

ITS-Chatbot Spring 2022

Project Proposal

Group Membership

Max Everest (meverest7@gatech.edu) 10:30-11:30 am	Skills	Java, Data Structures
	Credit	1 credit hour (4 hours/week)
	Responsibility	Research an efficient algorithm that uses the current questions dataframe to help complete user queries in correspondence with Nirjhar's competing algorithm. Implement that algorithm. Determine the more efficient one vs Nirjhar's. If needed: Front-end responsibilities.
Nirjhar Deb (ndeb6@gatech.edu) 10:30-11:30 am	Skills	Java, Python, HTML, CSS, Javascript, MATLAB
	Credit	1 credit hour (4 hours/week)
	Responsibility	Explore the Chatbot repository and run it against different inputs to understand how it works. Research various machine learning algorithms to predict the user's question by retrieving relevant questions from the database. Implement one or more algorithms to achieve this goal.
Zhen Hong Tan (ztan95@gatech.edu) 10:30-11:30 am	Skills	Java, Python, C++
	Credit	1 credit hour (4 hours/week)
	Responsibility	Explore and getting used to the current ChatBot, research the current data set and design and implement an efficient algorithm to parse the keyword from the queries and store questions in an array for each keyword. Assist in implementing algorithms if needed.

Project Timeline:

All the weeks are based on the official VIP-ITS [schedule](#).

- Week 1-3 (Jan 13 - Jan 26): Project planning, team building, proposal drafting
- Week 4: Explore the Chatbot, test it against different inputs to understand it fully
- Week 5-6: Research appropriate data structures and ML algorithms to implement
- Week 7-10: Build the data structure, sort the previously asked questions using the ML algorithm, and get the Chatbot to predict the user's question
- Week 11 (March 23): Spring Break
- Week 12-15: Continue working and general improvements and optimizations
- Week 16-17 (April 25 - May 3): Final Presentation

Project Description

We are primarily going to be working on the chatbot side to have direct access to the dataframe. We plan to utilize the existing dataframe that stores all the questions that have been asked to the Chatbot over the past few years.

We will then parse out unnecessary words in the questions to get keywords. Each question will then be sorted into various arrays, one array for each keyword.

Then, we will implement machine learning algorithms that will look at the most relevant keyword in the user's input, go to the array of questions containing that keyword, and select the most relevant question to autocomplete the user's question.

For example, if the user asks "What is a circuit...", the program will identify the most relevant keyword in the question (i.e., circuit), look at an array of previously asked questions about "circuit," and then select the most relevant question to autocomplete the user's question. In this example, let's say that the most relevant (based on ML algorithm) question related to "circuit" is "What is a circuit board?". The program would use this question to autocomplete the user's question with the word "board?".

NOTE: We will likely not build our own algorithm since that requires a very complex skillset. We will research various algorithms and use an appropriate one for this project.

NOTE: The specifics are tentative. We will modify our approach and be flexible as we get further into the project.

Overall Goal: Get the Chatbot to predict the user's question based on partial user input.

Foreseeable Challenges

For All Us:

The final product will interop multiple languages and distinct areas of development into a single feature. A big challenge may be finding a way to make each part utilize and synergize with all other parts. It could also be hard to properly use and integrate our peers' code with our own code without fully understanding their code and thought process.

For Max:

Lack of proficiency with non-java languages can delay the process. Hopefully, the concepts of java could translate over to all languages and might diminish this obstacle. The main repository also consists of several years of design and implementation, which could lead to complications in understanding how the pre-existing databases work.

For Nirjhar:

Understanding the complexities of the Chatbot may take a lot of time since this is my first large, multi-developer project. Since I have just a primitive understanding of machine learning in Python, implementing more complex ML might be very difficult at first, but hopefully, with assistance from the VIP team, it will work out.

For Zhen:

Understanding the current database can be time-consuming due to a lack of experience. Designing an efficient way to parse the keywords from the queries can be difficult.

Implementation and Collaboration

This project will be built using PyCharm. This project is mainly developed in Python and will continue to be so. We will fork the Chatbot-v2 repository on Github and manage our files in a new repository. Each member should create their own respective branches on the repo and add more if needed. Development should be done in member's respective repositories and merged into the master branch on a weekly basis. A review from the person other than the Pull Request initiator is required to merge changes into the master branch. For communication, we will use GroupMe and text.