

PATIENT'S CHATTING MECHANISM
SERIOUS GAME FOR INTERNAL MEDICINE PRACTICE

LEE SIM SIEW

UNIVERSITI MALAYSIA PAHANG

PATIENT'S CHATTING MECHANISM
SERIOUS GAME FOR INTERNAL MEDICINE PRACTICE

LEE SIM SIEW

THESIS REPORT SUBMITTED IN FULFILMENT OF THE DEGREE OF COMPUTER
SCIENCE

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING

2013

UNIVERSITI MALAYSIA PAHANG
BORANG PENGESAHAN STATUS TESIS

JUDUL:

SESI PENGAJIAN:

SAYA(HURUF BESAR)

Mengaku membenarkan tesis/laporan PSM ini disimpan di Perpustakaan Universiti Malaysia Pahang dengan syarat-syarat kegunaan seperti berikut:

1. Tesis/Laporan adalah hakmilik Universiti Malaysia Pahang.
2. Perpustakaan Universiti Malaysia Pahang dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institut pengajian tinggi.
4. **Sila tandakan (✓)

☐

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972) *

☐

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan) *

☐

TIDAK TERHAD

Disahkan oleh

Alamat tetap :

Penyelia :

.....
.....

.....
.....

Tarikh :

Tarikh :

.....

.....

*Sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis/laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

SUPERVISOR DECLARATION

I hereby declare that I have read this thesis report and in my opinion this thesis report is sufficient in terms of scope and quality for the award of the degree of Bachelor of Computer Science (Graphics & Multimedia Technology) with Honours.

Signature :

Name :

Date :

STUDENT DECLARATION

I hereby declare that the work in this thesis report is my own except for quotations and summaries which have been duly acknowledged.

Signature :

Name :

Date :

ACKNOWLEDGMENTS

I would like to express our deep gratitude to my research supervisor, En. Abbas Saliimi bin Lokman for his guidance, patience and support and useful critiques of this research work. Without his offering to accomplish this research, we would not be able to finish our study at University Malaysia Pahang.

I owe many thanks to my classmate, friends and all some lecturer who always support and give full attention to solve our problem. They always help us in exchanging any ideas and give the enjoyable studying environment. Moreover, I would like to thank all the lecturer, staff and colleagues from five faculties (FKPPT, FIST, FSKKP, FKASA and FKKSA) who were respondents for this study, for their full support and co-operation to make this study possible.

Also, I would like to express my high appreciation to my family for all their support and assistance along the way which cannot be expressed in words. Moreover, I am grateful to other colleagues that help me in my report writing and other related work.

Next, I would like to extend my thanks to my lecturer Rahmah Binti Mokhtar who always gives us guideline and many instructions during the whole project. I also thankful grateful to our parents who always loved us and supported us during this project.

ABSTRACT

This subject deals with an invention towards developing a Artificial Intelligent chatbot for a online serious game in the domain of medical and treatment. It lead to a creation of artificial intelligence chat-box which can communicate with user in human natural language. This robot is formulated to behave as a human patient. The artificial intelligent chatbot, Ill.Bot is being designed as a web-based language processor to analyze our human natural language by using fuzzy logic methods through web scripting. The trick of artificial intelligence is to locate the user input from the existing history or memory and match the response to related input through the use of search and match mechanism. Alongside the development of chatbot, an approaches, One-Match and All-Match Categories - algorithms and techniques had been used to enhance the process of keywords/pattern matching for chatbot.

ABSTRAK

Subjeck ini tertakluk dengan ciptaan ke arah pembangunan chatbot iaitu “Artificial Intelligent” untuk permainan talian serius dalam domain perubatan dan rawatan. Ia menuju kepada penciptaan kecerdasan buatan chat-kotak yang boleh berkomunikasi dengan pengguna dengan penggunaan bahasa manusia. Robot ini dirumus untuk bertindak sebagai pesakit manusia. Buatan pintar chatbot, Ill.Bot direka sebagai robot yang boleh menganalisis bahasa manusia berasaskan bahasa manusia dengan menggunakan kaedah logik kabur melalui web skrip Kecerdasan teknologi ini boleh bertindak balas kepada input dengan merujuk sejarah atau memori yang sedia ada dalam system Di samping itu, dengan pembangunan chatbot, One-Perlawanan dan Kategori-Match -penggunaan algoritma dan teknik digunakan untuk meningkatkan proses kata kunci / corak padanan untuk chatbot.

TABLE OF CONTENT

Page

THESIS STATUS VERIFICATION FORM	i
SUPERVISOR DECLARATION	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
CONTENT	viii
LIST OF TABLES	ix
LIST OF FIGURES	x

Section	Content	Page
1.0	INTRODUCTION	1
1.2	Introduction	1-2
1.3	Problem Statement	2
1.4	Objective	3
1.5	Scope	3
2.0	LITERATURE REVIEW	3-8
3.0	METHODOLOGY	8-9
4.0	DESIGN AND IMPLEMENTATION	10
4.1	Field Study	10
4.2	Previous Study And Case Study	11-12
4.3	Data Collection And Content Development	12-13
4.4	Architectural Model Design	13-15
4.5	Development Of Prototype	16-20
4.6	Testing	21
5.0	RESULTS AND DISCUSSION	22
5.1	Result Analysis	22
5.2	Challenges And Difficulties	23
5.3	Constraints	23
6.0	CONCLUSION	24
	REFERENCES	25-27
	APPENDIXES	28-29

LIST OF TABLES

Table Number		Page
4.1	Chatbot Model Comparison	11
4.2	Database Table Structure	17
4.3	Sample Database	18

LIST OF FIGURES

Figure Number		Page
3.1	SLDC Phase	9
4.1	Patients-Doctor Diagnosis Session Field Study Live Shot	10
4.2	Patient-Doctor Recorded Conversation	12-13
4.3	Chatbot Architecture Design Model	14
4.4	OMAMC Method Flow Chart	15
4.5	Chatbot Prototype Interface Design	16
4.6	Sample Conversations	19-20

1.0 INTRODUCTION

1.1 Introduction

A serious game is a virtual environment designed for education, training or medical application - using innovative design mechanics and technology that engage the participant, promotes learning, and conveys knowledge. It takes advantage of game engines in order to improve the realistic experience of users. Serious games present in many areas of knowledge, including defense, manufacturing, education and medicine, among others. With advances in technologies, medical is now becomes the second only to the military in directing the evolution of serious games.

In virtual world, serious games can provide crisis resource training, with a large variety of cases, in a relatively cheap, readily available environment that provides a viable alternative to expensive simulators. It supplies training environments for disaster situations and mass casualty incidents, including combat care. Serious games allow multiple professionals to train simultaneously on one case and allow one professional to train multiple cases simultaneously. These skills are recognized as critical in reducing medical errors in dynamic high-risk environments, such as the operating room or emergency department. Games need to be designed to fit into residency teaching programs if they are to be used as a way of preventing medical errors. Simulation and serious gaming represent ideal teaching methods to optimize the knowledge and skill of residents before they are entrusted with procedures in real patients. Educators and games designers should develop serious games that train professionals in order to maximize patient safety.

Research into and developments of medical simulators have seen significant enhancements in recent times. A serious game is formally defined as an ‘interactive computer’ application. The use of Human–Computer Interaction (HCI) through such methods as affective computing to portray a mid-to-high level of fidelity is essential. Gaming and in the serious gaming perform most common type of interactions - dialogue system. Typically, this type interaction is done between human and interactive online characters, the use of virtual characters or so-called Non Player Characters through the support of Natural Language Processing (NLP). In order for game characters to be believable, NPCs must appear to possess qualities such as the ability to learn and adapt as well as being able to communicate in natural language.

