

引用格式: Wang Xuemei, Li Xin, Ma Zhicheng, *et al.*. Development of the Bibliometric Geographic Information System based on Open Source Platform[J]. Remote Sensing Technology and Application, 2014, 29(5): 891-896. [王雪梅, 李新, 马志程, 等. 基于开源平台构建文献计量地理信息系统[J]. 遥感技术与应用, 2014, 29(5): 891-896.] doi:10.11873/j.issn.1004-0323.2014.5.0891

基于开源平台构建文献计量地理信息系统

王雪梅^{1,2}, 李新¹, 马志程³, 马明国¹, 赵国剑⁴

(1. 中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000;

2. 中国科学院兰州情报文献中心/中国科学院资源环境科学信息中心, 甘肃 兰州 730000;

3. 国网甘肃省电力公司信息通信公司, 甘肃 兰州 730000; 4. 武汉大学测绘学院, 湖北 武汉 430072)

摘要: 地理信息系统(Geographical Information System, GIS)在文献计量学中的应用是 GIS 应用拓展的一个新方向, 近 10 年来得到越来越广泛的应用, 特别是在空间相关信息展示和查询等方面发展迅速。为了方便非 GIS 专业人员应用 GIS 技术进行文献计量信息挖掘、制图和空间分析等操作, 利用开源 GIS 软件 DotSpatial 开发文献计量 GIS 工具, 可以满足用户数据导入、指标计算、图层和视图操作以及制图输出等需求。研究结果显示: 通过研发的文献计量 GIS 工具可以比较快捷地实现文献指标的空间展示和制图目的。

关键词: 地理信息系统; 文献计量; 空间展示; 制图; 开源平台

中图分类号: P 208.2 **文献标志码:** A **文章编号:** 1004-0323(2014)05-0891-06

1 引言

学科交叉、融合已成为当代科学发展主要特点之一, 多学科协同是科学发展的必然趋势^[1]。地理信息系统(Geographical Information System, GIS)和文献计量学的集成是一个新的发展方向^[2], 得到越来越广泛的应用, 特别是在空间相关信息展示和查询等方面发展迅速^[3]。

随着计算机技术的发展, 开源软件(Open Source Software, OSS)逐渐成为主流选择之一^[4]。越来越多的事实表明, 价格不是开源软件唯一的优势, 定制化能力、灵活性、安全和质量等更具有价值^[5]。开源 GIS 软件是开源软件家族中的一员, 近年来发展迅速, 越来越多的科研机构、企业和组织开展了开源 GIS 软件的开发和应用研究^[6-7], 在图书馆中也有应用案例^[8]。

本文利用开源 GIS 软件开发一套专门针对文献信息挖掘和空间展现的文献计量 GIS 分析工具, 该软件具有标准文献资料地理位置信息提取和汇总、指标自动计算、简单空间分析和结果空间展示等功能, 可为相关研究人员提供 GIS 支持下的科研影响力动态分析, 促进文献计量学与 GIS 技术的集成。此处介绍的文献计量 GIS 软件的版本主要针对 Web of Science(WOS)检索的文献数据, 研究单元主要为国家和城市两种尺度。

2 总体设计

2.1 系统简介

文献计量 GIS 工具包含文献信息字段提取模块、指标计算模块、地理位置相关信息数据库、空间信息可视化模块和空间分析模块。其总体结构如图 1 所示。

收稿日期: 2014-01-15; 修订日期: 2014-07-18

基金项目: 中国科学院文献情报能力专项项目“基于知识流地理扩散的科研活跃中心动态变化监测”, 国家科学图书馆业务发展领域前沿扫描项目“国际科技评价活动与前沿方法扫描”和国家自然科学基金项目(40701133)资助。

作者简介: 王雪梅(1976-), 女, 重庆永川人, 副研究员, 主要从事资源环境科学计量评价研究。E-mail: wxm@lzb.ac.cn。

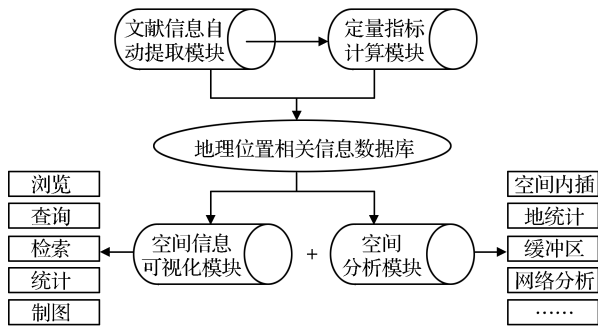


图 1 文献计量 GIS 软件总体结构

Fig. 1 Overall structure of the bibliometric GIS software

2.2 数据库设计

文献计量 GIS 工具的后台数据库主要包括属性数据和空间数据,其数据结构见图 2。其中属性数据以表格的方式存储,考虑用户的使用成本和数据库本身的复杂程度,选择 Microsoft Access 作为属性数据库管理系统。属性数据库主要包括:

