

基于信息出版价值链的学术数据库商 分类评价方法研究*

□ 周杰 丁道劲 周锐 / 中国科学技术信息研究所 北京 100038

摘要: 文章在信息出版价值链的基础上通过构建数据库商分类评价指标体系,对进入中国市场的30家国外主要学术数据库商的价值活动数据进行了实证研究。研究中根据构建的评价指标体系分别采用了层次分类和模糊C均值分类方法来分析各种类型数据库商之间的相似程度。分类结果表明,基于信息出版价值链理论的学术数据库商分类评价指标体系能有效划分数据库商类型。

关键词: 价值链, 学术数据库商, 模糊C均值, 分类评价

DOI: 10.3772/j.issn.1673—2286.2013.05.005

1 引言

学术出版是出版业的重要组成部分,而学术出版最主要的内容是STM(科学、技术与医学)出版,大多数学术成果都以期刊文章、图书或专题论文的形式出版^[1]。学术数据库则是这些资源经过组织加工之后的集合,以智能化的信息检索方式和统一友好的检索界面,通过各种途径向广大用户提供数字化的信息服务^[2]。一般来说,学术数据库商不仅包括以数据库产品形式出版学术成果的机构,也包括所有从事学术数据库产业活动的组织。因此,除了大众所熟知的Thomson Reuters、Elsevier等大型出版集团外,EBSCO、iGroup等内容集成商和代理商也应属于学术数据库商的范畴。

数字技术的发展不仅使信息资源生产、加工、销售和使用流程中的各个环节得以数字化,它还使学术数据库商能根据自身优势对该

流程的任意环节通过拆分、重组进行运作成为可能,造成了不同的学术数据库商在内容、销售模式等方面的明显差异。因此,仅用一种评价方法无法满足对所有学术数据库商的评价需求,分类研究是全面、客观评价学术数据库商的基础。国内有学者基于该流程对数字出版的功能特点进行了相应的分类研究,将生产内容和技术平台区别开来。金永成^[3]突出了提供信息技术和服务的平台运营商在数字出版活动中的重要性,周正兵^[4]则提出应以内容集成运营商的概念完成我国传统出版业向数字出版业的转变。由此可见,分类研究多从定性角度出发,定量性研究还比较少。

而基于价值链、供应链等理论对出版行业流程进行提炼,进而展开量化研究的成果则较多。欧阳峰等^[5]从经济学角度出发,基于价值本体论,利用e3—value工具对电子图书网络出版商业模式进行建模和结构化分析,对商业模式的盈利性

和生存性进行了定量评价。国外也有学者对数字出版的上述流程进行了相关研究,Hui-Yi Ho等^[6]研究了出版商的商业模式研究对其对供应链和价值链的影响,他们分析了美国排名前5的独立数字出版商的供应链数据,将Mitchell和Cole对商业模式的定义和Magretta的商业模式三要素重新组合构建了更为详细的独立数字出版的供应链模型。X. Tian和B. Martin^[7]经过对澳大利亚14家出版企业进行实证研究,用产业链图的形式详细表达了一个普通商业出版企业的商业模式。这说明将价值链理论用于出版行业研究具有一定的可行性。

在数字时代,数据库商之间的收购、拆分日益频繁,同时部分数据库商的产品线繁复,这些因素对数据库商的分类带来困难。因此,本文将研究限定于图书馆资源建设工作的实际需求,以DRAA集团采购和NSTL范围内采购的数据库作为数据库商的主要代表,根据信息

* 本文系国家自然科学基金青年项目“面向文献相关性度量的共词和引文综合分析方法研究”(编号:11CTQ019)的研究成果之一。

出版价值链定量分类和评价我国市场的主要国外学术数据库商。

2 基于出版信息价值链的学术数据库商分类评价指标体系

2.1 信息出版价值链

价值链是用于描述产品或服务生成所需要的所有活动,包括从最初生产到终端用户使用以及使用后处理等一系列过程。这一概念在20世纪六七十年代就被用来绘制矿物出口经济的发展路径,但是直到九十年代,由于Michael E. Potter和Womack等学者的一系列论著才使价值链的概念得到广泛的应用^[8]。随着信息技术的发展,传统的价值链内涵得到进一步的延伸,信息的创造和利用过程被纳入其中。在此基础上,Jefferey F. Rayport和John J. Sviokla于1995年针对互联网提出了开发虚拟价值链的观点^[9]。在

