

# 国外馆藏资源可视化研究综述

邱均平 (武汉大学中国科学评价研究中心 湖北 430072)

余厚强 吕红 李小涛 (武汉大学信息管理学院 湖北 430072)

**摘要** 文章利用 Web of Science 平台检索,对国外馆藏资源可视化研究做了全面而深入的调研,发现馆藏数字资源可视化研究最为活跃。从关联表示可视化、查询过程可视化、检索结果可视化、可视化评价四个维度对相关研究做梳理和归纳,据此提出馆藏数字资源可视化的四大趋势,即馆藏资源语义关联可视化、馆藏资源深度聚合可视化、开发专业可视化语言和支持新型数据可视化,最后提出要从技术角度和图情学科方法同时出发完善馆藏资源可视化模式。

**关键词** 馆藏资源 可视化 信息可视化 图书馆

Review of the Visualization Researches on Library- Collected Resources Abroad

Qiu Junping (Research Center for Chinese Science Evaluation, Wuhan University, Hubei, 430072)

Yu Houqiang Lv Hong Li Xiaotao (School of Information Management, Wuhan University, Hubei, 430072)

**Abstract** The Web of Science platform was used for retrieval. Full and deep review was done on the foreign library collection visualization studies. It was found that library digital resource visualization researches were the most active. Relative researches were sorted and summarized from four dimensions, i. e. relevance presentation visualization, query process visualization, search results visualization and visualization evaluation. Four trends were given based on the previous studies, i. e. library collection semantic relevance visualization, in- depth aggregation visualization, developing professional visualization language and support new type data visualization. In the end, it is suggested that library collection visualization shall be improved from both technical perspective and information and library discipline methods.

**Keywords** library collection, visualization, information visualization, library

馆藏资源可视化是近年图书馆学研究的重点和热门主题,这主要源于两个背景:一是馆藏资源尤其是馆藏数字资源的海量增加,为图书馆用户有效利用馆藏资源带来挑战;二是网络时代信息可视化技术的长足发展,为解决上述挑战提供了良好解决方案。信息可视化(Information Visualization, InfoVis 或 IV)是利用计算机支撑的、交互的、对抽象数据的可视化表示,来增强人们对这些抽象信息的认知<sup>[1]</sup>。将信息可视化应用到图书馆,能提供强有力的人工视野和空间感知,来帮助用户观念上组织、电子形式获取和管理大型复杂的信息空间,将用户缓慢阅读的精神压力转移到如可视化模式识别等快速的感知过程<sup>[1]</sup>,为馆藏资

源的检索、浏览和导航带来积极影响<sup>[3]</sup>。

具体来说,信息可视化技术主要从以下三个方面在馆藏资源有效利用上发挥关键作用:(1)显著降低用户信息搜寻成本。在图书馆信息搜寻初期,用户处于知识异常状态(Anomalous State of Knowledge, ASK),不能明确搜寻内容,造成关键词搜索时,术语选择存在模糊性,最终检索结果匹配度低。信息可视化的最终目的是一是浏览和交互,二是探索数据背后的规律。通过信息可视化支持检索结果结构的识别,促进用户理解检索结果之间的相互关系,进而达到精炼搜索的目的。对虚拟环境的使用进一步发展了可视化的概念,提供了完全的、实时的交互和从无限多视角来观

察对象间关系的能力,从而减少信息搜寻初期的不确定性,大大降低信息搜寻成本。(2) 提供馆藏资源概览界面的理想选择,并促进浏览。现代图书馆用户获取数字图书馆资源的首要方式就是搜索引擎,然而,搜索界面缺乏支持信息探索的能力<sup>[4]</sup>。并且图书馆用户通常不确定馆藏资源的覆盖范围,对理解和使用馆藏资源造成障碍。于是,现代图书馆面临的挑战之一就是设计有意义的概览,要求概览中的模式能被很容易地识别,创造出可理解的界面来明确用户所需<sup>[5]</sup>。一般认为通过语义关系来可视化信息,能促进用户信息空间认知图的开发和应用。这让搜索和浏览信息的任务和现实世界的导航相似,例如地图概览(对超文本导航的辅助)被证明对用户有益。信息可视化可用于概要地展示信息,通过允许用户交互地选择内容,成为高效的浏览工具,使得科学家和实践者能够获得数字图书馆的宏观视图,锁定相关资源,监测他们本身和其他知识领域的演变,以及追踪领域内和跨领域理论的影响,所以是数字图书馆信息可视化界面的理想选择。(3) 大幅提高图书馆知识服务水平。现代图书馆以数字图书馆为主流形态,正成为人类知识的主要存储库, Fast 等人<sup>[6]</sup>强调数字图书馆作为交互知识环境的概念重构,呼唤利用信息可视化技术带来革命性的转型。信息可视化通过对馆藏文献关联、查询过程及检索结果的图形化展示,不仅使得用户操作过程变得智能化和友好化,而且提供超出传统文本结果的整体把握,增强资源深度利用的洞见性。同时,可视化与可获取文献的用户交互数据能评价和提高数字图书馆的利用。

越来越多的图书馆员和研究人员认识到,将信息可视化技术和现有图书馆技术结合,是现代图书馆发展的必然趋势。掌握国外研究动态能为本国相关研究提供参照和依据。因此,本文对国外馆藏资源可视化研究进行了详细调研,以期反映国际上馆藏资源可视化研究的发展脉络与最新进展。

