

文章编号 1000-5013(2000)01-0092-04

# 基于书本知识的多媒体写作系统

陈永鸿<sup>①</sup> 陈一秀<sup>②</sup>

(① 华侨大学计算机科学系, ② 华侨大学机电工程系, 泉州 362011)

**摘要** 基于书本知识, 提出并实现一种高效率的多媒体信息获取方法. 利用在结构上类似自然语言的非常简单的计算机语言(SBL) 改写书本自然描述, 并应用一个简单的知识编译系统处理 SBL 程序, 便于高效率地实现书本知识获取. 在书本知识库的基本语义框架引导下, 采用多媒体求精系统对书本知识库进行少许求精并加入少量的其他媒介信息, 把书本知识库改装成符合专家个人经验的多媒体信息库.

**关键词** 书本知识描述语言, 章, 段, 知识编译, 语义图, 多媒体信息求精

**中图分类号** TP311.134.3

**文献标识码** A

目前, 多媒体写作系统面对庞大的信息源, 其信息获取效率很低, 出版一套多媒体信息库往往要花费巨大的人力物力<sup>[1~3]</sup>. 这个瓶颈的卡口处主要有2点. (1) 一个领域的多媒体信息大部分是文本信息, 它属于领域专家个人经验的比例很少, 大部分可在专业书本上找到. 同时, 这部分信息量非常庞大, 一般情况用原始的方式(几乎是手工)进行搜集和整理. (2) 对于其他媒介信息, 不但其搜集和整理方式非常原始, 而且在搜集和整理时比起文本信息更加杂乱无章. 为此, 本文探讨一种新的高效率的多媒体信息获取方法.

## 1 总体设计

如图1所示, 这种方法有如下5点要点. (1) 设计并实现一种书本知识描述语言 SBL, 它结构上十分接近书本自然语言, 但又是一种非常简单的计算机语言, 可以被无二义性地编译. (2) 把多媒体信息库的开发分为两个阶段, 即信息工程师(相对)独立工作阶段和领域专家(相对)独立工作阶段. (3) 在第一阶段, 由领域专家指定专业书籍, 信息工程师把专业书籍自然描述的知识用扫描仪高效自动地输入计算机, 然后, 把书本非常复杂的自然语言改为 SBL 语言. 它无需理解或读过这本书, 对概念的内容也无需改写, 仅把其当作 SBL 中的字符串, 且书本自然描述中描写概念内容的句子占篇幅的绝大部分. 另外, 对改写一些涉及概念之间关系的句子比从英文翻译到中文还容易, 仅需在句子中删掉、增加和顺序调整个别词, 同时把句子分词. 因此, 信息工程师可大段大段地快速改写. (4) 经过翻译的书本已经是一个简单的程序, 知识编译器

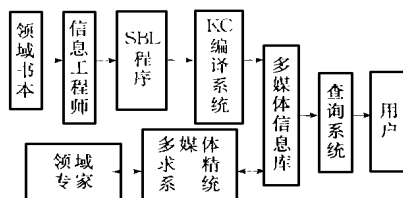


图1 基于书本知识的多媒体写作系统的体系结构

KC 利用高效和成熟的编译技术处理 SBL 程序, 自动地从中抽取领域知识来, 并组织成书本知识库。(5) 在第二阶段, 领域专家在书本知识库的基本语义框架的引导下, 利用多媒体求精系统对书本知识库高效地加入少量的其它媒介信息, 进行少许文本信息的求精, 把书本知识库改装成符合专家个人经验的多媒体信息库。

## 2 书本知识获取

### 2.1 SBL 语言

2.1.1 语法介绍 SBL 语言是非常接近自然语言的书本知识描述语言, 它的语法可以被认为汉语语法的一个子集。程序体对应于书本自然描述的结构, 而实际上程序体由一些章组成。章由章标题和章体组成。章标题与书本上的章节标题对应; 章体由一些子句、段和更小的章所组成, 对应于书本上自然描述的章节。段对应于书本上的对某一方面知识的集中描述, 一个段由段引导句、段体和段终结词‘段止’所组成, 段体则由一些子句、更小的段所组成。每个子句中的主要成分是字符串, 字符串表示概念的内涵或概念名, 有些子句还含有表示概念之间说明关系的一些关键词。书本自然描述的连续几个句子经常可直接用 SBL 中的一个字符串表示, 最简单的子句就是一个字符串, 用于表示概念的内涵。

2.1.2 语义说明 引导句必须被置于一句子(关键子句)之前或之后, 或者被置于段引导句之前, 这表明所引导的关键子句或段说明一个领域知识单元。另外还应注意, 引导句包括段引导句和说明引导句。为了解释说明引导句、章和段如何说明书本自然描述的结构及蕴涵的语义, 我们引入如下定义。

