Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

Supporting Literature Exploration with Granular Knowledge Structures

Yiyu Yao[1,2], Yi Zeng[2], and Ning Zhong[2,3]

[1] Department of Computer Science, University of Regina
Regina, Saskatchewan S4S 0A2, Canada
yyao@cs.uregina.ca
[2] International WIC Institute, Beijing University of Technology
Beijing, 100022, P.R. China
yzeng@emails.bjut.edu.cn
[3] Department of Information Engineering, Maebashi Institute of Technology
Maebashi-City, 371-0816, Japan
zhong@maebashi-it.ac.jp

Abstract. Reading and literature exploration are important tasks of scientific research. However, conventional retrieval systems provide limited support for these tasks by concentrating on identifying relevant materials. New generation systems should provide additional support functionality by focusing on analyzing and organizing the retrieved materials. A framework of literature exploration support systems is proposed. Techniques of granular computing are used to construct granular knowledge structures from the contents, structures, and usages of scientific documents. The granular knowledge structures provide a high level understanding of scientific literature and hints regarding what has been done and what needs to be done. As a demonstration, we examine granular knowledge structures obtained from an analysis of papers from two rough sets related conferences.

Keywords. Granular computing, research support systems, research methods, literature exploration, granular knowledge structures

1 Introduction

Literature exploration plays an important role in scientific research. Many scientists devote much of their valuable time exploring and digesting the scientific literature. With the over-increasing volume of scientific documents, the study and analysis of them becomes a real challenge for any scientist. Solso envisioned an intelligent system that "may tell us what research has been done, so we can avoid redundant studies, and it also may tell us what needs to be done, so we can put our valuable time to good use" [8]. Mjolsness and DeCoste suggested that machine learning can be used to support every phase of the research process [3].

Traditional information retrieval systems and Web search engines support the basic tasks of browsing and retrieval, so that a scientist can easily navigate the Web, browse digital libraries, and find relevant documents. They normally

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

do not support the knowledge intensive tasks of analyzing, organizing and digesting the retrieved documents. Although many authors have pointed out the ineffectiveness of retrieval systems and Web search engines, the real problems may not lie on the classical issue of "retrieval". That is, the real problems are no longer retrieval, but post-processing of retrieved results.

In order to resolve the difficulties of current retrieval systems and to better support scientists, many proposals of next generation intelligent systems have been made, including information retrieval support systems [9, 10] and research support systems [11]. The main objective of this paper is to propose a framework of literature exploration support systems, as a sub-system of a research support system. Such systems help scientists understand scientific literature in a structured and knowledgeable way.

Many authors have studied the problem of supporting literature exploration from different perspectives. Robert and Alfonso examined the connection and relation among different literature by domain characteristics [7]. Kuznetsov analyzed literature from its content view using concept lattice [2].

Based on these studies, we introduce the notion of literature exploration support systems. Such a system needs to analyze and organize scientific literature in multiple views. It supports a scientist to make explicit the granular knowledge structures embedded in scientific literature from its contents, structures, and usages perspectives. Techniques of granular computing are used to construct and represent granular knowledge structures.

2 An Overview

This section introduces the notion of literature exploration support systems and two important technologies for building such systems.

2.1 Literature Exploration Support Systems

Knowledge structures play a central role in problem solving [1, 6]. They may help scientists to see the contributions of a particular study and its relationships to other studies. One of the objectives of reading and literature exploration is to construct these knowledge structures or concept maps. This is evident from many survey papers and literature review sections in many scientific writings.

A set of documents from different sources (e.g., digital libraries, conference proceedings, journal databases, results from retrieval systems or Web search engines, etc.) may be viewed as the space of exploration. Through multi-view analysis of its contents, structures and usages, a literature support helps a scientist to organize the literature into a structured and knowledgeable way so that it can be better understood and used in future research. The results may be represented as granular knowledge structures.

