Supporting Literature Exploration with Granular Knowledge Structures

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

[1] Department of Computer Science,
University of Regina, Canada
http://www2.cs.uregina.ca/~yyao
[2] International WIC Institute,
Beijing University of Technology, P.R. China
http://www.iwici.org/~yizeng
[3] Department of Information Engineering,
Maebashi Institute of Technology, Japan
http://kis-lab.com/zhong

Acknowledgements

- Thanks for attending the talk.
- Special thanks to Yan Zhao, Guang Chen, the conference committee and all the reviewers.

- Thank you for Prof. Yao

  to bring me to the Science temple.

  to give me suggestions on how to do research.

  to help me find the way of my life.

Motivations

- **How** scientific researchers **do research** in the **past** and **now**?
- **Limited support** for reading and literature exploration in the process of scientific research: Identifying relevant materials.

- " Electronic access to information is expanding greatly in all fields, … Is there the danger of having too much data and not knowing what it all means?... We are at risk of becoming overwhelmed by the **plethora of information**….. It is likely that some type of program that **processes information in a knowledgeable, intelligent way-that is understands information-** it 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. "

Robert L. Solso, Cognitive Psychology, 7th edition, 2004.

- ***"machine learning can be used to support every phase of the research process"***

Mjolsness, E., DeCoste, D. Machine Learning for Science: State of the Art and Future Prospects. Science 14 (2001) 2051-2055.

What we are doing.

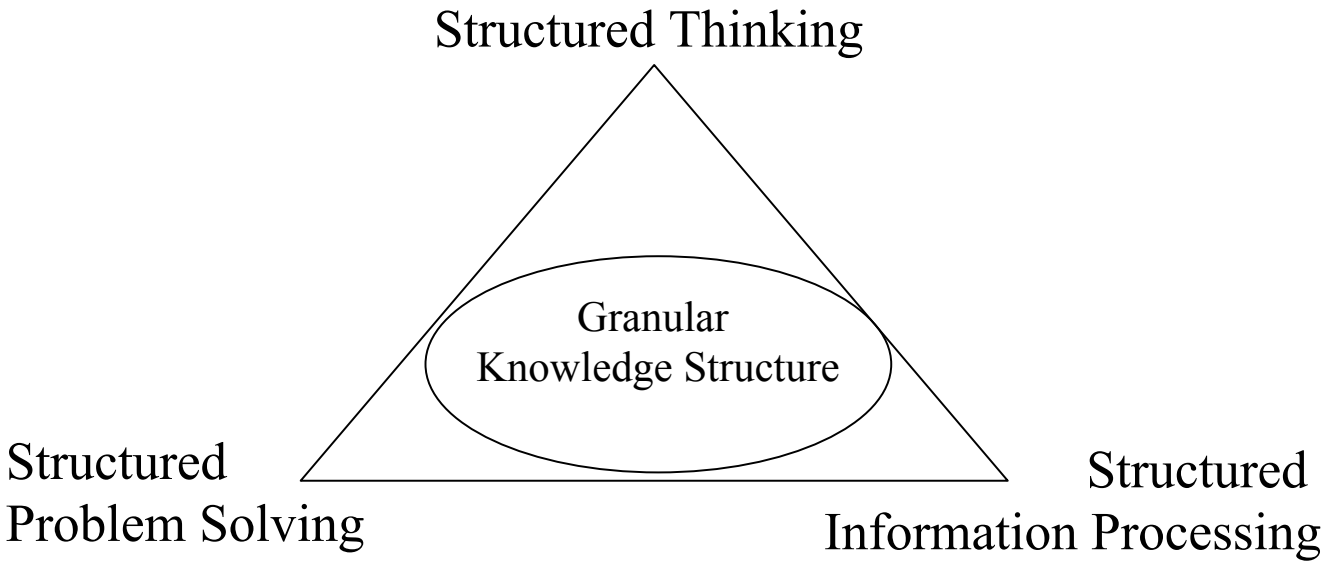
- **Not only "Retrieval", but "Post-processing".**
- **Literature Exploration Support System**: providing additional support functionality by focusing on **analyzing, organizing, and digesting** the retrieved literature.