定义 1 术语: 除引导句以外的句子。

定义 2 术语标记: 章的标题与段的引导句子中的中心成分。用  $a$ ,  $b$  和  $c$  等来表示术语标记。

定义 3 术语标记的作用域: 若术语标记为一个章的标题, 则其作用域为该章体; 若术语标记为一个段的引导句的中心成分, 则其作用域为该段。对每个术语标记  $a$ , 它的作用域记为  $FD\ a$ 。

定义 4 若对于任意  $a$  和  $b$ , 有  $FD\ b$  包含于  $FD\ a$ , 则称  $a$  大于  $b$ , 记为  $a > b$ , 或  $b$  小于  $a$ , 记为  $b < a$ 。而且, 如果某章或段对应于  $b$ , 而另一个对应于  $a$ , 则称后者大于前者, 或前者小于后者。类似地, 可定义术语标记的最大者和最小者、以及章和段中的最大者和最小者。

定义 5 若对于任意  $a$ , 比  $a$  小的术语标记为  $b_1, b_2, \dots, b_m$ , 则  $FD\ a$  去掉  $FD\ b_1, FD\ b_2, \dots, FD\ b_m$  所剩区域中的构成  $a$  的内涵的术语称为  $a$  的固有术语, 记为  $DG\ a$ 。所剩区域中的构成另一些知识单元内涵的术语称为  $a$  的不定术语, 记为  $UG\ a$ 。

通过分析书本知识自然描述的特征, 从隐含在书本层次结构和上下文关系中的说明关系结构特征出发, 赋予几点说明引导句、章和段在这方面的语义。(1) 对于任意  $UG\ a$ , 若它的前(或后)加上说明引导子句, 则有下列两种情况。(i) 说明引导句的中心成分必须是下面两者之一: (a)  $UG\ a$  包含于其中的章、段的术语标记; (b) 包含该不定术语的章和段中的最小者, 其对应术语标记设为  $a$ , 则  $a$  为作用域中的且在上下文关系中, 处于该不定术语前面(或后面)的术语标记(或术语)所对应的领域知识单元名称。(ii) 该不定术语所在知识单元说明引导

句的中心成分所标明的知识单元. 对于段对应的术语标记, 若在段前加上说明引导子句, 也具有同样情形. (2) 对于某章, 若对应术语标记为  $a$ , 且含于该章中极大的段的术语标记为  $c_1, c_2, \dots, c_m$ .  $FD\ a$  去掉  $FD\ c_1, FD\ c_2, \dots, FD\ c_m$ , 剩下的区域中极大的段的术语标记为  $d_1, d_2, \dots, d_n$ . 则  $FD\ a$  去掉  $FD\ c_1, FD\ c_2, \dots, FD\ c_m, FD\ d_1, FD\ d_2, \dots, FD\ d_n$ , 所剩区域中没有加上说明引导子句的所有术语为  $DG\ a$ . 若  $d_1, d_2, \dots, d_n$  中对应的段前面没有加上说明引导子句的术语标记  $d_{i1}, d_{i2}, \dots, d_{in}$ , 则  $d_{i1}, d_{i2}, \dots, d_{in}$  各自对应的知识单元都说明  $a$  对应的知识单元.  $a$  对应的知识单元分别说明  $c_1, c_2, \dots, c_m$  对应的知识单元. (3) 对于某段, 若对应术语标记为  $a$ , 且含于该段中极大段的术语标记为  $e_1, e_2, \dots, e_r$ . 则  $FD\ a$  去掉  $FD\ e_1, FD\ e_2, \dots, FD\ e_r$ , 所剩区域中没有加上说明引导子句的所有术语为  $DG\ a$ . 若  $e_1, e_2, \dots, e_r$  中对应的段前面没有加上说明引导子句的术语标记为  $e_{j1}, e_{j2}, \dots, e_{js}$ , 则  $e_{j1}, e_{j2}, \dots, e_{js}$  各自对应的知识单元都说明  $a$  对应的知识单元.

## 2.2 多媒体信息库的总体结构

书本知识自然描述的主要结构是篇章结构, 这对应于 SBL 程序的说明引导句、章和段所组成的结构. 对应的多媒体信息库的总体结构, 是由概念以及概念之间的说明关系组成的一个语义图, 每个概念由三层媒介内涵组成. 其中的一种媒介内涵由文本(字符串)组成, 它主要来自书本知识, 另外两层媒介内涵为多媒体.

