

基于用户需求驱动 的文献引证可视化 导航工具设计与实现

汪雪峰¹, 刘玉琴², 李荣荣¹, 李海若¹

(1. 北京理工大学管理与经济学院, 北京 100081; 2. 北京印刷学院绿色印刷包装产业技术研究院, 北京 102600)

摘要: 为引导和辅助用户快速地浏览和检索文献, 提高用户面对海量数据资源时对文献资源的利用效率, 设计实现了能够按照用户需求进行多层次文献引证可视化导航的系统工具. 通过采用双曲树可视化方法, 明确了该工具的技术思路、数据库设计、界面设计、主要功能和关键技术实现. 最后, 以专利文献引证为例进行了实证研究.

关键词: 文献引证; 双曲树; 可视化

中图分类号: TP 391 **文献标志码:** A **文章编号:** 1001-0645(2014)01--

Design and Implementation of Demand-Driven Visual Navigation System of Literature Citation

WANG Xue-feng¹, LIU Yu-qin², LI Rong-rong¹, LI Hai-ruo¹

(1. School of Management & Economics, Beijing Institute of Technology, Beijing, 100081, China; 2. Academic of Printing and Packaging Industrial Technology, Beijing Institute of Graphic Communication, Beijing 102600, China)

Abstract: To improve utilization efficiency of the mass literature resources, guide and assist user to browse and retrieval, the demand-driven visual navigation system of literature citation is designed and implemented based on hyperbolic tree, and its technical idea, database design, interface design, basic framework, main function and key technologies are elaborated. At last an empirical illustration about patent reference is proposed.

Key words: citation; hyperbolic tree; visualization

文献引证数据具有重要的参考作用. 通过引证数量统计可以发现特定技术领域核心文献的分布, 进而开展科技成果评价; 通过跟踪文献引证信息可以发现领域竞争对手和潜在竞争对手及其研发策略; 通过引证文献与参考文献的综合可以揭示文献间相互联系、相互影响与促进的关系, 揭示特定技术领域内的知识流动、技术扩散和技术融合的路径. 文献引证数据的分析利用在科技信息资源利用中占有非常重要的地位^[1-2]. 为提高用户在面对海量数据资源时对文献资源的利用效率, 引导和辅助用户快速的浏览、检索文献, 本文构建了基于用户需求驱

动的文献引证可视化导航工具, 提高了文献引证导航的效率.

1 需求与技术思路

1.1 需求分析

在面对海量科技文献资源时, 如何使用户高效利用引证数据进行资源的查找, 对信息资源服务机构提出了技术挑战. 在实践应用中, 基于文献间的引证关系, 应用信息可视化技术对引证文献和参考文献进行导航浏览, 是有效利用引证数据的方式之一.

收稿日期: 2012-12-14

基金项目: 国家自然科学基金资助项目(70903007); 国家部委基金资助项目(2011M010500)

作者简介: 汪雪峰(1977—), 男, 副教授, E-mail: wxf5122@gmail.com.

通信作者: 刘玉琴(1979—), 男, 高级工程师, E-mail: liuyuqin2004@126.com.

文献引证导航与可视化技术关系密切. 文献引证导航界面的呈现, 导航元素的设计、安排, 导航结构、导航过程的呈现与控制都需要借助可视化技术. 可视化是增强导航系统表现力、交互性和可用性的一个重要手段. 如汤森路透集团的 Web of Science 引文索引数据库提供了文献间一级、二级的前向引证(参考文献)、后向引证(引证文献)、双向引证关系图, 图中每个节点表示一个文献, 并用不同颜色按照节点文献所属作者、机构、期刊、学科等对节点加以区分, 并切换节点文本显示内容^[3]. 中国知网的引文网络图^[4], 由节点文献、参考文献、二级参考文献、引证文献、二级引证文献、共引文献和同被引文献等七个节点构成, 每个节点指向对应文献的文献列表, 方便用户对参考文献和引证文献的查看; 万方数据知识服务平台^[5]也提供了一级的引证文献和参考文献列表. 这些对于引证文献和参考文献的导航方式存在共同的特点: 引证文献和参考文献一般到二级为止, 并且相关的文献一次全部罗列出来, 用户被动地接受数据, 无法按照自己的需求对更深层级的文献进行快速地、选择性地浏览和查看. 虽然通过一些替代的方式也能实现相似功能, 但操作的复杂程度和系统交互时间显著增加. 基于以上分析, 设计开发一个能够按照用户需求进行多层次参考文献和引证文献的导航工具是十分必要的.

