DOI:10.15941/j.cnki.issn1001-0424.2014.14.016

信息可视化分析工具的比较分析*

——以 CiteSpace、HistCite 和 RefViz 为例

田 军

【摘 要】文章选取 SCI 中以“Digital Libraries”为主题的文献为源数据，基于用户视角和功能视角对 CiteSpace、HistCite 及 RefViz 进行分析比较，用户视角包括界面设计、软件操作、数据处理、节点控制、可视化图谱显示，功能视角包括国家与机构、著者、关键文献、研究热点及趋势。在此基础上，归纳总结各软件在学科知识领域应用中的共性与特性，期望为用户正确选择信息可视化工具提供有益的帮助。

【关键词】信息可视化工具 CiteSpace HistCite RefViz

Abstract: The paper selected the literatures on subject of “digital libraries” as the source data from SCI, and then compared CiteSpace, HistCite and RefViz from user perspective and function perspective. The former covered interface design, software, data processing, node control, and visualization map display. The later included countries and institutions, journals, authors, key literature, and research hotspots and trends. In addition, this paper summarized the specialties and commonness of CiteSpace, HistCite and RefViz in applying and anticipates it can provide help when users select information through visual tools.

Key words: information visualization tools CiteSpace HistCite RefViz

1 引言

大数据时代，信息呈现出数据量大、价值密度低及快时效等特点，人们保存、检索、分析及利用能力面临着巨大的挑战[1]。信息可视化工具的出现，使得大规模非数值型信息资源得以视觉呈现，为人们理解和分析数据提供了帮助。信息可视化工具种类繁多，CiteSpace、RefViz 及 HistCite 3 种广受关注[2]。本文在应用层面对这 3 款工具进行同源数据的对比分析，剖析各个分析软件在学科知识领域应用中的优缺点及各自的特色。

本文的文献数据来源于美国 Thomson Scientific 集团开发的 web of science 平台，该平台收录学科齐全，所收录的引文数据质量较高[3]。在 web of science 平台，时间跨度选择 all years，数据库来源选取 SCI-EXPANDED、SSCI、A&HCI、IC 以及 CCR-EXPANDED，数据类型选择 all types，检索词为：TITLE = “digital librar* ”，经过学科提炼得到 4961 条数字图书馆方向的文献数据，将这些数据以“txt”文本文档形式下载并保存，数据采集日期为 2014 年 2 月 12 日。

2 基于用户视角的引文分析软件的比较分析

2.1 界面设计比较分析

相同点：CiteSpace、HistCite 和 RefViz 3 款软件的主界面均包含菜单项、显示框和参数设置项。

CiteSpace 的菜单项包含了文件、项目、数据、网络、可视化等栏目，显示框包含了数据基本信息和软件运行结果报告两部分，参数设置项包含了时间切割、术语来源、术语类型、节点类型、图谱精简和可视化类型等部分。如图 1 所示。

HistCite 的菜单栏包含了文件选项、数据分析、可视化窗口等栏目，在主界面上可以依据记录、作者、期刊等参数对数据进行统计排序并显示，参数设置项包含了国家和地区、机构、语言、文献类型、出版年份等选项，如

