I RAPPORT DE STAGE		II M2

ELEVE

Nom : ZHU Prénom : YANWU Filière/Majeure : DTCS

SUJET : Research Intern

ENTREPRISE

Nom : Efrei Paris- Panthéon-Assas University Paris II

Adresse : 30 Av. de la République

DATE DU STAGE Date de remise du rapport de stage aux membres du Jury :

du : 02/05/2022 au : 21/10/2022 durée effective en semaines : 26

SOUTENANCE Date : Heure :

Composition du Jury :

- Président (responsable Efrei Paris) :
- Tuteur entreprise :
- Tuteur Efrei Paris :
- Invité(e) :

PUBLICATION DU RAPPORT DE STAGE

Le Responsable du stage : autorise le stagiaire à publier lerapport de stage sur l'Intranet de l'Ecole. Signature

Mots clés:

Academic, Research Oriented Internship, Business Process, Service-Substitution.

EFREI - Paris-Panthéon-Assas Université

End-of-study Internship Report

Yanwu ZHU

Group DTCS Promo 2022

20190637

Under the supervision of:

Professor Walid Gaaloul (Tutor from school)
Professor Hanen Ochi (Organizational Supervisor)

October 2022

Table of Contents

List of Figures

Figure 1: Photo of the School Efrei

1 Presentation of the work context

To begin with, My internship was carried out in the Efrei Research Lab, The lab is the laboratory of Efrei, the top-ranking higher education establishment in digital technology. It brings together some thirty teacher-researchers with different and complementary expertise, supervising as well Ph.D. students.

1.1 Introduction of Efrei Paris

The EFREI, formerly the French School of Electronics and Computer Science, is one of 204 French engineering schools accredited to September 1, 2020, to issue of an engineering degree and an established component of the University of Paris-Panthéon-Assas II.

Founded in 1936, the EFREI has been awarded the title of engineer since 1957. A private higher education establishment of general interest (EESPIG), located in Villejuif, it is a member of the Conférence des grandes écoles (CGE) and of the Union of large independent schools (UGEI).

The French School of Electronics and Computers was founded in 1936 under the

name “French School of Radioelectricity” (EFR), with the aim of training senior technicians specializing in the manipulation of radio waves. She moved to, rue Amyot in the 5th arrondissement of Paris.

In 1945, it was recognized by the State and created the “engineer” section in 1947. Ten years later, the Commission des Titres d’Ingénieur empowered the school to issue the title of engineer. The teaching of computer science was introduced in 1969, followed by that of networks.

In 1973, the school changed its name and became the “French School of Electronics and Computer Science” (EFREI).

Since September 2001, the EFREI occupies only a 10,000 m 2 site in Villejuif. In 2003, the school changed its name again and became the “School of Information Technology and Management Engineering”.

In May 2007, the school signed a partnership agreement with the University of Marne-la-Vallée, which allows third-year students who so wish to obtain a degree in computer science. In 2008, a partnership agreement was signed with the Audencia Nantes management school, allowing certain 5th-year students to obtain a double engineer-manager degree. EFREI also offers 6 double degrees with specialized masters from Grenoble Ecole de Management (GEM). In September 2017, EFREI and ESIGETEL merge to become EFREI.

On June 19, 2021, the University of Paris-II Panthéon Assas announces a project for an experimental public establishment (EPE), the statutes of which have been submitted to the Ministry of Higher Education, Research and Innovation, which will bring together the Higher Institute of Interpretation and Translation (ISIT), the EFREI, the Center for Training of Journalists (CFJ) and the W School as component establishments of the university, as well as the Item of the School of war as a partner institute. January 1, 2022.

On December 24, 2021, the decree relating in particular to the integration of EFREI as a component establishment of the University of Paris-Panthéon-Assas is published in the Official Journal for entry into force on January 1, 2022.

When announcing this merger to the press during a press conference at Paris-Panthéon-Assas University on January 12, 2022, EFREI unveiled its new common logo at the university and dropped the word ” Paris ” of its name on its website and social networks, thus continuing its policy of expansion in several cities in France.

Figure 2: Logo of Efrei Lab

1.2 Introduction to Efrei Research lab

The Efrei Research Lab is the laboratory of Efrei, which brings together around 30 teachers/researchers with different and complementary expertise, supervising as many doctoral students.

The Efrei teacher-researchers who work in the Research Laboratory have different, often complementary expertise. Together, they respond to national, European, or international calls for projects on themes allowing them to pool their own experiences.

Also, the various research themes come directly to feed, through courses and projects, the teachings provided in the training. The laboratory brings together around twenty teacher-researchers, and around twenty doctoral students and hosts many interns at the M2 level.

Since January 2022, the Efrei has integrated Pantheon-Assas University, leading to becoming a large establishment, combining the prestigious Pantheon-Assas University, 4 private schools recognized as being of public interest as well as the strategic research Institute of the Military School. This new establishment is made up of 26 laboratories covering the major themes of the contemporary world such as ethics, law and justice, security and defense, management, communication, new economies, and the digitization of companies. The Efrei Research Lab is the digital laboratory of Assas, attached to the Doctoral School EGIC 455, delivering a doctorate in computer science and a doctorate in electronics.

The areas of expertise focus on digital domains, such as Data and Artificial Intel-

ligence; Security and digital trust; Communication networks; Intelligent embedded systems

The lab is also involved with a project called Sciences2024, which is a research program to help athletes to improve their performance through science and innovation, in preparation for the Paris Olympic and Paralympic Games in 2024. Sciences2024 brings together 14 engineering schools, 1 university, and 2 research centers supported by the CNRS, the CNSD, and the INSEP. Efrei is the only private engineering school to participate in this partnership.

Efrei is positioned to cover the entire field of digital technology with its 13 specialization majors, almost all of which can be applied to sports technologies, including robotic systems, imaging, and virtual reality, Big Data and machine learning, and bioinformatics. The Efrei Research Lab has already conducted studies on video tracking (player and object tracking), gesture monitoring (fatigue detection), and individual location and field occupation.

The researchers develop research projects and supervise student projects with an approach to Intelligent Embedded Systems and Data-Driven Decisions for Sports in line with the laboratory's research axis. The chosen orientation will be signal acquisition and processing, video and data analysis, data modeling, processing, and visualization.

Sports technologies are an emerging field of digital technology, and it is also a theme of particular interest to our students. It is a source of very attractive projects for our students and an opportunity to compete with the best student projects in France through the Sciences2024 Challenges.

Efrei Research Lab also participates in Project Health Multi-Omics, with the aim to improve human health, through the development of artificial intelligence methods, focusing on translating multi-omic discoveries into precision diagnostics. The mission is to use the data we get from genomics, radiomics, patronymics, and its impact on pathological conditions using computational multi-omics approaches. Such methods rely on the statistical analysis and integration of big data (high-throughput sequencing, microarrays, proteomics, high-throughput screening), medical imaging, and clinical/phenotypic data. We look at both clinical data as well as data generated outside of hospitals and aim to support both medical providers and patients in their decision-making.

Figure 3: Linkedin page of my supervisor Hanen OCHI

The lab also works closely with companies, the collaboration includes Internships and research projects, Cifre, the Research contract, the scientific Council, Collaboration with asset sharing, and Collaboration within a large project. Also, some productions are involved such as patents, other intellectual properties, innovative products or services, and also outreach brand Images.