网络环境下,Joe Peppard和Anna Rylander打破价值链的传统线性思维,在价值链的基础上引入了价值网络的概念^[10]。数字信息技术发展使企业的运作方式发生了巨大变化,价值链内涵因此得到进一步的延伸与拓展。

信息出版价值链是价值链理论在出版行业的应用,Hanno Ronte^[11]从信息出版价值链的角度分析了技术发展对于出版业的影响,并将信息出版价值链的各个环节划分为内容生产、标引、集成、检索、界面和营销,最终到达终端用户。通常情况下,出版行业供货商分成信息提供商和集成商两类。然而,信息技术的发展给信息出版价值链带来一系列改变,例如出版行业内出现了网络信息集成商等新的出版商类型,以及以需求驱动的PPV(Pay Per View)等新的销售模式。

信息出版价值链准确地反映出版商的类型随着技术的发展而发生

的变化,特别是出版商之间的差异表现明显。本文在对信息出版价值链内容生产等环节进行整合与扩展的基础上,对具有代表性的学术数据库商进行分析、归纳,最终形成对学术数据库商进行分类评价的指标体系。

2.2 顶层指标

指标设计过程中,首先根据信息出版价值链的各个环节对进入我国市场的8个典型学术数据库商的主要经营活动进行了分析,以形成顶层指标。根据信息出版价值链理论及8个学术数据库商在各个价值链环节的覆盖程度进行分析,形成了明显的3个区间,即内容生产、数据库产品生成和营销,其中数据库产品生成包括了对原始信息资源的组织加工和技术平台两个方面(如图1所示)。此外,数据库商的商业性质及其用户情况也是区分数据库商的重要因素,最终形成了内容生