- Some related ideas:
  (1) Literature analysis in the early 1970'
  (2) Examining on the connection and relation among different literature by domain characteristics [7].
  (3) Analyzing literature from document content view [2].
  (4) Supporting exploratory search (Special Issue CACM, 49(4), 2006).

- "Knowledge = Information + **Information Relationships** + Inference rules." Andrzej Skowron. Toward Rough-Granular Computing, Keynote talk, JRS'07.

- **Granular knowledge structures** provide a structured, high level understanding of scientific literature and hints regarding what has been done and what needs to be done. (contents, structures, and usages perspectives).

Perspectives of Granular Computing for literature exploration



- Granular Computing (GrC) is consistent with human problem solving based on knowledge structures [13]. The three perspectives of granular computing are very relevant to literature exploration.
- **Philosophical perspective**: structured thinking for understanding and organizing scientific literature.
- **Methodological perspective**: language and methods to build and represent granular knowledge structures from the literature.
- **Computational perspective**: structured processing of granular knowledge structures.

Multi-view and Multi-level Exploration

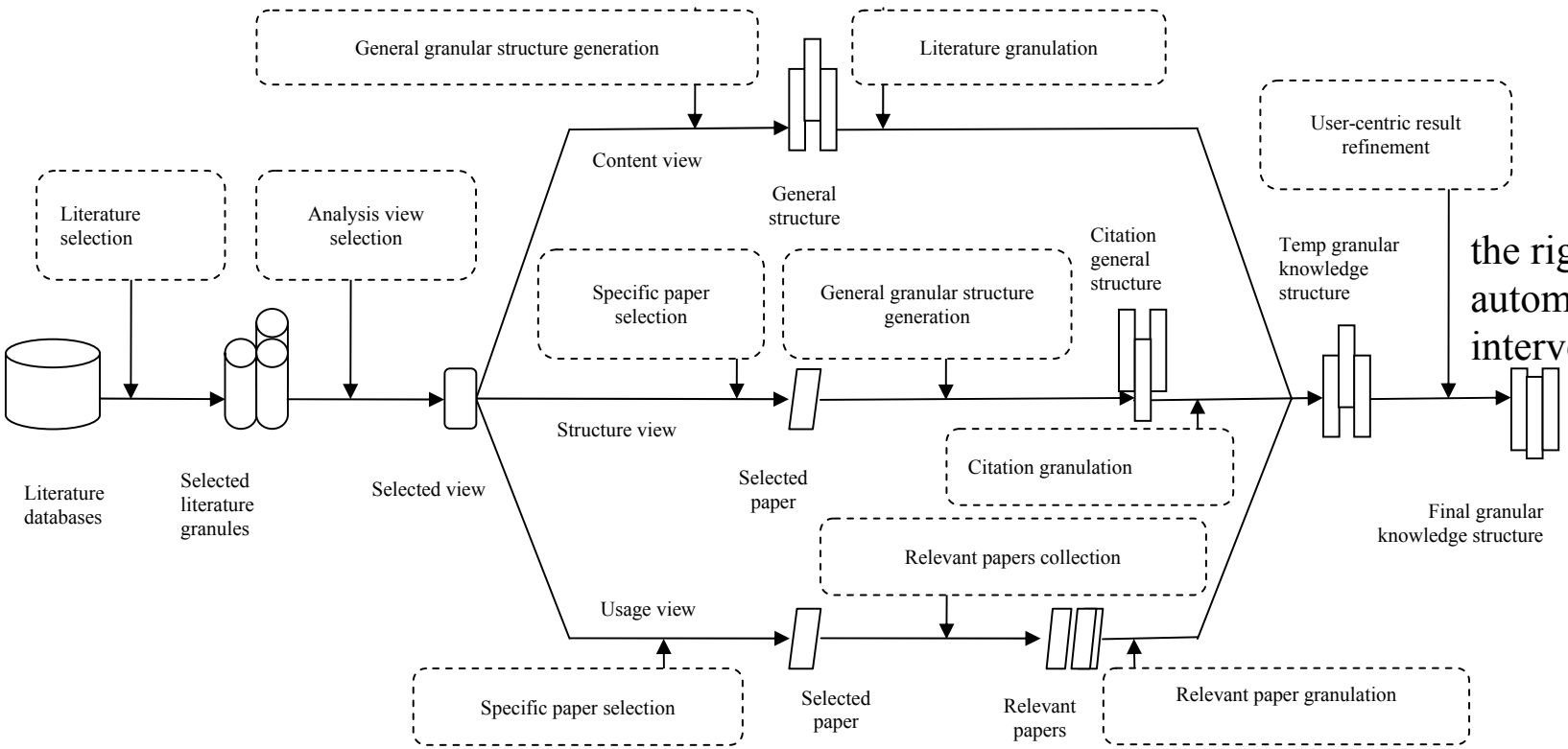
- The contents of literature can be organized based on different levels of granularity. Each granule represents a specific level of details of the literature.

