Chatbots: The next generation in computer interfacing – A Review

MS Walgama, B Hettige[#]

Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University, Kandawala Road, Rathmalana,Sri Lanka
[#] budditha@kdu.ac.lk

Abstract – Computer interfacing is now moving towards more machine driven communications than the Command-line or graphical user Interfacing. Human-Computer speech through communication is now gaining momentum as a technique of computer interaction which paved the way for enormous developments using natural language processing. Such systems are precisely designed to simulate how a machine could behave as a conversational partner. A Chatbot is a computer program that stimulates intelligent human conversation using a natural language. This could be done through textual or auditory mechanisms. These Chatbots facilitates easier learning of the domain concepts and their interrelated relationships which make them efficient in the use of general applications. Chatbot architecture integrates a language model and computational algorithms to emulate informal chat communication between a human user and a computer using natural language. Recently, Chatbots techniques are widely used for various practical purposes and made available to the public. The objective of this paper is to review significant conversational agents that have been developed in Sri Lanka as well as in the other parts of the world for various domains over the past years. The development techniques, approaches and functionalities are pointed out through this paper.

Keywords: AIML , Chatbot , Natural Language Processing

I. INTRODUCTION

User interfaces are referred to as the graphical, textual and auditory information that a program parses to a user. After decades of development, at present there exist several types of interfaces. The shift of computation intensive design to presentation intensive design is one of the largest steps in software development. The era of command line interfaces which occurred in late 1960s allowed users to respond using a visual prompt by typing in commands. The MS-DOS prompt application of Windows is one of the widely used. The story became interesting and wide spread with the invention of digital computers. As a result, Graphical User interfaces(GUI) emerged. GUI allows users to interact with devices through graphical icons instead of text- driven commands. Direct manipulation of graphical elements performs actions in a GUI. The next generation of innovation was taken to another level which allows users to interact using voice commands. The latest evolution of user interfaces is conversational agents. These use natural language to communicate with the user for eg Siri, cortana.

In 1950, Alan Turing distributed his well-known article "Computing Machinery and Intelligence" what is presently called the Turing test (Turing A.M, 1950)as a foundation of intelligence. This measure relies upon the capacity of a computer system to imitate a human in a constant composed discussion with a human judge. (French, 1990) In 1966, ELIZA which was developed by Joesph Weizenbaum is recorded as the first ever program to pass the Turing test. ELIZA is considered as a clinical agent which runs on a time-sharing framework which can automatically handle several patients in an hour.

Chatbot architecture integrates language models and computational algorithms (Setiaji and Wibowo, 2016)to emulate informal chat communication between a human user and a computer using natural language. Speech based search engines and assistants such as Cortana by windows, Siri by Apple and Google Chrome are gaining a surge in the commentary context.

The development of chatbots involve extensive knowledge acquisition which is stored into the system to function with user queries. Use of Artificial Intelligence Markup Language (AIML) is widely seen in the development of chatbots. AIML supports dialogs in natural language which is then matched to an input query

to given text pattern. The use of chatbots is visible in a miscellaneous set of applications both online and offline. Chatbots could be used as a tool to learn, to access information in a system, as a question and answering tool for a specific domain and in many different fields like medicine, education, entertainment, automobile or any other.

ALICE, Mitzuki which won the Leobner Prize in 2013, Cleverbot are some of key chatbots while Jabberwacky, Watson and botster some other chatbots used for different purposes. This paper reviews many exiting systems that use this chatbots techniques in their development. The presented systems are from different fields proposing to address different problem domains. The features, functionality and methodology of these systems are also presented in this paper.

This paper gives an overview of the existing chatbot systems including their approaches and functions. The rest of the paper is organized as follows. Section 2 presents about Chatbot fundamental design techniques and approaches. Section 3 reviews existing systems on chatbots. Finally, conclusion is given in section 5.

II. CHATBOT FUNDAMENTAL DESIGN TECHNIQUES AND APPROACHES.

Building a chatbots requires many different techniques to be implemented together. Several key techniques that are widely used in development are also looked in to within this content.