* 本文系教育部人文社会科学研究规划基金项目“信息用户在图书馆与社会网络使用中的双重融合实证研究”（项目编号：10YJA870029）的研究成果之一。

图 2 所示。

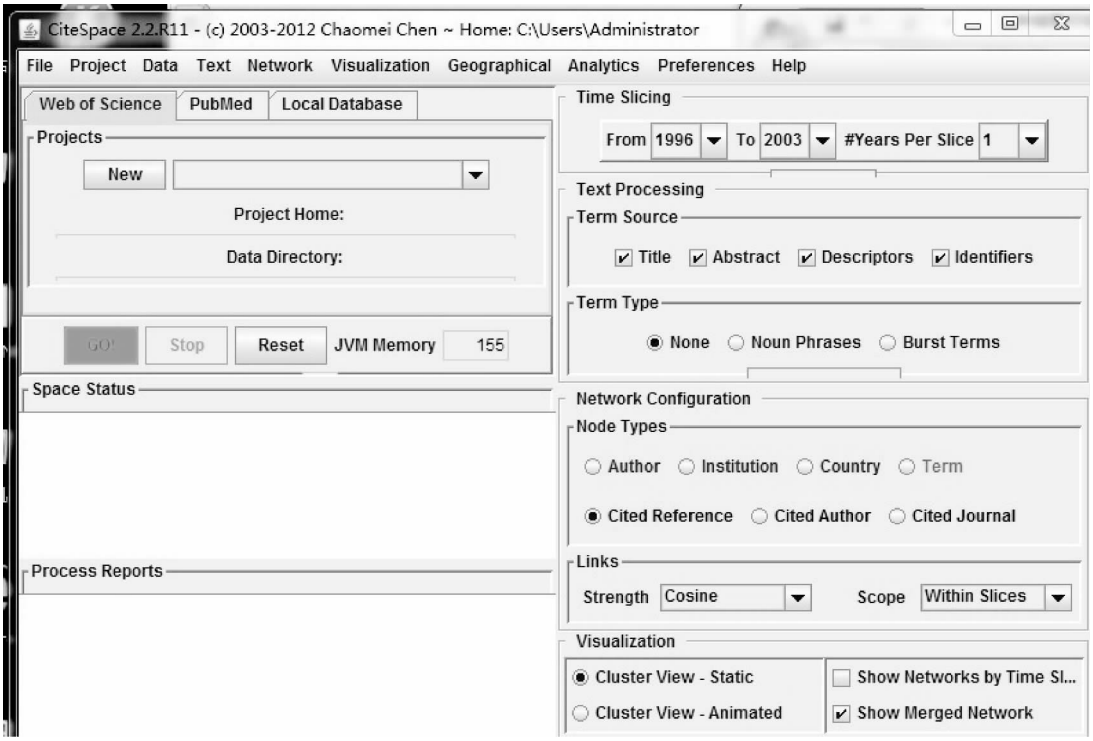

图 1 CiteSpace 主界面视图

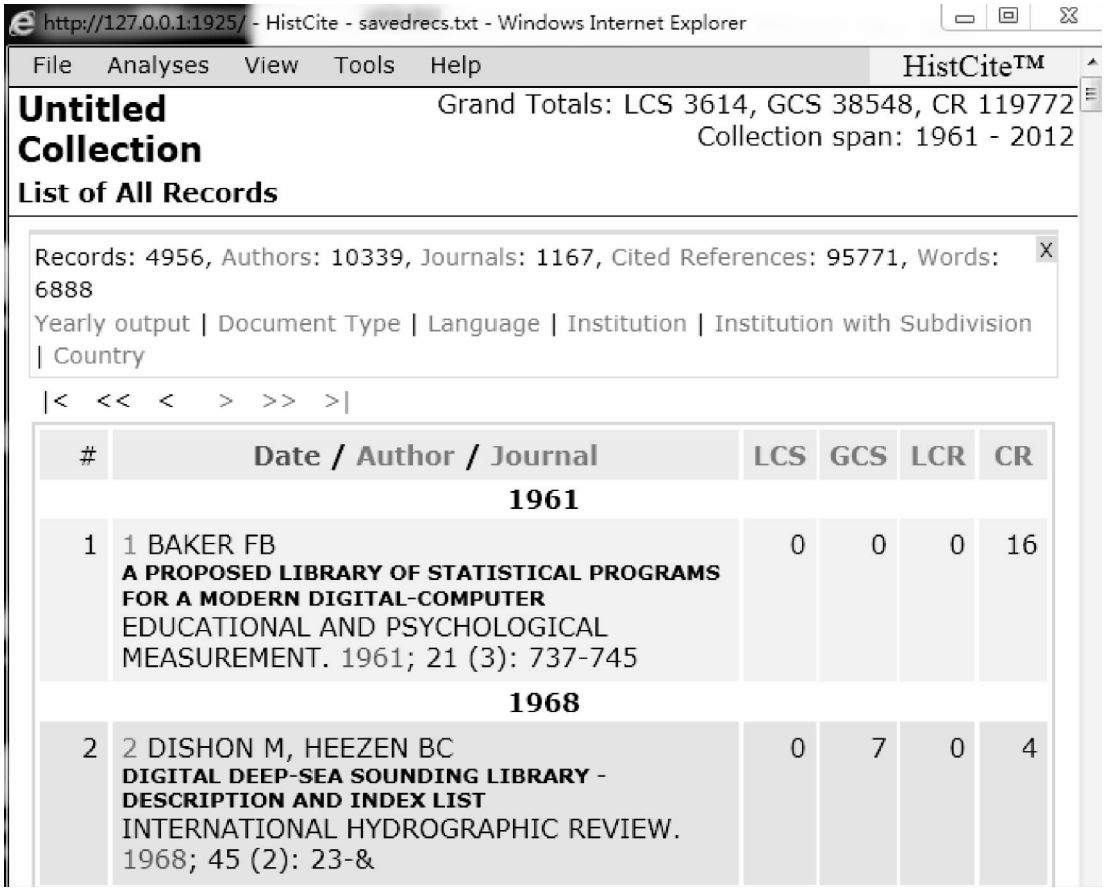

图 2 HistCite 主界面视图

RefViz 工具的菜单栏包括文件选项、可视化图谱类型选项、工具栏和帮助菜单栏目，在主界面上有 galaxy 和 matrix 两种显示类型，galaxy 显示的是文献聚类图谱，而 matrix 图谱则显示主题词之间或者主题词和文献分组之间的关系，RefViz 主界面提供两个基本参数控制图谱类型，分别为 galaxy 和 matrix，如图 3 所示。