A literature exploration support system may be viewed as a sub-system of a research support system. Such a system can be seamlessly integrated with a retrieval system or a Web search engine, by treating the retrieved results as a

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

collection of scientific documents. Thus, a literature exploration support system may also be viewed as a sub-system of an information retrieval support system.

2.2 Granular Computing

Knowledge of a well-established field can normally be organized in a hierarchical way [6]. More abstract knowledge can be built upon more concrete knowledge. Knowledge at different levels represents differing granularity. Furthermore, at each level of the hierarchical structure, one associates rules regarding how to apply such knowledge [6]. It becomes clear that a literature exploration support system must help us to construct such granular knowledge structures.

As an emerging field of study, Granular Computing (GrC) is consistent with human problem solving based on knowledge structures [13]. Granular computing covers theories, methodologies, and tools that explore data granules, information granules and knowledge granules in problem solving. By viewing literature exploration as a problem solving task, one can immediately apply granular computing to literature exploration support systems. The three perspectives of granular computing are very relevant to literature exploration. In the philosophical perspective, it leads to structured thinking for understanding and organizing scientific literature. In the methodological perspective, it offers language and methods to build and represent granular knowledge structures from the literature. In the computational perspective, it deals with structured processing of granular knowledge structures.

2.3 Multi-view and Multi-level Exploration

Different views provide various unique understanding of the literature. By drawing results from Web mining, we propose to support literature exploration in multiple views and at multiple levels, based on the contents, structures, and usages of the literature.

The contents of literature can be organized based on different levels of granularity. Each granule represents a specific level of details of the literature. By comparing different levels of granules, the system can find the relationships among different papers or between a specific paper and a given topic.

Scientific literature is closely linked together by cross references. Such structural information needs to be explored when generating knowledge structures. For example, citation information has long been used in many studies of the structures of the literature.

Literature usage is another source that may be useful for building knowledge structures. The relationships among different documents could be investigated through user access behaviors. For example, if some papers are always viewed or studied together, one may establish a connection between them.

Based on the multi-view and multi-level in each view, a literature exploration support system can provide visualization for knowledge navigation and browsing.

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

3 Granular Knowledge Structures Generation

The essential issue in implementing a literature exploration support system is the generation of granular knowledge structures.

3.1 Granular Knowledge Structures

Knowledge structures can be built based on concepts. A concept is considered to be the basic unit of human thought and knowledge. A concept can be conveniently interpreted as a granule, namely, the extension of the concept. The representation, interpretation, connection and organization of concepts lead to granular knowledge structures [12].

We can use an information table and a language with respect to the table to represent knowledge. Consider a generalized decision logic language [14], *GDL*-language, which is an extension of a decision logic language used by Pawlak [4]. A specific information granule is represented by an atomic formula (a, r, l), where r denotes a particular relationship between an attribute value and a label. Let L_a be the set of labels for all granules on the domain of attribute a. We have the relation set $R_a = \{=, \in\}$. Thus, the atomic formulas are of the two forms, $(a, =, l)$ and $(a, \in, l)$.

A concept in an information table can be jointly represented as $(\phi, m(\phi))$. The formula ϕ represents the intension of the concept, while the set $m(\phi)$ consists of those objects satisfying the formula and represents the extension of the concept [14].

Knowledge granules can be defined as relations on concepts:

$$G(\{\Re_i | i \in I^+\}, \{(\phi_n, m(\phi_n)) | n \in I^+\}), \tag{1}$$

where $\Re_i$ denotes the relations between concept granules and I^+ the set of positive integers. Different levels of relations among concepts induce a hierarchical structure called a granular knowledge structure. In particular, we need to consider three levels of structures, namely, internal structure of a granule, collective structure of a family of granules, and hierarchical structure of a web of granules [13]. They form the integrated knowledge structures of the literature.

3.2 A Granular Knowledge Structures Generation Process

Building knowledge structures based on isolated papers by using traditional Web mining methods may not be satisfactory. They may not be able to represent the connections between different levels of granules, such as subtopics and disciplines [5]. As a knowledge intensive system, a literature exploration support system must consider semantic information and user involvement.

