A Proposal of the Method to Identify Adverse Effects Based on Topic Maps

Kazuya Yokokawa[a], Masaomi Kimura[a], Michiko Ohkura[a] and Fumito Tsuchiya[b]

[a]Shibaura Institute of Technology
3-7-5, Toyosu, Koto City, Tokyo, 135-8548 JAPAN

[2]International University of Health and Welfare
2600-1, Kitakanemaru, Ootawara City, Tochigi, 324-8501, JAPAN

ABSTRACT

Medical experts need to properly identify the adverse effects. However, this is difficult for novice medical experts, since such identification need veterans' knowledge. One may consider reading of package inserts is helpful, since they are sole authorized documents. However, the novice medical experts will also feel difficulty, since the number of drugs is huge. In order to identify the adverse effect, patients' explanation about what they feel are usually helpful. However, the problem is that there is an semantic gap between patients' explanation and medical knowledge about adverse effects. The variations of patients' explanation also make it complicated to extract information about adverse effects from their words. In this study, we utilize information in drug information sheets and techniques of topic maps. We propose the method to extract adverse effect information from drug information sheets, to associate it with patients' explanation, and to create a database which has the topic map structure. We assume three elements to describe both patients' symptom explanations and initial symptom descriptions in drug information sheets. Moreover, we propose the method that starts from the three elements and follows the associations to find adverse effects.

Keywords: Semantic Gap, Adverse Effects, Topic Maps

INTRODUCTION

When a patient takes some drug, he/she may suffer from illnesses caused by it. The illness is called as an adverse event. A drug not only has beneficial effects but also can have adverse effects. One of major causes of adverse events is the adverse effects. However, a cause of adverse events can be the patient's primary diseases and complications. In the case that an adverse event does not originate from the adverse effect of a drug, it is dangerous for a patient to decrease amount of drugs, since the beneficial effects of the drugs decrease and his/her disease can get worse. In the case that the cause is the adverse effects, medical experts, such as doctors, pharmacists and nurses should decrease or stop the drug.

From these reasons, medical experts need to identify the adverse effects. However, this is difficult for novice medical experts, since such identification need veterans' knowledge. One may consider reading of package inserts is helpful, since they are sole authorized documents. However, the novice medical experts will also feel difficulty, since the number of drugs is huge (e.g. There are more than 10,000 drugs in Japan).

In order to identify the adverse effect that a patient suffers from, his/her explanation for feelings are usually helpful. However, the problem is that there is an semantic gap between patients' explanation and medical knowledge about adverse effects (e.g. information described in package inserts). The variations of patients' explanation also make it complicated to extract information about adverse effects from their words. Therefore, we need a specific method to associate patients' explanation with adverse effects.

In this study, we utilize information in drug information sheets and techniques of topic maps. A drug information sheet is an explanatory leaflet for patients about adverse effects. A topic map has a data structure to express relationships between concepts by means of a graph structure. We propose the method to extract adverse effect information from drug information sheets, to associate it with patients' explanation, and to create a database which has the topic map structure. We assume three elements, an objective element, a state element and an extent element, to describe both patients' symptom explanations and initial symptom descriptions in drug information sheets. We propose the method that starts from the three elements and follows the associations to find adverse effects. In our experiment, we confirm that adverse effects can be properly derived from the topic map database.

UTILIZED DATA IN THIS STUDY

Drug Information Sheet (RAD-AR Council, Japan, 2014)

In this study, we utilize the descriptions in drug information sheets disclosed by pharmaceutical companies in Japan. Drug information sheets describe adverse event information in the way patients can easily understand. They show the name of adverse effects accompanied by their initial symptoms. In the description, there are two formats for adverse event information: one is a basic format applied to whole of drugs (RAD-AR Council, Japan, 2011) and another is a format only for injection drugs (Shiga, 2002). In the basic format, square brackets are used to enclose the name of adverse effects. In the format for injection drugs, angle brackets are used to enclose the name of adverse effects. The followings show samples of the two formats in drug information sheets.

