7]

1) Every complaint in the database contained a dossier ID, which was used as a case ID

Additionally, each dossier contained information on the

2) Activities executed on the dossier (activity)
3) Whom the activities were executed by (employee)
4) Start date of each step of the dossier (timestamp)

After long preprocessing work of joining multiple tables in different formats from multiple departments into one large, scaled and process mining applicable table as a CVS file, imported to Disco containing 2511 cases and 7582 events, the event log was finally created.

Against the city‘s starting assumption of the employees deviating from the business model being responsible for the delays, an in-depth process mining analysis revealed otherwise:

1) Dossier transfer: The results of the process map, based on Total duration, revealed that the major delay occurred in the transfer of the dossiers between two departments. This transfer alone took between 18-24 weeks.

2) Documentation work: Further analysis revealed bottlenecks in the handling of the dossiers, mostly at very specific points in time. Interviewing the employees revealed that a certain classification task was responsible. Its very complex nature on the one hand, as well as physical documentation work on the other hand, required more trained staff. It turned out that because this task was done manually, only little to no data was recorded on how much time the sub activities of this documentation work took in reality.

3) Redundancies: Finally, the analysis showed major redundancies in the process model: Electronically stored information had to be re-registered manually in different steps. This was a very error-prone and time-costing task that had to be modified.

Hence, the recommendations were the following:

A) Adapting the business model: The classification task was divided into subtasks to point out its importance. Also, this enabled a more efficient distribution of the resources and the staff.

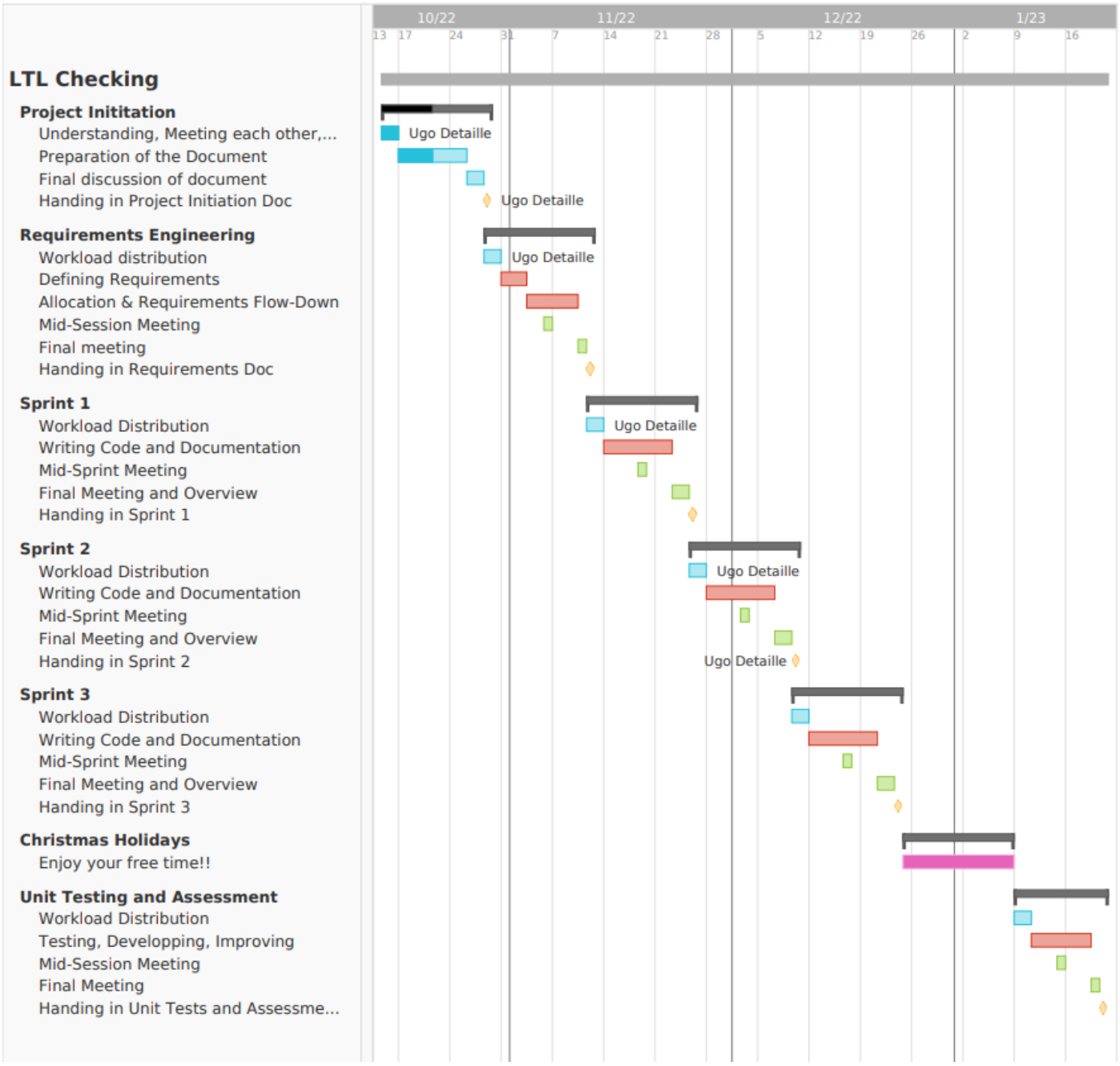
B) Improving the lack of traceability within the computer systems. It was not official neglect, but rather the lack of alignment with the business process that was responsible for making it so hard to identify the delay-causing activities.