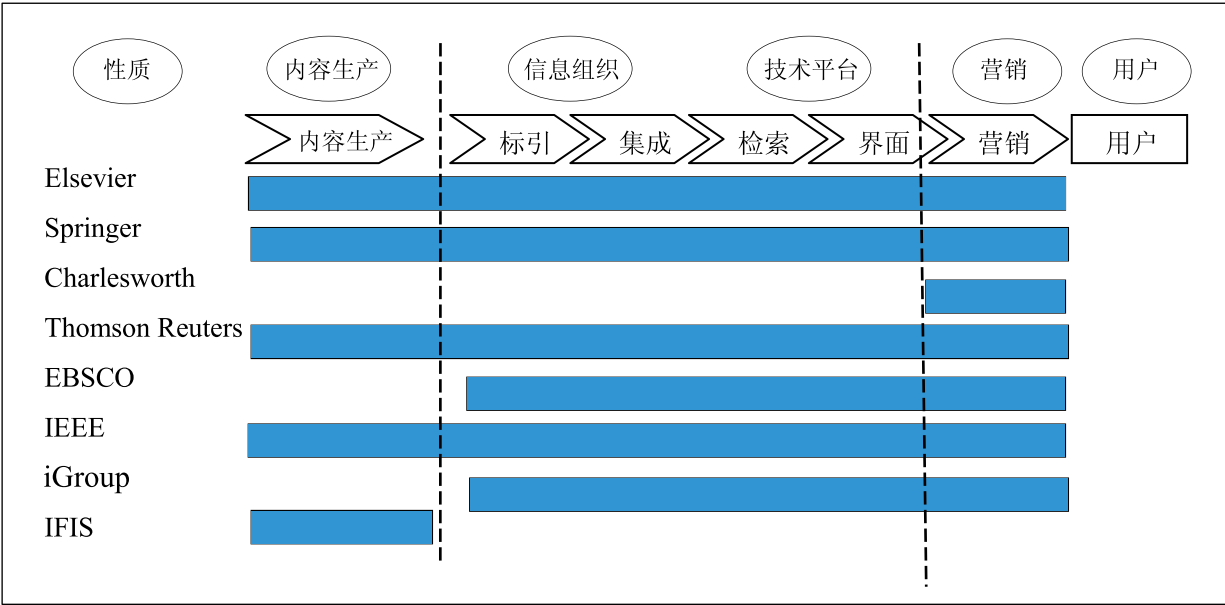


图1 典型学术数据库外商价值链分析

产、信息组织、技术平台和市场营销4个方面辅以商业性质和用户情况的学术数据库商价值活动顶层评价指标。

2.3 底层指标及其权重关系

在顶层指标的基础上,通过专家调查法,根据数据库商在各个顶层指标中所体现出来的特性,对顶层指标进一步细化后得到12个底层指标,并形成完整指标体系(见图2)。

为了确定底层指标之间的权重

关系,对上述指标体系的两层指标分别建立判断矩阵,并向来自于图书馆与数据库出版相关的专家发放调查表,调查进行了两轮,采用不记名反馈、再调查方式。对收集的调查表采用平均方法,形成最后的两层判断矩阵。然后根据判断矩阵采用层次分析法计算各层指标权重。各指标的具体内容和权重见表1。

3 学术数据库商分类方法实证研究

为了验证本文所建立的指标体系的合理性及其与我国学术数据库

的实际发展情况的一致性,进一步探索该指标体系是否存在改进之处,本文调研了主要学术数据库商的实际情况,在该指标体系的基础上进行实证分类分析。

3.1 数据来源

(1) 确定研究对象

本文从DRAA引入中国的数据库中筛选出30个主要的学术数据库商作为研究对象,其中既包括Springer等大型的商业性出版社,也包括中小型的学协会出版社,详见表2。

(2) 调研学术数据库商相关情况

在对上述30个数据库商实际情况进行整理的基础上,通过数据库商的官方网站和人工核实,并向上述10位专家发放了调查表,通过回收数据并去除异常数据,采用平均方法计算形成最终的指标数据。

3.2 分析方法与结果

为了对选定的学术数据库商进行分类研究,在指标矩阵的基础上,本文采用了层次分类和模糊C均值分类的方法对这些数据库商进行了定量分析研究。采用层次分类方法是为了观察本文提出的分类指标体系对选定数据库商的分类结果与实际情况的符合程度;采用模糊C均值分类方法来判断在类别数既定的情况下,每一样本与各个类别的隶属程度,由此可判断由指标体系确定的学术数据库商类别分布是否合理,以及指标是否具有重合性,进而加以合并,精简整个指标体系。