- (1)地名信息数据,根据不同级别研究单元,形成不同专题的地名信息数据表格,例如国家、城市和研究机构等,在同一个级别上针对不同区域可以生成各自所需的地名信息数据表格。
- (2)文献信息提取数据,主要是从原始文献字段中通过统计获取不同级别研究单元的发文、引文、被引和合作等方面的基本信息。
- (3)文献计量指标数据,在基本信息基础上计算出的常用文献计量指标表格。
- (4)辅助数据,存储研究单元的社会经济等方面的统计数据。

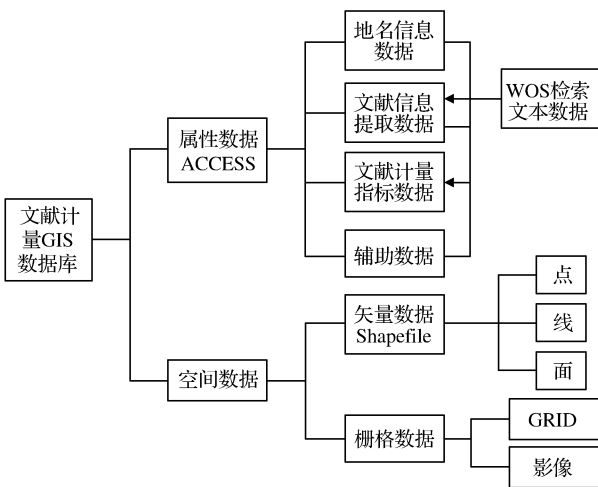


图 2 文献计量 GIS 的数据结构

Fig. 2 Data structure of the bibliometric GIS

空间数据主要包括矢量数据和栅格数据,矢量数据从格式上分为点、线、面 3 种基本类型,选择

Shapefile 格式存储。栅格数据包括用于空间分析的 Grid 格式数据和用于影像存储的 GeoTIFF 格式数据。空间数据库主要包括:

- (1)基础地理信息图层,包括经纬网、行政区划、公路铁路等矢量图层。
- (2)用于关联的基础图层,有世界范围的国家面状图和点状图、中国省份面状图和点状图、世界及中国的城市点状图。
- (3)关联后的图层,通过属性表将属性数据关联到矢量层,可以作为其中的字段直接用于空间操作。但关联的表格如果太多,矢量层的属性表就会变得过于庞大,字段太多不便于后期操作,因此可以将关联后的图层另外命名存储。

(4)空间分析输出的栅格图层,基于矢量数据进行空间插值、缓冲区分析、密度图计算等操作,生成 Grid 格式的栅格图层。

(5)影像数据,可以提前制备研究区影像图,或者从第三方网络数据库直接调用影像数据作为背景层显示。

2.3 制图范围与投影

根据当前的空间展示和分析需求,选择 3 个不同空间尺度的区域作为制图范围。地图投影均采用 WGS84 参考系下的经纬度坐标。

(1)全球范围,坐标范围为 180°W ~ 180°E、90°S ~ 90°N,在全球国家和城市级别上的空间展示和分析。

(2)中国范围,坐标范围为 70°E ~ 115°E、0° ~ 55°N,在中国各省和城市级别上的空间展示和分析。

(3)欧洲范围,坐标范围为 15°W ~ 35°E、30°N ~ 75°N,在欧洲范围国家和城市级别上的空间展示和分析。由于欧洲国家数量较多,在全球范围进行空间展示时不易分辨各个国家的特征,因此设置欧洲范围,在更大比例尺下对其进行展示,便于用户更直观地了解该区域的文献计量指标的空间分布特征。