C) Improving the departments‘ information and communication systems. This avoids the redundant manual re-entry of already existing data. A system of automatic transfer of dossiers between the departments involved would also reduce the massive delays found in the analysis.

The improvements included better transparency of the workload, the resources and the task management, a key factor in a functioning business model. Plus, workload was then scaled, so that it could be measured. Consequently, more precise responses about the expected completion date of a dossier could be given to the citizens. Also, the work of the government officials was now valued much more by not only the population, but also the city’s supervisors, resulting in a better work atmosphere. Finally, the metrics and formats put in place by the process mining team would enable new analyses of the municipality to be way more efficient and straightforward in the future.

5. Project Charter

Project Initiation (14.10.2022 - 28.10.2022)
Requirements Engineering (28.10.2022 - 11.11.2022)
Sprint 1 (11.11.2022 - 25.11.2022)
Sprint 2 (25.11.2022 - 09.12.2022)
Sprint 3 (09.12.2022 - 23.12.2022)
Christmas Holidays (24.12.2022 - 08.01.2023)
Testing and Assessment (09.01.2023 - 20.01.2023)



6. Tools

To ensure a productive environment and successful completion of the project we use the following tools:

6.1. Trello

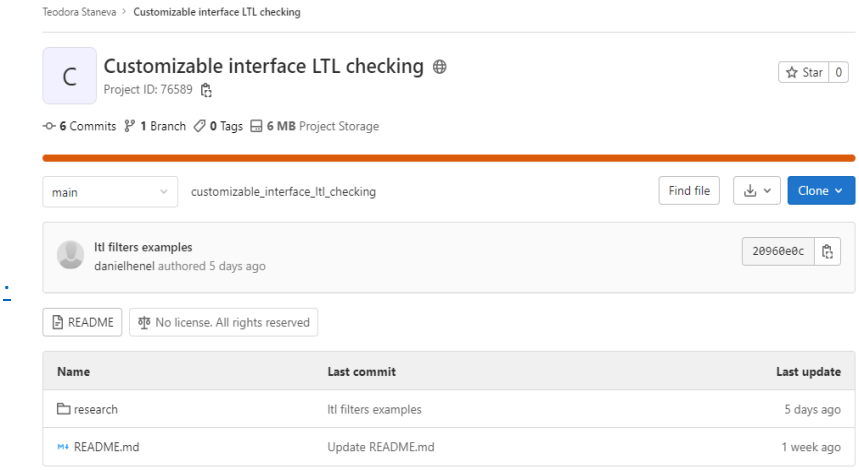
Trello is an online platform, where multiple users can plan, coordinate, and designate tasks. Our team has created a Trello Dashboard based on the SCRUM model. We have utilized the tools provided from the platform and have different lists for planed, running and completed tasks and a backlog. We use colour coded tags to differentiate Milestones and later on will use the latter also while programming and dividing bigger tasks into separate smaller parts. Tagging members has also proved useful and each member knows what their job is and using deadlines helps with keeping up with the workload. Working in a team can sometimes lead to miscommunication and confusion concerting the tasks that need to be done, but Trello is a useful platform that can minimize uncertainty and improve productivity.

Link to our Trello dashboard:

[Process Conformance Checking - Project Group 2]

6.2. GitLab RWTH

As students of RWTH Aachen we can take advantage of GitLab that the university provides us. It is an open-source version control system. We have created a repository, where we will upload our project files, and manage our code. Multiple people can have access to the project and our whole team will be managing and contributing via GitLab.
There we will be uploading important files and useful information about the project. We will take use of the version control and we will have the possibility to take a look at an older version of our code.



Link to our GitLab repository:

[Customizable interface LTL checking]

6.3. Discord

Discord is a platform where users can communicate with each other via direct messages or users can create servers with different types of channels. We have created our own Discord server with separate text channels for general discussion, materials and notes and two voice channels. We will hold our meetings there and everyone can join easily. Discussion about the tasks and project also takes place in Discord. Users can tag each other when needed and communication is faster and more efficient than communicating with emails.