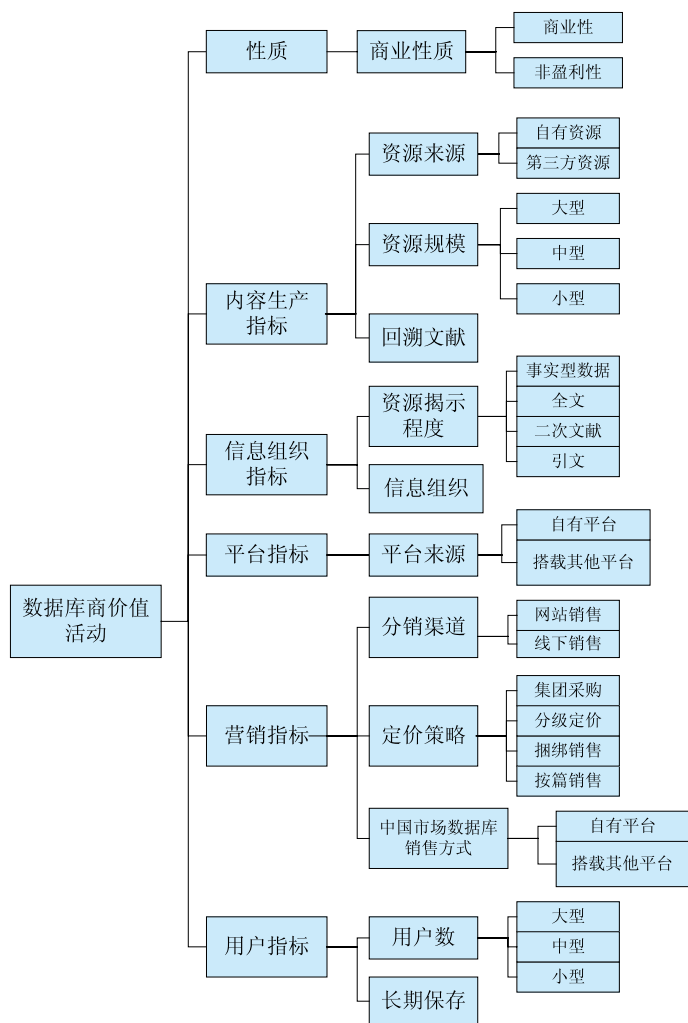


图2 数据库商价值活动指标体系

表1 底层指标内涵与权重关系

顶层指标名称	顶层权重	底层指标名称	底层权重
数据库商性质	0.05	商业性质——数据库商的经营目的是否为营利性性质	1
内容生产指标	0.3	资源来源——数据库商的原始内容资源为自有或从第三方购买	0.5
		资源规模	0.3
		回溯文献拥有情况——包含2000年之前的内容可认为拥有回溯文献	0.2
信息组织指标	0.15	信息组织——在内容上线之前是否采用叙词表等知识组织工具对内容资源进行再组织等	0.33
		资源揭示程度——指用户通过平台所获得的内容的加工程度	0.67
平台指标	0.15	平台来源——数据库平台是否为数据库商自建平台或搭载其他平台	1
营销指标	0.2	中国市场数据库销售方式——直接销售或通过国内代理公司销售	0.5
		分销渠道	0.25
		定价策略	0.25
用户指标	0.15	用户规模	0.67
		长期保存提供情况	0.33

表2 主要学术数据库外商列表

序 号	数据库商名称	序 号	数据库商名称	序 号	数据库商名称
1	Springer	11	Wiley	21	Elsevier
2	ACM	12	CUP	22	Emerald
3	IET	13	ASME	23	AIAA
4	ASCE	14	AIP	24	OECD
5	EBSCO	15	iGroup	25	Charlesworth
6	Taylor & Francis	16	Sage	26	ProQuest
7	Thomson Reuters	17	ACS	27	OUP
8	AGU	18	IWA	28	OSA
9	SPIE	19	IEEE	29	APS
10	IFIS	20	Ovid	30	SWETS

(1) 学术数据库商的层次分类

由于各指标所占权重不同,在分类过程中将指标数据(0或1与对应指标权重的乘积)进行了标准化处理,对处理后的数据进行层次分类。在算法执行中,采用类间平均法,即定义类间距离等于类间所有

样本对之间距离的平均值,采用欧氏距离作为距离计算方法,分类结果如图3。

图3的分类结果显示,30个数据库商在分成4个类的层次上,第I类包含的数据库商个数较多,如Springer等大型全文商业数据库商、CUP和OSA等大学和学协

会全文出版商、Thomson Reuters和OECD等非全文出版商均属于第I类;第II类包括IFIS非营利性二次文献数据库商;第III类包括EBSCO、iGroup等内容集成商和平台商;第IV类包括纯代理商Charlesworth。各类别的具体特征和内部差异如下:

第Ⅰ类：这一类别的共同特征是：产品内容均来自于数据库商本身，根据资源揭示程度和数据库商的性质可进一步划分为4类。其中第1小类包含了Sage等4个商业性大型全文数据库，Proquest由于同时具有第三方资源和自身资源，而Elsevier则在资源方面揭示方面包含了全文、文摘、引文等，因而它们未完全属于同一小类；第2小类为非全文大型商业出版商Thomson Reuters；第3小类为以全文产品为主的大学和学协会数据库商；第4小类为非盈利性组织OECD，其数据库产品为事实型数据。

第Ⅱ类：仅包含IFIS一个数据库商，它是一个非营利性的组织机构，其产品FSTA数据库是国际公认的食品科学和技术文献的首要数据库，搭载在第三方数据库平台中，在考察的数据库商中具有独特性。

第Ⅲ类：EBSCO、Ovid Technologies等数据库商的资源内容均来自第三方，其内部差异主要源自文献回溯年份以及平台资源的规模。iGroup和Swests均为代理商，两者具有较高的相似性，同时其资源基本上也都来源于第三方。这说明第三类中的产品主要是以第三方资源为主。