Natural language processing (NLP) is defined as a field of computer science, artificial intelligence, and linguistics devoted to create computer systems where interaction takes place between computers and humans in the most natural way. It is an area of research and application that explores the ways where computers can be used to understand and manipulate natural language text or speech to perform actions. NLP researchers aim to collect knowledge on how human beings understand and use language, so to develop appropriate tools and techniques which will make computer systems interpret and manipulate natural languages to perform desired tasks. With available technologies, researchers aim to simulate real-world environments by creating interactive dialogue systems.

The objective of our work is to design a conversational agent which is capable of understanding natural language statements in a restricted semantic domain. This feature is intended to allow a natural dialogue, especially in the context of medicine. This conversational agent will simulate a real-like-world communication and interaction for users to experiment real world environment in a medical situation. Therefore, a chatbot named *ill.Bot* is created to fulfill the requirement of this subject.

1.2 Problem Statement

Medical education faces a host of obstacles in coming decades requiring that it rethink the way it delivers medical curricula, especially with regard to critical thinking and differential diagnostics.

Traditional didactic curricula must be coupled with emerging technologies that provide experiential learning without risk to patients while not eroding the clinical effectiveness of advanced medical learners with the help of scenario practice in medical field.

Virtual-world technologies have to advance to a level where they must be considered as a method for delivering medical curricula effectively and safely; moreover, research must establish that such systems are reliable and valid means for delivering medical curricula.

1.3 Objective

The objectives of this research are:

- To come out with a creation of artificial intelligence chatbot which can communicate with user in human natural language.
- To use the algorithm of search and match method to analyze the user input and retrieve relevant responses.
- To make artificial intelligent robot posses the behaviour of a typical human being patient.

1.4 Scope

This research's scopes are:

- Web-based PHP application
- CSS styling interface design
- Using SQL database

2.0 LITERATURE REVIEW

Serious games engage players towards the acquisition of new skills rather than barely entertain them. Typically, serious games are designed for education, training or general counseling and assistance. But currently medical has become the second only to the military in directing the evolution of serious games (Illinios, 1976). This indicates that this application of digital games for training medical professional is on the rise. So-called 'serious' games for training tools that provide a challenging simulated environment, ideal for future surgical training (Dr M. P. Schijven, 2012). What could be more interesting than having communication with games like talking chatting with a real-life human beings? This incident could be happen with the participation of Natural Language Processing (Marilyn Walker, 2011). We present our current research activities associating natural language processing to serious games and virtual worlds by develop a reliable and valid learning platform delivering

medical curricula in virtual space for practitioners which tend to be simulation-based and used for training in medical field.

Natural Language Processing (NLP) enrich users' engagement to the game's world through interactive conversations (Cassell J. et al., 1999). Natural Language Processing (NLP) is an area of research and application that explores the ways where computers can be used to understand and manipulate natural language text or speech to perform actions. NLP researchers aim to collect knowledge on how human beings understand and use language, so to develop appropriate tools and techniques which will make computer systems interpret and manipulate natural languages to perform desired tasks (Gobinda G. Chowdhury, 2011). NLP have been proposed for representing conversational agent beliefs, desires and intentions. The levels of thrust, credibility and cooperativeness have been subject to a study in the last years, in which it can be collaborative or not (Roque, A., and Traum D.R., 2007). Our aim in the project is to provide natural language processing in serious games with conversation capabilities in collaborative environments via dialog where allow junior doctors to experience and train for a variety of acute medical scenarios using computerized mannequins as patients by the simulation of the real-world.

Dialogue systems are programs that can communicate with humans in natural language (Jurafsky, D. & Martin, J. H., 2001). A dialog is a verbal activity which involves at least two interlocutors and is used to accomplish a task in a given communication situation (Franck D., 2012). It is a coordinated sequence of actions (linguistic and non-linguistic) leading to a goal (Vernant, 1992). The system takes speech input and transforms it to a representation that can be understood by the system then the system will produce an appropriate response to the corresponding input (Jenny B., et al, 2007). Researchers believe that intelligent dialogue systems (also called advanced dialogue systems) may constitute a relevant answer to this problem (F. Dernoncourt, 2012). For example, if the business that we are interested in is a pharmacy or hospital, the dialogue between the simulated patient and the player, in this case a medical staff member, will help to get used to have dialogue with the patients and test their knowledge to solve usual, real-life situations that may be urgent and critical, where a mistake can be extremely serious, even fatal. In order to support conversations in serious games, it is necessity to implement a Dialogue Server (Jeremiah S., et al, 2011). Where it models goal-oriented conversations, in which

virtual characters know the type of knowledge the player must acquire to successfully complete a task in the game and the optional information the player should know. Dialogue systems have mainly been designed to collaborate with the user to solve a particular task (Aust, H., Oerder, M., Seide, F. & Steinbiss, V., 1995), i.e. to handle dialogues that (Allen, et al., 2001) refer to as practical. For instance, game players can achieve certain goals or mission in the game, such as, orienting players to find a particular place in a 3D environment, or providing relevant information of missions at each level of the game through human-computer interaction. These systems usually operate in a specific and simple domain, which enables them to perform robustly. In our case study, we integrated this dialogue framework in a medical serious game. We are going to create an application for training doctors and medical students to interact with the virtual patient using natural language by the creation of a more realistic and immersive dialogue with virtual characters in games.

In 1950, mathematician Alan Turing proposed a question - “Can machines think?” (Turing, 2009). Since then, a number of researchers attempt to tackle this question. They attempt to visualize intelligence machine which now called chatbot or chatter robot. Chatbot is a brilliant technology that enables interactions between machine and man using human natural language. First chatbot introduced by Weizenbaum in 1966 (Weizenbaum, 1966), she was named ELIZA. She later became the main inspiration in computer science field for linguistic researchers. A huge breakthrough in chatbot technology then happened in 1995 when Dr. Richard Wallace, an ex-Professor of Carnegie Mellon University created an Artificial Linguistic Internet Computer Entity which named A.L.I.C.E (Wallace, 2009). A.L.I.C.E. later being described as a modern ELIZA. She was the three times winner for Loebner’s annual instantiation of Turing’s Test for machine intelligence (Shah, 2006).

When chatbot technology evolves, in an aspect of managing knowledge-based data (chatbot’s brain), an evolvement in chatbots architecture can be justified. The first chatbot, ELIZA stored and embedded its data directly into application code, while more advanced A.L.I.C.E. uses custom design language, Artificial Intelligence Markup Language or AIML (a derivative of Extensible Markup Language or XML) to manage its datas (Shawar and Atwell, 2007)(Wallace, 2009). With the rapid evolution of Relational Database Model design together with Database Management System

(DBMS) technology, released more advance chatbots. One of the examples is VPbot, a SQL-Based chatbot designed for medical implementation (Ohno-Machado and Weber, 2005). VPbot was developed by Dr.Weber from Harvard University. It is a chatbot that takes advantage on Relational Database Model to store, manage and even use SQL language (database scripting language) to perform keyword and pattern matching process. Latest achievement in chatbot development, a further study has been done by Abbas Saliimawho's master in computer science. He created a chatbot for Diabetic patients in 2009. He then improved and named it as ViDi in "An Architecture Design of Virtual Dietitian (ViDi) for diabetic patients" in the following year. ViDichatbot was developed as an expansion towards E-CARE multimedia content.

Chatbot processing algorithms focus on the first developed chatbot (ELIZA), the most referred and used algorithms (A.L.I.C.E.'s AIML), the modern chatbot with the implementation of Relational Database Model design (VPbot) and the latest improved chatbot, ViDi.

