BUILDING COMPLAINT ONTOLOGY BASED ON KEYWORDS

Patcharaporn Panwong

School of Information Technology, Mae Fah Luang University, Thasud, Chiang Rai 57100, Thailand
*e-mail patcharaporn.pan@mfu.ac.th

Abstract

Ontology is a hierarchical structure belonging to specific concept which consists of main categories, topics and keyword terms in case of building for topic identification. However, the defined structure of ontology affects a relevant result of finding the topic. If ontology is not well defined one, this is difficult to find the topic by mapping nodes from ontology with the extracted keywords from the content. To get better performance, analyzing and defining the related keyword terms become one of the most significant process in building ontology. The existing complaints from "30 Bath Co-payment Scheme", Thai governmental policy related to citizen health insurance, are used as the case study. This study presents complaint ontology which is built to collect topics and keyword terms which are extracted from real complaints and also adjusted from the existing one by using old structure in separating main category (officer's guideline).

Keywords: complaint ontology; 30 Baht co-payment scheme; keyword terms of ontology

Introduction

For developing countries, many governments use the health insurance policy as citizen's facility (Guy et al. 2005). In Thailand, "30 Baht Co-payment Scheme" is one of our government's policies related to the health insurance of citizen. The citizen has to pay 30 Baht only, each time he undergoes treatment. Moreover, the National Health Security Office (NHSO) is mainly responsible for running and operating this scheme. This scheme has been undertaken for several years. However, there are some problems occurred from using and implementing this policy due to lack of required resources and properly management. This leads to lodge numerous complaints to the government via several channels such as phone call, mail (post or e-mail), Website and so on. Most commonly, this is due to service, impolite staff, and poor work by the public services. These complaints are significant feedbacks as opportunity for improving the service to achieve user's satisfaction (Janelle and Claus 2008). Therefore, some officers are assigned to review and summarize them manually before providing solution.

To support their work, the existing complaint ontology is built from officer's guideline for topic identification function which is a part of an integrated complaint management system. The result of testing provided an error of around 35% for finding topic. This could be caused because of many related factors. One of them is the accuracy of finding topic depends on the keywords defined in this complaint ontology (Patcharaporn and Vatcharaporn 2007). Therefore, this paper focuses on building ontology based on keyword extraction from the real

complaint at topic and Leaf node level. Moreover, this proposed complaint ontology will be adjusted from the existing one by using old structure in separating main category. For these data given by user, some of them are related to “National Health Security Act”, the proposed legislation passed by parliament in November 2002. Moreover, this is well defined in categorizing and analyzed for fitting with user work.

The Existing Complaint Ontology

The existing complaint ontology of “30 Baht Co-payment Scheme” is the hierarchical structure which is built from officer’s guideline as mentioned earlier. Figure 1 illustrates some examples of its class hierarchy. Lower ontology has 3 levels; main categories such as ‘Standard’, topic such as ‘Treatment’ and keyword term such as ‘Infect’.