第Ⅳ类：CharlesWorth在中国市场主要侧重于数据库产品的代理销售环节，基本不涉及在其平台提供集成等访问。

经专家分析和判断，本文认为该分类结果与我国学术数据库商的实际情况基本一致，应用该指标体系对学术数据库商进行分类具有合理性。表3为该30个数据库商分为8个类别的结果，相比于4个类别的分类结果，该结果更为细致。因此，本文选择该分类结果用于模糊C值分

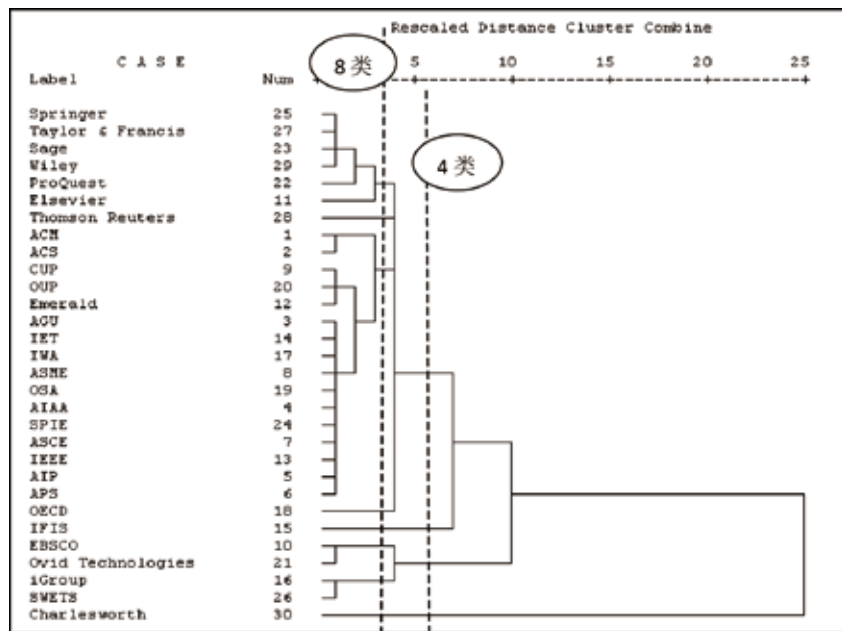


图3 层次分类结果

表3 各类别成员列表

类1	Springer; Taylor & Francis; Sage; Wiley; Proquest; Elsevier
类2	Thomson Reuters
类3	ACM; ACS; AGU; AIAA; AIP; APS; ASCE; ASME; CUP; Emerald; IEEE; IET; IWA; OSA; OUP; SPIE
类4	OECD
类5	IFIS
类6	EBSCO; Ovid Technologies
类7	iGroup; SWETS
类8	Charlesworth

类分析。

(2) 学术数据库商的模糊C均值分类

为了验证分类指标体系的准确性，研究采用模糊C均值分类方法进一步分析8类成员的内部相似程度。该方法分析的数据与采用层次分类法的数据相同，其主要思想是：不一味定义目标函数为两个向量之间的距离，而是在距离前面加入一个隶属度参数，其取值范围为[0,1]。模糊C均值方法就是要看隶属度小到什么程度，才能使目标函

数达到事先界定的取值范围。两个向量之间距离越小，越接近，隶属度就越接近1，即相似程度越大。算法的具体步骤为：

- 用[0,1]间的随机数初始化隶属矩阵；
- 计算各聚类中心；
- 计算目标函数。如果它小于某个确定的阈值，或它相对上次价值函数值的改变量小于某个阈值，则算法停止；
- 利用最小二乘原理重置初值，返回步骤b)。

表4 C均值类中心距离矩阵

类别	A	B	C	D	E	F	G	H
A	0.000	69.624	89.557	85.145	73.991	54.852	95.883	83.611
B	69.624	0.000	52.119	47.335	20.607	45.665	112.303	100.9718
C	89.557	52.119	0.0000	70.540	45.162	63.970	123.905	113.181
D	85.145	47.335	70.5406	0.000	52.217	66.683	101.465	89.0375
E	73.991	20.607	45.1629	52.217	0.000	42.351	113.700	102.745
F	54.852	45.665	63.9700	66.683	42.351	0.000	106.172	95.139
G	95.883	112.303	123.905	101.46	113.700	106.172	0.000	52.736
H	83.611	100.971	113.181	89.037	102.745	95.139	52.736	0.000

