文献题录信息挖掘技术方法及其软件 SATI 的实现

——以中外图书情报学为例

刘启元　叶　鹰

（浙江大学信息资源管理系，杭州，310027）

［摘　要］ 利用 C＃编程技术基于.NET 平台设计开发出文献题录信息统计分析工具软件 SATI，可导入处理 EndNote 格式、NoteExpress 格式及 NoteFirst 格式的国内文献题录数据和 HTML 格式的 WoS 国际文献题录数据，进行数据格式的转换、字段信息的抽取、词条频次的统计和知识单元共现矩阵、词条频率逐年分布矩阵及文档词条矩阵的构建，以辅助生成聚类图、多维尺度图谱、网络知识图谱、策略坐标图等可视化结果。以 2006～2010 年中外图书情报学各十种具有代表性的核心期刊刊载的 17440 篇论文数据为实例，基于聚类与多维尺度分析结果，呈现出中外图书情报学三大主要研究领域，并结合共词分析与社会网络分析方法，通过绘制共现网络知识图谱与策略坐标图，进一步揭示研究领域结构的内部联系及其特征。

［关键词］ SATI　共词分析　聚类分析　多维尺度分析　知识图谱　策略坐标图

［中图分类号］ G254.92　［文献标识码］ A　［文章编号］ 2095-2171(2012)01-0050-09

A Study on Mining Bibliographic Records by Designed Software SATI: Case Study on Library and Information Science

Liu Qiyuan　Ye Ying

(Department of Information Resource Management, Zhejiang University, Hangzhou 310027)

[Abstract] A bibliographic information analysis software named SATI (Statistical Analysis Toolkit for Informetrics) is developed using C＃ based on Microsoft .NET platform. The national data fitting EndNote, NoteExpress and NoteFirst can be imported into SATI as well as international data as HTML (output from WoS). For the purpose of getting the clustering graph, multidimensional scaling map, network knowledge map, and strategic diagram, four basic functions including transforming raw data into XML, extracting selected elements, counting terms frequency and building knowledge units matrices have been implemented. Taking 17440 articles published in ten core Library and Information Science journals both at home and abroad during 2006 to 2010 as the sample, this paper revealed three potential research fields of LIS research area based on the consistency between clustering analysis and multidimensional scaling analysis results, and figured out relations and features of subject areas by interpreting the network knowledge map and strategic diagram.

[Key words] SATI　Co-word analysis　Cluster analysis　Multidimensional scaling analysis　Knowledge mapping　Strategic diagram

对学科领域的研究总结有多种途径，其中通过对科技文献统计分析较为常用。学术期刊有科学研究成果公布、传播、积累、评价和学术导向等功能[1]。Law 等学者曾比较指出

［作者简介］ 刘启元，男，硕士研究生；叶鹰，男，教授，博士生导师。

2012
第1期
（季刊）

科技文献对于把握学科研究结构和发展的作用与优势，大部分研究领域的主要学者都将研究成果贡献于科技文献，而且从电子期刊数据库采集大量数据也更加低廉和便利[2]。随着期刊全文数据库的普及和信息处理技术的进步，文献题录作为描述文献外部特征的重要元数据集合，通过计算机技术和计量方法来对一定学科领域内的题录数据进行处理与分析，可揭示文献集合内外部特征并延伸挖掘出学科研究结构(Structure)与发展动态(Dynamics)。

本文基于.NET平台利用C#编程语言设计开发出具有通用价值的文献题录信息统计分析工具(Statistical Analysis Toolkit for Informetrics，SATI)。软件可导入处理四种格式国内外文献题录数据，具有题录格式转换、字段信息抽取、词条频次统计和知识矩阵构建等四大功能。结合文本预处理技术和基于共现分析的信息可视化技术，以图书情报学为实例进行关键词共现分析，借助SPSS、Ucinet等软件生成可视化结果，以揭示国内外图书情报学研究领域结构，直观呈现知识单元间的关系，并通过中外比较探讨国内外研究的共性和差异。

1 技术方法

国外关于文献信息统计分析的技术方法和应用软件相对较为成熟，已有社会网络分析软件Ucinet(内嵌开源软件Pajek，Netdraw和Mage)、科学计量学研究软件Bibexcel、文献可视化信息分析软件Citespace等，但这些软件都主要针对Web of Science (WoS)等国外数据库平台开发，需要专门的数据输入格式，对于国内期刊全文数据库题录数据不能直接处理。为兼顾处理国内期刊题录数据和国际WoS题录数据，本文尝试设计开发对国内外期刊全文数据库进行文献题录信息统计分析的统一软件。

技术方法的关键在于对国内和国际期刊全文数据库所导出题录数据的兼顾处理，设计思路是先将不同来源的数据格式统一转换为SATI处理的专用XML格式，抽取指定字段信息，得出条目元素(即词条Term，指语句元素的最小单元，可以是字、词或短语，包括关键词、主题词、文本预处理后的分词等受限词或自然词)的频次统计文档，再分析知识单元间的共现关系和频率分布，生成共现矩阵、分布矩阵和文档词条矩阵，继而实现对海量文献信息的定量分析和可视化呈现。

