

h 指数应用于专利影响力评价的探讨*

Research on the Application of h-Index in the Assessment of Patent

冯 君

(南京工业大学图书馆 南京 210009)

摘 要 以陶瓷膜分离领域的专利文献为研究对象,以“德温特世界专利创新索引”为数据源,探讨 h 指数在专利影响力评价中的应用。首先,利用 SPSS 分析软件研究 h 指数应用于专利影响力评价的有效性。在此基础上,选取 CHI 专利指标中用于衡量专利影响力的 CII 指数、TS 指数为参考,采用对比分析的方法研究 h 指数在专利影响力评价中的优势及存在问题。

关键词 h 指数 技术实力指数 当前影响力指数 专利分析

中图分类号 G306

文献标识码 A

文章编号 1002-1965(2009)12-0016-05

专利是集科技信息、经济信息、法律信息于一体的信息载体,是世界上反映科学技术发展最迅速、最全面、最系统的信息资源。国际上,一般通过专利的申请量与授权状况评价一个国家、机构或个人的技术创新能力,通过多角度的专利信息分析来揭示各大科研机构或公司在知识产权市场的竞争态势。在专利信息分析与评价的过程中,专利指标作为发明创造活动的“尺度”起着重要的作用,专利指标的研究一直受到研究者的关注。

众多专利指标中,专利的申请数量及其被引用情况是两个最基本的指标,也是评价专利影响力的两个重要指标。一般来说,一个机构在某研究领域内申请的专利数量越多,所申请专利被其他机构引用的次数越多,说明它在该领域内的影响力越大。赵亚娟等人^[1]以储氢领域为研究对象,通过统计专利的申请数量分析得到该领域最具有优势和影响力的专利权人。杨祖国等人^[2]通过统计分析中国专利的被引用情况,研究中国专利技术的重要性和影响力。实际上,单纯从专利数量或专利的引用情况进行专利影响力评价都具有一定的片面性。因此,以专利数量及其引证指标为基础,派生出一系列用于评价专利影响力的复合指标,最有代表性的当属美国著名的知识产权研究机构 CHI Research 提出的当前影响力指数(Current Impact Index, CII)和技术实力指数(Technology Strength, TS),这是目前国内较为常用的评价企业、国家或组

织专利影响力的指标^[3-7],美国国家科学基金会编写出版的《美国科学与工程指标》也采用了 CHI 的这套专利指标。在我国,杨中楷^[7]等人通过统计 1982~1996 年间美国专利商标局授予专利的 TS 指数,分析中国、美国、德国等国家在先进材料、汽车、信息技术等行业的技术实力。但此类复合指标在计算的过程中,受数据库、研究领域划分等条件的限制,数据的获取比较困难。

2005 年,美国加州大学圣地亚哥分校的物理学家 Jorge E. Hirsch 教授提出一个新的评价指标——h 指数,最初被应用于评价学者的学术成就^[8],该指数最大的优势在于它是一个兼顾论文发表数量和质量复合指标,目前已被广泛地应用于期刊^[9-10]及对研究机构的评价^[11-12]。近年来,一些学者开始探索基于专利申请和引用的 h 指数^[13-15],他们选择世界百强企业为研究对象,探讨专利权人 h 指数的计算方法,及其与单因素指标——专利申请数、专利总引用次数以及百强企业营业收入之间的相关性。已有研究更多的侧重于将 h 指数拓展到专利领域,缺乏与现有专利评价指标的对比分析。

本文以“德温特世界专利创新索引”为数据源,以陶瓷膜分离领域相关专利文献为研究对象,利用 SPSS 分析软件研究 h 指数应用于专利影响力评价的有效性。在此基础上,选取 CHI 专利指标中用于衡量专利影响力的 CII 指数、TS 指数为参考,采用对比分析的

收稿日期:2009-09-03

修回日期:2009-10-26

基金项目:南京工业大学 2008 年人文社科与管理学科群青年教师科研基金“基于专利情报分析的科技创新评价研究”(编号:39107103)研究成果之一。

作者简介:冯 君(1980-),女,博士,馆员,研究方向为学科情报、信息管理。

方法研究 h 指数在专利影响力评价中的优势及存在问题。

1 指标概述

1.1 专利权人 h 指数 借鉴 Hirsch 提出的用于评价科学家学术成就的 h 指数的概念,将专利权人的 h 指数定义为:专利权人所申请的 N 项专利中至少被引用了 h 次的 h 件专利数。

