

Chapter 4

Analysis and Design

4.1 Introduction

This chapter discusses the main requirements arranged through a thorough research on natural language processing, and the design of the proposed prototype and its functionality.

4.2 Identified Functional Requirements

- System shall provide an interface to customer to make his/her queries. (Similar to a chat application).
- System shall have facilities to get information from the organization data base.
- System shall have a knowledge base.
- Knowledge engineer shall have facilities to include new information, update knowledge base.
- System shall have the facilities to identify the customer queries.
- System shall have facilities to obtain the operator help if required.

4.3 Identified Non Functional Requirements

- Knowledge engineer should provide correct information.
- Chat bot should have a relatively large knowledge base to achieve higher level of accuracy in question identification.
- An operator's assistance is required to handle queries which chat bot fails to handle.

4.4 Proposed System

4.4.1 Main areas of the Chatbot

Figure 4.1 illustrates the proposed system design.

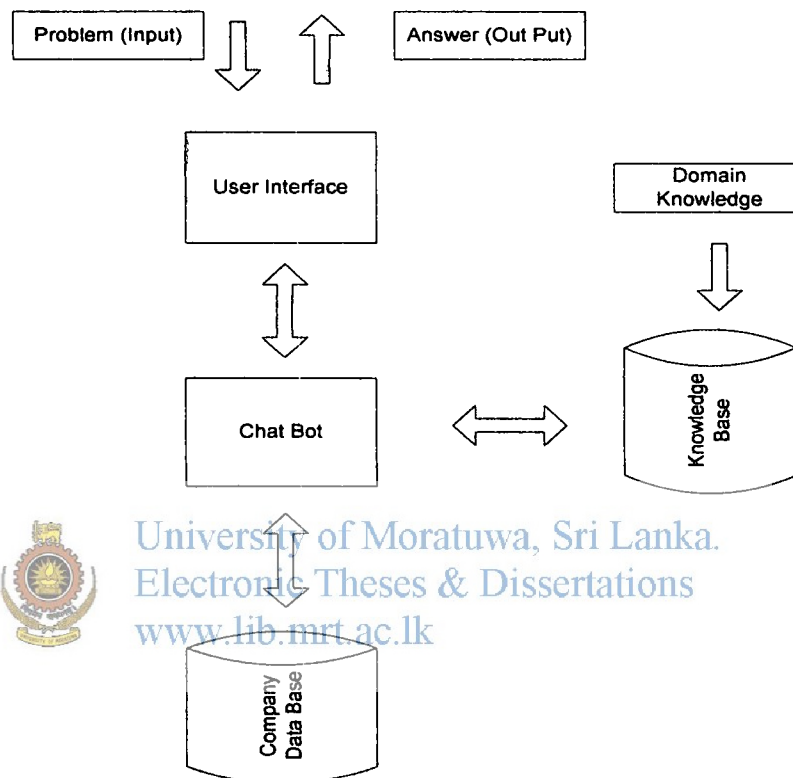


Figure 4.1 - Proposed System

The proposed system has four main areas, User Interface, Chat bot, Knowledge Base and Company Database.

User Interface – This is where the users of the Chat bot interact with it. This interface is similar to any chat window. This has 4 main areas as illustrated in Figure 4.2.

- Display area where all the past chats and answers are displayed.
- Chat Area where the user types the queries.
- Exit button to exit from the Chat bot.
- Close and Minimize buttons to close the chat window or to minimize the chat window.
- Submit button to submit the typed queries to the Chat bot.

Knowledge Engineer Interface – This is the area where the knowledge engineer manages the knowledge and stock information. The knowledge engineer has the ability to add and edit new knowledge/information to the two databases used by the Chat bot as shown in Figure 4.3.

Chat Bot – This is the heart of the chat bot where it identifies the customer queries, and generates answers.

Knowledge base – This is where all the supporting knowledge required to chat bot's operations is stored. Chat bot has the ability to update certain data (probabilities) and new knowledge should be given to chat bot through this database.

Company database – This is where the chatbot receives the stock and related information to generate answers for customer queries. This same database can be used for knowledge capturing about customer queries against time.