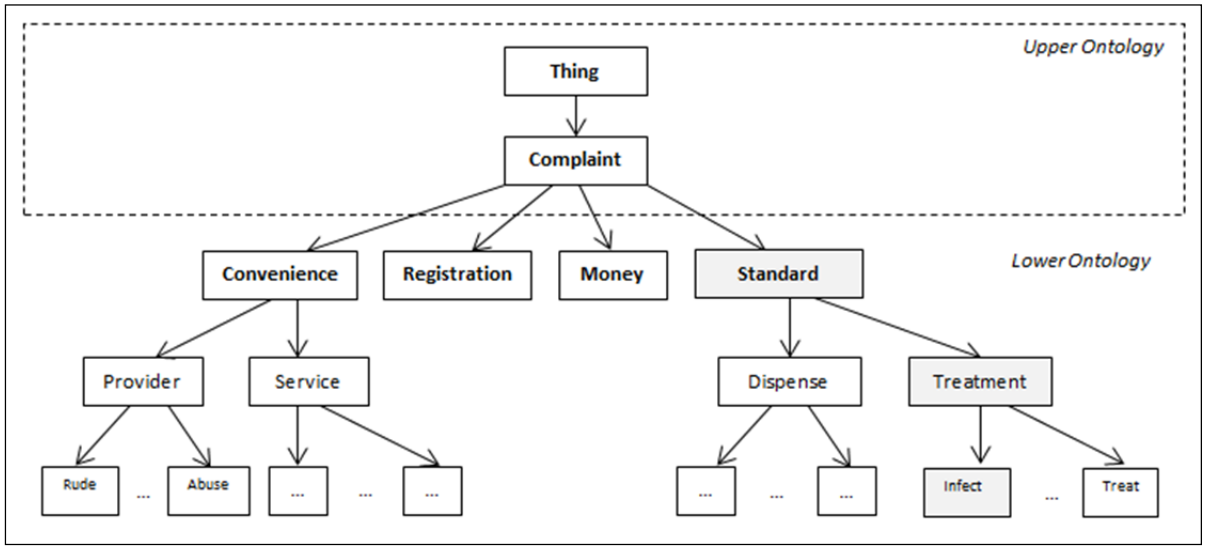

Figure 1 Some parts of class hierarchy of the existing complaint ontology

Methodology

The real cases of complaint, 100 complaints, from unsatisfied patients who live in Chiangmai and Chiangrai province are used for building proposed complaint ontology but all are in English version as a case study.

1. Specify Keyword Terms
The real complaint data is segmented into an individual word, apply Stoplist approach for removing common words, and then use heuristic rules to remove or transform English suffixes (M.F. 1980). In addition, a list of some keywords that do not use as Leaf node are “right”, “hospital”, “patient”, “treatment”, “doctor”, “staff”, and disease name. Because those words often occur in the detail of the complaint. This means that they are not specific keywords that help to represent the superclass.

After listing all candidate keywords belonging to all topics, there are two processes that have to be done. Firstly, keyword in case of occurring in many topics is set as Leaf node for one topic by considering the highest percentage value of occurrence compared with all complaints. Table 1 illustrates the comparison of keyword: “accident” occurrence in related

topics in term of the counted number and percentage. This shows that keyword: “accident” occurs in many topics but only can be a candidate keyword terms (Leaf node) referred to topic: “Right”.

Table 1 Comparison of keyword: “accident” occurrence in related topics

Topic	Keyword	Num# Occurrences	% of all
Behavior	accident	1	7.69
Doctor diagnosis	accident	2	15.38
Money	accident	1	7.69
Right	accident	6	46.15
Transfer report	accident	1	7.69
Treatment	accident	2	15.38
Total		13	100.00

Secondly, some keywords which have the lower percentage value than 80% are eliminated from a list of candidate keywords belonging to each specific topic.

Table 2 A list of some candidate keywords belonging to topic: “Registration”

Topic	Num# complaint	Candidate key words	Num# Occurrences	% of all
Registration	44	verify	35	97.22
		show	34	97.14
		change	14	93.33
		transfer	8	23.53
		register	10	100.00
		web	6	100.00
		expense	2	12.50

Table 2 shows a list of some candidate keywords belonging to topic: “Registration”. After eliminating process, the keyword terms of this topic are “register”, “web”, “verify”, “show”, and “change”.

2. A Button-Up Development Process

A button-up development process is selected to develop the class hierarchy of ontology after specifying relevant keyword terms as Leaf node of ontology from previous section.

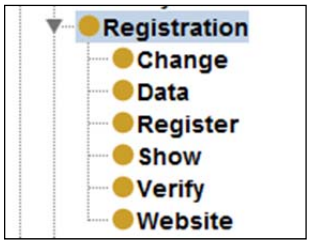

Figure 2 An example of class hierarchy from bottom level of topic: "Registration"

Figure 2 illustrates an example of class hierarchy of topic: "Registration". There are six keyword terms at the bottom level of this topic. After defining the leaves of the hierarchy, subsequent grouping of them into general concepts is required (Sanjay and Akshat 2010). The general concept in this case is category or the first level of lower ontology. This proposed complaint ontology will use almost same category names and topics with the existing one but some of them have been revised or added for better meaning and clear grouping due to the result of specifying keyword terms.

Results

Some parts of class hierarchy as final result, the proposed complaint ontology, can be represented as shown in Figure 3. The main structure of this proposed one is based on the existing complaint ontology and still has main 3 levels; main categories, topic, and keyword terms (Leaf nodes). But the keyword terms are selected by filtering only the significant one. Therefore, some keywords from the existing one are eliminated and some keywords are added to reduce the ambiguity of referring topic. In addition, some topics e.g. topic: "Diagnosis" and "Lack of coordination" are added due to new keyword terms extracted from the real complaint.