专利权人 h 指数的计算方法如下:在带有引文信息的专利数据库中,如德温特世界专利创新索引(Derwent Innovations Index)等,检索某一专利权人的所有专利,按照被引频次的降序排列,当专利序号大于等于该论文被引用次数时,这个对应的专利序号就是该专利权人的 h 指数,如表 1 所示。h 指数是一个累积指标,不会随时间的推移上下波动,只会逐渐增大或保持不变。

表 1 专利权人 h 指数的计算方法

专利	按照被引频次排序	被引用次数
A	1	5
B	2	4
C	3	3
D	4	2
E	5	1

1.2 当前影响力指数 (Current Impact Index, CII)

当前影响力指数(以下简称 CII),是基于研究机构过去 5 年内申请专利以及被引用情况测算其专利影响力的指标,具体的计算方法如下^[3-4]:

$$CII = \frac{C_{ij}/K_{ij}}{\sum_i C_{ij}/\sum_i K_{ij}} \quad (1)$$

其中: C_{ij} 代表研究机构(i)今年以前连续 5 年授权专利被引用的总次数; K_{ij} 代表研究机构(i)今年以前连续 5 年的授权专利数量; $\sum_i C_{ij}$ 代表该研究机构所在行业(j)今年以前连续 5 年授权所有专利的总引用次数; $\sum_i K_{ij}$ 代表该研究机构所在行业(j)今年以前连续 5 年的授权专利总量。

CII 是一个相对指标,代表一个研究机构在过去 5 年时间里,专利平均被引用情况与同期同技术领域专利平均被引用情况的相对高低情况,CII 值越大,代表在相应领域内的影响越大。CII 值会随着时间的变化而变化,当 CII 值开始变小的时候,说明该研究机构的研究领域发生了调整。

1.3 技术实力指数 (Technology Strength, TS) 技术实力指数(以下简称 TS),这是由 CII 引申出来的指标,具体的计算方法如下^[3-4]:

$$TS = \text{企业申请专利总数} \times CII \quad (2)$$

TS 指数是专利数量在质量方面的加权,用于评估某专利权人的专利技术组合力量。TS 值越大表示专利权人在相应领域内的实力越强。

2 数据来源和研究对象

2.1 数据来源 德温特世界专利创新索引^[16](Derwent Innovations Index,以下简称 DII)是由美国汤姆森路透集团(Thomson Reuters)下属的汤姆森科技信息集团(Thomson Scientific)推出的基于网络的专利数据库,它将德温特世界专利索引(Derwent World Patents Index[®],简称 WPI)和专利引文索引(Patents Citation Index[®])有机地整合在一起,用户不仅可以通过它检索专利信息,而且可以检索到专利的引用情况。DII 数据库收录了来自 40 多个专利机构授权的 1460 万项基本发明,3000 多万条专利,每周更新并回溯至 1963 年,是目前全球最具规模与影响力的专利检索数据库之一。

与免费访问的网络专利数据库相比,Thomson Scientific 组织 350 多名专家,对来自全球 41 个国家和地区的专利进行深加工,去除了专利说明书以及权利要求书中晦涩难懂的技术性词汇,利用通用的技术词汇重新对标题和文摘进行了改写,并增加了经过规范化的公司代码,Derwent 分类和 Derwent 手工代码等标引字段,大大提高了专利检索的查全、查准率。

2.2 研究对象 陶瓷膜分离技术^[17]是 20 世纪开发成功的新型、高效、环保的分离技术,适合现代工业对节能和环保的要求,该技术出现以来发展十分迅速,相继涌现出大量专利,应用也越来越广泛,目前在石油冶金工业、气体分离与净化、环境工程、制药等领域都得到了较好的应用,为此本文选择陶瓷膜分离领域为研究对象。根据膜材质来分,陶瓷膜主要有氧化硅膜、氧化铝膜、氧化锆膜和氧化钛膜四种。陶瓷膜以其优异的材料性能在很多苛刻的应用体系中(较高温、强酸碱环境和有机溶剂)表现出有机膜无可比拟的优势。

在 DII 数据库中综合运用“主题”和“德温特手工分类代码”构造检索式为“Derwent Manual Code=(A12 - E06B OR A12 - W11A OR D04 - A01D OR D04 - A01E OR H02 - 04 OR J01 - C03 OR J01 - C03A OR J01 - C03B OR J01 - C03B1 OR J01 - E02C OR J01 - E02C1 OR J01 - E03E OR L03 - E04B2 OR N06 - C03 OR S03 - E03C OR U11 - C18C OR V03 - C01A1A OR V03 - C01A2A OR X16 - E06A5C OR X16 - F02 OR A12 - E06B OR A12 - W11A OR D04 - A01D OR D04 - A01E OR H02 - D04 OR J01 - C03 OR J01 - C03A OR J01 - C03B OR J01 - C03B1 OR J01 - E02C OR J01

