

《中图法》可视化研究*

靖培栋

摘要 首先用XML对《中图法》分类体系进行存储,利用Java与DOM实现此体系的截词检索,并且集成普通树型和双曲树模式,采用普通模式、长类名模式,以及在显示区域内的节点上停留鼠标显示其详细类目标名称的方式显示检索结果,对整个分类体系及检索结果实现可视化,促进了《中图法》分类体系的方便使用。图4。参考文献13。

关键词 信息可视化 《中国图书馆分类法》 截词检索 双曲树
分类号 G350.7

ABSTRACT This paper does the research on the storage of *Chinese Library Classification* system using XML. The author realizes the word truncation retrieval of the system by Java and DOM, and makes the visualization of *Chinese Library Classification* and the visualization of query results. 4 figs. 13 refs.

KEY WORDS Information visualization. *Chinese Library Classification*. Word truncation retrieval. Hyperbolic tree.

CLASS NUMBER G350.7

《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,是我国图书馆与情报单位普遍使用的一部综合性分类法。该分类法历经四版,相对成熟完善,在文献分类、编制分类检索工具、组织文献排架等方面有广泛应用。

信息可视化是在计算机的协助下,对信息的可见的、交互的表示。可视化的目的是洞察数据、发现信息、作出决策或解释数据。信息可视化的应用领域是非常广泛的,可以应用可视化的信息类型也是多种多样的,如多维信息、层次(树型)信息、网状信息与文献信息等。我国情报学术界对信息可视化系统设计与实现的探讨还比较少。《中图法》分类体系是典型的大型层次结构,笔者按照信息可视化流程对此体系进行了可视化研究。笔者使用XML与Java技术实现了《中图法》分类体系的可视化检索系统,首先用XML对《中图法》分类体系进行了存储,利用Java与DOM实现了此体系的截词检索,对整个分类体系及检索结果实现了可视化,是文献[1]中结果的发展与完善。

1 层次信息可视化

大型层次信息的可视化研究主要解决大型层次信息显示和计算机显示器屏幕尺寸有限之间的矛盾,寻找最佳的可视化表示方式。传统的显示方式是将整个层次信息结构显示在一个比显示器更大的显示区域内,利用滚动条等可视化控件来调整可见区域。该方法的问题在于,如果不借助辅助视图,用户在关注局部层次信息的同时,不能同时看到该局部与整个层次结构的关联。层次型结构的可视化多采用树型,如普通树型、双曲树(Hyperbolic tree)^[2]、圆锥树(Cone tree)^[3]、树图(Treemap)^[4]、空间树(Spacetree)^[5]等。笔者根据中图法的特点,采用普通树型及双曲树对其进行可视化。普通树型的优点是可以清楚地展示局部类目标细节,缺点是不易看出局部类目标在整个分类体系中的位置;双曲树的优缺点与普通树相反。因此,笔者将普通树与双曲树集成,对《中图法》分类体系进行可视化。

* 本文为国家自然科学基金资助项目“信息可视化研究”(项目编号:02BTQ019)成果之一。

2 信息可视化流程

信息可视化流程是将信息转换为能够与用户交互的可视形式的计算过程,具体步骤如图1所示^[6]。文献[7]中的流程模式流传较广,在我国学术界也有介绍^[8],图1给出的是对它的改进模式。

下面对图1中的步骤作一些说明^[6]。

第一步,将原始信息转换为组织良好的规范的数据形式,这种形式通常构成包含数据实体的数据集,每个数据实体具有多个属性值。根据需要,各种数据处理步骤(如:数据挖掘)可用于处理数据,导出数据往往对于洞察结果有帮助。