2.1 Artificial intelligent Markup Language (AIML)

AIML is an XML complaint language that is widely used in designing chatbots. This is mainly based on the technology used for ALICE. The goal of AIML language is to direct the process of conversational modeling in to a stimulus response process. AIML characterizes the type of data object whose responsibility is to model conversational patterns. These could be named as the frequent tags and the bases used in the design of AIML chatbots who respond intelligently. Below shows the structure of category, pattern and template object used in AIML.

```
<category>
  <pattern> User Input</pattern>
    <template>
Corresponding Response to input
   </template>
       </category>
```

2.2 *Pattern matching*

Pattern matching is the techniques of checking a given sequence of tokens for the presence of some patterns. Chatbots use this techniques as a common practice and it is frequently seen in question and answering systems. One of the key ways of developing a chatbot is by analyzing the input and finding the best match. Chatbots systems practice these patterns matching techniques mainly because to make the users feel that they are chatting with a real human. ELIZA use a matching keyword and application of transformation rules. In this process, the conditions of all topics are compared to the goal information, the query and the keyword list. As a result, the topics are scored and the highest score out of all is returned as the answer.

2.3 *Parsing*

Parsing is the process which is used in analyzing a string of symbols either in computer languages or natural languages. In computational linguistic parsing is used to analyze a sentence or another set of strings in to its constituents which may contain semantic and other information. This technique includes analyzing the input text and manipulating it by using several NLP functions. One such example is trees in Python NLTK.

2.4 Chat scripts

Chat scripts are often used when the AIML does not provide any matching in its context is a sequence of expect strings. Chat scripts mainly focus on the best syntax out of all the provided syntaxes given to provide a sensible answer to system. Variable concepts, logic and facts are some of the functionalities that chat scripts address.

2.5 ontology

Ontologies are used in chatbots to compute the concepts of synonyms, antonyms, hyponyms and other co concepts that prevail in any natural language

III. EXISTING SYSTEMS

Through further studies for years the researchers could develop various efficient systems with the use of chatbots techniques which is a catalytic section in the field of Natural language processing. This section describes the systems that was developed in the recent past which are applicable for various fields such as medicine, education, entertainment etc.

ALICE (AbuShawar and Atwell, 2015)is one of the earlier developed chatbots. Shawar and Atwell presents an overview of ALICE chatbot, its AIML format and their experiments they used to generate different prototypes of ALICE based on a corpus approach. A description of developed software which converts readable text (corpus) (Shawar and Atwell, 2005)into AIML format is presented alongside with describing the different corpora that is used in this system. A Java program that converts a readable text to the chatbot language model format is developed. The entire program has been divided in to four phases to handle the linguistic annotations and filters. The approach used is a Dialog Diversity Corpus of English along with an after-text reprocessing and filtering using the developed java program. A monologue bilingual corpus was also used in this process.

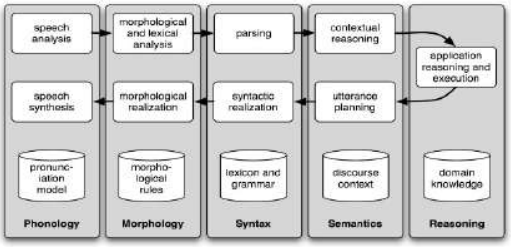

Fig1 -Overview of ALICE chatbot

Shawar, Eric Atwell (Shawar and Atwell, n.d.)present a program developed to convert a machine-readable text (corpus) to a specific chatbot format, which is then used to retrain a chatbot and generate a chat which is closer to human language. Different corpora such as dialogue corpora, monologue corpora such as holy Qur'an (Bayan Abu Shawar and Eric Atwell, n.d.)and FAQ s have been used here. The main goal of this process is the ability to generate different chatbot prototype that spoke different languages.

