

# 大容量知识的组织质量标准及实现技术

鲁东明 何志均 (浙江大学人工智能研究所, 杭州310027)

TP18

## 摘 要

Based on analyzing large scale knowledge's characteristic, this paper will propose and detailly discuss the basic standards for evaluating knowledge organization qualities, such as, logical clarity and understandibility, excellent physical access efficiency, capability of multi-sharing and modularity of organization, etc. Then, It will explore knowledge organization implementation techniques, Knowledge Dictionary, Integration of RDBMS and Item techniques, separation of logical and physical storage, etc. and describe briefly their applications in ZKBE system. Finally, we present some proposals for further research work.

在知识库系统中,良好的知识组织结构既能有效地改善系统的运行效率,又能方便知识库的管理、维护、共享和理解<sup>[1]</sup>。知识组织研究是知识库系统(KBS)的重要课题之一,随着KBS研究和应用的发展,已经历了几个不同的阶段。在初始阶段,由于知识库能够全部地驻于内存,人们强调的是如何改进动态的知识组织结构,提高知识的查询、匹配和触发等效率;近期,大容量知识的提出,内、外存并存的组织结构已成为研究的难点。人们对知识形式的结构化,非结构化知识的存贮及如何利用现存的数据库技术来组织知识等方面进行了广泛的研究<sup>[1]</sup>,但至今仍未有成熟的结果。

本文研究的重点放在利用知识表达模型OOS<sup>[2]</sup>,形成合适的知识库逻辑和物理组织结构,达到知识工程和效率上的要求。

### 一、大容量知识对组织的挑战

大容量知识不仅具有知识所固有的演导性、通用性、隐喻性,表现出表达上的复杂和不规整性,而且还具有一系列新的特点:

- 知识量大:系统知识数量之巨大,不

可能全部地放入内存,从而内、外存同时存放的组织结构成为必要;

- 存贮的长久性:知识一般有很长的生命周期,并且需要不断的扩充、维护和管理,其组织结构应支持用户的经常性操作;
- 共享性:知识的领域范围较宽,能支持多种形式的领域问题求解。知识库的多用户、多系统共享必须有良好的组织支持。

无疑,大容量知识给组织带来了新的困难和挑战,尤其是知识的存贮、查询速度和效率。因此,大容量知识的组织问题涉及以下方面:

- 大容量知识库组织结构的质量标准。
- 知识库组织的具体实现方法和技术。
- 大容量知识库组织的体系结构。

### 二、知识组织的质量标准

知识组织的质量标准是研究的基础,简单地说,是指一个知识库的结构在计算机处理环境下的逻辑清晰程度,知识组合的好坏程度,及对知识库访问、知识维护的支持程度。

### 1. 组织的逻辑清晰度和可理解性

知识库组织的逻辑清晰度和用户可理解性在组织结构中显得至关重要，因为这种知识库需要经常性的维护、扩充，并且要支持多用户共享。以OOS表达模式中的CLASS作为最基本的知识组织单元，大量的知识在逻辑上呈现如图1所示的层次结构。

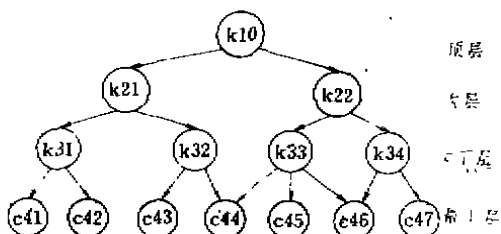


图1 知识的逻辑层次结构

- 顶层知识模式，代表大容量知识库所对应的领域中某个子领域，包括领域的范围、性质、功能及领域的常识性知识；
- 次层知识模式，是组成子领域的主要对象和任务；
- 次下层知识模式，表示同领域问题相关的宏对象；
- 最下层知识模式，是领域中最基本对象的知识。

