

# 论实现《中图法》可视化的几种途径\*

欧阳宁

(中国药科大学图书馆, 南京 210009)

摘要: 对近年来国内学者实现《中图法》可视化系统的几种途径进行分析, 讨论了各自的优缺点, 并提出建议。

关键词: 可视化; 《中图法》

中图分类号: G35

DOI: 10.3772/j.issn.1673—2286.2014.10.005

## 1 前言

信息可视化是用可视的形式表示抽象数据, 将大量抽象的信息映射到图像上, 以利于人们通过图像的结构、特征等来洞察数据、发现信息, 作出决策或解释数据。可视化将数据用有意义的图形表示出来, 为用户直观、方便地获取所需信息提供了有效途径<sup>[1,2]</sup>。

文献分类法的分类体系一般为等级列举式结构, 对知识的组织采用从总到分、从宽到窄、层层划分的方式, 形成比较纵深的等级结构, 其优越性在于较强的系统性和族性检索功能。但是典型的树型结构体系不能反映现代科学的交叉和综合性发展特征, 不能满足类目的多维检索, 在一定程度上限制了用户的多途径检索<sup>[3]</sup>。

1979年《杜威十进分类法》(简称DDC)第19版在其新出版的手册中, 用地图来显示DDC的地区复分表, 在地图的某一区域内标明其地区分类号, 便于用户查找, 这是可视化显示在分类表的首次尝试。到21世纪初, 在网络、计算机技术的支持下, 分类法的可视化技术也迅速发展起来, 可视化系统主要采用树(Trees)、图(Graphs)及虚拟现实(Virtual Reality)等隐喻方式, 知识组织工具包括主题地图(Topic map)以及本体(Ontology)等<sup>[2]</sup>一系列可视化显示技术。目前国内外

学者对不同的文献分类法已经进行了多种尝试, 例如: 由丹麦、芬兰、德国、荷兰、瑞典、英国的图书馆与研究发展中心共同合作开发的集成化的网络信息资源门户网站Renardus项目, 将自身网站的结构和DDC结合成一个开放的公用系统; OCLC研究署的研究项目OverView, 采用了信息可视化技术, 将检索结果表现为由虚拟本体模型语言(VRML)构筑一个DDC的三维信息空间, 用柱形图显示检索结果在各个类的相关度<sup>[4]</sup>; 还有国内的《中国分类主题词表》电子版, 检索系统界面上各种浏览方式之间和各窗口之间可随意跳转, 用户可根据个人习惯和查询需求自主切换窗口; 这些例子对于其他体系分类表的可视化, 都具有很好的借鉴意义。

## 2 实现《中图法》可视化的几种途径

《中国图书馆分类法》(简称《中图法》)是一部大型列举式等级体系分类法, 该分类法历经五版, 相对成熟完善, 是国内图书馆编目标引、检索查找和排架的权威工具之一; 在文献分类、编制检索工具、组织排架等方面有广泛应用; 是我国图书馆与情报单位普遍使用的一部综合性分类法<sup>[5]</sup>。国内学者们对《中图法》的可视化一直在进行研究, 采用的技术也有所不同, 本文对近十年来国内学者研究出的几种《中图法》可视化检索系统进行总

\*本研究得到江苏省现代教育技术研究2013年度课题“基于用户服务的《中图法》网络化研究”(编号: 2013-R-25888)资助。

结,分析一下各自的优缺点,以期能抛砖引玉,对今后的研究有所帮助。

## 2.1 基于本体的《中图法》可视化系统的实现

基于本体的可视化系统,首先是利用本体编辑工具Protégé对《中图法》大类进行了本体构建,以W3C标准中的OWL语言作为本体的描述语言实现了对本体的描述并保存为XML文件格式,然后利用Jena的持久性存储组件对本体文件进行存储,在创建本体模型时采用了OWL推理。最后借助Java开发平台,设计了可视化查询系统的结构框架,实现了《中图法》类目的可视化检索系统(如图1)<sup>[6]</sup>。

