An Integrated Complaint Management System for Thai E-Government

Patcharaporn Panwong and Vatcharaporn Esichaikul
School of Engineering and Technology, Asian Institute of Technology (AIT)
Pathumthani, 12120, Thailand
Email: panwong.p@gmail.com, vatchara@ait.ac.th

Abstract

Complaint service systems are provided by numerous Thai governmental agencies through their websites. These websites provide various channels for lodging a complaint such as web boards or electronic forms to be filled in online. Currently, most complaints are handled manually, assigning people to read and summarize each complaint. This way of handling a huge amount of complaints in the form of raw data gets very cumbersome and time consuming because prompt decisions cannot be undertaken.

This study presents an integrated complaint management system with important features like the grouping and ranking function. Besides, this system assigns a topic using the complaint ontology automatically to each complaint. This ontology is built to collect the existing categories and keywords. The "30 Bath Co-payment Scheme" of the Thai governmental policy related to health insurance of citizens is used as the case study for this study.

Key Words: Complaint management system, E-government, Complaint ontology, Topic identification, Classification of complaint, Simple Multi – Attribute Rating Technique (SMART)

1. Introduction

There are many ways to make a complaint to the government. Complaints can be made by phone call, mail (post or e-mail), as well as using a complaint service system from the official website.

In Thailand, the government launched a portal, http://www.1111.go.th, to enable citizens to lodge complaints about the country's public services. Besides, some government agencies have developed their own websites which allow citizens to lodge complaints. A fusion of web board and e-mail are used as communication tools in most of these websites. When using the web board and electronic form, the complaint data is stored in a database and the government or its agencies usually assign some staffs to handle it. Unfortunately, the complaints are processed manually, where they are reviewed, summarized and a resolution is crafted accordingly.

There are two main problems with the current system. The first problem is the time consumed in managing the complaint data. Consequently, this leads to a slow response to the complainer. As the complaints are often made on the same services and topics, the data will need to be classified into main categories. If the complaints are frequent, there will be an enormous accumulation of raw data which can further exacerbate the problem of data handling and management. The second problem involves prioritizing the complaint topics. Due to these problems, it is better to apply grouping and ranking technique (decision support technique) to manage the complaint data.

In this study, "30 Baht Co-payment Scheme", the existing government project of Thailand, is used as a case study and the data source. "30 Baht Co-payment Scheme" is one of the Thai government's policies, related to the health insurance of its citizens. In the proposed scheme, the Thai government pays for its citizen's medical expenses, therefore the citizen has to pay 30 Baht only, each time he undergoes treatment. The National Health Security Office (NHSO) is mainly responsible for running and operating this scheme.

2. Related Work

In this research, the proposed complaint management system has three main features: topic identification, grouping and ranking.

For topic identification, it can be assigned by a number of methods such as occurrence of certain keywords, their frequency [2], and the context of the document [4]. From related works, most of them have similar steps of topic identification that differ a little in details. To identify a topic, the following steps are performed:

Extracting keywords from textual data: In this part, it also removes the stop word from extracted

words. Applying some techniques for finding the topic / mapping with existing topic: In this part, several approaches can be applied to find the topic. For example, computing the similarity, and probability [1], [2], [8]. In case of mapping, Tiun et al. [12] and Charoenkijipaiboon [11] use mapping with the words of ontology concepts. That ontology, a set of concepts within a specific domain and also the relationships between those concepts [10], should consist of targeted topics. Other approach [7] applied Neural Networks for topic identification in natural language dialogues, suitable for long document.
Optimizing suitable topic: This is a selection of topics fit for each document.

The approach used is based on the work by Tiun et al. [12] and Charoenkijipaiboon [11] which used ontology for topic identification. In addition, they applied WordNet, an external linguistic database [5], which is used to expand query term by using a set of its synonym.

To group similar topics together, using ontology provides the classification of categories automatically, because the ontology contains class hierarchies of categories. This is one of many benefits in using ontology.

For ranking function, it relates to decision making process based on alternatives and criteria. When the number of the criteria and alternatives is finite, this problem is called Multi – attribute decision making problem. The two approaches given are based on the Multi-attribute Utility Theory (MAUT). They are Simple Multi – Attribute Rating Technique (SMART) and The Analytic Hierarchy Process (AHP). Zhao [9] stated that AHP is suitable technique for complex decisions, and require checking because of some redundant data from its pairwise comparison matrix. Edwards et al. [13] and Fülöp [6] described that SMART is easy to implement, and a suitable technique If new alternatives are likely to be added to model later. Therefore, SMART is selected for implementing this function.

