B.Sc. (Hons.) Degree in Computer Science

University of Sri Jayewardenepura

Model a Academic Counselling chatbot for CS and ICT students of University of Sri Jayewardenepura.

K.G.S. Manimekala

AS2019445

Academic Year 2023

Supervisor: Prof. T.G.I. Fernando

Professor in Computer Sciences, Department of Computer Science, University of Sri Jayewardenepura.

This independent research proposal is submitted in partial fulfillment of the requirement for the B.Sc. (Hons.) Degree in Computer Science at the University of Sri Jayewardenepura.

Contents

1 Introduction 3
1.1 Chatbots 3
1.2 Problem Background 3
1.3 Aim and Objectives 3
1.4 Motivation 4
2 Literature Survey 4
3 Methodology 6
3.1 Data Collection 6
3.2 Preprocessing 6
3.3 Model Implementation and training 6
3.4 Model Evaluation 6
3.5 Deployment 6
4 Timeline 7
4.1 Gantt Chart 7
References 8

1 Introduction

1.1 Chatbots

A chatbot is a conversational agent which is basically a computer application that is designed to simulate and process human like conversation with humans using text or speech capabilities. A chatbot program can communicate with humans as it is a real human being by utilizing natural language processing (NLP) and natural language understanding (NLU) technologies. These chatbots use algorithms to understand intents of a user from a user input prompts and provide replies based on their pre-defined rule set or data that it has been trained on. With advancements in AI and ML, chatbots have evolved from rule-based systems to more sophisticated NLP models.

1.2 Problem Background

Academic advising and counselling is a crucial and challenging task in a university at the beginning of each term specially when a new batch of students are being selected to the university and throughout the whole year. It remains as a manual process where students always seek guidance to their academic concerns through academic staff, senior students or their fellow batchmates. This issue becomes even more complex due to the diverse backgrounds of students, including those from rural areas and various personality types, such as introverts who may hesitate to seek assistance. Consequently, addressing the diverse array of academic-related inquiries can be difficult for both students and university staff. Recognizing this common difficulty experienced by all students, including myself, this research aims to provide a systematic solution to minimize these challenges.

1.3 Aim and Objectives

Aim: To develop an AI-based chatbot to provide academic counseling and support to CS and ICT students at University of Sri Jayewardenepura.

Objectives:

- Investigate different architectures and frameworks used in the development of AI-based chatbots.
- Research and identify existing pre-trained NLP models suitable for chatbot development.
- Study about different evaluation metrics and techniques for assessing chatbot performance.
- Create a Suitable FAQ Dataset for Academic Counseling and Support.
- Develop an AI Model or Utilize Pre-trained Models for the Chatbot.
- Evaluate the Developed Chatbot.
- Create an efficient and user-friendly chatbot that enhances the learning experience for students and reduces their dependency on human support.
- Integrate the developed chatbot into the Department of Computer Science website to make it accessible to students.

1.4 Motivation

The motivation behind this research stems from the need to provide timely and accurate support to students who follow CS or ICT as a subject. Students often face challenges in finding immediate assistance for their academic queries, which can be time-consuming and frustrating. By developing a chatbot, students can access information anytime and anywhere, empowering them to make informed decisions and navigate their academic journey more effectively.

2 Literature Survey

This paper presents a chatbot implementation for college students and teachers (Lalwani et al., 2018). They utilized AIML for dataset creation and a rule-based approach. Preprocessing involved WordNet for lemmatization and word similarity measures Path Similarity and Wu-Palmer (WUP) Similarity for semantic sentence similarity.

(Ayanouz et al., 2020) highlighted key concepts and AI techniques necessary for developing an intelligent conversational agent using deep learning models, the underlying principles and technologies involved in building effective chatbots. Also, they proposed a functional architecture for an intelligent chatbot specifically designed for healthcare assistance.

The authors developed "MyAdvisor," a chatbot-based advising system (Kuhail et al., 2023). The paper mainly focuses on system usability and usefulness. Dialogflow, intent manager, database, sentiment analysis API, and text similarity API has been used for the system's architecture.