ELIZA, the first chatbot developed by Professor Joseph Weizenbaum who's from Massachusetts Institute of Technology (MIT). It is described as a program that makes conversation with a computer by natural language possible. In ELIZA, input sentence are analyzed by the basis decomposition rules which are triggered by keywords within the input texts. Then, response are generated by reassembling rules associated from selected decomposition rules.

A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) is an artificial intelligence natural language processing chat robot invented by Alan M.Turing in 1950. It can utilizes AIML, an XML language which designed for stimulus-response chat robots. Some view A.L.I.C.E. and AIML as an extension from the old ELIZA psychiatrist program. The comparison is fair regarding the stimulus-response architecture. But the A.L.I.C.E. bot has at present more than 40,000 categories of knowledge, whereas the original ELIZA had only about 200. Another innovation was provided by the web, which enabled natural language sample data collection possible on an unprecedented scale. A.L.I.C.E. won the Loebner Prize, an annual Turing Test, in 2000 and 2001. Although no computer has ever ranked higher than the humans in

the contest, she was ranked “most human computer” by the two panels of judges (Wallace, 2009).

VPbot, a SQL-Based chatbot designed for medical applications. VPbot stores ‘language rules’ into a relational data model. It's having many same features as A.L.I.C.E., but it is often easier to define new language rules in VPbot than that in AIML. Whereas A.L.I.C.E. is designed to produce generic responses to a wide range of topics. While VPbot is best suited to a targeted topic of conversation. The VPbot's algorithm accepts three types of input parameters, a vpid, the current topic, and a sentence. The vpid is a unique identifier for each VPbot instance. The output of VPbot is a new sentence or a new topic. Same with AIML, the output sentence can be dynamically constructed using parts of the input sentence; the database does not have to store every possible response (Ohno-Machado and Weber, 2005).

ViDi was developed as web-based application. Content presented in ViDi is originated from the collected diabetes educational data provided by medical professionals. The data stored in ViDi's knowledge-based storage is temporary as It can be updated or edited by admin from time to time. Several additional approaches have been improved from ELIZA, A.L.I.C.E and VPbot in the development of ViDi, such as extension and prerequisite and OMAMC (One-Match and All-Match Categories) matching methods. The different is that ViDi's All-match keywords had no limit over how many keywords can a single set have (VPbot limitation is three keywords for each single set).

OMAMC technique comprises of two components. 1) Keywords arrangement for matching precedence, and 2) keywords variety for matching flexibility. Describing the fundamental idea of OMAMC, One-match category describes a set of keyword which is build from a single word or a single phrase. While in All-match keywords matching category, keywords are presenting themselves individually which means that there will be many single words or many single phrases. Therefore, All-match keywords can be in a form of combination between single word and phrase, producing either multiple one-words keyword, multiple phrases keyword or both one-word/s and phrase/s keywords in a keywords set.

For both One-match and All-match categories, each keywords set will be stored as a single variable. Therefore, for All-match category which contains multiple

keywords within the same set, the symbol commas (“,”) is needed to separate each keywords. One-match is equal to an exact-match where it has to be exactly the same as the input sentence in terms of word/s and location. While All-match is equalized to a flexible-match where words’ location is a flexible factor. Which means the sequence of the keywords can be differ from the input sentence.

For the matching process, the precedence sequence will start from the keyword that had the most frequent word’s count to the keyword with lowest word’s count. Precedence word-match process is over OMAMC keyword search. While One-match keyword category will has the priority to be processed first then All-match category, as it holds the exact keywords data for matching including the sequence location of the words (exact match).

After comparing the four Chatbot model and their features from the previous studies, ViDichatbot will be selected as the reference model for this thesis and research development. While the techniques considered is OMAMC (One-Match and All-Match Categories) matching method.

3.0 METHODOLOGY

This chapter will cover the details and explanations on the methodology used to complete this project. In order to achieve the objective of the project and accomplish a perfect result, the methodology will be use is System Development Life Cycle (SDLC).

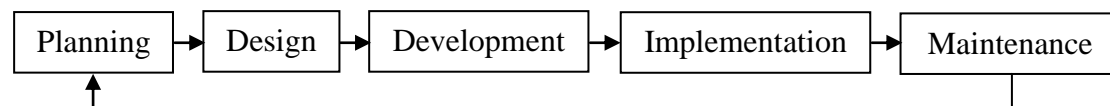


Figure 3.1 SLDC Phase

This project goes through three major steps during development process starting from planning, design and development. These phases had been identified to be the milestones in completing the subject. Each phase will undergo series of tasks that will contribute to the completion of the main task.

To identify all the information and requirements such as hardware and software, planning must be done in the proper manner. The planning phase has two main elements namely data collection and the requirements of hardware and software

In this phase, medical real-time education and management practice had been studied. The field study took place at Hospital Tengku Ampuan Afzan, Kuantan, Pahang, Malaysia. With the help from IIUM medical doctors and hospital staffs, several visits had been done to the patient/s-doctor session in which medical education and management with the real patients take place. Data collection is a stage in any area of study. At this stage I planned about the projects resources and requirements, literature studies and schedule to get more information in this study. All the materials are collected from journal, texts book and research papers gathered from libraries and Internet.

Chatbot is being coded using Hypertext Preprocessor (PHP) programming language together with Asynchronous Javascript + XML (AJAX) technology which contains the technology of Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), XMLHttpRequest (XHR) and Document Object Model (DOM) that being accessed via JavaScript. jQueryjavascript's library also being used for DOM manipulation as well as CSS. As for database, Chatbot uses MySql database with phpMyAdmin as the Database Management System (DBMS). Chatbot had been tested across multiple web browsers including Mozilla Firefox, Google Chrome, and Microsoft Internet Explorer. The hardware used for chatbot's development is Intel Core i5 750 Processor Window 7 Operating System.

4.0 DESIGN AND IMPLEMENTATION

This chapter will discuss on the overall approach and framework of the research. It describes method, technique or approach to be used in this project. It explains on the method used during design and implementation phases.

The following are the considered activities during the research development:

- (i) Field study on medical and treatment education and previous study of prerequisite case study.
- (ii) Design architectural model.
- (iii) Develop prototype.
- (iv) Testing.

4.1 Field Study On Medical And Treatment Education

Medical real-time education and management practice had been studied. The field study took place at Hospital Tengku Ampuan Afzan, Kuantan, Pahang, Malaysia. With the help from IIUM medical doctors and hospital staffs, several visits had been done to the patient/s-doctor session in which medical education and management with the real patients take place.



Figure 4.1 Patients-Doctor Diagnosis Session Field Study Live Shot

4.2 Previous Study On Literature Analysis Of Prerequisite Case Study

	ELIZA	A.L.I.C.E	VPbot	ViDi
Direct/exact relation/s between response/s	No	No	No	Yes
Matching precedence	Whole keyword's database with accordance to keyword's rank or precedence number	1. <that> tag 2. Others (whole keywords' database)	Whole keywords' database (lack of true support of context)	1. Extension response/s 2. Others (whole keywords' database)
Support for same keywords representing different meaning	Yes (same keywords can be used differently according to the Transformation Rules)	Yes (AIML complexity, writing previous utterance for each pattern is a tedious activity)	No	Yes (Extension and Prerequisite)
Possible link for the whole conversation	No (did not draw a path)	No (did not draw a path)	No (did not draw a path)	Yes (draw a path)
Human Working Memory imitation (storing previous utterance)	Not particularly (cannot store the whole utterance)	Yes (can store the whole utterance)	Not particularly (cannot store the whole utterance)	Yes (can store the whole utterance)
Longer keywords effect on matching precedence	No	No	Yes	Yes
Precedence analysis while keywords matching	While matching	While matching	After matching	While matching
Exact-match precedence priority against other matching types	N/A (matching are based on Decomposition Rules)	Highest (same category)	Highest (same category)	Highest (respective category)
Benchmark for stopping matching process if a match is found	Delimiter (comma/period) and/or higher RANK	No benchmark (stop instantly)	No benchmark (stop instantly)	Different category and/or lower word count
Generic keywords support	No	No	Yes (maximum three)	Yes (unlimited)
Response selection/generation precedence if there is a tie (more than one final response)	N/A (response generation is based on Transformation Rules)	1. Exact-match (atomic) 2. Other match (default and recursive)	1. Exact-match 2. Other match 3. Longest keywords	1. Exact-match 2. Other match 3. Longest Keywords 4. Lowest GWP value

Table 4.1 Chatbot Model Comparison

Chatbot processing algorithms focus on the first developed chatbot (ELIZA), the most referred and used algorithms (A.L.I.C.E.'s AIML), the modern chatbot with the implementation of Relational Database Model design (VPbot) and the latest improved chatbot, ViDi.