One issue is the weights of different documents in the literature. It is a well known fact that some scientific papers are more important than others, because they have major impact on later research. Therefore, those documents should play a major role in forming the knowledge structures. A set of such documents

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

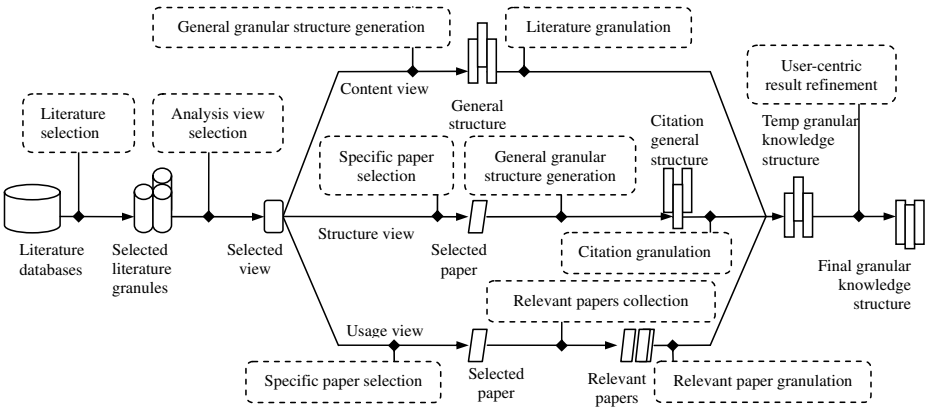

Fig. 1: Main Steps for Granular Knowledge Structures Generation

may be easily obtained from citation information. A related issue is the definition of semantic relations between concepts, documents, and sets of documents. Typically, a scientific document has a well defined granular structure, consisting of title, abstract, section titles, and subsection titles. Such in formation may be incorporated. In other words, we can associate different weights to different concepts in a document.

A literature exploration support system must incorporate domain knowledge and user background knowledge. Although the construction process is the same, the knowledge base used is domain specific and personalized. We take a human-centric approach that allows a scientist to add new, to improve existing knowledge, and to refine granular knowledge structures. A support system needs to seek for the right balance between automation and user intervention [15].

Figure 1 shows the main steps for generating granular knowledge structures:

- In the *literature selection* step, the system or a user collects a set of documents to be explored.
- In the *view selection* step, a user selects a particular view for building multi-view based granular knowledge structures.
- In the *structure generation* step, the system generates different granular structures.
- In the *user-centric result refinement* step, a user can refine the results from the previous steps.

For exploration from content view, we build the general granular structures according to information about a single document and a sub-collection of documents, as well as domain knowledge. For exploration from structure view, we

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

Table 1: A Partial Information Table for Generating Figure 2

Paper	Initial Page	Theory	Application	Domain
No.05	p1-94	Rough-Algebra	–	Rough Set
No.12	p1-345	Rough-Fuzzy Hybridization	–	Rough Set
No.25	p2-342	Logics and Reasoning	Medical Science	Rough Set
No.21	p2-263	Data Reduction	Image Processing	Rough Set
No.29	p2-383	Logics and Reasoning	Bioinformatics	Rough Set
No.97	p3-522	Formal Concepts	–	Rough Set
No.30	p2-430	Data Reduction	Bioinformatics	Rough Set

focus on building citation relation structure. For exploration from usage perspective, we find the connections and external structures of relevant papers based on literature access logs and domain knowledge.

4 An Illustrative Example

To demonstrate the proposed framework, we extract related information from RSFDGrC 2005 and RSKT 2006 proceedings to form the granular knowledge structures of Rough Sets.

The diagram shown in Figure 2 is formed based on information granules at different level of granularities from the two proceedings. Table 1 contains some examples of the information table used to construct Figure 2.

