Available online at http://www.ajol.info/index.php/njbas/index
Nigerian Journal of Basic and Applied Science (March, 2013), 21(1): 65-69
DOI: http://dx.doi.org/10.4314/njbas.v21i1.10
ISSN 0794-5698

Full Length Research Article

An Integrated Textual Case-Based System

A. Almu
Dept. of Mathematics, Computer Science Unit , Usmanu Danfodiyo University, P.M.B 2346, Sokoto – Nigeria
[*Corresponding author: Email ; almul2003@yahoo.com; ☎: +2348064267939]

Abstract: Textual Case-Based Reasoning as a problem solving approach allows knowledge source to be integrated with a view to improving the effectiveness of the system during retrieval. The earlier proposed Textual Case-based System depends on statistical similarity alone and most of the time does not retrieve the solution to the problem even if it exists. In this paper, the WordNet is being integrated to the developed Textual Case-Based Mobile Phone Diagnosis Support system in order to take the synonyms similarity of the problem terms into account while diagnosing a given problem. Thus, the integration will makes the system not to depend on statistical similarity alone but rather take synonyms similarity of the problem term into consideration. The result of the experimental evaluation using some set of problems has demonstrated that retrieval by incorporating WordNet works better since it diagnosed 95% of the problems with relevant solutions than the retrieval without WordNet which diagnosed 75% of the problems with relevant solutions.
Keywords: Textual Case-Based Reasoning, jColibri, WordNet

INTRODUCTION

The problems experienced over the years by many mobile phone users in Nigeria resulted to the development of a textual case-based mobile phone diagnosis support system that the repairers can use to diagnose various phone problems in an easier and efficient manner (Almu and Maiyama, 2010). Since whenever a repairer submits a similar problem, its corresponding solution can be reused and presented to the repairer. Moreover, the system may not retrieve relevant solutions to a given problem especially if the repairer submits a problem that contains some terms that are entirely different with the terms in the system case-base, even if the solution to the problem exists. This is due to the fact that natural languages such as English contains terms that can mean different things in different contexts. To address this kind of problem, the WordNet is being integrated to the system in order to take the synonyms similarity of the problem terms into account while diagnosing a given problem by searching through a collection of previous cases and use similar case to diagnose the problem.

MATERIALS AND METHODS

WordNet is a lexical database that consists of about 118,000 words and also about 90,000 senses (meanings) of words with the view of providing semantic relations between the words (Miller 1995). This semantic nature of the WordNet will be very useful in comparing the documents similarity since they could have the same form but different meanings. This could serve as the basic idea to be applied to an application in order to improve its capability to behave in natural language form so as to enhance its performance.

The WordNet would be used to get synonyms of problem terms for the purpose of expanding the user given problem to the system. WordNet was described as a lexical database that serves as an electronic dictionary designed to be accessible under the control of program in identifying the semantic relationships or connections between the words (Miller 1995). This semantic feature of WordNet made it useful to be considered in this project for the purpose of extracting the synonyms sets of the problem terms needed.

Part-Of-Speech (POS) Tagging

Tagging is a way of assigning descriptors (also called a tag) to the given words or tokens automatically. The tag is used to indicate one of the parts-of-speech, semantic information etc. The process of assigning one of the parts of speech to the given word is called Parts Of Speech tagging (Martinez 2012). It is also commonly referred to as POS tagging. POS tagging is one of the steps of Natural Language Processes (NLP) and the tags are used to indicate the grammatical behavior of the words in a sentence. For the purpose of tagging in this paper some methods of the jColibri framework has been adopted. These includes: the OpennlpPOStagger() method which assigns part of speech tags to tokens and lookupWordNetPos() method which looks for the appropriate tags of each tokens.

The semantic relationship in WordNet links the Four-Part-Of-Speech of Nouns, Verbs, Adverbs and Adjectives together to synonyms sets (Miller 1995). Therefore, the words or terms of the problem have to be tagged with their appropriate POS before passing them to the WordNet assessment part of the system. For this reason, the OpennlpPOStagger() method is used to carry out the task of identifying the correct POS of each term in the problem. Having tagged the terms with their POS, then the lookupWordNetPos() method is used by the WordNet to look for the tag of every term or word in the problem. For instance, a term with a tag "NN" or "NNS" is a noun, a term with a tag "V" is a verb, a term with a tag "J" is an adjective and finally a term with a tag "RB" is an adverb (Source Forge 2009).