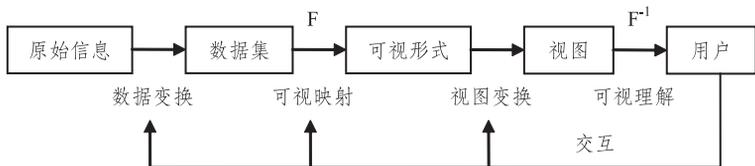


图1 信息可视化流程

第二步,将数据集映射为可视形式,是可视化过程的主要步骤,可视形式包含对应于数据集实体的标记(glyphs)。

第三步,将可视形式嵌入到视图,视图显示在屏幕上,并提供各种视图变换,如导航。视图通过人的视觉系统呈现给用户。

第四步,用户能够与上述任何一步进行交互,以改变可视化结果并作进一步解释。

上述第二步中的可视映射是可视化的核心,应小心设计。此映射的作用是人机通信,通信的媒介是信息的可视表示。通过某函数 F ,数据集被映射为可视形式,此函数将数据集作为输入,将可视表示作为输出。当可视表示传递给用户时,他们必须能从认知上对可视映射求逆,以从可视表示解码信息。

3 《中图法》可视化检索系统的实现

我们按照上面的可视化流程,实现了《中图法》分类体系的可视化检索,开发出了《中图法》可视化检索系统原型。首先将《中图法》原始信息转换为组织良好的规范的数据形式。由于《中图法》分类体系是层次信息,用XML表示这一体系是自然的选择。

3.1 《中图法》的XML文档存储

我们为了将《中图法》分类体系存储成XML文件,制定了下面的规则:

- 1) 引进标记 \langle 中国图书馆分类法 \rangle , 作为XML文档的根元素;
- 2) 《中图法》中的一个类对应XML文档中的一个元素,用分类号作为XML文档中的标记名,类名作为相应标记的文本;
- 3) 若B是A的子类,则B对应的元素是A

对应的元素的子元素;

- 4) 对于有类目注释的类,给其对应的元素添加属性note,note的值就是类目注释。

由于XML标记名称必须以字母或下划线开头,名称的剩余部分可以由字母、数字、下划线、点或连字符组成,有些分类号不符合XML文档标记的命名规则,不能直接作XML标记名。为此,我们做了一些处理。例如:

对于有中括号“[]”的分类号(交替类号),我们将“[]”去掉作为标记名,为此标记增加属性another,another的值是目标类号;

对于包含“/”的分类号,我们选择“/”前的部分作为标记名,为此标记增加属性to,to的值是“/”后面的数字。

3.2 《中图法》分类体系的检索

我们将《中图法》的XML文档解析为DOM,使用JAXP(Java API for XML Processing)技术^[9]实现对文档内容的检索。对于类名我们实现了

截词检索,截词可以是后截断、前截断与中截断^[10]。对于分类号也实现了后截断检索。用户通过系统界面,每次可以输入一个查询,输入可以是一个中文词语或者是一个要查找的分类号。若输入是一个中文词语,则查找出相匹配的类目;若输入是一个分类号,则查找出相匹配的分类号。针对用户的输入,系统对整个 DOM 树进行遍历^[9,11],查找出所有匹配结果。遍历就是对 DOM 树的每个节点进行验证,我们设计了 DOM 树的遍历算法。

3.3 可视形式

作者对《中图法》整个分类体系给出了两种可视形式,可以结合使用。一种是分类体系的树型结构表示(称为树型视图),另一种是分类体系的双曲树表示(称为双曲视图)。对于查询结果,我们给出了从根类目到每个查询结果类目的路径,并将这一路径用可视的形式表示出来。

(1) 树型视图

层次信息的可视化,最直观的方式就是使用树型结构。许多层次结构,像组织机构、目录结构,往往非常庞大,不能将它们整个显示在计算机屏幕上。常规的显示方式是将整个层次信息结构显示在一个比显示器更大的区域内,利用滚动来调整可见区域。我们正是使用这一技术,利用 Java Swing 中的 JTree 组件实现了《中图法》分类体系的树型结构表示,作为《中图法》整体结构的一种显示方法。该方法的问题在于,用户无法看到可见部分与整个层次结构的关联。

(2) 双曲视图

J. Lamping 与 R. Rao 实现了基于双曲几何的大型层次结构的可视化^[7],其主要思想就是将层次关系显示在一个双曲平面上,这样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,同时又能够显示整个层次结构,这就是信息可视化理论中著名的“focus + context (聚焦 + 关联)”技术^[12]。可以用一种自然的方式将双曲平面映射到单位圆上,此映射使得在单位圆中,给予距离圆心越近的部分更多的显示

