WEEKLY INTERNSHIP REPORT

18th July – 24th July

HUYNH THANH QUAN

# Plan of this week

* To evaluate AI system by current data set
* To create training set
* To evaluate AI system by latest data set
* Self-introduction presentation in July 18th
* HR interview in July 19th
* Business trip to Numazu July 20th

# Detail schedule of this week

## Wednesday

* + Encrypted text data in type UTF-8 to avoid of losing information
  + Preprocessed the raw text data (removed inappropriate punctuations)
  + Read and clean text message from CSV of Sakai-san, Akasofu-san, and Oyu-san
  + Aggregated information and stored into MongoDB without overlapping data
  + Exported MongoDB data into local CSV files (2 database to avoid of lost data)
  + Concluded first SVM model (result is not good)
  + Analyzed the result and found the reasons and solutions
    - Amount of data is not sufficient
    - Noise in data
    - Features are not clear
  + Self-introduction presentation in 1:30PM
  + Continued to find more data (bug commits)
  + Tried to find some distribution diagrams of the data

## Thursday

* + HR midterm interview (9:00AM – 9:30AM)
  + Enhanced Data Visualization with Seaborn library (box plot, join plot)
  + From box plot visualization, I recognized a better pattern that data distributed in exponential function.
  + Applied the logarithm base 10 to normalize data (better visualization)
  + Built the second SVM model with new features (normalized data)
  + Recorded the result into CSV and compared with the first SVM model
  + Even though, the result was better than the first SVM model (higher precision and recall) but it was not good enough
  + Attended the Japanese training class

## Friday

* + Business trip to Numazu in whole day

## Monday

* + Recognized that linear solution is not flexible enough to solve this problem (many data is overlapped)
  + Concluded to use non-linear solution (neural network model)
  + The result was outstanding in compared with using SVM model
  + Recorded the result
  + Tried different potential initial starting points to find the best neural network model
  + Saved configuration, weight, parameters of model into SAV file

## Tuesday

* + Attended the Sprint Review/Retrospective and Sprint Planning
  + Pair-worked with Akasofu-san and wrote wiki of Requirement, Testing and Result

# What I have learned

**Technical**

* Preprocessing raw text data (remove punctuations before storing or they will cause problems)
* Handling with different data structures
* Improving data visualization to help recognize hidden patterns
* Finding hidden patterns in data
* Experiencing on neural network models (deep learning)
* Saving into SAV file and loading model

**Social**

* Self-presentation and QA with colleagues in Numazu
* Travelling business trip in Numazu
* Experiencing on Fujitsu technologies and facilities
  + Server and liquid immersion cooling system *(I really like it)*
  + FACOM 128B
  + Fujitsu wireless, wire devices, electronics, telecommunications, software, hardware…
  + Factory facilities
* Experiencing on Fujitsu history and activities
  + FACOM 100
  + History with German conglomerate Siemens
  + Founders of Fujitsu…
  + Historical events
  + Green farm, Charity…
* HR interview

# Plan of next week

* Writing wiki (Requirement, Implementation, Testing and Result)
* Continue to increase bug commit data
* Weekly report
* Presentation for our division and Quality Assurance Division
* Prepare for final Presentation
* Attend Final Interview

HUYNH THANH QUAN