按此思路，我们首先对国内三大期刊全文数据库知网、万方和维普的题录数据格式进行细致的分析，找出了三大主流输出格式EndNote格式、NoteExpress格式和NoteFirst格式题录数据的字段信息特征，主要体现在用于区别不同字段的标识符和词条的分隔符(如知网新平台EndNote格式题录数据中，关键词字段的标识符是“%K”，关键词之间的分隔符是“;;”或“,”，但不同数据库平台和期刊会稍有不同，需进行特殊处理)，利用同样的方法再同时对WoS导出的HTML格式题录进行特征分析，通过编程实现抽取不同字段信息，转换生成为SATI软件专用的XML格式文件；在自动导入转换后的XML文件后，基于抽取出的相应字段信息，再利用频次统计算法得出词条频次统计文档；然后将频次降序排列表中相应数量的条目元素(词条)作为知识单元按照适当的算法模型构建出共现矩阵、分布矩阵和文档词条矩阵。设计思路如图1所示。

为便于后期数据的进一步处理和可视化呈现的需要，软件可同时生成Excel格式矩阵和.txt文本格式全矩阵。只要将共现矩阵文档导入相应的数据分析软件(如Ucinet、SPSS等)，即可构建出知识单元聚类图、多维尺度分析图、共现网络知识图谱和策略坐标图等。

2 功能实现

我们把自主设计开发的文献题录信息统计分析工具命名为SATI，作为免费辅助软件，软件官方网址为：http://sati.liuqiyuan.com，主界面如图2所示。软件采用中英文两种界面，DataGridView和TextBox两种视图。

目前软件主要实现了以下四大功能：

(1)题录格式转换：支持输入WoS数据库平台导出的HTML格式、国内期刊全文数据库导出的EndNote格式、NoteExpress格式和NoteFirst格式题录数据。对英文题录关键词、主题词、标题和摘要字段进行文本预处理(Tokenization，Stop Words[注1] & Stemming[注2])操