3. Methodology

This proposed complaint management system have front – end and back – end components. For the front – end component, the website is the interface and an electronic form is used as a communication tool. The back – end component has the administration part which is used to process data.

In addition, this proposed system consists of three main modules. These are topic identification, classification of similar topics, and ranking main categories according to their priority.

3.1 Topic identification module

3.1.1 Building the complaint ontology. This complaint ontology is the class hierarchy of the existing categories, subcategories and keyword terms from NHSO's data collection as shown Figure1.

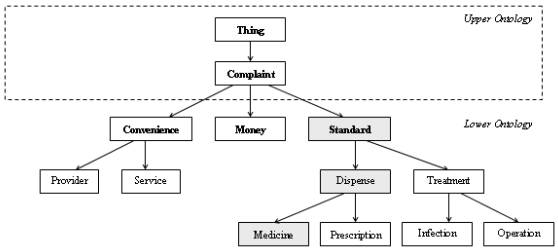

Figure1. Some examples of class hierarchy related to this complaint

3.1.2 Extracting keywords. In general, keywords characterize a content of text and become candidate term to be a topic of document. This module handles extracting keywords from the detail of each complaint.

3.1.3 Mapping extracted keywords with ontology. This proposed system sets the weight of each node in ontology, following an order of node importance. The order of node importance (level of node importance) comes from how specific each node or word is and can refer to the complaint topic; this is the way for distinguishing the difference in word importance. The proposed system uses four levels of organizing nodes into each level of node importance. Table1 shows the lists of word levels and their node examples.

Table1. The lists of word levels and their node example

Level	Level description	Example nodes in ontology
Level 1	– Lowest level of the node importance	Hospital
Level 2	– Almost all of words in this level are main categories of complaints	Payment
		Right
Level 3	– Almost all of words in this level are sub-categories of complaints	Web
		Civil servant
Level 4	– Highest level of the node importance – Almost all of words in this level are keyword terms (Leaf nodes in ontology) – Keyword terms are specific definition	Medicine
		Impolite

XPath is used as a query language for this complaint ontology. Furthermore, WordNet is used for generating a set of synonym words related to the query term. This is helpful when the original extracted word could not map directly with node term of this complaint ontology.

For the web interface of filling a complaint, users should select the main category and sub-category of the complaint from the list box. This category data is used as a main keyword [3] for searching and retrieving the single path of the ontology tree. Afterwards, the extracted keywords (output of the extraction module) are checked with nodes (word concepts) in this single path. At least, one of all keyword terms mapped with the extracted keywords is acceptable, because the assumption that each main concept (category) has different keyword terms. As an example, Figure1 has {Complaint: Standard: Dispense: Medicine + Prescription}. This means 'Dispense' class has two keyword terms ('Medicine' and 'Prescription'). If the extracted keywords can be mapped with one of those keyword terms, topic identification module will finish processing and set this selected sub-category as the complaint's topic. This is the first case for mapping extracted keywords with ontology.

If those extracted keywords can not be found in the above single path, another way is using extract keywords to be the main keywords in searching. The next step is to optimizing the topic concept (node). This is the second case that should be handled for this sub-module.

3.1.4 Optimizing the topic concept (node) in second case. For this sub-module, the main keyword for searching and optimize the topic node is the extracted keywords. After mapping the extracted keywords with ontology, the mapped keywords have their own level of node importance. The topic referral of each complaint comes from finding the mapped keyword which has the maximum level of that complaint, as it has highest probability for referring to the node topic. When the system finds that keyword, it will query the path from ontology again for referring to its sub-category. In case of finding more than two keywords which have the same maximum level from a complaint, the system will ask a user to select only one topic and add comments for admin users to read and check again. This prototype assumes that a citizen lodges only one topic of the complaint each time. Thus, finding more than one topic is considered in the future work.

The final output of the processing topic identification module is the "topic" of each complaint which equals to the sub-category within this complaint ontology. To assign a main category to be the topic node is too board, as to set a keyword term (Leaf node) to be the topic node is too specific. This shows that sub-category is suitable to be the topic. Figure 2 illustrates the process of finding topic from each complaint.

