International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:5, No:9, 2011

Method for Concept Labeling Based on Mapping between Ontology and Thesaurus

Kazuki Sonoda and Masahiro Hori

***Abstract*—**When designing information systems that deal with large amount of domain knowledge, system designers need to consider ambiguities of labeling termsin domain vocabulary for navigating users in the information space. The goal of this study is to develop a methodology for system designers to label navigation items, taking account of ambiguities stems from synonyms or polysemes of labeling terms. In this paper, we propose a method for concept labeling based on mappings between domain ontology andthesaurus, and report results of an empirical evaluation.

***Keywords*—**Concept Labeling, Ontology, Thesaurus, Vocabulary Problem

I. Introduction

DUE to the rapid growth of information society, the amount of information people can accessfrom web sitesandinformation systemsisexponentially increasing. To make it easier for users toreach informationthey need, it is importantto provide navigation labels such as headings and menusthat facilitateusers'understandingof what labels means[1]. In such a situation, system designers need to consider ambiguities stems fromsynonyms or polysemesof labeling terms. Users maynot understand an intended meaningofa polysemouslabel when they are not well acquainted with thedomain knowledge.In such situations, users may fail to forecast contents indicated by item labels, and expectinformation different from the one intended by the system designer [2].

To helpsystem designers to label navigation items withappropriatedomain vocabulary, we propose a method for conceptlabeling based on mappings between ontology[3] and thesaurus[1].In our method, on the basis of an ontological engineering method[3], domain ontology is constructed to deal with in-depth semantics of domain concepts for a target information system. In addition, thesaurus is prepared to provide a collection of terms familiar to users. Items in these two constructs are then mapped each other to clarify relationships between termsand concepts. These mapping relations are exploited for assisting system designers to select appropriate domain termsfor users when labeling conceptsfor navigating information items.

K. Sonoda is with Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292,JAPAN (e-mail: sonoda@jaist.ac.jp).
M. Hori is with Kansai University, 2-1-1 Ryozenji-cho, Takatsuki, Osaka 569-1095, JAPAN (e-mail: horim@res.kutc.kansai-u.ac.jp)

II. Ambiguity of Vocabulary

In general, the terms have various features in representations and meanings like synonym and polyseme[4]. Polysememeansa term that expresses two or more concepts.Synonymmeanstwo or more terms thatexpress common concept. Relations betweenterms and concepts would not be so straightforward since they depend on various contexts. Explicit contexts include academic, technical and special domain, whileimplicit contexts include individual experience, property, or region. For example, aninformation system that deals with disaster-prevention knowledge for general users hasa domain term*flood*, whichentails two meanings: rise of river and overflow of river (Fig.1).

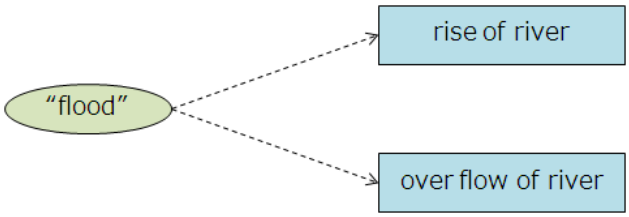

Fig.1 Polysemous word*flood*

In an information system, whena polysemous word*flood*is used in the sense of rise of river, there might be a possibility of providing misleading information ifa user takesthe meaning as overflow of river (Fig.2). This is the reason why system designers need to selectan appropriate label for navigation designtaking account of the variations and ambiguities in the domain.

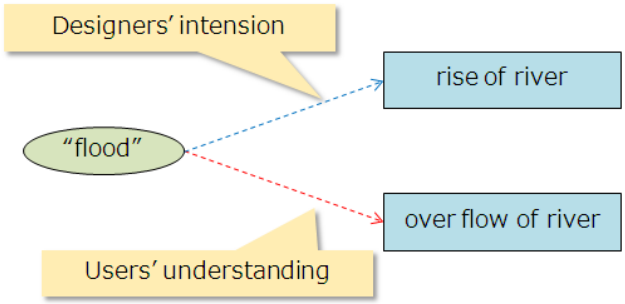

Fig.2 Semantic mismatching between designers and users

III. Approach

We propose a method for conceptlabeling based on a mapping between ontology and thesaurus (Fig. 3). By means of