1.3 Introduction of the supervisor (Team leader)

Professor Hanen OCHI obtained her doctorate in computer science from the University of Paris 13 in December 2015. She then held a teaching and research associate position at the Galilee Institute (educational institution) and at the LIPN (laboratory computer science from the University of Paris-Nord). After that, She was an RD engineer in Big Data and worked on different innovation projects (autonomous cars, predictive analysis, opinion analysis,...). She is currently working on combining her dual proficiency in model-based process analysis with data-oriented analysis by mixing data science algorithms with formal verification techniques to link data to process analysis. Her research focuses on the development of core approaches for data analysis, data mining, and formal verification, and their application to various fields including process mining, healthcare, service-oriented architecture, etc

Her fields of interest include Formal Verification, Model-based process Analysis, Services Analysis, Data Analysis, Decision aid tools, Artificial Intelligence, etc. Her publication can be found in this page: https://eng.efrei.fr/allianstic-research-laboratory/research-team/hanen-ochi-2/

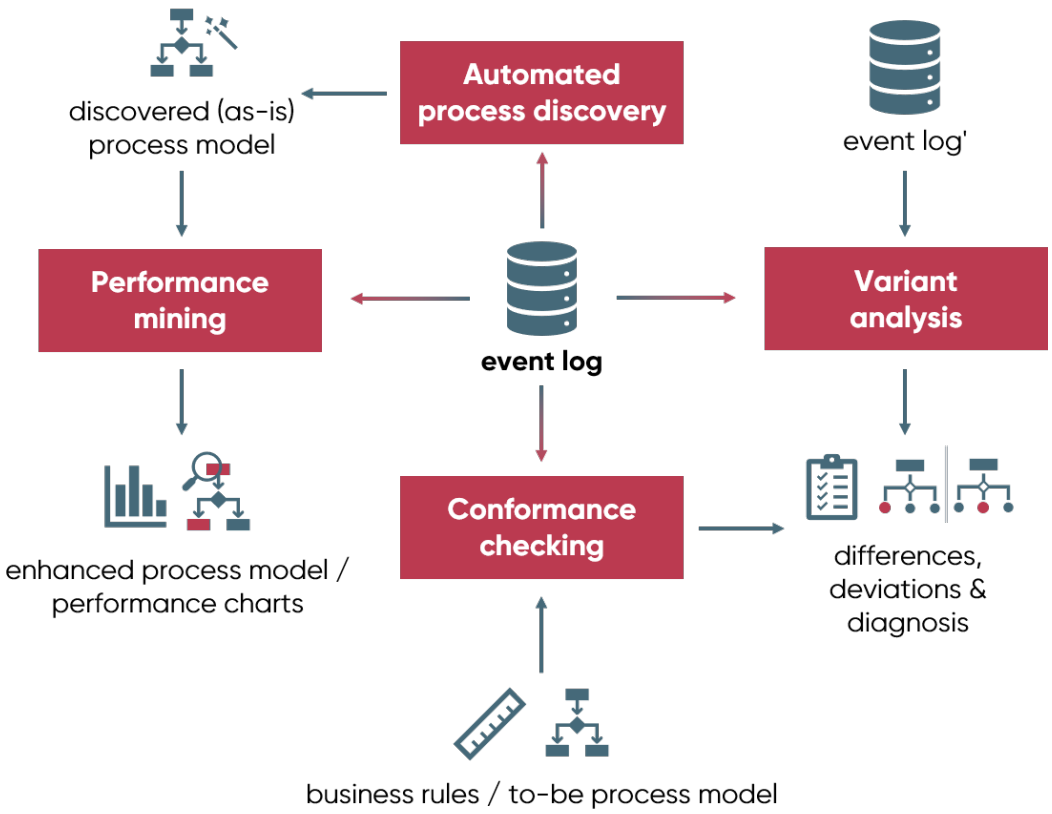

Figure 4: The possibility process mining is capable of doing

1.4 Introduction to the Research Topic

Process mining aims to bridge the gap between Business Process Management (BPM) Business Intelligence (BI) or/and Data Science (DS). Traditionally, BPM focuses on process models and the various aspects, e.g., design, execution, and optimization, of business processes rather than dealing with the generated event data. In contrast, data-oriented analysis (BI/DS) tends to look at particular decisions or patterns and typically is not end-to-end process-centric.

Process mining connects these two disciplines by adapting data-oriented analysis to improve processes. Event data from the event logs of information systems, such as ERP systems (SAP, Oracle, etc.) and BPM systems (Pegasystems, Bizagi, Appian, BPM BPM, etc.), are used in process mining as a starting point to "discover, monitor, and improve real processes".

There are three main areas in process mining. First, process discovery takes an event log and produces a model through different discovery algorithms. Second, conformance checking relates a process model with event data of the same process to

identify commonalities and discrepancies. Third, enhancement enriches an existing

This research internship focuses on conformance checking in the process mining area. There are many reasons for performing conformance checking. For example, conformance checking can support the audit process of organizations so that they can strive towards Continuous. Moreover, organizations might also want to know whether their set processes, which are no longer available or that no longer satisfies the quality requirements, can be substituted by others. By identifying discrepancies between the recorded event data and the process model, conformance checking can aid in process re-design and modification of the BPM lifecycle.

The internship on this topic will focus on the process of mining in the context of services. Some cloud services/processes may be invalid since they are located in a dynamically changing network environment. Service substitution is necessary when a cloud service cannot be used. Typically, there are n services/partners which are involved in one 'global' process since a single component is unable to satisfy all the user/organization's requirements. Each of the partners has its own 'local' process which is private and does not know about the other partner's models (privacy). Amongst the most critical issues that should be handled towards the success of this paradigm is:

- (1)The service composition raises the need for design-time approaches to check the correct interaction between the different components of a composite service.
- (2) However there is a need to use an adequate representation of the different exposed services to be able to hide the internal behavior with respect to its privacy
- (3) and at the same time, a service may be deployed, or un-deployed at any time; its implementation, along with its interface may change without prior notification. Moreover, multiple services may be available with different quality features such as response time, MTBF, MTTR, etc.

The straightforward approach for dealing with a service that is no longer available, or a service that no longer satisfies the quality requirements of the client software that uses it, is to try to substitute it with another one that provides the same result.

The Symbolic Observation Graph is an abstraction of service in order to represent a service behavior. In fact, the SOGs allow us to hide internal activities and expose

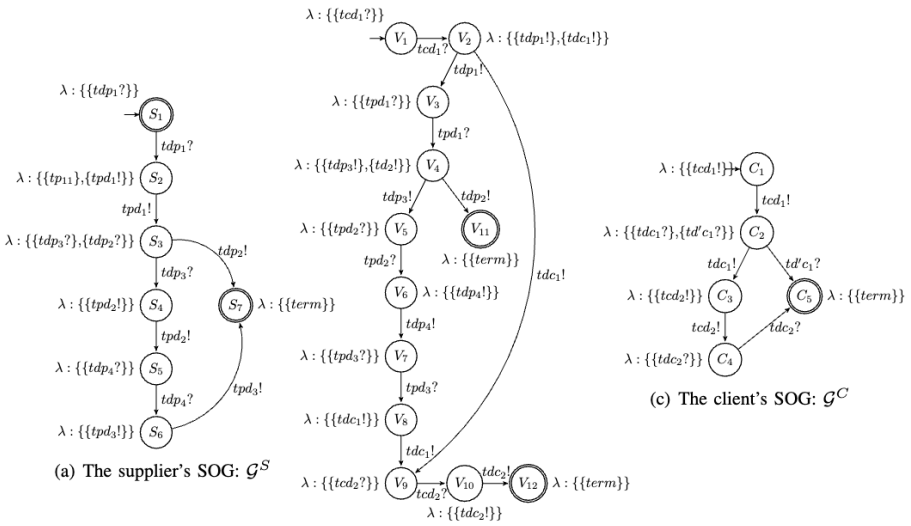

Fig. 4. $\mathcal{G}^S$, $\mathcal{G}^V$ and $\mathcal{G}^C$

Figure 5: An example of SOG as the representation of the service

only the collaborative ones so that we deal with the privacy issue. These observed activities correspond to the transitions that are connected to the interface/resource places.

2 About a Research Internship

2.1 Job Scope for starting a research-oriented internship

The reasons for me to choose a research-oriented internship over an internship in the industrial field are abundant. This decision is based on my future planning, and job scope such as salary and Work-life Balance.

To start with most important, the choice of my research internship was based on the fact that I have decided to pursue a Ph.D. degree after my graduation from EFREI. It is generally considered to have some experience of researching, or research-related work would increase the chance of being admitted. Additionally, since I decided to have a research career as a doctoral student, it's never bad to have a try in advance, to see what research life is like.

Secondly, although the salary or general revenue of an associate professor in a university in China is generally a bit lower than the engineering position in IT companies, the invisible welfare of a professor in Chinese universities is tremendous. Professors are often granted by the universities with an apartment and a considerable kick-off fund to settle them down and to have a good start, not to mention the IT job opportunity in China often have problems such as long working hours and risk of getting optimized because of an increasingly fierce competition happening on IT job market.

