INTERNATIONAL JOURNAL OF ADVANCE RESEARCH AND DEVELOPMENT
IJARnD
(Volume 4, Issue 8)
Available online at: www.ijarnd.com

Replacing same meaning in sentences using natural language understanding

Tin Htar Nwe, Zar Lwin Phyo, Zin Mar Myo, Aye Thidar Win
University of Computer Studies, Magway, Myanmar

ABSTRACT

NLP (Natural Language Processing) can be used to communicate with computers by means of intelligence methods in a natural language. In other words, it is very useful for classification and analysis. Although there are many natural language syntax analyzers that existed before, it still remains to fulfill the requirements for analyzing English text for English to English Machine Translation. In this paper, we have proposed a chunk-based analyzer for English to English Machine Translation System. The keyword search Technique is one of the techniques of Natural Language Understanding. A simple keyword based matching technique is used for classification. Domain-specific dictionaries of keywords are used to reduce the dimensionality of feature space. The output result is that the word matched to a known word will be replaced in the given sentence, which will not be changed the original meaning.

Keywords— *NLP (Natural Language Processing), A chunk based syntax analyzer, Matching technique, Intelligence methods*

1. INTRODUCTION

Natural Language Processing is one of the most important pieces of researches carried out in the world of Artificial Intelligence. AI-based programs allow a computer to understand and generate appropriate natural language output. With natural language processing, it is much easier to communicate with computers. AI researches in natural language processing expect their work to lead both to the development of practical, useful language understanding systems and to a better understanding of language and the nature of intelligence.

Keyword Analysis is used in early natural language processing programs. The NLP programs search through an input word looking for key words program based on keyword analysis scans the text, looking for words that it has been programmed to recognize. This program is able to know, only selected words. Once a keyword is recognized, the program responds with specific canned responses. The goal of NLP is to enable people and computers to communicate in a "natural" (human) language, such as English, rather than a computer language. In this system, there are two modules. First, it scans through an input sentence, looking for keywords. Second, it generates an appropriate output response to a keyword in the user's input.

Natural Language task is concerned primarily with understanding. We are trying to get from some input sensor data to some representation of what that data really means. We're just concerned with getting a sufficient interpretation for our purpose. It should be that there is no single right approach to NLP. There are many approaches in order to create a solution that is effective for a particular domain and task.

The simplest method of analyzing the content of a sentence is a pattern-matching technique called keyword analysis. NLP program searches through an input sentence looking for keywords or phrases. The program is able to identify, or "knows", only selected words and phrases. Once a keyword or phrase is recognized, the program responds with specific canned responses.

Alternately, the program may actually construct a response based on a partial reply coupled with the keyword or selected phrases from the input. The program recognizes very specific inputs that it uses to construct an output response or initiate some other actions [10].

A syntax analyzer has emerged as an important component in a variety of natural language processing applications. Our presented work aims to build a syntax analyzer for English to English Machine Translation [7].

Natural Language technology gives computer users the ability to communicate with the computer in their native language. This technology allows for a conversational type of interface, in contrast to one of computer jargon, syntax, and commands. Limited success in this area is typified by current systems that can recognize and interpret written sentences relating to very restricted topics.

The field of natural language processing is divided into two sub-fields:

- Natural language understanding investigates methods of allowing the computer to comprehend instructions given in ordinary English so those computers can understand people more easily.
- Natural language generation strives to have computers produce ordinary English language so that people can understand computers more easily.

2. RELATED WORK

Chunking has been studied for English and other languages, though not very extensively. The earliest work on chunking based on machine learning goes for English. Ramshaw and Marcus [6] used transformation-based learning using a large annotated corpus for English. Kudo and Matsumato [8] used support vector machine for chunking. [9] Presented an attractive finite-state cascade architecture for parsing unrestricted text and show that its distinct processing advantages. These advantages explained why the human parser might adopt a chunk-by-chunk strategy. An approach to parsing phrase grammars based on rule sequence is presented by Marc and David [11]. A new formal grammatical system called link grammar was defined by Daniel Sleator and Daby [13] for efficient parsing. This formalism is lexical and makes no explicit use of constituents and categories. Waston and Carrall [12] presented an approach based on the Inside Outside Algorithm for producing weighted grammatical relation output directly from a unification-based parse forest. Zaveral and W. Dadlemans [14] presented a memory-based learning approach to shallow parsing in which POS tagging, chunking, and identification of syntactic relations are formulated as memory-based modules. But their system identified only subject and object relations.