an ontological engineering method, fundamental concepts are elicited from domain knowledge, and concepts andrelationships of those conceptsare defined without regard to lexical labels[3]. Although the terms expressing concepts are diverse, concepts behind terms are canonical since they are prescribed as domain ontology. Domain ontology is organized here with anupper ontology that consists of general concepts that are not specific to the domain at hand[3].In addition to the domain ontology, to deal with domain-specific terms, we construct thesaurus that is a systematization of termsorganized by means of relations between terms such as narrower, broader, synonymous,and polysemous[1] (Fig.4). Thesaurus is a collection of domain term, and is concerned with neither concept definitions nor semantic relations between concepts behind terms.Each domain terms in the thesaurus is mapped to a corresponding conceptin the domain ontology, so that the terms can be given with semantic definition.In particular, synonymous terms are mapped to a common concept, and a polysemous term is mapped to more than one concept. According to this mapping, relationships between terms and concepts are made explicit.It is thenexpected that the mappingallows system designers to remind ways users understandor confuse meaning behind the terms, and help the designers to select more adequate navigation labels.

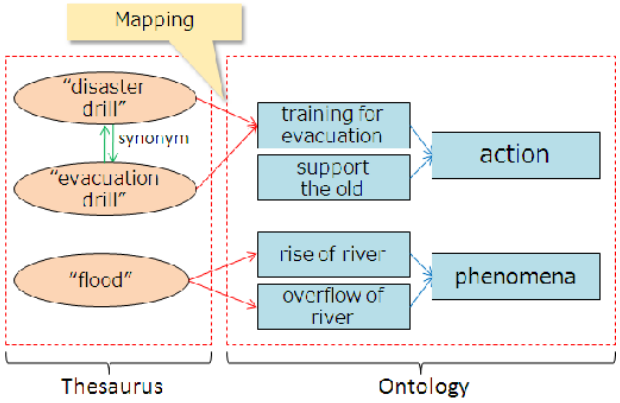

Fig.3 Overview of the mapping approach

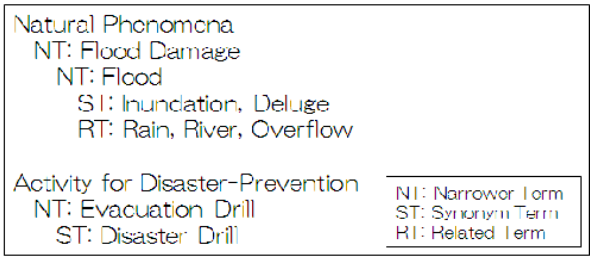

Fig. 4 Example of thesaurus

Relations between terms and concepts are diverse. However, concepts correspond to polysemousterms are specialized in individualsituation, and their meanings can be clarified in relation to context-dependent concepts(called *role concepts*[5]).The role concept, therefore, is a key to cope with the multiple, potentially ambiguous relations between the terms and concepts. Taking account of the role concept and *is-a* relation in the representation of ontology, we extracted essential relationships between terms and conceptsas possible configurations, and classified them into 8 patterns (Figs.5 and 6). In our method, an ontology developer establishes the ontology-thesaurusmapping by using thesemapping patterns.