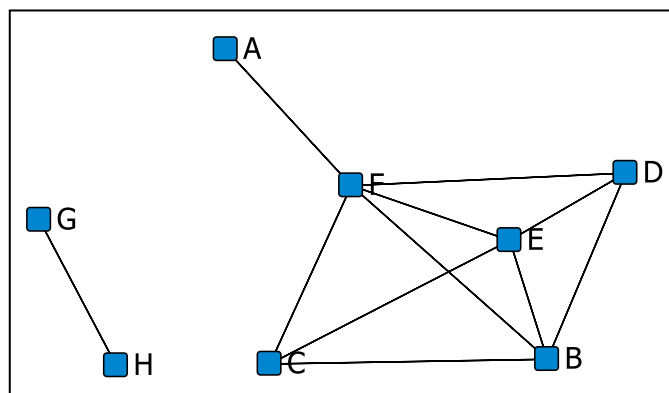


图4 C均值各类别中心关系图

前面的层次聚类中，对30个数据库商分成的8类做了分析解释，因此在隶属度分析中将C值设置为8，以观察在采用模糊C均值方法将数据库商分为8类后，各类别类中心之间的关系以及单个学术数据库商与各类别的隶属关系。

● 模糊C均值分类的类中心关系分析

当C值为8时，模糊C均值分类结果的8个类中心之间的距离矩阵（为便于观察，数值已扩大100倍，同时为与前面层次分类的结果进行区分，用A-H表示C均值方法的分类结果），具体见表4。为便于

观察，定义矩阵中的类中心距离小于65时，两个类中心比较靠近，对应结果见图4。从图中可以看到，第G、H类的中心之间的距离较近，而第A类的中心仅与第F类的中心距离较近，第B、C、D、E、F类的中心彼此都比较靠近。这说明第B、C、D、E、F类有很高的相似性，某种程度上可归结为一个大类。

● 各类别成员隶属度分析

为便于观察被模糊C均值法分为8类中各成员与各类别的隶属度，本文用雷达图来展示它们之间的关系。

图5表示的是层次分类中的第

1类成员与C均值分类中A-H各个类之间的隶属程度。从图中可以发现，第1类中的内部成员属性较为一致，它们与第A类中心隶属程度较高。这些出版商均为以全文出版为主的大型商业性数据库商。其中Sage、Springer、Taylor & Francis、Wiley与A类的隶属度均达到70%以上，Elsevier和Proquest的隶属度虽仅有30%，但仍可发现它们与其他类别具有较大差异。

图6表示的是层次分类中的第3类成员在A-H中的隶属程度。从图中可以看到第3类成员分别与B、C、D、E、F类有较高隶属度，并结合前面C均值分类中B-F这5个类的中心点极为靠近这一结论，可知第3类成员尽管分属在B-F不同的类别中，但仍然可划分为一个类，但内部成员之间可能存在一定差异；而既处于B-F分类中且属于同一类的成员应该极其相似，例如：ACM与ACS，OUP与CUP均为此种情形。

图7表示的是层次分类中的第6、7类成员与A-H类的隶属情况。从图中可以看到，这部分数据库商主要是代理商，EBSCO、Ovid Technologies与iGroup、SWESTS