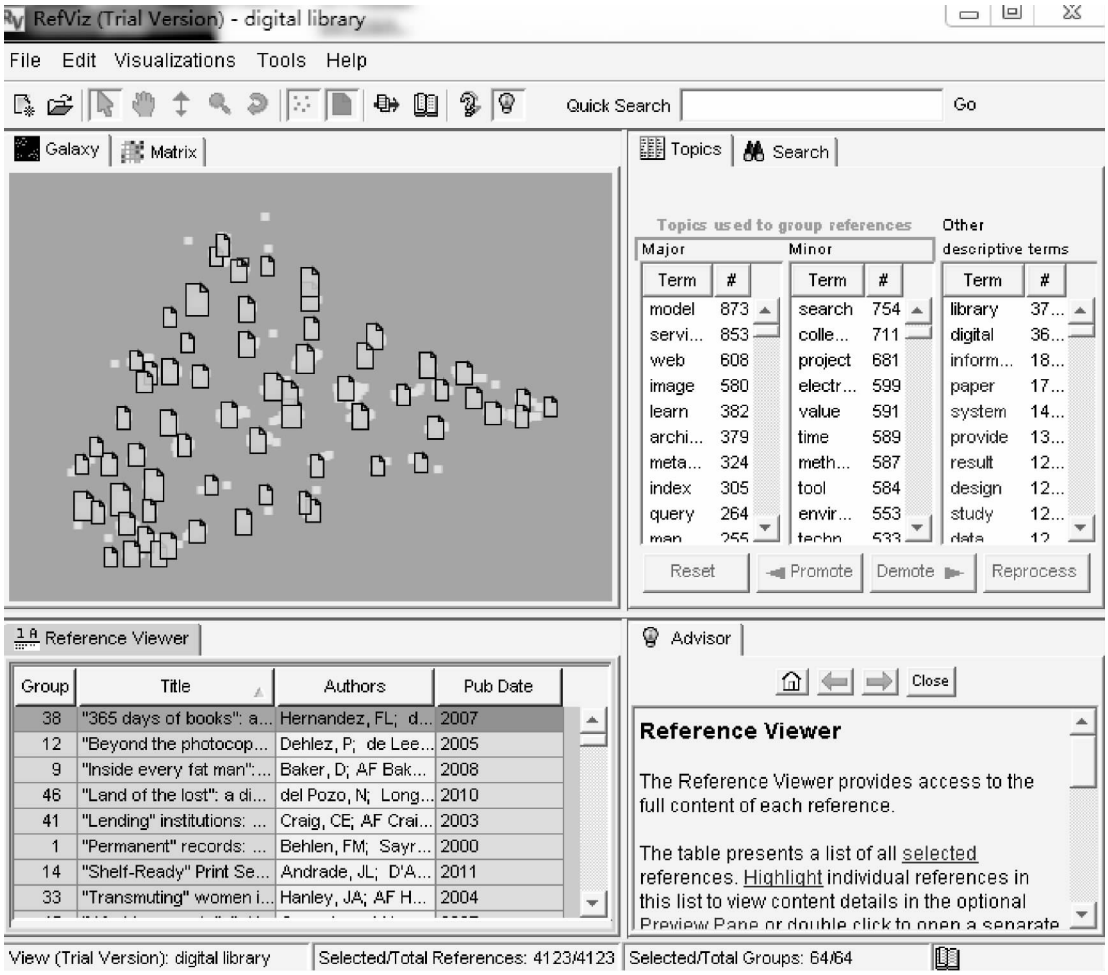

图 3 RefViz 主界面视图

不同点:

CiteSpace 的操作界面提供了数据库类型选项，而且提供了项目位置、数据存放位置、数据分析过程和结果报告等选项。

CiteSpace 可以同时运行多个窗口，以不同窗口显示各节点的引文历史轨迹图; HistCite 仅支持运行一个窗口; 而 RefViz 的节点显示窗口都集中在主界面上，一次只能打开一个窗口。

CiteSpace 具有良好的提示功能，当软件无法运行数据时会弹出消息框，提示数据格式是否正确、或是否以“download. txt”命名; RefViz 没有相应提示，主界面给用户提供网络或本地数据库两种选择，当无法将数据导入 HistCite 时，软件仅提供“文件无效”、“格式有误”或“路径错误”等信息，没有向用户提供更为明确的提示信息。

参数设置方面，CiteSpace 有时间段分割、网络精简类型、Threshold Interpolation、Top N per slice; RefViz 参数方式少，仅按照 galaxy 和 matrix 控制图谱类型; HistCite 提供的参数设置方法单一，仅提供根据如作者、引文、地区显示结果。

2. 2 软件操作难易程度分析

相同点: CiteSpace、HistCite 和 RefViz 3 款软件均为英文版本，尚没有简体中文版本，这无疑加大了用户学习和使用的难度。

不同点: 在人机交互方面，HistCite 和 RefViz 这两种工具界面简洁、操作步骤简单，软件参数设置简单，比较

容易掌握；CiteSpace 在环境支持方面，其运行需要 JAVA 环境的支撑，不同的软件版本对操作系统有不同的要求；功能方面，CiteSpace 中通过很多参数干预结果。

2.3 数据处理功能分析

2.3.1 数据处理与转换功能

相同点：3 款软件均可处理 web of science 平台的数据，必须以“download.txt”的形式命名数据文件，仅支持英文字母或者数字。

不同点：如果在 CiteSpace 中处理 web of science 数据，则需要将从 SCI 下载的原始数据集以“download.txt”格式另存[4]。打开 CiteSpace，利用引文数据转换器将数据集导入转换并输出单独的文件夹中；HistCite 也需作数据处理和转换，对比 HistCite 的样本数据可发现，每条记录之间均存在一个空行，而从 SCI 下载的数据缺空行，因此需要 Notepad++软件处理，利用替换功能，将“ER\nPT”换为“ER\n\nPT”，实现了在每条记录之间加空行，该软件通过原始数据所在文本文档导入数据，但是文本文档的命名不能出现中文；RefViz 通过原始数据所在文本文档直接导入数据，保证文本文档是“download.txt”格式即可。

2.3.2 支持的语言类型比较

相同点：CiteSpace、HistCite 和 RefViz 3 款软件均支持英文数据的识别和处理，英文数据来源主要为 web of science 数据库平台。

不同点：与 HistCite 和 RefViz 相比，CiteSpace 可以支持中文数据的处理，中文数据的来源为南京大学中文社会科学引文索引数据库（CSSCI），只是需要做格式和编码的转换。