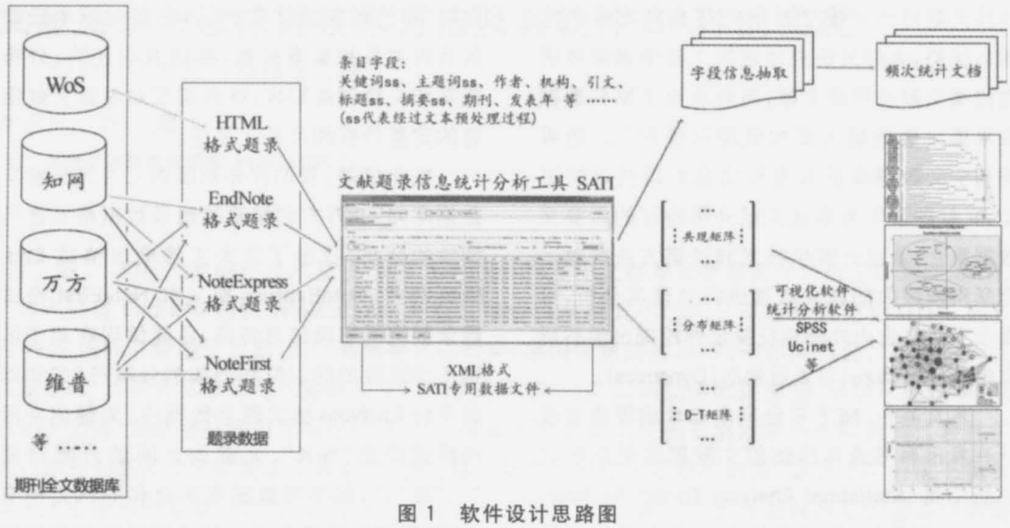

图 1　软件设计思路图

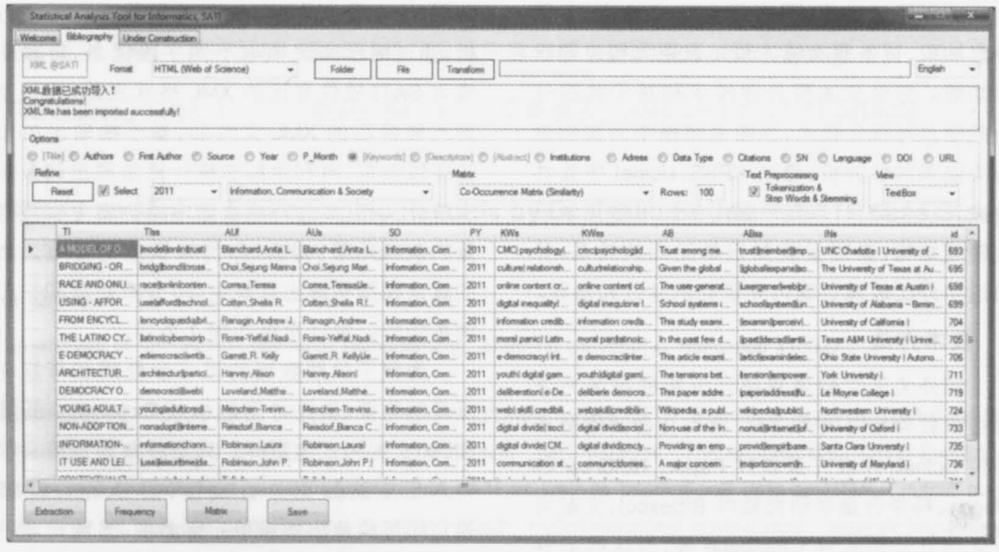

图 2　文献题录信息统计分析工具 SATI 主界面

作，中文题录标题、摘要进行中文分词[注3]和停用词[注4]处理后，将题录自动转化为 XML 格式的 SATI 专用数据文件，以为后期题录数据的存储、交换和分析提供便利。

SATI 专用数据文件（XML 格式）采用简洁的三层树状结构，实例如图 3 所示。用户可将期刊全文数据库导出的题录文件导入 SATI 自动生成 XML 格式专用文件，或根据实际需要将相关数据文件自行转换为 SATI 处理所需格式。

（2）字段信息抽取：在“Options”面板可以选择抽取标题、作者、第一作者、文献来源、出版年、关键词、主题词、摘要、机构、地址、文献类型、引文、语种、DOI 和 URL 等字段信息，并可保存为 .txt 文本文件。支持勾选“Text Preprocessing”选项，选取经文本预处理后的标题、关键词、主题词、和摘要等信息。还可利用“Refine”面板按照出版年和文献来源进行数据集合的限定，并在此基础之上进行下一步的统计分析。

（3）词条频次统计：根据抽取到的字段信息对条目元素（包括：关键词、主题词、作者、引文、机构、发表年、期刊、文献类型等）的频次进行统计和降序排列，同样可以按照时间和期刊对数据进行限定，生成相应的频次统计文档，

```
<?xml version="1.0" encoding="utf-8" ?>
<!--注释信息-->
<Newell>
<Bibliography id="1">
    <AUs>Bar-Ilan, J</AUs>
    <AUf>Bar-Ilan, J</AUf>
    <TI>Informetrics at the beginning of the 21st century - A review</TI>
    <TIss>informetr|||begin|||21st|centuri||review|</TIss>
    <SO>JOURNAL OF INFORMETRICS</SO>
    <LA>English</LA>
    <DT>Review</DT>
    <KWs>informetrics| bibliometrics| scientometrics| webometrics</KWs>
    <KWss>informetr|bibliometr|scientometr|webometr|</KWss>
    <DEs>AUTHOR COCITATION ANALYSIS| JOURNAL-CITATION-REPORTS| UNIVERSITY WEB SITES| ...</DEs>
    <DEss>author cocit analysi|journal citat report|univers web sit| ...</DEss>
    <AB>This paper reviews developments in informetrics between 2000 and 2006. ...</AB>
    <ABss>|paper|review|develop||informetr||2000||2006|...</ABss>
    <INs>Bar Ilan Univ | </INs>
    <AD>Bar Ilan Univ, Dept Informat Sci, IL-52900 Ramat Gan, Israel.</AD>
    <CIs>PINSKI G, 1976, INFORM PROCESS MANAG, V12, P297 |MOED HF, 1985, RES POLICY, V14, P131 | ...</CIs>
    <SN>1751-1577</SN>
    <PY>2008</PY>
    <VL>2</VL>
    <IS>1</IS>
    <SP>1</SP>
    <EP>52</EP>
    <DOI>10.1016/j.joi.2007.11.001</DOI>
</Bibliography>
<Bibliography id="2">
    ...
</Bibliography>
</Newell>
```

图 3　SATI 专用 XML 格式数据文件实例(部分)

并可保存为.txt 文本文件。

(4)知识矩阵构建:软件可生成三类共八种矩阵。

①词条共现矩阵。可自行设定共现矩阵输出行列数,将频次降序排列表中的相应数量条目元素作为知识单元进行运算,以构建知识单元共现矩阵(分相似矩阵、相异矩阵、多值矩阵和二值矩阵四种,包括关键词共现矩阵、主题词共现矩阵、引文共现矩阵、作者共现矩阵和机构共现矩阵等)。