6.4. PM4Py

Pm4py is an open-source python library built by Fraunhofer Institute for Applied Information Technology to support Process Mining. We are going to use this library to implement conformance checking.

7. About us

7.1. Vishisht Choudhary

I am a bachelor computer science student in the 5th semester at RWTH Aachen University and I am taking Medicine and Business Administration as my secondary subjects. I have coding experience in Java, C and Python and have already had some contact with data science topics. In January I will be going abroad and will also be takin courses in Artificial Intelligence, Data Science and Statistics. In general, I enjoy learning about new topics and believe this course can help me get a very thorough introduction to process mining. Since I have never really programmed a proper team, I hope that this course can also help me get insights into an agile developer environment and improve my python skills.

7.2. Daniel Henel

I'm an Erasmus student at RWTH University, but actually my home university is AGH University of Science and Technology (Cracow, Poland). I study Computer Science and Intelligent Systems. It's my 5th semester (bachelor's degree). I'm particularly interested in widely understood artificial intelligence and data science. I have some experience with machine learning, data engineering, cloud computing and virtual reality. I have just joined this course because process mining is one of the branches of data science and it’s also used to collect data for machine learning. We can often extract a lot of surprising information from the data and that's very exciting. During the course I want to develop my analytic skills and work together nicely with my new friends.

7.3. Mohamed Amine Kooli

I am a bachelor student in computer science in the 5th semester at the RWTH Aachen with business administration as a secondary subject. So far, I have worked with some programming languages such as java, c, python etc. Solving problems and getting involved in projects are such challenges that I always want to take on. During my previous semester, I took data-oriented courses such as databases and the Process and Data Science proseminar, in which I focused on the topic of data mining techniques. I was so interested in the field of data science and wanted to learn more and as process mining is part of this vast field, now I am taking "process mining: data science in action" on coursera and chose to work on this project focusing on conformance checking so that I have the chance to learn and master this process mining technique as well as improve my python programming skills. What I expect from this course is to work as a team to carry out our project and be able to gain my first practical experience in process mining.

7.4 Ugo Detaille

I am a Computer Science and Mathematics student at RWTH Aachen University. In Computer Science, I am currently in my 5th semester. Therefore, my specialization lies in mathematics. So far, I have mainly coded in Java and C, as well as some experimental work in C++. Java and C are the programming languages I have personally used on private projects as well as over the course of my studies. As for C++, I have read guidelines, tutorials and books for self-interest

As a mathematics passionate, I really enjoy the theoretical aspects of Computer Science. Nonetheless, I am aware of the importance of learning how to apply theoretical frames into real life applications. In fact, Data Science is a field I have not really had the chance to discover yet and therefore I was really excited to enroll in a programming course on a Data Science related, real-life applicable topic. Additionally, conformance checking uses frameworks of theoretical Computer Science to try and solve large scale problems in many fields. So, both of my interests seem to be covered!

I hope that working in a group of 5 people will help me improve working with other people, as well as my communication and organization skills that will be tested throughout the non-coding, documentative part of this project. Moreover, I am looking forward to finally getting to know the infamous language called Python that I have yet had the chance to work with. So, there are many upsides for me in this course and I am cheerful I can attend it.

7.5 Teodora Staneva

I am currently a bachelor student at RWTH Aachen, and I am studying computer science with business administration as a secondary subject. I have some coding experience in Java, but I chose this course so that I could improve my programming skills with Python. I have taken a course in Artificial Intelligence, Implementation of Data Bases and had interests to learn more about Data Science. Process Mining sounds like a great challenge to start learning more about the subject. Since I am new to the topic, I hope I will gain new knowledge and skills that will help me with my further development. I also like to work with other people so I’m glad that the course is structured in a way that we can use our communication and soft skills.

8. Roles in the team

8.1 Software Manager

During this Project, Vishisht will be working as the Software Expert. His ongoing Computer Science degree and previous experience in Python will prove helpful in the implantation of the LTL Checking Tool. His responsibilities will be to master the PM4PY Library and other Process Mining or GUI packages that we will use throughout our project

8.2 Test Manager

Testing is an important part of the project development. As our Test Manager Teodora will be responsible for coordinating, planning and performing the tests. She will notice if they are any problems and solve them in the process and will analyse the test results and present them.

8.3 Communication Manager & Product owner

Communication is the responsibility of Mohamed Amine. He is responsible for arranging the internal meetings of the group members as well as establishing a communication channel between our group and the teaching assistants. He also seeks to maximize our product's value by managing and optimizing the product backlog.