A java program that converts the corpus to the chatbot language model format has been developed which aims to create ALICE knowledge base automatically and based on specific corpus or domain. To handle the linguistic annotations and fillers, the program is composed of four phases. The system could generate AIML training data for ALICE chatbot to serve in different domains. (Shawar and Atwell, 2003a)

Chatbots could be used in various fields such as education, medicine, and entertainment, business etc. Several applications that could be used in day today life has been developed by the researchers in modern days. The first Sinhala chatbot (Budditha Hettige and Karunananda, 2006)developed by Hettige and Karunananda is a useful system that has been developed to communicate in Sinhala language. This system has been developed as an application of a Sinhala parser that comes under a major component of the project in English to Sinhala machine translation system. The system has been developed using JAVA and SWI-PROLOG supporting both Linux and Windows. As the major components, this system comprises of a Sinhala Morphological analyzer and a Sinhala Language parser. (B. Hettige and Karunananda, 2006)

The Sinhala Morphological analyzer connects with three dictionaries namely, base dictionary, rule dictionary and concept dictionary. Knowledge identification engine reads all the information given from Sinhala language parsing system. It uses simple pattern matching algorithm to identify user input and find the appropriate solutions from knowledge base. This has been developed as an automatically updating system where the knowledge base is updated whenever the users use this. Further it is stated to extend the chatbot to operate on a more specific domain.

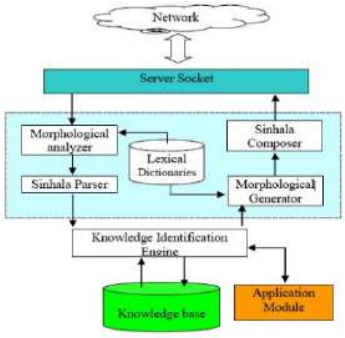

Figure 2 - Overview of the Sinhala chatbot

Abdul-Kader and Woods(Abdul-Kader and Woods, 2015)presents a survey on the techniques used in designing chatbots. As per survey most research work has focused on improving recognition rates of the human voice and the technology is now approaching viability for speech based human computer interaction. Different techniques that could been seen in these applications are compared. Features and functions of the NLTK is also described here.

A responder, classifier and graph master are three main parts in a chatbot. Parsing, pattern matching, AIML(Richard S. W A L L A C E, 2003)chat script, SQL and relational database, Markova chain, language and ontologies are some of the key techniques of chatbots that are described. Review of recent chatbots which are used in different sectors along with their functions are also discussed here. The ranking of several chatbots along with their functions are presented through this.

Shawar and Atwell presents a comparison of two chatbot systems, ALICE and Elizabeth(Shawar and Atwell, 2002),(Shawar and Atwell, 2003)illustrating the dialogue knowledge representation and pattern matching techniques of each. ALICE was found to be better suited for training using dialogue corpora because of its simple patterns templates and simple matching technique. A general description about ALICE chatbot, the AIML data objects, types of categories in ALICE, pattern matching in Alice and its algorithms are described here. ALICE is a software robot or program that you can chat with using natural language. Elizabeth is an adaptation of the Eliza program, in which the various selection, substitution, and phrase storage mechanisms have been enhanced and generalized to increase both flexibility and adaptability.

It is concluded that they decided to train ALICE rather than Elizabeth to learn from human dialogue corpora for two reasons. Firstly, the AIML format is closer to the markup format used in annotated corpora. Secondly, the simplicity in generating patterns/templates, and applying simple pattern matching technique. The conclusion relating to Corpus Linguistics is that the Dialogue Diversity Corpus (DDC) illustrates huge difference in dialogues.

ELIZA (West et al., 1985) by Joseph Weizenbaum is a system that is implemented within the MAC time sharing. ELIZA can converse in natural language. ELIZA analyzes the user provided statement and generate the corresponding response. The input is read and analyzed with the presence of a keyword and when a specific word is matched, the other words will be deleted and the sentence will be transformed according to the associated rules.

ELIZA scripts exist in Welsh and German other than English. One major problem in this system is text manipulation. ELIZA script contains a list of structures. The actual keyword directory is created once the script is read. The performance of ELIZA is significant when the first input is given by the user.