## 1 数据获取与分析方法

### 1.1 数据检索式

馆藏资源可视化的研究对象是图书馆,其文献主要分布在图书馆学和情报学的期刊上。Web of Science (简称 WoS) 是全球最大的索引库,几乎覆盖了世界范围内所有具备影响力的国际期刊。WoS 将期刊分为 56 个大类,其中 Information Science Library Science 大类(简称 LIS 大类)中的期刊多达 83 种。因此本文选取 WoS 中的 LIS 大类中的期刊作为文献调研范围,能够保证涵盖所有重要的和有效的相关文献。在 WoS 的高级搜索中,使用检索式 TS= "LIBRARY" AND TS= "VISUAL\*" AND WC= "INFORMATION SCIENCE LIBRARY SCIENCE",检索共得到 329 篇文献,逐篇阅

读,去除研究主题如“向视力障碍用户提供图书馆服务”等与本研究相关度较低的文献,剩下 118 篇核心相关文献。在对这些文献进行考察的基础上,将每篇论文的参考文献和引用文献纳入调研范围,从而进一步提高了文献调研的全面性和可靠性。

### 1.2 研究对象界定

馆藏资源主要分为实体资源和数字资源,实体资源包括图书馆建筑、内部设备、纸质书刊、珍藏典籍、声像光盘等,数字资源包括电子化的书刊、馆藏书刊的元数据。具体说来,涵盖各种全文数据库、索引数据库、文献数据库、电子书数据库、联机目录等。馆藏资源的可视化因而分为两种:第一种是对馆藏实体进行数字化模拟的可视化,其目的是生动形象地揭示馆藏资源的分布,方便读者迅速熟悉图书馆环境和搜寻到目标信息源;第二种是对馆藏数字资源在深度聚合和挖掘的基础上进行可视化,其最终目的是实现知识服务。馆藏数字资源和数字图书馆是紧密关联的,数字图书馆正是专门针对馆藏数字资源进行信息组织的集中体现,因而对数字图书馆的可视化研究,实际上是对馆藏数字资源的可视化研究,相关研究主题所得到的结论都完全适用于馆藏数字资源,甚至对数字图书馆的可视化是国外馆藏资源可视化研究的主要体现方式。正是出于这个原因,本文将对数字图书馆的可视化研究视同对馆藏数字资源可视化研究,对相关研究方法和分析结论进行综合处理。

### 1.3 综述分类框架

Shneiderman<sup>[7]</sup>从技术角度出发,将信息可视化分为二维可视化、三维可视化、多维可视化、网络可视化、时序可视化和树状可视化。Zaphiris<sup>[8]</sup>等人根据数字图书馆的三大任务,将馆藏资源可视化研究分为相应的三大类,即信息检索可视化、信息导航可视化和信息浏览可视化。Shiri<sup>[9]</sup>则根据信息可视化在馆藏资源中的应用前景,将馆藏资源可视化分为四大类,即馆藏资源表示可视化、查询构建和修改可视化、检索结果可视化和关联性可视化。笔者认为 Shiri 的分类更符合用户对馆藏资源使用过程的理解,但是关联性可视化与资源表示可视化的最终目标一致、研究方法重叠,因而可以合为一类。在此基础上,对馆藏资源可视化的评价研究是改善馆藏资源可视化的重要环节,也列成一类。至此,可将国外馆藏资源可视化研究文献分为四类,如表 1。

表 1 馆藏数字资源可视化综述框架

本文接下来的分析思路是,首先论述馆藏实体资

源的可视化研究进展,然后重点综述馆藏数字资源的关联表示可视化、查询过程可视化、检索结果可视化和可视化评价研究,在此基础上发掘馆藏数字资源研究趋势,最后对已有研究成果进行总结,为后续研究思路 and 方向提出笔者的思考和展望。

## 2 馆藏实体资源可视化

馆藏实体资源可视化主要从两个方面着手:一是馆藏对象的可视化展示,例如 Rauber 等人<sup>[1]</sup>的 SOM-Lib 项目,界面展示是书架上的书,文档的大小用书的厚度表示,上次借阅的时间用书与架子边缘的距离表示,文档不被利用用灰尘或蛛网表示,使用频率高通过让书显得有很多指印来表示,崭新程度通过让书看上去闪亮来表示,不同文献类型在书架上用不同媒体来展示;二是图书馆空间的虚拟展示,通过三维空间展示来说明馆藏 in 图书馆中的分布,为读者提供向导服务。较为流行的是基于 GIS 的可视化<sup>[11-13]</sup>,Buchel 等人<sup>[13]</sup>在此基础上进行拓展,使其包含非地理空间属性,以支持用户进行初级和概要可视化任务,并实际用 Google Maps 开发了基于地图的原型系统,展示了额外表现形式的作用方式。

馆藏实体资源可视化往往不涉及文献内容属性,主要在物理空间的导航和搜索中发挥作用。对馆藏实体资源可视化固然能提升读者使用体验,但不是馆藏资源可视化的研究重点,因为图书馆用户所面临的主要挑战是数字文献空间的海量性和复杂性带来的认知负担。

## 3 馆藏数字资源可视化

馆藏数字资源逐渐成为馆藏资源的主体部分,以数字图书馆的服务形式呈现,是国际上可视化的主要研究对象。按照 1.3 中介绍的综述框架,馆藏数字资源可视化研究从以下几个方面开展。

### 3.1 馆藏数字资源关联表示可视化

馆藏数字资源关联表示可视化,实质上是对原有馆藏数字资源信息组织的可视化,在提供概览的基础上,将数字资源的关联呈现出来,达到提高浏览效率的目的。Brad 等人<sup>[14]</sup>详述了数字图书馆界面中二维和三维信息可视化技术支持多源和信息馆藏的展示与导航,用户可以任意方式操作、搜索和查看。