1.2 技术思路

面对海量的科技文献资源, 实现按照用户个性化需求的多层级文献引证导航需要解决两个关键问题: 一是提取引证数据并可视化显示的系统响应速度问题, 特别是在网络环境下. 二是较大规模的文献数据及其引证关系可视化显示问题.

文献引证的目的可以分为正常动机的引证和非正常动机的引证, 而且即便是正常动机的引证, 仍然会存在文献联系程度上的差异, 使得引证文献和参考文献间往往具有不同的学科、技术类别、机构、作者等信息. 对于文献的读者来说, 并不是所有的引证文献和参考文献都是其关注内容. 因此, 在解决提取引证数据并可视化显示的系统响应速度问题上, 本文的技术思路是基于用户的需求进行引证文献和参考文献的提取. 即在提取某个文献的引证文献或参考文献时, 只有那些满足用户设定条件的文献才被提取出来并可视化显示, 且每次仅对一个文献进行操作, 其余文献则不提取也不显示.

在解决较大规模的科技文献数据及其引证关系

可视化显示问题上, 与 Web of Science 引文数据库的文献引证可视化技术相同, 采用操纵大型层次结构的焦点+背景技术即双曲树可视化方法, 并不再对树的层级进行限制.

2 系统设计

2.1 数据库设计

文献引证可视化导航工具数据库中存储了文献的基本信息和文献间的引证关系信息, 为此设计了文献信息表和文献引证关系表, 如图 1 所示. 文献信息表中存储了文献题目、作者、机构、分类、时间、储存路径等, 文献引证关系表中存储了文献引证关系中的父文献 ID 和子文献 ID, 并与文献信息表中的文献 ID 构成外键约束.

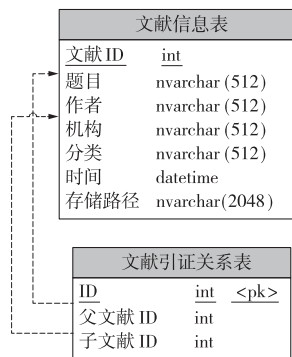


图 1 可视化导航工具数据库结构

Fig. 1 Database structure of visual navigation tool

2.2 界面设计

文献引证可视化导航工具的菜单栏主要包括: 文件、查询、图形显示、语言、帮助 5 个菜单项; 工具栏主要包括: 打开、保存、提取参考文献、提取引证文献、链接、背景设置、图形样式、根节点、截图、主页链接、退出等十项操作; 界面下方为可视化图形输出区域. 工具支持输入文献 ID 后在数据库中实时查询文献及其引证关系并可视化呈现, 也支持图形文件“打开”对历史数据进行可视化呈现; 并且通过鼠标的移动切换不同的节点到屏幕中心, 使得在文献数量较大时仍能对文献间的引证关系进行浏览.

2.3 功能设计

文献引证可视化导航工具的功能设计如下:

① 初始文献查询. 在工具栏的查询文本框中输入文献 ID 号, 点击提取引证文献或提取参考文献按钮, 搜索数据库中该文献的引证文献 ID 或参考文献 ID, 即一级引证数据.