4.3 Data Collection And Content Development

To proceed this phase, me and my team had visit to Hospital Tengku Ampuan Afzan, Kuantan, Pahang to record some sample conversation of a doctor diagnose a patient. Their conversation has been used as a reference on how a verbal diagnosis is made and what question will normally asked and the possible answers from patient. From this, we can set up the database with potential questions and answers.

Patient: Good afternoon.

Doctor: Good afternoon. Have a seat. So, how are you today?

Patient: Thank you. I'm feeling ill, I've got quite a bad cough, but I don't seem to have a fever.

Doctor: I see. How long have you had these symptoms?

Patient: Oh, I've had the cough for two weeks, but feeling ill just these past few days.

Doctor: Are you having any other problems?

Patient: Well, I've got a headache. I've also had a little bit of diarrhea.

Doctor: Do you produce any phlegm when coughing?

Patient: Sometimes, but it's usually pretty dry.

Doctor: Do you smoke?

Patient: Yes, a few cigarettes a day. Certainly no more than a half a pack a day.

Doctor: How about allergies? Do you have any allergies?

Patient: Not that I'm aware of.

Doctor: Does your head feel stuffy?

Patient: Yes, for the past few days.

Doctor: You should stop smoking, you know?

Patient : OK, doctor. I'll try my best. Thanks.

Doctor : It's all right.

Patient : May I come in, Sir?
Doctor : Yes, you may.
Patient : I've been suffering from fever for the past two days.
Doctor : Did you take any medicine?
Patient : No, I didn't. I've got a cold too.
(Doctor examines the patient.)
Doctor : Your body temperature is 102° F. You must take medicines for at least three days.
Patient : Can I attend office?
Doctor : You shouldn't. You must take rest.
Patient : What about the diet?
Doctor : You should be on liquid diet.
Patient : Should I see you again?
Doctor : You needn't. But if the fever persists for more than three days, come and see me.
Patient : Sure.
Doctor : Here is the prescription.
Patient : Thank you, Sir.
Doctor : You're welcome.

Figure 4.2 Patient-Doctor Recorded Conversation

4.4 Architectural Model Design

Figure shows the schematic representation of chatbot architectural design model. In use case context, actor Users interact with chatbot by providing utterance input while it will react back by providing response output which correspond to that particular input. Actor Authors can add, edit or delete data from knowledge-based database contains of Root-Words, Synonyms, Keywords, and Responses tables data which have to be updated from time-to-time.

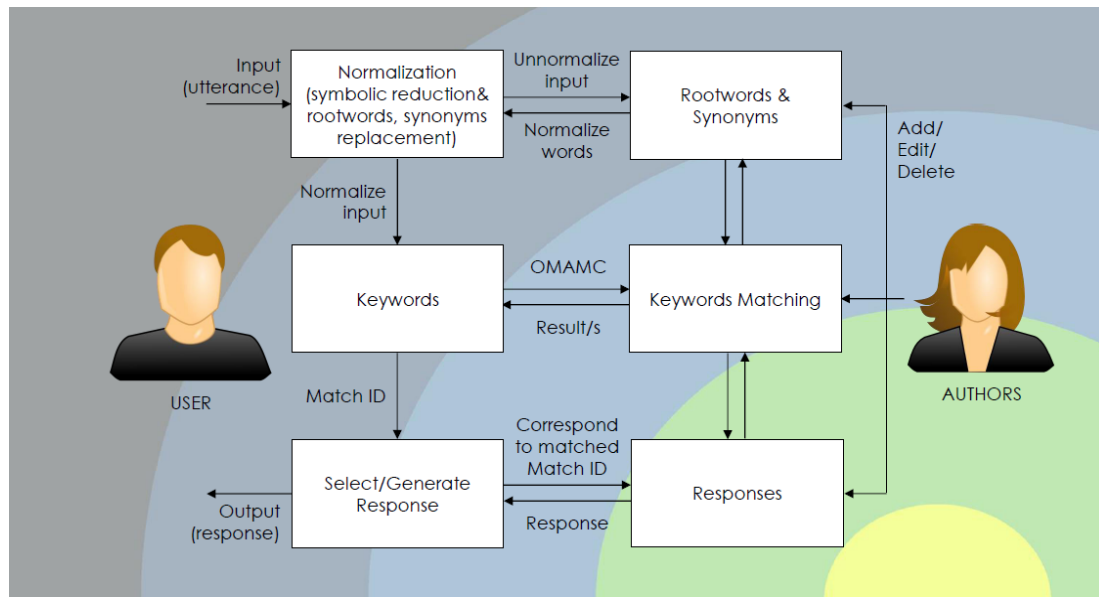


Figure 4.3 Chatbot Architecture Design Model

Generally, it takes about three main processes to generate an output from a user's input. The processes include 1) Normalization process, where input query is being normalized into standard form by removing unwanted symbols and words and synonyms replacements, 2) Matching process, is where input query are being matched with the keywords from the database, and lastly 3) Generate Response process, that correspond to the Responses database reflected according to the matched keywords. There are four tables are involved in this architecture. Rootwords and Synonyms tables to be uses in input normalize process, Keywords table to be used in keyword matching process, and Responses table that store every possible responses.

For the flow of whole process, it started with an user's input, then chatbot will perform Normalization onto the input query while interference with the Rootwords and Synonym table. Normalized words will be sent to Keywords Matching for matching process with the Keywords database table. This process involve OMAMC (One-Match and All-Match Categories) technique. All results have their own Match ID, then it will be sent to Select/Generate Response process where Responses database will execute the response which is associated with the Match ID back to the process as the final output. While in author context, author can perform add, edit, or delete action onto the database system.

