

基于粒结构的知识检索

曾毅(1), 姚一豫(1,2), 钟宁(1,3)

- (1) 北京工业大学国际WIC研究院, 北京, 中国
- (2) 里贾纳大学计算机科学系, 里贾纳, 加拿大
- (3) 前桥工业大学信息工程系, 前桥, 日本



粒计算的一种认识 ----- > 具体领域(检索)



以粒计算的三个视角开始

粒计算是一种问题求解的统一方法论

*** 哲学观点：结构化思维**

人类的问题求解决定于对现实世界的认识，它由我们所拥有的世界观决定。

*** 方法论观点：结构化问题求解**

我们需要建立一套完整的，系统的理论及方法论。这些方法论将我们的世界观转换成对具体问题的理解和表示、求解策略、求解过程。

*** 计算模式观点：结构化信息处理**

从计算机科学角度讲，我们希望所设计的系统可以直接给出问题的解或帮助我们获得问题的解。

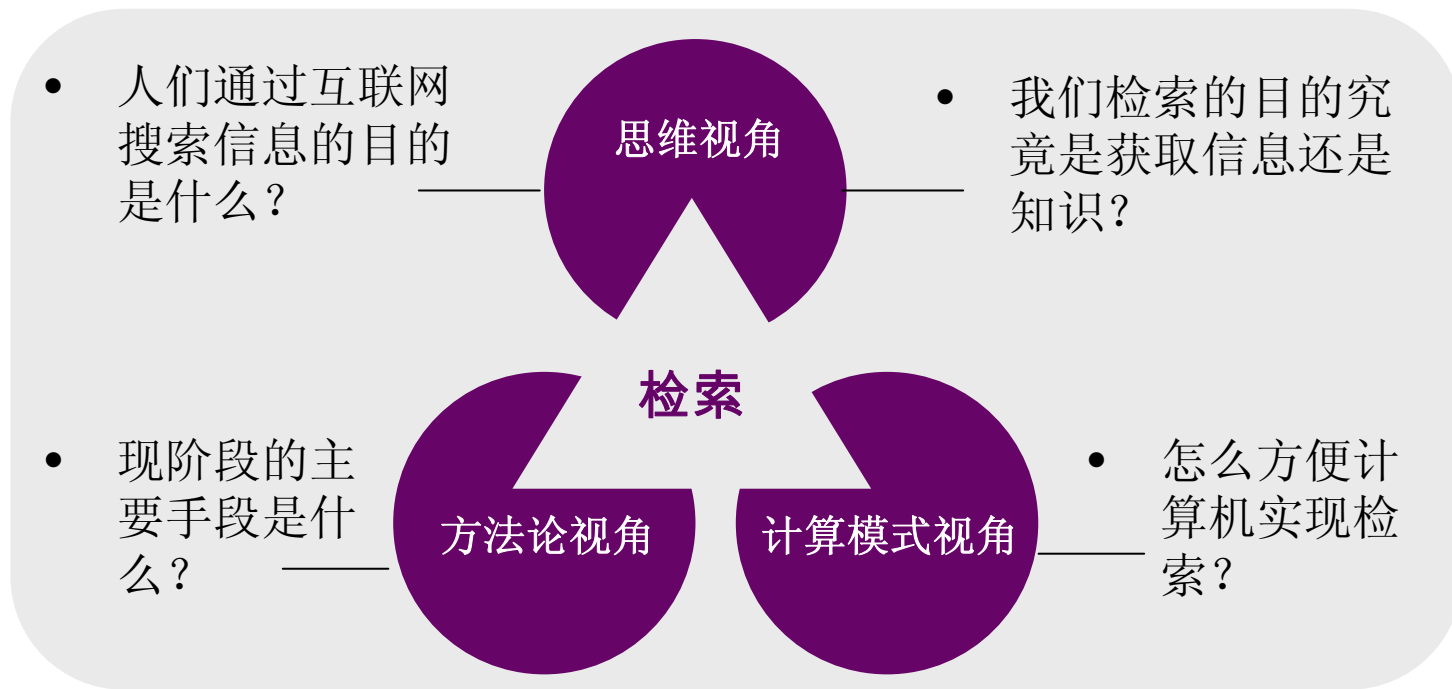


图 1. 粒计算的三个视角

引自姚一豫，粒计算的艺术，《粒计算：过去、现在与展望》，科学出版社，2007年。



对检索问题的一些思考



- (1) 现阶段的困难是什么？
- (2) 下一步应当怎么办？

