

Final Internship Report

The information in this report is confidential.

Submitted by

<<Gao Yunyi>>
<<1002871>>
<<ISTD>>
<<2017>>
<<20 Aug 2018>>

Submitted to

<<ExactCure>>
<<Fabien Astic >>
<<President >>



Final Internship Report

1. Acknowledgement

The internship opportunity I had with ExactCure was a great chance for learning and professional development. I would like to express my deepest gratitude and special thanks to Sylvain Benito who despite being extraordinarily busy with his duties, took time out to arrange the administrative matters for my internship, hear, guide and keep me on the correct path and allow me to carry out my project.

I express my deepest thanks to Mattieu Basset, my direct supervisor, for designing my projects and giving necessary advices and guidance. I want to thank Lola Rousseau, my colleague and Fabien Astic, the president, and all my other colleagues for showing hospitality and care for me throughout my stay in Nice. I am grateful to SUTD career development center teachers, such as Christina Chee and Eleanor Loh for guidance and making this overseas internship possible.

I perceive as this opportunity as a cheerful and fulfilling one. Hope that I could continue cooperation with ExactCure in the future.

2. Executive Summary

This internship report records the work done at ExactCure between the period of 18 June 2018 to 31 August 2018 in Nice, France. I was mainly focused on automation of literature mining and constructing database for future application development. I was first given opportunities to gain familiarity with the literature mining of pharmacokinetic papers for construction of mathematical modelling of drug. Then I consolidated the basics of web scrapping and data mining through independent hands-on learning. In first two weeks, I did projects to extract drug resume from solidarites-sante.gouv.fr and automated search of papers based on drug substance from PubMed and returned them in excel format. Both were using HTML parsing. For the following two weeks, I enhanced resume extract with more efficient approach and returned information gathered in word document individually for each drug, to create a name card for the drug. Under my supervisor's guidance, I modified the literature mining based on XML parsing to extract papers and added the functions to prevent repetitive papers to exit in the database. These projects are designed to facilitate the literature mining for ExactCure by automating the process. I spent another two weeks to understand relational database management system, getting myself proficient with SQL language and SSMS, an intermediate software to process SQL data. During 7th week and 8th week, I wrote code to populate and make request from SQL database using python and designed the database organization for better user experience. For the final weeks, I populated user samples for trial tests for the application demonstration and enhanced automatic literature reading from PDF file. In the end, under my colleagues' guidance, I learnt to use Knime, an analytic platform, for better arrangement of the software product development projects.

3. Table of Content

Commented [YG1]: You are expected to write the final report as though the reader has not read your interim report. You may reference some of the contents from the interim report and provide a summary of the contents where relevant in this report.

The intended audiences for the report is for immediate supervisor, faculty mentor and career centre staff. The length of this report should not exceed 2,500 words.

Commented [YG2R1]:

Commented [YG3]: should acknowledge the support and help of key personnel who helped in the completion of your internship and the preparation of the reports. These could include your supervisor, department colleagues, pillar mentors and others.

Commented [YG4R3]:

Commented [YG5]: executive summary provides a preview of the main objective of the final internship report. It should help the reader to get a quick overview of the report before reading the details. In not more than half a page, you should briefly mention the key points of what you have done, learnt, or experienced.

Commented [YG6R5]:

Commented [YG7]: the important section headings and sub-headings in the report with the respective page numbers in a tabular format.

Commented [YG8R7]:

Content	Page Number
1.Acknowledgement	1
2.Executive summary	1
3.Table of content	2
4.Introduction	2
5.Work experience	3,4,5
5.1 Pharmacokinetics Basics Learning	
5.2 Literature Review Practice	
5.3 Literature Mining Programming	
5.4 Database Construction Automation	
5.5 SQL language Learning	
5.6 Database Structure Conceptualization	
5.7 Project Analysis and Management	
6.Challenges and action plans	6
7.Learning Experiences	6
a. Knowledge acquired	
b. Skills acquired	
c. Observed attitudes	
d. Achievements	
8.Conclusion	6,7
9.Reference	7
10.Appendix	7

4. Introduction

ExactCure is a start-up of three co-founders who motivated to fight against iatrogenic medication through their application. Their future Twin Digital will simulate the effects of drugs in the patient's body based on their personal characteristics to help them avoid underdoses, overdoses and drug interactions. ExactCure takes business to business model. In the application development stage, Mathematic modelling of numerous types of drugs is required based on literature mining and analysis of historical data record of patients. After the models are constructed and stored in the database, the application should be able to return the users the personalized model for the drugs they take. Literature mining, hence, requires massive amount of work. Automation of the literature mining is thus very important for populating models that are needed for the drug model library. With my work, my colleagues are able to work on literature mining with more efficiency.