Additionally, the research topic of this internship is fairly interesting, and also could be widely applied to several fields, such as business management, IoT, cloud service, and so on. This is really beneficial when it comes to the selection of the job, the job, or the internship has to be of some future help to your personal development rather than just an errand to earn money or to pass time. This internship topic opens enormous possibilities for me considering future job hunting.

In general, although an industrial internship is more 'traditional' and closer to real-life working experience, the research-oriented internship is as good, it also links theory with practice and gains real-world experience outside of the classroom to put on a professional resume. One can gain exposure to your field of interest by working in various organizations to explore career possibilities. Through the assignment of tasks and mentoring of the internship program, there is immense skill development that can help us in future jobs.

2.2 The reason for choosing Efrei Research Lab

The reason I chose the Efrei Research Lab as my internship destination is complicated, there are several parts to it, for one thing, and most importantly, the topic of the internship is intriguing, and can open more opportunities for future development. In addition, the job requirement matches my background. Due to the internship was carried out in Efrei, the curriculum and training I've had in Efrei made me a suitable candidate for the internship, in regard to the topic of business process, data processing, etc.

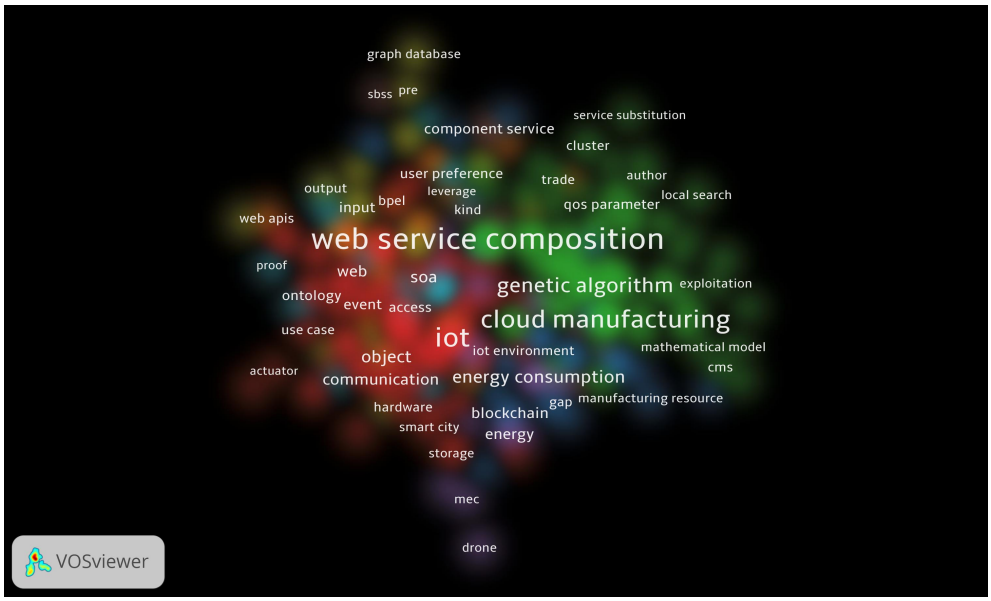

Figure 6: Keyword Analysis - Relevance

3 Technical environment

Systematic Report of the State-of-the-art

This report is composed during my internship as a state-of-the-art report, also a result of my paper review.

It's composed in a systematic way through the analysis of Vosviewer and relevance analysis of papers and publications during the last 10 years on related fields of service substitution and service composition.

Through the relevance analysis6, it is clear that the criteria of QoS, scalability, adaptability, and runtime are mentioned in various related works in the past five years. Related fields are also displayed clearly, such as IoT, Cloud manufacturing, blockchain, and some sub-topic under those great topics, such as drones, mobile edge computing(MEC), smart city, etc.

Through the analysis sorted by time7, it is explicitly demonstrated that applied fields in the industry have become the mainstream of works regarding service composition and substitution. Additionally, some roles are noticeably studied, for example, service provider, user case, user preference, customers, enterprise, and developer. Those newly published studies are based on roles mainly on the novel trends of the research on service substitution and composition.

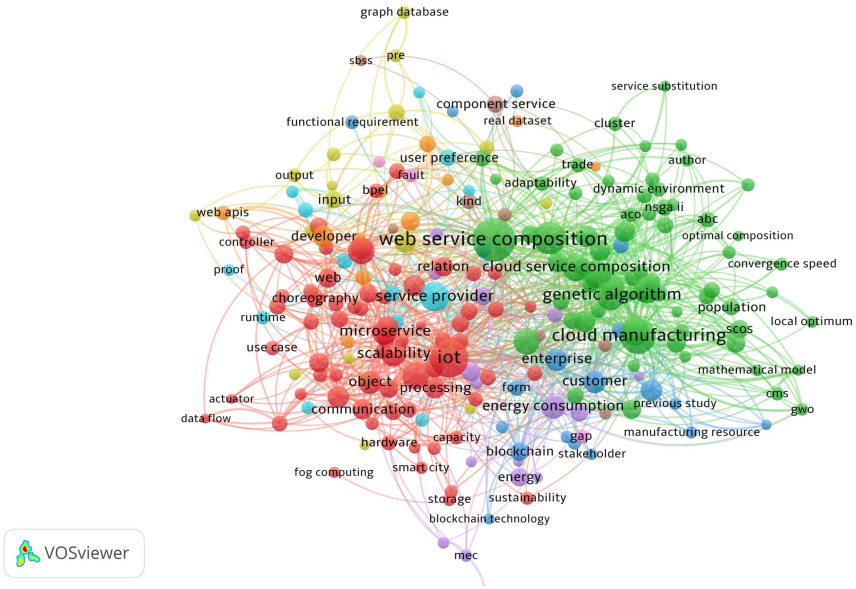

Figure 7: Keyword Analysis - Time

The following section will follow the order of the relevance demonstrated in the result above. The basic concepts and definition will be introduced, and then follows with a survey of different service composition languages(SCL), the analysis of different features and traits valued in SCL, and finally the usage in related fields.

3.1 SoC and SoA

Service-oriented Computing (SOC), as well as Service-Oriented Architecture (SOA), as paradigm concepts for distributed service structure in various fields, have provided theoretical basis research on service composition, management, optimization, and substitution. Such concepts can be dated back to when Papazoglou and Georgakopoulos 2003 brought up the definition of SoC, with four criteria namely Coordination, Monitoring, QoS, and Conformance. When Booth et al. 2004 published such architecture as SOA on the working notes of W3C which provided a common definition of a Web service and define its place within a larger Web services framework to guide the community. The WSA provides a conceptual model and a context for understanding Web services and the relationships between the components of this model. Works regarding those topics have been sprouting in literature ever since over the years.

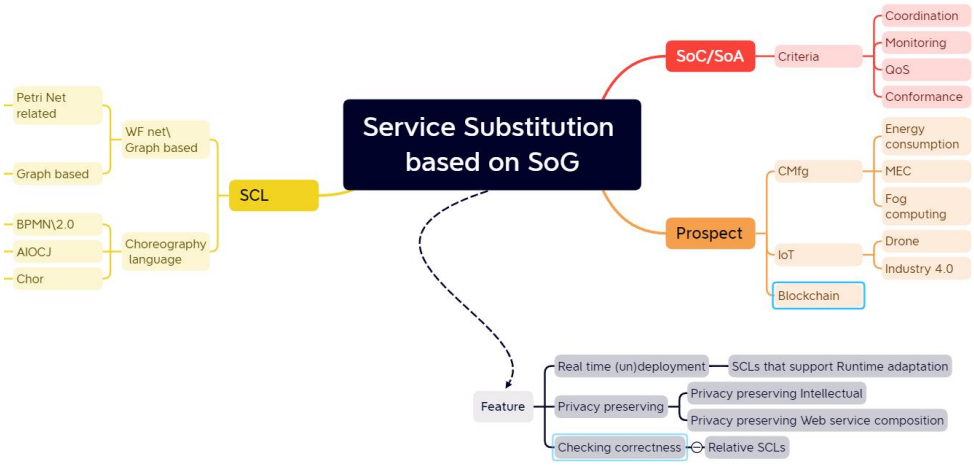

Figure 8: Mind Map of the topic

3.2 Service Composition Language

Among those research outcomes under the greater service-oriented topics, our result is most relevant to the field of Service composition language (SCL), the type of language dedicated to the realization of service composition. SCL is a heated field, with a plethora of outcomes in the literature as well as the industry comes in different formations, see, e.g., the survey by Nikoo et al. 2020, such as choreographic language, graph-based languages, etc.