依据上述逻辑组织结构，用户容易利用知识来解决具体的领域问题。同时，由于对知识的组织有了深层的理解，终端用户能方便地对大容量知识库进行更改和扩充。

### 2. 良好的物理存取效率

大容量领域知识不仅要求有极大的物理存贮容量，同时也希望有良好的物理存取执行效率。从目前状况看，外存的容量可以说是海量，内、外存统一的组织体系结构，基本上能解决存贮的容量问题，因此物理存贮的效率成为该标准的核心问题。知识物理存贮的效率标准又具体地体现为知识查询速度和知识更新效率两个子目标。

1) 知识查询速度。对于满足一定约束和要求的知识体，系统能以较短的路径快速定位。对知识的查询分两种基本情况，一是问题求解过程中的动态知识调度，二是用户

进行维护、管理时对知识的查询。前者基于知识内容的查找路径(Content-Sensitive Path)，以知识动态的逻辑和语义上的关系信息制导查询；后者基于知识结构的查找路径(Structure-Sensitive Path)，以知识间的结构关系信息制导查询。

2) 知识更新效率。对于需要日常维护和管理的大容量知识库，知识更新效率与知识查询速度同样具有重要的意义。高质量的组织要支持如下两个方面的工作：一是对每个知识体CLASS的更改，其作用效果仅局限于尽量小的范围，存贮结构要求模块化；二是知识物理组织空间的安排，须面向修改操作，避免不必要的知识块的大幅度搬家。在实现技术研究一节中，将进一步讨论如何支持这一点。

### 3. 知识的可共享性

知识的可共享性是大容量知识库的主要特点，也是研究大容量知识库组织的基本出发点之一。良好的知识组织结构，能使系统根据实际领域问题的特征和需要，形成逻辑的知识库视图，以便从大容量的知识库中分离出面向特定领域任务的知识集。多功能知识字典(KD)<sup>[1]</sup>的提出和应用，使知识组织有可能保证知识的视图结构(图2)。从某种意义上说，知识字典构成了大容量知识库抽象的概念层次上的索引。

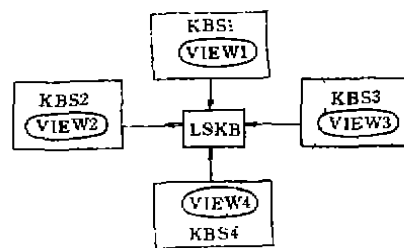


图2 LSKB的视图和共享

### 4. 知识组织的模块化

单个知识的可分离性和模块化是大容量知识库组织质量的又一标准。正如前面所提及的，大容量知识库的知识为某一大领域内多个面向特定任务的基于知识的系统所共

享,因此仅要求把任务相关的知识体从大容量的知识库中分离出来,形成逻辑上有一定独立性的知识子集;另一方面,经常性的知识库扩充和维护,还要求知识组织结构中单个知识体具有可分离性和模块化。一般说来,保证这一知识组织质量有以下一些方法。

- 根据知识体的内容和功能,建立层次、分类型的大容量知识库的整体组织结构,对于某一子领域中与特定任务相关的知识基本上落在了较小的分枝内,而无需关心其它分枝上的内容,能大大减少知识查询和搜索的空间。

- 根据知识体信息和功能,模块化组织,模块内部的知识体之间的联系较紧密,而模块之间的语义关系相对减少,使得大容量的知识库便于维护和管理。

### 三、知识组织的方法论和实现技术

从大容量知识库(LSKB)的要求和特点分析可知,在许多方面同大型的软件系统及数据库有许多相似之处。诸如,知识的大容量存贮、并行存取、多用户的可共享性以及可维护和可理解性。单就组织研究而言,可以从多方面借鉴软件工程和数据库系统设计的思想和技术。<sup>[2][8][4]</sup>

- 研究思想的借鉴。软件工程和数据库的许多系统设计思想在宏观上对大容量知识库的组织研究有很大的指导作用,诸如软件工程的结构化、模块化的系统设计,数据库的规范化模式,数据的物理组织结构,数据索引及数据字典等思想;