Examples of information granules from Table 1 are given as:

$$G(Theory, =, Formal\ Concepts) = \{No.97\},$$
$$G(Application, \in, l_1) = \{No.25, No.29, No.30\},$$
$$G((Theory, =, Data\ Reduction) \wedge (Application, \in, l_1)) = \{No.30\},$$
$$G((Page, =, 2-383) \Rightarrow (Application, =, Bioinformatics)) = \{No.29\}.$$

The label l_1 is the granule containing {Medical Science, Bioinformatics}. The symbol $\Rightarrow$ denotes the connection between two granules [14]. The concept granules for the first two formulas can be represented as:

$$((Theory, =, Formal\ Concepts), m(Theory, =, Formal\ Concepts)),$$
$$((Application, \in, l_1), m(Application, \in, l_1)).$$

An example of granular knowledge structure based on the partial ordering is given as:

$$((Theory, =, Formal\ Concepts), m(Theory, =, Formal\ Concepts))$$
$$\subseteq ((Domain, =, Rough\ Sets), m(Domain, =, Rough\ Sets)).$$

Figure 2 shows a multi-level granular structure from the content view. For example, the coarsest granule is "Rough Sets", finer granules are subtopics related to "Rough Sets", and papers falling under each subtopic form the basic

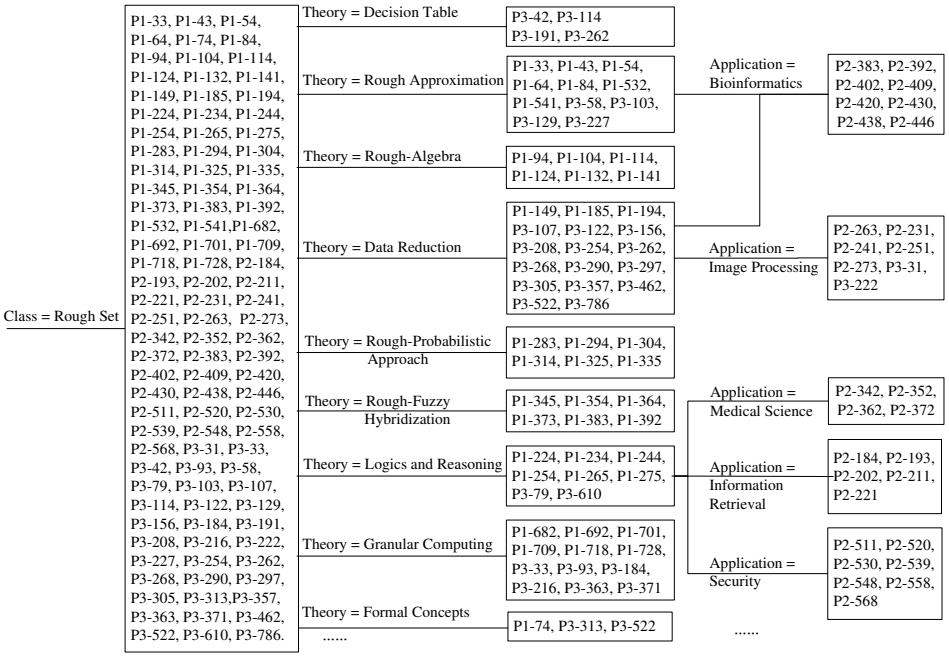

Fig. 2: Granular Knowledge Structure of Rough Sets from the Content View of the RSKT 2006 and RSFDGrC 2005 Proceedings

granules. The fact that "Nine theory subtopics are related to Rough Sets" reflects a coarser knowledge structure. The fact that "Bioinformatics and Data Reduction are related" reflects a finer knowledge structure. Figure 3 provides a structural view of a single paper's citations. Other views of granular structures could be further investigated.