Use of technology in the field of medicine (Vales and Sukhanya, 2009) is stepping forward rapidly. One such initiative is “Pharmabot”(Comendador et al., 2015), which is defined as a pediatric generic medicine consultant chatbot. Developed by Comendador , this Pharmabot will converse in order to prescribe, suggest and provide information on medicine for children. A left to right parsing algorithm is used in the development. Details about some chatbots such as Erica which is developed for dental practices, ELIZA which stimulates a psychotherapist and about PARRY are described in this paper.

Visual C# has been used in the development and is designed to run as a standalone system. The user will be asked several questions there by the properly answered ones will be directed to the parsing algorithm. As a significant fact of the system this provides a dictionary database which contains technical and medical terms for any novice user. The calculation which has been done through a weighted mean method depicts that this system function and generate the same results that is expected from a manual system.

Another application of chatbots in the medical field is MedChatBot(Kazi et al., 2012). This is based on AIML and Unified Medical Language System (UMLS) based. UMLS contains around two million medical concepts including variety of medical domains (Webber, 2005) and about 135 semantic types.

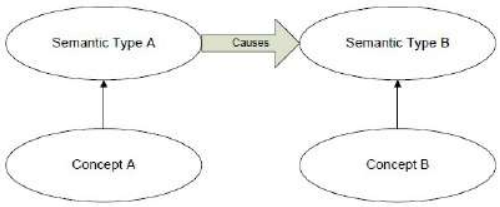

Figure 3 - Relationships between Semantic Types of Concepts in MedChatBot

This use SQL queries and a knowledge base to return the result in natural dialog. Chatterbean, the open source java based AIML interpreter has been used in this system for its further development. There are three major components in the architecture of the MedChatBot namely the front end, the AIML parser and the database. The user inputs the question.

One of the limitations of the system is that the at times the system will be unable to produce accurate responses to the queries due to lack of casual relationship between concepts of UIML. 150 queries were gathered from five random students accounting to 30 from each. Then these collected queries have been tested by expert and the system has generated results at a higher rate.

“Octopus” (B Hettige and AS Karunananda, 2015)which is a multi-agent chatbot is another application of chatbot which use Sinhala language. This is structured with 8 sub multi agent systems namely core system, GUI system, Natural Language Processing system, communication

system, learning system, action system, searching system and data access system. The development is done through the framework MasMT which is a multi-agent framework. Octopus functions through text input which also provides action searching facilities such as to execute commands, open or close applications. This has been developed as a java application running on Windows and Linux. The current version of Octopus is said to have limited capability in semantic processing. Improving such aspects and passing the turning test is stated as future work of the system.

Bradeško and Mladenić (Bradeško and Mladenić, 2012)compare the technologies of the chatbots that have won the Loebner Prize. The technologies and approaches of those chatbots are presented in this paper. The Loebner Prize Competition is an annual competition held for chatbots to check its capabilities, through Turing test method. Chatbots like ALICE, CLEVERBOT, Suzette, Rosette are some of the award winning chatbots (Konferenca Jezikovne tehnologije et al., 2012) The technologies, language tricks, technical approaches and their respective algorithms are presented through this paper. Pattern matching, Markova chain models, parsing, ontologies and chat scripts of different chatbots are also explained through this paper.

Ontologies such as OpenCyc (Lenat, 1995)have been widely used in chatbots. ALICE's main technology, AIML is most often used in many other chatbots too. Non-sequitur, simulation of keystrokes and canned response are some of the language tricks described here which are used by many intelligent chatbots.

Using chatbots in the field of education can assist people in many ways. One such application is using Bots as Language learning tools. Wang and Petrina (Wang and Petrina, 2013)in their paper presents how to predict and advise the design of a language tutor called LUCY using learning analytics. Further the paper describes student learning methods, data trails, chat log architecture which could be useful in designing more sophisticated language learning bots.

Lucy comes in two forms as a commercial chatbot and as an intelligent chatbot. Commercial chatbot Lucy is introduced as a digital language tutor to carry extensive conservation with learners as they speak. Lucy provides the users with a feedback regarding their pronunciation and provides exercises. Hosted on Pandorabots, Lucy is an online tool which guides learning in reviewing grammar and vocabulary (Jørgensen and Phillips, 2011).This chatbot Lucy provides several modules such as travel English, hotel English, restaurant English, causal talks and helping visitors. The paper describes the importance of applying learner analytics for understanding the design of the intelligent chatbots.

