Team 4: Bi-Weekly Report 6

Project Title: PEACH Data Mining Date: 26 January 2018

Gavin Shek, Saqib Jahangir, David Stepanovs

Overview

For these past few weeks, we have began redesigning the user interface for the existing data anonymisation and generation tool, which was created by a previous UCL PEACH team. Following a client meeting, we are developing a program designed for users by providing features and functionalities as requested by our client. In terms of the data visualisation aspect of the project, we have begun connecting the existing PEACH core analytics system with Kibana and Elastic Stack. We have also created a video, added to our website and presented an elevator pitch regarding the project.

Meetings

Client Meeting - 19th January

We spoke with our client on Friday 19th and were given an in-depth explanation of the features of the proposed system. The client gave a rough idea of what the UI of the anonymisation system should look like and how the anonymisation should work. The client added that the software must take in a data sets, of varying formats, and allow the user to choose which columns to remove, randomise or anonymise. Therefore since the tool is random, it must be non-deterministic.

Aside from this, the client said the system must be able to run on most common platforms. The client suggested we should possibly try and and deploy the anonymiser using React Native so that the anonymizer works on all platforms, however, the anonymisation libraries are written in Java.

For the other aspect of the project, the data visualisation, we discussed the use of Kibana and Elastic Stack. Navin suggested us to research into Spark Connectors and Elastic Stack Hadoop to be able to connect with the existing system. Finally, we talked about how the entire core analytics system is to be deployed in the future, and possible orchestration engines.

Numerous Meetings During and Outside of Lab Sessions

We have had various meetings where we have planned and delegated our tasks. Together, we prepared a script for the video and chose a template for the website. We also discussed the tasks required and set internal deadlines for team members to meet.

Completed Tasks

- Began redesigning the data anonymiser user interface
- Set up Kibana to run locally on development machines
- Finished research of existing data anonymisation tools
- Created a video and website about the project
- Presented an elevator pitch about the project

Problems to be Solved

- To ease deployment by dockerising the entire Elastic Stack
- To optimise the deployment of the existing core analytics system

Plan

- Complete the anonymisation user interface
- Connect PEACH core analytics with Kibana and Elastic Stack
- Research alternative data visualisation tools
- Create a better design for the project website

Individual Section

Gavin Shek

I set up Elastic Stack locally on my laptop and have begun researching and using connectors to connect Kibana to the existing core analytics system. I have also been attempting to dockerise Kibana but have faced some problems which I am currently working on.

Saqib Jahangir

I worked on researching all existing anonymization tools and wrote a report on possible solutions. I concluded which technologies are best to pursue for this project.

David Stepanovs

I have worked on the data anonymisation part. I have looked at how the previous team has implemented their data anonymiser and which aspects can be improved. I have started redesigning the UI using JavaFx and material design library.