等级结构通过将相关信息组织成可理解的结构来减少复杂性,由于和图书馆分类法、主题法等知识组织方法相契合,成为最经典的馆藏资源关联表示可视化对象,其具体可视化方式很多,用于展示等级结构的技术有双曲树、扩展树、瓦条图(TileBar)和基于交叉的聚类图。Hearst 等人<sup>[15]</sup>开发的 用户界面原型,将大型分类等级结构连同它们相关文本集合的搜索和浏览集成起来,分类标签和所检索文档同时相互独立的

在三维空间中展示。Hascoet 等人<sup>[16]</sup>则以导航树和邻接树的形式来展示文献和一个焦点文献的关系,将检索出的文献按缺省设置自动生成图,图是由一系列图层构成,每个图层分成不同区域,区域里是代表文献的图标。如果分类系统是单个等级的,则只用一个图层,每个区域对应分类系统中的一个分类;如果分类系统是两个等级,每个等级由一个图层表示,每个子分类由区域表示。用户可以修改图的设置和对图进行编辑(经典的操作如剪辑、粘贴、移动、删除、保存、加载视图和扩展视图)。

在以等级结构进行可视化的基础上,学者做了许多新的改进,融入了其他元素,目标是更加智能化和精细化:

(1) 结合同义词典。Costabile 等人<sup>[17]</sup>开发了原型智力数字图书馆界面,使用 INSPEC 同义词典和科学文献集合,通过交互动态地图来展示同义词典及文献馆藏的语义内容。Pfeffer 等人<sup>[18]</sup>的 Seminal 原型系统则考虑了基于同义词典索引的相关领域,将同义词典的树图视图和文献集合上的实时统计分析结合起来,单个树图元素对应同义词典树形结构中的结点,元素大小取决于子结点数,颜色取决于统计分析结果。该方法提供了特定馆藏方面结构和现行使用的直观理解,让专家可以交互地阅读自动索引结果,在图书馆馆藏的发展和管理中都发挥了重要作用。

(2) 改造传统目录。Julien 等人<sup>[19]</sup>认为传统基于文本的 OPAC 目录是馆藏资源组织的重要形式,但是缺少关键词搜索和结果过滤工具,忽视了受控主题词表丰富的网络关系,新一代的 OPAC 应当在现有信息组织基础设施和提供在线搜索者全新的浏览和搜索环境之间达到平衡,于是设计并开发了虚拟现实主题浏览和信息检索工具,将 LCSH(Library of Congress Subject Headings)转变成有用和可用的树结构,提供包含现实世界中的科技领域馆藏的可视化隐喻,正式的测试表明用户能有效的浏览 LCSH 树,并根据关键词搜索查询将其分块。Osinska<sup>[20]</sup>认为有必要将可视化研究中的定性属性和数字图书馆资源的定量特征相结合,实际对 ACM 数字图书馆采用的分类和通用分类法的可视化进行描述和展示,用非线性图形过滤器处理给定的地图,用分形分析(fractal analysis)产出图形的量化,让调整可视化参数和评价最初分类策略和动态特征成为可能。

(3) 图像关联展示。Cunningham 等人<sup>[21]</sup>将从馆藏文献中抽取的图像构成流动拼图,作为获得馆藏资源综合概览的一个有效工具,定性研究表明用户在浏览这种流动拼图时对馆藏实际理解的内容更佳。Borner 等人<sup>[22]</sup>主持的 LVis 项目采用 Latent 语义分析(LSA)来自动抽取图像间的语义关系,再用聚类算法将图像分组成为若干有语义相似描述的类,接着用修正的 Boltz-



man 算法在空间中展示图像。

(4) 融入概率模型。De Stefani 等人<sup>[23]</sup>认为数字图书馆管理系统 (Digital Library Management System, DLMS) 的信息获取组件, 例如自动分类和数字对象的检索组件, 应当允许用户和系统进行交互, 从对象集合中浏览、探索和检索资源。为此设计了二维 Naive 贝叶斯概率模型的可视化工具, 在给定兴趣分类后, 定义对象的概率和二维空间的点之间的关系, 把整个馆藏对象都图像化到笛卡尔平面上, 并设计算法在二维展示上分类和检索文献, 这两者在探索原始数据和解释结果时作用显著<sup>[23]</sup>。

(5) 其他改进方向。Waltinger 等人<sup>[24]</sup>受网络社会标签启发, 提出在逻辑文献结构分析基础上, 进行语义内容分析, 形成可用于主题追踪和数字文献标签语义链的算法模式的概览, 用于主题标签和文献分类, 提高了文献的可视化表示的效率。Farney<sup>[25]</sup>将点击分析用于图书馆, 可视化展示了用户在图书馆网页上点击的热点区域和内容, 从而帮助图书馆更加容易地确定页面中使用高频和低频的区域。

### 3.2 馆藏数字资源查询过程可视化

查询过程可视化既包括馆藏资源浏览过程的可视化, 也包括检索过程的可视化, 是用户有效获取目标馆藏资源的关键环节。浏览任务可以是对特定事实的搜寻行为, 也可以是对已知数据库无结构的开放浏览和相关主题信息可用性的探索<sup>[26]</sup>。

早在 20 世纪末, 信息可视化技术就被用于改善用户浏览和搜索馆藏信息, 例如动态查询<sup>[27]</sup>、可视化信息搜寻和查询预览, 这些界面的共同特征是对数据库的可视化展示 (典型的是散点或者星域展示) 和利用一系列工具的查询可视化展示 (如 sliders), 以及这两个方面的紧密耦合。Jones<sup>[28]</sup>就报道了新西兰数字图书馆的可视化界面, 核心是利用 Venn 图来促进和改善查询构件。