There has been recent work on integrating keyword search with structured. XML querying [3] [4][2]. S. Pandey and B. Bajracharya [5] proposed sharing geographic information on the Internet using ARCIMS. They present an integrated metadata server system has been designed and developed with an aim to provide a platform to share databases and metadata residing in ICIMOD (International Center for Integrated Mountain Development) and its GIS partners and stakeholders through the Internet. J. Feng and T. Watanabe presented index structure for managing multi-levels for road networks on distributed environment. They propose an index structure, MOR-tree (Multi-Levels-Object-Relation tree), for organizing integrated maintenance procedure. MOR tree is an extension of R-tree index structure with the ability of indexing spatial objects of multi-levels. O.Ozdilek and D.Z.Seker presented a web-based application for real-time GIS. Their system has two main differences from other web-based GIS applications. It is real-time and gets data from sources simultaneously in exact time intervals that user-defined previously and secondly, all the developing and publishing tools used in the system are open-sourced. Their aim is to create a web-based, free and open-sourced GIS that can work with different data formats by exchanging and presenting data as a real-time map on the web [1]. Artificial intelligence is behavior by a machine that, if performed by a human being, would be called intelligent. A thought-provoking definition is provided by Rich [17].

Mark Fox of Carnegie-Mellon University often says that AI is basically a theory of how the human mind works, Winston and Prendergast [14]. NLP program searches through an input sentence looking for keywords or phrases. The program is able to identify, or “knowns”, only selected words and phrases.

3. BACKGROUND THEORY

3.1 Natural Language Processing and Techniques.

NLP refers to communicating with a computer in a natural language (English) rather than using special commands, syntax, or menus. AI-based programs allow a computer to understand and generate natural language. With natural language processing, it is much easier to communicate with computers. Two main techniques are widely used in natural language processing programs.

There are:

(a) Keyword Search Analysis

(b) Syntactic and Semantic Analysis

Natural Language Processing (NLP) is one of the prominent research areas carried out in Artificial Intelligence. NLP offers possible solutions to the problems of communication between humans and computers. It has been very difficult to develop computer systems capable of generating and understanding Natural Language such as English. A computer system is capable of understanding a message in a Natural Language. Such a system requires both the contextual knowledge and the process of making inference. The message is understood when a successful inference is made between the contextual knowledge and the message itself. To understand a natural language inquiry, a computer system must have knowledge to analyze, then interpret the input. It must understand grammar and definitions of words. It must also have significant (sometimes very large) amounts of general. Commonsense, world knowledge, which is sometimes needed to understand even the simplest sentences used commonly by people. The implemented computer system needs a general dictionary and general inference algorithms. In addition to the general knowledge base, a natural language processing system must have a large domain-specific knowledge. This base should contain terms and algorithms. Once the computer understands the input, it can take the desired action. Once the computer takes action, it usually provides some output. In most cases, it is desirable to provide that output in natural language. The easiest way to do this is to provide the computer with scanned sentences, phrases, paragraphs, or other inputs. When a particular input is detected, an appropriate output response is accessed and delivered.

3.2 Keyword Search Analysis

The natural language processing program searches through an input sentence looking for words or phrases. The program is able to identify only selected words and phrases. Once a keyword or phrase is recognized, and the program responds with specific canned responses. This technique is suitable for analyzing formal languages that are constructed with reserved keywords. The simplest method of analyzing the content of a sentence is a pattern matching technique called keyword analysis. A program based on keyword analysis scans the text, looking for words that it has been programmed to recognize. When it finds one of these keywords, it responds by manipulating the text in some predetermined fashion.

3.3 Syntactic and Semantic Analysis

It performs a detailed analysis of the syntax and semantics of an input statement. In this way, the exact structure of an input sentence and its meaning can be determined. It is a straight forward approach for natural language translation. Although keyword pattern matching is a widely used natural language technique, its usefulness is restricted because it simply cannot deal with the large variations in language that naturally occurs. For that reason, AI researches have looked for and developed more sophisticated ways of analyzing an input sentence and extracting meaning from it. The most obvious and straight forward approach to the problem is to perform a detailed analysis of the syntax and semantics of an input statement. In this way the exact structure of an input sentence and its meaning can be determined. Of course, this is easier said than done. Even sophisticated systems for analyzing syntax and semantics fall short of the job because there are too many words with multiple meanings and an enormous number of ways to put those words together to form sentences.

3.4 Natural Language and Natural Language Understanding

Most human beings spend a large part of their time talking to one another, reading or writing. We communicated our needs and desires to one another. Language is one of the most useful things we know. Without language we would not be able to teach one another about our experiences or discovers. Language is so useful, so uniquely human, and so difficult. It is hardly surprising that it has attracted an enormous amount of interest from AI researchers. Any program capable of generating and understanding unconstrained, fluent English speech or text will have be to solve a large number of difficult problems that any mature English speaker appears to find easy. To write such a program, we will have to find out how these problems can be solved. Once we have done it, we will have a computer system which is far easier and more comfortable to use than anything which is currently available, at least for some tasks and domains. We will also have a rather better understanding of what we ourselves do when we are acting as language users.