- 功能和技术上的吸收和增强,如耦合现有的数据库和知识库形成大容量知识库的存贮框架<sup>[5,6]</sup>。现有的DBMS的数据组织存贮的技术对规范化程度较好的知识的组织存贮也是非常有效的,因此我们以关系型的数据库管理系统为基础,并扩充以其它类型的组织存贮功能,作为大容量知识库组织存贮的后端来存贮知识的内容。

下面,我们介绍在开发ZKBE时所采用的一些关键技术,详细的分析和讨论可参见

有关文献<sup>[8][9]</sup>。

**A. 知识字典。**它用来描述基本知识的内部模式、知识的组织存贮信息及知识体之间的逻辑关系<sup>[8]</sup>。以Beam、Beam-Support-Type、I-beam、Beam-Section-Type四条知识所组成的局部机械CAD知识库为例,知识字典的主要信息如下:

(1) CLASS Beam

DESCRIPTION

Goal, "To transfer linear loads over a span to support"

ROLE

A-Kind-Of, Linear-Structural-Member

Partitions, (Support-type Section-type)

METHOD

Type, Rule

Base, Rule10

Position, Position11

ARGU

Base, Frame12

Position, Physical13

CONSTRAINT

Base, Cons14

Position, Cons15

(2) CLASS Beam-Support-Type

DESCRIPTION

Goal, "To provide the support system of a beam"

ROLE

A-Partition-Of, Beam

METHOD

Type, Rule

Base, Rule10

Position, Position21

ARGU

Base, Frame12

Position, Physical22

CONSTRAINT

Base, Cons14

Position, Cons23

(3) CLASS Beam-Section-Type

...

(4) CLASS I-Beam

...

显然，字典的信息体现了如下的知识库逻辑结构：

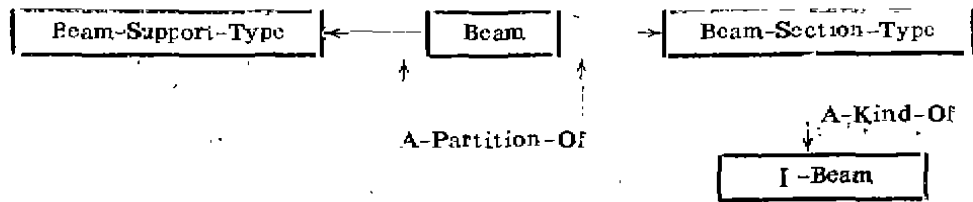


图3 知识库逻辑结构

字典信息对每个知识体内部提供了详细的内容，能有效地进行查询和利用。它不仅包含了知识库的全局模式，又能提供实际领域问题所需的逻辑视图，因此，知识字典是 ZKBE 组织和管理大容量知识的核心和基础。

**B. 知识体 CLASS 表层的规范化、结构化形式和内部非规范化的复杂形式的分离。**为了提高知识库的实际存取速度，可分别采用面向规范化和面向非规范化的知识存储技术。对于大容量知识库中的说明性、良结构和规范化的部分知识，利用关系数据库管理系统的技术来组织存储；而对于知识的非结构化部分，用 Item-方法转变成为 Item 形式，然后再用特殊的 Item-组织技术来组织知识<sup>[6]</sup>，其 Item 的物理存储结构如下：

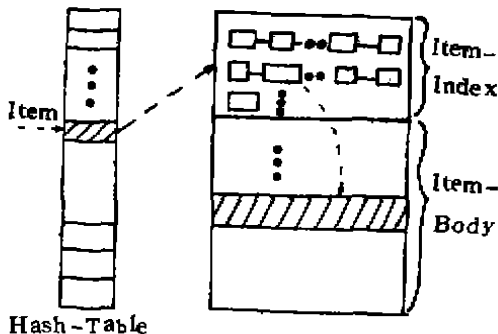
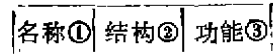


图4 Item物理存储结构

**C. 知识标识的语义化。**为了支持知识的高效查询和更新速度，我们采用一种称为语义化知识标识的方法。它的基本思想是将知识标识的编码语义化。在一般的知识库系统中，知识体的标识除能唯一地区别和确定整个系统中的某一个知识之外，基本上无其它作用。语义化的知识标识却不同，它还能体