随后, 信息检索可视化界面纳入了知识组织系统, 如同义词典和其他受控词表作为主题元数据源, 例如 MeSHBrowse、Cat-a-cone、Visual MeSH 和 Integrated Thesaurus-Results Browser<sup>[29]</sup>。Eichler 等人<sup>[30]</sup>展示的数字图书馆助手 (Digital Library Assistant, DiLiA) 为用户控制的交互搜索开发了先进的信息可视化方法, 旨在从多个维度加强在数字图书馆中的搜索, 该交互模型设计对用户透明且易用。Fuhr 等人<sup>[31]</sup>开发的数字图书馆界面, 采用可视化技术来表示搜索策略和元信息服务, 如同义词典、分类和参考文献。Janecek 等人<sup>[32]</sup>开发的数字馆藏可视化界面用的是鱼眼视图, WordNet 的语义结构被纳入该界面来支持用户的探索性搜索体验。Julien 等人<sup>[33]</sup>针对用户搜索词和信息检索系统能识别的词匹配度不足问题, 设计了一种支持受控主题标题浏览的界面, 对馆藏主题标题和它们之

间关系进行交互可视化图谱的设计和开发, 设计出来的可视化原型对关键词搜索来说是个有力补充, 通过提供给用户馆藏概览, 提示小部分专门馆藏合法主题查询术语, 使得主题浏览更加便捷。

与静态的、预先定义好的等级导航结构不同, Papadakis 等人<sup>[34]</sup>提出的导航依靠动态的、交互的图形结构、主题标题及其之间的关系分别构成了图中的结点和边, 网络图形化用户界面和构成主题标题及其相互关系的底层本体沟通, 用户则通过和该界面进行交互来选择主题标题, 将主题标题应用到 OPAC 中来语义化地描述文献集属性。这种交互式导航通过提供更广阔或更细致的可供选择概念, 或者提示他们可能事先没有觉察到的替代性或相关概念, 辅助搜索者准确地形成自己的查询。这种可视化主题标题的全新方法可以扩展可视化图书馆领域其他一系列概念化内容, 让隐藏在图书馆领域却出于各种原因未能呈现给图书馆用户的隐藏知识得到利用。Wong 等人<sup>[35]</sup>认为可视化分析技术是通过交互可视化和可视化分析系统实现的推理科学, 能同时支持感知和信息检索, 揭示现有在线搜索工具隐藏的底层信息组织框架, 使用户利用覆盖主题进行馆藏浏览变得容易。实际设计的 INVISQUE 是实现交互可视化搜索与查询的方法和实验软件, 利用索引卡隐喻来展示图书馆内容, 可视化集成如引文和出版日期等属性来组织, 保证选出最新和被引最高的论文更加容易。它使用焦点结合情境的设计技巧来揭示文献间的关系, 避免了搜寻目标模糊的问题。Lee 等人<sup>[36]</sup>设计了 NetLens 来探索内容—行动者网络数据模型, NetLens 界面被设计成允许用户提出系列基本查询, 且循环精炼可视化概览和分类列表, 这就支持了在传统结点——链接可视化中很难确定的复杂查询。因为它适用于任意可以用我们抽象内容—行动者数据模型表示的数据集, 所以 Netlens 是通用的和可量度的。

查询过程可视化的两个热点是利用用户日志分析和聚类搜索引擎。Kyriakidi 等人<sup>[37]</sup>设计了一种基于欧洲图书馆门户需求开发的个性化图书馆系统, 利用一种叫 madIS 的环境, 实现了对门户日志文件广泛集合的数据处理、分析和挖掘技术, 该系统也展示了日志分析结果基于网络的可视化特征。Burt 等人<sup>[38]</sup>对聚类搜索引擎和基于列表搜索引擎进行对比, 来获得用户感知负担的相关洞察力, 表明聚类特征可以成为信息搜寻用户工具箱中的额外工具, 聚类在与传统搜索引擎串联使用时最为有效。结论也表明在传统列表式搜索引擎中结合聚类元素的设计具备优良品质, 例如 Bing 现在包含相关研究条, 允许搜索者一步步精炼自己的查询。

### 3.3 馆藏数字资源检索结果可视化

检索结果可视化是馆藏数字资源可视化的最后

环节,直接影响用户对资源的使用效率,经历了静态展示向动态展示的演变过程。

静态可视化展示存在多种构图方式,由 Liu 等人<sup>[39]</sup>实现的独立于语言的文献分类系统,为数字图书馆环境提供搜索帮助,用向量模型计算文献间的相似性,并图形化的帮助用户分析搜索查询结果。Shneiderman 等人<sup>[4]</sup>开发的 GRIDL 二维可视化界面,利用范畴和等级轴创立了简洁的二维展示,在展示的每个网格,呈现颜色编码的点或者条形图,用户能查看整个结果集,然后点击标签来在等级中下移一个层次,据此将 ACM 计算分类系统的技术图书馆进行布局,来提供等级式浏览结构。Brettlecker 等人<sup>[40]</sup>开发的下一代数字图书馆管理系统原型 DelosDLMS,采用了 Konstanz 大学开发的自组织图(Self-Organizing Map, SOM)应用,将用户查询结果集进行可视化,通过从高维特征空间向低维展示空间映射,使得多个对象集合,以及它们的结构和相互关系,一次性地在可能的不同特征空间中可视化出来。这些静态展示结构清晰,内容丰富,但在交互性上存在明显不足。