2.4 节点控制分析

2.4.1 节点的缩减功能比较

CiteSpace 具有良好的节点选择缩减功能[5]，第一种方法在主界面中有 4 个选项框用来控制节点的形成和数量。第一个为 Top N Slice，提取时间段被引频次最高的前 N 个，系统初始值为 30，N 越大，则形成的图谱相对更加全面，第二个为 Top N% per slice，将每个时间段的节点按照被引频次降序排列，仅保留前 N%，第三个为 Threshold Interpolation，可设置 C，CC 和 CCV，最后一个选项框为 Select Citers，按 Continue，再设定方法 1，2 或 3。第二种为在形成的图谱中右击某个不被显示的节点，选择“hide node”可隐藏节点，通过隐藏节点便可以达到控制图中节点数量的目的。

HistCite 则通过 LCS 和 GCS 两种模式和“count”和“value”控制节点的数量，例如在 LCS 模式下，选择“count”则表示显示在当前文献集合中被引频次最高的节点数量，软件初设值为 30，选择“value”代表显示在当前文献集合中被引频次超过设置值的节点。通过“count”和“value”可以控制形成图谱中显示的节点数量和权值。

RefViz 尚没有节点数量缩减控制的功能。

2.4.2 节点间的位置控制比较

CiteSpace 通过 3 种方法控制图谱中节点的位置，一种是依据被引频次控制节点位置，以节点被引频次的多少控制图谱中节点的数量和位置。另外一种是依据“centrality”控制图谱中节点的数量和位置，节点的中心性越高，反映了网络中任意两点之间经过该节点的最短路径越多。最后一种方法为通过在图谱中拖动节点来控制节点的位置，这表明节点在图谱中的位置并非绝对，CiteSpace 图谱的动态性更强，CiteSpace 没有对节点的绝对位置进行控制。

HistCite 及 RefViz 不能进行节点间位置的控制。

2.5 可视化图谱比较

2.5.1 图谱的显示方式

CiteSpace 的图谱显示方式多样，比如聚类图（cluster）、时间图谱（timeline）和时区图谱（timezone）[6]。RefViz 以文件夹的形式将所有文献分组并编号形成文献聚类视图。HistCite 则依据时间分区形成了节点之间引用关系视图。

2.5.2 可视化图谱的类型比较

通过 CiteSpace 可以形成很多可视化图谱，比如文献聚类视图、国家和地区合作网络、著者合作网络、时间和时区图谱等等。RefViz 的可视化图谱类型也较多，比如在 Galaxy 视图下的文献聚类视图，Matrix 视图下根据相关

度或者数值形成的主题词共被引图谱和主题与文献分组共引视图等。而 HistCite 的可视化图谱仅为引文编年图，但是 HistCite 具有很强大的文献计量统计功能。

2.5.3 可视化图谱色彩比较

CiteSpace 形成的图谱为彩色图，如图 4 所示，通过不同的颜色反映节点不同的被引年份，而且用大红色和玫瑰红色代表突变性很强的节点，用来反映突变术语和研究前沿。节点之间的连线代表它们之间的共被引关系，连线的颜色代表节点之间首次共被引的年份。通过节点之间连线的粗细和各个节点的大小，颜色的差异很容易辨识节点的被引以及共被引情况。