空间,距离圆心越远,分配更少的空间。将层次结构信息显示在此圆上,正好聚焦于圆心附近,在双曲平面上移动焦点,就是在双曲平面上平移层次结构。

我们以 J. Lamping 与 R. Rao 的研究为基础,使用 Java 开放源代码 HyperTree^[13],实现了《中图法》分类体系的双曲树表示。源代码不是针对 XML 文档的,没有 XML 接口。由于本系统采用 XML 存储《中图法》类目,为此,我们开发了系统的 XML 文档接口。针对《中图法》的情况,我们对源代码的功能进行了扩充,并对可视效果进行了改进。下面 4.2 节中的长类名显示模式和鼠标停留在节点上时显示完整类名的模式,是笔者的发展,取得了更好的可视效果。

(3) 检索结果的可视表示

对于用户输入的查询词,检索出的结果一般是多个类目。对每个查询结果类目,首先通过 DOM 接口与 JAXP 找出其父类,再找出父类的父类,直到根类为止。我们使用 Java 中的 Graphics 技术,对于每个查询结果类目,用自定义的标签序列,显示出从根到结果类的路径,标签之间用从父类标签到子类标签的箭头连结。可以通过这个序列,在树型结构或在双曲树中查看更详细的信息。检索结果的可视化是笔者设计的,这样可以将检索结果非常清楚地展示给用户。

4 系统功能与可视化效果

我们将“中国图书馆分类法”作为根结点,使用上一节中描述的方法,实现了《中图法》可视化检索系统。图 2 是系统界面,图中双曲树聚焦于根结点,根结点在圆心处,类名显示采用长类名。

4.1 检索功能及结果可视化

在系统界面的文本框中输入一个查询,然后点击查询按钮或者按回车键,就可以得到查询结果。图 3 是输入“情报学”后的查询结果,共有三个匹配类目,图中显示的是从根到相应类目的路径。