图 2. 与检索有关的一些问题



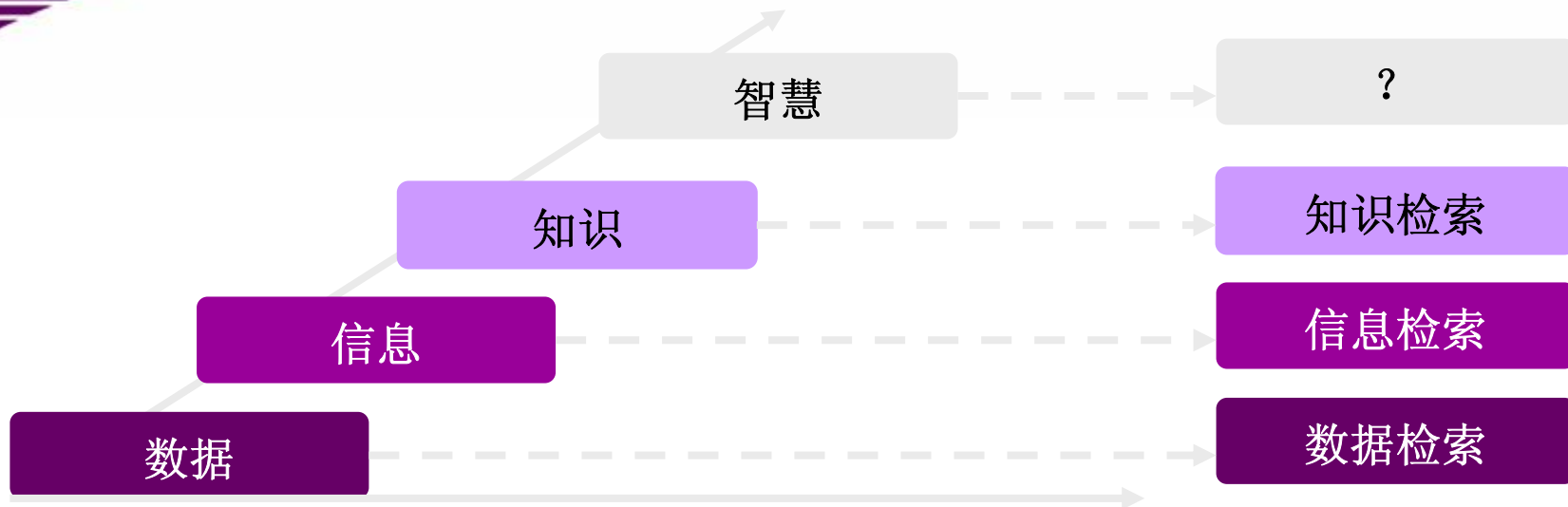


图 3. 系统分层结构与检索系统

数据层面 存储介质：数据库，检索系统：数据检索系统(DRS)，所处理的问题是结构化的，定义的概念是清晰的。

信息层面 信息检索系统(IRS)，处理的问题：半结构化，定义的部分概念不清晰。

- 输入查询关键字或其某种形式的组合检索到与此相关的信息，并通过对这些信息的学习获取知识。
- 海量信息条件下不可行：
 - (1) 逐一访问检索记录不可行。
 - (2) 排序算法推荐的检索结果并不一定满足用户的个性化需求。



进一步的认识

- (1) 科学研究中的文献检索 < ---- > 计算机和人类思维中知识表示方式的差异。
- (2) 文件、信息流 vs. 知识单元、知识结构。
- (3) 从信息中发现隐藏的知识以及构建知识结构的任务可以由计算机完成，从而缩小计算机中信息表示和人类知识表示的差距。

将检索信息的过程变为直接检索知识的过程，使检索结果由相关文档变为符合检索条件的知识。



粒知识结构与知识检索

一些前人的工作:

假设接受来自不同文献的提法, 则近期这个领域的进展主要包括:

- 大规模知识的获取与组织[8];
- 网络环境下的知识表示与描述方法[9, 10];
- 知识检索模型与方法[7, 9, 10]。

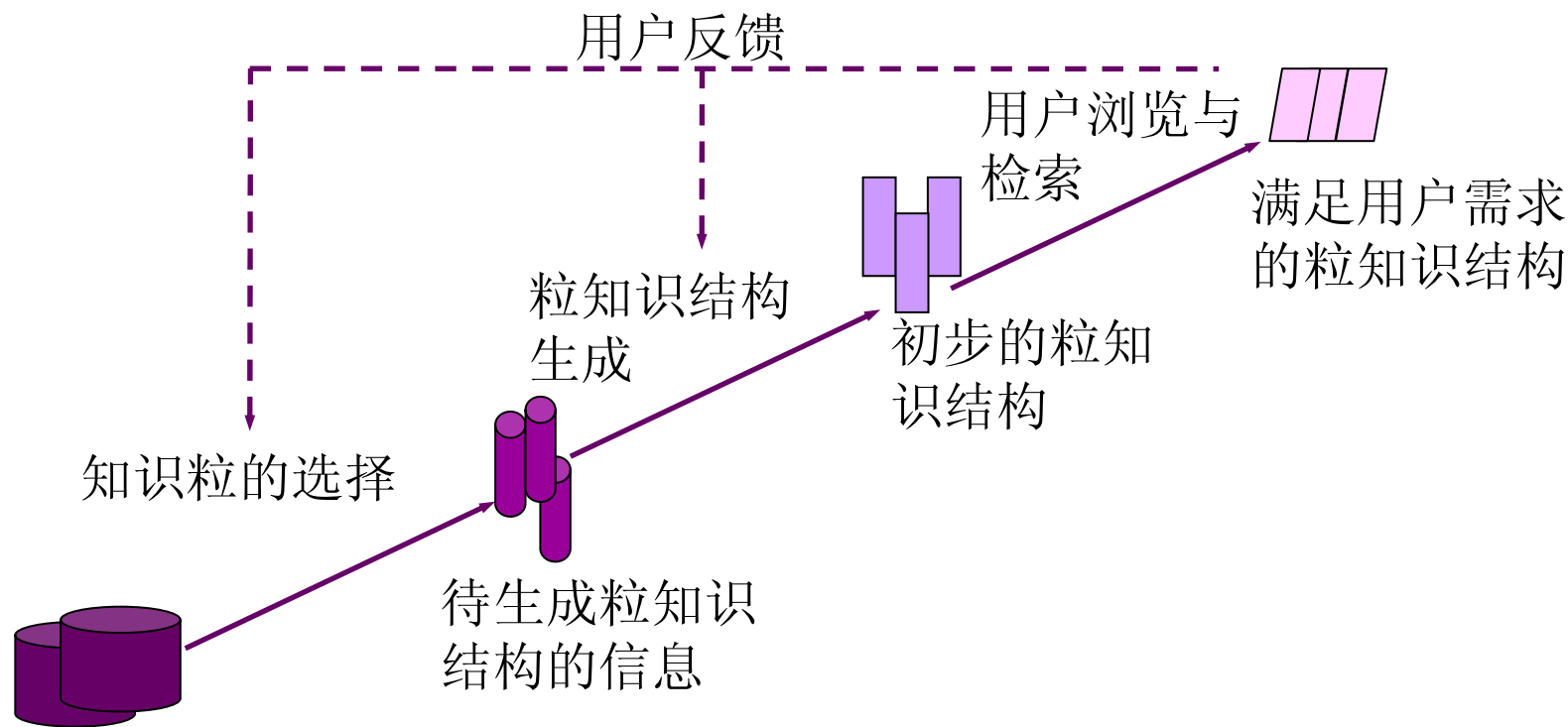
我们的认识:

考虑用户仅对要检索的内容有一个概要了解, 并不清楚需要检索的具体信息的情况, 若能以结构化的方式将相关知识呈现出来, 则有利于引导用户确定所需检索的内容[10, 11]。因此我们认为, **知识检索应当是围绕知识结构的生成、表示、浏览与检索展开的。**

- 树结构和概念图 [10, 12, 13], 语义网络 [11, 14], 形式概念分析。
- 粒计算独立于这些具体的工具, 为知识结构与知识检索提供了更抽象层次的表示和描述, 由此我们提出**基于粒结构的知识检索**。



基于粒结构的知识检索的概念模型



数据、信息源

图 4. 基于粒结构的知识检索

更具有普适性的知识检索模型请参考: Yiyu Yao, Yi Zeng, Ning Zhong, Xiangji Huang. Knowledge Retrieval (KR), The 2007 IEEE/WIC/ACM International Conference on Web Intelligence, to appear.