The paper (Ishita Shah et al., 2023) discusses the evolution of chatbots, development techniques, and evaluation methodologies of chatbots such as DeepProbe, MilaBot, SuperAgent, Meena, and ChatGPT.

(Hettiarachchi & Gamini, 2023) has implemented virtual conversation agent based on deep learning algorithm which is multilayer perceptron neural network and a special text dataset for conversations about CEB (Ceylon Electricity Board) services.

(Nga et al., 2022) have implemented a chatbot for university admission consulting in English and Vietnamese using few pre-trained multilingual BERT models to predict answers from the input question. To create word representations and find out semantic similarity Skip, Word2vec, GloVe word embedding algorithms have been used. They have compared four BERT models with three methods of catching semantic similarity and one without using any word embeddings. Out of these 4 models with 4 methods *salti/bert-base-multilingual-cased-finetuned-squad* model with Word2vec has given the best results.

Paper name	Methods	Datasets	Evaluation metrics	Pros	Cons
Eliza (Weizenbaum, 1966)	Pattern matching and rule-based approach.	No pre-existing dataset was used like in modern chatbots.	N/A	• Laid the groundwok for future research and development in NLP and AI. • Simple rule-based approach	• Limited understanding of complex language and context.
Implementation of a Chat Bot System using AI and NLP. (Lalwani et al., 2018)	Used NLP, Pattern matching and retrieval using sentence similarity algorithms.	Question and answer pairs stored in AIML files.	N/A	• Simple and easy-to-implement. • fast and efficient search for answers.	• Limited understanding of complex language and context.
Chatbot for Ceylon Electricity Board Website. (Hettiarachchi & Gamini, 2023)	NLP techniques for data preprocessing & a MLP model for intent classification.	JSON format dataset on FAQs based on CEB.	80% / 20% Split & Combination of Monte Carlo and normal k-fold cross validation.	• ML optimization for improved performance.	• Limited knowledge base with 250 patterns. • Inability to respond to new/unprovided questions.
Chatbot system for Da Lat University. (Nga et al., 2022)	Used NLP and BERT models for intent classification.	Admission-related pairs of 800 questions and answers.	F1 score EM score Accuracy	• High F1 (88.6%) and EM (79.6%) scores. • High intent classification accuracy (99.9% & 100%) • Immediate & automatic response to users' questions.	• Low corpus size. • Limited to admission-related questions in the university domain

Table 1: Comparative Analysis of Implemented Chatbots in Research Papers.

3 Methodology

3.1 Data Collection

To create the dataset, we will conduct a survey using a Google Form targeting students at the Department of Computer Science, University of Sri Jayewardenepura. The questionnaire will consist of a series of questions designed to gather information about the students' day-to-day queries and concerns related to course registration, program requirements, assignment deadlines, course details, faculty information, scholarships, and other academic-related queries. The responses obtained from the survey will serve as the input texts for the dataset.

Various sources such as the department websites, student handbooks, and academic resources will be used to gather the necessary information and extract the corresponding output texts, which will form the answers to the students' queries. Then the gathered information will be stored in the JSON format.

3.2 Preprocessing

Necessary NLP techniques will be used for preprocessing tasks such as tokenization and lemmatization. Then the dataset will be split into training set, testing set and validation set.

3.3 Model Implementation and training

We intend to develop a Deep Learning Algorithm or use existing algorithms and it will be trained on the dataset that will be created using a survey.

3.4 Model Evaluation

User based evaluation will be conducted and other existing evaluation methods will be used accordingly to the model evaluation (E.g., cross validation, F1 Score, Accuracy).

3.5 Deployment

The chatbot will be deployed as a feature in the Department of Computer Science website, gather user feedback to assess its effectiveness and make necessary improvements.