Commented [YG9]: an overview of the department that you interned in. This may include the department organisation chart, the types of positions available, and the upstream and downstream stakeholders, if any. You are also required to demonstrate understanding of how the department contributes to the organisation's mission and vision.

Commented [YG10R9]:

5. Work Experiences

5.1 Pharmacokinetics Basics Learning

Before arriving in France, I was tutored by my supervisors on fundamental knowledge of pharmacokinetics and pharmacodynamics, which mean effect of body on the drug and effect of drug on the body respectively. Two books recommended by my supervisor which are entitled to Basic Pharmacokinetics and Essentials of Pharmacokinetics and Pharmacodynamics (Second Edition) were read to facilitate my understanding of the topic. Weekly video meetings with my supervisors were conducted to enhance my understanding and solve my confusion. I did presentations with slides for each of the meeting and problems in my presentations were explained.

Commented [YG11]: 2 frameworks you could adopt under this section;

i. Summarise the work and assignment done during the first six weeks from the interim report, and add on any new work and assignment,

or

ii. Summarise the entire 16-week experience in the organisation with appropriate categories.

Commented [YG12R11]:

Through the process, I have a better idea of the interaction between body and drug and ADME (Absorption to the gastrointestinal walls, Distribution between organs through the blood compartment, Metabolism by the liver mainly, Elimination or Excretion by the kidneys) of drugs. This helps me know how the mathematical model of the drug taking is constructed and what are the main variables and equations for the modelling.

5.2 Literature Review Practice

To complete this internship, this first learning period was necessary and it gives me tools and theoretical knowledge in hand to address pharmacokinetic literature articles. The aim of a pharmacokinetic study will be to model and predict the temporal evolution of the drug in the body according to the ADME scheme. Upon arriving, I practiced some literature review under the guidance of my colleagues to know what the sources for finding articles are, what elements to reach out for in the articles and the difficulty of the mining process.

5.3 Literature Mining Programming

During the first two weeks of my attachment, I consolidated the basics of web scrapping and data mining through independent hands-on learning in several projects. In first project, I used python to write a program to make http request and analyze the html text downloaded using BeautifulSoup and use Pandas and Openpyxl to return excels of the information useful for further analysis. In this project, pharmacokinetic information from RCP (RÉSUMÉ DES CARACTÉRISTIQUES DU PRODUIT) of different medicine were supposed to be collected. I was recommended to use Ansm as the source, but after several test run, the limitation of the database made me switched to solidarites-sante.gouv.fr.

In the second project, I programmed to automate search of scholastic articles from PubMed based on drug names and keywords, and returned relevant information, like authors, titles, abstracts, ids and so on into excels named by drug names with multiple sheets named by keywords for search. I manipulated webdriver of Chrome through this project, which is a very good tool for form submission in web scrapping process. Both programs used HTML parsing. This gave me an approachable start for my future projects and it gave a lot of very good learning points about web scrapping.

5.4 Database Construction Automation

For the following two weeks, I enhanced resume extract with more efficient approach in the third project. As the program from the previous weeks were based on trial-and-error and manipulation of strings of ids to get the right URLs for request. This was time-consuming and thus I wrote a program to make use of the index search to require all proper URLs for future mining.

And in the fourth project, html content downloaded was processed and categorized in excel and then to the word format based on substance of the drug. Creation of the name card of substance in

word document format was automated. The result was hence more readable and easier for comparison and selection for my colleagues.

In the fifth project, the program for PubMed article search was changed from html parsing to xml parsing. Bio, the library recommended by my supervisor, which was perfectly designed for web scrapping using xml brought great help. The information gathered was processed in such a way that my colleague only need to write the product names and keywords they want to search in text format, then the program will return numerous excel file named under the drug name with articles wanted. No repetition of the drug name or ids of the article visited will happen due to the addition of function of checking the repetition. To better automate the literature mining process and construct internal database for company future research, these projects, suggested by my supervisors, gave me good view of the directions.