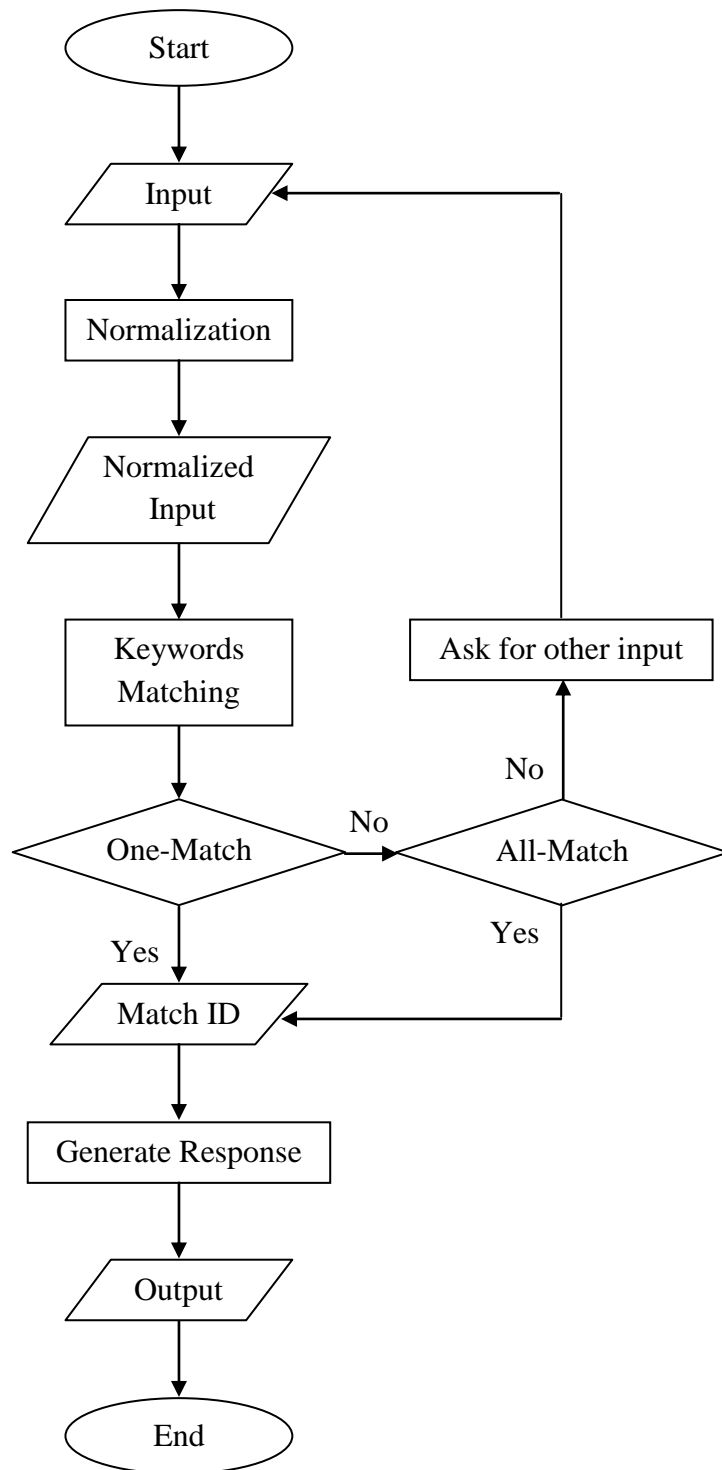


Figure 4.4 OMAMC (One-Match and All-Match Categories) Flow Chart

For OMAMC technique, when the process received the normalized input, it will refer to the One-Match database and search for the exact keywords, if not it will search from All-Match database and get the MatchID of the particular keywords. By holding the ID, it will then generate the correspond respond.

4.5 Development Of Prototype

CSS style sheet language is used to create chatbot interface.

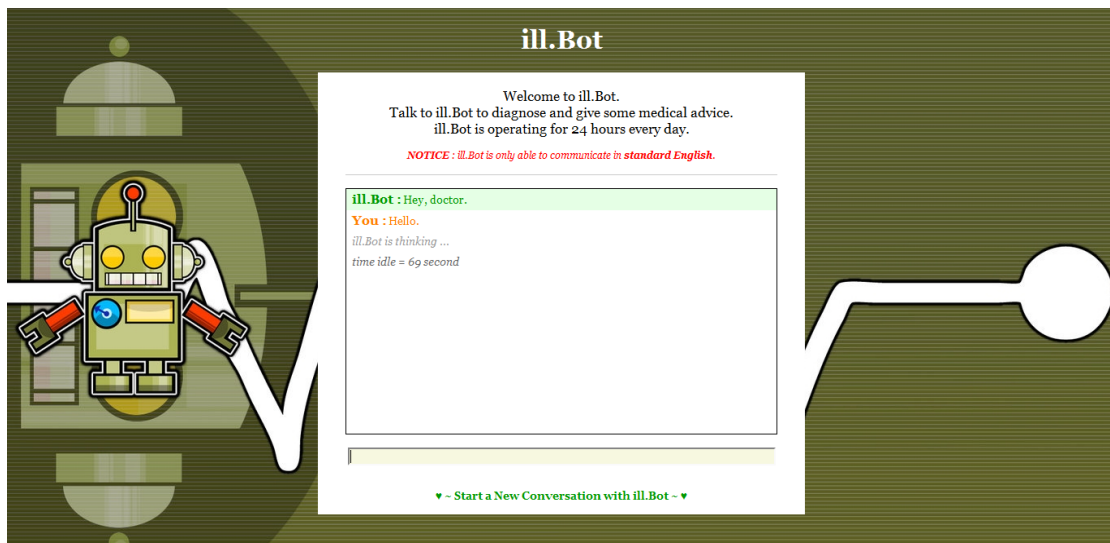


Figure 4.5 Chatbot Prototype Interface Design

Total of four tables in database.

- 1) keywords - table contains keywords for triggering responses.
- 2) responses - table contains potential reaction, reply or answer.
- 3) synonyms - having synonyms words to be swapped.
- 4) rootwords - used in user's input queries normalization process.

rootwords

Name	Data type	Description
word	text	Affixed or root/base word in input queries. Affixed word to be swap with “swap” while root/base word will be stayed with deleted affixation from original word in input queries.
swap	text	Root/base word to be swapped with “word” that is an affixed word in input queries.

synonyms

Name	Data type	Description
input	text	Word/phrase in input queries to be swapped.
swap	text	Word/phrase that will be swapped with “input” data in input queries.
wordCount	integer	Synonyms’ total word’s count (to determine the length of input words to be swap with synonyms data).

keywords

Name	Data type	Description
category	text	Keyword’s category (matched with response’s category).
words	text	Data that represents the keywords set. One-match or allmatch keywords’ types.
matchId	text	ID that matches variable “matchId” in table “responses”.
wordCount	integer	Keywords’ total word’s count (to determine the length of keywords).
allMatch	enum (‘yes’, ‘no’)	‘yes’ for all-match or ‘no’ for one-match keywords’ types.
prerequisite	enum (‘yes’, ‘no’)	‘yes’ for keywords with prerequisite data (matched with response’s prerequisite data).

responses

Name	Data type	Description
category	text	Response’s category. e.g. “diagnose”, “idle”, etc.
words	text	Data that represents the response.
matchId	text	ID that matches variable “matchId” in table “keywords”.
type	enum (‘A’, ‘Q’, ‘S’)	Type of response, ‘A’ for answer, ‘Q’ for question or ‘S’ for statement.
extension	integer	Extension data for response (if any, one or many). Only response type ‘Q’ is eligible to have extension/s.
prerequisite	integer	Prerequisite data for response (if any, one or many).
general	float	General Words Percentage (GWP) value.

