

Question analysis and answering system for the internet usages by using Ontology and Concept Weighting

Narroup Ragnam, Anongnart Srivihok

Department of Computer Science, Faculty of Science, Kasetsart University, Thailand

email: {g4764229, anongnart.s@ku.ac.th}

Abstract: This study proposed a system for an automatic question analysis and answering that applies information extraction techniques for extracting the relevant information from questions of the Internet connections. The study uses both ontology, and concept weighting to extract, and classify questions. The information extractions are divided in two steps as follows. First, ontology is used to create a decision tree according to the concepts of questions. Second, concept weighting is applied; the weight of each concept is given by its frequency in the past documents. The results from this study reveal that it is possible to develop an automatic question answering system. The proposed system had been tested with sample questions from the Internet service provider: True Corporation Call Center. Precision of the system is about 0.77 which is good. This study offers useful information regarding the areas of information extraction for Call Center or Web Based Support Systems.

Keywords: Question analysis; Ontology; concept weighting; Call Center.

I. Introduction

The number of internet usages is increasing tremendously every year. Problems of users with the Internet connections are also increasing. Thus, the Internet Service Provider also needs to help users solve these problems. It is claimed that 80% of internet connection problems are simple and can be solved by users without the assistance of call center. These problems include telephone cord is not plugged tightly, or misspelling password for authentications. In July 2005, the number of phone calls related to internet usage problems in True Corporation which is one of the largest internet service in Thailand are about 71,628 calls. This is huge and these calls consume a lot of service time.

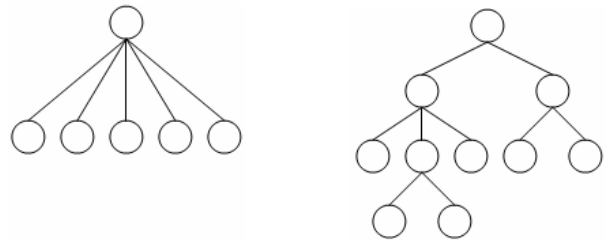
The objective of this study is to offer the proposed question answering systems by using ontology and concept weighting. This study include four main sections. Section 2 is the related theories and studies in the past. Section 3 is research methodology. Last, section 4 is conclusion of the study.

II. Literature Review

II.1 Text Classification

The text are classified by using a classifier which might be

a decision tree. The structure of decision tree can be divided in two categories: hierarchical and flat structure (Figure 1). It was reported that using hierarchical structure for text classification provides higher performance than flat structure [3].



(a) Flat Structure

(b) Hierarchical

Figure 1. Structure of decision trees.

Ontology is a formal, explicit specification of a shared conceptualization [2]. Also, Ontology has been applied in various studies. Miller (1993) introduced WordNet which is an on-line lexical database. This study used ontology for conceptualizing words by lexical order such as verbs or nouns. Wordnet uses semantic in categorising the questions. Yeh and coworker (2003) applied semantic inference on ontology for data mining of medical FAQ. In Thailand, Mounpai and Koatrakul (2003) applied ontology in the classification of One Tambol One Product (OTOP) database which includes large variety of Thai handicrafts and food products.

II.2 Concept weighting

In 2004, Jinarat and Soonthornphisaj [2] used probability method-Naïve Bayes and feature weighting in text classification. The feature weights are changed according to their frequency in the documents. This study reveals that the classifier has better performance after adjusting concept weights. Further, Reinforcement theory is applied in feature weighting of an intelligent agent for tourism. Features are awarded and increased their weights if the user selects the provided trips. On the other hand, they will be penalised by decreasing feature weights if they are not selected. By this way, the weights of trip features are adjusting according to the user selections [7].

II.3 Searching

After the classifier weights are adjusted, then the answers matching for each concepts have been linked. By this way, the customers who want to find the clues for trouble shooting can use the system.

III. Research Methodology

This study presents methods for question analysis and answering system of call center. The system can help customers in trouble shooting of errors in the Internet connections without the assistant of call center. The reseach design includes four steps (1) question extracting by Latent Semantic Analysis (LSA), (2) Ontology, and (3) Concept weighting. Ontology is applied for classification of concepts in knowledge base and store in hierarchical structure.