3.2.1 Choreography language- BPM and its extensive language

With similar goals to our method of optimizing the repetitions in well-structured, predefined standard tasks and processes using activity-centric and control flow-oriented methods, Business process management (BPM) solutions, brought up by Dumas et al. 2013, already have several languages and frameworks based on it. Camuunda, a mature BPM platform Fernandez 2013 based on BPMN 2.0 as one of them, can be used for microservice orchestration and human task management. decision workflow modeling. Just like SOG of our work, it also has a deadlock-free feature and graphical modeling. There are other works in the field of BPM that have a resemblance to ours, AIOCJ by Preda et al. 2014 and Chor by Carbone and Montesi 2013 are two deadlock-free frameworks based on a choreographic approach, focusing on the service composition of distributed applications. They are based on

JolieMontesi et al. 2007, a Java orchestration language interpreter engine. It is also worth mentioning that AIOCJ provides run-time adaption, which is much like the trait of real-time deployment in our work. Admittedly, those were excellent works, however, unlike the SoG, the approaches are in a choreographic language manner rather than graph-based, also the technologies and notions adopted from the above works are slightly obsolete.

3.2.2 Net-based Language

Also, there are several works using Net-based methods, namely YAWL and free-choice net to deal with service and process management. Van Der Aalst and Ter Hofstede 2005 worked on YAWL, a workflow language handling complex data transformations and web service integrations. YAWL is a state-based language and is based on Petri-nets to provide a basis for formal analysis of services. Aalst 2021 also designed another Net-based workflow language, free-choice net, dedicated to BPM and process mining in particular. Those two works are all net-based just like OWF-net in our work, whereas there are differences in the design purposes.

3.3 Features and attributes of the Service composition

3.3.1 Privacy-preserving

Regarding process management with privacy-preserving, Mannhardt et al. 2019 explains potential privacy leakages and means to protect against them and invented a protection model for event data privacy which applies the well-established notion of differential privacy. There are also a number of recent works dealing with privacy preservation in service management, Carminati et al. 2015 worked on a framework of Privacy-preserving composition under the context of web service, while Hübscher et al. 2022 came up with a graph-based way of process mining which can preserve privacy under the context of intellectual property, but rare are works that have the privacy-preserving traits in the field of graph and net based service composition.

3.3.2 Correctness-Checking and Run-time deployment

In the previous discussions of this report, specifically in the section on SCL, the features of correctness-checking and run-time deployment, which can also be called real-time adaptability, are already studied. It has been discovered that nowadays many algorithms and SCLs have already supported the function of correctness-checking, such as those net-based languages and also the two languages, AIOCJ and CHOR based on Jolie. However, the feature of run-time deployment is relatively not so often implemented in related works. In regard to SCLs, only AIOCJ claims to be run-time adaptive. A more detailed investigation can be found in Nikoo et al. 2020 and the former sections of this report.

3.4 Applications of service substitution and composition

Both academia and industry have seen the implementation of service substitution and service composition, not only in business processes or web service but also in other fields,Alinani et al. 2020 surveyed in the application of service composition in cloud manufacturing, whileHamzei and Navimipour 2018 andAoudia et al. 2019 researched on the implement of service composition in the field of IoT, among all those novel applications of service composition technique, several works are similar to ours. Yuan et al. 2020 investigated cloud manufacturing (CMfg) service composition. On the basis of the hierarchical structure of CMfg service integrated management, the position of CMfg service composition and the process of CMfg service composition is comprehensively examined. While the method of Grey relation adopted by Yuan can be also used to analyze the privacy-preserving requirements, this work adopted a model of linear service composition rather than a graphical, and dimensional one with more interfaces like ours. Barakat et al. 2018 designed a novel adaptive execution method, which efficiently handles service changes occurring at execution time, for both repair and optimization purposes. Such execution time adaptation, regardless of the changes happening, resembles the trait of SOG's Real-time deployment.

By analyzing the related work using VOSviewer 9, it is illustrated in the picture of analysis by the cluster that the IoT (the red one) and Cloud manufacturing(the green one) are two major related fields or applied fields of service substitution and composition. As for IoT, it is always related to scalability, communication, hard-

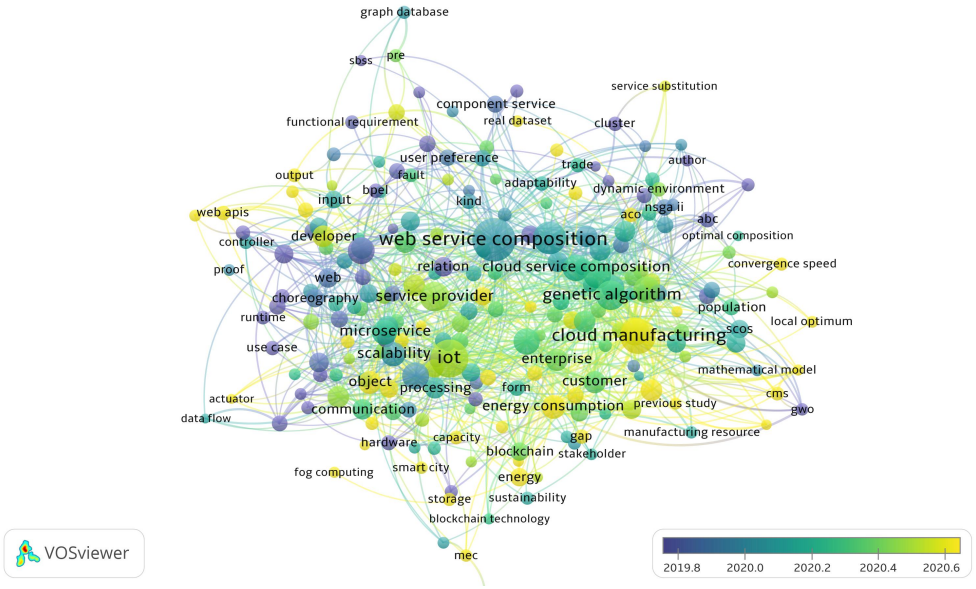

Figure 9: Keyword Analysis - Cluster

ware, controller, and storage. It is understandable because the Service composition algorithm in the IoT environment will have to take the limitation of the device into account. While for the cluster of cloud manufacturing, dynamic environment, adaptability, genetic algorithm, and convergence speed is frequently mentioned and studied.

3.4.1 IoT

The aim of Asghari et al. 2018 is to analytically and statistically categorize and analyze the current research techniques on the service composition in the IoT (published between 2012 and 2017). The service composition in the IoT does not have any systematic and complete study about examining its significant techniques. Hamzei and Navimipour 2018 aims to investigate the available methods in this field using a systematic manner.

At present, the problems of service composition in the IoT environment mostly focus on the evaluation research based on the quality of service (QoS), ignoring the overall energy consumption in the process of dynamic configuration of service composition Gao 2020. Alammari et al. 2020 review several types of compositions and give a view on future smart city service providers.

3.4.2 CMfg

Charro and Schaefer 2018 presents both a first step and a baseline point of reference toward bridging the gap between advanced manufacturing technology and new business development in the context of I4.0 (Smart Manufacturing). Research work done to date regarding service aggregation and composition in CMfg has been investigated in detail from different viewpoints such as selection criteria, solving algorithms, correlation consideration, etc. The purpose of Bouzary et al. 2018is to provide a brief guideline for researchers who are aiming to do similar studies and assist them towards a better understanding of related research work done to date. The key characteristics, concepts, challenges, open issues, and future trends of cloud manufacturing are presented to direct future research Ghomi et al. 2019. In order to adapt the decentralized decision-making management of manufacturing service allocation in practice, a multi-user-oriented manufacturing service allocation framework is proposed by Wang et al. 2019.

4 Tasks during internship

I have had multiple and various tasks and missions during my internship at Efrei Research Lab. These missions can be divided into research-related tasks, software engineering, and communication. The tasks during the internship are not completely sequential but sometimes parallel and inter-depend on each other. A timeline showing my main mission during my internship in a monthly order can be seen in the picture.