为消除多值共现矩阵中频次悬殊对统计结果造成的影响,软件采用 Equivalence 系数[3](式 1 所示)将多值矩阵转化为元素值在[0,1]区间取值的相似矩阵,在此基础之上再生成二值矩阵和相异矩阵。相似矩阵中的数字代表矩阵元素间的相似性,数值越大关联程度越强。又因相似矩阵中的 0 值过多,统计时容易造成误差过大,软件在此基础之上自动生成相异矩阵,即相似矩阵元素值与 -1 的和为相异矩阵元素的值:

$$E_{ij} = \frac{F_{ij}^{\ 2}}{F_i * F_j} \qquad (式\ 1)$$

其中,E_{ij}为相似矩阵元素的值,对于词条 T_i 和 T_j,F_{ij}为 T_i 与 T_j 的共现次数,F_i 为 T_i 出现总频次,F_j 为 T_j 出现总频次。

②频率分布矩阵。可自行设定条目元素(词条)数,生成词条的逐年分布矩阵(分频次矩阵和频率矩阵两种)。分布矩阵的行与词条元素对应,列与发表年相对应。其中,频次矩阵元素值为词条在某年出现的频次,频率矩阵元素的值(式 2 所示)为词条在某年的频次与当年所有词条频次总和的商):

$$R_{ij} = \frac{F_{ij}}{\sum_{k=1}^{n} F_{kj}} \qquad (式\ 2)$$

其中,R_{ij}为频率分布矩阵元素的值,F_{ij}为第 i 个词条在第 j 年的频次,F_{kj}为第 k 个词条在第 j 年的频次,共有 n 个词条。

③文档词条矩阵(Document - Term Matrix):依据文本预处理结果,生成文档——词条矩阵(分多值矩阵和二值矩阵两种,包括文档——标题词矩阵、文档——关键词矩阵、文档——主题词矩阵和文档——摘要词矩阵)。多值矩阵元素的值为词条在文档中出现的频次,二值矩阵元素的值为其布尔值。文档词条矩阵的行与文档 ID 相对应,列与词条相对应,文档词条矩阵可用于文本向量的构建,利用向量空间模型(VSM)做进一步数据挖掘。

待生成 Excel 格式或.txt 文本格式的知识矩阵数据后,可将相应矩阵文档导入数据分析软件(如 Excel、SPSS、Ucinet、Netdraw 等)以生成各种基本图表、聚类图、多维尺度分析图、共现网络知识图谱和策略坐标图等。

3　实证研究

3.1　数据采集

依据 SSCI-INFORMATION SCIENCE & LIBRARY SCIENCE-JOURNAL LIST[4] 和《中文核

心期刊要目总览》2008 年版选取了 2006～2010 近五年来国内外图书情报学(LIS)各十种具有代表性的核心期刊所刊载的论文作为数据来源。十种国内外图书情报学核心期刊及其统计数据如表 1 所示。这些国内外图书情报学具有代表性的核心期刊基本覆盖了 LIS 学科的主要研究,对于数据的分析、统计与挖掘具有较好的参考价值和较高的可信度。

针对国内外期刊全文数据库的差异,从 WoS 检索并下载所需国际期刊的论文题录数据,并以 HTML 格式导出,从 CNKI 检索并下载所需国内期刊的论文题录数据,再将国内外数据分别导入 SATI 软件进行处理,分别进行数据格式的转换、关键词字段抽取、词频统计分析和共词矩阵的构建。基于 SATI 软件自动生成的 Excel 格式共词矩阵文档,利用 SPSS 和 Ucinet 最终生成国内外图书情报学高频关键词聚类图、多维尺度分析图,并结合绘制出的共现网络知识图谱和策略坐标图进一步得出和验证结论。

表 1　2006～2010 年中外图书情报学十种核心期刊论文采集处理结果

国内核心期刊	论文数	国际核心期刊	论文数
中国图书馆学报	611	Journal of Information Science	249
大学图书馆学报	698	Journal of the American Society for Information Science and Technology	1151
情报学报	809	Scientometrics	845
图书情报工作	3433	Journal of Informetrics	172
图书情报知识	681	Information Processing & Management	513
图书馆杂志	1549	Information & Management	292
情报科学	2008	Journal of Documentation	345
情报资料工作	781	Libri	128
图书与情报	1037	College & Research Libraries	376
图书馆	1415	International Journal of Information Management	347
论文量(总计/具有关键词)	13022/13001	论文量(总计/具有关键词)	4418/1867
关键词总频次	57034	关键词总频次	8654
关键词总个数	22689	关键词总个数	4838
平均每篇关键词数	4.39	平均每篇关键词数	4.64