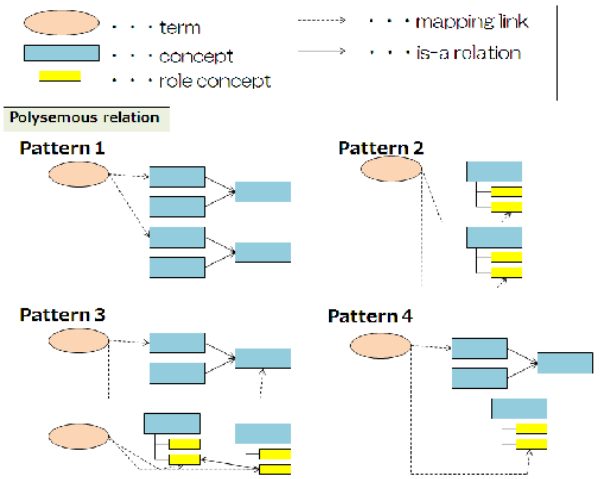

Fig.5 Mapping patterns for polysemous relation

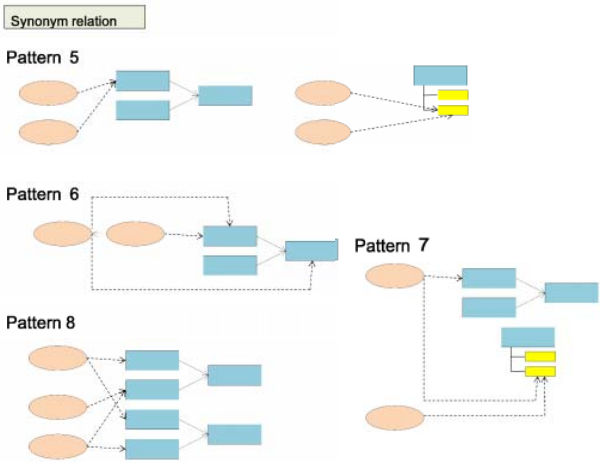

Fig. 6 Mapping patterns for synonymous relation

Fig.7 showsan example in disaster prevention domain. That is mapping between a term *flood* and concepts‘flood’ and ‘inundation.’‘Flood’whose meaning is rise of river is a role conceptthat depends on a ‘prevent flood’ task.This role concept is specialized in the sense that a general phenomenon‘rise’occurredparticularly at river and becomes a subject of prevention task. ‘Rise’ is defined as a process that a certain substance increases in a certain space. ‘Inundation’whose meaning is overflow of river is also a role concept that depends on ‘flood’ concept. This role concept is specialized in the sense that a general ‘overflow’phenomenon is occurred at river, which is different from ‘flood’.‘Overflow’

is defined that certain amount of substance goes out of a space when the amount of substance exceeds the capacity of the space. In the thesaurus side, term *flood* has multiple meanings, which are rise of river and overflow of river. A term *inundation*also has the same meanings.Furthermore, another meaning of theterm *inundation*is simply 'overflow'. Therefore, they are both mapped to 'flood' and 'inundation' concepts and term *inundation* is also mapped to the concept of 'overflow' in ontology. These mapping relations fall into the patterns 2 and 4depicted in Fig. 5.In this way, the term *flood*leads to both rise of river and overflow of river, and the meaning of the latter is expressed by the term *inundation* as the synonym of term *flood*. It is important to note here that the mapping patterns allow system designers to pay more attention to ambiguityof terms, and provide opportunities of clarifying the relations between terms and concepts.

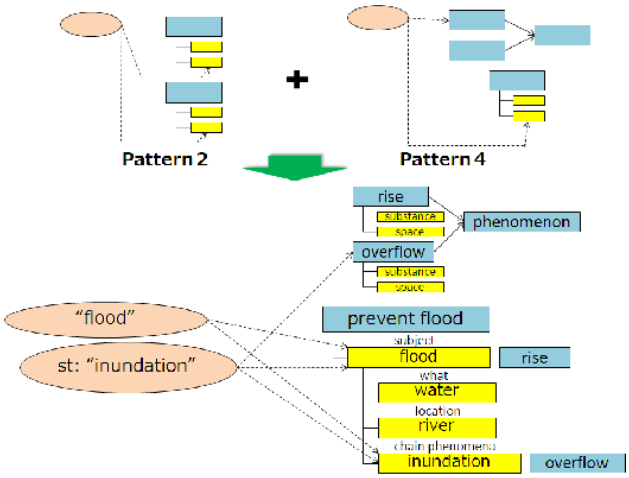

Fig.7 Mapping example

IV. Evaluation

We conductedempirical evaluation as a user study to investigate the two points: (1) designers' prerequisite knowledge and understanding on ontology engineering required touse the proposedmethod, and(2) usefulness of the proposed concept- labelingmethod.

A. Participants

Evaluation was conducted with 9 participants belong to Japan Advanced Institute of Science and Technology (JAIST). As depicted in Fig. 8, we considered two situations where system designers are given with a domain ontology in different manners: ontology with mapping to thesaurus (task condition A),and ontology without mapping to thesaurus (task condition B). As an evaluation method, think-aloud protocol method [6] was used.

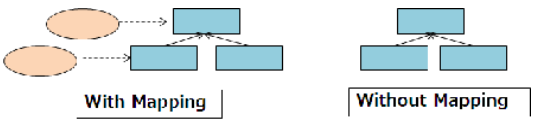

Fig. 8 Situations considered in the evaluation

In orderto avoid order effect, we dividedparticipants into two groups: one group begins with the condition A, and another begins with B.To balance theexperience period of ontological engineering, the participants are allocated to different experimental conditions considering their period of experience (TABLE I).