The basic format of adverse event information in drug information sheets

出血が止まりにくい、体がだるい、発熱、のどの痛み[骨髄抑制、溶血性貧血]
(Bleeding tendency, dullness, fever, sore throat[bone marrow depression, haemolytic anaemia])

The format of adverse event information for injection drugs in drug information sheets

<間質性肺炎> 熱がでる、痰がからまない咳、呼吸をしにくい、頭が痛い、全身がだるい。
(<interstitial pneumonia> fever, dry cough, breathing difficulty, headache, general dullness)

In drug information sheets, adverse effects have already corresponded to initial symptoms. Since descriptions of initial symptom are similar to patients' explanation, we can easily relate the descriptions to patients' explanation. Therefore, Relationships between the descriptions and patients' explanation and relationships between the description and adverse effects give us the bridge to connect patients' explanation and adverse effects. Figure 1 shows schematic explanation of these relationships.

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2093-0

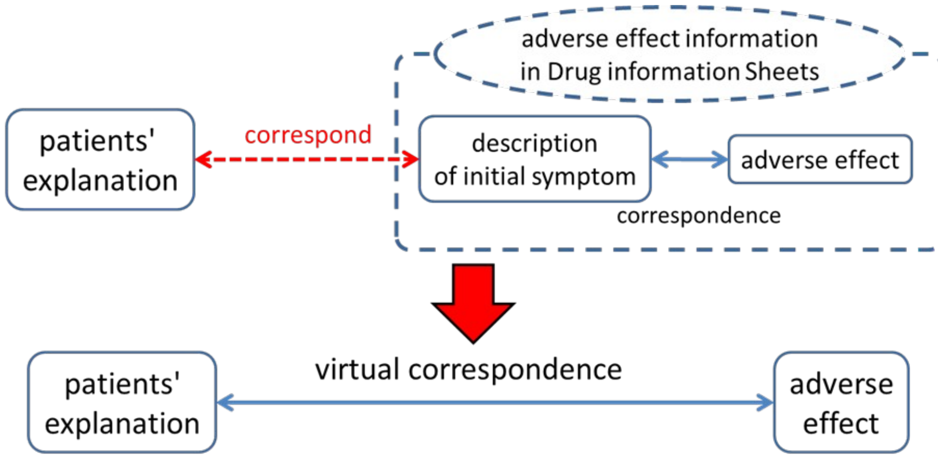

Figure 1. Relationship between patients' explanation and adverse effects

DATA STRUCTURE

Topic Map (Naito, Kato, Kiriyama, Komachi, Setogawa, Nakabayashi, Yoshida, 2006)

A topic map is a ontorlogical data structure to express concepts and their relationships, defined in ISO13250:2003, *Topic Maps*. A topic map mainly are composed of two elements: topics and associations. A topic represents a specific and abstract concept, and have names and types as its property. An association represents a relationship between topics and can also have its types as properties. Figure 2 shows an simple example of a topic map. This topic map expresses that a cough suppressant affects cough and have headache as an side effect.