② 引证信息的可视化表示. 搜索到的文献数

据以双曲树形式显示在图形区,通过鼠标双击或拖拽图形区的文献节点将其布置在屏幕中心,从而获得用户最大的关注。

③ 文献引证导航。用户依据文献节点的题目、作者、机构、时间、类别信息,通过鼠标双击将其置于屏幕中心。此时,工具栏的查询文本框中显示屏幕中心的文献 ID。重复之前的操作步骤,进行二级引证文献或参考文献的查询和显示。依次类推,进行第 N 级文献引证的可视化呈现,进而实现文献引证的导航功能。

④ 查看文献详细信息。在导航的过程中,对于屏幕中心的文献,单击菜单栏的连接按钮,如果数据库中存储了该文献的存储地址,则直接打开文献所在的存储路径。

⑤ 可视化图形样式设置。工具栏中的背景和图形样式设置,可用于修改背景、图形节点、连线和文字颜色。

⑥ 图形文件的保存。可视化的图形结果可直接截取为 JPG 格式,也可以将其通过类的序列化保存到硬盘,再通过文件打开进行可视化输出。

3 系统实现

3.1 引证关系查询与存储

在工具栏的文本框中输入初始文献节点,应用 SQL 语句获取其引证文献和参考文献,以 C# 树形结构类 `TreeView` 存储文献引证关系,树节点类 `TreeNode` 存储文献节点。输入的初始文献作为 `TreeView` 的根节点,一级参考文献和引证文献均作为根节点的子节点,并附加状态属性表示子节点是父节点的参考文献还是引证文献。对于用户感兴趣的一级参考文献,通过鼠标双击使其在工具栏的文本框中显示出来,再次采用 SQL 语句获取引证文献和参考文献。依次递归,形成了以初始文献为根节点的引证关系树。

3.2 双曲树算法技术实现

双曲树算法主要原理是将树结构在双曲空间进行布局,然后映射到欧式空间的庞莱卡圆盘进行显示。欧式空间中两个相同大小的区域离庞莱卡圆盘中心越近,在双曲空间中所占用的空间越小;反之,双曲空间中两个大小相同的区域离原点越近在庞莱卡圆盘中所占用的空间越大。当关注的树节点被放到双曲空间的原点后,在欧式空间该节点显示在圆盘中心(电脑屏幕中心),且占用的空间最大。

算法包括双曲空间树节点布局,双曲空间向欧式空间映射两个主要步骤。首先,将树的根节点坐标设置为双曲平面的原点(0,0);然后,把根节点的扇形区域平分给根节点的子节点,每个二级子节点拥有自己的扇形区域,再把每个二级子节点的扇形区域平分给其所拥有的三级子节点,依次递归进行节点的分布;最后,采用庞莱卡投影把双曲空间中的点映射到欧式空间的庞莱卡圆盘上。具体技术实现步骤如下:

① 以复数类 `HTCoordinate(double x, double y)` 表示双曲空间中点坐标,复数的实部、虚部分别与双曲空间中点的横、纵坐标对应,扇形类 `HTSector(HTCoordinate p1, HTCoordinate p2)` 表示双曲空间中的扇形区域;

② 在双曲空间中对树结构中的每个节点进行布局,除根节点布局在原点(0,0)外,其它子节点调用如下布局过程进行坐标的递归设置;

```
HTCoordinate w = parent.Coordinates(); // 当前父节点坐标
double angle = sector.Angle(); // 当前节点父节点所拥有的扇面
z.x = length * Math.Cos(angle); // 当前节点相对于父节点的位置坐标
z.y = length * Math.Sin(angle);
z.Translate(w); // 经过变换设置当前节点在双曲空间中的坐标
```

其中,坐标变换函数 `HTCoordinate.Translate(HTCoordinate w)` 的变换规则如下:

```
double deltax = x * w.x + y * w.y;
double deltax = y * w.x - x * w.y;
double newx = x + w.x;
double newy = y + w.y;
x = (newx * deltax + newy * deltax) / (deltax * deltax + deltax * deltax);
y = (newy * deltax - newx * deltax) / (deltax * deltax + deltax * deltax);
```

③ 将双曲空间点 Z 的坐标映射到欧式空间,映射规则如下:

```
x = Math.Round(z.x * (double) max.x) + org.x; // max 为庞卡莱圆盘的大小
y = -Math.Round(z.y * (double) max.y) + org.y; // org 为庞卡莱圆盘中心的欧式空间坐标。
```