3.2　分析方法

本文基于 SATI 自动生成的数据结果,结合共词分析法和社会网络分析法,尝试探讨中外图书情报学的研究领域结构和内部联系。共词分析法是内容分析法的一种,其认为两个能够表达文献主题内容的词条在一篇文献中同时出现,则表明二者具有一定的共现关系,共现次数越多,则关系越强。考虑到关键词的彼此孤立性,如果将一个研究领域在一定时期内的所有文献搜集起来定义为文献集合,将这个文献集合里的所有关键词抽取出来定义为关键词集合,那么对关键词集合内的关键词词频进行统计分析,频次高的关键词可被用来确定这个研究领域的研究热点主题词。将社会网络分析方法引入,可将关键词作为结点,结点位置越居中则越核心,词与词之间的关系表征于结点间的连线,连线越粗则关系越紧密。利用聚类分析和多维尺度分析法,构建聚类图和多维尺度图谱,聚在一起的若干主题词可构成一个研究主题领域。利用社会网络分析方法,绘制网络知识图谱可呈现出各个研究主题在相互作用下的分布情况(核心与边缘),因知识图谱并不能很好的反映词团(主题领域)的成熟度,难以判定某研究领域的成长趋势[5],还将基于共现矩阵构建策略坐标图,进一步解析各个研究领域的特征以验证结论。

3.3　结果呈现

3.3.1　词频统计列表

在利用技术手段进行词频统计的过程中,为了最大限度消除人为定性因素的影响,体现和反映作者群体对特定关键词的共识度,对于国内文献题录关键词并没有进行删减或对同义相似词的词频进行合并等操作,对于国际文献题录,为解决"一意多词"现象对统计结果的影响,利用文本预处理技术,进行 Tokenization 与 Stemming 操作,即只进行形变处理(标点符号、

大小写、单复数及词干提取），最终得到了如表2所示的国内外图书情报学高频关键词列表，限于篇幅只列出前20位，对于后续分析中使用到的关键词频率分布列表不再列出。

表2　国内外图书情报学高频关键词列表

	国内高频关键词（共22689个）		国际高频关键词（共4838个）	
序号	关键词	频次	关键词	频次
1	图书馆	1223	inform retriev	146
2	高校图书馆	586	knowledg manag	62
3	数字图书馆	445	bibliometr	54
4	公共图书馆	417	h index	51
5	知识管理	349	citat analysi	48
6	信息服务	289	inform scienc	46
7	图书馆学	278	inform manag	44
8	竞争情报	267	internet	39
9	情报学	194	digit librari	39
10	信息资源	180	user	37
11	本体	149	citat	35
12	知识服务	147	evalu	30
13	图书馆员	141	search engin	28
14	图书馆服务	125	knowledg shar	28
15	资源共享	124	case	27
16	信息检索	123	trust	26
17	电子政务	118	inform research	23
18	大学图书馆	116	classif	22
19	中国	108	inform system	22
20	引文分析	106	web search	21

3.3.2　聚类树状图与多维尺度图谱

聚类分析是通过聚类算法将关联密切的主题聚集在一起形成类团（研究领域）的过程，用于揭示某学科领域的研究主题结构。设定SATI软件Rows/Cols选项知识单元数为30，得出国内和国际高频关键词共现相似矩阵，将相似矩阵分别导入Ucinet进行层次聚类分析，得到如图4所示的国内外图书情报学高频关键词聚类树状图。

多维尺度分析通过测定主题词之间的距离来发现主题结构[6]，高维空间数据变换后的低维数据（二维数据）仍能近似地保持原数据间的关系。与聚类树图相比，多维尺度分析可以在较低维空间中直观地判断出某研究领域在学科内的位置[7]。将相异矩阵导入SPSS进行多维尺度分析，得到如图5所示的多维尺度图谱。

从以上聚类图和多维尺度图谱的一致性出发，国内图书情报学研究可大致分为三大主题结构：①资源导向研究：包括信息资源、资源共享、数字资源、数字图书馆、图书馆、知识产权、开放存取、对策、信息服务和网络环境共计10个主题关键词；②服务导向研究：包括图书馆服务、图书馆管理、知识服务、图书馆事业、读者服务、学科馆员、图书馆员、图书馆学、高校图书馆、公共图书馆和大学图书馆11个主题关键词；③技术导向研究：包括知识管理、信息检索、搜索引擎、情报学、引文分析、电子政务、本体、竞争情报和中国等9个主题关键词。

国际图书情报学则可大致分为以下三大主题结构：①信息技术研究：包括information retrieval、search engine、information search、web search、information science、internet、information、evaluation、information system、classification、user、digital libraries等12个主题关键词；