	数据检索	信息检索	知识检索
匹配	布尔匹配	部分匹配, 最佳匹配	部分匹配, 最佳匹配
推理	演绎推理	归纳推理	归纳推理, 演绎推理, 关联推理, 类比推理
模型	确定性模型	不确定性模型	语义模型+推理模型
查询	人工语言	自然语言	知识结构+自然语言
组织	表格, 索引	表格, 索引	知识单元与知识结构
表示	数值、信息表等	自然语言、标签语言等	概念图、谓词逻辑、产生式规则、框架、本体、语义网络等。
检索结果	数据集	文本段落或文档	知识单元

表 1. 几类检索系统的比较



粒知识结构的定义与表示

概念是人类思维和知识表示的基本单元，我们可以使用概念按照某种规则构建出知识结构。在知识表示的过程中，我们首先从信息粒中将概念粒显式地表示出来，随后对其进行解释、组织，则可以形成知识粒。在文献[15]中，我们基于概念之间的关系给出了知识粒一种可能的表示。

知识的粒结构至少包括三个层次的含义[16]:

- **知识粒内部的结构**: 构成整体结构的每一个知识粒可能由概念粒构成，也可能是粒度更小的知识粒构成的，它们之间的关系可能是序关系，也可能是其它特殊的联接关系[17]。
- **同层次中粒族的结构**: 同一层次中的粒体现出来的结构则是知识粒在这一层次中具有局部结构。
- **粒网络的分层结构**: 对知识粒从粗粒度到细粒度(或从细粒度到粗粒度)变化的整体结构的表示。



粒知识结构的发现与构建

知识的两个要素是知识因子和知识关联[7]，知识因子可被认为是一个基本知识粒，复杂的粒知识结构则是建立在知识关联的基础上的。知识具有层次性，但其分层结构并不一定是显而易见的[18]，我们需要使用某种方法将隐藏在信息粒中的粒知识结构显式地表现出来，将无结构、半结构的信息粒转化为结构定义良好的知识粒。

Collins和Quillian在通过反应时间对思维结构的研究工作中还指出：人类的思维结构往往是由多棵语义树表示的，其中包括一棵默认树和至少一棵备选树[11, 19]。

--- > 知识及其粒结构的表示应当站在多视角、每一个视角应当采用多层次。文献[15]从内容、结构、用法三个角度对文献的粒知识结构进行了构建和分析。



知识的来源可能会影响其粒结构，知识的权威性和正确性无法在复杂的Web环境下得以保证，所以生成粒结构时可信来源的选择是一个关键问题。例如在研究领域，粒知识结构的生成就可以依据引用率首先对文献的重要程度排序，也可以指定其来源(如特定的期刊、会议)[6, 15]。

由于领域背景的差异和个人理解的不同，即使针对同样的事物，不同人对知识的描述也可能截然不同。粒知识结构的形成过程应当融入用户的背景知识，并允许其根据需要对粒知识结构进行改进。