Even though the chatbots are designed in a specific procedure, through research it is found that the linguistic style of chatbots is altered over time(Ali et al., 2012). The study done by Ali shows that chatbots depicts a certain behavioral drift in their styles. Alice, CleverBot, Hal, Jeeney, SkyNet, TalkBot, Alan, MyBot, Jabberwock, Jabberwacky, and Suzette have been used for the experiment. The data for the current study has been collected over years and contains only the chatbots that were used in their previous study about "Evaluation of authorship attribution software on a Chat bot corpus."(Ali et al., 2011) The experiment has been conducted through a model built for authorship identification called RapidMiner.

The confidence value of the chatbot has been taken as the key parameter. Jabberwacky had shown more positive results compared to the others. These experiments had several variations in the styles while some styles were steady. As stated the reason could be their intelligent algorithms. Performing additional research on these chatbots are focused as future work in this.

Chai(Chai et al., 2001)provides an effective solution for information access through a web based natural language dialog system. This system allows users to find relevant products on e-commerce sites. The main aim of this system is to provide a better experience for the users who are frustrated by menu driven navigation and keyword search. The dialog system is built on a traditional rule based technology.

The architecture is a support mixed initiative dialogue with multiple modalities. The natural language assistant (NLA) in this system uses a hub and spoke architecture. The authors have chosen C for the implementation of the statistical parser due to its efficiency. The system reports that the average number of clicks has been reduced by 63%. Defining quantitative and objective measures of system's success is stated as further work by the authors.

At present chatbots are developed for various domains for various languages. Poongkuzhaki(T.Kalaiyarasi et al ,2003) is an intelligent chatterbot developed for Tamil language by T.Kalaiyarasi. The main function is to produce responses to given inputs for any existing topic. If the user is in a pause the system will initiate the conversation. Identifying key words in Tamil and discovering the

minimal context in the question is stated as the prominent technical issues faced. The means to overcome such issues are also addressed here. The time complexity of the system is of order n where n is the input word count. Voice enabling and advanced grammar Handling is mentioned as future work of this project.

IV. CONCLUSION

This paper reviewed a wider range of applications where chatbots can be used. After various advance studies chatbots are now used in fields like education, health care, language learning etc. Through a proper study of the techniques and tools that prevail researches can develop efficient and useful systems to the mankind. Rule based approaches or pattern matching techniques could be used to develop a successful system. There is clearly an increasing trend matching towards to develop wide range resources and tools to be used in this field. This paper presented unified overview of some selected papers which involve techniques and approaches in their development. Using such approaches along with theories of implementing chatbots could be used to develop numerous systems to assist the world.

REFERENCES

Abdul-Kader, S.A., Woods, J., 2015. Survey on chatbot design techniques in speech conversation systems. Int J Adv Comput Sci ApplIJACSA 6.

AbuShawar, B., Atwell, E., 2015. ALICE Chatbot: Trials and Outputs. Comput. Sist. 19. doi:10.13053/cys-19-4-2326

ALICE(2016),<http://alice.pandorabots.com>

Ali, N., Hindi, M., Yampolskiy, R.V., 2011. Evaluation of authorship attribution software on a Chat bot corpus, in: Information, Communication and Automation Technologies (ICAT), 2011 XXIII International Symposium on. IEEE, pp. 1–6.

Ali, N., Schaeffer, D., Yampolskiy, R.V., 2012. Linguistic

Profiling and Behavioral Drift in Chat Bots., in: MAICS. pp. 27–30.

B Hettige, AS Karunananda, 2015. Octopus: A Multi Agent Chatbot, in: 8th International Research Conference, KDU, Published November 2015.

Bayan Abu Shawar, Eric Atwell, n.d. An arabic chatbot giving answers from the Quran.

Bradeško, L., Mladenić, D., 2012. A survey of chatbot systems through a loebner prize competition, in: Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies. pp. 34–37.

