Microsoft Capita Team 2 / Bi-Weekly Report 1

Date: 10/10/2016

Team: Lambros Zannettos, Nathan Liu, Junwen He

Overview

This first week has been mostly spent researching the project and the relevant technologies that we will be potentially using. Our current understanding of the project is that Capita, a Microsoft software partner, has a product called SIMS. This is a school management system which is used by most schools in the UK and as such, Capita has access to a very large set of data which can potentially be used to answer significant questions and solve big problems that schools and the education system has in general. Our project, therefore, will be involved with the mining and manipulation of that data.

The programming languages we are considering using are C# or F#. The factors affecting our choice of language include API support for the language, online support for common problems and bugs associated with the language (availability of resources, StackOverflow for example). We have not yet had an opportunity to go through requirements with the client so the possibility of functional languages such as F# is still open.

We have also looked through various APIs to use. In terms of machine learning we are looking to use Accord.NET and with a possibility of Word2Vec.NET to help with natural language processing (NPL) for a machine learning model in C#. For data visualization in C# we will most likely use the Microsoft Graph Class in C#.

Meeting summary

Tue, 4 October 2016
During the Systems Engineering Project practical session, we were given some information about the Microsoft projects in general, and then a bit more information specifically regarding the Capita teams, from a Microsoft representative.

Thu, 6 October 2016
The team met up to discuss the project and any initial research done regarding Capita and the SIMS software package, as well as any relevant Microsoft technologies that we will potentially be using during the project.

Tue, 11 October 2016
We further explored possible technologies we can use, and, having received some more detail on what the project will be about, started brainstorming about potential uses of the data Capita has access to.

Wed, 12 October 2016
Meeting with Steven Law from Capita. We were given some more information on the nature of the project, and a few more contacts to get in touch with in terms of the technical aspects of the project.

Tasks Completed

1. Researched Capita/Sims.

2. Started familiarizing ourselves with relevant Microsoft tools (PowerBI/Microsoft Office Group).
3. Looked into potential languages and APIs to use, although the project brief is not clear yet so these are subject to change.
4. Had a meeting with our contact from Capita.

Problems to be resolved

1. The project brief is still not very clear.

Plan for next two weeks

1. Arrange a meeting with James Randall from Capita, who is the project mentor.
2. Clarify the brief.
3. Assign clearer roles within the team.
4. Get in touch with Geoff Hughes about possibly borrowing some Windows machines to work on.
5. Get a copy of SIMS with a training data-set from Steven Law.

Individual reports

Lambros Zannettos:

During these first few days, I have spent most on my time on this project researching Microsoft technologies, and looking into Capita and SIMS. Capita's official YouTube channel helped to better get acquainted with SIMS and what it can do. I also found some relevant courses on Lynda that could benefit the group (courses on C#/Data science etc.) and have also reminded myself of the use of PowerBI and Azure Machine Learning. I was fortunate enough to have used both PowerBI and Azure Machine Learning at the Microsoft Data Challenge hackathon last year, so they are both relatively familiar to me.

I have also been working with Xamarin/C# for the past few months, so I have started thinking about possible ways we could use this to perhaps develop a phone app which would be beneficial to the client and relative to the project end-goals.

Nathan Liu:

I have been doing research into languages and machine learning tools that may be used for the project. I previously worked with Java, DeepLearning4J and Word2Vec in Java and Python (gensim) so I initially wanted the project to be in either Java/C# or Python. Upon further research I came across a handful of .NET machine learning libraries for C# and F# and so I chose Accord.NET and Word2Vec.NET.

For data visualization I suggested the Microsoft Graph Class with the possibility of using the principal component analysis (PCA) for dimensionality reduction since Accord.NET already has a PCA Class.

Junwen He:

I don't have too much experience which is relevant to this project, but since we are going to use C# or Java, my plan is to learn as much as I can on these languages and other relevant subjects like big data and machine learning. I will go through these lessons by watching videos on Lynda. And I will try my best to work with this team.