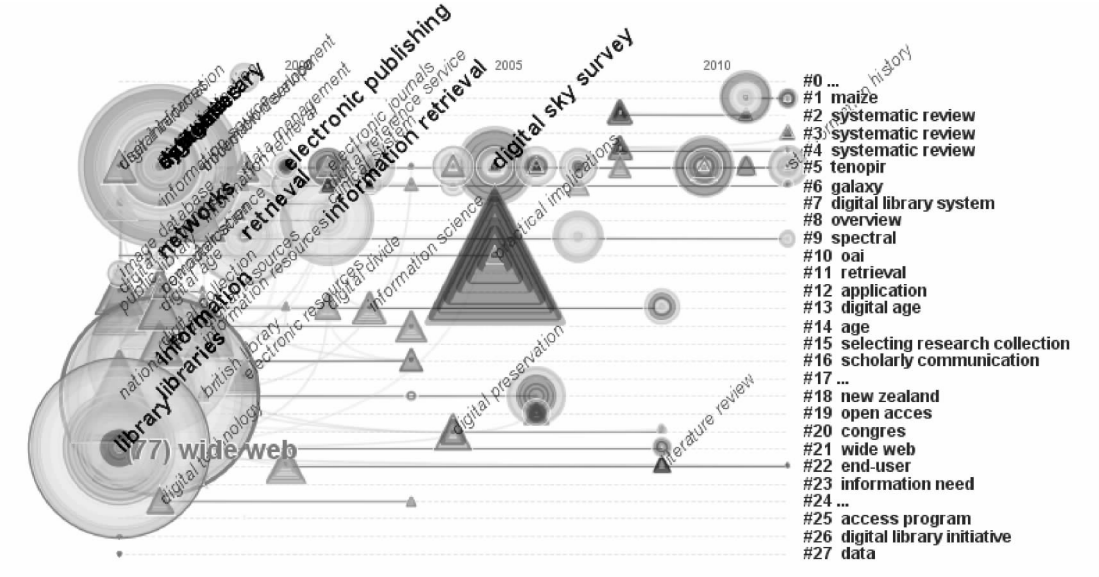

图 4 数字图书馆研究前沿时间图谱

借助 HistCite 形成的引文编年图颜色较单一，为黑白色，不易辨识节点之间的种种关系以及节点的被引历史，缺乏生动性。

RefViz 的 Galaxy 视图中，绿色代表选中的文献分组，蓝色代表未选中的记录，通过节点的颜色及大小很容易辨识不同的文献组和文献，Matrix 视图中颜色更丰富。Relevance 模式中红色表示主题词有强关联性，蓝色表示关联性较弱，白色区域表示没有关联性。

2.5.4 可视化图谱解读难易度

CiteSpace 的可视化图谱清晰，易解读，通过不同的颜色、节点的大小、节点之间的位置很容易辨认被引频次高的节点，很容易了解数字图书馆文献的力量分布、核心作者、热点等，想了解数字图书馆的研究前沿，根据“burst”值对图谱进行操作，就会形成以红色的节点。

HistCite 的图谱解读难易程度取决于节点间的被引频次和节点的数量，如果引文编年图中的节点过多且被引频次过多，则图谱中很多连线交织，形成一个很复杂的网络图，且图谱是黑白色，增加了图谱解析的难度。此外，因其图谱中节点之间的连线粗细一致，研究人员无法辨析关键节点之间的共被引关系。

在 RefViz 的 Galaxy 视图中，软件将所有文献进行分组编号并以文件夹的形式展现出来[7]。文献分组的大小反映了该组文献数量的多少，文献分组之间的距离反映它们的研究内容的相似程度，通过文献分组密集的区域很容易辨识数字图书馆领域的研究方向。在 Matrix 视图中，图谱以不同的颜色反映主题词之间或主题词同文献分组间的关联。

3 基于功能视角下的引文分析软件比较分析

从数字图书馆方向的国家与机构、著者、期刊、关键文献、研究热点及趋势几个方面对 CiteSpace、HistCite 和 Refviz 3 款软件进行对比分析。

3.1 国家与机构分析功能

相同点：CiteSpace 与 HistCite 都有揭示国家和机构的功能；都提供国家和机构统计分析；两款工具都可以对国家或者机构节点信息进行排列，Citespace 可依据被引频次的高低对文献信息进行排列；两者的国家和机构统计

信息均以表格的形式显示，但是两者的统计表格均不能被直接输出。

不同点：CiteSpace 的国家与机构显示多样化，以表格和视图的形式显示国家与机构的信息，HistCite 只能以表格的形式显示；CiteSpace 可以视图的形式揭示国家与机构的合作情况，以节点和连线的形式反映国家与地区间的关联度，依据 HistCite 无法揭示国家与机构的合作关系；在排列依据方面，CiteSpace 除了依据频次统计，还提供中心度，HistCite 可揭示国家或者机构在当前文献中的被引数；在时间方面，HistCite 的国家与机构统计中缺乏时间因素。