Synonyms Sets

The synonyms sets are the sets of synonyms of a given problem terms to be extracted by the WordNet. These words are also referred to as synsets. The synsets of a term might consist of one or more words together with the term itself. The WordNetBridge.SynsetWords() method of the jColibri framework is adopted since it extracts the synonyms of the given token or word. The WordNetBridge.SynsetWords() method is used in this case to get the synsets of a particular term or token needed. For instance, a term "work" in a problem "keypad doesn't work" might have another synsets of "function" and "operate" to be displayed in the expanded diagnosis area of the system interface. So when a user included those problem terms synsets in the expanded diagnosis to be submitted to the system, it would made it possible for the system to retrieve those relevant solutions in the case base that might have either "work" , "function", "operate" or even all the terms in them and present them to the user as a solution.

Integration of WordNet

WordNet is incorporated into the application to improve the solution retrieval. The output of the preprocessing stage as described in Figure 1 is to be sent to the WordNet. The WordNetBridge.init() is a method of the jColibri framework that loads the Wordnet dictionary file to the memory for it to be accessible by the program. The WordNet is made available to the application by calling the WordNetBridge.init() method that initializes it by loading the dictionary file into the memory. Before passing the problem tokens to the WordNet, the OpennlpPOStagger.tag() method performed the part-of-speech tagging. This involved giving each token a tag (i.e. a noun, a verb, an adverb or an adjective tag). Having given each token a tag, then the lookupWordNetPos() method identifies these tags of the tokens to the WordNet with a view to providing their synonyms. Finally, the WordNetBridge.SynsetWords() method supplies the synonyms to those problem tokens.

The getSynsets() is a method created in one of the application class called the SystemProcessor to store the extracted synonyms sets from the WordNet of the problem tokens. Then, this method is called in the application main class known as the SystemInterface together with the created instance "systemprocessor" of the SystemProcessor to make is possible to display the synonyms sets in the expanded diagnosis area of the system interface.

System Evaluation

After integrating WordNet to the diagnosis support system, the performance of the system was evaluated based on the relevant of the textual solution retrieved to a given problem. The evaluation is done using some collection of 42 mobile phone problem cases compiled from some phone repairers within Sokoto metropolis whom normally repairs the users' phone. Face-to-face interviews were conducted during the compilation process of the problem cases to enable the researcher to identify some of the problems with clear solutions. Among all the cases gathered during the survey only 42 problems were found to have clearly defined solutions.

The evaluation experiment is typically designed to provide an appropriate method to be used in measuring the performance of the system. According to Bruninghaus and Ashley (1998) evaluation of Textual Case-Based Reasoning system is particularly challenging since the system may consists of different components that will form the overall system. Therefore, it has to be carefully designed so as to assess those individual components to capture the overall performance of the system.

In this situation, since the problems case-base to be used consists of 42 mobile phone problem cases and each problem has only one right solution to be retrieved as a solution. In order to improve the possibility of the system to retrieve more than one solution that could be relevant to a given problem and also to take care of the problem that might be expressed in different words by the users. Each problem in the case base is rephrased into two additional problems that contains some of the terms from the original problem and even with some

terms that are quite different from the original words of the problem.

The rephrased problems were used as the sample problems cases to form the case base for testing the system. Ten (10) problems with the possibilities of having some terms with more than one synonym were selected from the original problems to run the experimental trials. But, when running each of the trial the original problem was taken away from the case-base since it will be used as the input to the system. Similarly, each trial was run using the two different diagnosis techniques that the system uses namely normal diagnosis and expand diagnosis.