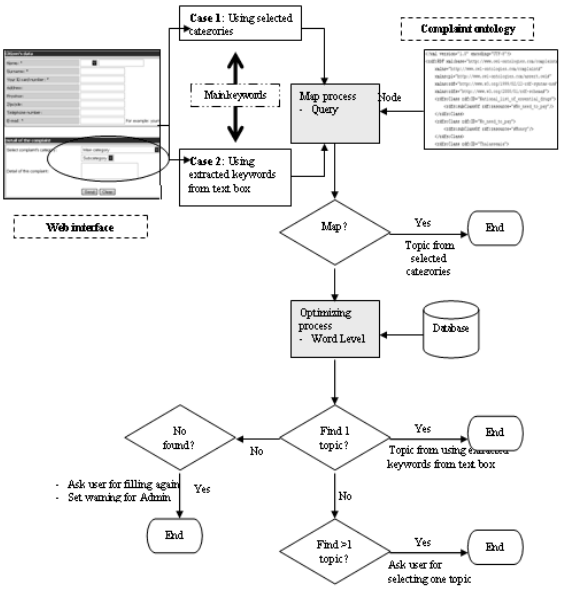

Figure2. Process of finding topic from each complaint

3.2 Classification module

In fact, the categories have already been classified in this complaint ontology. The topic of each complaint is assigned from one sub-category of this complaint ontology. When the topic has already been set into complaint document, the proposed system can refer to its main category from this ontology and classify similar topics together followed by its main category.

3.3 Ranking module

From data collection, the response time for citizens should be approximately within seven days. In addition, each daily report shows complaints posted in the current day and old complaints that have not been considered or solved yet. Therefore, this prototype should consider at what date the complaint was posted.

Ranking is the last main process of this prototype and produces the daily report that lists the priority of each category. Simple Multi-Attribute Rating Technique (SMART) is used to rank the priority of categories. Besides, two selected criteria are identified and used with SMART: priority of

categories and time period. Each criterion has its weight and sub-criteria. The weight of each criterion comes from comparing importance of criteria. The criterion, which has highest value (importance) among them, provides the highest weight from calculation. In addition, an admin user can change or set weights of criteria and priorities of sub-criteria by filling numerical values or score (values of 0 – 100) into an admin user interface.

The way to calculate preference ranking of many alternatives is by using a Matrix table; the alternative is the belonging category of complaints which citizens lodged, and the office has not considered yet. The preference value will be recommended to staff for consideration and solving that alternative (problem) first. Table2 shows an example of a Matrix Table for calculating SMART value. In addition, the examples of data which are used in this example of calculation (Table2) are listed below:

- Complaint about inconvenient service (2 complaints): one is 21 days and another is 1 day since users lodged these complaints.
- Complaint about the rights (1 complaint): 22 days since a user lodged this complaint.

4. Result

4.1 Result of testing

This prototype has been tested with 50 complaints that citizens lodged on 155 node categories plus keyword terms of the complaint ontology. Precision equation is selected to measure the accuracy of this proposed system as expressed Equation 1 [12]:

$$\text{Precision} = \text{Hits} / (\text{Hits} + \text{Mistakes}) \quad \text{Equation 1}$$

where, Precision is the result of accuracy

Hits are the relevant retrieved results

Mistakes are the retrieved results that are not relevant

Table2. An example of Matrix Table for calculating SMART value

Alternatives	**Criteria (Weight)**		**Total**	**Rank**
	Category	**Time period**		
	0.625	0.375		
Convenience	0.7	(1.0+0.4) /2	0.7 (0.4375 + 0.2625)	2
Rights	1.0	1.0	1 (0.625 + 0.375)	1

Table3. Result of testing system of each main module

	Three main module		
	Finding topic	**Classification**	**Ranking**
Percentage of results by calculating Precision equation	59%	59%*	70%
Average Total	64.5%		

*The output of finding topic module is the input of the classifying module. So, both modules provide same percentage of result.

From calculations using the Precision Formula (see Table3), 59% success in finding topic was obtained. This percentage is equal to the precision value of classifying process, because assigning a wrong topic leads to error in classifying the topic as well. Besides, accuracy at 70% comes from the ranking part which is tested individually in order to check the relevant retrieved result of SMART matrix calculation. Finally, the average precision percentage of the whole system is around 64.5%.

4.2 Analysis

Testing the prototype resulted in an error of around 35.5%. This could be caused because of many related factors. The first factor is the existing complaint that has ambiguous meaning, such as using wrong spelling, non-structured writing, and using informal words or slang. Another factor is keyword terms in the complaint ontology that are not adequate in explaining and referring to topic nodes or defining wrong keyword terms. The cause of unsuccessful topic identification due to the second factor could be because of the fact that the system can not find keyword terms. Therefore, defining keywords is very important process and this requires expertise in this compliant domain for developing the relevant complaint ontology about “30 Baht Co-payment Scheme”.

