Microsoft Capita Team 2 / Bi-Weekly Report 3

Date: 18/11/2016
Team: Lambros Zannettos, Nathan Liu, Junwen He

Overview

The week has been particularly tough because the team was dedicated towards the scenario week project. We have done a lot of research into Natural Language Processing and other possible methods of transforming user inputs into SQL queries. We have also been setting up SIMS but are still having problems with authentication and importing the data provided by Capita. Our client is aware of the problem and steps are being taken to fix this.We also had a meeting with Scott Bell, our client representative, to discuss our ideas about the project. He liked the direction we were taking the project in, and has encouraged us to continue in this direction. This was also taken as a further refinement of the project requirements, and we are now confident that we have set ourselves some clear goals.

Meeting summary

Mon, 28 October 2016

We had a quick meeting after our first scenario week lecture. We discussed that we would be researching technologies while Junwen attempted to set up SIMS, Nathan doing research into NLP and Lambros helping out with both.

Thu, 3rd November 2016

During this meeting with met with Scott to help us resolve our problems with the SIMS setup. Scott gave some hints and pointers to set us in the right direction. We attempted to reinstall SIMS but ran into more problems with the set up.

Tue, 8th November 2016

We briefly met to fill each other in regarding our individual research. Junwen was set to do work on the website while Nathan and Lambros continue with research into NLP. We also arranged a Skype meeting the following week with Scott and a member from SIMS service desk to help install SIMS onto our laptops.

Wed, 16th November 2016

We had a meeting with our TA to discuss our research and progress. He suggested we implement a minimum viable product by December. We were informed that the code had to be delivered by December so we made a command line application where users could automatically generate questions from a decision tree that would in turn map to a SQL query. This command line application would eventually be replaced with the web-app that will be deployed on Azure.

Thurs, 17th November 2016

We help a Skype meeting with Scott and a member from SIMS service desk to install SIMS onto our windows machine. We ran out of time to install SIMS so we have arranged another Skype session with Scott next Monday afternoon to install SIMS.

Tasks Completed

1. Research

The team has been researching techniques for natural language processing (NLP). We considered Restricted Boltzmann machines (RBM) or a Latent Dirichlet Allocation (LDA) for topic modelling. We were not sure about the LDA because we were not sure how to create an unsupervised machine learning model using this technique. So we opted with the RBM since we are confident they can support unsupervised NLP. We are currently seeing whether a deep belief net would be better suited for unsupervised NLP (since they are RBMs stacked on top of the other). On another note, if an unsupervised model cannot be feasibly made then we can consider a supervised machine learning model for generating SQL queries from an input string or perhaps look to the Azure Machine Learning platform.

We also looked at how search engines interpret the semantics of a user input since we believe that a search engine carries out a similar same task of transforming a user input into a SQL query. We have looked some of the theories of sentence parsing and semantic analysis from cognitive psychology and also looked into the search techniques used by Google. We have looked at concepts in semantic search including Resource Description Framework (RDF) path traversal and Fuzzy concepts.

We also looked at existing NLP libraries and products, like Kueri.me, OpenNLP, SharpNLP etc.

2. Reached out to Professor Richard Noss at the Institute of Education, and waiting to hear back to set up a meeting and start discussing the project.

Problems to be resolved

1. Finding a feasible natural language processing algorithm to transform human inputs into SQL queries.

2. Research into decision trees, how to implement it into a database and use them to create questions for users.

Plan for next two weeks

- We plan to continue with getting familiar with NLP concepts. Most of the time spent on research will be devoted to going through the NLP course taught at Stanford and at University of Michigan (Coursera).

- Get to grips with MS Azure and figure out the best way to deploy a Web App.

- Keep working on project website.

- More research into existing technologies.

Individual reports

Lambros Zannettos:

Following our client's approval of the idea to use a form of Natural Language Processing to create a system for asking the database questions, I continued my research on the subject. I have enrolled on an online course offered by the University of Michigan titled *Introduction to Natural Language Processing* (found on coursera.org). I have also done further research on existing NLP libraries and products, and ways to use them for our purposes. Since our project will most likely take the form of a Web App, I also started looking into the best way to achieve this using MS Azure (trying to use the lessons learned from Scenario Week where Amazon Web Services was used to deploy a Web App.)

Nathan Liu:

I proposed the possibility of using a RBM or LDA as a starting point for an unsupervised NLP (or semi-unsupervised NLP) model. I opted for the RBM since I am more familiar with using RBMs however I am aware that RBMs are already very old and so I suggested there are better ways for NLP (or interpreting the semantics in a search term) other than RBMs. As a result I looked into how cognitive psychologists attempted to explain NLP. Most of the ideas from cognitive psychology were very high level, the theories I looked at (e.g. garden-path model) did not seem like reliable models for sentence parsing and so I looked into how search engines analysed a user input and came across concepts such as RDF path traversal. I am currently on the Standford NLP course to see more viable alternative of NLP for this project.

Upon the meeting on the 17th I generated a decision tree that could be implemented into a MySQL database. The database was then connected to a C# application where users could automatically generate questions by navigating through the decision tree. When a question was fully generated the database would return a SQL query (that could be used to call SIMS data) and return the answer to the question for the user. I migrated the database from localhost onto a MySQL database in the cloud. I am now looking to integrate NLP features into the application such as word correction.

Junwen He:

This week I look into our project website, and I have uploaded a demo through FileZilla to the department web server and it worked well. So the following week, I'm going to edit the website and upload our process to record what we have done. I also watched the data mining course on Lynda and will keep learning to help doing this project. Further more, I am watching online course about NLP to gather more information about it and help with the project.