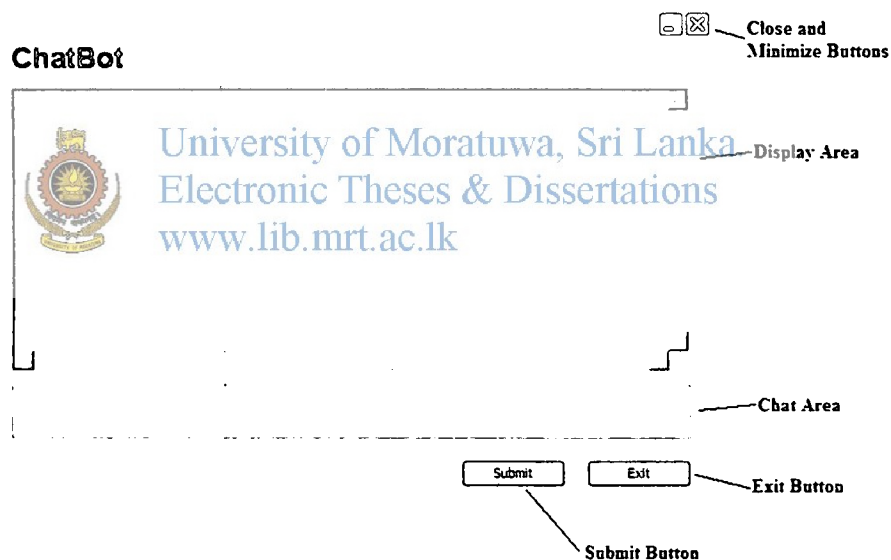


Figure 4.2 - Chat Window

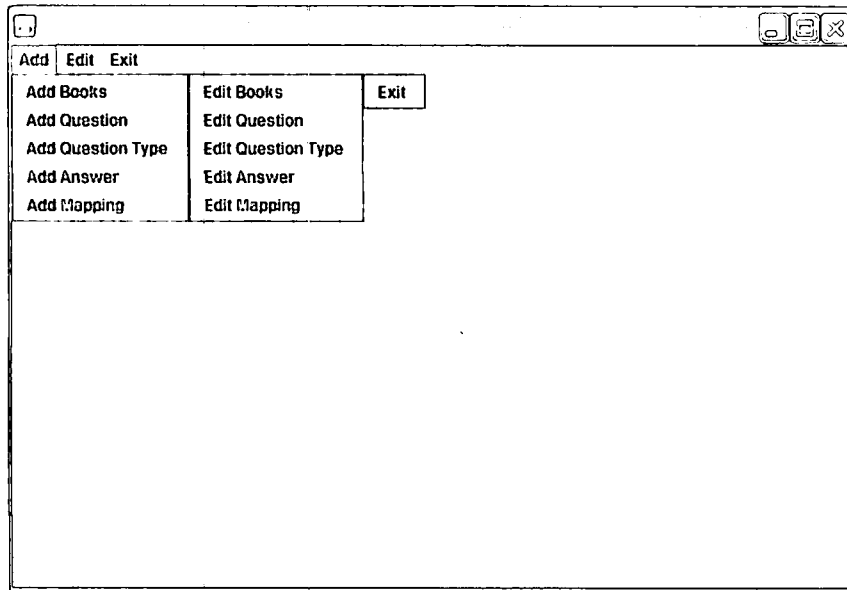


Figure 4.3 - Knowledge Engineer Interface

4.4.2 Architectural Diagram

Figure 4.4 shows the architectural diagram of the system.

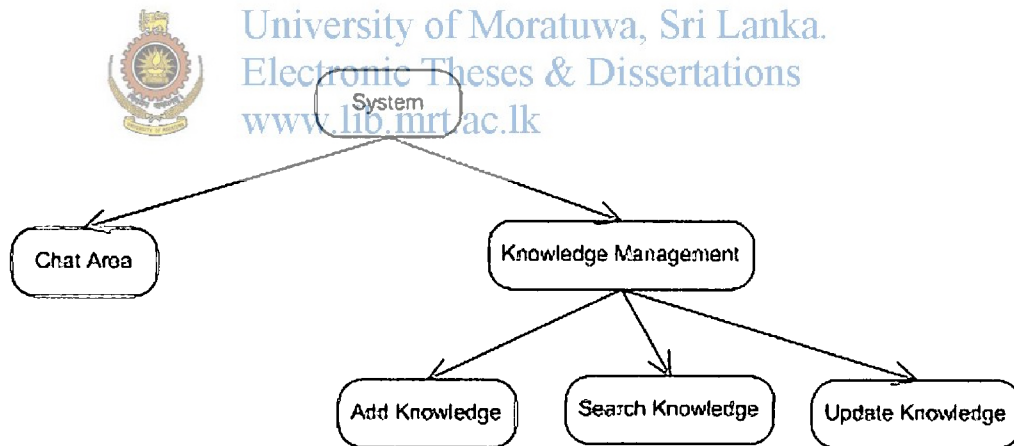


Figure 4.4 - Architectural Diagram

Chat Area – This is where the customer interacts with the system. Customer can type queries and will get answers through this window.

