

信息可视化技术在数字化图书馆中的应用

徐 刚 徐 聪

[摘 要] 信息可视化是情报学研究的新领域,它的核心是将现实的空间形式引入因特网,将现实的空间与虚拟现实技术叠加一起。将信息可视化技术引入到数字图书馆领域,解决信息需求与服务的个性化、信息提供的个性化等问题,可以通过信息可视化在一定程度上尝试解决发展中遇到的问题。

[关键词] 信息可视化;数字化图书馆;信息资源

[中图分类号] G230 [文献标识码] A [文章编号] 1672-7320(2008)06-0919-03

信息可视化是情报学研究的新领域,其核心是将现实的空间形式引入因特网,将现实的空间与虚拟现实技术叠加一起。在虚拟环境的平台上,利用可视化编程技术,通过将数据图形化、地理化形象真实地表现出数据背后蕴含的信息。信息可视化技术可实现对信息数据的分析和提取,以图形、图像、虚拟现实等易为人们所认识、可视的方式,展示数据间的复杂关系、潜在信息和发展趋势,以便更好地掌握和利用信息资源。

一、信息可视化概念与研究发展过程

信息可视化是目前情报学研究的热点。1786 年苏格兰 W. Playfair 是最早利用线和面作为可视化表示数据的人。1967 年法国 J. Bertin 发表了图形理论。1983 年美国 E. R. Tufte 发表了数据图形理论。1987 年美国国家科学基金发表了一份研究报告《科学计算中的可视化》,标志着科学可视化的产生。“信息可视化”最早出现在 1989 年 G. Robertson S. Card 与 J. Mackinlay 的论文中,成为一个与科学可视化并列的研究领域^[1](第 1 页),他们在科学可视化的基础上,运用计算机图形学和图像处理技术,把数据转换成图像在屏幕显示出来,并进行交互处理的理论、方法和技术。国外对信息可视化进行了广泛而深入的研究,并取得了丰硕进展。比较有影响的国际会议是国际电气与电工工程师学会 IEEE(Institute for Electrical and Electronic Engineers)所组织的两个系列国际研讨会:一个是 1997—2002 年间在英国每年 7 月召开的“International Conference on information Visualization”;另一个是 1995—2002 年间在美国召开的“IEEE Symposium on Information Visualization”每年一度的国际研讨会。这两个国际会议集中代表了当代该领域的研究水平。讨论的主题主要有:信息可视化、知识可视化、可视化应用、可视化与图形、可视化与数据库、地理信息系统可视化、虚拟现实技术的扩展与应用等^[2](第 4-7 页)。

(一)信息可视化与科学可视化

信息可视化与科学可视化的主要区别在于:信息可视化是通过计算机支持的、人机交互、视觉的表现形式,对抽象数据的认识的扩展。

科学可视化是指空间数据场的可视化,人们需要在计算过程和数据处理流程中了解数据的变化,然后通过图形、图像、图表以及其他可视化手段来检查、分析处理结果数据^[3](第 94-96 页),主要通过观察基于物理的、几何属性的数据,并且用图形、图像来解释海量数据。科学可视化适于高层次专业人士,而信息可视化的主要用户是非专业技术人员,信息可视化使用信息服务必须摒弃专业的操作程序和格式,采取大众化的可视化方式,它是信息生产与消费的发展趋势。

(二)信息可视化与知识发现和知识管理

数据挖掘又被称作数据库中的知识发现,从技术上讲,则是从大量的、不完全的、有噪声的、模糊的、随机的实际应用

数据中提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程。

数据的挖掘可视为从各种网络资源中,例如政府信息服务、数字图书馆、电子商务数据及数据库中,利用信息可视化技术找到有用信息的过程。Web 的挖掘内容对象包括文本、图像、音频、视频、多媒体和各种类型数据。利用 Web 结构挖掘是来挖掘 Web 潜在链接的结构模式,建立自身链接结构。

二、信息可视化在数字图书馆中的应用

20 世纪 90 年代初美国科学家提出数字图书馆概念后,以驱动多媒体海量数字信息组织与互联网应用问题各方面研究的技术领域开始在全球迅速发展起来^[4](第 32 页)。数字图书馆以电子格式去存储海量的多媒体信息并对这些信息资源进行高效操作,如插入、删除、修改、检索、提供访问接口和信息保护等。经过数十年的发展,已经进入一个相对稳定的阶段,同时也面临着多方面的挑战。将信息可视化技术引入到数字图书馆领域,解决信息需求与服务的个性化、信息提供的个性化等问题,可以通过信息可视化在一定程度上得以解决。

(一)信息检索过程可视化

信息检索由两个步骤组成:构建和使用。用户作为信息使用者,同时也是信息的构建者,通过对增加检索路径到信息空间,这些增加的路径给其他用户检索其他路径提供了有价值的信息。

在运作中,首先通过邻近一映射方法设计空间一语义界面,通过信息可视化分类法来对检索路径进行分析,主要采用 Shneiderman 的信息可视化分类法,对每个用户的行为根据检索过程进行精确测量,然后通过隐性马尔可夫模型(HMM)对信息检索行为进行序列行为分析,在这个过程中包括网络寻址(Pathfinder Network Scaling)和多维尺度(Multidimensional Scaling)等可视化技术,成功的用户检索路径将插入隐性马尔可夫模型程序中,最后通过隐性马尔可夫模型和在相应的语义空间中动态生成合成的用户路径。