## 2.3 知识编译系统

采用成熟的编译技术进行书本知识获取. 整个编译过程紧紧围绕书本知识自然描述的篇章结构进行, 由此生成一个粗的书本知识库. 当编译进入下一层章节时, 如果该章名(概念名)不在概念名表中, 则在其后面添加一条新概念名记录, 建立上一层章名指向本层章名的说明关系. 在对本层的内容进行编译时, 则(1) 对于搜索到的每一个段, 如果该段名(概念名)不在概念名表中, 则在概念名表后添加一新概念名记录, 建立该概念名指向本章名(概念名)的说明关系. 如果该概念名是一个新概念名, 则该概念名的内涵(字符串)直接加在存放文本式概念内涵的文件之后, 并把其地址存入新概述名的记录中. 如果该概念名为一个旧概念名, 则根据旧概念名记录的地址字段从文本式概念内涵文件中找到其原有内涵, 并拷贝到该文件. 然后, 在其后再添加入新增的概念内涵, 同时把新地址存入旧概念名记录相应字段中. 对搜索到的该段中所含的更小的段, 如果它的前面没有加上说明引导句, 这时必须做如下处理: 如果该子段名不在概念名表中, 则在概念名表后面添加一新概念名记录, 最后建立本子段名指向上一层段名的说明关系. (2) 对于搜索到的每一个说明引导句, 把相应的具有说明关系的两个概念名加在一个辅助概念名表之后. 当编译完成之后, 才根据辅助概念名表中的说明关系完善概念名表中概念名之间的说明关系. (3) 对于搜索到的每一个章(更小的章), 其处理方式与上一层的章类似. (4) 本层的内容中, 不是段、说明引导句和更小的章的字符串是属于本层章名(概念名)的内涵, 对章名内涵的处理与(1)中段名内涵的处理方式类似.

## 3 多媒体求精系统与多媒体信息库的使用方法

由领域专家在多媒体求精系统的帮助下, 对书本知识库加入其他媒介信息及对本体信息的少许求精. 整个工作(多媒体求精系统的运行)是在书本知识库的基本语义框架(语义图)引

导下进行,其过程如下:依深度优先搜索,由代表书名的概念(无入弧顶点)出发,按出弧方向遍历各孩子顶点.同时,对遍历到的每个新顶点,围绕着它依深度优先搜索,按入弧方向遍历各“孩子”顶点.这两个过程是有机的统一体,是同一个算法的两个侧面,也是求精工作的主要过程.另外,求精系统还配有一导航仪,领域专家可借助此工具从代表书名的概念出发,迅速搜索到需要求精的语义图的顶点.用户可用两种方法使用多媒体信息库.一是类似求精系统工作的主要过程,另一是类似求精系统导航仪.在用户查询信息库的某一个概念时该概念的三层媒介内涵(如果有的话)要同步输出.显然,这两种查询信息库的方法效率极高.

## 4 总结

(1) 基于书本知识的多媒体写作系统的最大特色,是可以高效地获取信息.(2) 可以减少信息工程师和领域专家合作的紧密程度.(3) 直接从书本上获取大部分信息,以及少许文本信息由领域专家利用多媒体求精系统直接修改.这样可大大地缩短准备和输入领域信息源(此时还未装入信息库)的时间和费用.

## 参 考 文 献

- 1 刘正捷. 人机交互中的多媒体信息——希望与问题[J]. 计算机科学, 1995, (4): 5 ~ 8
- 2 刘隆达, 陈启泉, 陈维斌等. 多媒体文档系统 MDS-1 的设计与实现[J]. 华侨大学学报(自然科学版), 1999, 20(2): 200 ~ 203
- 3 董轩明, 徐兴佑. 支持协同编著系统的协作机制研究[J]. 计算机科学, 1996, (3): 33 ~ 36

## A Multimedia Writing System Based on Book Knowledge

Chen Yonghong<sup>①</sup>      Chen Yixiu<sup>②</sup>

(<sup>①</sup> Dept. of Comput. Sci., Huaqiao Univ; <sup>②</sup> Dept. of Electromech. Eng., Huaqiao Univ., 362011, Quanzhou)

**Abstract** The multimedia writing system now available fails to acquire efficiently the information from huge information source. Based on book knowledge, the authors present and implement an efficient method for acquiring multimedia information. As the very complicated first step, the natural description of book is rewritten by information engineers into SBL language which is similar to natural language in structure and is a very simple computer language. And then, the SBL programs are treated by using a simple knowledge compiling system so as to efficiently implement book knowledge acquisition. Finally, leading by the basic semantic framework of book knowledge base, domain experts refine somewhat book knowledge base and put in a little information of other media by using multimedia refining system, the book knowledge base is thus reformed into multimedia information base in keeping with experts' personal experience.

**Keywords** SBL language as describing language of book knowledge, chapter, segment, knowledge compiling, semantic graph, multimedia information refining