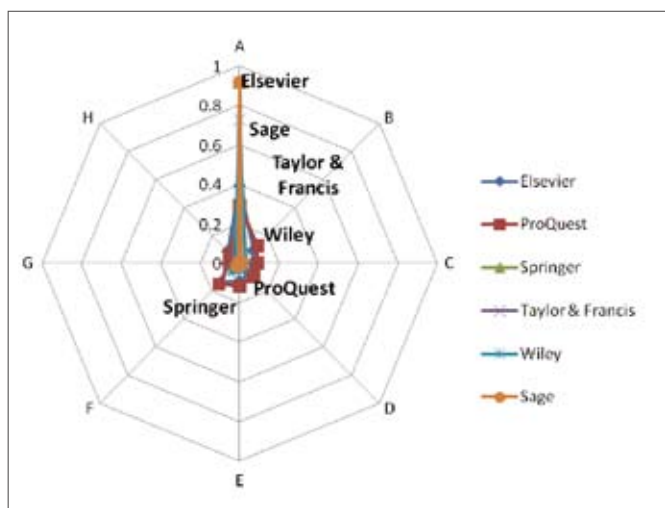


图5 第1类成员隶属度分析

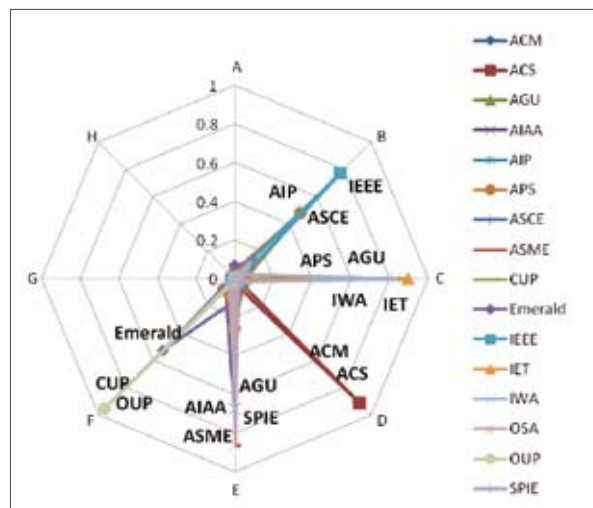


图6 第3类成员隶属度分析

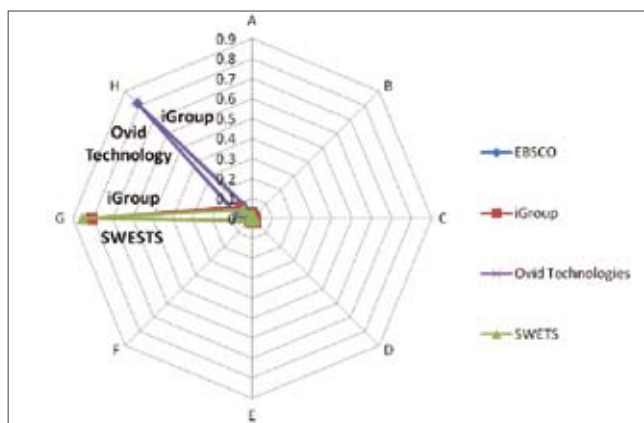


图7 第6、7类成员隶属度分析

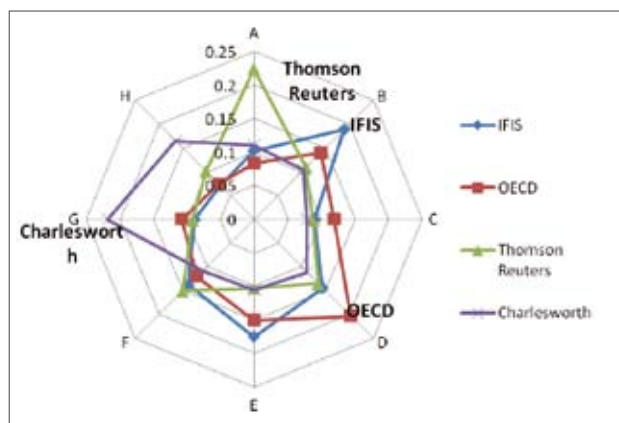


图8 第2、4、5、8类成员隶属度分析

之间有相当差异，其分别与G、H类的隶属程度达到了80%以上。

图8表示的是层次分类中第2、4、5、8类成员与A-H类的隶属情况。2、4、5、8类的成员均仅有一个，分别为Thomson Reuters、OECD、IFIS、Charlesworth。这类学术数据库商各自具有独特的属性特征，与A-H的隶属情况相对比较均衡，但仍然存在对A-H中某一类隶属程度相对较高的情况。