- E02C OR J01 - E02C1 OR J01 - E03E OR L03 - E04E2 OR N06 - C03 OR S03 - E03C OR U11 - C18C OR V03 - C01A1A OR V03 - C01A2A OR X16 - E06A5C OR X16 - F02) AND Topic = (Ceramic * or Alum * or Zirco * or titan * or silic *)", 检索得到自 1963 至 2009-07-24 有关陶瓷膜分离的所有专利共计 13159 条, 本文选取其中专利申请数量排名前 20 位的专利权人为研究对象。

3 数据统计与分析

3.1 统计结果 选取“陶瓷膜分离”领域专利申请数量排名前 20 位的专利权人, 统计其从 1963 至 2009-07-24 期间申请的专利总数、被引用总次数, 并计算它们各自的平均引用次数、h 指数、CII 指数以及 TS 指数, 结果见表 2:

表 2 “陶瓷膜分离领域”专利申请数量排名前 20 名的专利权人

专利权人代码	专利数	总引用数	平均引用次数	h 指数	CII	TS
MATU-C	276	404	1.46	10	0.744	16.360
BOSC-C	250	1026	4.10	19	1.223	28.127
SMSU-C	227	258	1.14	7	0.896	31.363
NPDE-C	170	825	4.85	15	1.259	16.373
NIGA-C	164	683	4.16	12	0.741	2.965
TOYT-C	116	105	0.91	5	0.261	3.920
KYOC-C	113	100	0.88	5	0.306	1.224
TOKE-C	112	230	2.05	8	1.089	15.246
ASAH-C	110	318	2.89	9	0.296	0.593
NITS-C	104	189	1.82	8	0.264	3.562
SHIH-C	104	208	2.00	5	0.396	1.321
HITA-C	98	295	3.01	8	0.828	1.656
NITL-C	86	179	2.08	7	0.256	0.000
NIOD-C	84	123	1.46	5	0.096	0.384
GENE-C	79	206	2.61	8	0.932	0.000
MITO-C	78	189	2.42	7	0.384	2.306
SONY-C	77	72	0.94	4	0.283	0.000
AGEN-C	75	311	4.15	10	0.000	0.000
MATW-C	72	51	0.71	4	0.409	0.818
FUJD-C	67	58	0.87	3	0.000	0.000

注: 专利权人代码是经过规范化的公司或高校代码。DII 对全球 22000 家大公司标识了一个唯一的四位代码, 即专利权人代码, 这个代码就包含了该公司所有的子公司专利。

分别根据专利权人 h 指数和 TS 指数的大小, 对表 2 中的 20 个专利权人进行重新排序, 结果见图 1。

3.2 h 指数与现有指标的相关性分析 应用 SPSS 软件分别研究复合指标专利权人 h 指数、CII 指数、TS 指数与专利申请数、专利总引用次数以及专利平均引用次数间的相关性关系, 结果见表 3。