动态可视化展示加入了布局、过滤等选项,可以从多个角度查看结果集。Collins 等人<sup>[41]</sup>描述信息可视化如何在大型存储库的界面设计工具中得到应用,其为 Los Alamos 国家实验室设计了 SearchGraph,通过二维和三维的交互散点图形式来展示搜索结果,用户可以看到一个密集化的、抽象的搜索结果的可视化,通过操作若干布局、分类和过滤选项,可以从多个角度来浏览相同数据集,观察数据集中的不同模式,对大型数字对象存储库数据集深入分析。Yoshida 等人<sup>[42]</sup>设计了可视化有效关键词的系统,该系统从搜索结果中抽取重要术语,将他们在二维图上展示出来,用户可以查看搜索结果的趋势,系统允许用户通过移动图上的术语并动态地重新排列和重新搜索结果。该系统还通过展示与查询相关的补充信息来试图扩展搜索区域,方法被证明是可行的。

除了简单地可视化搜索结果以外,有些研究者进一步提供工具帮用户探索文本文献之间复杂的关系。例如由西北太平洋国家实验室(Pacific Northwest National Laboratory)开发的 SPIRE(Spatial Paradigm for Information Retrieval and Exploration),亚利桑那大学的自组织特征图<sup>[43]</sup>等。

### 3.4 馆藏数字资源可视化评价研究

馆藏资源可视化理论和实践的发展都离不开可用性研究的实证证据支撑,而可用性研究的基础正是馆藏资源可视化评价。馆藏数字资源可视化评价不仅直接决定可视化系统的实际可用性和推广价值,而且决定未来可视化系统的改进方向和完善路径。其基础评价通常分为两部分:(1)用户更倾向于使用哪个系统;(2)在给定相似任务下,哪个系统表现更优。从细

粒度的角度观察,馆藏资源可视化的评价研究主要在以下两个方面:

(1)可视化界面选择评价。早期用户对馆藏资源可视化界面的评价并不高,例如,对图形化概览研究发现用户认为聚类令人困惑,图形反而没有文字清楚<sup>[44]</sup>;对 SOM 的研究发现,用户在浏览多层 SOM 时容易迷失,而转向使用传统基于文本的字顺结构<sup>[45]</sup>;另一项研究也发现大多数用户偏好使用单词列表而不是可视化<sup>[46]</sup>。随着时间的推移,对馆藏资源可视化界面的积极评价也逐渐增多。Morse 等人<sup>[47]</sup>的比较实验发现在给定的五种检索方式中,用户更倾向于使用可视化界面。对众多信息可视化界面研究的元分析也表明,用户使用可视化空间界面比使用传统方式表现要好,当可视化界面简洁时,用户使用效果更佳<sup>[48]</sup>。近年,随着研究者展示了用可视化搜索界面辅助用户进行信息搜寻后,若干主流图书馆供应商在他们的产品上添加了可视化搜索界面功能。Fagan 等人<sup>[49]</sup>对可视化界面与传统基于文本的搜索界面做了比较研究,对 EBSCOhost Basic Search 和 Visual Search 进行了八项全尺度使用测试,发现受测试者在与搜索结果交互后识别的子主题质量,与所选总体主题相比子主题选择的质量和子主题搜索结果的质量都有显著提升。

(2)可视化系统性能评价。聚类视图(分类视图)是馆藏资源可视化呈现的主要视图之一<sup>[50]</sup>,聚类的精准度成为聚类视图的重要评价标准<sup>[51]</sup>,所谓聚类成功是指通过机器算法形成的聚类与人工分类结果保持一致。当然,人工分类本身就存在缺陷,分类人员选择分类时由于文化和个体差异,会导致分类结果不尽相同。数字图书馆用户搜索行为研究表明,用户倾向浏览关键词搜索结果,而不是事先制定好的聚类<sup>[52]</sup>,有些人甚至认为聚类限制了搜索信息时的开放创新过程,图书馆的大量信息也应该像网络信息那样,通过关键词自由搜索来实现,而不是人为进行逻辑分类。然而实践表明,聚类系统的缺陷是可控的,通过维护自动聚类揭示不明显的关联,让用户理解分类的数量和含义能显著提高用户创造力<sup>[53]</sup>。Leouski 等人<sup>[54]</sup>使用一种非交互式方法来评价可视化文献组织系统,设计了一种方法用于评价反馈特定查询的文献集时信息组织的质量,并基于从用户获得的信息介绍了两种完善该组织的方法,展示了这样的反馈如何提高可视化组织系统质量。

### 3.5 馆藏数字资源可视化研究趋势

从上述研究可以看出,馆藏数字资源可视化研究是全方位、多角度且不断深化的,关联表示可视化、查询过程可视化和检索结果可视化,共同构成了完整的馆藏资源可视化利用环境,经历了由静态展示到动态展示、由小范围简单的映射可视化到大规模智能的语义可视化的转变过程。据此,未来馆藏资源可视化研



究有四大趋势:

(1) 馆藏资源语义关联可视化。随着数字信息可获得性的增长,语义网技术被用于构建语义数字图书馆来减轻信息理解的压力,语义网的使用让用户在语义环境下搜索和可视化资源。生产语义网数据作为语义数字图书馆建设的关键过程,是指将数字资源的元数据转换成语义网数据,许多文本挖掘技术,如关键词抽取与聚类,已经被用于生产语义网数据。为了促进语义数字图书馆查询,Huang等人<sup>[5]</sup>引入了一种内容和服务推理模型来演绎出内容和服务之间的15种关系,并定义了函数来操作这些关系。Kruk等人<sup>[59]</sup>开发了JeromDL,让用户可以在语义化标注的目录下为有趣的书、论文或其他材料添加书签。Li等人<sup>[5]</sup>开发了名为oreChen ChemXSeer的语义数字图书馆,用语义元数据对化学论文进行建模。Jiang等人<sup>[59]</sup>提出基于标准化压缩距离(normalized compression distance)的聚类算法,解决语义数据生产中论文作者单位的模糊性问题。