Although today's computers have become much easier to use, communicating with a computer is not yet as simple as the most "natural" kind to communication: communicating with another person. We don't have to use any kind of specialized, technical language to communicate with other people we use natural language, such as English. If computers use natural language, you could tell them what you wanted them to do in ordinary, everyday English. If computers could generate natural language, they could ask you questions and give you information in language that would be easy to understand. Natural language processing generally refers to language that is typed, printed or displayed, rather than being spoken. Programming a computer to understand a natural language, such as English, should be easy because people can understand English with very little effort. However, it is difficult to accomplish because computers require more precision in communication than people do. In trying to create programs that allow computers to understand natural language, scientists are trying to teach computers to emulate a skill that nearly all of us perform without any trouble. The goal of natural language understanding is not to have computers understand everything we say; after all, even people misunderstand each other occasionally. The goal of natural language understanding is to allow computers to understand people, as well as people, to understand people. It is not possible to define understanding with any more precision than that with which we were previously able to define intelligence. Rather than trying to develop a formal definition of understanding, let's consider one useful characteristic: understanding allows appropriate action to be performed.

4. PROPOSED SYSTEM

When a user inputs a sentence, it is stored in input file as an input sentence. The input sentence that the user inputs is analyzed by sentence analyzer. After that, the main verb (Tense) must be searched in this input sentence. Here, the main verb is posting tense, it needs to be converted to verb simple form. This system consists of database in which the same meaning verbs are stored and all these verbs are simple present tense (V1). The main verb is converted to verb simple present tense (V1). The main verb which is converted to verb simple form is compared with the same meaning in the database and retrieved it. In checking with the database, the system performs two functions the first is found when the same meaning is e found in database and the second one is not found when it is not found in it. If it is e found, the main verb needs to be changed the user given verb form, when the main verb in the input sentence is past tense (V2), the present tense (V) in the database must be changed into past tense (V2). The next step is "Relocation". The post tense (V2) which is changed by Relocation is replaced in the original main verb of the input sentence. Then, the user can get it as an output sentence.

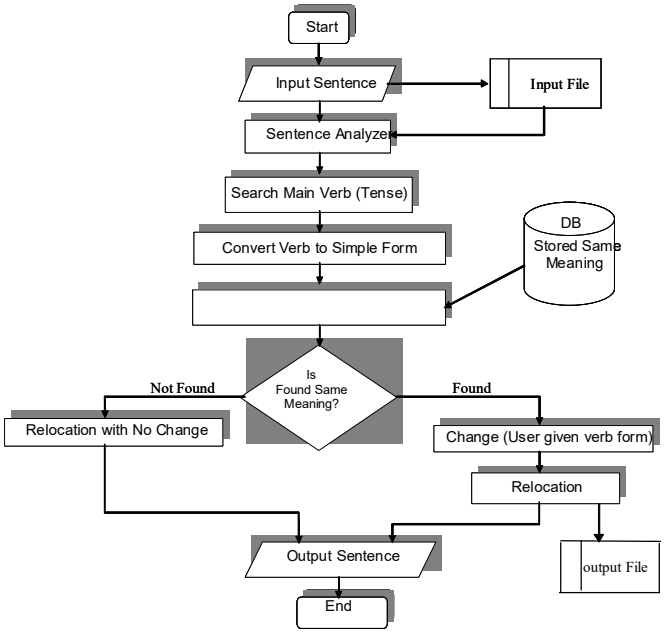

Fig. 1: Association Language Analysis on Computerized System

Otherwise, the main verb is not found in the database, that is, it is e not found, relocation cannot do any change this main verb. So, it is sent to the user as an output sentence. For example, the user inputs a sentence, "Mg Mg" bought a new file yesterday. At that time, sentence analyzer analyzes the main verb. When it finds that the main verb "bought", it needs to be converted into verb simple form, "buy". After that, it is checked whether it contains in the database or not. When it is checked with the same meaning in the database, "purchase" can be seen as the same meaning of "buy". Then, the verb simple form "purchase" is extracted or database and it is changed again into a past tense, "purchased". The next step is this converted verb is replaced with the main verb of the input sentence. Then the user will get the output sentence, "Mg Mg purchased a new file yesterday".

If the main verb of the input sentence is not found in the database, the user will get it as an output sentence without having any changes.

4.1. Algorithm: Association of Language Analysis

Input: Accept a user input sentence.
Output: Display a sentence containing the changed verb with a similar meaning from the input sentence.
Step 1: Accepts an input sentence and stores it in a file
Step 2: Search the main verb with tokenizer function and convert it into the simple form, easy to find.
Step 3: The changed form of the verb must be matched with a verb belonging the same meaning as it, while the demand factor is not satisfied and the file termination is not completed.
3.1 if the same meaning of the main verb is found in the database, it must be relocated in the given sentence with changing desired input form.

