SindhUniv. Res. Jour. (Sci. Ser.) Vol.50 (001) 153-158 (2018)
27http://doi.org/10.26692/sujo/2018.01.00
SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Artificial Intelligence Mark-up Language Based Written and Spoken Academic Chatbots using Natural Language Processing

ADNAN. A. ARAIN, A. MANZOOR*, K. BROHI**, K. HASEEB***, I. A. HALEPOTO****, I. A. KOREJO**

Department of Telecommunication Engineering Systems Quaid-e-Awam University Nawabshah

Received 24th October 2017 and Revised 25th February 2018

Abstract: Chatbot is an intelligent conversational system that interacts with humans via natural languages, where natural language processing utilizes computational methods to learn, understand, and generate human language contents. The scope of this paper is restricted towards Chatbot. Accessing the required and right type of information from a website is a time-consuming process. In this paper a particular focus is placed on the academic organizations. Sometime an employee or an academic administration staff spends several hours to reply a query. To address this issue, this paper proposes two forms of Chatbots i.e. written and spoken. They provide user interaction and consultation regarding admission process, and other related information pertaining to an academy. This would be a quick way to acquire information without the need to traverse the entire website and is also applicable to any academic or educational websites, such as schools, colleges, universities, or any institution. Both Chatbots are implemented using artificial intelligence mark-up language (AIML).

Keywords: Artificial intelligence, computational linguistics, natural language processing, human computer interaction, chatbot.

1. INTRODUCTION

Natural language processing (NLP), or computational linguistic (CL), is an interesting domain of research in computer science that carries a unique perspective on fundamental topics in cognitive science (CS), artificial intelligence (AI), and human computer interaction (HCI) (Figure 1). NLP can be defined as a hypothetically inspired variety of computational approaches for analysing and signifying naturally-occurring texts at one or more phases of linguistic analysis for the purpose of accomplishing human-like language processing for a variety of applications". NLP facilitates the interaction between human and computers via oral (dialogue) and written (monologue) of languages such as English, Urdu, German, Hindi, French, Malay, and Italian, among others.

NLP consists of two main areas, which are (i) natural language generation, and (ii) natural language understanding. One of the promises of natural language processing is to enable human-machine interaction for a variety of tasks. A Chatbot can address variety problems in natural language processing for instance they can provide a person social and emotional support, they can also assist customers on a B2C website e.g. to track a product, a Chatbot can even help you find a better health insurance plan for your next insurance. The concept of phases of linguistic analysis refers to the point that there are many linguistic processing that are known to require effort when creating or understanding languages. It is assumed that people usually use all of these phases, since each phase expresses dissimilar meanings. However, numerous NLP systems use different phases or mixtures of phases for linguistic analysis, which manifests itself amongst numerous NLP tasks.

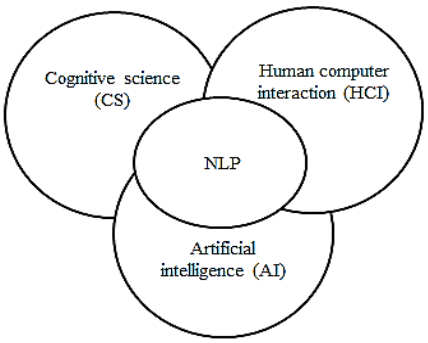

Fig. 1: Natural language processing (NLP).

To understand natural languages and extract meaning from monologues or dialogues, it is significant to distinguish between the different phases of languages or knowledge (Fig. 2), such as (a) Phonetics or phonology phase, dealing with pronunciation or concerns how words are realized as sound, and automatic speech understanding systems. (b) The ecology phase, dealing with language conventions for punctuation text markup and encoding. (c) The

++Corresponding adnan@quest.edu.pk
*Department of Information TechnologyQuaid-e-Awam University Nawabshah
**Department of Computer Science University of Sindh Jamshoro
***Department of Computer Science Islamia College University Peshawar
***Department of Computer Systems Engineering Quaid-e-Awam University Nawabshah

morphology phase, dealing with the meaningful component of the word or are concerned with how words are created from meaningful units (also named mophemes). (d) The syntax phase deals with the structural relationship between the words or the fact on how words can be combined to form sentences that are correct in the language. (e) The semantic phase deals with the meaning of the words or sentences. This level is subdivided into (i) lexical semantics concerning the mopheme meaning, and (ii) compositional semantics concerning the sentence meaning. (f) The pragmatic phase concerning how sentences are used in dissimilar contexts and how it affects the interpretation of the sentence. (g) The discourse phase deals with the conventions of dialogue or knowledge coming from the outside world. Global knowledge contains the universal knowledge about the structure of the world that people use to keep a conversation going.