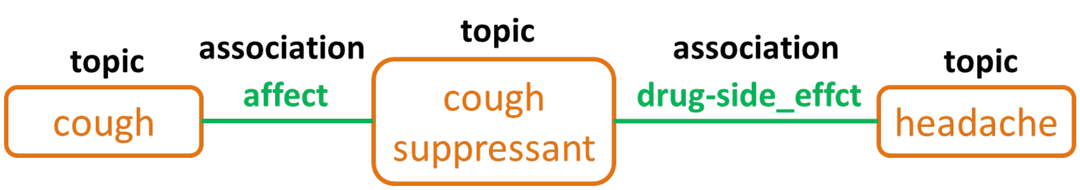

Figure 2. A sample of Topic maps

ELEMENTS COMPOSING A SYMPTOM

The Objective Elements, the State Elements and the Extent Elements

In this study, we assume we need three elements, an objective element, a state element and an extent element, to describe both patients' symptom explanations and initial symptom descriptions in drug information sheets. The objective element is the body region where the symptom was developed or the discharged matter resulting from the symptom (e.g. *head, body, blood* and *urine*). The state element is a word with the meaning of a symptom, such as *pain, tired* and *bleed*. The extent element is a word expressing the extent of a symptom such as an onomatopoeia. An onomatopoeia is a word that phonetically expresses the state of things, which is usually a source of sound. In Japanese, the onomatopoeia "がんがん" expresses the loud noise and/or its source. It is also used to express hard head ache, because we feel as if some imaginary loud noise sourse causes it. We, therefore, assumed an onomatopoeia is often used to express the extent of a symptom.

We obtained a set of Japanese terms appropriate to the three elements from two Japanese dictionaries, *Japanese Lexicon* (NTT Communication Science Laboratories, 1997) and *Japanese onomatopoeia dictionary* (Ono, 2007). Japanese Lexicon, a dictionary that classifies Japanese words into 3000 semantic classes, contains 1767 words appropriate to an objective element in the class, *Animal / Body Part*. It also contains 1057 words appropriate to a state element in the classes, *Change in body*, *State of Consciousness* and *Body conditions*. We obtained 173 onomatopoeias appropriate to an extent element from Japanese onomatopoeia dictionary.

A METHOD TO RELATE PATIENTS' EXPLANATION TO ADVERSE EFFECTS

A Method to Relate Patients' Explanations with Initial Symptom Descriptions

We utilize the objective and the state element to relate patients' explanations to initial symptom descriptions. We apply morphological analysis to a description of initial symptom and obtain a series of words. If each obtained word is in the lists of an objective element or an state element, we extract it as an objective or state element of the description. If patients' explanation share both the objective element and the state element with an initial symptom description, we regard that they are mutually related. Figure 3 schematically shows this method.

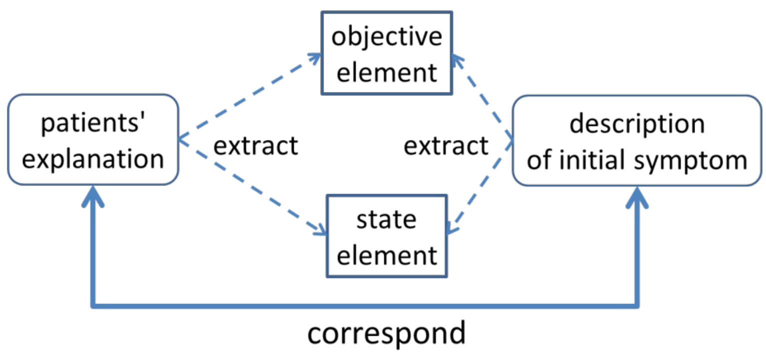

Figure 3. Relationship between patients' explanations and descriptions of initial symptom

Associating Onomatopoeia to Correspondences

In order to add information about the extent of a symptom to relationship between patients' explanations and adverse effects, we associate extent elements not with descriptions of initial symptom but with adverse effects. This is because, even if adverse effects cause the same symptom, the extent of the symptom can be different. Figure 4 shows the way to associate the extent element to an adverse effects.