3.2 unless the found statement is satisfied, the input sentence should be displayed without changing anything.
Step 4: As a result, the system stores the desired sentence and replies the message to the user.

5. Case Study

5.1 Example Relocation

Input: Mg Mg buys a new file every month.
output: Mg Mg purchases a new file every month.
Input: Thiri called her old friends in the supermarket.
output: Thiri invited her old friends in the supermarket.
Input: They push the main door.
output: They drive the main door.

Table 1: Same Meaning Table.

Sr.no	gv1	sv1	sv2	sv3	sving
1. 2. 3. 4.	buy	purchase procure acquire corrupt	purchased procured acquired corrupted	purchased procured acquired corrupted	purchasing procuring acquiring corrupting
5. 6. 7. 8.	call	invite summon convence convoke	invited summoned convenced convoked	invited summoned convenced convoked	inviting summoning convencing convoking
9. 10. 11. 12.	push	press drive propel impel	pressed drove propelled impelled	pressed driven propelled impelled	pressing driving propelling impelling
13. 14. 15. 16.	listen	hear attend mind hearken	heard attended minded hearkened	heard attended minded hearkened	hearing attending minding hearkening
17. 18. 19. 20.	look	see watch view gaze	saw watched viewed gazed	seen watched viewed gazed	seeing watching viewing gazing
21. 22. 23. 24.	make	produce perform do execute	produced performed did executed	produced performed done executed	producing performing doing executing

By analyzing associate language on computerized systems, users can easily know the synonyms of some verbs which cannot easily be found in dictionaries. Thus, it can save time for users. It can solve the problem of ambiguous meaning for the user.

6. CONCLUSION

The aim of thesis is to analyze the input sentence syntactically and then searches the type of tense to convert another word of same meaning based on Natural allowing the computer to comprehend instructions given in ordinary English so that stored in the internal knowledge base. It will reduce the time-consuming, costs and will give efficient management without requiring much computer-skill. In this thesis, the system will be developed based on a keyword analysis of Natural Language Processing.

The system can be extended for searching other parts of speech, nouns, adjectives, adverbs, etc. The system can extend for checking sentence, paragraph, and other grammatical facts. Although it can display some synonymous verbs, it can far all details of other items.

7. REFERENCES

[1] D. G. Cottrell, "XML & GIS Application Integration"
[2] A. Gupta, R. Marciano, I. Zaslavsky, C. Baru, "Integrating GIS and Imagery through XML-based Information Mediation"
[3] Y. Chen, J.Gong, W.Jia, Q.Zhang, " XMl-Based Spatial Data Interoperability on The Internet"
[4] D.Florescu, L.Manolescu, D.Kossmann: "Integrating Keyword Search into XML Query Processing"
[5] S.Pandey, B.Bajracharya, "Sharing Geographic Information on the Internet IC/MOD's Metadata/data Server System Using ARCIMS"
[6] L.A Ramshaw and P.M Marcus, "Text Chunking Using Transformation-Bsed Learing", In proceedings of the 3rd Workshop on Very large Corpora, 1995.
[7] Myat Thuzar Tun, Ma An English Syntax Analyzer for English-to-Myanmar Machine Translation Ph.D Tesis, February, 2007
[8] T.Kudo and Y. Mastumoto, "Chunking with Support Vector Machines", In proceeding of CoNLL-2000, 2000.
[9] S. P. Aboney, "Parsing By Chunks", Kluwer Academic Publishers, Dordrecht, 1991.
[10] Turban E., Expert Systems and Applied Artificial Intelligence, New York, Macmillan Publishing Company (1992)
[11] M. Vilian and D. Day, "Finite-state phrase parsing by rule sequence", In proceeding of the 2nd Workshop on Language Learning in Logic and 4th Conference on Computational Natural Language Learning, 2000.
[12] R. Waston, J. Carrall and T. Briscoe, "Efficeent extraction of grammatical relations", In proceedings of 9th International Workshop on Parsing Technologies, Canada, 29 September, 2005.
[13] D. Daniel Sleator and D. Temperley. "Parsing English with a Link Grammar", Third International Workshop on Parsing Technologies, August, 1993.
[14] J. Zavereal and W. Daelemans, "Memory-Based Learning: Using Similarity for Smoothing", In proceedings of ACL'97, pages 436-433, Madrid Spain, 1997.
[15] T. James, "An Introduction to Natural Language Processing" Some Basic Concepts in Neuro-Linguistic Programming, 2002
[16] W. Daelemans, S. Buchholz, Jorn Veenstra, "Memory=Based Shallow Parsing", In proceedings of the Third Conference on Natural Language Learning, 1999
[17] Y.Tsuruoka and Jsujii, Bidirectional Inference with the Easiest-First Strategy for Tagging Sequence Date, In proceedings of HLT/EMNLP 2005, pp.467-474,2005.