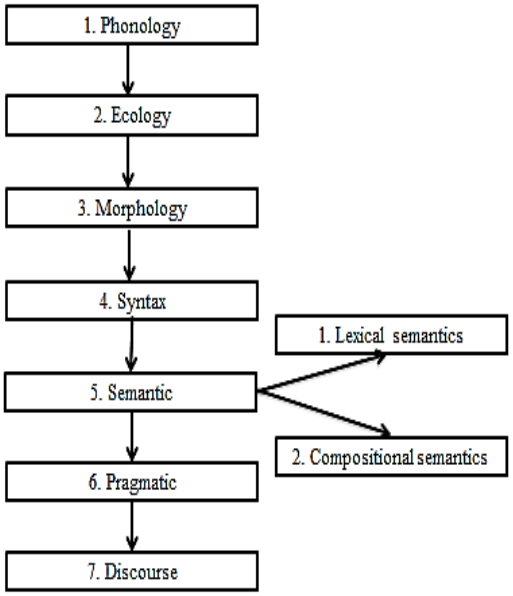

Fig. 2: Phases of knowledge or language.

As shown in Fig. 3, language flows from the parsing stage towards world knowledge. Parsing comprises of the analysis of the syntactic structure of sentences, and also determines that a sentence follows the syntactic rules of the language. The output of this stage is a parse tree. Semantic interpretation includes the production of a depiction (conceptual graph and frames or other knowledge representation technique) of the meaning of a sentence, whereas the incorporation of global knowledge contains the generation of an extended representation of a sentence's meaning to completely understand the meaning of that sentence, upon which the output will be utilized by the application system. NLP delivers the concepts for multiple applications, with the ones utilizing it the most being listed in Table 1. uncertainties in the project schedule in the research presented in this article.

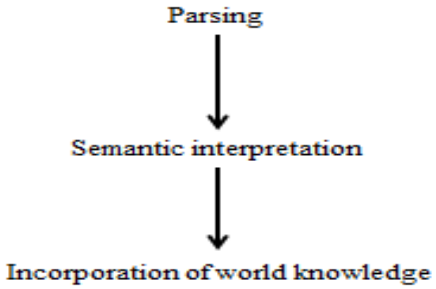

Fig. 3: Flow of language analysis.

Table 1: Most frequent applications utilizing NLP

Applications of NLP	
1	Information extraction
2	Question answering
3	Automatic summarization
4	Machine translation
5	Dialogue system
6	Speech recognition
7	Word segmentation
8	Parts of speech tagging
9	Parsing
10	Human computer interaction (HCI)
11	Spelling and grammar checking optical character recognition (OCR)
12	Natural language interfaces to database email understanding

Human computer interaction represents an application of NLP technology, where designers enable human interaction with computers via NLP. A significant progress in this realm is the Chatbot (chatterbot); a conversational agent that uses natural languages to interact with people. Generally, the communication with users/administration and accessing required information is time consuming, as it requires an employee or an academic administration staff to spend several hours obtaining the appropriate information. However, this work proposes two Chatbots i.e. written and spoken that is responsible for user interaction, consultation regarding the admission process and other information pertaining to academy, provide an applicable website association if the users disagree with the answers. This would make information acquisition a lot faster and applicable to all

academic or educational websites such as schools, colleges, and universities.