(2) 馆藏资源深度聚合可视化。文本挖掘、聚类分析等技术进一步被用于揭示文献之间的关联,目前的馆藏资源可视化已经不局限于馆藏资源结构的直观反映,而是试图在展示文献信息组织结构的同时,反映文献之间的深层次关联,这个过程会继续深化。事实上,馆藏资源深度聚合可视化也是提高可视化效果的必然途径。

(3) 开发专业可视化语言。数字图书馆可视化系统的设计和开发涉及不同的利益相关者,例如信息构建师、图书馆员、领域专家,这些利益相关者们需要有共同的语言来描述、讨论、商量图书馆可以提供的服务。通过开发专门的可视化语言供数字图书馆系统设计者(包括信息架构师、图书馆员、领域专家)使用,可以为定义的数字图书馆自动生成代码,让可视化系统设计者不必纠缠于布局的技术问题。Malizia等人<sup>[59]</sup>开发了CRADLE( Cooperative- Relational Approach to Digital Library Environment)可视化语言,其定义的工具集合允许若干服务自动生成,并提供终端用户使用图形化用户界面的生成,认知维度框架评价证实了其有效性。

(4) 支持新型数据可视化。数字图书馆对文本型和若干其他类型非文本文献的支持,在过去几年显著进步。数字图书馆支持意味着许多方面,包括整个图书馆工作流程建模、交互的和可视化的检索,提供有效的查询构建和结果展示。近来开始出现应当得到数字图书馆支持、却没有现成技术实现的新型非文本文献,例如由科学实验、地球观测或者模拟仿真产生的科学一手数据。Bernard等人<sup>[60]</sup>报道了一种对数字图书馆功能的概念和首次实现,支持可视化检索和探索一种重要类型的科学一手数据,即面向时间的数据。该方法具备跨学科属性,由图书馆学家、自然科学家和可视化分析学家们共同完成,除了展示概念和讨论相

关挑战以外,还展示了用该方法首次实现现实世界科学一手数据集后的表现。

#### 4 结语

综上所述,笔者利用Web of Science平台对国外馆藏资源可视化文献做了全面而深入的调研,发现馆藏数字资源可视化研究最为活跃。从关联表示可视化、查询过程可视化、检索结果可视化、可视化评价四个维度对相关研究做整理和归纳,得到的主要结论有:(1) 馆藏数字资源关联表示可视化以等级结构表示为基础,通过结合同义词典、改造传统目录、融入概率模型、图像关联展示等手段进行改进和发展。(2) 馆藏数字资源查询过程可视化早期以查询预览、动态查询为起点,后来纳入了知识组织系统如同义词典作为主题元数据源,使得主题浏览更加便捷,逐渐形成基于底层语义交互的动态图形导航结构,用户日志分析和聚类搜索引擎是近年来的研究热点。(3) 馆藏数字资源检索结果可视化经历了静态展示到动态展示的演变过程,以从多个角度查看结果集。(4) 相关可视化评价研究表明,用户对可视化界面存在偏好,但要经历一个心理接受过程,可视化界面能否达到最终目标,还需要用户反馈基础上的不断改进。(5) 据此总结出馆藏资源可视化的四大趋势,即馆藏资源语义关联可视化、馆藏资源深度聚合可视化、开发专业可视化语言和支持新型数据可视化,它们是未来研究的重点。这些结论表明,国外馆藏资源可视化研究已经度过了探索阶段,取得了丰富的成果。

最后需要指出的是,馆藏资源可视化不仅是图书馆学情报学的研究重点,同时也是计算机科学、人机交互、界面设计等学科的交叉研究主题,然而现有研究大多基于技术视角,忽视了从图书情报学的本源出发,Heuvel等人<sup>[61]</sup>提出要关注图书情报学视野下的界面设计。这说明馆藏资源可视化既要从技术上突破,更要从馆藏资源本身的研究入手,二者不能偏废。后续研究笔者将沿着这个思路,一方面从技术角度完善馆藏资源可视化模式,另一方面从图书情报学的学科方法(如计量分析法)出发去改进可视化的对象基础。

#### 参考文献

- [1] 周宁,张玉峰,张李义.信息可视化与知识检索[M].北京:科学出版社,2005:3.
- [2] Börner K, Chen C. Visual interfaces to digital libraries[C]. Proceedings of the 2nd ACM/IEEE- CS Joint Conference on Digital Libraries. ACM, 2002: 425- 425.
- [3] Shen R, Vemuri N S, Fan W, et al. Exploring digital libraries: Integrating browsing, searching, and visualization[C]. Proceedings of the 6th ACM/IEEE- CS Joint Conference on Digital Libraries. ACM, 2006: 1- 10.
- [4] Börner K, Chen C. Visual interfaces to digital libraries:

- Motivation, utilization, and socio-technical challenges [M]. Visual Interfaces to Digital Libraries. Springer Berlin Heidelberg, 2002: 1-9.
- [5] Shneiderman B, Feldman D, Rose A, et al. Visualizing digital library search results with categorical and hierarchical axes [Q]. Proceedings of the Fifth ACM Conference on Digital Libraries. ACM, 2000: 57-66.
- [6] Stanney K M, Salvendy G. Information visualization: Assisting low spatial individuals with information access tasks through the use of visual mediator [J]. Ergonomics, 1995, 38(6): 1184-1198.
- [7] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualization [Q]. IEEE, 1996: 336-343.
- [8] Zaphiris P. Exploring the use of information visualization for digital libraries [J]. New Review of Information Networking, 2004, 10(1): 51-69.
- [9] Shiri A. Metadata-enhanced visual interfaces to digital libraries [J]. Journal of Information Science, 2008, 34(6): 763-775.
- [10] Rauber A, Merkl D. SOMLib: A digital library system based on neural networks [Q]. Proceedings of the Fourth ACM Conference on Digital Libraries. ACM, 1999: 240-241.
- [11] Buckland M, Chen A, Gey F C, et al. Geographic search: Catalogs, gazetteers, and maps [J]. College & Research Libraries, 2007, 68(5): 376-387.
- [12] Coyle A. Interior library GIS [J]. Library Hi Tech, 2011, 29(3): 529-549.
- [13] Buchel O, Sedig K. Extending map-based visualizations to support visual tasks: The role of ontological properties [J]. Knowledge Organization, 2011, 38(3): 204-229.
- [14] Brad E. Practical applications of 2D and 3D information visualization for information organization [J]. Library Technology Reports, 2005, 41(1): 55-64.
- [15] Hearst M A, Karadi C. Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy [Q]. ACM SIGIR Forum. ACM, 1997, 31(S): 246-255.
- [16] Hascoet M, Soinard X. Using maps as a user interface to a digital library [Q]. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1998: 339-340.
- [17] Costabile M F, Esposito F, Semeraro G, et al. Interacting with IDL: The adaptive visual interface [M]. Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 1998: 515-534.
- [18] Pfeffer M, Eckert K, Stuckenschmidt H. Visual analysis of classification systems and library collections [M]. Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2008: 436-439.
- [19] Julien C A, Guastavino C, Bouthillier F. Capitalizing on information organization and information visualization for a new-generation catalogue [J]. Library Trends, 2012, 61(1): 149-161.
- [20] Osinska V. Fractal analysis of knowledge organization in digital libraries [J]. New Trends in Qualitative and Quantitative Methods in Libraries, 2011: 17.
- [21] Cunningham S J, Bennett E. Understanding collection understanding with collage [M]. Digital Libraries: Universal and Ubiquitous Access to Information. Springer Berlin Heidelberg, 2008: 367-370.
- [22] Borner K, Dillon A, Dolinsky M. LVis-Digital Library Visualization [Q]. Information Visualization, 2000. Proceedings. IEEE International Conference on. IEEE, 2000: 77-81.
- [23] De Stefani L, Di Nunzio G M, Vezzaro G. A visualization tool of probabilistic models for information access components [M]. Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2009: 481-482.
- [24] Waltinger U, Mehler A, Heyer G. Towards automatic content tagging-enhanced Web services in digital libraries using lexical chaining [Q]. WEBIST'08. 2008: 231-236.
- [25] Farney T A. Click analytics: Visualizing website use data [J]. Information Technology and Libraries, 2011, 30(3).
- [26] Shneiderman B. Designing information-abundant websites: Issues and recommendations [J]. International Journal of Human-Computer Studies, 1997, 47(1): 5-29.
- [27] Ahlberg C, Williamson C, Shneiderman B. Dynamic queries for information exploration: An implementation and evaluation [Q]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 1992: 619-626.
- [28] Jones S. Graphical query specification and dynamic result previews for a digital library [Q]. Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. ACM, 1998: 143-151.
- [29] Sutcliffe A G, Ennis M, Hu J. Evaluating the effectiveness of visual user interfaces for information retrieval [J]. International Journal of Human-Computer Studies, 2000, 53(5): 741-763.
- [30] Eichler K, Hensen H, Neumann G, et al. DiLiA: The digital library assistant [M]. Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2010: 534-537.
- [31] Fuhr N, Klas C P, Schaefer A, et al. Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries [M]. Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2002: 597-612.
- [32] Janecek P, Schickel V, Pu P. Concept expansion using semantic fisheye view [M]. Digital Libraries: Implementing Strategies and Sharing Experiences. Springer Berlin Heidelberg, 2005: 273-282.
- [33] Julien C A, Cole C. Capitalizing on controlled subject vocabulary by providing a map of main subject headings: An exploratory design study [J]. Canadian Journal of Information and Library Science, 2009, 33(1-2): 67-83.
- [33] Julien C A, Cole C. Capitalizing on controlled subject vocabulary by providing a map of main subject headings: An exploratory design study [J]. Canadian Journal of Information and Library Science, 2009, 33(1-2): 67-83.
- [34] Papadakis I, Stefanidakis M, Tzali A. Visualizing OPAC subject headings [J]. Library Hi Tech, 2008, 26(1): 19-23.
- [35] Wong B L W, Choudhury S T, Rooney C, et al. INVISQUE: Technology and methodologies for interactive information visualization and analytics in large library collections [M]. Research