Knowledge Management – This is where the knowledge engineer maintains the knowledge of the system. He/she can add, search and update knowledge. Access to this section is only for selected persons.

4.4.3 Use Case Diagram

Figure 4.4 Figure 4.5 shows the use case diagram of the system.

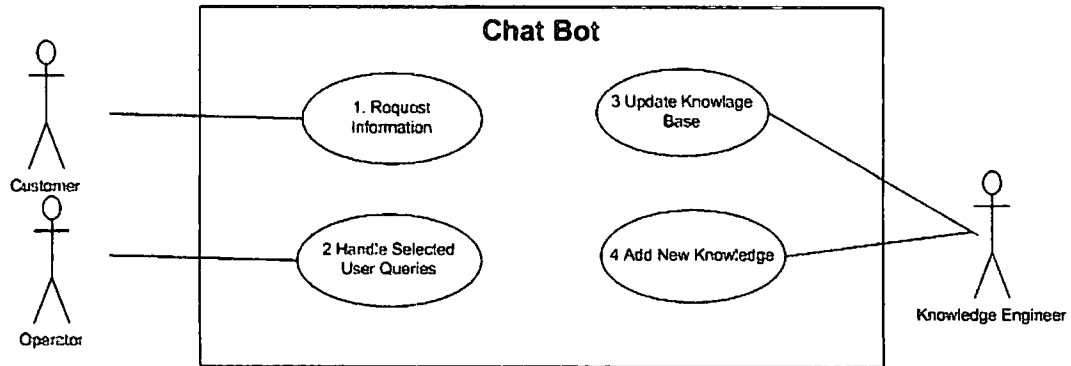


Figure 4.5 - Use Case Diagram of Proposed System

4.4.4 Data Base Design

The Following ER diagram (Figure 4.6) shows the RDBMS tables and their relationships.

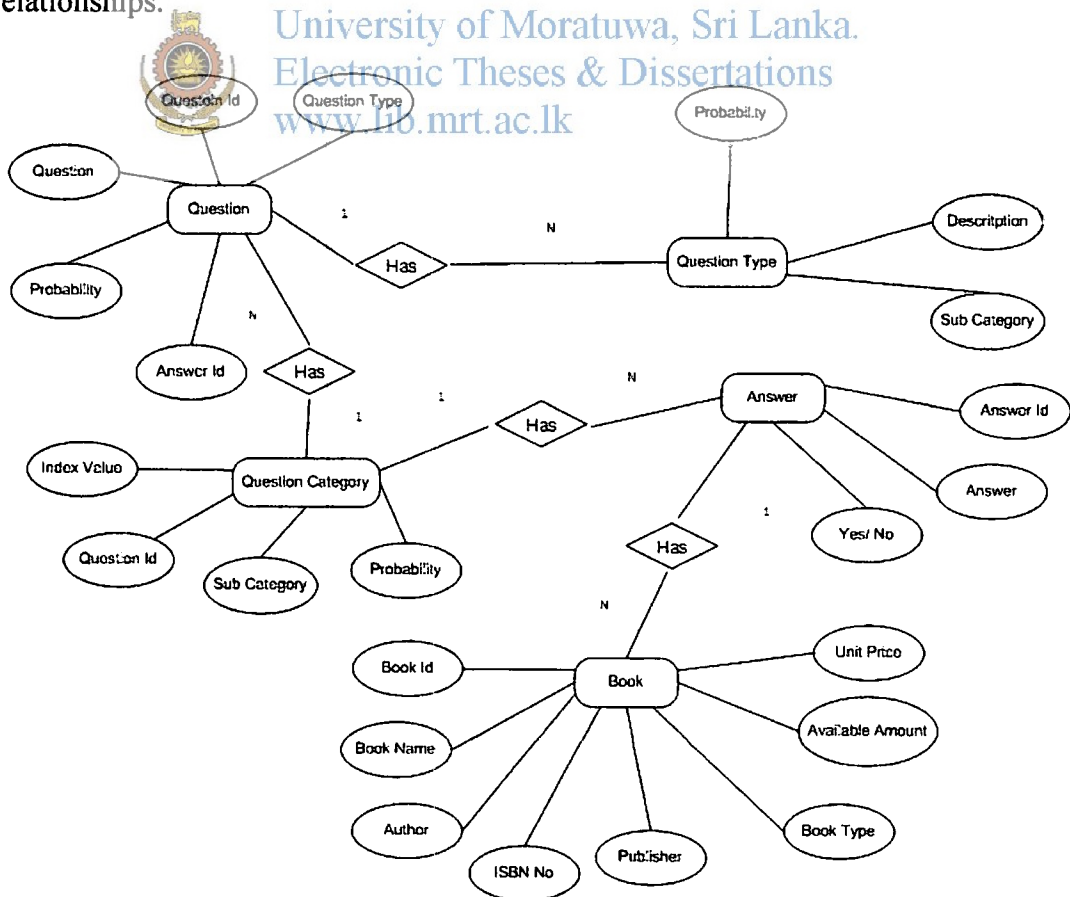


Figure 4.6 - ER Diagram


The system maintains a database to store details about the domain. In-addition to details of the domain, the same database is used to store sample questions, sample answers and other related data required by the application to work smoothly.

