Title	An Analysis of Research Topics within a Community : the Example of Knowledge Science
Author(s)	NIE, Kun; JI, Zhe; NAKAMORI, Yoshiteru
Citation	
Issue Date	2007-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/4138
Rights	
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, Proceedings of KSS'2007 : The Eighth International Symposium on Knowledge and Systems Sciences : November 5-7, 2007, [Ishikawa High-Tech Conference Center, Nomi, Ishikawa, JAPAN], Organized by: Japan Advanced Institute of Science and Technology

An Analysis of Research Topics within a Community: the Example of Knowledge Science

Kun NIE Zhe JI Yoshiteru NAKAMORI

Graduate School of Knowledge Science,
Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
{ niekun, zhe-ji, nakamori}@jaist.ac.jp

Abstract

This Paper presents a new approach how to analyze research topics within a given research community. Under the guidance of the I-system methodology, this paper conducts both top-down and bottom-up analysis. For the bottom-up analysis, similarity measurement and hierarchical clustering are applied to obtain a tree-like dendrogram structure of research topics; for the top-down analysis, the experts' knowledge is included. Then resulting from the iterative dialogue between the above two stages of automatic construction and expert- supervision, an ontology structure of research topics is finally achieved.

Keywords: I-system, research topics, top-down, bottom-up, ontology structure

1 Introduction

Graduate School of Knowledge Science (KS School) of JAIST specializes in this unique position in the world to have a variety of interdisciplinary or multidisciplinary research. Meanwhile, as a result of vast research topics in KS school, it is hard to have a clear picture of what has been done in KS school. In order to improve current research in KS school and accelerate knowledge innovation in KS school, it is very important to integrate the past research achievements (Ji, 2007).

Therefore, the purpose of this research attempts to map the relationships among past research topics in KS school, and farther construct an ontology structure of research topics for KS school. The objectives are:

- ✓ Collecting research topics information from papers/articles in KS school
- ✓ Measure the similarity and map the relationships among these research topics
- ✓ Cluster the research topics into a certain number of groups
- ✓ Building an ontology structure for KS school.

Two groups of data are collected; one is master thesis and doctoral dissertation by students in KS school with the purpose to know what has been done in the community of students, the other is papers/articles by faculty of KS school with the purpose to know what has been done in the community of KS school faculty. This paper only concentrates on the first group of data.

The remainder of this paper is organized as follows: Section 2 introduces I-System and its application in the context of our work; with the help of I-System methodology, section 3 designs an algorithm for building ontology structure; Section 4 provides a case study; Section 5 summarizes this paper.

2 I-System Methodology

Nakamori (2003) proposed I-System methodology which includes five sub-systems: Intervention, Intelligence, Involvement, Imagination and Integration. I-System methodology stresses that most uncertain complex problem couldn't be solved only from scientific front; social front and cognitive front need to be considered as well. That is, we have to integrate scientific, social and cognitive dimensions in order to arrive at a good solution for an uncertain problem.

Figure1 puts I-System in the context of our work and explains it in more depth.

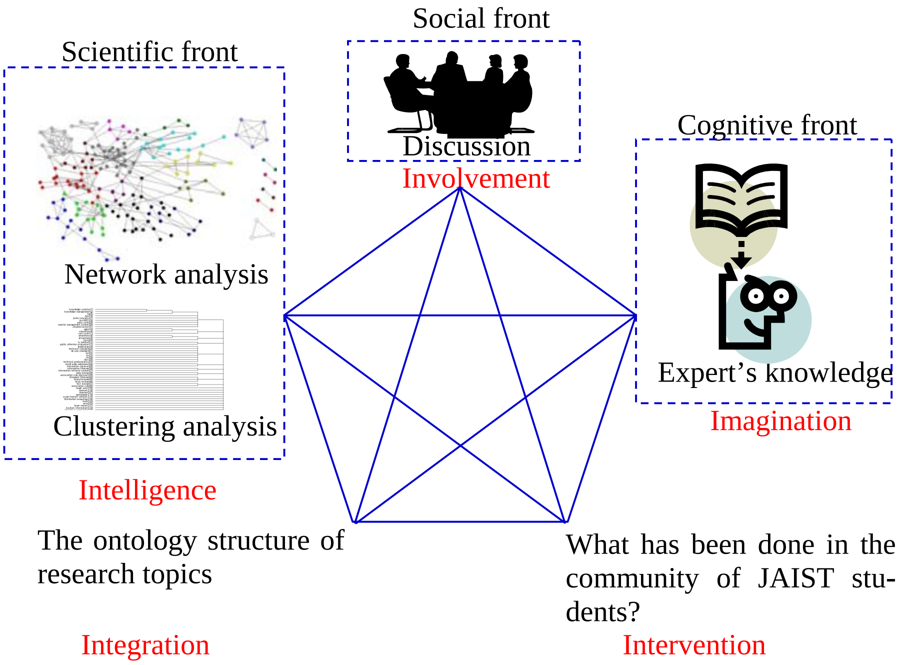

Fig. 1 I-System Methodology

In our work, I-System is used to assist thinking and working on how to build ontology structure of research topics.