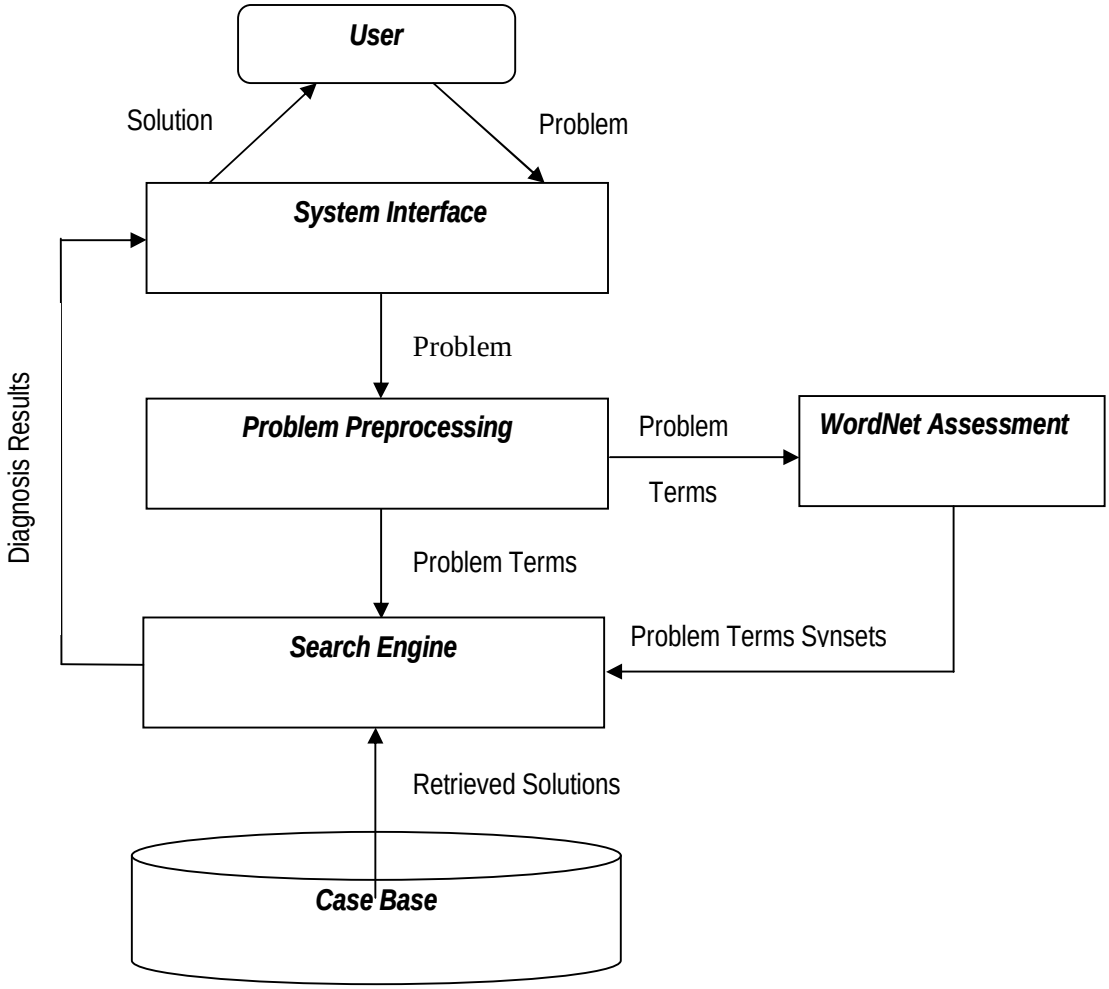

Figure 1: The Extended Architecture of the Mobile Phone Diagnosis Support System

Normal Diagnosis: This would retrieve some solutions that might be relevant to the user given problem using the problem terms. This is the diagnosis technique of the earlier system.

Expand Diagnosis: This would retrieve some solutions that might be relevant to both user problem terms and extracted synonyms of the problem terms from the WordNet. This is the diagnosis technique of the integrated system.

The experiment is carried out using five (5) domain experts (repairers) and five (5) users, by allowing each expert or user to submit one problem for the system to diagnose. So whenever each of the problems is submitted, the expert or user was asked to identify the relevant solutions being retrieved by the system to that problem. The performance of the system is captured by taking the precision of the retrieved solutions into account.

RESULTS AND DISCUSSION

The result of the experiment was recorded for the precision at 2 since there are only two possible relevant solutions to be retrieved for each problem. Therefore, the possible values that the result can have are 0, 0.5 and 1. For instance, if the system retrieves one solution that answer a given problem then it is precision is 0.5. Then if it retrieves the two solutions that answer a given problem then the precision would be 1. But, if none of the retrieved solutions answer a given problem then the precision would be 0. The summary of the results obtained when running the 10 problems by using the two mentioned different diagnosis techniques are represented in form of graphs as shown in Figure 2 and 3 respectively.