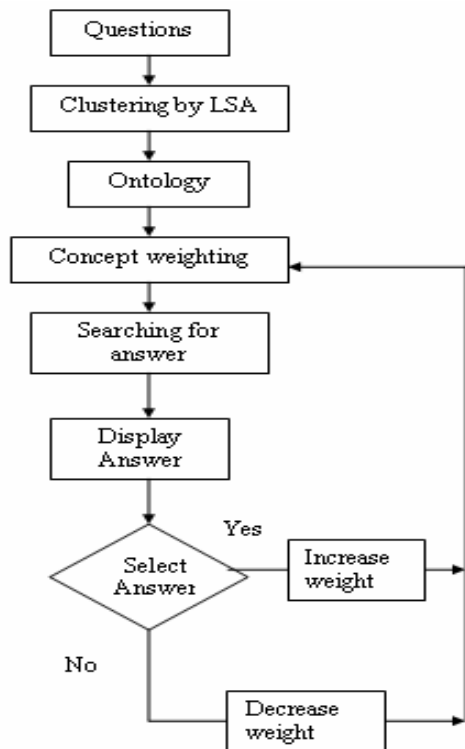


Figure 2: Flow of question analysis and answering system

III.1 Question Extracting

Thai documents are written continuously without stop words or spaces between words. First the questions need to be separated into single words by using Swath program developed by Paisarn (1997). Then Latent Semantic Analysis is applied for solving problems of words which have many meanings namely polynymous and the case of many words which have the same meaning synonym. Then similar words are grouped in one concept. Figure 3 depicts the word synonym.

Word synonymy	Intension
Connect, เชื่อมต่อ, ต่อเน็ต, connected	การเชื่อมต่อ / Connect
คิด...ดับ, หลุด, เน็ตตัด, disconnected	เน็ตหลุด / Frequently disconnected
Hisp, high speed internet, ไสปีด, hi speed internet	Hi-speed internet
เออเรอ, Error, Errors	Message Error
...ช้า, เร็ว...k, speed...MB, slow	ความเร็ว / Speed

Figure 3: Word synonym and grouping

III.2 Document Classification by Ontology

Ontology is applied to build the hierrachical structure classifier. In this study, problems in the Internet connections are classified and built in decision tree which have hierrachical structure (Figure 4). This ontology is developed by the assistant of experts and engineers in the Internet and networking. The problems in internet connections are divided in two main categories: (1) internet can not be connected and (2) internet can be connected but there are some problems. Then each category is divided in subcategories. Each node of the decision tree contains clues or answers which help customers to solve their problems in the Internet connections. Figure 4 reveals the structure of decision tree used in this study.

III.3 Concept Weighting

The frequency of occurrences of each problem is differrent, thus each concept in the decision tree has different weight. As well, concept weight is depended on the frequency of it occurrence in the Internet connection. From 71,628 calls at call center, about 81% (57,803 calls) complain about unable to connect the Internet. Whereas, the one which can perform the Internet connection still has problems.

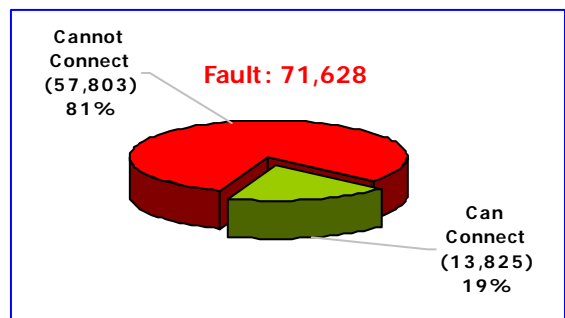


Figure 5: Probabilities of errors in the Internet usages reported at Call Center

Next step is adjusting the concept weights. In this step, each concept will be assigned with a weight according to its occurrence from historical data (Figure 6). For example, the highest frequency of concept which causes internet disconnect is due to no installation of splitter in the computer system which is 42%. The second higher is broken connection line or noise which 21%. Next is connection time out is 19%. Last is virus computer or spyware is 18%.

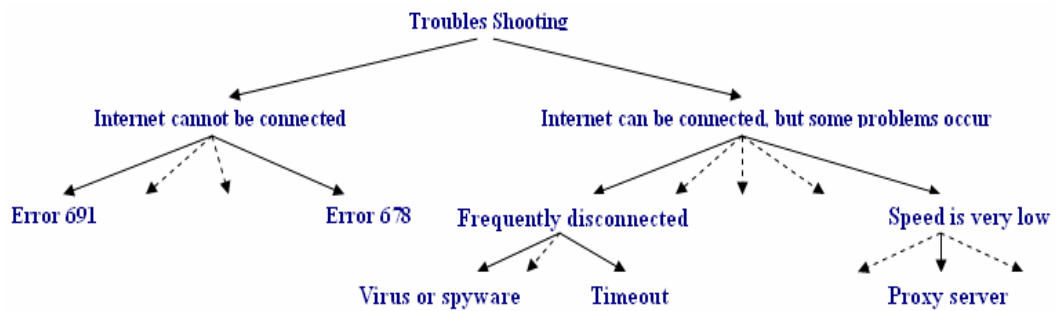


Figure 4: Hierarchical Structure of trouble shooting of the Internet connections

Probability	Cause
0.42	Splitter installation
0.21	broken connection line or noise
0.19	connection time out
0.18	virus computer or spyware

Figure 6: Probability of the cause of the Internet Disconnection

Customer who has problems with internet connection requests for assistants. The system provides answers to customer, then customer responds whether the answers work or not. Customer responses are used in weight adjusting.