Chai, J.Y., Budzikowska, M., Horvath, V., Nicolov, N., Kambhatla, N., Zadrozny, W., 2001. Natural Language Sales Assistant-A Web-Based Dialog System for Online Sales., in: IAAI. pp. 19–26.

Comendador, B.E.V., Francisco, B.M.B., Medenilla, J.S., Nacion, S.M.T., Serac, T.B.E., 2015. Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. J. Autom. Control Eng. 3, 137–140. doi:10.12720/joace.3.2.137-140

Eliza(2016),<http://www.manifestation.com/neurotoys/eliza.php3>

Filho, H. P. (2009). ChatterBean: Flexible Alicebot. http://www.geocities.com/phelio/chatterbean/

French, R.M., 1990. Subcognition and the limits of the TuringTest. Mind 99, 53–65.

Hettige, B., Karunananda, A.S., 2006. First Sinhala chatbot in action. Proc. 3rd Annu. Sess. Sri Lanka Assoc. Artif. Intell. SLAAI Univ. Moratuwa.

Hettige, B., Karunananda, A.S., 2006. A Parser for Sinhala Language - First Step Towards English to Sinhala Machine Translation. IEEE, pp. 583–587. doi:10.1109/ICIIS.2006.365795

Hettige, B., Karunananda, A.S., 2015. Octopus:A multi agent chatbot. Proc. 8th International research confernece , KDU

Jabberwacky. (2016). Jabberwacky-live chat bot-AI Artificial Intelligence chatbot.< http://www.jabberwacky.com/>

Jørgensen, M., Phillips, L., 2011. Discourse analysis as theory and method. Sage, Los Angeles, Calif.

Kazi, H., Chowdhry, B.S., Memon, Z., 2012. MedChatBot: An UMLS based Chatbot for Medical Students. Int. J. Comput. Appl. 55.

Lenat, D.B., 1995. CYC: A large-scale investment in knowledge infrastructure. Commun. ACM 38, 33–38.

Loebner prize". <http://www.loebner.net/Prizef/loebner-prize.html lists>

Mitsuku(2016),http://http://mitsuku.com/

Richard S. W A L L A C E, 2003. The Elements of AIML Style.

Setiaji, B., Wibowo, F.W., 2016. Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling. IEEE, pp. 72–77. doi:10.1109/ISMS.2016.53

Shawar, A., Atwell, E.S., 2005. A chatbot system as a tool to animate a corpus. ICAME J. Int. Comput. Arch. Mod. Mediev. Engl. J. 29, 5–24.

Shawar, B.A., Atwell, E., 2003a. Using the Corpus of Spoken Afrikaans to generate an Afrikaans chatbot. South. Afr. Linguist. Appl. Lang. Stud. 21, 283–294.

Shawar, B.A., Atwell, E., 2003. Using dialogue corpora to train a chatbot, in: Proceedings of the Corpus Linguistics 2003 Conference. pp. 681–690.

Shawar, B.A., Atwell, E., 2002. A comparison between ALICE and Elizabeth chatbot systems. University of Leeds, School of Computing research report 2002.19.

Shawar, B.A., Atwell, E., n.d. CHATBOTS: CAN THEY SERVE AS NATURAL LANGUAGE INTERFACES TO QA CORPUS?, in: Proceedings of the Sixth IASTED Internationa Conference. p. 183.

Turing A.M, 1950. Computing Machinery and Intelligence.

T.Kalaiyarasi, Ranjani Parthasarathi, T.V Geetha, 2003. Poongkuzhaki- An intelligent tamil chatterbot.

Vales ,A. and Sukhanya, K “Personal healthcare assistant/companion in virtual world, ”Department of Electrical Engineering and Computer Science, Los Angeles, California, 2009.

Wang, Y.F., Petrina, S., 2013. Using learning analytics to understand the design of an intelligent language tutor–Chatbot lucy. Editor. Pref. 4.

Webber, G. M. (2005). Data representation and algorithms for biomedical informatics applications. PhD thesis, Harvard University.

West, L.H., Fensham, P.J., Garrard, J.E., 1985. Describing the cognitive structures of learners following instruction in chemistry. Cogn. Struct. Concept. Change 29–49.