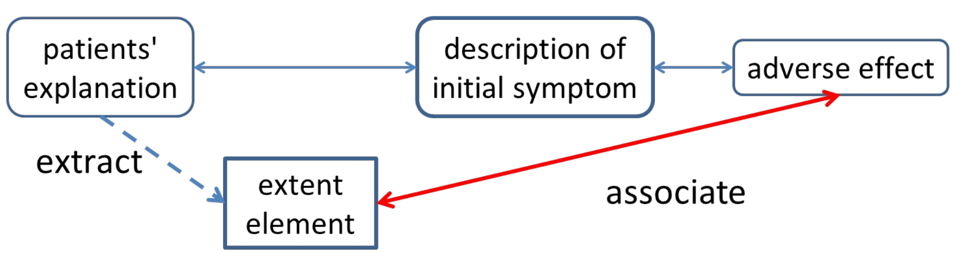

Figure 4. The way to associate the extent element to the name of adverse event

In this study, we utilized a web search engine to measure strength of relationship between an extent element and an

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2093-0

adverse effect. We obtained the number of search results of an extent element and an adverse effect, and applied Simpson’s coefficient (Matsuo, 2005) to it. The Simpson’s coefficient is defined as follows:

$$Simpson's\,Coefficient = ¿A \cap B \vee \frac{¿}{min\ ¿¿}¿$$

where $|A|$ and $|B|$ are the number of searched web pages containing the extent element the adverse effect respectively. The $¿A \cap B \vee ¿$ denotes the number of searched web pages containing both the extent element and the adverse effect. we associate the pairs of the extent element the adverse effect with the value of Simpson’s coefficient larger than the threshold defined in this study. In order to decide the value of the threshold, we randomly chose extent elements and adverse effects and calculated their values of Simpson’s coefficient. Based on the results, we determine the value of threshold to be 0.01.

The Process to Create Topic Maps of Correspondences

We define 6 topic types: *drug*, *side_effect*, *symptom*, *objective*, *state* and *extent*, and 5 associations: *drug-side_effect*, *causal_relationship*, *hasPart_objective*, *hasPart_state* and *cooccurrence_relationship*. Table1 shows the meanings of the defined topic types and their examples. Table2 shows the defined associations and the types of topics which the associations have at their terminal.

Table 1: The types of topics

Types of Topics	Meaning	Examples
drug	brand name of drug	‘’エヌケーエスワン配合カプセル(NKS-1 combination capsule)
side_effect	name of adverse event	‘骨髄抑制’(bone marrow depression)
symptom	Symptom	‘のどの痛み’(sore throat)
objective	objective element	‘のど’(throat)
state	state element	‘痛み’(pain)
extent	extent element	‘ずきずき’(zukizuki, an onomatopoeia expressing hard pain)

Table 2: Associations and the types of topics which they have at their terminal.

Associations	The types of topics	
drug-side_effect	drug	side_effect
causal_relationship	side_effect	symptom
hasPart_objective	symptom	objective
hasPart_state	symptom	state
cooccurrence_relationship	side_effect	extent

The procedure to create a topic map is as follows:

1. We extract the brand names of drugs and the names of adverse effects from drug information sheets, and create *drug*-type topics whose names are the brand names and *side_effect*-type topics which have the names of adverse effects. We connect them with a *drug-side_effect* association.

2. We extract initial symptom descriptions from drug information sheets, and create their *symptom*–type topics. We connect them to their related adverse effects by *causal_relationship* associations.

3. We extract objective elements and state elements from the initial symptom descriptions, and create *objective*–type topics corresponding to the objective elements and *state*-type topics corresponding to the state elements. Then, we connect the *symptom*-type topic to the *objective*-type topic by a *hasPart_objective* association and to the *state*-type topic by a *hasPart_state* association.

4. We create *extent*-type topics corresponding to onomatopoeias obtained from Japanese onomatopoeia dictionary, and calculate Simpson's Coefficient to measure the strength between the onomatopoeias and adverse effects. If the values of Simpson's Coefficient are larger than 0.01, we link the *extent*-type topics to *side_effect*-type topics with *cooccurence_relationship* associations.

Figure 5 illustrates a topic map that expresses the topics and the associations obtained by the above procedeure.