Table 4.2

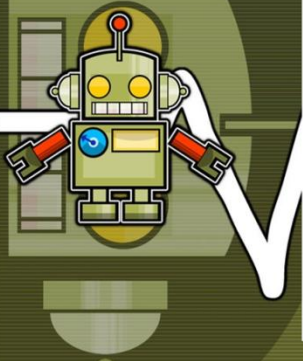
Database Table Structure

rootwords			synonyms			
id	word	swap	id	input	swap	wordCount
67	allergies	allergy	88	constantly	regularly	1
94	breathing	breath	91	drug	medicine	1
88	coughing	cough	71	hey	hello	1
93	difficulty	difficult	85	meal	diet	1
81	eating	eat	83	med	medicine	1
89	examination	exam	92	test	exam	1
78	medication	medicine	6	u	you	1

keywords						
id	category	words	matchId	wordCount	allMatch	prerequisite
470	diagnose	how are you	7	3	no	no
471	diagnose	how, are, you	7	3	yes	no
496	diagnose	how long, symptom	10	2	yes	no
498	diagnose	take, medicine	11	2	yes	no
499	diagnose	phlegm, cough	12	2	yes	no
518	diagnose	last, physical, exam	17	3	yes	no
525	diagnose	how pain	22	2	no	no

responses							
id	category	words	matchId	type	ext.	pre.	gen.
8	idleAns	You are still here.	0	A	0	0	0
66	dntKnow	Sorry. I don't understand @.	0	A	0	0	0
2	greeting	Hello, doctor.	3	A	0	0	0
5	idle	Are you still there?	1	Q	0	0	0
6	idle	I am leaving this conversation since there is no response received.	2	A	0	0	0
275	diagnose	Doctor, I am having a very bad cough and it is making me feel sick.	8	A	0	0	0
262	diagnose	It started in June. So for more than five months now.	10	A	0	0	0
257	diagnose	I had my last physical two years ago.	17	A	0	0	0
276	diagnose	No, I don't have fever but there is a lot of restlessness.	28	A	0	0	0
291	diagnose	It is somewhere near lower abdomen.	32	A	0	0	0

Table 4.3 Sample Database



ill.Bot

Welcome to ill.Bot.
Talk to ill.Bot to diagnose and give some medical advice.
ill.Bot is operating for 24 hours every day.

NOTICE : ill.Bot is only able to communicate in *standard English*.

ill.Bot : Hello, doc.

You : Good afternoon.

ill.Bot : Good afternoon, doctor.

You : How are you today?

ill.Bot : I am sick, doctor.

You : What is wrong with you?

ill.Bot : I've been sick after I fly back from Portland.

ill.Bot : I've got this yellow stuff I keep blowing out my nose.

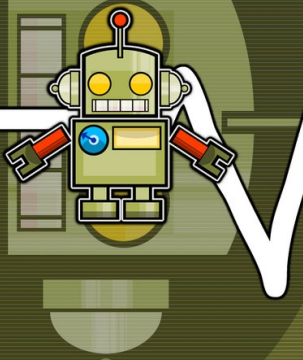
ill.Bot : My sinuses are so stuffed up, I can hardly breathe on one side if my nose.

You are typing ...

time idle = 82 second

How long does the symptom last?

▼ - Start a New Conversation with ill.Bot - ▼



ill.Bot

Welcome to ill.Bot.
Talk to ill.Bot to diagnose and give some medical advice.
ill.Bot is operating for 24 hours every day.

NOTICE : ill.Bot is only able to communicate in *standard English*.

ill.Bot : Hey, doc.

You : What can I help you?

ill.Bot : I have a terrible stomach ache, sir. It makes me can't sleep at night.

You : Where is the pain exactly?

ill.Bot : It is somewhere near lower abdomen.

You : How strong is the pain exactly?

ill.Bot : Can't say. I mean sometimes its severe and sometimes mild.

You : How long does the pain last?

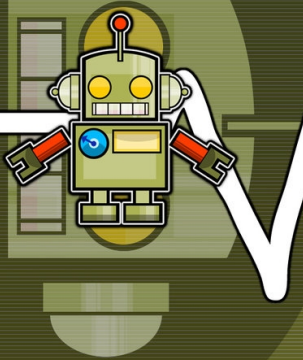
ill.Bot : It comes and goes. Sometimes, I hardly feel anything.

ill.Bot : Other times, it can last up to half an hour or more.

time idle = 82 second

Taip your response here and press Enter

▼ - Start a New Conversation with ill.Bot - ▼



ill.Bot

Welcome to ill.Bot.
Talk to ill.Bot to diagnose and give some medical advice.
ill.Bot is operating for 24 hours every day.

NOTICE : ill.Bot is only able to communicate in *standard English*.

ill.Bot : Hello, doc.

You : How do you feel today?

ill.Bot : Doctor, I am having a very bad cough and it is making me feel sick.

You : Do you have any fever?

ill.Bot : No, I don't have fever but there is a lot of restlessness.

You : Are you passing phlegm while coughing?

ill.Bot : It is most of the time dry cough but sometimes I do pass phlegm.

You : What is the color of the phlegm?

ill.Bot : Green in color.

You are typing ...

time idle = 41 second

Do you have any breathing difficulties?

▼ - Start a New Conversation with ill.Bot - ▼

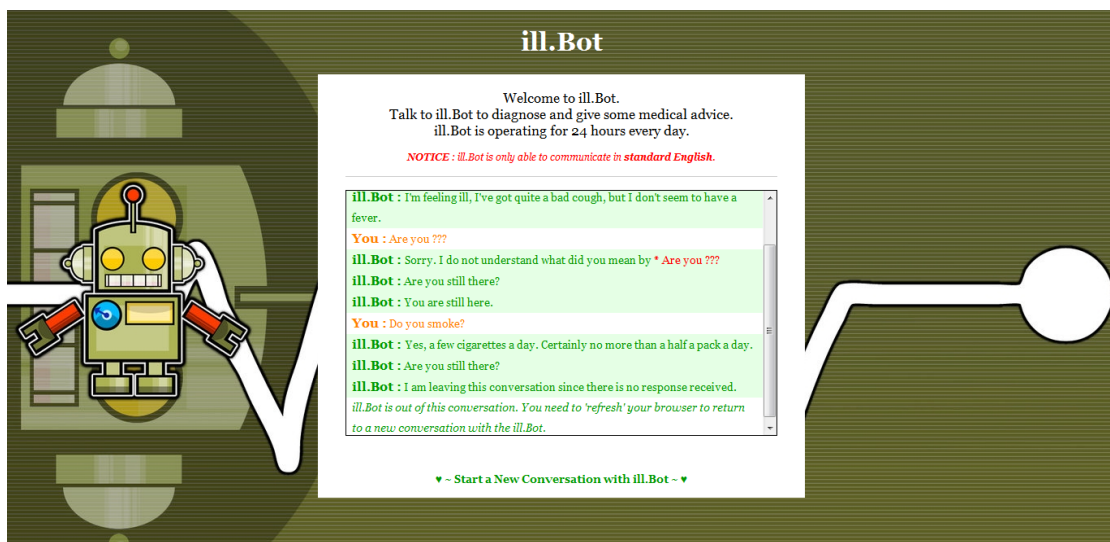
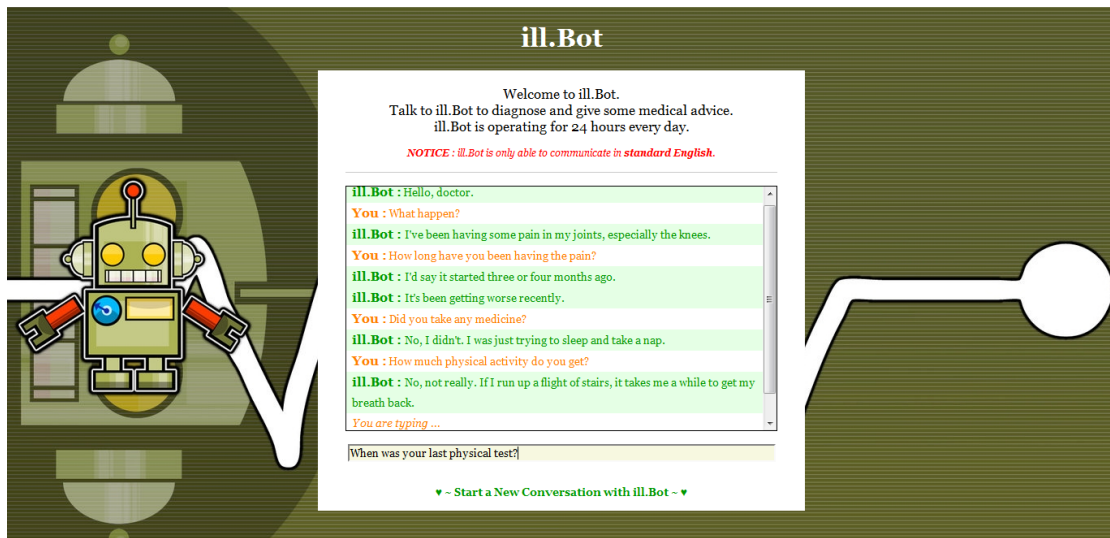