(二)信息检索结果可视化

信息可视化技术在交通运输、航空航天、科学研究、通信、工业生产,包括在地理信息系统、城市地图、建筑平面图、三维军事地形图、区域规划、天气预报、数字艺术、医学信息、网络图形、远程教育、生态环境、图形图像、多媒体、动漫、虚拟现实技术、数字图书馆等领域得到初步的发展和利用。在信息管理系统中,信息检索的结果可视化所取得的结果最引人注目。例如,美国马里兰大学的人机互动实验室可视化 See Soft 系统、Xerox PARC 的科研人员开发的数字图书馆研究^[5](第 685 页)。

(三)知识发现可视化

数字图书馆理论和实践的不断成熟,可供获取的知识也在不断地成几何指数增长,研究知识和科学的结构、科学发展的动力(发展速度、发展模式等)、寻找科学前沿等问题,越来越受到人们的重视。信息可视化技术的发展为研究知识领域的相关问题提供了新的思路,尤其是在最近几年时间里,将信息可视化技术与科学计量学方法相结合,生成具有各种属性的科学地图,表达学科、领域、专业、文献、著者之间的关系、解释知识领域的结构、映射知识领域的发展趋势,促进信息获取,使知识结构变得更加明显,帮助用户达到他们的目的。在国外有的学者将其称为科学映射(Science Mapping)。基于动态数据分布的并行 Shear-Warp 体绘制技术是科学计算可视化的一个重要手段,它是处理三维数据内部结构的信息的有效方法,尽管它无法满足交互分析,但在探索规模的科学数据的要求,快捷、实时的体绘制技术成为可视化的重要目标。信息可视化理论内容相当丰富,例如多维数据可视化技术、动态探索技术、“聚焦+关联”技术、信息检索与可视化、文献信息可视化、层次信息可视化、网络数据可视化等。

(四)数字图书馆的文献信息检索结果可视化

1. Scatter/Gather。Scatter/Gather 系统是检索结果可视化中基于分类的文档簇法。这是通过使用一个活动的目录表来帮助用户掌握大量搜集的文档,系统使用文档聚集把搜集的内容分解成小数量的连贯的文档簇,并通过这种简短概述的文档簇法形式表达给用户^[6](第 7 页)。基于这些概述,用户选择一个或更多的文档簇以供将来学习使用。将这些选择出的文档簇再进行聚集或联合并形成一个二次搜集,系统重新运用聚集技术来分散新的二次搜集以形成一个新的文档,再将这些依次表达给用户。在每一次连续重复下,这些文档簇将变得更小,因而也更具体化。

文档聚集法充分运用速度来加强相互作用,而不能保证准确度。它的主题思想是找出具有共词的文档,把包含共词的最多文档放在同一簇中。每个簇根据簇中文档的主要语义内容给出一个总的标题,以便让用户能找到所需要的信息。当然,簇还只是完成了将文档进行归类的任务,为了揭示文档簇(集)之间的逻辑关系,还需要解决如何对簇进行排列。在簇的排列上,有的将簇作为结点排列成层次结构,有的排列成网状结构。

2. TileBARS。TileBARS 是加州大学数字图书馆项目中基于 Web 的分布式检索课题 Cheshire II 的一部分,是当前

比较方便、实用的检索可视化系统。这个系统利用 Texttiling 算法,将每篇文档按页、段落划分为分主题块。检索结果针对相关度最大的簇(从 Scatter/Gathe 结果中得出)。TileBARS 被显示,横列反映了特定界面的标准:文档首先按在所有的检索词组中有多少页面数被子命中来排序,然后是按检索词被子命中的总数来排序,最后是通过相似的搜索来排序。

文献同引(Co-citation)和共词(Co-word)是目前在研究知识领域可视化中与信息技术结合的比较成熟的两种方法。文献同引产生于 20 世纪 70 年代,随着《科学引文索引》的成熟,它在科学映射的历史上也具有独特的地位,主要有文献同引(DCA)和著者同引(ACA)等,相应的软件有 Pajek、Citespace、Histcite 等。共同分析方法产生于 20 世纪 80 年代,这种方法能够在不借助知识专家的帮助提供各种客观的测量方法,而是通过聚类分析来观察高频词之间的亲疏关系,以分析这些词所在的学科和主题结构的变化。

[参 考 文 献]

- [1] 李 纲、郑 重:《情报科学核心领域研究进展》,载《海外人文科学发展年度报告》,武汉:武汉大学出版社 2007 年版。
- [2] 周 宁:《信息可视化在信息管理中的新进展》,载《现代图书情报技术》2003 年第 4 期。
- [3] 李学静、谢 蓉:《信息可视化与 Web 信息检索》,载《图书馆理论与实践》2004 年第 3 期。
- [4] 胡永刚:《数据挖掘中可视化技术综述》,载《计算机与现代化》2004 年第 10 期。
- [5] 靖培栋:《信息可视化-情报学研究的新领域》载《情报科学》2003 年第 7 期。
- [6] 黄如花:《网络信息的检索与利用》,武汉:武汉大学出版社 2005 年版。

(责任编辑 杜 剑)

Application of Information Visualization Technology in Digitized Library

Xu Gang¹, Xu Cong²

(1. Physical Education Department, Wuhan University, Wuhan 430072, Hubei, China;

2. School of Economics & Management, Wuhan University, Wuhan 430072, Hubei, China)

Abstract: This article has summarized the information visualization basic concept and the information visualization engineering research developing process. It elaborated the information visualization technology unit process and the spatial multi-dimensional characteristic main research area and the scope. Between the information visualization technology and the knowledge discovery, the knowledge management relations, the information visualization variable, the symbolism and so on the research, is the information visualization main object of study.

Key words: information visualization; digitization library; information resource