3.2 作者分析功能

相同点：CiteSpace 和 HistCite 均具有揭示重要作者的功能；均以图表的形式反映某领域的重要作者；两者都能以被引频次等属性对作者进行统计；两者形成的统计信息表格均不能直接被输出，需要人工辅助统计或者用截图软件导出。

不同点：CiteSpace 以可视化视图展现著者的共引情况，HistCite 不能揭示著者间的合作情况；CiteSpace 依据中心性和被引频次共同确定重要作者，而 HistCite 依据发文量和在当前文献集合中的被引数判断重要作者；CiteSpace 提供了突变率检测的功能，通过 burst 值可寻找短期内引用次数激增的作者，HistCite 不能揭示著者的变化程度。

3.3 关键文献分析功能

相同点：CiteSpace、HistCite 和 RefViz 都有揭示学科领域关键文献的功能；都是从某领域文献集合与被引文献集合的关系中寻找关键文献；均能够以可视化图谱的方式展现文献数据之间的种种关系；均具有对关键文献进行统计并显示的功能。

不同点：从分析方法来看，CiteSpace 软件以被引用次数和中心性为标准来判断文献的重要程度，HistCite 的关键文献分析功能以当前文献集合为分析对象，RefViz 将文献按内容和关联度分组，以文件夹的方式展现文献组情况；从重要性参数看，CiteSpace 以被引频次、共引次数、突变率及中心性作为衡量标准，HistCite 以 LCS、GCS 为参数，RefViz 以文献分组数量及分组的位置作为衡量文献重要性的标准。

3.4 研究热点和趋势分析功能

相同点：CiteSpace、HistCite 以及 RefViz 均具有识别某一个学科领域的研究热点的功能；都是从文献数据的题录部分提取词组的方式来确定热点主题词；均能对提取的主题词或者关键词进行词频统计并分析。

不同点：从分析方法来看，CiteSpace 以词频统计及共词方法分析主题词之间的引用是共引关系，HistCite 以词频分析方法统计热点主题词，RefViz 以词语加权方式将词语按照重要性分布；从热点词组的显示来看，CiteSpace 以图谱方式显示关键词间的引用关系，从引用历史轨迹可查看主题词年份分布，HistCite 只能以表格的方式对主题词进行统计，无法揭示主题词之间的关系，RefViz 的 Matrix 视图能够揭示热点主题词之间的关系；从主题词的衡量指标看，CiteSpace 以中心性和被引频次作为衡量依据，此外依据突变率分析词语的变化趋势，HistCite 以 TLCS 及 TGCS 作为衡量热点主题词的依据，RefViz 依据词权确定主题词，另外对热点词语进行权值干预；从揭示研究前沿的程度看，CiteSpace 依据主题词的 burst 值来判断某领域的研究前沿和发展趋势，而 HistCite 和 RefViz 对研究前沿术语的揭示功能方面比较弱。

4 结论

经过以上的比较，可以发现这 3 种软件在用户设计、功能设计方面都具有相同的特性，但具体从每一个比较标准来看，各软件之间确实又存在着差别，也正是这些差别使得不同的软件有其各自不同的使用领域与范围。CiteSpace 能够显示一个学科或知识域在一定时期发展的趋势与动向，形成若干研究前沿领域的演进历程；RefViz 可以确定和精炼领域关键词，展示发展的热点趋势，确定顶尖杂志发表的文章主要是什么方向；HistCite 能够用图示的方式展示某一领域不同文献之间的关系，帮助我们绘制出一个领域的发展历史，定位出该领域的重要文献，以及最新的重要文献。

注释

[1] 刘泽渊，陈悦，侯海燕. 科学知识图谱：方法与应用 [M]. 北京：人民出版社，2008：60－70.

[2] 侯剑华，陈悦. 战略管理学前沿演进可视化研究 [J]. 科学学研究，2007（25）：15－21.

[3] 栗春娟，侯海燕. 科技政策研究代表人物与核心文献可视化网络 [J]. 科学学研究，2008（6）：1164－1167.