Figure 4.6 Sample Conversations

4.6 Testing

- a. Code-Based Testing - Execute every code at least once. Try to run chatbot on browser every time new code added.
- b. Code Inspection - Analyzing the program logic and its compliance with coding standards. Think of the logic on flow of the data and the preciseness of the designed architecture .
- c. Debugging - Finding and reducing the number of bugs or defects. Solve the issue when error occur. Surf internet for solution.
- d. Mutation Testing - Testing done on the application with purposely added bugs into it in order to detect the more big bugs.
- e. Functional Testing - Testing the features and operational behaviour of the applications. Comply conversation with chatbot.
- f. Dynamic Testing - Giving the input values and checking if the output is as expected. Input utterances with exist keyword data.
- g. Negative Testing - Testing using negative data to discover application's limitation and constraint. Attempt to input unavailable data keyword sentence.
- h. Compatibility Testing - Testing the applications in various elements of a system with which it should operate. Run ill.Bot on different browser such as Google Chrome and Mozilla Firefox.
- i. Alpha Testing - Among public users within the research group. Asking teammate to test for the system.
- j. Public Beta Testing for stable prototypes - Directly adjust and fix any emerging issue or error identified from public's feedbacks. Released new adjusted and fixed prototype straightaway to be beta tested again. Recursive phase until no new issue or error is emerging. Request medical pharmacy students from AIMS University to try for the application.

5.0 RESULTS AND DISCUSSION

5.1 Result Analysis

The developed application, chatbot has met all the objectives of this project, which are :

- To come out with a creation of artificial intelligence chatbot which can communicate with user in human natural language.
- To use the algorithm of search method to analyze the user input and retrieve relevant responses.
- To make Artificial Intelligent robot posses with basic medical and health knowledge.

The artificial intelligent chatbot, ill.Bot is being designed as a web-based language processor to analyze our human natural language by using fuzzy logic methods through web scripting. The robot is develop by using php as the main programming language. It is use for scripting the algorithm and all the main function of the program. While css is use for styling the interface so to make the interface more attractive and most importantly organize the page layout structure. Css file is used so to separate the styling script with the content.

The trick of artificial intelligence is to locate the user input from the existing history or memory and match the response to related input through the use of search and match. Because the human language is enormous and existing the potential of one question could be present in variety ways by a user so the one of the technique used here is break the input sentence structured and search for relevant possible responses, and from there, we sort the potential answer by the number of word match in the input sentence and the response sentence.

To make AI robot a health advisor where act like a doctor, it should posses with the basic medical and health knowledge. The storage of the robot brain is in database, so we have to instill it with general medical and health information. Me and my team have visit to Pahang, Kuantan General Hospital to record the conversation of a doctor diagnose a patient. Their conversation has been use as a reference on how a oral diagnosis is made and what question will normally asked and the possible answers from patient. From this, we can set up the database with potential questions and answers.

5.2 Challenges And Difficulties

The challenge faced through the development process are :

Firstly is defining the structure of the database. Since the database is the brain of the robot, so we have to figured out how should it think. The way how it think is the way how it is going to recall his memory. So all the database table is working like the brain cell is connected to each other and form the database structure.

Secondly, is the algorithm of the of the word search part. We have to understand how the robot is going to analyze an input data. After that, how it knows what to respond back. All these steps have to be script with the proper manner and sequences.

Lastly, is the setting up of the database's data. To set up the memory of the robot, we have to predict what question will be faced by it or else it will not understand about what the user says and return a helpless answers responses. And from the question, we must have to define several answer or response ways so that will not repeating the same sentence all the time, so that to shows it is not like a dead robot but can speak and act like a human.

5.3 Constraints

This research's constraints are:

- (i) Chatbot not capable of supporting true conversational input. It does not address the natural language understanding in its generality since the semantic domain of a game is generally well defined. The types of sentences found in the dialogue are restricted.
- (ii) Parameters as the user's vocabulary, syntax, or speed, and free-form English input must occasionally be supplemented with commands. Lingua franca - English are used as the standard language for all conversation.
- (iii) It does not support multi-users. The application limits the number of different users. Every dialogues only involve two party which are the user and a patient.

6.0 CONCLUSION

In this project, I present to you the creation of a patient chatbot which is able to communicate in human natural language.

Through the creation of chatbot, it extends a proven interaction technique and make it useful for generate game content. The existence of chatbot dialog provide a mechanism for involvement for the user through the process. This chatbot system will be implemented within the platform for serious games of medical practice, in which it will be evaluated through experiments with medical students.

Significant contribution from this research towards computer science field is the development of a dialogue between user and computer leads to potentially very many applications that are not limited to serious games. For example, designing bots, and questions and answers systems, keyword matching technique etc.

In addition, work on conversational agents have many common issues with document analysis, data mining, machine translation and the semantic web: all of these areas represent indirect applications of our work and interactions are considered with some of them.

For the recommendation, is regarding managing knowledge-based data. As known, chatbot needs to have a huge collection of corpus in order to response effectively. As such, managing corpus can be hard if there is no indexing mechanism. Therefore, such mechanism needs to be designed together with the corpus creation (at an early stage).

Instantly, this system presents many opportunities for future enhancement. A goal-driven dialog model with a sensible personality, mood and emotion model will offer obvious advantages over the currently implemented Q&A interaction.

7.0 REFERENCES

- A. Abramczyk, J.B. Lewoc, A. Izworski. (2005). Nurse training case study: e-training of nurses in diabetes problems. *Nurse Education in Practice*, 5, 70-77.
- Abbas S. Lokman and Jasni M. Zain, "Enhancement Algorithms for SQL-Based Chatbot," *Communications in Computer and Information Science*, 2011, Volume 180, Part 5, 470-479, DOI: 10.1007/978-3-642-22191-0_41.
- Abbas S. Lokman and Jasni M. Zain, "One-Match and All-Match Categories for Keywords Matching in Chatbot," *American Journal of Applied Science* 7 (10): 2010, pp. 1212-1218, DOI: 10.3844/ajassp.2010.1406.1411.
- Abbas S. Lokman and Jasni M. Zain, "Extension and Prerequisite: An Algorithm to Enable Relations Between Responses in Chatbot Technology," *Journal of Computer Science* 6 (10): 2010, pp. 1212-1218, DOI: 10.3844/jcssp.2010.1212.1218.
- Abbas S. Lokman and Jasni M. Zain, "Chatbot Enhanced Algorithms: A Case Study on Implementation in Bahasa Malaysia Human Language," *Networked Digital Technologies*, F. Zavoral et al., eds., Springer-Verlag, 2010, ISBN: 978-3-642-14291-8, pp. 31-44, DOI: 10.1007/978-3-642-14292-5_5.
- Abbas S. Lokman and Jasni M. Zain, "An Architectural Design of Virtual Dietitian (ViDi) for diabetic patients," *Proc. 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT 09)*, IEEE Press, August 2009, pp. 408-411, DOI: 10.1109/ICCSIT.2009.5234671.
- Abbas S. Lokman and Jasni M. Zain, "Designing a Chatbot for diabetic patients," *Proceedings of International Conference on Software Engineering and Computer Systems, ICSECS 2009*.
- Alan Baddeley. (1992). Working Memory. *Science*, New Series, 255(5044), 556-559.
- Alan Baddeley. (2003). Working Memory. *Nature Reviews Neuroscience*, 4, 829-839, DOI: 10.1038/nrn1201.
- Caplan, D. & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and brain Sciences*, 22, 77-126.