从表 3 可以看出, CII 指数、TS 指数仅和专利申请数、总引用次数间具有显著相关性, 与平均引用次数间

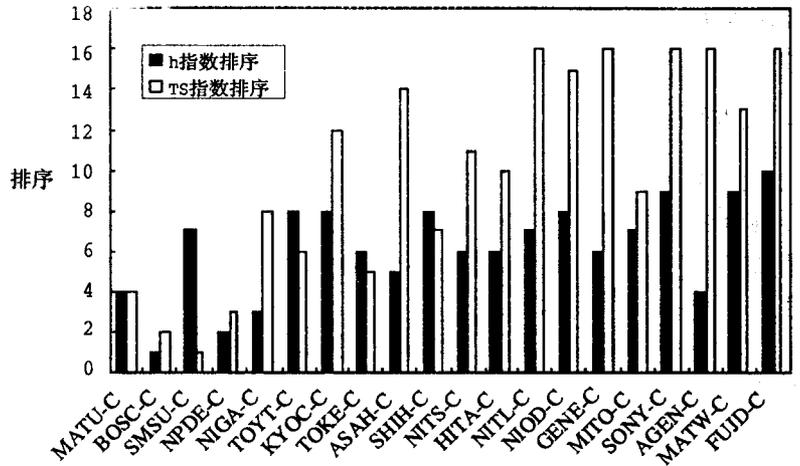


图 1 专利权人 h 指数与 TS 指数的排名对比

不具有相关性。而专利权人 h 指数与专利申请数、总引用次数以及平均引用次数间都具有显著正相关性, 其 Pearson 系数分别为: 0.642、0.966、0.839, 尤其是 h 指数和总引用次数以及平均引用次数间具有高度相关性, 显著性检验均为 0.000 (sig. (2-tailed)), 即专利权人申请的专利数量、总引用次数、平均引用次数均影响专利权人 h 指数的大小。同时, h 指数和 CII 指数、TS 指数间也具有显著相关性, Pearson 系数分别为 0.578 和 0.681。

通过 h 指数与现有指标, 如专利申请数、总引用次数、平均引用次数、TS 指数等的相关性分析看出, 专利权人 h 指数是兼顾专利申请数及其被引用次数的复合指标, 可以作为专利影响力的评价指标。

3.3 h 指数用于专利影响力评价的优势 对比表 2、图 1 中专利权人的 h 指数、TS 指数以及相应的排名, 可以看出 h 指数在专利影响力评价的过程中具有以下优势:

a. 相对于其他复合指标, 如当前影响力指标 (CII)、技术实力指标 (TS) 等, h 指数十分简单且易于确定。只要将专利权人的所有专利按照被引频次的高低排序, 就很容易得到相应的 h 指数, 他的提出者——Hirsch 称只需花 30 秒钟。而当前影响力指标 (CII)、技术实力指标 (TS) 的计算过程中, 需要计算专利权人所申请专利的总引用次数、相应研究领域的专利总数及总引用次数, 其中大量的统计工作需要人工完成, 要耗费大量的时间和人力。

b. h 指数能给予专利权人更有效、全面的评价。h 指数与其他“有意义”的专利指标, 如专利申请数、总引用次数、平均引用次数、CII 指数和 TS 指数均具有显著的正相关性。而 CII 指数和 TS 指数只与专利申请数、总引用次数间具有显著相关性, 与平均引用次数间不具有相关性。

表 3 专利权人 h 指数、CII 指数、TS 指数的相关性分析

		Correlations					
		h 指数	TS	CII	专利数	总引用次数	平均引用次数
h 指数	Pearson Correlation						
	Sig. (2-tailed)						
	N						
TS	Pearson Correlation	.578**					
	Sig. (2-tailed)	.008					
	N	20					
CII	Pearson Correlation	.681**	.716**				
	Sig. (2-tailed)	.001	.000				
	N	20	20				
专利数	Pearson Correlation	.642**	.847**	.624**			
	Sig. (2-tailed)	.002	.000	.003			
	N	20	20	20			
总引用次数	Pearson Correlation	.966**	.589**	.692**	.676**		
	Sig. (2-tailed)	.000	.006	.001	.001		
	N	20	20	20	20		
平均引用次数	Pearson Correlation	.839**	.187**	.471**	.232**	.813**	
	Sig. (2-tailed)	.000	.429	.036	.326	.000	
	N	20	20	20	20	20	

** . Correlation is significant at the 0.01 level(2-tailed).

* . Correlation is significant at the 0.05 level(2-tailed).

c. h 指数是一个累计指标,有利于表现专利权人在相应领域内形成的持久影响力,避免了 TS 指数、CII 指数只反映专利权人最近 5 年内影响力的缺陷。例如,澳大利亚的公司 AGENCY OF IND SCI & TECHNOLOGY(专利权人代码 AGEN-C)由于在过去 5 年内没有在“陶瓷膜分离领域”申请相关专利,所以其 CII 值为 0,TS 指数为 0,根据 TS 指数的排名为 16,但基于 AGEN 以前申请专利的影响力,它的 h 指数为 10,h 指数的排名为 5。

3.4 h 指数用于专利评价的缺陷 区分度是指指标对评价对象的区分程度,是反映指标有效性的参数之一。由表 2 可见,所研究的 20 个专利权人中,h 指数的跨度为 3~19,其中德国 BOSCH 公司(专利权人代码为 BOSC-C)的 h 指数最高,为 19;h 指数为 10 的专利权人有 2 家;h 指数为 8 的有 4 家;h 指数为 7 的有 3 家;h 指数为 5 的有 4 家;h 指数为 4 的有 2 家。