4.1 Research-related tasks

Researching ability is actually composed of many trivial abilities, different fields of research also require different types of research abilities. For example, in literature or the philosophical field of research, the ability to empathize or appreciate poems can become a crucial skill regarding research. However, some of the research abilities are shared in different fields and areas of academic research, which are academic reading and writing. Especially for a beginner in his/her research career, the ability to read and appreciate, analyze, simulate, and re-implement other scholars' works become very significant. Because as a beginner, reading papers, essays, reports, or reviews of other scholars' work in this field can give you a quick prospect of what's happening now in this very specific field and also how the fundamental knowledge and theories are applied in the research of this direction.

Being a newbie in the field, the main mission before publishing or releasing any valuable works is to accumulate enough knowledge and ideas about the research topic worked on. That's why these tasks are divided mainly into **familiarization** (concerning independent learning and paper reviewing) and also **report composition**.

During the first two months of the internship, my main job was to get familiar with the research topic since there were still some concepts confusing me, which includes studying some papers and also finding some useful sources of information in order to get a deeper understanding of the topic. I used the platform Coursera to follow the course of process mining, and I set up a repository on my personal GitHub account to track the progress of learning and personal growth during the first half of the internship.

Figure 10: Monthly main missions/progress

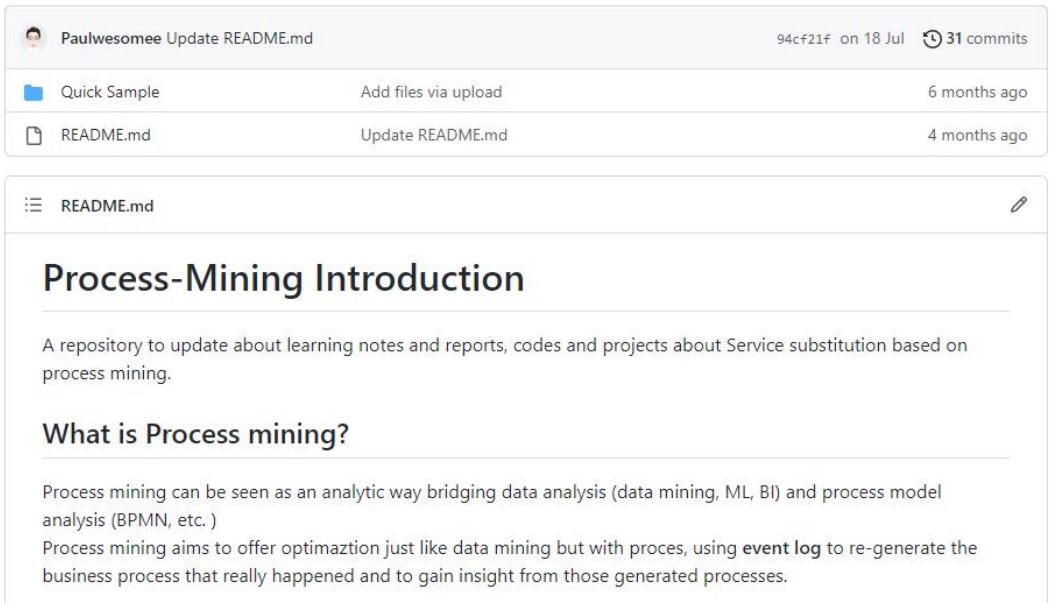

Figure 11: A screenshot from my GitHub page

With the in-depth study of over twenty referencing papers, I presented several reports to my supervisor in which I analyzed the state of the art nowadays regarding service substitution and composition, and also service compositing language. Part of this report can be found in the section on analysis of the technical context.

4.2 Overseas communication

In order to implement the algorithm our research team worked out on the topic of service composition, a related model (an example model) needed to be found to demonstrate the functionality of the algorithm. During the process of reviewing papers, a scholar from Qingdao Technology University, China utilizes a big enough model to illustrate his work, which intrigues the research team. I was requested to get in contact with this professor due to my proficiency in Chinese.

Finally, the professor from China offers some advice about how to fetch data from several information sources, and also provided some data sets already labeled in order to help with the research work in our team. He also showed great interest in having future collaboration on the current research topic.

Communication skills are vital everywhere. Those who have not developed proper communication skills run the risk of falling behind their peers or being emotionally overwhelmed. . Effective communication helps in building and maintaining

胡强教授/同学：

您好，我是朱严吾，是来自巴黎第二大学的科研助理，之前我们课题组在研读一篇您与您的同事一起完成的关于服务替换的（A Cluster and Process Collaboration-Aware Method to Achieve Service Substitution in Cloud Service Processes）论文，发现您在论文中使用了一个体量比较庞大的服务数据集，我们很感兴趣，希望也能使用这个数据集完成一系列实验，所以想请教这个数据集是如何从航天云网获取的，如果邮件不方便回复，可以加我微信：Paulwesome，期待和您的合作。

祝好，

Yanwu ZHU / Research Intern.

M2 Digital Transformation @ Efrei Paris- l'université Paris-II Panthéon-Assas

efrei

PARIS PANTHÉON-ASSAS UNIVERSITÉ

将消息翻译为 英语 | 始终不翻译 中文 (简体)

H huqiang <huqiang200280@163.com>

收件人: Yanwu ZHU

周日 2022-06-19 07:22

朱老师您好：

感谢您的关注，我们的工作做的非常肤浅和简单，让您见笑了。数据是仿真数据，没有实际应用价值。

国内现在有不少类似于航天云网的网站，但很多都已经不再支持数据爬取了。从这些网站上也无法爬取到具体的云制造服务，只能或许一些服务的描述。

已经有几年没再关注航天云网，好像这个网站目前向工业互联网平台转型的幅度比较大，已经看不到上面发布的服务和需求。

我们当时爬取了一些服务的描述（包括一些需求），将其作为一个服务，然后自己仿真生成的一些接口数据，所以数据集没有什么实际价值。

真实一些的服务可以去RapidAPI去爬取。

请海涵我们工作的不足， 欢迎您对我们的工作进行指导和帮助。

胡强

Figure 12: The email between me and Professor HU from Qingdao Technology University

relationships by building trust, strengthening bonds, and reducing conflict. So as can be seen, communication skills are vital in everyday life, especially in academia and another academic or educational contexts. Communication means expectable collaboration and connections. Research fields are highly subdivided nowadays, and more connections and communications might well be of great use to your research work, in regard to the exchange of ideas, brainstorming, or any collaboration in cross-domain topics.

4.3 Research-oriented Engineering

Surely as a student who majored in IT relevant major and had an internship at a CS-related laboratory. Programming is inevitable, I used **C++** programming language as the main language to program the tool for experiment implementation under **Linux**. During the earlier part of the internship, I also have done some experiment samples in **Python** regarding process mining.

The mission regarding software engineering can be divided into three parts, namely Re-constructing the Tool for implementation, Debugging, Adapting the algorithm, beautifying the UI, and Technical Document Editing.

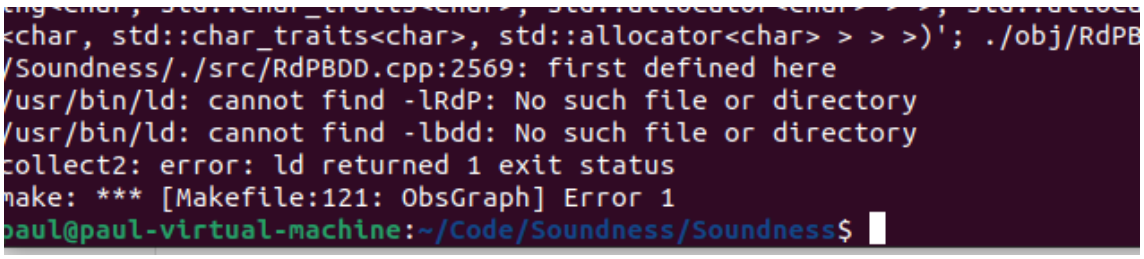

Figure 13: One of the makefile errors I encountered when compiling the program

4.3.1 Debugging Testing

After two months of studying the relevant papers, and exploring related fields, I have adequate knowledge to move on to the next step of the internship, which is the implementation of an algorithm that would be used in a work to be published. The idea was to add some new functionality to an already-exist piece of the program from my supervisor. However, due to some incompatibility issues, the C++ program can not be run on my machine.