3 功能设计

3.1 数据导入和信息提取

数据导入主要分为 3 种类型,第一种是文本文件导入,为 WOS 检索结果输出的文本文件格式,每篇文章主要包含标题、作者、作者地址、摘要、关键

词、期刊、卷、期、页、出版年、学科分类和被引频次等信息。通过各要素的特征识别符可以自动将每篇文章的信息录入到表格中,便于后期的统计分析。针对作者地址信息要做标准化处理,首先增加国家和城市两个字段,将作者地址中的国家和城市信息分别提取和存储到这两个字段中,然后与国家和城市地名信息系统进行匹配。国家和城市地名信息系统后台也是一个表格,其标准地名与矢量数据中的地名一致,从而可以支持关联。进行匹配时,有的地名可能存在多种表现形式,通过在表格中增加一个存储变化形式的字段可以实现成功匹配,如果某个地名存在多个表达形式时,用分号进行分割列入即可。城市地名众多,基础数据库中很难包含所有作者的城市信息,通过增加功能可以添加城市记录,输入城市名称和经纬度坐标,即可把该城市添加到矢量数据和地名信息系统中。

第二种导入的数据类型是汤森数据分析器(Thomson Data Analyzer, TDA)统计的结果, TDA软件可以将WOS检索下载的文本文件导入进行统计分析,并将统计结果以表格形式输出。这里TDA输出统计结果主要指国家和城市的发文篇数、合作篇数、被引次数和学科分类等信息。同样在TDA中国家和城市的名称也存在标准化问题,通过地名数据字典可以进行编辑,也可以针对新出现的形式进行新增和补充。

第三种类型是国家和城市的社会经济统计数据,以数据表格的形式存储,主要包括不同年代的人口数量、国内生产总值、科研投入和国土面积等信息。

第一种文本数据录入的结果为单篇文章的信息,需要进行信息提取和统计才能使用,最终和TDA的输出结果相似。后两种导入的数据类型为国家和城市两个尺度的统计结果。

3.2 文献计量指标计算

如果想使用其他文献计量指标来定量评价研究单元更多的文献特征,系统需要具备计算现有主要文献计量指标的功能。计算时需要关联到相关的表格,例如计算国家人均发文量,就需要先将发文统计结果的表格与人口数量的表格通过国家名称进行关联,然后计算得到结果,并作为矢量数据表格的一个新增字段添加进去,用于后期的空间展示。有些指标还需要对文本导入的每条文献的记录进行计算才能输出指标结果,例如h指数。由于指标众多,新的

指标还在不断涌现,不可能将所有指标都提前定制到系统功能中,因此在矢量数据表格操作中设置了编辑公式的功能,用户可以根据指标算法调用所需字段和编辑计算符号,最终输出自定义的指标计算结果。

3.3 空间关联

调用用于关联的基础图层,通过国家或者城市名称字段实现矢量数据与统计结果或者文献计量指标的空间关联。关联字段是进行空间关联的重要桥梁,矢量数据属性表和属性数据表格必须同时包含这一字段,而且针对同一目标的名称必须具有唯一性。因此前面数据录入时,地名的规范化显得十分重要。针对城市名称经常存在同名的情况,例如英国和美国都有城市Cambridge,为了实现其唯一性,可以对同样名称的城市通过增加国家信息或者邮编信息进行区分。

3.4 空间展示

空间展示方面包括以下常用功能:①工程操作,主要是建立工程文件,存贮相关设置,便于再次打开进行修改。②图层操作,主要是实现图层的添加、移动和删除等操作。③视图操作,是在视图界面进行操作和测量等。④图例设置,对各个图层进行符号、颜色、图型、尺寸等设置。⑤制图输出,配置图形输出的参数,例如视图、图例、比例尺和指北针等。

3.5 空间分析

GIS的奠基人之一Goodchild(1987年)曾指出地理信息系统真正的功能在于它利用空间分析技术对空间数据的分析^[9]。空间分析技术在文献计量学中的应用还处于初期阶段,亟待进一步拓展。尝试在文献计量GIS中增加空间分析功能,其目的是为了从发文、合作和引用等文献数据中挖掘科研活动的时空分布特征和动态变化趋势等信息,为相关决策者和研究人员把握学科发展态势提供参考。可以考虑以下空间分析功能:

(1)叠加分析:将从科研文献提取的空间相关数据与地理要素按图层进行叠加,进而分析科研文献的地理分布特征及可能的空间影响因素。

(2)缓冲区分析:主要用于分析发文或者引文数量与研究单元或研究区域的空间位置关系,度量随着与研究区的空间距离变化,文献数量的密度变化规律。

(3)网络分析:主要基于文献间的引用关系构建

引文网络,或看作者间的相互合作构建地理网络,挖掘时空关联性等。

(4)密度计算:通过对城市或者研究机构发文或引文特征的密度计算,可以获取研究力量或科研活动热点的时空分布特征。

(5)趋势分析:通过对研究单元科研文献统计指标的时间序列分析,例如回归模型模拟,可以探讨和预测其在时空方面的发展趋势。

图 3 是空间分析模块的流程图。虽然此处提到了较多可以在文献计量中发挥作用的空间分析功能,但要在文献计量 GIS 软件平台中实现这些功能,开发的工作量比较大,目前只做了个别尝试,这里主要是对空间分析潜在功能的实现设计,更多的开发工作需要以后继续开展。

