

EE/CPR E/ SE Weekly Report 8

April 12-April 19

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:**

The team proposed an adjustment to the project plan this week to make the AI process more efficient. The team mostly waited on Collins Aerospace decision for the new proposed design. Overall the idea was shot down but concepts such as standardization of requirements can be applied still to our algorithm decisions if it will help. Work started on the design document as well as dividing work for experimentation on other approaches for a similarity algorithm.

### **Past Week Accomplishments:**

#### Conceptualize and Design Requirement Hosting Tool (Ryan):

Ran with idea for updating Collins' requirement hosting platform and using it to standardize their requirements, proposed it to collins supervisor and designed the system for if it were to be used.

#### Began Design Document (Ryan):

Began laying out the design document and writing for the requirements hosting tool. However the idea was scratched for the requirement hosting tool so the document will have to be restarted.

#### Project Planning & Clarification (Apurva & Ryan):

Apurva and Ryan thought of the alternatives that could be worked on. As a team, Ryan and Apurva came up with the best possible solution for moving forward to Collins Aerospace expectation.

#### Research on Requirement Tracing(Apurva):

Apurva research on requirement tracing and how it works. Compared the solution that team has been implementing so far with similar project done by several other university or organization. As a result, we came to a conclusion that we are somewhat on the correct part.

However, there are few techniques that needs to be changed. The finding has been shared with the team by Apurva on slack channel.

#### Dataset(Apurva):

As Collins Aerospace restricted a bit on providing the data. Apurva looked for various dataset that could help us to train our model. There were quite a few dataset available and has been shared with the team by Apurva on the slack.

#### Research on Alternative Solutions (Jonathan)

Jonathan had a very full schedule this week with work for other courses, so was unable to complete much research about alternative algorithm solutions. However, over the weekend Jonathan will research GloVe and perform experiments for feasibility testing. He will share his findings during the sprint meeting with Collins on Monday.

#### Topic modeling experiment (Takao)

##### Problem

- How do I compare the results of topic probability distributions?
- It involves finding similar words and replacing them before doing LDA since for example LDA doesn't tell the difference between 'user' and 'users' → we can use google word2vec model for this. → how to actually decide which word to use consistently?
- I gave up with Pine data analysis (topic-modeling-experiment-1) because Pine data is way too small to automatically decide stop words. → but maybe even in real world situation, automatically deciding stop words maybe difficult. It might be better to let engineers choose stop words
- Approach Using google doc2vec model to unify similar words might unify the words too much. → maybe better to just simply use lemmenization??
- I am feeling like there is a better algorithm to do topic modeling. More comprehensive and no worry about this stuff.

##### Current roadmap

1. find top N most used words and prompt engineers to choose stop words (this technique would apply to any kind of algorithms we would use)
  - eg. we probably don't need "shall" in REQ and "user" in UC
  - maybe completely remove any verbs
2. somehow unify similar words
  - eg. unify "misspell", "misspells", "misspelling", "misspelled" into "misspell"
3. Apply LDA and compare the probability distribution for each req for each uc
4. each req chooses top X similar UCs

#### **Individual Contributions:**

Name	Individual Contributions	Hours	Hours Cumulative
Jonathan Murphy	-Researching alternative solutions -Began design document	6	Approx. 76
Apurva Patel	-Dataset - Research on Requirement Tracing -Team Bonding	8	86
Ryan Cerveny	<ul style="list-style-type: none"> <li>● Thought up and designed prototype for requirement hosting tool</li> <li>● Proposed requirement hosting tool with collins supervisor</li> <li>● Discussed assignments with team members</li> <li>● Began design document</li> <li>● "Team bonding"</li> </ul>	8	Approx. 92
Takao Shibamoto	Continue experiments with topic modeling, specifically using LDA Stuck at the problem of how to effectively compare topics.	8	84

**Plans for Upcoming Week:**

- Takao and Apurva will be focusing on prototyping the research performed by Takao and Jonathan.
- Everyone will be focusing on developing the design documentation and presentation

**Summary of Weekly Advisor Meeting:**

Apurva had a meeting with Dr. Simanta Mitra. He has updated Dr. Mitra about the project been a continuation based on research and getting something that could work in the process. Thus, the team will be focusing on research, prototype, and developing the design documentation and presentation for the upcoming week.