After a week of troubleshooting, with many trail and errors on different systems, VMs, and cloud servers, the problem still wasn't settled so I shifted the focus of work to composing a report of state-of-the-art. Then in September, I re-take this mission and this time after continuously trying and studying the C++ environment and dependency, I was finally able to propose a correct environment with a respondent dependency that would let the program run successfully. The process of 'repairing' contains the cleaning of the cache, the research on which dependence is missing, and understanding the inter-depend relationship between libraries used by the program. Though it sounds like a small error-fixing job, there are still several difficulties behind it.

Even though the program can be executed, some of the outputs of the program weren't correct. So it is also my responsibility to adjust the algorithm and re-construct the code blocks where the output seems to go wrong. Several classic testing techniques were used in this step, for example, the breakpoints test.

To conclude, through a long period of setting up dependency, and environment, then testing and debugging, some of the function blocks that will be used in future work were finally adjusted and were able to output desired data.

```
9/12 14:50
./ObsGraph -S ModularModels/Contractor+SubcontractorAlt/Prod/Contractor.net -
OModularModels/Contractor+SubcontractorAlt/Prod/Obs_Contractor
ModularModels/Contractor+SubcontractorAlt/Prod/Fplace_Contractor
ModularModels/Contractor+SubcontractorAlt/Prod/inet_ContSubCont.net -
OModularModels/Contractor+SubcontractorAlt/Prod/Obs_inet_ContSubCont
ModularModels/Contractor+SubcontractorAlt/Prod/Fplace_inet_ContSubCont
ModularModels/Contractor+SubcontractorAlt/Prod/Subcontractor.net -
OModularModels/Contractor+SubcontractorAlt/Prod/Obs_Subcontractor
```

Figure 14: The input to initiate the tool before modification

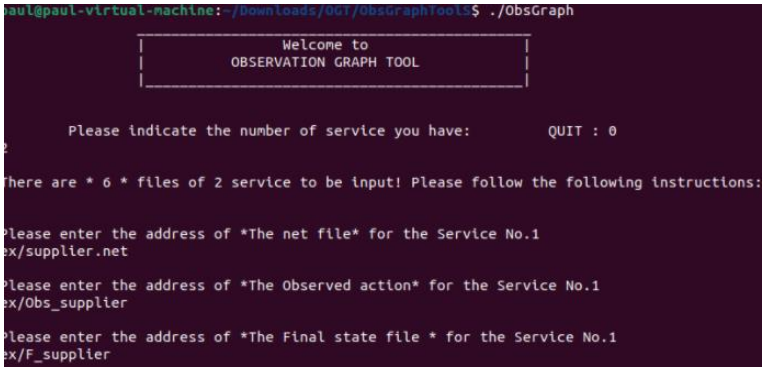

Figure 15: The User interface after modification

4.3.2 Re-constructing and testing of the tool for demonstration

During the last month of the internship, I was requested to adjust the tool in order to demonstrate some of the functionality that will be displayed in a work to be published, so I have to re-construct the main program in order to be more user-friendly and fit the functionality description mentioned in the paper.

Following the course work from my M1 curriculum, the information system modeling. I then transform the experiment tool's main program and its User interface into a more interactive and user-friendly one. The program used to be initialized by a fairly long text message of input files, after modification, it now has the freedom to select the file one by one with step-by-step instructions to be followed.

4.3.3 Technical Document Editing

The last part is also important, which is the composition of a technical document. Often it's manual to a product or a readme in the GitHub repository.

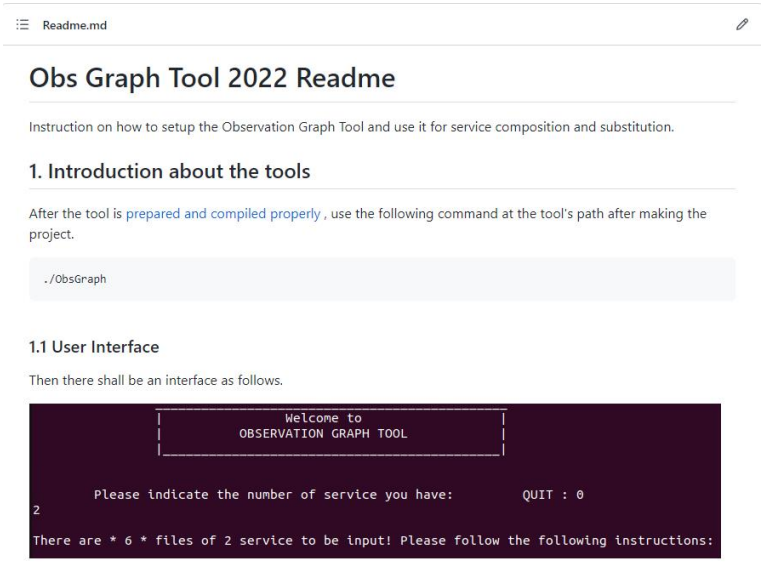

Figure 16: Screenshot of the readme file on the demo page of the tool

Technical documentation is important because it provides essential information about a product or service to a person or organization who needs it. This information allows all parties to achieve their goals. For end-users, the goal might be to enjoy a product while optimizing its functionality and maximizing its life span, which well-written technical documentation allows them to do. Meanwhile, the company may save money on customer service and replacements if end users follow the documentation, use the products correctly and refer to the user guide to troubleshoot their own problems.

Internally, technical documentation can increase productivity and efficiency. For example, in product development, a technical document might explain how to carry out a procedure. If the document is clear and easy to follow, it helps to ensure accurate adherence to the steps with no time wasted. Then, as development progresses, the documentation can help align the goals and understanding of the various teams involved.

After finding the solution to correctly set up the running environment for the tool and fixing the dependency issue, I was requested to compose a technical manual guiding others who will get their hands on the tool to correctly use it. Including the dependency and compilation tools to be installed, the command to be typed in, and the expected errors with corresponding solutions.

Furthermore, after the program was re-constructed for demonstration, a more de-

tailed readme file with the use case and also instructions to run locally are edited as a readme file in the GitHub repository, which can be seen as follows.

4.4 DevOps - Maintaining a collaborated repository on GitHub

In order to work with more flexibility and agility, I take advantage of my major's course work, digital transformation, and cloud service, and move the tool's source code to the cloud, the GitHub repository, in order to collaborate on it without sending emails of Zip file containing the latest version of the code.

After this transformation, my supervisor and I can commit code modifications from different local hosts without overlapping each other's modifications and were able to raise issues, create new branches, and test versions before settling down which version to use.

5 Challenges during internship

5.1 A new field to explore

Although the topic of this research internship is business process management, which I am quite familiar with due to the curriculum from Efrei, and I am no stranger to a research-oriented internship, there are still a great number of new concepts and new technologies that I haven't known. For example, the **Petri Net**, the concept of **Service substitution and symbolic observation graph**, etc.

At the very beginning, I panicked because there is simply not so much time for me to familiarize myself with such an amount of concepts so I was worried about if I would have poor performance in future work during this internship. But later on, things are getting better as I reviewed some papers, and also followed the courses on Coursera in order to better understand the topic and relevant fields.

One way I found fairly useful is to get a few hands-on exercises when exploring a new topic. Overly spending one's time on reading, or watching tutorial videos can sometimes be not so efficient. It is better to get our hands on the technology, it could be a 'Hello world' program or an easy function block, as long as we are using it and taking advantage of it, the familiarity would increase faster than just receiving knowledge without using it.

The other important thing to note is that whenever there's an unsolvable problem or some concepts which can't be understood, always go to the supervisor to give feedback and ask for guidance and a solution. Yes, this would lead to a re-evaluation of your ability, but it's nothing when it comes to the progress of the project or the efficiency of the team. There's no shame in asking for guidance and help, especially when it's a teamwork context, and holding off would only lower the efficiency of the whole team.

Such circumstance often occurs when we enter the job market, no matter how perfect our educational background and courses cover our professional field, there are always new things to learn, understand, and get familiar with.

Figure 17: Screenshot of the weekly meeting on Teams

5.2 Time Management and procrastination

Due to the fact that the office on site was being renovated, I was working from home in the first few months of the internship. Indeed, working from home can be fairly convenient and more flexible. It also causes some troubles when it comes to efficiency and time management.