The same database keeps domain information (Book shop) separate, which will be used to generate answers to customer queries.

4.5 How Chatbot Works

4.5.1 Identify the Question.

The chat bot will maintain a sample set of questions related to the domain. Each of those questions has its own probability which will be changed/updated on each customer query. The following steps will be followed to identify the customer query:

- 
- University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk
- When the Chat bot receive a customer query it will select the sample predefined question with the highest probability and tries to match it with the customer query. If the matching results are lower it will select the next probable sample question and match it.
 - If the Chat bot identifies a successful match its next step is to extract the core details of the customer query. This is achieved by removing the identical words of the sample question and the customer query and taking the remaining word/s.
 - Each sample question has a predefined sub category. With the help of the sub category the Chat bot tries to find a suitable matching with the core details in the company database.

As an example if we take the bookshop domain.

If the selected sample question is “Do you have _?”, the most suitable sub category will be “Book Name”, “ISBN Number” or “Book Type”. Other categories like “Author”, “Publisher”, etc are not common with the selected sample question.

Each sub category has their unique probabilities (Updated as the sample question probability after each successful matching).

Using the same example if “Book Name” has the highest probability for the selected sample question, the Chat bot will search under the “Book Name” column of the company database to find any matching.

If a successful matching is found by the Chat bot, it will update the probabilities for the selected question by adding 0.01 and subtracting 0.001.

- At this stage the Chat bot has the following details. Sample answer for the identified question, details about the core information, and the core information available in the company database.
- The sample questions used in the chat bot are obtained through a small survey done among a selected set of people. Where the selected people are asked to provide various types of questions they are going to ask in the selected domain (Book shop).

4.5.2 Prepare the Answer



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mru.ac.lk

With the identification of the customer queries and the required data the Chatbot has to have the ability to generate successful answers to complete the dialog, and uses the following method to do so.

- At the sample question design stage the knowledge engineer has to identify the possible positive and negative answers related to each sample question. The Chatbot using the assigned answers to the already identified question determines the correct positive or negative answer and displays it to the customer.
- The positive or negative identification are made based on the search results of the company database. If it is having matches for the core information searched in the database will result a positive answer, vice versa will result in a negative answer.

4.5.3 Update Probability Based on Frequency

The Chat bot requires some sort of mechanism to identify the most likely sample question first. It is highly unlikely to obtain a large set of training data, to set the initial probabilities of each sample question which is used to identify the customer queries. Hence, the method used to set initial probabilities of the sample questions is as follows.

- Give same probability “1” or a pre decided value for all sample questions. And provide the Chat bot with a mechanism to adjust each questions probability based on each customer queries. If the Chat bot successfully matches a customer query to a sample question it will automatically increase the probability of that sample question by 0.01, making that question’s chances of getting selected next time higher. At the same time the Chatbot will decrease the probabilities of all other sample questions by 0.001 to reduce their chances of selecting next time. With this method we expect to get more and more stable probabilities for sample questions based on customer queries.
- Sample survey results (frequencies of each question type) not used for setting up the initial probabilities, due to small number of people used in the survey.

The above mentioned methods are used to update the sub category probabilities. This way the Chat bot tries to learn from the experience as discussed under learning networks in Chapter 2.

4.6 Main Issues Faced

- The coding language used is Java and only partially familiar with it. Eclipse and Net Beans are IDEs which require putting an effort to learn them.
- Had to refer larger number of documents/research papers to get a good understanding about natural language processing and its techniques.

- Amzi prolog and SWI prolog languages were learnt as AI languages, but due to the new areas, complexities, and new way of coding used in those languages (logic programming) and from time limitations, this was not used for chat bot design and development.
- Statistical techniques were not used for planning and development since they require larger corpuses and training data, which are harder to obtain freely.
- Most of the existing systems were designed for various competitions and therefore very little technical and design information were available.
- Certain high quality research papers were not freely available which resulted in some limitations of the research of existing technologies.

4.7 Summary

Under the design phase we discussed about the design of the proposed system through activity, use case and ER diagrams. Furthermore, we explained how the Chat bot works to achieve its target and how it learns (update its knowledge base) and the problems faced in developing the system.