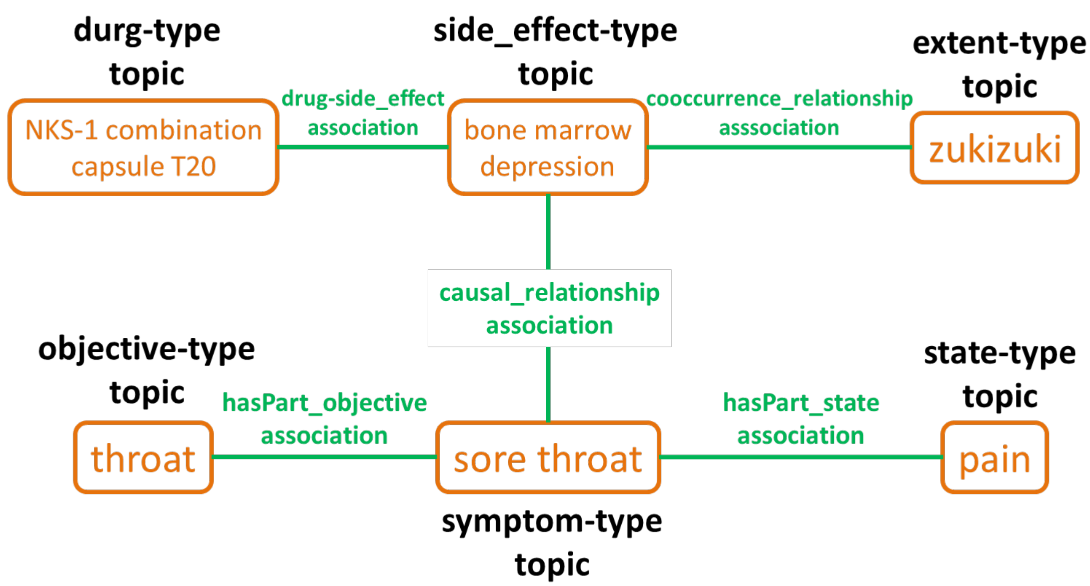

Figure 5. An example of obtained topic maps

The Process to Search Adverse Effects in a Topic Map

Input data are an objective element, a state element, an extent element and a drug name extracted from a patient's explanation.

The procedure to search adverse effects in the topic map is as follows:

1. In the target topic map, we search the topics corresponding to an objective element, a state element, an extent element and a drug name extracted from a patient's explanation.

2. We search symptom-type topics linked to the objective topic by hasPart_objective associations and linked to the state-type topic by hasPart_state associations.

3. We search side_effect-type topics linked to the obtained symptom topics by causal_relationship associations.

4. We limit side_effect-type topics obtained in Step 3 to the ones connected to the extent-type topic by coocurrence_relationship associations.

5. If the side_effect-type topics obtained in Step 4 are connected to the drug-type topic in Step 1 by drug-side_effect associations, the name of the resultant side_effect-type topics is what should be output as adverse effects .

Figure 6 illustrates this procedure.