Weight Adjustment:

$$W_{ti} = W_{ti} + Q_i$$

W_{ti} = weight of each concept

Whereas Q_i is defined as:

0.1 if user agree that the provided information works

-0.1/n-1 if user does not agree that the provided information works ; n is total number of features

$$\text{Precision} = \frac{\text{Number of correct retrieved answers}}{\text{Total number of retrieved answers}}$$

Example of Weight adjusting for user agrees that splitter installation advice works:

$$W1 = 0.42 + 0.1 = 0.52$$

$$W2 = 0.21 - 0.033 = 0.177$$

$$W3 = 0.19 - 0.033 = 0.157$$

$$W4 = 0.18 - 0.033 = 0.147$$

Figure 7 reveals the advice for splitter installation according to customer requests. Also the system asks for the feedback after customer has solved the problem with the provided information.

Splitter should be installed to prevent disturbing signal during hi-speed Internet access, by following the below instructions.

1. Connect True telephone line to Line socket of the Splitter.
2. Connect telephone line from Phone socket of the splitter to the telephone set.
3. Connect the line from Modem socket of the splitter to the ADSL modem.

[Click here to see how to install splitter](#)

After following the above steps, can the problem of frequent disconnections be solved?

Yes, it can.

No, it cannot.

Figure 7. Suggestion by system for trouble shooting.Conclusion

IV. Results

The system is evaluated in two steps (1) performance of question extraction (Pf) and (2) performance of retrieving the answer by using precision. The two steps are conducted by using the following formular.

Performance in question extraction (Pf) :

$$Pf = \frac{\text{Number of the correct extracted questions}}{\text{Total number of extracted questions}}$$

Sample data in this study is obtained from Call Center at True Corporation. Data was cleaned and selected for the one that related to the Internet connection problems. Then, experts in computer network answered questions from call center. These questions-answers are stored in the knowledgebase. Then, 2,800 samples are tested with this system on the performance of question extraction and precision in retrieving correct answer. Results show that question extraction performance is very good, the correctness is about 91%. Further, the precision of retrieving correct answer is about 77% ,which is good.

% correct question extraction	% correct searching
91 %	77 %

Figure 8: Comparisons of the performance of word extraction and searching

The precision of the system is not high as performance of question extraction. Since, in the process of question answering, there are some questions that should not be answered in public so the clues are not provided. Further, some provided answers in solving the Internet connection are too complex and can not be operated by novice customers.

V Conclusion

This study proposed question analysis and answering system for call center. Ontology and concept weighting are applied in this system. The system has been evaluated in two categories: performance of question extraction and precision. Then, it was tested with 2800 samples. The performance of question extraction is very good, the correctness is 91%, while system precision in searching for correct answer is 77% which is good. Future research should be done on the development of intelligent systems/agent for call center, this system can be an automatic system that help customers and company to work more efficiently and effectively.

References

- [1] Charoenpornasawat, P., Swath, <http://www.links.nectec.or.th/~yai/>, 2005.
- [2] Gruber, T. R., "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 1993, 199-220.
- [3] Jinarat, S. and Soonthornphisaj, N., "Improve Hierarchical Naïve Bayes Classification using Feature Weight Adjustment", Thailand Computer Science Conference, 2004, 55 - 62.
- [4] Yeh, Jui-Feng, Ming-Jun Chen and Chung-Hsien W. Yeh, "Semantic Inference Based on ontology for medical FAQ mining". Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering, 2003, 170-175.
- [5] Meanprai, N. and Kawtrakul, A. "Natural Language Processing Tech-based Knowledge Management for e-Organization", Proceeding of the Third National Conference in eBusiness, NCEB2003, Chulalongkorn University, Bangkok, 2003.
- [6] Miller, G. R. Beekwith and C. Felbaum, "Introduction to WordNet: An On-line Lexical Database", MIT Press, 1993.
- [7] Srivihok, A. and Sukonmanee, P., "Intelligent Agent for e-Tourism: Personalization Travel Support Agent using Reinforcement Learning", Proceeding of The 14th International World Wide Web Conference (WWW2005) Workshop . 10-14 May 2005. Chiba: Keio University, Japan.