TABLE I
Allocation of Participants

Participant No.	Experience period of ontological engineering	Task execution order
1	Short (less than 2 months)	AB
2		
3		BA
4		
5	Medium (more than 3 months, and less than one year)	AB
6		
7		BA
8		
9	Long(6 years)	AB

B. Materials

We usedflood damage measures in disaster preventionas a domain of ontology and thesaurus forthis evaluation.The domain ontology was constructed with the upper ontology YAMATO [7] which is modeled based on the principle of role concept and Activity First Method (AFM) [8].The domain ontology here includes task concepts for theflood damage measures, as well as ways of taking measures relevant to the task concepts, flood damage phenomena, and general phenomena. Examples of the task concept hierarchy and task concept definition are shown respectively in Figs. 9 and 10.

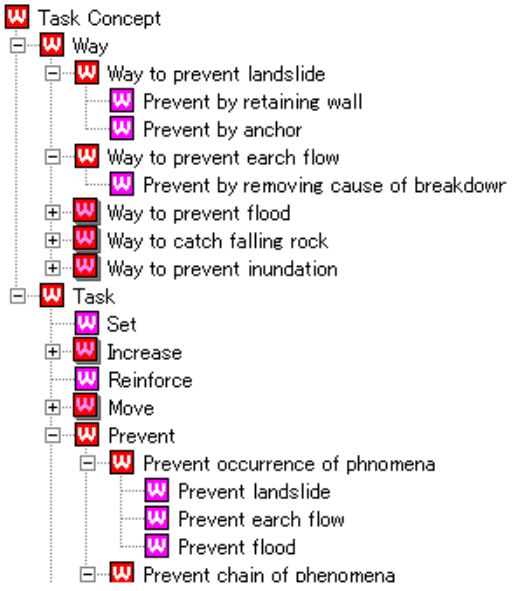

Fig. 9 Task concept hierarchy

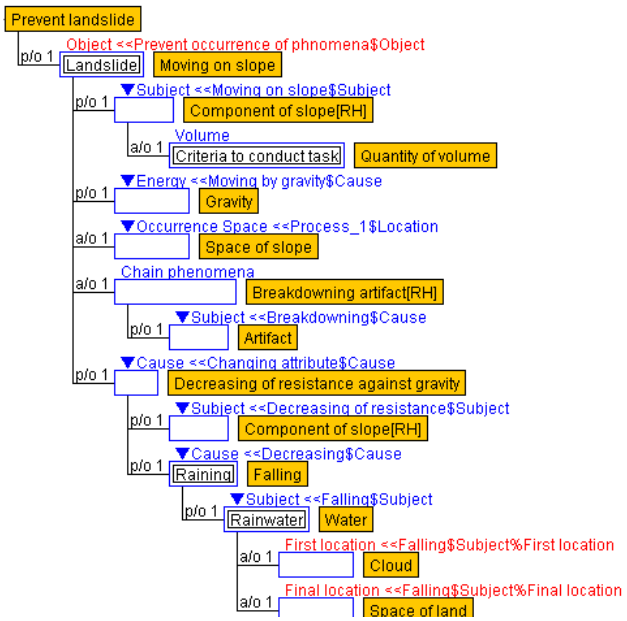

Fig. 10 Example of task concept definition

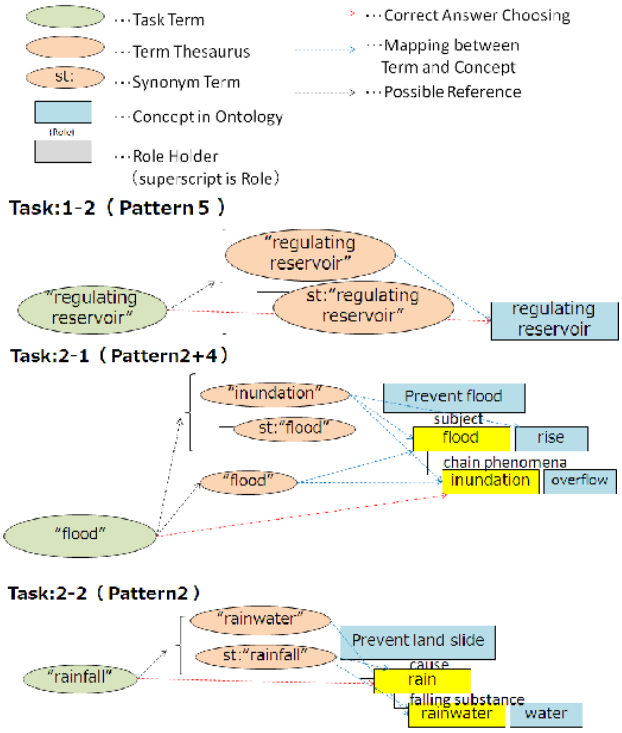

Fig. 11 Associations between task terms and concepts

A thesaurus was constructed following the construction guidelines [9]. Fig. 11 shows associationsbetween terms in the thesaurus and concepts in the ontology, which are created based on the mapping patterns in Figs. 5 and 6. Meanings of task terms are handed out as printed glossary and participants are allowed to see the glossary whenever they want.

C. Tasks

In evaluation tasks, participants are given task terms (TABLE II)chosen from the thesaurus, and asked to find concepts semantically correspond to those terms from the ontology.

TABLE II
TASK TERMS AND CONCEPT LABELS WHICH SEMANTICALLY CORRESPOND TO TASK TERMS

Task No.	Task term	Concept label