几个实例

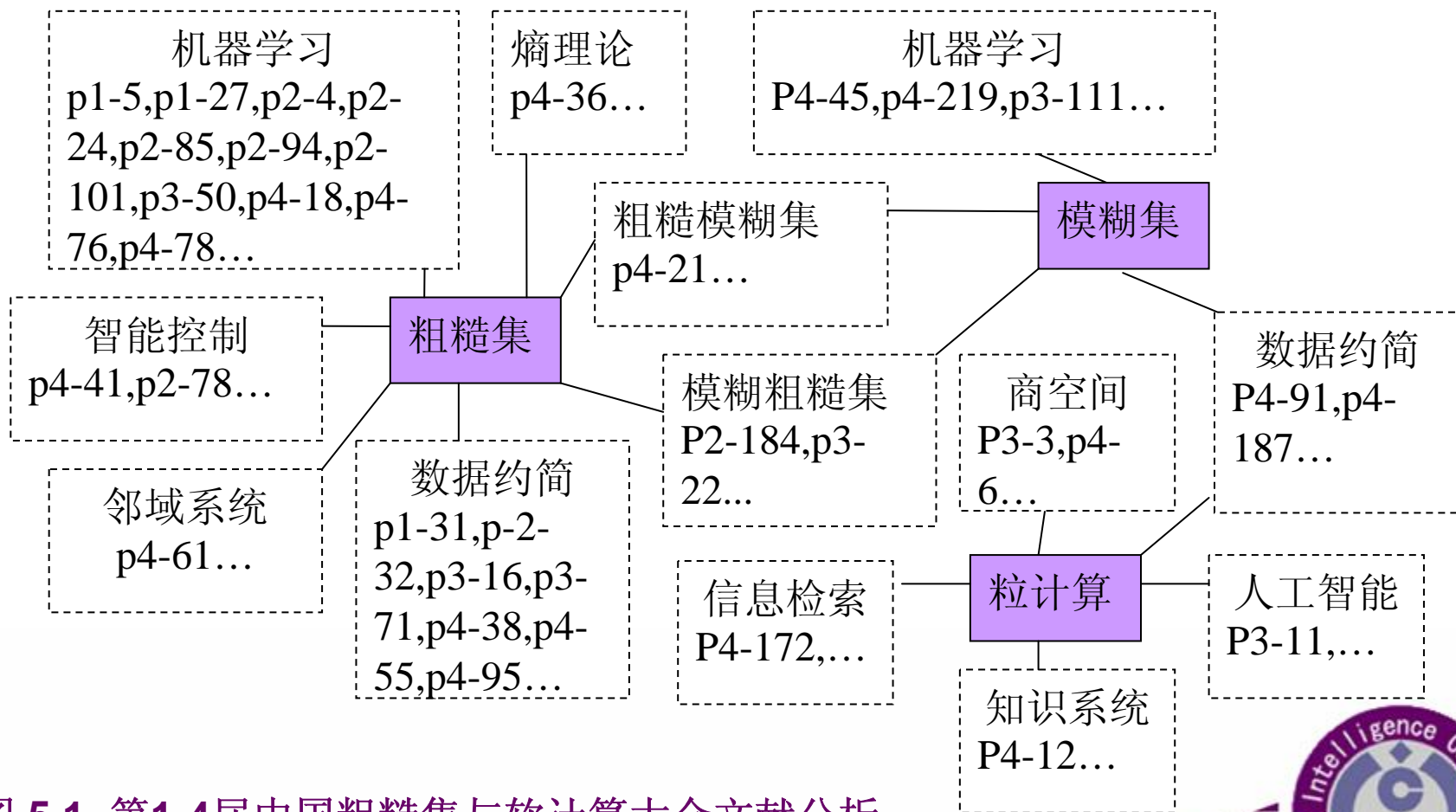


图 5.1. 第1-4届中国粗糙集与软计算大会文献分析
(多视角的一个实例)



几个实例

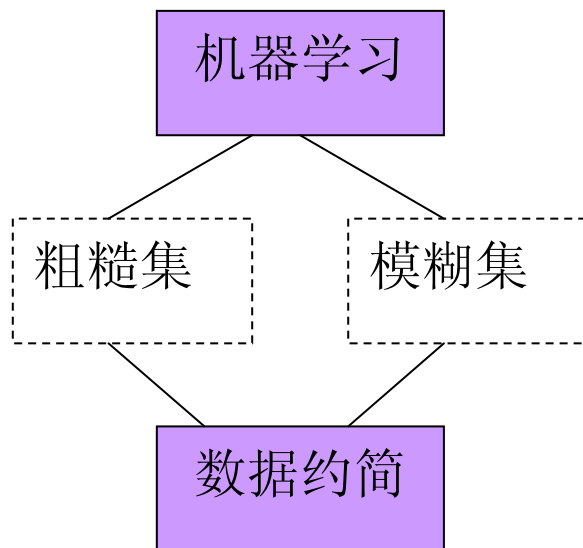


图 5.2. 第1-4届中国粗糙集
与软计算文献分析
(多视角的一个实例)