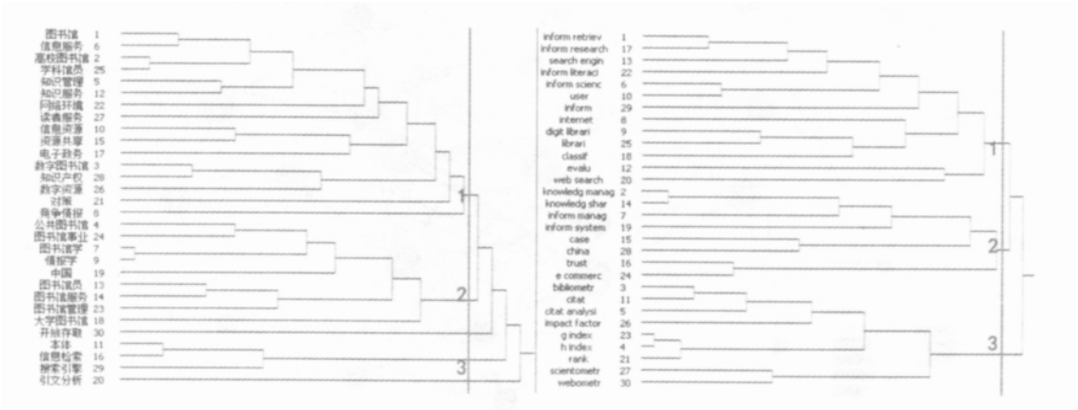

图4　国内外图书情报学高频关键词聚类树状图

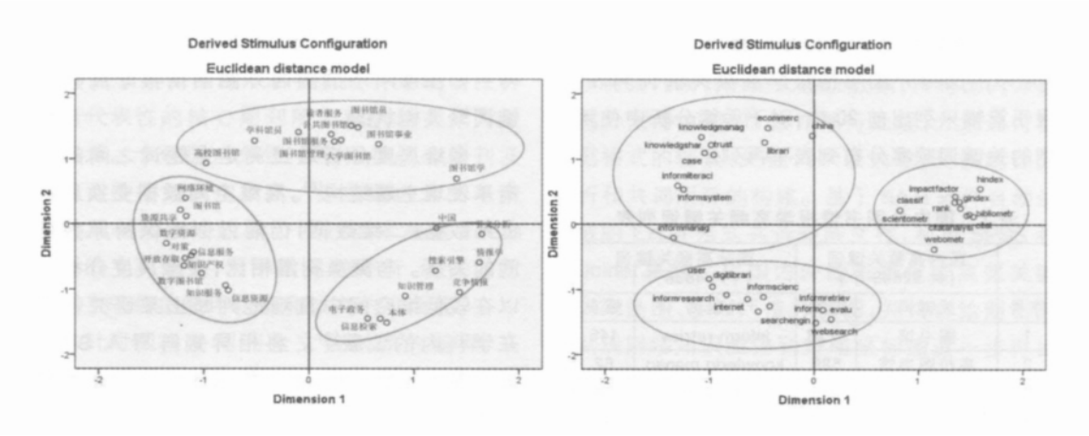

图 5　国内外图书情报学高频关键词多维尺度图谱

②信息管理研究：包括 information management、knowledge management、knowledge sharing、information literacy、library、e-ecommerce、china、trust 和 case 等 9 个关键词；③信息计量研究：包括 bibliometrics、webometrics、scientometrics、citation analysis、impact factor、rank、citation、h-index、g-index 等 9 个关键词。

3.3.3　网络知识图谱与策略坐标图

将社会网络分析方法应用到共词分析，可以探讨一个学科或主题的研究结构（核心与边缘）。社会网络分析将关键词视为结点（actor），结点位置越居中越核心，关键词之间的共现关系被视为结点间的连线（tie），连线越粗关系越强。设定 SATI 的 Rows/Cols 选项知识单元数为 30，得出国内和国际高频关键词共现相似矩阵和词频年度分布矩阵。将矩阵文件导入 Ucinet，计算各个结点的中心度（Degree Centrality）并以此为属性控制网络结点大小，形状统一选择 circle，结点颜色按照以上聚类结果进行着色和标记（R 代表资源导向研究结点，T 代表技术导向研究结点，S 代表服务导向研究结点，M 代表信息管理导向研究结点，I 代表信息计量导向研究结点），结点边缘颜色深浅度则对应于该关键词词频出现最高的年份，颜色越浅则侧重对其研究的年份越近，得到如图 6 和图 7 所示的国内外图书情报学高频关键词共现网络知识图谱。