1-1	land slide	land slide
1-2	regulating reservoir	regulating reservoir
1-3	overflow of river	inundation
1-4	super bank	high-grade bank
2-1	flood	inundation
2-2	rainfall	rainwater
2-3	overflow	brim over
2-4	rain cloud	rain cloud

D.Equipment

We developed a viewer that runs on a web browser (Figs. 12 and 13), and made it available for all the participants. The viewer consists of anontology viewer and a thesaurus viewer.

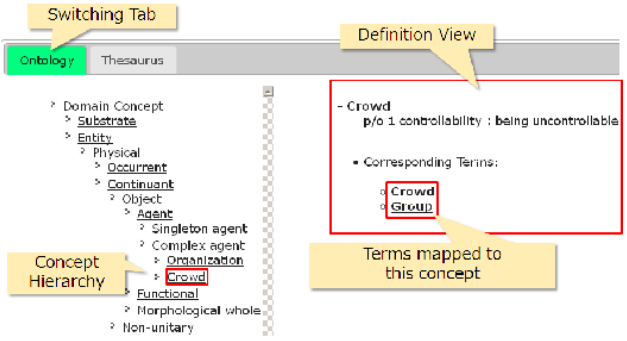

Fig. 12 Screen copy of Ontology viewer

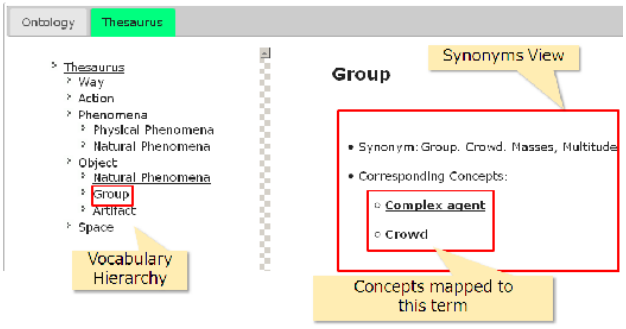

Fig. 13 Screen copy of Thesaurus viewer

These two view panes can be switched alternatively over by a switching tab.This viewer provides functions and user interface similar to existing ontology editors such as Hozo[10].All the evaluation sessions were videotaped with permission from the behind. Participants' utterances and behavior during the task execution were recorded, and used for protocol analyses following the think-aloud evaluation method.

E. Evaluation Procedure

Evaluation tasks were performed individually.After the explanations on the task execution procedure, we asked participants to think aloud, especially the points as follows.

a) Meanings that reminded from task terms
b) Reasons to narrow down target concepts
c) Reasons to select a concept

We then instructedparticipants to practice thinking aloud using a shopping web site. A sample task was to look for items. Actual evaluation tasks consist of 8 tasks, and the four tasks are provided thesaurus while remaining four tasks are done without thesaurus. Finally, weasked participantsfor their comments on the mapping method anddifficulty/ease of selecting target concepts in the interview session.