4.3 Limitation of a Prototype of Complaint Management System

From prototype testing, it shows that the accuracy of the prototype system depends on the keywords that we defined in this complaint ontology. If appropriate keywords are defined, the system will assign a wrong topic or will not find the relevant one. In addition, this prototype does not extract the meaning from each complaint. Therefore, if the original extracted keyword cannot be mapped, it is difficult to find the relevant synonyms of each extracted keywords from WordNet and to map them with the keyword terms in this complaint ontology. This is another limitation of this prototype.

5. Conclusion

This study presents the prototype of integrated complaint management system with important features of the grouping and ranking functions. The proposed system is also based on topic identification using the complaint ontology to assign topics automatically to each complaint. This ontology is built to collect the existing categories and keywords, as the topic of complaint among others. Besides, using ontology helps gather same topics of the complaint. If a topic was set, then the system can refer to its main category from ontology or hierarchical structure. Concerning ranking, SMART (Ranking technique) is used to rank main categories according to their priorities (to consider and solve the problems) and provide daily report to the staff in form of tables. These features will improve the work efficiency of staff.

From prototype testing, the average precision percentage of whole system is around 64.5%. The part which leads to errors is finding topic by using the complaint ontology. Further studies are required to analyze the related keyword terms related to the domain of this complaint ontology and specialist who has enough expertise in this complaint domain should be involved to give an idea about building this complaint ontology. After implementation, we found that the main problem is the existing complaint data which often have ambiguous words and structures.

In future, the performance of this proposed system could be improved, if some features are added. These are Thai language support, function for adding more keywords into this complaint ontology from admin user interface, and assigning more than one topic to the complaint in case of complaint addressing many problems. Besides using ontology for topic identification, Neural Network could be used to find the relevant topic of document for the future research work.

6. References

[1] B. Bigi, A. Brun, J.P. Haton, K. Smaili and I. Zitouni, "A Comparative Study of Topic Identification on Newspaper and E-mail", *String Processing and Information Retrieval*, Nov. 2001, pp. 238-241.

[2] C.Y. Lin, "Robust Automated Topic Identification", PhD Thesis, University of Southern California, 1997.

[3] C.Y. Lin and E. Hovy, "Identifying Topics by Position", *In proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP--97)*, 1997, pp. 283-290.

[4] C.Y. Lin, "Knowledge-based Automatic Topic Identification", *In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL--95)*, 1995, pp. 308-310.

[5] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K.J. Miller, "Introduction to WordNet: An On-line Lexical Database", *International Journal of Lexicography*, 3(4), 1990, pp. 235-244.

[6] J. Fülöp, "Introduction to decision making methods", Working paper of the Laboratory of Operations Research and Decision Systems. (LORDS) WP05-6, 2005.

[7] K. Lagus and J. Kuusisto, "Topic identification in natural language dialogues using neural networks", *In Proceedings of the 3rd SIGdial workshop on Discourse and dialogue - Volume 2*, 2002, pp. 95-102.

[8] K. Seymore and R. Rosenfeld, "Large-scale Topic Detection and Language Model Adaptation", Technical Report CMU-CS-97-152, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1997.

[9] L. Zhao, "The Integration of Geographical Information Systems and Multicriteria Decision Making Models for the Analysis of Branch Bank Closures", PhD Thesis, University of New South Wales, Sydney, Australia, 2002.

[10] M. Horridge, H. Knublauch, A. Rector, R. Stevens, C. Wroe, "A Practical Guide to Building OWL Ontologies Using the Prot´eg´e-OWL Plugin and CO-ODE Tools Edition 1.0", The University Of Manchester, 2004.

[11] N. Charoenkijipaiboon, "Chat recognition for medical problem-based learning", Master Thesis, Asian Institute of Technology, Thailand, 2006.

[12] S. Tiun, R. Abdullah, and T.E. Kong, "Automatic Topic Identification Using Ontology Hierarchy", *In Proceedings of the Second International Conference on Computational Linguistics and Intelligent Text Processing*, 2001, pp. 444-453.

[13] W. Edwards and F.H. Barron, "SMART and SMARTER: Improved simple methods for multiattribute utility measurement", *Organizational Behavior and Human Decision Processes*, 1994, pp. 306-325.