Content View: comparing different levels of granules, the system can find the relationships among different papers or between a specific paper and a given topic.

Structure View: Scientific literature is closely linked together by cross references. (e.g. Citation information has long been used in many studies of the structures of the literature.)

Usage View: The relationships among different documents could be investigated through user access behaviors. (If some papers are always viewed or studied together, one may establish a connection between them.

Granular Knowledge Structures Generation Process



- **Literature selection** : the system or a user collects a set of documents to be explored.
- **View selection** : users select a particular view for building multi-view based granular knowledge structures.
- **Structure generation :** the system generates different granular structures.
- **User-centric result refinement:** user refines the results from the previous steps.

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].
- Knowledge Representation: information table + GDL- language [14].
- A specific **information granule** is represented by an atomic formula:

$$(a, r, l),$$

r : a particular relationship between an attribute value and a label.

L_a : the set of labels for all granules on the domain of attribute a.

Ra = {=, ∈}. (the relation set) Thus, the atomic formulas are of the two forms,

(a, =, l) and (a, ∈, l).

Granular Knowledge Structures (cont.)

- **Concept granule** : (ϕ , m(ϕ)).

  ϕ : intension of the concept

  m(ϕ) : objects satisfying the formula and represents the extension of the concept [14].

- **Knowledge granule**:

  $G(\{R_i \mid i \in I^+\}, \{(\phi_n, m(\phi_n)) \mid n \in I^+\})$

  Different levels of relations among concepts induce a multi-view, multi-level hierarchical structure called a **granular knowledge structure**.

  Integrated knowledge structures of the literature:

  (1) internal structure of a granule

  (2) collective structure of a family of granules

  (3) hierarchical structure of a web of granules [13].