3.3 可视化图形渲染技术实现

可视化图形的输出基于微软 .NET Framework 3.0 绘图技术开发框架 Windows

Presentation Foundation(WPF)进行,分别进行节点和连线的图形渲染:

```
DrawingContext g=new DrawingContext();
g. DrawRectangle ( nodeFillColorBrush, nodeBorder
ColorPen,x, y, width, height);
g. DrawLine(lineColorPen,StartPoint,EndPoint).
```

4 实证分析

以美国专利文献的引证数据为例,随机选择专利号为 7395421 的文献作为用户关注的初始文献.该专利文献是 Intel 公司有关计算机双操作系统启动技术和设备的专利,技术主分类为 713/2. 利用本文构建的可视化导航工具,设定用户筛选条件为“与 7395421 具有相同的申请机构和技术类别”,对其引证专利和参考专利进行导航和浏览.

首先,在工具栏的查询文本框中输入专利号(数据库中以专利号作为文献 ID),点击提取引证文献

按钮,获取一级引证专利;点击提取参考文献按钮,则获取一级参考专利.

其次,为了按照设定的条件获取二级引证专利和参考专利,切换可视化图形节点显示内容,使每个节点显示节点文献的所属机构.

搜索显示:观察图形发现:一级引证专利中 8239667 为 Intel 公司的专利,参考专利中 6367074、6484309、6631469、7162629 为 Intel 公司的专利;其中引证专利 8239667 和参考专利 6367074、6631469 的技术分类包含 713/2. 依次点击这三个专利文献使其置于屏幕中间,获取二级引证专利和参考专利. 查看专利 8239667 的引证专利时,未发现该专利的引证专利;查看专利 6367074 的参考专利时,虽然有多个参考专利,但这些参考专利均非 Intel 公司所有. 查看专利 6631469 的参考专利时,发现满足预先设定的筛选条件,如图 2 所示.

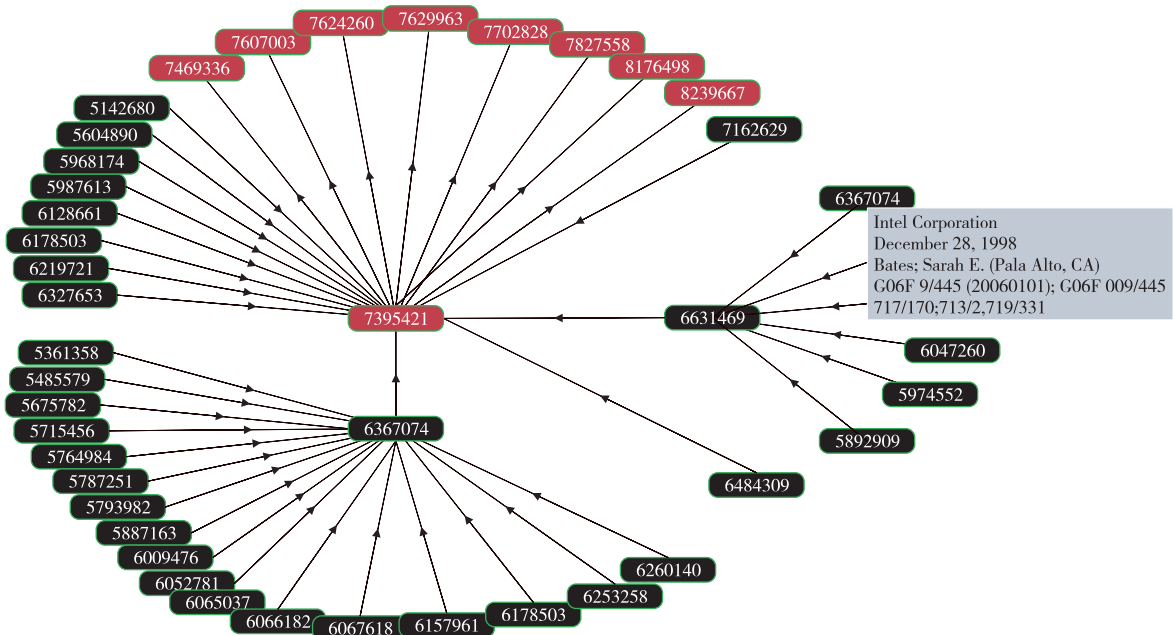


图 2 二级引证专利和参考专利文献
Fig. 2 Secondary citation patent and patent reference literature

进一步获取其三级参考专利,发现没有满足预先设定条件的三级参考专利,如图 3 所示.至此,完成了满足预先设定筛选条件的引证专利和参考专利的导航和浏览.