现其在整个库中的结构位置和功能特性的作用。语义知识标识 (SKI) 码由三部分组成：



其中，①由用户取的名称；②刻画知识在知识库中的结构位置；③反映知识的功能和作用。有了 SKI 技术，我们可以进一步研究 SKI 的 HASH 函数，并使 HASH(SKI) 能体现知识的结构和功能属性。这样，知识在物理上的实际组织存储就能根据知识的结构和功能上的特性来安排，避免了原来的盲目堆砌，为高效的知识查询和维护铺平了道路。

#### 四、大容量知识库组织的体系结构

在第一代或早期的知识库系统中，知识库仅仅是作为知识体的简单堆积，很难甚至不可能满足大容量知识库的一系列新的特点和要求，尤其是难以有效地解决逻辑概念上面向问题领域，面向用户可理解性和物理组织存储实现高效性这对矛盾。为了克服传统知识库系统中的组织的弱点，我们从两个层次上来研究：

- 宏观层次上的知识组织，必须能体现领域问题知识自然的构成特点，同时充分表现知识间的丰富的语义关系；
- 微观层次上的知识组织，必须面向知

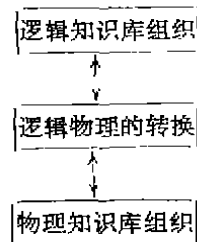


图5 多层次的组织体系结构

识体的具体的内部结构,以便于实现知识体的修改、删除和增加等操作。

因此,对于大容量的知识库,我们采用如图5所示的多层次的组织体系结构。

### 五、小结和展望

由于大容量知识库不同于传统知识库系统,有一些新的特点和要求,因此对大容量知识的组织研究带来了新的困难和挑战。大容量知识的组织结构需要考虑知识的存取效率、知识体的可分离性、多用户的共享及维护管理等多种因素。本文提出了组织的基本质量标准,并介绍在ZKBE系统开发中所采用的知识组织的方法论和实现技术。实践表明这些方法是有效的。目前,进一步的工作可以从以下两个方面深入:

1) 由于目前的存贮方法仍建立在O.S.文件系统的基础上,对速度造成了限制,因此需进一步深入O.S.的低层,对知识信息直接管理组织。

2) 结合具体的实际应用领域,如机械产品CAD系统,根据特定的知识表达模式进一步深入研究某些技术,取得经验,完善通用的大容量知识库组织的方法论和思想。

### 参考文献

[1] Michael L. Brodie, "Large Scale Knowledge-Based System, Concluding Remarks and Technological Challenges", On Knowledge Base Management System, Michael L. Brodie and John My-

lopoulos (ed.) 1986 by Spring-Verlag pp579—586

- [2] Amit P. Sheth, "Does Loose AI-DBMS Coupling Stand a Chance?", Readings in AI and Database, Morgan Kaufman San, Mateo CA (1988) pp252—254
- [3] Rob Jansen, "A Database Dictionary Approach to the Software Engineering of Rule-Based Expert System", Artificial Intelligent Developments and Applications, Gww J.S. and Stanton R. (Eds) 1988 by North Holland, Netherlands pp101—117
- [4] H. Craig Howard and Daniel R. Rehak, "KADBASE, Interfacing Expert Systems with Database", IEEE Expert, 4(3), (1989) pp65—76
- [5] Robert M. Abarbanel, et al., "KBE Connection, A Bridge between Database and Knowledge Bases"
- [6] Robert R. Heffman, "The Problem of Extracting the Knowledge of Expert", AI Magazine Summer 1987
- [7] 秦彤, "模式知识库理论与组织技术研究", 1990. 哈工大博士论文
- [8] 鲁东明, "第二代通用知识库系统及知识模型——大容量知识库系统研究", 技术报告, 浙江大学人工智能研究所, 1991年5月
- [9] 鲁东明, 何志均, "知识库系统Z KBE及大容量知识组织研究", 《知识工程进展1991》, pp84—92