5.4 SQL Language Learning

I spent the next two weeks familiarizing myself with SQL language and the intermediate tools to manage database stored in SQL server. SQL (Structured Query Language) is a programming language designed for managing data in a relational database. It is very easy to read, manipulate, and change data. Guided by my supervisor, I had better understanding of its application in massive interactive database management such as Clouds service, social media like Facebook and Twitter. I also learnt how to use SSMS (Microsoft SQL Server Management Studio) to manage data via graphical interface. I searched and studied on python library such as Pyodbc to make connections with server and I wrote two simple python program to populate the database and send request to the database.

5.5 Database Structure Conceptualization

To manage the data better with consideration of the future use of the company product, a good design of the database is very important. My supervisor gave me an interesting tutorial about the primary key and foreign key concept in database. We worked together and came out with a simple structured database and did a demonstration of future possibilities on SSMS.

5.6 Project Analysis and Management

In the end, Knime analytics platform is introduced. Knime is a graphical user interface which allows assembly of nodes blending different data sources, including preprocessing for modeling, data analysis and visualization without, or with only minimal, programming. I learnt the existing node repository to create several workflows for practice. I learnt how to make extension to incorporate python to create my own nodes and use them to construct a workflow that mimic the future use of application and arrange my previous work with more clarity. My colleague who has previous experience with Knime gave me guidance in learning and insights to make the modification of the workflow.

6 Challenges & Action Plans

Commented [YG13]: updates on the challenges and actions plans previously listed in the interim report.

- Have the action plans been diligently followed through?
 - Were the challenges resolved or any improvement made?
- Describe any new challenges you faced at work from week 7 onwards, and describe in detail the action plans undertaken to overcome these challenges.

Commented [YG14R13]:

Two main challenges were met during. The first one is the language barrier when I did the web scrapping. To make better manipulation of website skeleton and content, I need to understand the page and its element first. However, since I was working with French website half of the time, it took me longer time to get familiarized with the websites. The obstacle was mitigated with help of my local colleague who willingly explained the content to me and my supervisor who introduced me to the web in a more structured computer language.

The second challenge was lack of enough knowledge to fulfill the requirement for the task. Despite the distress, I actively immersed myself in independent learning in online open-source courses for skills and did not hesitate to ask my supervisors for advice. My supervisors are all patient and friendly, giving me time and encouragement for practice and improvement. I have learnt a lot from both to tackle the challenge.

The problem that I found about myself is that I lack the knowledge to think what else I could do to help with the company when the tasks given are already finished. I try to tackle this problem with deepening the learning with existing projects and reporting more frequently to my supervisor for feedback and progression.

7 Learning Experiences

a. Knowledge acquired:

I have learnt how to design and manage data using new software such as SQL server management studio, Knime analytics platform and got to understand and practice new library for python programming.

b. Skills learned:

I have learnt to clarify what does my supervisor really wants, listen to their requirements and know the potential constraints or extension. I now know how to manage time for the working hours and try to maintain a good relationship with my colleague despite cultural differences, learn from them and talk to them.

c. Observed attitudes and gained values:

I understand the importance of corporation with your colleagues to help each other and compensate for each other's shortcoming with your own strength to improve efficiency. I learnt to be more prompt to the supervisor's arrangement and be ready for each presentation no matter it is casual or very official ones.

d. Achievements

I have learnt a lot of technical skills and sharpened my programming skills while exposing to a lot of interesting and useful applications. I felt that I could manage data relevant tasks with more confidence and proficiency. In addition, witnessing the organization and the environment of startup in France is also very valuable.

8 Conclusion

Commented [YG15]: on your entire internship journey and describe the knowledge gained through your internship. Relate this knowledge to what you were taught and learned from the pillar and courses.

Commented [YG16R15]:

Commented [YG17]: the manners, mindsets, or values that you observed to be important during the internship, for example, Accountability, Honesty, Integrity, Empathy, Teamwork, etc, and relate them to how they could contribute to a successful career in the future.

Commented [YG18R17]:

Commented [YG19]: and briefly describe the quantifiable achievements you have had during your 16 weeks of internship.

Commented [YG20R19]:

Commented [YG21]: the overall experience in the internship, keeping in mind the tasks performed and the learning experience.

I really appreciated this overseas internship opportunity at ExactCure and am thankful to my supervisors and colleagues who treated me with hospitality and patience. Nice, France, where the company is located, is also a very charming city with beautiful architecture and landscape of mountains and crystal clear and blue sea. I spent a wonderful time here.

9 References

Company About Us[company website content].Retrieved from

<https://www.exactcure.com>

<https://gust.com/companies/exactcure>

10 Appendices

(Word count 2250 words)

CONFIDENTIAL