Background

A substantial amount of research has been done in this domain, involving marketing, entertainment, and education. The notions of a Chatbot is not limited to any single use, but normally work in multiple sectors; some predetermined event triggers the bot, and the bot responds by writing out a suitable message that is nominated by its artificial intelligence (Shawar and Atwell, 2007). Through SOFIA Chatbot (Duangnamol, Suntisrivaraporn, Supnithi, and Ikeda, n.d.), student can solve mathematical expressions (general). (Watanabe, *et al.,* 2005) develop a knowledge base for a QA system that answers "how" type questions. Huang and Zhou (Huang, *et al.,* 2007) proposed an approach for extracting Chatbot knowledge from online discussion forums. Wang (Wang, 2008) designed Chatbot interfaces for language learning: ethnographic research in to affect and users' experiences. Deryugina (Deryugina, 2010) developed Chatbots as tutoring systems. Several programming clubs have proposed coding Chatbots as an approach to increase their respective students' interest in computer science. Miroslawa Lasek (Lasek and Jessa, 2013) developed Chatbots for customer service on hotel websites. Shaw (Shaw, 2012)used Chatbot in computer science courses to explain computing principles.. Keegan (Keegan, *et al.,* 2012) presented Chatbot (Turi) for schools. Huang and Rebedea (Huang *et al.,* 2007) designed a Chatbot that simulates a historical figure. Kowalski and Goldstein (Kowalski, *et al.,* 2009) developed two case studies for security training using (Gunning and Forslund, 2013) developed a self-learning dialog system that has the ability to interpret natural languages, and they use Twitter® as a source of knowledge to generate responses. Augello and Pilato (Augello, *et al.,* 2014) described the different Chatbotarchitectures and exploited the ontologies in order to create clever information suppliers and overcome the main limitations of Chatbots. Shaw (Shaw, 2012) proposed a framework of sentence describing rules to get answers while chatting with a Chatbot using an "answer matching" strategy.

2. IMPLEMENTATION OF CHATBOTS

These days a number of open source APIs exist which can be used to implement a Chatbot, moreover it can also be integrated with an existing chatting or messaging platform. Once it is integrated with an existing platform its feasibility can be tested by conducting an experiment. A significant part of any Chatbot is its knowledge base, or database. It needs to be equipped with vast amounts of phrases, both as cues to contest the user input, and as appropriate answers to each cue. Many works desire to analyse a Chatbot's knowledge base and construct a Chatbot with character, with have knowledge associated to its intended use with enough answers to simulate a real discussion. Both written and spoken chatbots were implemented using AIML. To construct the AIML documents for the knowledge base, the data was obtained from admission staffs and their respective websites. AIML comprises of data objects named AIML objects made up of units called topics and categories (**Table 2).** Topics are optional, while categories are the basic entities of information in AIML (Wallace, *et al.,* 2003). Each category is a rule for converting an input (matching) to an output, consisting of a pattern that matches the operator input, and a template to generate the Chatbots' answer. The AIML interpreter attempts word by word matching to acquire the best response.

Table 2: Format of AIML

```
<aiml version="1.0">
<topic name="the topic">
<category>
<pattern>PATTERN</pattern>
<that>THAT</that>
<template>Template</template>
</category>
....
....
</topic>
</aiml>
```

There are three different forms of categories available such as recursive, default, and atomic (Shabaz, *et al.,* 2015) **(Table 3-4).** Default categories configuration have wildcard signs *, or _, whereas atomic categories configuration lack these signs. The wildcard signs match any inputs; nevertheless, it can vary in their arranged alphabetical orders. The recursive categories templates have recursive tags, such as <srai> and <sr>. Typically, response from all Chatbots depends on the linguistic data being stored and the used algorithm for pattern matching. The system looks for a lexical matching between the user query and the query answer units stored in their data base. The execution of such schemes varies from using keyword matching, string similarity, or complex natural language processing techniques. Prior to the matching procedure, normalization method is applied to eliminate punctuation; the input is subdivided and transformed to uppercase letters. Post normalization, an AIML interpreter generates the longest pattern match (es) by matching them word by word. When a contest is found and the procedure ends, the template that belongs to that type is processed by the cryptographer to produce an output.

Table 3: Kinds of AIML categories.

Kinds of AIML categories	Atomic	<category><pattern>hello </pattern> <template>hello </template></category> If the user inputs 'hello', then BOT replies 'hello'.
	Default	<category><pattern> HELLO* </pattern> <template>hello friend </template></category> If user input 'hello robot', then BOT replies 'hello friend'.

Table 4: Kinds of AIML recursive categories.