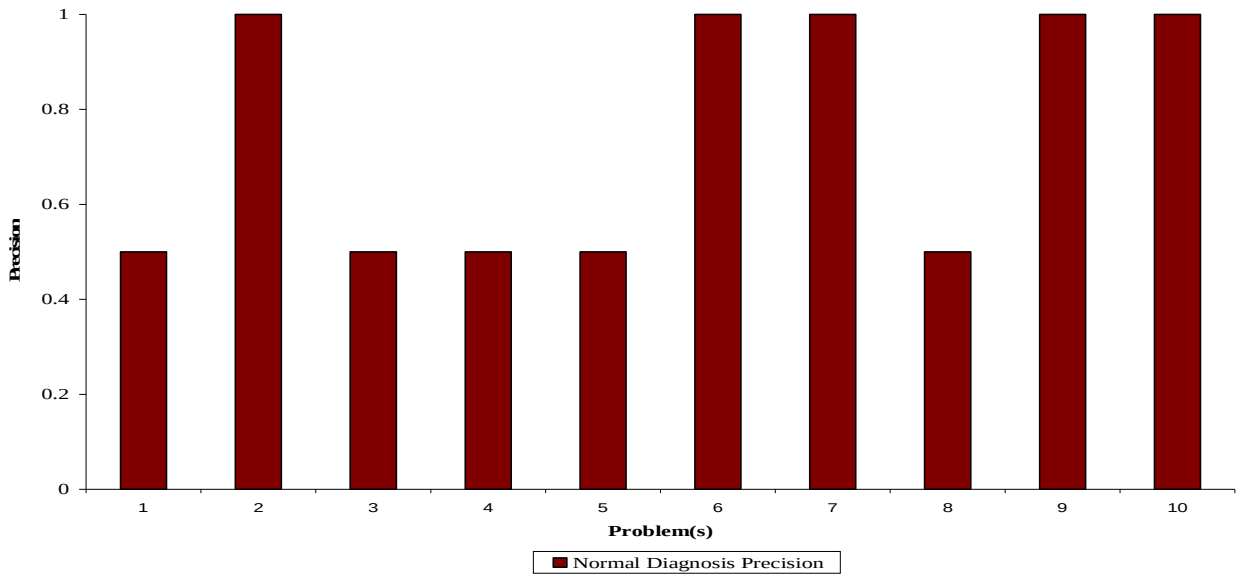

Figure 2: Precision Result of the Scheme Using Normal Diagnosis

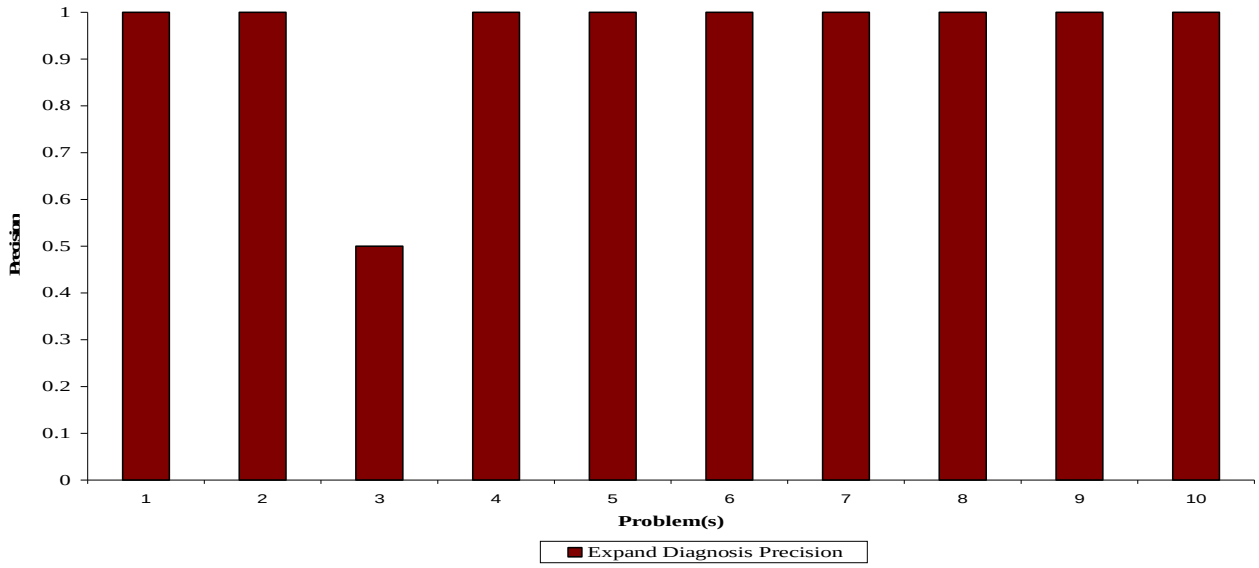

Figure 3: Precision Result of the Scheme Using Expand Diagnosis

From the results of the two graphs above only problem 3 has not improved with the expand diagnosis which happens due to the "keypad" term in that problem has only one synonym as identified by the WordNet.

The result of the evaluation shows that the diagnosis support system demonstrates better effectiveness in both the two diagnosis techniques for retrieving the relevant solutions to answer the user's given problem.

The two diagnosis techniques that the system uses performed better and the result obtained has shown that expand diagnosis was significantly better than the normal diagnosis based on the 10 sample problems used for the experiment. Therefore, based on the system precision result obtained, it can be inferred that the expand diagnosis retrieves only one (1) irrelevant solution to the user given problems (i.e. problem 3) than the normal diagnosis which retrieves five (5) irrelevant solutions (i.e. problems 1, 3, 4, 5, and 8). Similarly, it can be inferred that the expand diagnosis retrieves more relevant solutions to the user given problems than the normal diagnosis.

CONCLUSION

This work has proposed an integrated diagnosis support system which incorporates Wordnet in order to expand the user given problem for the purpose of improving the retrieval solutions. The difference between normal and expand diagnosis techniques of the mobile phone diagnosis system demonstrated that the system can diagnose a problem even if it is expressed in a quite different words. Although, the better result of the expand diagnosis depends on the appropriate choice of the synonyms for the problem terms by the user. Also, this significant difference happened due to the fact that, the normal diagnosis can perform the retrieval by using the statistical similarity to retrieve relevant solutions from the system case-base