The communication between my supervisor and me was mainly carried out by Microsoft Teams, the collaborative meeting software. We met on teams about once a week, to report and summarize the progress of my work.

Because the working environment wasn't face-to-face or physically adjacent to each other, the communication and the management of the progress can be really a problem. I was procrastinating for a long time when I was faced with some errors and bugs in the program, or having no papers to review without being motivated to look for new ones on my own.

That's when I learned that the moment when we are faced with some problems that look impossible for us to solve, the best way is to go to your supervisor or leader (if you have any) to clearly address your difficulty. Surely independence matters, but what matters more is the progress of your work, if you stop for a long time or lag the whole team to get along with your independence, then you are totally wrong. It's not shameful to ask for help, in fact, the ability to ask for help is very important and precious in a period of teamwork.

5.3 In times of accidents and errors

As is often the case, the code would have unfixable bugs, the experiment you've spent hours doing would go because of the collapse of the Virtual machine, the reimbursement wasn't permitted and the team couldn't get the money back. There were lots of errors and problems that happened every single week during my internship.

During the first few weeks, I would panic and don't know what to do. As I get to know the team, know the cooperating parties from other schools, know how the whole project is going and know who has a better coding knowledge base in the office. The daily problem you met will be solved easily as your experience was gained through your familiarity with your work.

It's inevitable that there will be all kinds of accidents and errors in our work, compared to panicking all the time, it's always better to stay calm and seek the best solution.

6 Conclusion

During this long journey of working as a research assistant in Purple Mountain Laboratory affiliated with Southeast University, I have learned many things not only with research skills and abilities but also with soft skills such as project management and interpersonal communication skills, to make a team more solid and united.

I realized that the research field is no different from other job occasion whether it's IT companies or the IT departments of some other companies, where techniques that we mastered certainly have some weight regarding the assessment and efficiency of an employee, but the ability to fit in a team or more importantly to lead a team is definitely more important considering the long-term development of a professional career.

I am glad that in this period of one and a half years, I was trained both technically and non-technically, from which I benefited a lot.

6.1 Skills that I harvested

From these almost one and a half years long internships, I certainly acquired some skills, some skills like programming, testing, and operation under Linux, editing technical documents, and also some academic reading and writing are the hard ones. Unlike interpersonal communication, team management, leadership, etc., those are soft skills.

6.1.1 Soft skills

One soft skill I acquired is **multi-tasking**, the ability to handle multiple tasks at a time. It's true that we more or less experienced such circumstances at school, but it's another story at work, where the pressure of failing your tasks can be very serious. The key to multitasking is to prioritize each task according to its importance and its due date, the list of priorities has to be changed every week or every few days, according to the progress of each task. I remembered when I was buried with tons of work such as organization of the incoming conference, handling reports from different universities, making an academic report, and still having to make appointments and go physically to complete the reimbursement procedure. My ability to multi-tasking

was well-trained during my internship period.

Some other soft skills such as **teamwork and project management** are too complicated to be finished in just one paragraph, so they are listed in the last few subsections.

6.1.2 Hard skills

Certainly, during the internship, I also acquired some hard skills, to my knowledge, hard skills are those you can actually apply to your work and get results without the interference of personal relationships.

academic reading

One major hard skill that I learned during my academic internship is **academic reading**, after in-depth reading of over fifty papers in the relevant field of the lab and the project, I finally have some grasp of what to do when I'm reading an academic paper. I also was taught an efficient way to do it, using such things as called Paper-reading Form as a record of the target papers, which can keep track of the papers you have done researching and keep organized and recorded so that your thought would also be clear and organized, which will, in turn, have a better understanding of your research target.

Programming in C++ under Linux

Surely as a student who majored in IT relevant major and had an internship at a CS-related laboratory, the hard skill that I certainly got strengthened is **programming**, I used C (programming language) as the main language to program the demo sample for the laboratory members when I was in UNIS lab, under the operating system based on Linux. This experience largely exercised my ability to code in various languages (C, Java, Python, etc.) and to smoothly operate Linux OS, for example, different commands in the command line.

Editing technical documents

Technical documentation is significant and can't be ignored, as is often the case. From my perspective, technical documents, such as manuals and readmes, are just as important as the product to be delivered.

Externally, it gives vital information about a product or service to those who require it. This information enables all parties to fulfill their objectives. The objective for end users may be to enjoy a product while enhancing its usefulness and increasing its life cycle, which well-written technical documentation enables them to achieve. Meanwhile, if end customers read the instructions, use the items appropriately, and refer to the user guide to fix their own problems, the firm may save money on customer care and replacements.

Internally, technical documentation has the potential to boost productivity and efficiency. In product development, for example, a technical paper may explain how to carry out a technique. It helps to guarantee proper adherence to the instructions with no time spent if the paper is clear and easy to follow. The documentation can then assist align the aims and knowledge of the various teams engaged as development continues.

6.2 Thoughts on research work

Having an academic internship, or a research-oriented internship is somehow different from another kind of internship, for that if there's a large proportion of actual academic investigation and paper reviewing, chances are that you will work alone, not to say that you have no teammate on your project, but to say that you are on a new direction where no one else has to lay their foot on before because that's what researching is all about innovation and novelty.

This is not like, for example, developers in IT companies, where, though innovation is also needed, the major part of the work is to operate in the set track using knowledge and information that are already there. There will be short in a spirit of novelty and creation.

I choose this internship out of many reasons, but mainly because I plan to pursue my Ph.D. degree in a similar direction, so I find that an academic internship would be more suitable for me. I would like to say that when it comes to the decision-making of job offerings or other choices you have to make, it's always important to make decisions according to your final goals and future planning. Though sometimes things are not going to develop as you imagine, what we can at least do is play the best out of what we have.

6.3 Reflection: Programming in Academia VS Industry

After having programmed both in the academic setting and in the industry. In college, I've been on research projects, directed readings with professors, and developed apps and experiments. I've also done programming in the industry and actually worked on products that would have end-to-end users.

A lot of people seem to think that programming in academia isn't real-world programming. Indeed programming in academia solving theoretical problems in C/C++, VHDL, or Matlab isn't going to make you a rockstar in implementing RESTful web services in Java or .NET, I do think the core programming skills and principles carry over nonetheless. However, there are still some similarities as well as differences between those two fields' programming works.

6.3.1 Differences

One of the major differences in my experience between the two domains has been guidance. If I'm stuck on something in an academic context, even if this is as trivial as figuring out the best practices, I can bounce ideas off my professors, and they understand that I'm learning by doing and are usually accommodating. An elegant solution is preferred over a hacked-together solution, and you are encouraged to learn best practices doing it. Not only because it's the spirit of looking for the best solution, but also for the time-cost and complexity in research work is crucial, when it comes to comparison to other algorithms and other methods, those parameters are exactly what you need to outperform.

While in the industry when some product is better than no product rule takes precedence over anything else. If you cannot figure out the most elegant way to do something, hack something together to have a working solution. If time and budget permit, you may have a chance to later go back and fix it, although sometimes you may not. You may have a mentor assigned to you, but he's busy doing his stuff and will usually try to point you to something, and if it doesn't help, and you are on your own to figure it out. If you ask TOO MUCH, you're probably a bad hire. The focus is to **make more out of less cost**, the time-complexity or the elegancy of the code were rarely cared, especially when it's an unimportant function block with a limited time to go online.

6.3.2 Similarity

However both coding in academia and industry, in essence, are coding and software engineering. The difference and changes we made in each field are only the change we made according to the context and the request from our 'customers and boss'. Coding rules, documentation, and technical documents, such as readme are necessities in both fields, after the product, whether it's a piece of software or an experimentation tool, is to be delivered and used by humans. Therefore, user-oriented design and modeling, though sometimes will be less-valued, can't be ignored.

6.4 About Project Management

During the internship in The Project, I gained over one 1-year experience of in managing a project with teams from different facilities even different countries and different cultural backgrounds. It was certainly not easy to manage such a diverse project and the experience I gained was precious and important to my future path. I would like to mainly divide project management into three major parts.

The first part is about the division of labor. If it is a task that can be completed by one person, the project manager will generally assign someone to be responsible.

There is no division of labor for individuals to work independently. In the two-person collaboration, they can effectively distribute the workload and work content through equal consultation and communication.