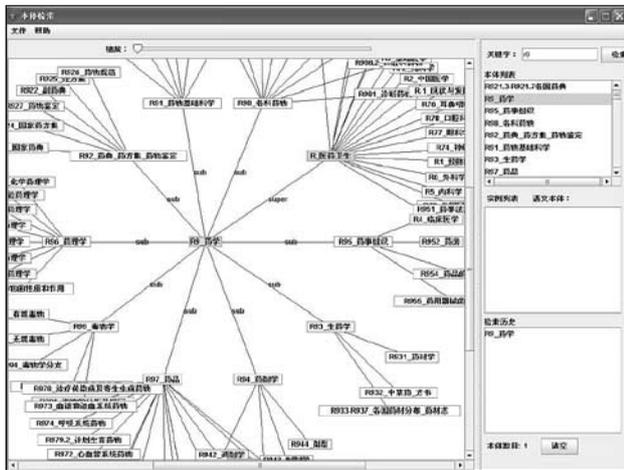


图1 基于本体的《中图法》可视化检索系统

因这个系统是本人研究设计的,所以对其不足之处感触较深。此可视化系统采用XML文件方式来进行本体库的存储,每次调用不同的文件数据都需要修改文件的URL来进行操作,不适合较大的本体库。而且在系统的检索途径方面,随着检索点的增多,内存占用越来越大,调用速度很慢,达不到的效果。另外,本体工具Protégé毕竟是国外开发的软件,设计思路和原则与《中图法》有一定的出入,并且部分插件不支持中文输出,需要对其进行改造。

## 2.2 基于双曲几何的《中图法》可视化系统的实现

基于双曲几何的可视化系统,是使用了J. Lamping和R. Rao的“基于双曲几何的大型层次结构的可视化”

的理论,采用了信息可视化中著名的“focus+context”(聚焦+关联)技术,主要思想就是将层次关系显示在一个双曲平面上,使得既能将更多的显示空间给予当前层次结构中受关注的部分,同时又能显示整个层次结构。在这个系统中使用Java的开放源代码HyperTree,将《中图法》分类体系存储成XML文件,然后将XML文档解析为DOM,使用JAXP(Java API for XML Processing)技术实现对文档内容的检索。针对用户的输入,系统对整个DOM树进行遍历,查找出所有匹配结果(如图2)<sup>[7]</sup>。

这个系统可以清楚地看出局部类目在整个分类体系中的位置,用户可以通过点击任何可见的节点,或鼠标拖拽可见点到任何位置。但是本人认为不足之处在于DOM技术在解析XML文档时需要把整个结构化的XML文档加载到内存之中,建立一个反映XML文档间层次关系的DOM树,而《中图法》的类目体系非常繁杂,所以使用传统的DOM技术来解析以XML格式存储的《中图法》分类体系存在数据处理效率低、与用户交互功能简单等缺点。

## 2.3 基于SAX解析的《中图法》可视化系统的实现

基于SAX的可视化系统,集成了传统树型结构简单易懂的特点和新技术的动态交互功能,采用SAX解析技术来处理以XML形式存储的《中图法》,并用JAVA技术实现了放大、缩小、拖拽、检索、颜色定位等