based on the occurrence of the problem terms in those solutions. Similarly, the expand diagnosis used the statistical similarity to retrieve relevant solutions from the case-base but by incorporating the knowledge source from the WordNet to extract the synonyms of the problem terms needed. This would enable those solutions that have the same words with the user problem and those that have different words in the case-base but the same meaning with the user given problem to be retrieved by the system. Typically, the expand diagnosis has really improved the retrieval performance of the overall system by retrieving about 95% of the relevant solutions to the problems than the normal diagnosis which retrieves about 75% of the relevant solutions to the problems.

However, as a future work, the research should consider more problems instead of ten (10) such as 20, 30, 40, 50 etc. Secondly, the work should be extended by developing the domain electronic dictionary that will capture the specific synonyms related to mobile phone problems instead of using Wordnet. This will actually take care of the same mobile phone problems that could be expressed using different words or synonyms.

REFERENCES

Almu, A. and Maiyama, K. M. (2010). A Textual Case-Based Mobile Phone Diagnosis Support System. *Nigerian Journal of Basic and Applied Sciences,* **18(2):** 260-268

Bruninghaus, S. and Ashley, K.D. (1998). Evaluation of Textual CBR Approaches. In: *Proceedings of the AAAI-98 Workshop on Textual Case-Based Reasoning.* Pittsburgh: AAAI Press. Pp. 30-34 [online] Available from: http://www.geocities.com/bruninghaus/papers/eval-tcbr.pdf [Accessed 22 May 2009]

Martinez, A.R. (2012). Part-of-Speech Tagging. *Wiley Interdisciplinary Reviews: Computational Statistics,* **4(1):** 107–113.

Miller, G.A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, **88(11):** 39–41.

Source Forge. (2009). jCOLIBRI: *CBR Framework*. [online] Available from: http://sourceforge.net/projects/jcolibri-cbr/files/jCOLIBRICBR/jCOLIBRI21.zip/ [Accessed June 22, 2009]