The granular knowledge structures not only provide a relation diagram of specified discipline, but also help researchers to find the contribution of each study and possible future research topics. For example, as shown in Figure 1, many studies concentrate on data reduction and rough set approximations, and

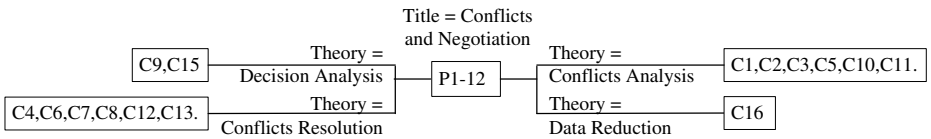

Fig. 3: A Single Paper's Granular Knowledge Structure from Citation View

Yiyu Yao, Yi Zeng, and Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures,
In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,
Lecture Notes in Artificial Intelligence 4482, Springer, Toronto, Canada, May 14-16, 2007, 182-189.

research of applications does not receive much attention. It can also be concluded that one may apply some of the theoretical studies (e.g., Rough-Algebra).

5 Conclusion

This paper proposes a framework of literature exploration support systems. Such a system constructs granular knowledge structures of the literature by using the theory and techniques of granular computing. This enables scientists to explore literature in multiple views and at multiple levels, in order to see the contributions of a particular study and its relationships to other studies.

Literature exploration support systems focus on the post-processing of retrieved results of current retrieval systems and search engines. These systems may have great impact in helping scientists to meet the challenge of over-increasing literature growth.

References

1. Gordon, S.E., Gill, R.T.: *The Formation and Use of Knowledge Structures in Problem Solving Domains*. Idaho University, Moscow (1989).
2. Kuznetsov, O.S.: Galois Connections in Data Analysis: Contributions from the Soviet Era and Modern Russian Research. In: *Formal Concept Analysis*. Springer, Berlin (2005) 196–225.
3. Mjolsness, E., DeCoste, D.: Machine Learning for Science: State of the Art and Future Prospects. *Science* **14** (2001) 2051-2055.
4. Pawlak, Z.: *Rough Sets, Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991).
5. Pedrycz, W.: *Knowledge-Based Clustering: From Data to Information Granules*. John Wiley & Sons, Inc., New York (2005).
6. Reif, F., Heller, J.: Knowledge Structure and Problem Solving in Physics. *Educational Psychologist* **17** (1982) 102-127.
7. Robert, H., Alfonso, V.: Implementing the iHOP Concept for Navigation of Biomedical Literature. *Bioinformatics* **21** (2005) 252-258.
8. Solso, R.L., MacLin, M.K, MacLin, O.H.: *Cognitive Psychology*. Pearson Education, Inc. (2004).
9. Yao, J.T., Yao, Y.Y.: Web-based Information Retrieval Support Systems: Building Research Tools For Scientists in the New Information Age. In: Proc. of the IEEE/WIC Int. Conf. on Web Intelligence 2003, Halifax, Canada (2003) 570-573.
10. Yao, Y.Y.: Information Retrieval Support Systems. In: Proc. of FUZZ-IEEE'02, Hawaii, USA (2002) 773-778.
11. Yao, Y.Y.: A Framework for Web-based Research Support Systems. In: Proc. of COMPSAC'03, Washington, DC, USA (2003) 601-606.
12. Yao, Y.Y.: Concept Formation and Learning: A Cognitive Informatics Perspective. In: Proc. of the IEEE-ICCI'04, Victoria, Canada (2004) 42-51.
13. Yao, Y.Y.: Three Perspectives of Granular Computing. In: Proc. of IFTGrCRSP2006, Nanchang, China (2006) 16-21.
14. Yao, Y.Y., Liau, C.-J.: A Generalized Decision Logic Language for Granular Computing. In: Proc. of FUZZ-IEEE'02, Hawaii, USA (2002) 1092-1097.
15. Zhao, Y., Chen, Y.H., Yao, Y.Y.: User-centered Interactive Data Mining. In: Proc. of the IEEE-ICCI'06, Beijing, China (2006) 457-466.