Chris Evans, Nicola J. Gibbons. (2007). The interactivity effect in multimedia learning. *Computers & Education*, 49, 1147-1160.

Christy, A. and P. Thambidurai. (2008). CTSS: A Tool for Efficient Information Extraction with Soft Matching Rules for Text Mining. *J. Comput. Sci.*, 4, 375-381. DOI: 10.3844/2008.375.381.

Ericsson, K.A. and W. Kintsch. (1995). Long-term working memory. *Psychol. Rev.*, vol. 102(2), 211-245, PMID: 7740089, DOI: 10.1.1.20.2523.

Frazier. (1987). Sentence processing: A tutorial review. In M. Coltheart (ed.), *Attention and Performance XII*. Hillsdale, NJ: Lawrence Erlbaum Associates, 559-586.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76.

Jia, J. (2004). The Study of the Application of a Web-Based Chatbot System on the Teaching of Foreign Languages. *Proc. Society for Information Technology and Teacher Education International Conference (SITE 04)*, AACE Press, USA, 2004, pp. 1201-1207, LINK: <http://www.editlib.org/p/13634>.

King, J. & Just, A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580-602.

Kirschner, P.A., J. Sweller and R.E. Clark. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86, DOI: 10.1207/s15326985ep4102_1.

Ohno-Machado, L. and G.M. Weber. (2005). *Data Representation and Algorithms for Biomedical Informatics Applications*. 1st Edn., Harvard University, Cambridge, MA., USA., ISBN: 0-542- 11988-9, pp: 184.

Pickering, M., Clifton, C. and Crocker, M. W. (2000). *Arcitectures and Mechanisms in Sentence Processing*. *Arcitectures and Mechanisms for Language Processing*, Cambridge University Press, ISBN: 0-521-63121-1.

Raju, S.V. and A.V. Babu. (2007). Parallel Algorithms for String Matching Problem on Single and Two Dimensional Reconfigurable Pipelined Bus Systems. *J. Comput. Sci.*, 3, 754-759. DOI: 10.3844/2007.754.759

Shah, H. (2006). ALICE: An ACE in digitaland. *tripleC*, vol. 4(2), 284-292, ISSN: 1726-670X, LINK: <http://triple-c.at/index.php/tripleC/article/view/46/44>.

Shawar, A.B. and E. Atwell. (2007). Chatbots: Are they really useful. *LDV-Forum Band*, vol. 22, 31-50, DOI: 10.1.1.106.1099.

Shawar, A.B. and E. Atwell. (2005). A chatbot system as a tool to animate a corpus. *ICAME Journal*, 29, 5-24, DOI: 10.1.1.106.1099.

Shawar, A.B. and E. Atwell. (2005). Using corpora in machine-learning chatbots systems. *International Journal of Corpus Linguistics*, 10(4), 489-516. ISSN 1384-6655/E-ISSN 1569-9811.

Susan J. Jones, Gilbert Cockton. (2004). *Tightly Coupling Multimedia with Context: Strategies for Exploiting Multi Modal Learning in Complex Management*

Topics. *Advanced Learning Technologies*, Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04), 813-815.

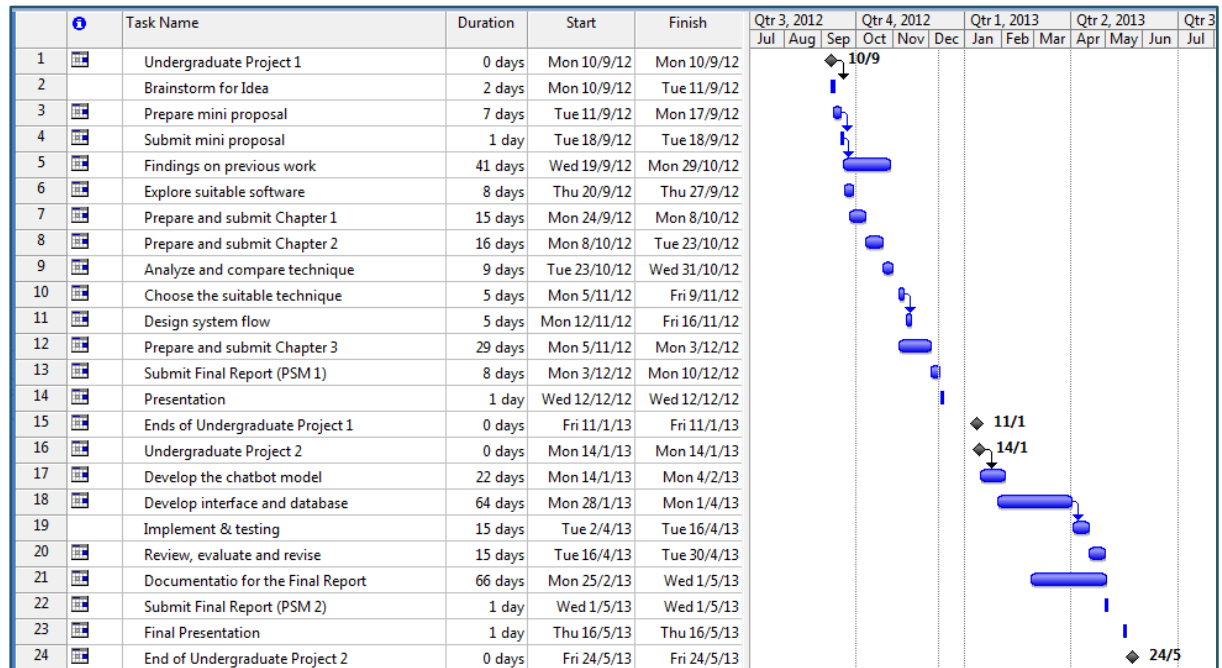
The third National Health and Morbidity Survey (NHMS III). (2006). Retrieved May 4, 2010, from http://www.nih.gov.my/NHMS/abstracts_17.html

Van Eygen Luk, Sunaert Patricia, Feyen Luc, Borgermans Liesbeth, De Maeseneer Jan. (2008). Priorities for diabetes primary care in Europe. *Primary Care Diabetes*, 2, 3-8.

Wallace, R.S. (2009). *The Anatomy of A.L.I.C.E. Parsing the Turing Test*, R. Epstein et al., eds., Springer Science + Business Media B.V., ISBN: 978-1-4020-9624-2, pp. 181-210, DOI: 10.1007/978-1-4020-6710-5_13.

Weizenbaum, J. (1966). *ELIZA- A Computer Program For the Study of Natural Language Communication Between Man And Machine*. *Communication of the ACM*, 9, 36-45, DOI: 10.1145/365153.365168.

APPENDIX A: GRANT CHART



APPENDIX B: PLAGIARISM REPORT

Preferences

turnitin

Originality Report

Processed on: 10-Dec-2012 22:20 MYT

ID: 251895981

Word Count: 5853

Submitted: 1

serious game for
internal medicine
practice (...)

By Siew DeLee

Similarity Index

30%

[What's this?](#)

Similarity by Source

Internet Sources: 19%

Publications: 19%

Student Papers: 7%

Document Viewer