EE/CPR E/ SE Weekly Report 6
March 25 - March 29, April 1 - April 5
Group Number: 08
Project Title: Artificially Intelligent Requirements Analysis Tools
Client & Advisor: Collins Aerospace & Dr. Simanta Mitra
Team Members & Roles:
- Ryan Cerveny: Scrum Master, Meeting Scribe/Facilitator, Project Lead, Communicator
- Apurva Patel: Project Lead, Report Manager, Technical Support, Communicator, AI Training Lead
- Jonathan Murphy: Testing Engineer, Researcher, Requirement Lead
- Takao Shibamoto: Chief Engineer, Researcher, UI Lead

Weekly Summary:

Over spring break and the week after, our team focused on creating the design document. This past week we put significant effort into the first working prototype. We shifted from the theoretical design stage to implementation stage. From the client's perspective, their focus significantly shifted towards the AI/ML portion of this project. We won't be doing the GUI and other stuff and instead we all will focus on what we can do with AI/ML.

Past Week Accomplishments:

Design Document (Ryan & Apurva)
Ryan and Apurva wrote the design document.

Found Mock Data to Use for a Prototype(Ryan & Apurva):
Found requirement data that we can use in to create the prototype since collins requirement data needs cleaning to be used.

Parsing the Training Data Files(Apurva):
Apurva began to look at parsing the csv files using python, and have been thinking on the algorithm as per different requirement and how to overcome it. Started coding on parser. Completed the parser for the example data. Currently, he is continuing working with the original dataset provided by the Collins Aerospace.

Design thinking on prototype design and implementation of word2vec training, and logistic regression model training (Ryan)
Parsing/cleaning the data has continuously given us grief. To negate this for now, we found mock data to create a prototype with. The mock data is a set of requirements used in a different requirement tracing project so and was much easier to clean (contained minimal acronyms and abbreviations).
As for training the model, we need all of the links to be classified. However, the files we were provided with have no classifications associated with them. Not only will we have to assign classes to each link, but the files are also only examples of good links. This means we will have to mock bad links which could be risky because they might not represent true mistakes that will be encountered in the real world.

Design thinking on training our model using external data (Takao)

Discussed the lack of data problem and suggested solutions about how to deal with it (developing Wikipedia crawler to collect data, replacing acronyms with proper English texts etc.)

Began to develop WebCrawler (Jonathan)
Jonathan began writing a program that would gather data from Wikipedia pages that related to Aerospace topics in order to train our model. After some discussion though, the team is not sure if this is the method we are using for training, so it may not be used.

Edited GUI (Jonathan)
Jonathan made some slight edits to the GUI for a cleaner look and better user experience.

Researched Transfer Learning (Jonathan)
Jason had mentioned to the team that Transfer Learning may be an important topic going forward. So, Jonathan and Takao found information online about topics related to Transfer Learning to develop a plan on how to apply it to this project.

Individual Contributions:

Name	Individual Contributions	Hours 3 Weeks	Hours Cumulative
Jonathan Murphy	• Design document • Began developing WebCrawler • Made some slight changes to GUI • Began researching transfer learning	20	Approx. 62
Apurva Patel	• design documentation • Developing parser for csv files	Approx 20	66
Ryan Cerveny	• Design document • Designed Prototype • Found mock data to temporarily replace collins data • Researched library options for logistic regression training • Implemented model training, word2vec training, and visualization of model/data for prototype • Communicated problems with advisors and supervisors • Began thinking about alternate solutions	Approx. 28	Approx. 74
Takao Shibamoto	• Discussed the problems we are currently facing and provided solutions about what should be done now. The solution I discussed is that instead of worrying about lack of data and hard-to-parse requirements data. We should just	20	61

| | | assume the data structure and we should focus on AI/ML portion of the project as the client suggests.

• Cleaned the sample data and helped other teammate implement the parser using Jupyter Notebook
• Experimented with RNN
• After the first prototype is done, I summarized what the problems are and suggested possible solutions. The problem concentrates on what to do with bad links and preprocessing the data. I suggested pre-categorizing the req/uc and do the prediction within a category. Also I suggested we shouldn't deal with bad links, and instead use threshold values for the likelihood that wmd provides. | | |

Plans for Upcoming Week:
- Finish up the prototype
- Brainstorm what needs to be done next
- Meet Collins Aerospace Representative
- Research/Experiment transfer learning

Summary of Weekly Advisor Meeting:

- No meeting with advisor. Will be meeting with the client over the weekend to discuss where to take the project since the approach for the original experiment needs to be rethought