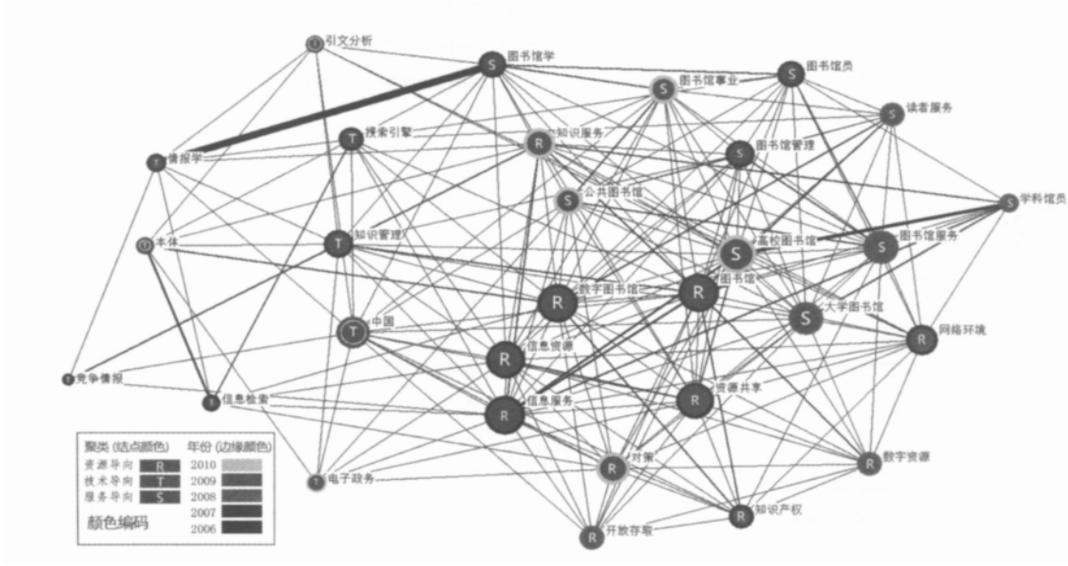

图 6　国内图书情报学高频关键词共现网络知识图谱

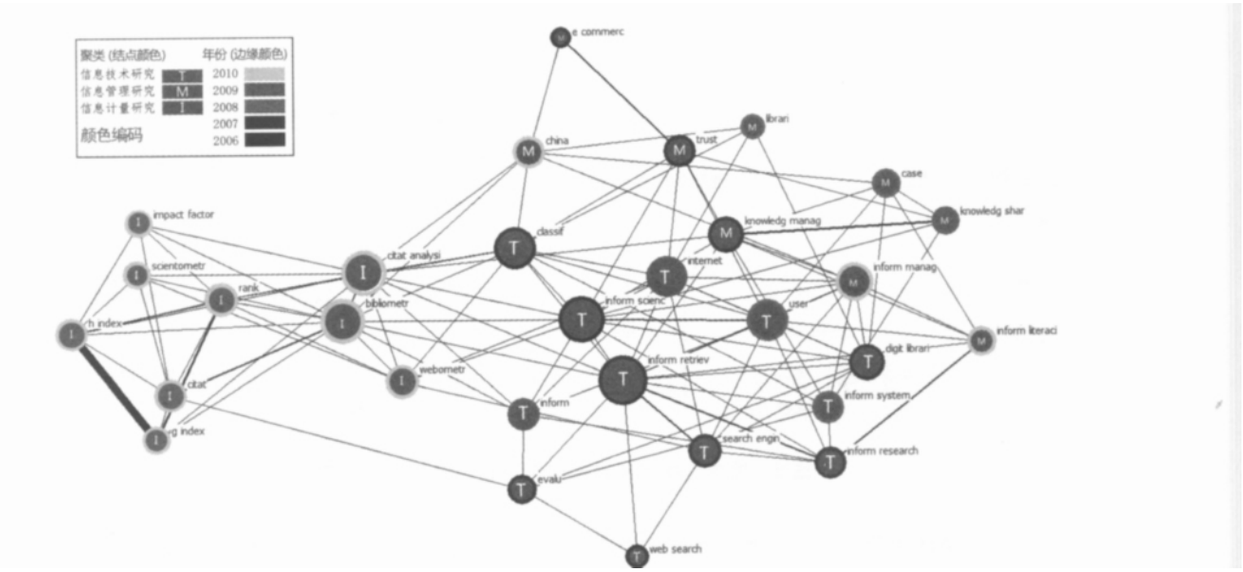

图 7　国际图书情报学高频关键词共现网络知识图谱

从图 6 可以看出，图书馆处于国内图书情报学高频关键词共现网络的中心位置，是网络中最重要的结点，说明图书馆在国内图书情报学学科的研究中占有举足轻重的地位。其中，图书馆学和情报学两个结点间的连线最粗，说明图书馆学与情报学虽为独立的两个不同学科，但两者关联性最大。观察结点颜色及标记还可得出，相对于资源导向研究（结点标记为 R，结点边缘多为深色）和技术导向研究（结点标记为 T，结点边缘多为中等深度颜色），服务导向研究（结点标记为 S，结点边缘多为浅色）近年来关注较多。

从图 7 可以看出，信息检索（Information Retrieval）处于国际图书情报学高频关键词共现网络的中心位置，是网络中核心结点。其中，结点 h-index 与 g-index 之间连线最粗，说明两者的关联性最大。从颜色编码可以得出，与信息技术研究（结点标记为 T，结点边缘颜色多为深色）和信息管理研究（结点标记为 M，结点边缘颜色多为中等深度颜色）相比，信息计量研究（结点标记为 I，结点边缘颜色多为浅色）近年来研究较多。