对各类别成员的隶属度分析结果表明，层次分类结果中的部分隶属于同一类别的成员仍具有一定的

差异性。Elsevier、Proquest是资源类型较为多样化的大型商业数据库商，这说明信息组织指标项对大型商业数据库商的分类结果影响较大；大学、学协会性质的数据库商数量较大，但在资源规模、平台来源等多重因素的影响下形成了内部差异，因此针对学协会数据库商的分类指标项与权重相比于商业性的数据库商应做出一定调整，并有可能增加未被提及的影响因素；文中构建指标体系对区分EBSCO、SWESTS等代理商效果较为明显，并可有效识别出特殊的数据库商

类型；如Thomson Reuters与大型商业全文数据库、Charlesworth与代理商的隶属度均在23%左右，但各自成类。

4 结语

本文建立的数据库商分类评价指标体系以定量方式描述进入中国市场的主要国外数据库商的特征。经验证，基于信息出版价值链理论学术数据库商分类指标体系，层次分类的结果经与领域专家确认，符合事实，可有效划分学术数据库商

类型。该分类评价指标体系是对国内图书馆资源建设中对数据库商归类的定量描述,有利于图书馆采购人员快速、准确定位新进数据库产品和数据库商,形成类似的资源采购策略,同时也为国内数据库和国

外数据库的归类比较提供了依据。

本文的局限性在于,在数据库商的分类评价中,没有从数据库商的商业模式各个层面构建指标体系,而主要从国内图书馆资源建设的视角来构建指标,同时,模糊C均

值分类的结果也显示该指标体系的部分指标可能存在部分重叠,因此还可以进一步合并简化。此外,后续使用中的版权因素对数据库商分类的影响也需进一步求证。

参考文献

- [1] 刘银娣,唐敏珊.欧美大型学术出版机构营销战略研究[M].广州:华南理工大学出版社,2012:5.
- [2] 肖强.商业数字图书馆信息资源建设服务模式[M].北京:中国出版集团,2012:39.
- [3] 金永成.B-B-C:数字出版商业模式的创新与启示[J].出版发行研究,2011(3):46-49.
- [4] 周正兵.我国传统出版单位内容集成运营模式研究[J].出版发行研究,2011(4):24-28.
- [5] 欧阳峰,赵红丹,郑广录.基于价值本体论的网络出版商业模式之模拟分析[J].科技管理研究,2008(6):232-234.
- [6] HO H Y, WANG L W, CHENG H J. Authors, Publishers, and Readers in Publishing Supply Chain: The Contingency Model of Digital Contents Production, Distribution, and Consumption [J]. Systems Engineering Procedia, 2011(2): 398-405.
- [7] TIAN X M, MARTIN B. Business Model in Digital Book Publishing: Some Insights from Australia [J]. Publishing Research Quarterly, 2009 (25): 73-78.
- [8] KAPLINSKY R. Globalisation and Unequalisation: What Can Be Learned from Value Chain Analysis? [J]. Journal of Development Studies, 2000, 37(2): 117-146.
- [9] 迟晓英,宣国良.价值链研究发展综述[J].外国经济与管理,2000,22(1):25-30.
- [10] PEPPARD J, RYLANDER A. From Value Chain to Value Network: Insights for Mobile Operators [J]. European Management Journal, 2006, 24(2-3):128-141.
- [11] RONTE H. The Impact of Technology on Publishing [J]. Publishing Research Quarterly, 2001, 16(4): 11-22.

作者简介

周杰 (1965-), 男, 副研究馆员, 研究方向: 信息资源管理。E-mail: zhouj@istic.ac.cn

The Classification and Evaluation of Foreign Academic Database Providers Based on Information Publishing Value Chain

Zhou Jie, Ding Qiujiing, Zhou Rui / Institute of Scientific and Technological Information of China, Beijing, 100038

Abstract: An indicator system based on the theory of information publishing value chain is articulated in this paper. 30 foreign academic database providers' value activities are clustered first and some FCM classifying analysis is conducted to know every case's membership degree. The results give the conclusion that the combination of the indicator system and analysis method can make off the academic database providers effectively and quantify their differences.

Keywords: Value chain, Academic database providers, FCM, Classification and evaluation

(收稿日期: 2013-03-25)