V. Results

It took 96minutesfor one participant to complete all the eight tasksin an average (evaluation tasks: 56 minutes, explanation and interview: 40 minutes).We analyzed the recorded utterance and behavior interacting with the viewer.Since the fourparticipantsdid notlook up the thesaurusat all during the evaluationsessions,the following analyses were made for the other five participants who referred to both the thesaurus and ontology.

A. Data Analysis

First, we transcribed utterances and extracted access log data of items and concepts manipulated by participants on the viewer screen. After the data collection, we classified types of utterances. The classification of utteranceswas made with regard to the types of interaction steps in the extended model of HCI[11].The results are shown in showed in TABLE III.

TABLE III
Correspondence of Utterances Contents Category to an Extended Model of HCI

Steps in Extended HCI model	Category of utterances	Explanation
Interpret the object	Reason	reason to select objects
	Choice	selectedobjects
Interpret the outcome	Forecast	forecast meaning of task term
	Interpretation	interpretation of the result at hand
Evaluate the outcome	Evaluation	evaluation of the entireresults

B. Results of Analysis

Relations between experience period of ontological engineering and the number of correct answers are given in TABLE IV.There were few differences in the number of correct answers.TABLE V shows the relations between experience period and the number of utterance types classified into the interpretation and evaluation.

TABLE IV
The Number of Correct Answers

Experience period of ontological engineering	Thesaurus	
	with mapping	without mapping
short (n=3)	2.7	1.3
medium (n=1)	3	2
long (n=1)	2	0

TABLE V
The Number of Utterances Types Classified into Interpretation and Evaluation

Experience period of ontological engineering	Thesaurus	
	with mapping	without mapping
short (n=3)	0.3	0.7
medium (n=1)	0	3
long (n=1)	4	3

The shorterexperience period of ontological engineeringparticipants have, the fewer utterances the participants made for the interpretation and evaluation. The participants with less experiences only talked about shallow reasons such that “a term with the label same as given task term was found,” even if they made utterances about evaluation. Further, in the interview session, we got comments about criteria of similarity between task terms and concept labels, and realized ways of selecting answers from thesaurus withoutlooking up and grasping concepts in the ontology. From these results, it is supposed that as the participants with less experiencesjudge relying merely on superficialsimilaritiesbetween terms and labels.One participant with medium experience made utterancesabout interpretation and evaluation only in the cases where the mapping was given. In cases without thesaurus mapping, it was not clear if the participant understood themeanings of concepts and evaluatedthe similarity appropriately.

On the other hand, the participant with the longest experiences made utterancesabout interpretation and evaluation. For example, a task term *regulating reservoir* in the task 1-2matches the mapping pattern 5(Fig. 6),and is contained in the thesaurus as a synonym of a term *balancing reservoir*. This term semantically corresponds to a concept 'regulating reservoir' which is mapped to the term *balancing reservoir*inthe thesaurus.All the participants chose the correct answer for this task. The participantwith the longest experience wasinspired by a concept label as well as the lexical similarity of the task term, and then judged that the concept 'regulating reservoir' semantically corresponds to the task term *regulating reservoir*. In contrast, all the other participants with less experiencejudged merely considering the lexicalsimilarity.

A task term *overflow of river* in the task 1-3matches the mapping patterns 2 and 4(Fig. 5), and semantically corresponds to a concept 'inundation' which is mapped to the term *inundation*inthe thesaurus.This is a role concept and specified

depending on the other role concept‘flood’. The participantwith the longest experiences came up with aterm*inundation*from the task term and looked upthe concept definition of 'inundation' mapped to the term *inundation*inthe thesaurus. However, the experienced participant onlylooked up the concept definition of‘flood’depended by the concept ‘inundation’ and did not regard itas over flow of river. Finally, the participant selecteda similar concept 'overflow' that is also mapped to *inundation* in the thesaurus.All the other participants with less experienceselecteda concept either ' inundation ' or 'flood', but they made that decision merely on the basis of the lexical similarity between a term *inundation* or *flood*came up with the task term and concept labels. In this task, the concept ‘flood’ had only the meaning of rise of river.A task term *rainfall* in the task 2-2matches the mapping patterns2 and 8 (Figs. 5 and 6), and semantically corresponds toa concept'rainwater'. In general, the term *rainfall*entails the meanings of both rain phenomena and water of rain. In the task 2-2, *rainfall* means the latter which corresponds to concept ‘rainwater’. However, in cases of the task condition without thesaurus mapping, the participants with short experience as well asthe participant with the longestexperiencecame up with the former meanings and selectedthe concept 'fall' and 'water'.The participants with short experience made that decision merely on the basis of thelexical similarity. Theparticipant with the longestexperience selected the concept based on the semantic similarity.