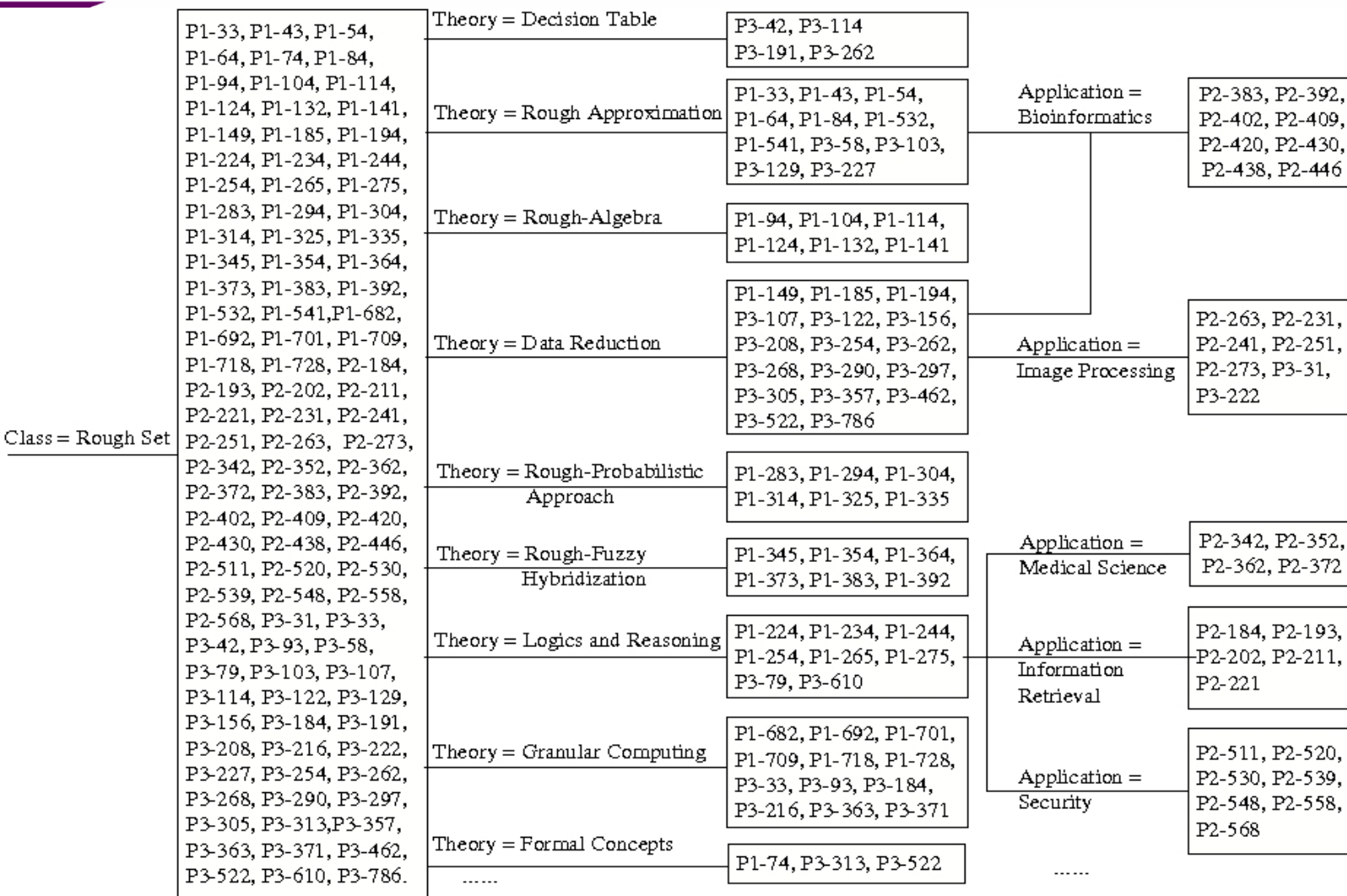


图 7. RSKT2006, RSFDGrC2005会议论文分析 Yiyu Yao, Yi Zeng, Ning Zhong. Supporting Literature Exploration with Granular Knowledge Structures, RSFDGrC 2007, LNAI 4482, Springer, 182-189.

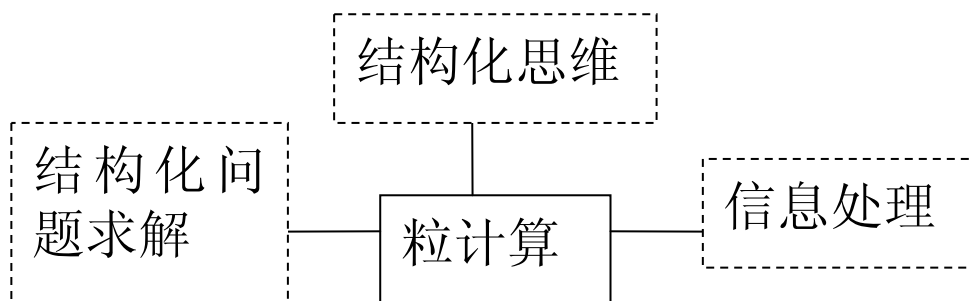


图 6. Yiyu Yao, Three Perspective of Granular Computing, Journal of Nanchang Institute of Technology



总结

- 基于粒结构的知识检索不是一个终止，希望是一个开始。
- 探讨的方法不一定正确，但是希望问题的本身有意义。
- 基于粒结构的知识检索也许是粒计算的一个好的应用领域，可能值得大家更深入的探讨。
- 希望明年有机会在第二届粒计算大会上介绍进一步的结果。

感谢各位老师 and 同学！

