Weeks Report (2018/02/15 – 2018/02/28)

Issued by: Sijin Cheng

2018-02-28

**Time Slot 1**

**What has been done in the past two weeks?**

Last week, our focus has been to make the thesis project start and get a clearer picture of what to deliver.

In short:

1. Has finished my thesis plan

2. Has contacted the examiner, Professor Arne, and got the approved for my thesis plan in WExUpp

3. Has a discussion by Slack with Mr. Falah that delivered a clear overview of his expectations

4. Has a video meeting with Mr. Falah on the details regarding the search query and the ranking principles

5. Has set up Python environment on Anaconda

6. Has preprocessed the Patent files with Scikit-learn: (tokenize, remove stopwords and punctuations, stemming, and build tf-idf matrix, clustering with KMeans)

**What is planned to be done in the next two weeks?**

Focus: Write the first draft of introduction part and theory part, find the appropriate cluster algorithm

In short:

1. Compare different cluster algorithms and choose the appropriate one

2. Find the method to consider weight into while clustering

3. Version 1 of the introduction part will be finished

4. Conclude the key techniques and write the theory part

**Which problems exist?**

We currently not have any problems with our project.

**Time report**

Is attached to the next page of this document.