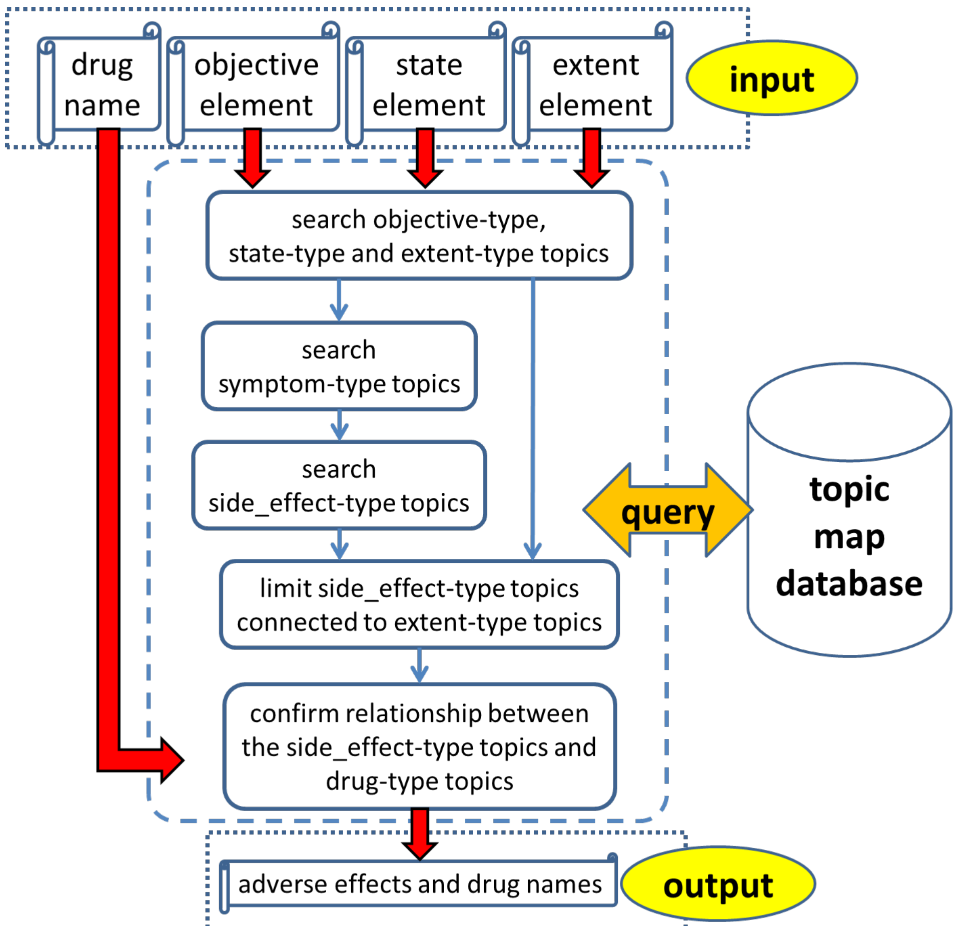

Figure 6. Process to search adverse effects in Topic Maps

EXPERIMENTS

Taget Data and Method

In order to verify that we can extract adverse effects of a drug from patients' explanation, we input patient explanation data about adverse events dislosed in the Internet into the system implementing our method.

Target data are 19 sentences about symptoms occurring after taking medicines, which are disclosed in the web site, *YAHOO ANSWERS* in Japan. We extracted three elements and a drug name from the sentences and input them to the system.

An example of the way to extract input data from the patient's explanation is as follows:

The input sentence: 'タミフルを飲むと、こめかみがガンガン痛む'(When I take "Tamiflu", I feel a "gangan" pain in my temple)

1. Extract 'タミフル'(Tamiflu) as a drug name from the patient's explanation

2. Refer to the drug information sheet of 'タミフルカプセル'(Tamiflu Capsules)

3. Check the adverse effects information on the drug information sheet

4. '頭痛'(headache) in the adverse effects information is similar to 'こめがみがガンガン痛む'(I feel a "gangan" pain in my temple) in the patient's explanation.

5. Extract 'こめかみがガンガン痛む'(I feel a "gangan" pain in my temple) as the symptom from the patient's explanation

6. Extract 'こめかみ'(temple) as the objective element from the patient's explanation

7. Extract '痛み'(pain) as the state element from the patient's explanation

8. Extract 'ガンガン'(gangan) as the extended element from the patient's explanation

9. The input data is 'タミフル'(Tamiflu), 'こめかみ'(temple), '痛み'(pain)' and 'ガンガン'(gangan)

We obtained 43 input data from the 19 sentences. If the elements of the input data don't exist in topic map, we substituted the elements to words similar in meaning (e.g. replace ‘temple’ to ‘head’). A Target topic map is created from 320 drug information sheets.

We input the two kinds of data into the system using the topic map. We list the search results of adverse effects of drugs from output data and compare the number of the result. After we show the results, we discuss the cause why the system could not search adverse effects on the input data changed.