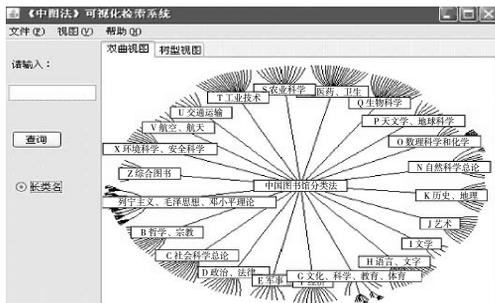


图2 《中图法》可视化检索系统界面

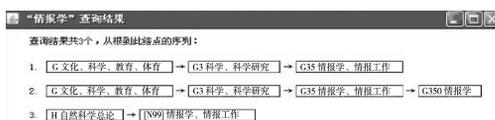


图3 “情报学”查询结果可视形式

得到以上结果后,可以在双曲视图或树型视图中查看相应类目的详细内容,图4是“情报学”在树型视图中的情形。

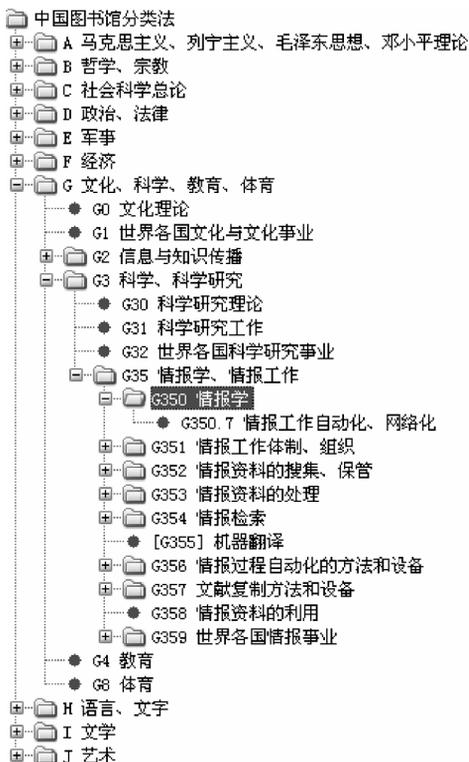


图4 树型视图

4.2 双曲视图效果

(1) 焦点转移

用户可以点击任何可见的节点,将其移动到显示区域的中心,或者用鼠标拖拽任何可见点到任何其他位置。两种情况下,整个层次结构的显示都会进行相应的调整。移到中心位置的部分将被放大,周围的部分则越接近区域边界而逐渐缩小。这就是“聚焦+关联”的效果。焦点的转移是通过双曲平面上的平移变换,将选定映射到中心来实现的。

在信息可视化应用中适当引入动画技术,使得不同显示视图之间用动画平滑过渡,可以保持对象变换的连续性,充分利用用户的视觉感知能力,生动感受视图的变换。在双曲平面上,通过适当的中间视图来体现焦点的平滑过渡。动画序列通过 n 次平移来实现,即用 n 次平移实现本来的一次平移, n 次平移的过程产生了中间视图。

(2) 类名显示

对于节点信息的处理,实现了普通模式、长类名模式,以及在显示区域内的节点上停留鼠标就可以显示其详细类目名称的模式,为用户使用提供了多种选择。在普通模式中,显示类名长度随空间的大小而变化,此方式视觉效果较好。在长类名模式中,若无遮挡,可以显示完整类名,不过这种显示方式看上去比较拥挤。当鼠标停留在节点上时显示完整类名,是前两种显示方式的折中。

5 结语

《中图法》可视化检索系统的建立,为《中图法》分类体系的使用带来了极大的方便。用户通过输入关键词可以查询到相关类目,通过输入分类号可以查询到相应类目信息。通过整个类目体系与查询结果的可视化表示,不仅可以提高《中图法》使用效率,还可以激发用户的使用乐趣。

参考文献:

- [1] 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.
- [2] John Lamping, Ramana Rao, Peter Pirolli. A Focus + Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies[C]. Proc of the ACM SIGCHI Conference on Human Factors in Computing System, 1995.
- [3] G. G. Robertson, J. D. Mackinlay, S. K. Card. Cone trees: Animated 3d Visualizations of Hierarchical Information [C]. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM, April 1991: 189 - 194.
- [4] B. Johnson, B. Shneiderman. Treemaps: a Space-filling Approach to the Visualization of Hierarchical Information Structures[C]. In Proc. of the 2nd International IEEE Visualization Conference, October 1991: 284 - 291.
- [5] Plaisant. C. , Grosjean. J. , Bederson. B. B. SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation[J], INFOVIS 2002. IEEE Symposium on Information Visualization, 2002 . Boston, 2002 (10): 57 - 64.
- [6] G. Salvendy. Handbook of Human Factors and Ergonomics[M]. New York: John Wiley & Sons, 2005: 1222 - 1246. .
- [7] Stuart K. Card, Jock D. MacKinlay, Ben Shneiderman (Eds) Readings in Information Visualization: Using Vision to Think[M]. San Francisco: Morgan Kaufmann Publishers, 1999.
- [8] 隆捷. 信息检索可视化研究[J]. 情报学报, 2006, 25(5): 365 - 368.
- [9] E. R. Harold. Java 语言与 XML 处理教程[M]. 北京: 电子工业出版社, 2003.
- [10] 赖茂生等. 计算机情报检索(第2版)[M]. 北京: 北京大学出版社, 2006.
- [11] 严蔚敏, 吴伟民. 《数据结构》(C语言版)[M]. 北京: 清华大学出版社, 1997.
- [12] J. Lamping, R. Rao. The Hyperbolic Browser: A Focus + Context Technique for Visualizing Large Hierarchies[J]. Visual Languages and Computing, 1996, 7(1): 33 - 55.
- [13] HyperTree Java Library [EB/OL]. [2008-08-10]. <http://hypertree.sourceforge.net/>.

靖培栋 北京师范大学管理学院教授, 博士生导师。通讯地址: 北京师范大学。邮编 100875。
(收稿日期: 2008-09-17)