- and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2011: 227– 235.
- [ 36 ] Lee B, Kang H, Plaisant C, et al. Exploring content– actor paired network data using iterative query refinement with NetLens [ Q ] // Digital Libraries, 2006. JCDL 06. Proceedings of the 6th ACM/IEEE– CS Joint Conference on. IEEE, 2006: 372– 372.
- [ 37 ] Kyriakidi M, Stamatogiannakis L, Triantafyllidi M L, et al. A prototype personalization system for the European library portal [ M ] . Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2010: 542– 545.
- [ 38 ] Burt M, Liew C L. Searching with clustering: An investigation into the effects on users' search experience and satisfaction [ J ] . Online Information Review, 2012, 36( 3 ) : 278– 298.
- [ 39 ] Liu Y H, Dantzig P, Sachs M, et al. Visualizing document classification: A search aid for the digital library [ J ] . Journal of the American Society for Information Science, 2000, 51( 3 ) : 216– 227.
- [ 40 ] Brettlecker G, Milano D, Ranaldi P, et al. DelosDLMS: A next– generation digital library management system [ Q ] . Image Analysis and Processing Workshops, 2007. ICIAPW 2007. 14th International Conference on. IEEE, 2007: 83– 88.
- [ 41 ] Collins L M, Hussell J A T, Hettinga R K, et al. Information visualization and large– scale repositories [ J ] . Library Hi Tech, 2007, 25( 3 ) : 366– 378.
- [ 42 ] Yoshida T, Nakamura S, Oyama S, et al. Query transformation by visualizing and utilizing information about what users are or are not searching [ M ] . Digital Libraries: Universal and Ubiquitous Access to Information. Springer Berlin Heidelberg, 2008: 124– 133.
- [ 43 ] Chen H, Schuffels C, Orwig R E. Internet categorization and search: A self– organizing approach [ J ] . Journal of Visual Communication and Image Representation, 1996( 1 ) : 88– 102.
- [ 44 ] Hearst M A. User interfaces and visualization [ M ] . Modern Information Retrieval. New York, 1999: 257– 323.
- [ 45 ] Chen H, Houston A L, Sewell R R, et al. Internet browsing and searching: User evaluation of category map and concept space techniques [ J ] . Journal of the American Society for Information Science, 1998( 7 ) : 582– 603.
- [ 46 ] Cole C, Leide J E, Large A, et al. Putting it together online: Information need identification for the domain novice user [ J ] . Journal of the American Society for Information Science and Technology, 2005, 56( 7 ) : 684– 694.
- [ 47 ] Morse E, Lewis M, Olsen K A. Evaluating visualizations: Using a taxonomic guide [ J ] . International Journal of Human– Computer Studies, 2000, 53( 5 ) : 637– 662.
- [ 48 ] Chen C. Information visualization: Beyond the horizon [ M ] . Springer Berlin Heidelberg, 2004.
- [ 49 ] Fagan J C. Usability testing of a large, multidisciplinary library database: Basic search and visual search [ J ] . Information Technology and Libraries, 2013, 25( 3 ) : 140– 150.
- [ 50 ] Geffner S, Agrawal D, Abbadi A E, et al. Browsing large digital library collections using classification hierarchies [ Q ] . Proceedings of the Eighth International Conference on Information and Knowledge Management. ACM, 1999: 195– 201.
- [ 51 ] Krowne A, Halbert M. An initial evaluation of automated organization for digital library browsing [ Q ] . Digital Libraries, 2005. JCDL 05. Proceedings of the 5th ACM/IEEE– CS Joint Conference on. IEEE, 2005: 246– 255.
- [ 52 ] Das Neves F A, Fox E A. A study of user behavior in an immersive virtual environment for digital libraries [ Q ] . Proceedings of the Fifth ACM conference on Digital Libraries. ACM, 2000: 103– 111.
- [ 53 ] Adams K C. Information architecture as narrative: The function and benefits of taxonomies [ J ] . Internet Librarian 2001 Collected Presentation, Pasadena, California, 2001, 8: 6– 11.
- [ 54 ] Leouski A, Allan J. Evaluating a visual navigation system for a digital library [ M ] . Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 1998: 535– 554.
- [ 55 ] Huang S H, Ke H R, Yang W P. Enhancing semantic digital library query using a Content and Service Inference Model ( CSIM ) [ J ] . Information Processing & Management, 2005, 41( 4 ) : 891– 908.
- [ 56 ] Kruk S R, Decker S, Zieborak L. JeromeDL– Adding semantic web technologies to digital libraries [ M ] . Database and Expert Systems Applications. Springer Berlin Heidelberg, 2005: 716– 725.
- [ 57 ] Li N, Zhu L, Mitra P, et al. oreChem ChemXSeer: A semantic digital library for chemistry [ Q ] . Proceedings of the 10th Annual Joint Conference On Digital Libraries. ACM, 2010: 245– 254.
- [ 58 ] Burkhard R A. Towards a framework and a model for knowledge visualization: Synergies between information and knowledge visualization [ M ] . Knowledge and Information Visualization. Springer Berlin Heidelberg, 2005: 238– 255.
- [ 59 ] Malizia A, Bottoni P, Levialdi S, et al. A cooperative– relational approach to digital libraries [ M ] . Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 2007: 75– 86.
- [ 60 ] Bernard J, Brase J, Fellner D, et al. A visual digital library approach for time– oriented scientific primary data [ J ] . International Journal on Digital Libraries, 2010, 11( 2 ) : 111– 123.
- [ 61 ] van den Heuvel C, Rayward W B. Facing interfaces: Paul Otlet's visualizations of data integration [ J ] . Journal of the American Society for Information Science and Technology, 2011, 62( 12 ) : 2313– 2326.

[ 作者简介 ] 邱均平,男,1947年生,武汉大学中国科学评价研究中心主任,教授,博士生导师。

余厚强,男,1990年生,武汉大学信息管理学院博士研究生。

吕红,男,1987年生,武汉大学信息管理学院博士研究生。

李小涛,男,1986年生,武汉大学信息管理学院博士研究生。

收稿日期:2013– 11– 04