（下转第 54 页）

村与村、村与屯之间农家书屋的图书流转交换，使农民群众既有“新书”看，又能最大限度地发挥书屋的作用。

3.2 探索数字书屋建设，降低资源的配置成本

利用数字农家书屋，可以降低图书资源的配置成本。“数字农家书屋”是利用卫星数字发行系统，将高品质的多媒体文件、电子图书、杂志、报纸、音像等内容以数字方式投递到任何一个“农家书屋”中，农民可以通过电视、投影、电脑等设备阅读和观看。数字农家书屋具有海量存储、传输快捷、出版成本低廉、节约印刷耗材、媒介形态多样、突破时空限制等优点，这些优点正好可以用来解决实体农家书屋出版物数量有限、不能及时更新以及边远地区报刊传递速度过慢等问题，完善了实体农家书屋功能，有助于实体书屋的可持续发展。

3.3 以农民为主体，提高文献利用率

3.3.1 加大宣传力度，吸引农民参与

农家书屋毕竟是面对文化层次较低的农村群众，只有想方设法扩大其影响力和知晓度，让更多的农民群众知道、了解并积极参与到读书借书的行动中来，农家书屋才能真正发挥作用，才能真正实现惠民、利民、为民。因此，可以采取办宣传橱窗、发宣传单、开村民会等各种形式，向农民群众宣传农家书屋，让每一个农民朋友都知道村里有一个农家书屋，在农家书屋能找到致富门路、致富技术，这些书报、杂志是免费阅读的，空闲时候大家都可以来看书报。

3.3.2 整合文化资源，提升服务效能

农家书屋既是农民读书、看报、查资料和学技能的集散地，同时又是开展农技知识培训、文化活动的主要阵地。通过与我市各农村文化活动中心、现代远程教育、文化信息资源共享工程、科普“一站、一栏、一员”等资源的有效整合，不断提升服务效能。

3.4 加强书屋管理员队伍建设

农家书屋要管好用好，图书管理员是关键。首先要严格图书管理员的选拔和任用。调查显示，专职人员管理的书屋无论从看书人数和借书数量都比兼职人员管理的书屋强得多。建议有条件的地方要在农村选择文化程度较高、责任心较强、热爱书屋管理工作的人员来担任书屋的专职管理工作。二是要加大培训力度。部分书屋的图书管理员是由村干部兼任，对图书管理、分类、上架、借阅等相关业务不甚了解。要通过定期开展检查评比活动，以增强服务能力。三是要适度解决好工资待遇问题。吉林地区农家书屋多数建在村委会，管理员大部分由村干部兼任，从长远看，应该给予管理员适当的补助。可参照村干部的工资标准解决其工资待遇，从而调动工作积极性。这一问题不仅是农家书屋后续保障资金的重要方面，同时也是农家书屋可持续发展的重要因素之一。

4 结语

农家书屋工程是一项复杂的系统工程，它的建设任重而道远。实现“全覆盖”只是农家书屋工程“万里长征”迈出了第一步，工作中的“管”“用”“活”，样样都比“建”来得更为复杂。因此，今后将对农家书屋可持续发展的路径进行不懈的探析，努力使这项惠民工程发展、壮大。

参考文献

1 中宣部，新闻出版总署．“农家书屋”工程实施意见．

2 程亚男．关于总分馆建设的几点思考[J]．图书与情报，2010(3)：14.

3 彭飞．农村图书馆事业发展模式研究[J]．图书馆，2010(2)：39-41.

4 高巾，刘兹恒．农村图书馆服务的可持续发展路径．图书馆论坛，2011(5)：138-140.

孙 薇 吉林市图书馆。

(上接第95页)

[4] ChenC. Visualizing Semantic Spaces and Author Co-Citation Networks in Digital Libraries [J]. Information Processing and Management，2010(35)：401-420.

[5] Nelson M J. Visualization of Citation Patters of Some Canadian Journals [J]. Scientometrics，2009(2)：279-289.

[6] Reid E F，ChenH. Mapping the Eontem Porary Terrorism Research Domain [J]. International Journal of Human-Computer Studies，2011(6)：42-56.

[7] 张国海．电子政务研究文献的量化可视分析[J]．情报杂志，2011(6)：82-86.

田 军 西安航空学院图书馆。