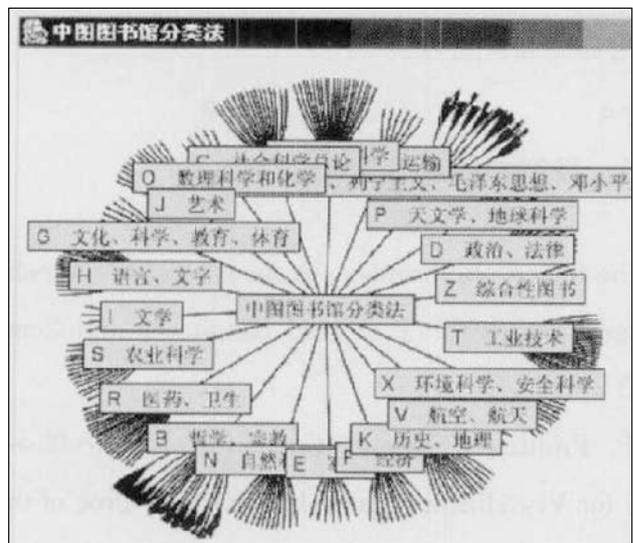


图2 基于双曲几何的《中图法》可视化检索系统

交互功能,利用开源的动态交互可视化工具Prefuse,参考IVRM模型,实现了动态交互的《中图法》可视化系统(如图3)<sup>[8]</sup>。

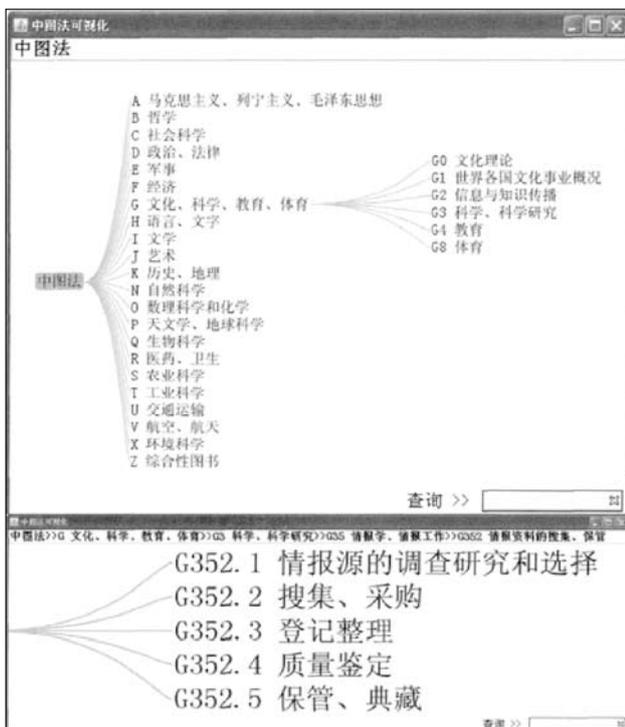


图3 基于SAX解析的《中图法》可视化系统

这个系统使用了处理XML文件时占用内存小、速度快的SAX技术,克服了DOM技术的占用系统资源大、效率低、速度慢等缺点,可以用来解析例如《中图法》这样拥有庞大类目体系的大型的、复杂的XML文档。但是,本人认为此系统的不足在于,没有突破分类法传统的树形结构,类目之间的关系只是简单的定义,相关类目之间的关系需要进一步的明确。另外如果增加用户的检索历史查询,可以更好地实现系统与用户的动态交互功能。

### 2.4 基于放射树的《中图法》可视化系统的实现

放射树能够达到与双曲树类似的效果,但是放射树的空间利用率更高,展示的效果比双曲树好。基于放射树和传统树的可视化检索系统,利用数据库技术来存储《中图法》的类目体系,采用了一个开放源代码的对象关系映射框架hibernate技术,然后利用Java语言编写的开源可视化工具Prefuse和JUNG (Java Universal Network/Graph Framework) 来对数据进行建模、分析

和可视化,实现《中图法》可视化系统(如图4)<sup>[9]</sup>。

基于放射树的系统提供了多种形态的放射树,能够根据用户的选择来可视化分类体系,而不仅仅是可视化整个分类体系;而且图形之间可以直接转换,便于用户随时以各种形式查看类目体系。另外数据库持久层使用了hibernate技术,整个系统没有写具体的SQL语句,对数据库的访问都是通过数据对象的方法,大大减少了代码的编写量,同时也提高了系统的运行速度。此系统实现时参考了其可视化模型和动态交互式技术,放射树提供了放大、缩小、关联邻近节点、显示节点的详细信息等交互式操作。本人认为这是以上几种实现《中图法》可视化方法中最为成熟的一个系统。但是可能《中图法》类目体系太过庞大与复杂,所以在有限的显示空间中显示海量信息的整体视图时,可视化视图不够清晰和直观,有时会出现显示信息不完整的现象,不是所有的类目及其上、下位类都能全部显示出来;但是双击图中任一类目,可以链接到新窗口查看该类目的详细上、下位类。

### 3 结语

从以上四种途径可以看出,不同的可视化算法所要求的格式不统一,所用的存储方式和解析方法各有不同,生成的可视化系统也各有其优缺点。本文最后提出两点建议:①《中图法》的类目众多,大类少则数千,子类多则数万;而且类目的上下位之间还要有统一性、继承性,类目之间有纵向、横向关系,这就需要类目之间不只是简单的定义,除类名、类号及上下位关系