In a large project team, due to a large number of members, it is obviously difficult to reach an effective and satisfactory plan through equal consultation and communication with each other on the distribution of workload and work content. Even if the project manager can arrange and coordinate, it itself requires the project manager to know how to work as a team.

The second part would be cooperation. If there is a division of labor, cooperation is required, that is, mutual cooperation. In peer collaboration, due to the simple composition of personnel, the difficulty of cooperation, coordination, and communication with each other is far lower than that of team collaboration. In a large project group, due to the differences in their membership backgrounds, complex interpersonal relationships, and unfamiliarity with each other's work, it is difficult to cooperate with each other.

The third part, which is also the part that many forget or neglect, the supervision. Without effective oversight, all project team members cut corners and the project failed completely.

When individual works independently, all work must be borne by oneself, so there is no possibility for others to share. In teamwork, they can supervise each other simply and effectively, so there is less possibility of problems in this regard.

Therefore, in a large project group, it is very important to establish good teamwork. Therefore, in the process of managing a project, teamwork is obviously an important link that cannot be ignored. When dealing with team collaboration problems, establishing a reasonable and complete team mechanism is the first problem to be solved.

Simply put, it is the regulation of the division of labor, cooperation, and supervision of the members of the team. While it is specifically reflected in the establishment of a series of rules and regulations, it focuses on the implementation of the establishment of each member of the team within the team.

6.5 The importance of teamwork

Work not only creates benefits for the greater good (such as for the company, facility, or team) but also improves oneself. Just like our classmates who have just entered the internship period, we still need to do more to accumulate experience. Especially now that internships don't have the same clear scope of work as regular employees. If the work attitude is not positive enough, there may be nothing to do, or they don't have a clue about how to do it.

Therefore, actively strive for more work to accumulate and improve. At the same time, improve teamwork: work is often not the work of one person, it is a team that completes a project, and how maintaining communication with other colleagues in the team is also important.

A reasonable division of labor can enable everyone to do their best work, unite and cooperate, cooperate with each other, and achieve common success. In order for individuals to be successful and achieve good grades, one rule must be kept in mind: We can never put our personal interests ahead of our team's. In teamwork, there are situations where we simultaneously benefit from our own assistance. This is the

most important factor to ensure your own success. Teamwork is a spirit of resource sharing and collaboration to achieve set goals. It can mobilize all the resources and talents of the team members, and it will automatically get rid of all discord and injustice. At the same time, it also has excellent performance. Rewards are given in a timely manner, making teamwork powerful and lasting.

For enterprises and other facilities, the importance of teamwork is mainly reflected in the following three aspects: Teamwork is conducive to improving the overall efficiency of the enterprise. By promoting team spirit and strengthening team building, domestic consumption can be further reduced. If you always spend time on how to define responsibilities, who should be dealt with, and let customers and employees move around, it will weaken the affinity of company members and destroy the cohesion of the company.

Teamwork helps achieve business goals. The realization of corporate goals requires the efforts of every employee. A team with teamwork spirit respects the individuality of the members, values the different ideas of the members, stimulates the potential employees of the members, and truly allows every member to participate in the teamwork and share the risks. Share benefits, cooperate with each other, and accomplish teamwork goals.

Teamwork is a huge driving force for enterprise innovation. Humans are the only active resource of all kinds. The development of an enterprise must reasonably allocate people, money, and materials, and mobilizing people's enthusiasm and creativity is the core of resource allocation. Teamwork is to rationally mobilize people's wisdom, strength, experience, and other resources to generate the greatest scale benefits.

7 Appendix

7.1 Links and of some of my works

https://github.com/Paulwesomee/Process-Mining
https://github.com/ObsGraphTool/Version2022

Bibliography

Aalst, Wil MP van der (2021). 'Using free-choice nets for process mining and business process management'. In: *2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, pp. 9–15.

Alammari, Abdulsalam, Salman Abdul Moiz and Atul Negi (2020). 'TOWARDS TRULY SMART CITY SERVICE PROVIDERS: A VIEW ON ON-DEMAND EVERYTHING AS A SERVICES'. In: *Congress on Intelligent Systems*. Vol. 1.

Alinani, Karim et al. (2020). 'Service composition and optimal selection in cloud manufacturing: State-of-the-art and research challenges'. In: *IEEE Access* 8, pp. 223988–224005.

Aoudia, Idir et al. (2019). 'Service composition approaches for internet of things: a review'. In: *International Journal of Communication Networks and Distributed Systems* 23.2, pp. 194–230.

Asghari, Parvaneh, Amir Masoud Rahmani and Hamid Haj Seyyed Javadi (2018). 'Service composition approaches in IoT: A systematic review'. In: *Journal of Network and Computer Applications* 120, pp. 61–77.

Barakat, Lina, Simon Miles and Michael Luck (2018). 'Adaptive composition in dynamic service environments'. In: *Future Generation Computer Systems* 80, pp. 215–228.

Booth, David et al. (2004). *Web Service Architecture http://www. w3. org*. Tech. rep. TR/ws-arch.

Bouzary, Hamed, F Frank Chen and Krishnan Krishnaiyer (2018). 'Service matching and selection in cloud manufacturing: a state-of-the-art review'. In: *Procedia Manufacturing* 26, pp. 1128–1136.

Carbone, Marco and Fabrizio Montesi (2013). 'Deadlock-freedom-by-design: multiparty asynchronous global programming'. In: *ACM SIGPLAN Notices* 48.1, pp. 263–274.

Carminati, Barbara, Elena Ferrari and Ngoc Hong Tran (2015). 'A privacy-preserving framework for constrained choreographed service composition'. In: *2015 IEEE International Conference on Web Services*. IEEE, pp. 297–304.

Charro, Alberto and Dirk Schaefer (2018). 'Cloud Manufacturing as a new type of Product-Service System'. In: *International Journal of Computer Integrated Manufacturing* 31.10, pp. 1018–1033.

Dumas, Marlon et al. (2013). *Fundamentals of business process management*. Vol. 1. Springer.

Fernandez, Allan (2013). 'Camunda BPM platform loan assessment process lab'. In: *Brisbane, Australia: Queensland University of Technology*.

Gao, Jianhao (2020). 'Green Energy Strategic Management for Service of Quality Composition in the Internet of Things Environment'. In: *Complexity* 2020.

Ghomi, Einollah Jafarnejad, Amir Masoud Rahmani and Nooruldeen Nasih Qader (2019). 'Cloud manufacturing: challenges, recent advances, open research issues, and future trends'. In: *The International Journal of Advanced Manufacturing Technology* 102.9, pp. 3613–3639.

Hamzei, Marzieh and Nima Jafari Navimipour (2018). 'Toward efficient service composition techniques in the internet of things'. In: *IEEE Internet of Things Journal* 5.5, pp. 3774–3787.

Hübscher, Gerd et al. (2022). 'Graph-based managing and mining of processes and data in the domain of intellectual property'. In: *Information Systems* 106, p. 101844.

Mannhardt, Felix et al. (2019). 'Privacy-preserving process mining'. In: *Business & Information Systems Engineering* 61.5, pp. 595–614.

Montesi, Fabrizio et al. (2007). 'Jolie: a java orchestration language interpreter engine'. In: *Electronic Notes in Theoretical Computer Science* 181, pp. 19–33.

Nikoo, Mahdi Saeedi, Önder Babur and Mark Van Den Brand (2020). 'A survey on service composition languages'. In: *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings*, pp. 1–5.

Papazoglou, Michael P and Dimitrios Georgakopoulos (2003). 'Introduction: Service-oriented computing'. In: *Communications of the ACM* 46.10, pp. 24–28.

Preda, Mila Dalla et al. (2014). 'AIOCJ: A choreographic framework for safe adaptive distributed applications'. In: *International Conference on Software Language Engineering*. Springer, pp. 161–170.

Van Der Aalst, Wil MP and Arthur HM Ter Hofstede (2005). 'YAWL: yet another workflow language'. In: *Information systems* 30.4, pp. 245–275.

Wang, Tianri et al. (2019). 'An evolutionary game approach for manufacturing service allocation management in cloud manufacturing'. In: *Computers & Industrial Engineering* 133, pp. 231–240.

Yuan, Minghai et al. (2020). ‘Service composition model and method in cloud manufacturing’. In: *Robotics and Computer-Integrated Manufacturing* 61, p. 101840.