VI. Discussion

To use the proposed method, it isnecessary for system designers to have prerequisite knowledge and understandingon domain ontology. The shorterexperienced period of ontological engineeringthe participants have, the fewer concept interpretation and semantic term-similarity evaluation the participants made. It was observed that participantswith less experience relied merely on superficial lexicalsimilarity between task terms and concept labels. Therefore, it is probable that inexperienced designers cannotfind out semantic difference betweentermsthat has semantic similarity.Therefore, it would be beneficial for such designers to be provided with assistance of reminding semantically similar concepts and terms with tool support by means of the mapping patters presented in this paper.

It was found that the proposed method madeit possible for the participantwiththe longestexperienceto pay attention to multiple meanings or representations of terms, and that allowed the participantto come up with the reasonable selection in the case of terms with multiples meanings. In the task condition without thesaurus mapping, it was observed that participantscould not notice another meaning of terms.In theinterview session, the participant with the longestexperience said that he could understandthat similar termswere semantically different due to the thesaurus mapping. However,in complex casessuch thatrole conceptsare specifiedin relation to the other role concepts, the participant overlookedsubsequent role concepts and only looked upimmediate role.

VII. Conclusion

In order to help system designers to select appropriate domain terms for users when labeling concepts for navigating information items, we proposed a method for conceptlabeling based on mapping between ontology and thesaurus. As results of the empirical evaluation, it was confirmed that the proposed method should be provided with system designers acquainted with ontological engineering to some extent.In addition, we found that the proposed method would facilitate systemdesigners to notice multiple meanings or representations of terms. Since this evaluation was made only the small number of participant, further investigation is needed to make sure the above points,collecting more empirical data from a number ofparticipants. Moreoverit is necessary to conduct evaluation with different domains.

References

[1] L.Rosenfeld, and P.Morville, “Information architecture for the world wide web”, Third Edition, O’Reilly Media, Inc. 2006
[2] G.W.Furnas, T.K.Landauer, L.M.Gomez, and S.T.Dumais, “The vocabulary problem in human-system communication”, Comm. ACM, Vol.30, No.11, 1987, pp.964-971
[3] R. Mizoguchi, “Tutorial on an ontological engineering - part 1: introduction to an ontological engineering”, New Generation Computing, OhmSha&Springer, Vol.21, No.4, 2003, pp.365-384
[4] J.Sowa, “Knowledge representation: logical, philosophical, and computational foundations”, Course Technology, 1999
[5] R.Mizoguchi, “Tutorial on an ontological engineering - part 3: advanced course of an ontological engineering”, New Generation Computing, OhmSha&Springer, Vol.22, No.2, 2003, pp.198-220
[6] K. A. Ericcson, and H. A.Simon,“protocol analysis: verbal reports as data”, revised ed, Cambridge, MA: MIT Press, 1993
[7] R.Mizoguchi, “YAMATO: Yet Another More Advanced Top-level Ontology”, http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/YAMATO101216.pdf,2010
[8] R.Mizoguchi, M. Ikeda, K. Seta and J. Vanwelkenhuysen, “Ontology for modeling the world from problem solving perspectives”, Proc. of IJCAI-95 Workshop on Basic OntologicalIssues in Knowledge Sharing,1995,pp.1-12
[9] J.Aitchison, A.Gilchrist, D.Bawden,“Thesaurus construction and use: apractical manual”, Routledge; 4 edition, 2002
[10] K.Kozaki, Y. Kitamura, M. Ikeda, and R. Mizoguchi, “Hozo: an environment for building/using ontologies based on a fundamental consideration of “role” and “relationship”” Proc. of the 13th International Conference Knowledge Engineering and Knowledge Management (EKAW2002),2002,pp.213-218
[11] M.Hori, and T.Kato,“Mobile web and accessibility”, In S. Harper and Y. YesiladaEds, Web Accessibility: A Foundation for Research, Springer-Verlag London Limited, 2008, pp. 301-313