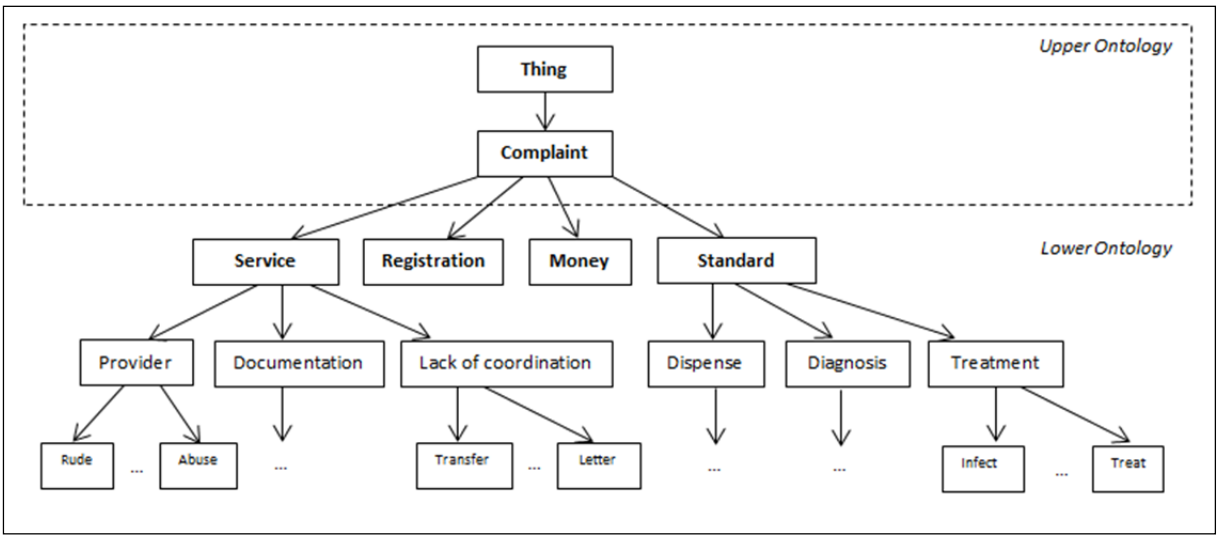

Figure 3 Some parts of class hierarchy of the proposed complaint ontology

Discussion and Conclusion

This work presented in this paper focuses on building ontology based on keyword extraction from the real complaint at topic and Leaf node level. For main category level, this proposed complaint ontology is still built based on old structure from the existing one (officer's guideline). The real complains are used to specify the related keyword terms which are Leaf nodes in ontology of each topic node. This proposed complaint ontology is eliminated the inappreciable keywords and added more relevant keywords. Furthermore, some topics are added due to new keyword terms extracted from the real complaint. This helps to reduce the ambiguity of keyword terms referred to topic and may provide higher possibility to discover topic by adding more relevant topics. Moreover, the maintenance of the entire ontology model tree (Ma et al. 2006) is required due to the changes such as policy.

References

1. Sanjay Anand and Akshat Verma (2010) Development of Ontology for Smart Hospital and Implementation using UML and RDF. IJCSI International Journal of Computer Science Issues, vol. 7, issue 5, pp. 206-212
2. Janelle Barlow and Claus Moller (2008) A complaint is a gift, 2nd ed. California: BK
3. Patcharaporn Panwong and Vatcharaporn Esichaikul (2007) An Integrated Complaint Management System for Thai E-Government. The 6th International Conference on Business
4. Ma Hui-nan, Wu Jiang-ning, and Dang Yan-zhong (2006) Ontology-Driven Information Retrieval System for Customers' Complaints of Mobile Communication Services. Management Science and Engineering, pp. 30-35
5. Guy Carrin, Maria-Pia Waelkens, and Bart Criel (2005) Community-based health insurance in developing countries: a study of its contribution to the performance of health financing systems. Tropical Medicine & International Health, vol. 10, issue 8, pp. 799–811
6. M.F. Porter (1980) An Algorithm For Suffix Stripping. Program, 14(3), pp. 130-137