(1) Subsystem of Intervention: “Intervention” is the first subsystem in which the faced problem has to be shaped or clarified clearly. To us, the problem needed to be solved is “what has been done in the community of JAIST students”. Once has a problem, this subsystem request the following three subsystems to concentrate on it from scientific front, cognitive front and social front respectively.

(2) Subsystem of Intelligence: “Intelligence” is bottom-up approach to analyze research topics. In our work, two important techniques, namely, network analysis and clustering analysis are applied.

(3) Subsystem of Imagination: “Imagination” is experience-based or top-down approach to analyze research topics. .

(4) Subsystem of Involvement: “Involvement” is from social front, we believe that both scientific method and cognitive method do have their advantages and disadvantages, and a conflict between them often happens. And this subsystem attempts to build a bridge between scientific and cognitive front.

(5)Subsystem of Integration: “Integration” is final subsystem. The tasks of this subsystem is to integrate results from the above four subsystems, and submit the final report.

3 Algorithm for Ontology Construction

Here we explain how to build the ontology structure based on the I-system methodology that we mentioned in the above section. Our procedure that combines stages of expert-supervised and automatic construction is articulated below:

Step 1: Start by selecting an ontological category that needs to be divided. This category can be determined either by expert or automatic construction.
Step 2: In the expert-supervised stage, the experts specifies several examples objects for the ontological category given in step1.
Step 3: In the automatic construction stage, all objects that are similar to those example objects are clustered to the same ontological category automatically.
Step 4: The resulting division in step3 may again be submitted for the approval of the experts, if the experts disapprove, go back to step 2.
Step 5: Steps 1-4 forms one iteration. The entire procedure is repeated for as long as there are no

more categories that need to be divided, or until another stopping condition.

Step 6: The final version of ontology is achieved and submitted to experts for evaluation of looking for incompleteness, inconsistence, and redundancy. Future maintenance and refinement are allowed.

In automatic construction stage, two important techniques, network analysis and clustering method, are specifically used. Network analysis allows measuring the degree centrality of a research topic which is defined as the number of other research topics directly connected to it (Hanneman 2005; Wasserman 1999). Because degree centrality can speak the power of a research topic in the network, that is, the higher degree centrality is, the more powerful a research topic has. By this reason, we also found that research topics with higher degree centrality are always top-level concepts, like *knowledge management*, *knowledge creation*, *system*, and vise versa, see Table 1. So network analysis assists assigning research topics into different layers of ontological category.

Table 1: Top Keywords Ranked by Degree Centrality

Keyword	**Degree Centrality**
knowledge creation	17
knowledge management	16
system	16
leadership	15
simulation	13
innovation	11
data mining	10
community	10
groupware	9

Our clustering method is based on network similarity which can be understood as the same pattern of connectivity in the network (Hanneman 2005; Wasserman 1999). That is, two research topics are similar if they are connected to the same other research topics. As an example, two research topics, *brainstorming* and *brain writing*, both of them are connected to research topics *divergent thinking* and *groupware*, they are considered having high similarity and thus they are clustered together into the same ontological category even they don't have a direct connection between them. Therefore, to measure similarity of two research topics, firstly, co-ocurrence matrix which describes how many times pairs of research topics appear together in one or more papers is calculated; secondly, classical similarity measuring algorithm, in our work, Euclidean distances-based algorithm (Formula 1 and Formula 2, before Formula 2, the role of Formula 1 is to standardize data otherwise the values from co-ocurrence matrix are un-comparable with each other because the values are dependent on their related research topics), is performed on co-occrence matrix which is then converted to similarity matrix; finally, classical cluster analysis method (in our work, single-link, or nearest neighbor method) is performed on similarity matrix to group those research topics that are most similar first, then similarity matrix is then re-calculated, and the next most similar pair are then joined, this process continues until all research topics are joined together and hierarchical dendrogram including all research topics is produced.

$$\mathrm{Re}l(k_i, k_j) = \frac{[Fre(k_i, k_j)]^2}{Fre(k_i) \times Fre(k_j)}$$

(Formula 1)

$$Sim(K_i, K_j) = \frac{1}{\sqrt{\sum_{s=1}^{n} [\mathrm{Re}l(k_i, k_s) - \mathrm{Re}l(k_j, k_s)]^2}} \quad (s \neq i, s \neq j)$$

(Formula 2)

4 A Case Study

In this case study, we are intending to build ontology structure of research topics within the community of JAIST Knowledge Science School students, including master and doctoral students. Research topics are considered as building blocks and are collected from master theses and doctoral dissertations at *Graduate School of Knowledge Science, Japan Advanced Institute of Science and*

Technology which is believed to be the world's first research and education institute under the theme of knowledge.

In total, 415 papers are collected and the top 200 research topics are selected depending on their frequency in the total number of papers. Then these 200 research topics are used as resources to build domain ontology of knowledge science.