共现网络知识图谱展示了研究主题结构的内部关系，策略坐标图则可对各个主题领域的重要程度及其特性做出解释。策略坐标图由 Law 等人提出[2]，用于揭示各主题聚类内部的强度和类间的联系。其中横轴代表向心度（Centrality），即某研究领域在整个学科的核心程度，揭示研究领域与其他主题领域之间的关联程度，纵轴代表密度（Density），即某研究领域的内部强度，揭示某研究领域维持和发展自身的能力。基于 SATI 生成的多值共词矩阵，计算得出国内外图书情报学各三大研究领域的向心度和密度值（本文采用“总和均值法”来计算向心度和密度，即聚类向心度为类内所有结点与其他类团内所有结点的边数总和的均值，聚类密度为类内所有结点之间边数总和的均值），再以两个指标的均值作为分割线得出了如图 8 所示的策略坐标图。

基于向心度与研究领域核心程度一致，密度与研究领域成熟程度一致的思想，从图 8 可以得出，国内图书情报学的三大研究领域中，资源导向研究和服务导向研究都较为核心和成熟，即已被关注与很好的研究过，而技术导向研究尚存不足，没有受到广泛关注且自身发展不够成熟。国际图书情报学的三大研究领域中，信息技术导向的研究处于核心地位，且发展较为成熟。信息计量导向研究处于边缘位置，但已经受到关注，且发展很成熟。信息管理导向研究虽也是核心，但发展不够成熟。

4　小结

综上所述，我们开发出文献题录信息统计

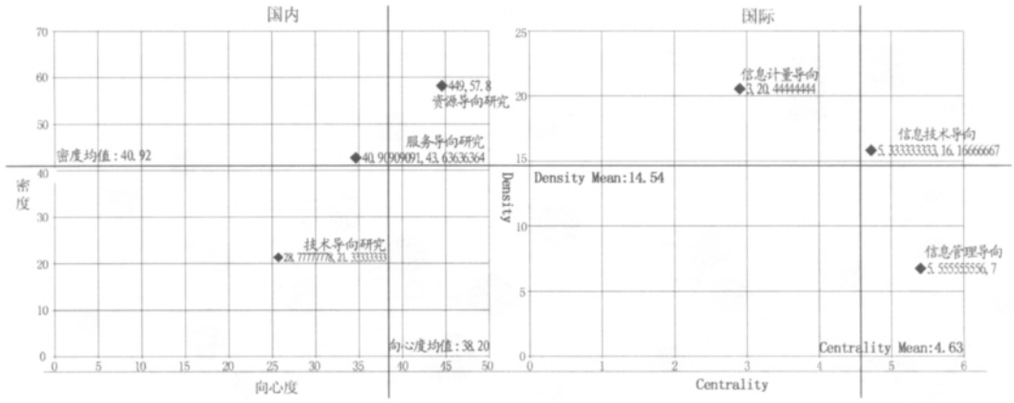

图 8　国内外图书情报学研究主题领域策略坐标图

分析工具 SATI 并用于揭示中外图书情报学研究领域的潜在结构和内在联系，虽仅以图书情报学为实例进行关键词共现分析，但技术方法和软件功能适用于各个学术领域，尤其适合在中外文献题录信息基础上挖掘出研究主题领域结构及其内部关联，并在比较中揭示其差异。期望能对信息资源管理及开发利用研究提供方法论参考。

注释

[1] 英文停用词表采自 Web of Science. http://images.webofknowledge.com/WOK46/help/WOS/ht_stopwd.html

[2] SATI 采用开源 Snowball Stemmer 的.NET 版. http://www.iveonik.com/blog/2011/08/snowball-stemmers-on-csharp-free-download/

[3] SATI 采用开源中文分词组件——盘古分词. http://pangusegment.codeplex.com/

[4] 中文停用词表基于网络检索并进行相应删减和补充。

参考文献

[1] 邱均平，杨思洛，王明芝，等. 改革开放 30 年来我国情报学研究论文内容分析[J]. 图书情报知识，2009(3)：5-17

[2] Law J, Bauin S, Courtial J, et al. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification[J]. Scientometrics, 1988, 14(3-4): 251-264

[3] Callon M, Courtial J P, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry[J]. Scientometrics, 1991, 22(1): 155-205

[4] Reuters T. Social science citation index - information science & library science - journal list[EB/OL]. [2011-03-18]. http://scientific.thomsonreuters.com/cgi-bin/jrnlst/jlresults.cgi?pc=j&sc=nu

[5] 杨颖，崔雷. 基于共词分析的学科结构可视化表达方法的探讨[J]. 现代情报，2011，31(1)：91-96

[6] 张勤，马费成. 国外知识管理研究范式——以共词分析为方法[J]. 管理科学学报，2007，10(6)：65-75

[7] Cottrill C A, Rogers E M, Mills T. Co-citation analysis of the scientific literature of innovation research traditions: Diffusion of innovations and technology transfer[J]. Knowledge, 1989, 11(2): 181-208

（收稿日期：2011-10-30）