5 结束语

鉴于文献引证数据的利用在信息资源建设和服务中的重要作用,文中设计实现了文献引证可视化

导航工具,按照用户的需求对多层级的文献引证进行可视化导航,以双曲树结构进行可视化呈现,从而引导和辅助用户快速地浏览、检索文献,有助于推进信息资源建设和服务实践.

参考文献:

- [1] Lin Yan, Chen Jian, Chen Yan. Backbone of technology evolution in the modern era automobile industry: an analysis by the patents citation network [J]. Journal of

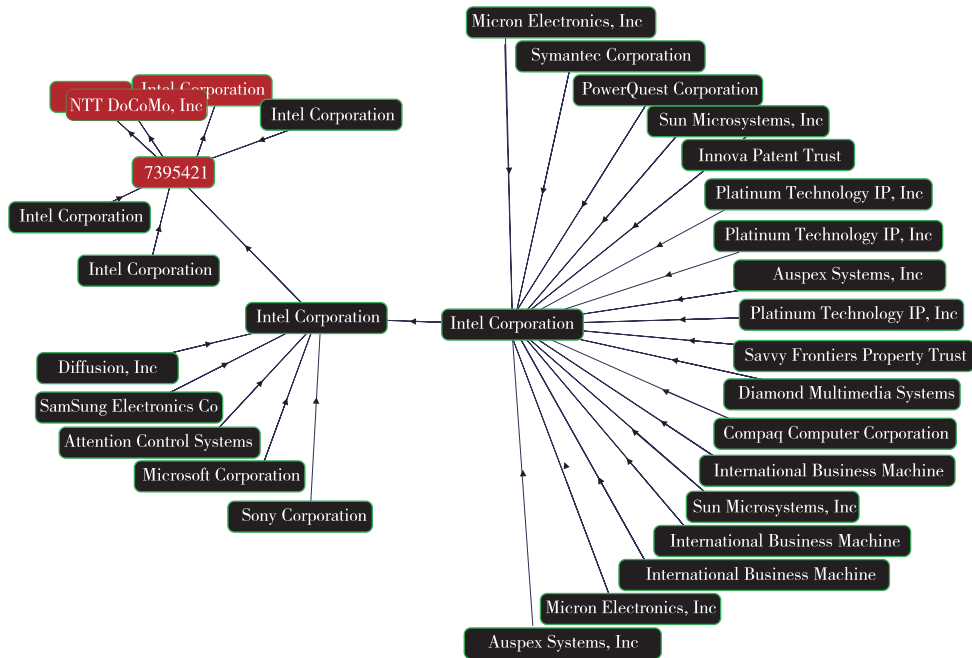


图 3 三级参考专利文献所属机构
 Fig. 3 Affiliation of tertiary patent reference literature

Systems Science and Systems Engineering, 2011,20(4): 416 - 442.

[2] 陈菊英. 文献引证关系研究简要评述[J]. 图书馆理论与实践, 2003(1): 47 - 49.

Chen Juying. Overview of the study on citation relations [J]. Library Theory and Practice, 2003(1): 47 - 49. (in Chinese)

[3] Web of knowledge. [EB/OL]. [2012. 8. 15] <http://apps.webofknowledge.com/>

[4] CNKI [EB/OL]. [2012. 8. 15] <http://acad.cnki.net/Kns55/brief/result.aspx?dbPrefix=CJFQ>

[5] 万方数据知识服务平台. [EB/OL]. [2012. 8. 15] <http://www.wanfangdata.com.cn/>

Wanfang knowledge service. [EB/OL]. [2012. 8. 15] <http://ww.wanfangdate.com.cn/>

(责任编辑: 孙竹凤)