8.4 Scrum Master

As our Scrum Master, Ugo holds the responsibility to precisely organize and monitor the team over the course of this project. Distributing the workload, keeping track of everyone's progress as well as ensuring a harmonic team atmosphere are some of the major tasks Ugo will have to deal with. His communication and organization skills will be tested, but he will surely live up to it.

8.5 Theory Expert, Advisor

Having a theoretical background and a detailed knowledge of LTL filters and conformance checking in general are the basis of the success of this project. Thus, our theory expert Daniel should be able to respond to any questions that other members might have concerning the algorithms used and the underlying scientific papers.

9. Phase Review

9.1. Vishisht Choudhary

The Project Initiation phase went well in my opinion, considering that the whole phase was conducted online. Straight after the first event we were able to establish contact and create groups on Whatsapp and Discord.

9.2. Daniel Henel

I'm very happy after the first phase of our project. We have created a magnificent and powerful team because each of us was really involved in the project. Thanks to this, our work was successful. We managed to complete all the planned tasks. We created our GitLab repository to store all the data related to our project, Trello Dashboard to schedule our task better and Discord channel for communication. We started working in accordance with the principles of the Scrum methodology. We had several meetings, but unfortunately, they were online. For me it would be much better to meet each other face to face. From my side, therefore that I'm a Theory Expert, I was reading a lot about LTL checking and I prepared some coding examples for the team.

9.3. Mohamed Amine Kooli

For me, this is the first project I am working on in a team. From the beginning the communication between all the members of the group seemed to work well, we quickly built Discord and WhatsApp groups where we get to know each other better and made it easy to understand what the next steps are, that we should take in our project. In addition, we had the opportunity to learn each other's skills and based on this, we divided the roles of each member of the team. In the meantime, I've researched and learned more about LTL conformance checking and process mining in general and familiarized myself with the tools we will be using throughout our project. Overall, I'm happy with the good working atmosphere we have created and I'm looking forward to seeing how we can bring our project to fruition.

9.4. Ugo Detaille

Overall, I think the Project Setup has worked out fine. On the one hand, our group was able to meet regularly and discuss important content and organization issues. On the other hand, I think the workload was not distributed conveniently. We will have to distribute the workload better and more clearly in the future. As the scrum master, it will be my task to improve that and to keep track of everyone doing their assigned task

Still, I believe we have developed a good team energy and a friendly atmosphere, which is important for efficient group work. Plus, I feel we have created a really good Project Setup document, design- and contentwise. In fact, the material that Mr. Berti shared with us was very helpful in order to get an idea of what such a document may look like, but we were able to broaden our horizon and develop new ideas, find new content and develop our very own good Setup document. For future work, I hope we will be able to keep the friendly atmosphere alive, but also improve on organization and distribution issues that we experienced in this phase.

9.5. Teodora Staneva

I think the first part of our project went well. We established contact with each other on the first day our team was assigned. Then we created communication channels on two platforms (Discord, Whatsapp), scheduled meetings to discuss important tasks and organise and also discuss more the topic of our project. When we arranged meetings, everyone was on time and the planning went smoothly. Our Trello Dashboard is efficient, and the tasks are presented properly. I'm also glad that the atmosphere is really friendly and we can ask and help each other without any worries. I hope we can keep the same pace and finish the project successfully.

10. References

[1] Barot, P. (2022, March 30). *How is Python transforming enterprise app development in 2022?* BoTree Technologies. https://www.botreetechnologies.com/blog/python-for-enterprise-app-development/

[2] Barot, P. (2021, September 28). *Top 10 real-world python use cases and Applications - BoTree*. BoTree Technologies. https://www.botreetechnologies.com/blog/top-10-python-use-cases-and-applications/

[3] Barot, P. (2022, January 13). *Flask vs Django: Django flask comparison 2022 - botree*. BoTree Technologies. https://www.botreetechnologies.com/blog/flask-vs-django-comparison/

[4] Welcome to flask. Welcome to Flask - Flask Documentation (2.2.x). (n.d.). https://flask.palletsprojects.com/en/2.2.x/

[5] Escobar, A. M. (n.d.). *Analyzing the complaints process at Granada City Council*. IEEE Task Force on Process Mining. https://www.tf-pm.org/resources/casestudy/analyzing-the-complains-prociess-at-granada-city-council

[6] van der Aalst, W. M. P., de Beer, H. T., & van Dongen, B. F. (2005). Process Mining and Verification of Properties: An Approach Based on Temporal Logic. In Lecture Notes in Computer Science (pp. 130–147). Springer Berlin Heidelberg. https://doi.org/10.1007/11575771_11

[7] Berti, A., van Zelst, S. J., & van der Aalst, W. (2019). Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1905.06169