With these 200 research topics and the guideline discussed in the above section, it is able to construct the ontology structure of research topics for JAIST Knowledge Science School Student Community. A part result is given below:

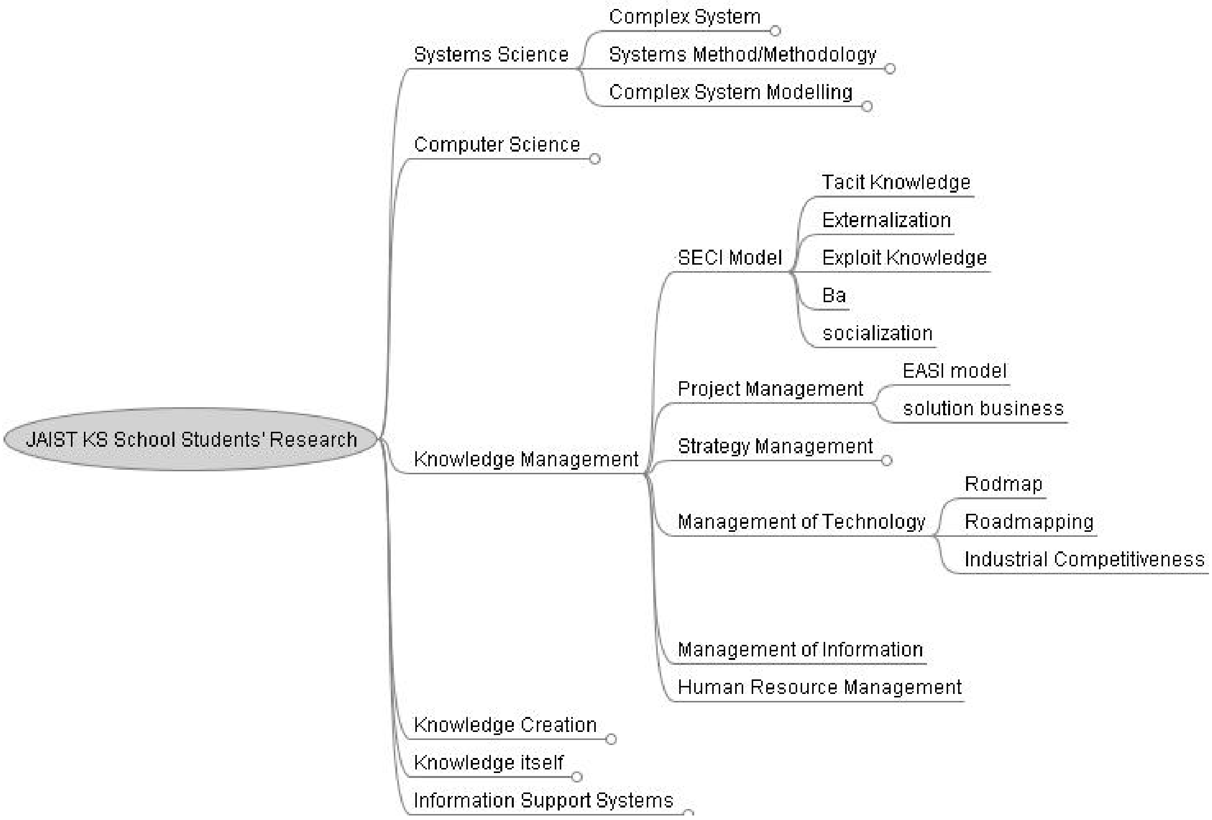

Fig 2 Ontology Structure for JAIST KS School Students' Research

5 Conclusion and Future work

Ontology Construction is a very difficult issue, apart from most previous work on this issue which either from complete expert-supervised method or complete automation, this paper proposed a semi-automatic method which combine both bottom-up and top-down analysis in order to take full advantage of computer and experts. A case study of JAIST Knowledge Science School Students Community was also given to test our algorithm for ontology construction.

The further test and improvement of our algorithm for ontology construction and our case study are needed. We believe in that the best test would be practice, to do that, we are planning to develop an Ontology Driven Semantic Search Engine in which we have to carry out all the steps of ontology creation, document annotation, and creation of a prototype Ontology-Driven Semantic Search Engine (*OSSE*) that works using a Domain Ontology.

Acknowledgment

The research is supported by the 21st COE (Center of Excellence) Program "Technology creation based on knowledge science: theory and practice" of Jaist, a funds by Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1]. Yoshiteru NAKAMORI. Systems Methodology and Mathematical Models for Knowledge Management. Journal of Systems Science and Systems Engineering, 12(1): 49-72, 2003.

[2]. Hanneman, Robert A. and Mark Riddle. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/), 2005.

[3]. Stanley Wasserman and Katherine Faust. Social Network Analysis: Methods and Applications. Cambridge University Press. 1999.

[4]. Zhe Ji. The Knowledge Mapping of Knowledge Science School of JAIST, Master thesis. 2007.