An Illustrative Example

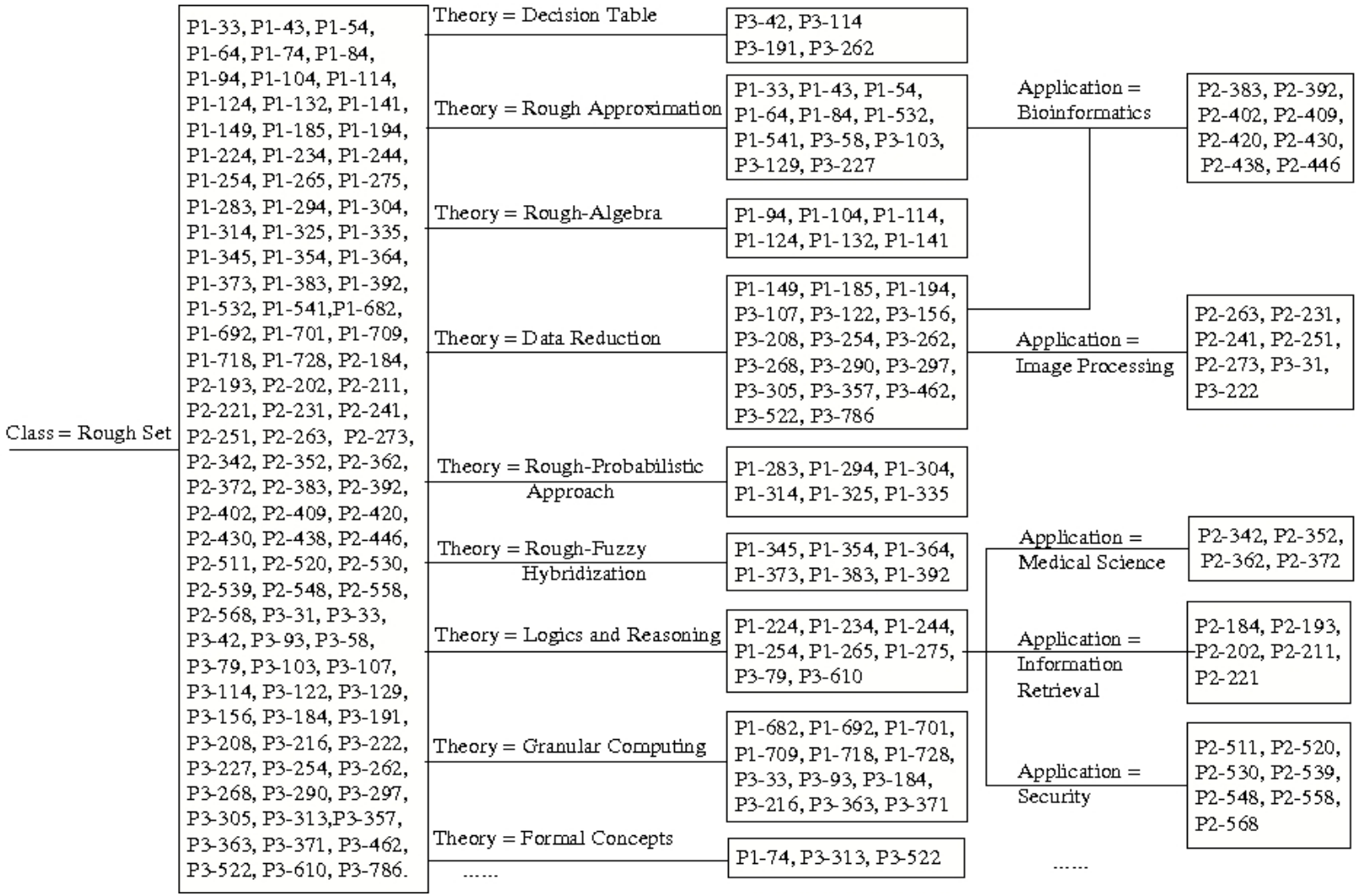


Fig. 2: Granular Knowledge Structure of Rough Sets from the Content View of the RSKT 2006 and RSFDGrC 2005 Proceedings

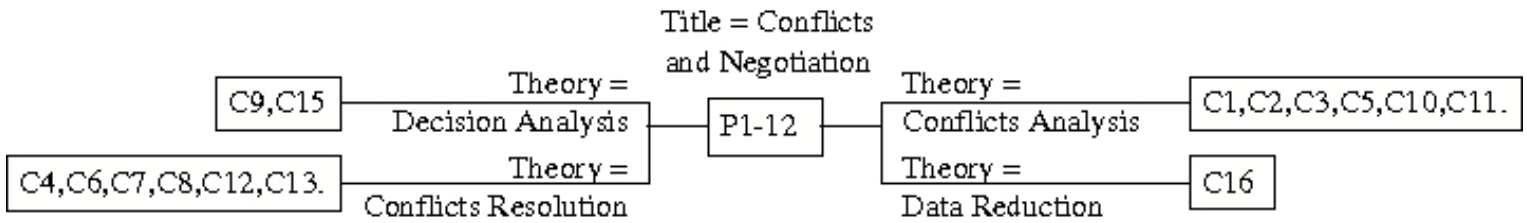


Fig. 3: A Single Paper's Granular Knowledge Structure from Citation View

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

Description of Granules

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, =, p2−383) $\Rightarrow$ (Application, =, Bioinformatics)) = {No.29}.

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)).

What we mentioned in this talk

- Some current problems for literature exploration and search engine support for scientific research.
- A framework of literature exploration support systems.
- Generations of Granular Knowledge Structure
- Descriptions of Information, Concepts, and Knowledge for results of post-processing(analyzing, reasoning, etc).

Paper 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.

Thank you!