Recursive AIML categories		
Symbolic reduction	Divide and conquer	Synonyms
<category> <pattern>DO YOU KNOW WHAT THE * IS</pattern> <template><srai>What is <star/></srai> </template> </category> <srai> is used to reduce the input to simpler form "what is *".	<category> <pattern>YES*</pattern> <template> <srai>YES</srai><sr/> <template></category> The input is partitioned into two parts, "yes" and the second part; * is matched with the <sr/> tag. <sr/> = <srai><star/></srai>	<category> <pattern>HI</pattern> <template> <srai>Hello</srai> </template></category> The input is mapped to another form, which has the same meaning.

3. CONCLUSION

One of the difficult applications of NLP is the creation of a dialog system or conversational agent. These days the advancements in the field of machine learning have led to the development of various algorithms and also the availability of large corpora would enable us to create better conversational systems or Chatbots. This paper presents the human computer interaction models i.e. written and spoken Chatbots. These Chatbots are implemented by AIML, with both having similar knowledge bases. The written Chatbot interacts via text, whereas the spoken Chatbot uses a virtual agent that verbally interacts with users and provides session related to admission information of academy via the pattern matching technique. The implemented Chatbots used only one communication medium, which is English, however, in the future, Chatbots will be developed to have the capabilities of utilizing more than one communication mediums at a time. In the future work other possibility of creating a Chatbot would be explored, for example these days a number of open source NLP platforms have been introduced which can be used to implement a Chatbot. Moreover, the Chatbot can also be integrated with a messaging app which would provide an opportunity to use this Chatbot in a study. The objective of this kind of study could be to provide customer support or to provide a social support to a person who is in a need.

REFERENCES:

Augello, A., G. Pilato, G. Vassallo, and S. Gaglio, (2014). Chatbots as interface to ontologies. In Advances onto the Internet of Things (pp. 285–299). Springer.

Deryugina, O. V. (2010). Chatterbots. Scientific and Technical Information Processing, 37(2), 143–147.

Duangnamol, T., B. Suntisrivaraporn, T. Supnithi, and M. Ikeda, (n.d.). Engaging Reflective Thinking in Learning Mathematics Word Problems by Using a Personal Tutoring Agent.

Gunning, C., D. and Forslund, (2013). Automatic Twitter ResponseGenerator.

Huang, J., M. Zhou, and D. Yang, (2007). Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI Vol. 7, 423–428.

Keegan, M., R. D. Boyle and H. M. Dee, (2012). Turi: Chatbot software for schools in the turing centenary. In Proceedings of the 7th Workshop in Primary and Secondary Computing Education (pp. 153–154). ACM.

Kowalski, S., K. Pavlovska, and M. Goldstein, (2009). Two case studies in using chatbots for security training. In IFIP World Conference on Information Security Education 265–272. Springer.

Lasek, M., and S. Jessa, (2013). Chatbots For Customer Service On HoTELS'websites. Information Systems in Management, 2(2), 146–158.

Shabaz, K., D O'Shea, J., Crockett, K. A., and Latham, A. (2015). Aneesah: a conversational natural language interface to databases. In World Congress on Engineering 227–232.

Shaw, A. (2012). Using chatbots to teach socially intelligent computing principles in introductory computer science courses. In Information Technology: New Generations (ITNG), 2012 Ninth International Conference on 850–851. IEEE.

Shawar, B. A., and E. Atwell, (2007). Different measurements metrics to evaluate a chatbot system. In Proceedings of the workshop on bridging the gap: Academic and industrial *research in dialog technologies* (89–96). Association for Computational Linguistics.

Wallace, R., H. Tomabechi, and D. Aimless, (2003). Chatterbots Go Native: Considerations for an eco-system fostering the development of artificial life forms in a human world. Published Online: Http://Www. Pandorabots.Com/Pandora/Pics/Chatterbotsgonative. Doc.

Wang, Y. (2008). Designing chatbot interfaces for language learning: ethnographic research into affect and users' experiences (PhD Thesis). University of British Columbia.

Watanabe, Y., R. Nishimura, and Y. Okada, (2005). Confirmed knowledge acquisition using mails posted to a mailing list. In International *Conference on Natural Language Processing* 131–142. Springer.