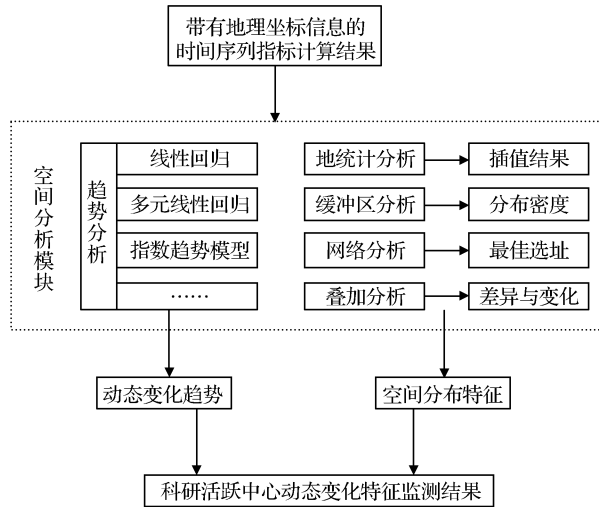


图 3 空间分析模块流程图

Fig. 3 Flowchart of spatial analysis modules

4 软件实现

4.1 开发平台介绍

DotSpatial 是基于 .Net Framework 4.0 平台开发的可扩展的开源 GIS 类库,可以作为桌面 GIS 库用于制作通用的 GIS 数据,提供空间数据分析。它不仅可以对数据进行浏览,高级用户或者开发人员还可以编写嵌入式的函数、模型等来满足个性化的需求^[10]。DotSpatial 除了开源免费外,还具有跨平台、插件兼容、组件模式、地图绘制、渲染效率高和专题功能强大等优点^[11]。DotSpatial 可以对 Shapefile 文件进行编辑和数据转换,对 DBF 属性表进行编辑,还支持大量的 GIS 数据格式,包括 GeoTIFF、ArcInfo ASCII 和 binary grids 等^[12]。

4.2 文献计量 GIS

图 4 展示的是开发的文献计量 GIS 系统视图操作界面,从图中可见,栏目部分包含了菜单和快捷按钮;左侧为图层显示和操作界面,右侧为视图显示;底端为状态显示栏,提示程序运行进度、视图显示比例尺和鼠标所在位置坐标等信息。

用户首先新建任务,选择导入文本数据的目录即可自动导入从 WOS 下载的文献记录。通过选择已定制文献计量指标,用户可以获取感兴趣的统计结果。

属性表可以进行字段添加、删除和记录修改等操作,并可根据某个字段进行排序。通过工具菜单可以实现查询和检索,以及进行指标计算。

通过可视化的参数设置来实现图例的定制,可选择唯一、定制、特征值和数量这 4 种形式来进行图例分级。在颜色设置中可以设置起始色和终止色,也可以设置渐变色。点状要素可以自动设置点的大小系列,线要素可以自动设置线宽系列。在通过某一字段数值进行分级时,属性字段可以通过右下角统计结果的分布进行分界线的快速调整。

图 5 为热点中心渲染,针对城市发文等点状数据,通过设置起始和终止颜色以及热点半径生成热点分布图。图 6 为制图界面,在此界面下用户可以添加视图、图名、图例、比例尺和指北针等要素,并可对这些要素的属性进行设置。

5 结语

为了便于非 GIS 专业人员应用 GIS 技术进行文献计量的空间信息挖掘、制图和分析等操作,本文尝试利用开源 GIS 平台 DotSpatial 开发文献计量 GIS 工具。在总体设计和功能设计的基础上,利用 DotSpatial 平台初步实现了软件的开发,可以进行数据导入、指标计算、图层操作和视图操作等功能,最终实现输出可视化的制图结果,基本达到预期的软件开发目的。但在空间分析模块开发方面还有待下一步工作的开展。

虽然现在有较多的商业 GIS 软件可以支持文献计量的空间数据分析和制图等功能,但由于这些系统功能强大,用户熟悉软件需要花费较多精力。同时,这些软件使用成本也很高,不便于后期广泛推广和应用。利用开源平台开发文献计量 GIS 工具是当前有效的开发模式之一,可以降低用户后期使用