4 Timeline

4.1 Gantt Chart

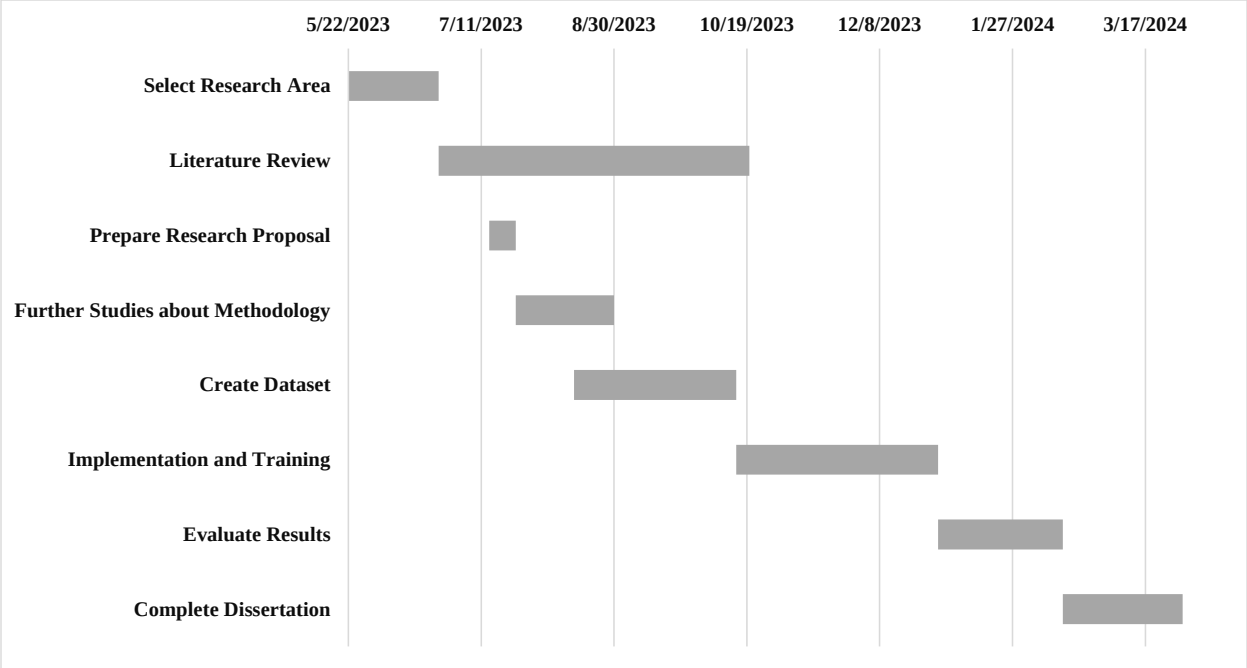

Figure 1: Gantt Chart of the Timeline.

References

Ayanouz, S., Abdelhakim, B. A., & Benhmed, M. (2020, March 31). A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3386723.3387897

Hettiarachchi, D. N. M., & Gamini, D. D. A. (2023). Using a Machine Learning Approach to Model a Chatbot for Ceylon Electricity Board Website. In *Current Scientia* (Vol. 26, Issue 01).

Ishita Shah, Shrihari Jhawar, Aashna Khater, Asher Jacob, & Dr. Girish Potdar. (2023). Chatbot Development Through the Ages : A Survey. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 262–271. https://doi.org/10.32628/cseit2390329

Kuhail, M. A., Al Katheeri, H., Negreiros, J., Seffah, A., & Alfandi, O. (2023). Engaging Students With a Chatbot-Based Academic Advising System. *International Journal of Human-Computer Interaction*, *39*(10), 2115–2141. https://doi.org/10.1080/10447318.2022.2074645

Lalwani, T., Bhalotia, S., Pal, A., Bisen, S., & Rathod, V. (2018). Implementation of a Chat Bot System using AI and NLP. *International Journal of Innovative Research in Computer Science & Technology*, *6*(3), 26–30. https://doi.org/10.21276/ijircst.2018.6.3.2

Nga, P. T. T., Lương, N. T., Thắng, T. H., & Quý, T. D. (2022). AN APPROACH FOR BUILDING A CHATBOT SYSTEM FOR THE ADMISSION PROCESS OF DA LAT UNIVERSITY. *TNU Journal of Science and Technology*, *227*(15), 23–32. https://doi.org/10.34238/tnu-jst.6056

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28.