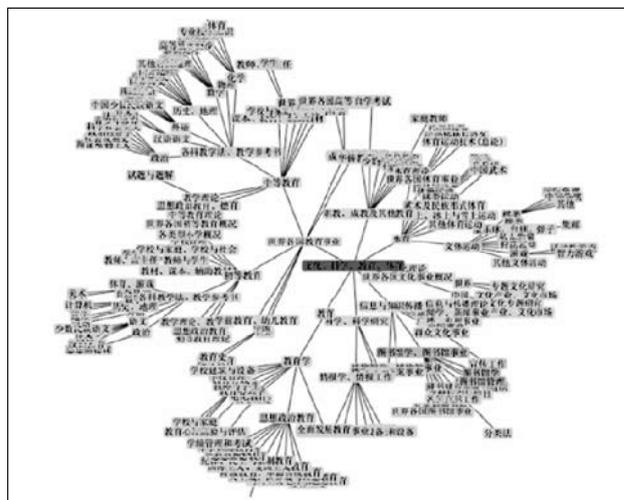


图4 基于放射树的《中图法》可视化检索系统

外,还要为类目定义若干的属性和注释,而且类目若能设计出有针对性的、特色鲜明的图符,就具有比较丰富的语义,若将此《中图法》可视化检索系统应用到图书馆的OPAC系统中,与用户之间的互动也更完美。②以文件方式存储数据轻便快捷,因为它不需要过多的配置,而且便于备份、复制,还可以随时编辑修改,而且查询速度快,所以适合于小型的库。这也是本文中讨论的三个可视化检索系统都采用了以XML文件存储的原因。基于数据库的存储方式虽然需要复杂的数据库配置,但是比文件存储方式能处理更大的模型,而且不需要显式的保存数据模型,效率更高。考虑到系统的可扩展性,本人认为《中图法》的可视化系统应采用基于数据库的存储方式,因为利用数据库可以生成符合要求的不同类型格式的《中图法》,用于不同的可视化算法。同时因为不用直接将整个树加载到内存中,而直接从数据库中查询和检索,利用数据库可以提高查询和处理的效率。但是基于数据库的存储方式对服务器等硬件的配置以及后期的维护要求都比较高,希望学者们在今后的研究中对这一方面能进行进一步的尝试和探讨。

本文通过总结《中图法》可视化的几种途径,旨在抛砖引玉,希望能在目前的研究基础上,随着以后进一步工作的展开,通过不断的研究,逐步地解决有问题的

地方,综合各家之长,逐步地完善检索系统,最终能找到最适合《中图法》的可视化方法。

#### 参考文献

- [1] MERKL D. Text classification with self-organizing maps [J]. Wien: Neurocomputing, 1998(21): 61-77.
- [2] KOSHMAN S. Visualization-based information retrieval on the Web [J]. Pittsburgh: Library & Information Science Research, 2006(28): 192-207.
- [3] 杨达.数字图书馆信息可视化的研究框架[J].沈阳教育学院学报,2005(3):127-130.
- [4] 王绍平.数字图书馆中的文献分类法[J].上海交通大学学报,2003(增刊):131-138.
- [5] 靖培栋.《中图法》可视化研究[J].中国图书馆学报,2009(3):42-46.
- [6] 欧阳宁,胡飞燕.基于本体的《中图法》类目可视化查询系统的设计[J].图书情报工作,2009(3):43-46.
- [7] 靖培栋,郭宇峰.基于双曲几何的《中图法》分类体系的可视化[J].现代图书情报技术,2005(1):40-42.
- [8] 李国俊,肖明.基于SAX解析的《中图法》分类体系可视化系统的实现[J].大学图书馆学报,2010(5):31-35.
- [9] 李国俊,肖明.基于放射树的《中图法》可视化实现[J].图书情报工作,2012(4):81-85.

#### 作者简介

欧阳宁,女,1980年生,硕士,中国药科大学图书馆馆员,研究方向:数字图书馆、专利,E-mail: oynwj@163.com。

### Comparative Study on Approaches to Visualization of Chinese Library Classification

OUYANG Ning

(Library of China Pharmaceutical University, Nanjing 210009, China)

Abstract: This paper analyzes domestic scholars' several ways to achieve the visualized query system of "Chinese Library Classification" in recent years, discusses their advantages and disadvantages, and puts forward some suggestions.

Keywords: Visualization; Chinese Library Classification

(收稿日期: 2014-07-14)