RESULTS

Table 3 shows the results of the experiment. According to the table3, we obtained adverse effects associated with drugs from 30% of the original input data, which we did not substituted to the words similar in meaning in the topic map. On the other hand, we could obtain the adverse effects from 72% of the substituted input data. This shows the necessity to include thesaurus in the topic map and to add a function to change the words of the input data to relevant words in the topic map automatically.

Table 3: The result of this experiment

Input data	The number of data not to get adverse effects	The number of data to get adverse effects	
		Not to associate with the drug	To associate with the drug
Extracted from sentence about symptoms occurring after taking	22data	8data	13data
Changed to words similar in meaning	5data	7data	31data

According to the results of subsituted input data, we can confirm 5 data which we cannot obtain adverse effects from. Since some patients' explanation only have extent elements and we cannot search adverse effect by using the process to search adverse event in this study. Some onomatopoeia in Japanese contain state elements. For this reason, we need to consider the process to search by only extent elements. Moreover, we can confirm 7 data which we cannot extract adverse effects associated with drugs from. We consider that one of the reasons is that the threshold to associate adverse effects with extent elements is not suitable. Though some adverse effects (e.g. fever and headache) are used in colloquial expression and easily exist with onomatopoeia, others (e.g. bone marrow depression) hardly exist with onomatopoeia. We need to set thresholds depending on each adverse effects.

CONCLUSIONS

In this paper, we proposed the method to extract adverse effect information from drug information sheets, to associate it with patients' explanation, and to create a database which has the topic map structure. We assumed three elements to describe both patients' symptom explanations and initial symptom descriptions in drug information sheets. Moreover, we proposed the method to follow the associations to find adverse effects.

In order to create a topic map, we extract the brand names of drugs, the names of adverse effects and initial symptom descriptions from drug information sheets, and create corresponding topics and associations. Moreover, we create topics corresponding to onomatopoeias obtained from Japanese onomatopoeia dictionary, and create associations between the onomatopoeias and adverse effects under the condition that the values of Simpson's Coefficient are larger than 0.01. The key idea is to directly connect onomatopoeias and adverse effects. This is because adverse effects may cause different extents of symptom even if the symptom is same.

In order to seek adverse effect information, we search topics expressing symptoms linked to the input topics expressing a body part and its state. We search adverse effect topics linked to the obtained symptom topics and limit them to the ones connected to the topic expressing the extent and the topic expressing the input drug.

In the future study, we will extend the topic map obtained in this study to the one applicable to the whole drugs. The variety of patients' expression about symptoms should be absorbed by adding synnonimous words to the topic map. Since the topic map is expected to be a huge one, we will need effective topic map database management system.

REFERENCES

Matsuo, Y. (2005), "*Social Network Extraction from the Web information*", Transactions of the Japanese Society for Artificial Intelligence, Volume 20 No. 1. pp. 46-56

Naito, M. Kato, H. Kiriyama, T. Komachi, Y. Setogawa, M. Nakabayashi, K. Yoshida, M. (2006), "*An introduction to topic maps*", Tokyo Denki University Press

NTT Communication Science Laboratories. (1997), "*GoiTaikei --- A Japanese Lexicon*", Iwanami Shoten

Ono, M. (2007), "Japanese onomatopoeia dictionary", shogakukan

RAD-AR Council, Japan. (2011), "*Kusuri-no-Shiori Basis of Preparation*"(3rd ed.), Kusuri-no-Shiori Website: http://www.rad-ar.or.jp/siori/sioriclub/document.html

RAD-AR Council, Japan. (2014), *Kusuri-no-Shiori*; *Drug information Sheet*. Kusuri-no-Shiori Website: http://www.rad-ar.or.jp/siori/

Shiga, N. (2002), "*Kusuri-no-Shiori Injection edition; Format and Basis of Preparation*", RAD-AR News, Volume 12 No. 6. pp. 12-13