以 h 指数同为 8 的四个专利权人为例,进一步比较 h 指数相同的专利权人的其他指标:专利数、总引用数、CII 值以及 TS 等,如表 4 所示:

表 4 h 指数为 8 的四个专利权人

专利权人代码	专利数	总引用数	平均引用次数	CII	TS
TOKE-C	112	230	2.05	1.089	15.246
NITS-C	104	189	1.82	0.264	3.562
HITA-C	98	295	3.01	0.828	1.656
GENE-C	79	206	2.61	0.932	0.000

从表 4 可以看出,虽然这四个专利权人的 h 指数相同,但其他指标却大不相同,且具有较大的跨度,专利数介于 79~112 之间,总引用数介于 189~295 之

间,平均引用次数介于 1.82~3.01 之间,CII 介于 0.264~1.089 之间,TS 介于 27.481~121.965 之间。可见,基于专利文献的 h 指数存在区分度不高的缺陷。

金碧辉^[18]提出的 R 指数能解决 h 指数存在的区分度不高的问题。R 指数是指对 h 指数起贡献作用的那部分论文总被引频次的平方根,具体的计算公式如下:

$$R = \sqrt{\sum_{j=1}^h cit_j} \quad (3)$$

其中, cit_j 表示对 h 指数起贡献作用的第 j 篇论文的被引用次数, cit_j 至少大于或等于 h。

借鉴金碧辉提出的 R 指数的计算方法,对具有相同 h 指数的专利权人进行区分。根据式(3)计算得到表 4 中四个专利权人的 R 指数分别为:12.207(TOKE-C);10.247(NITS-C);13.379(HITA-C);11.180(GENE-C)。在不改变专利权人 h 指数的前提下,R 指数通过对所有对 h 指数起贡献作用的专利被引频次总量平方根的计算,能对具有同样 h 指数的专利权人进行区分,对 h 指数起到了较好的补充作用。

4 结 语

h 指数的引入将会为专利评价提供新的视角。相对于其他复合指标,如当前影响力指标(CII)、技术实力指标(TS)等,h 指数十分简单且易于确定,有利于表现专利权人的持久影响力,但也具有区分度不高的缺陷,需要借助一些补充指标,如 A 指数、R 指数等。总之,在众多的专利评价指标中,h 指数作为一个新的评

价指标,有其自身的优势,也有一定的局限性,在专利分析的过程中不能孤立使用,要注意与其他技术指标的组合使用。

参 考 文 献

- [1] 赵亚娟,董 瑜,朱相丽. 专利分析及其在情报研究中的应用[J]. 图书情报工作, 2006, 50(5): 19-22
- [2] 杨祖国,李文兰. 中国专利引文分析研究[J]. 情报科学, 2005, 23(5): 700-703
- [3] 陈 燕,黄迎燕,方建国. 专利信息采集与分析[M]. 北京: 清华大学出版社, 2006
- [4] Dar-Zen Chen, Wen-Yan Cathy Lin, Mu-Hsuan Huang. Using Essential Patent Index and Essential Technological Strength to Evaluate Industrial Technological Innovation Competitiveness[J]. Scientometrics, 2007, 71(1): 101-116
- [5] 2003-2004 MIT Patent Scorecards Show Gains in Technology Strength By ATP-Funded Biotechnology Companies[EB/OL]. 2009-08-01. <http://www.atp.nist.gov/factsheets/1-h-2.htm>
- [6] 杜邦公司专利影响力仍居榜首[EB/OL]. 2009-08-01. <http://210.82.89.169/web/info/2009/0424/43.html>
- [7] 杨中楷,孙玉涛. 基于专利引用的国家技术力量指标比较[J]. 科学学与科学技术管理, 2005(10): 11-14
- [8] J E Hirsch. An Index to Quantify an Individual's Scientific Research Output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572
- [9] 郑惠伶. 运用 h-指数评价期刊影响力—以图书馆情报学期刊为例[J]. 情报科学, 2008, 26(3): 409-413
- [10] Andras Schubert, Wolfgang Glanzel. A Systematic Analysis of Hirsch-Type Indices for Journals[J]. Journal of Informetrics, 2007(1): 179-184
- [11] 蒋 新. 基于 h 指数: 课题组研究力评估的量化指标[J]. 图书情报工作, 2009, 53(4): 50-52
- [12] Jean-Francois Molinari, Alain Molinari. A New Methodology for Ranking Scientific Institutions[J]. Scientometrics, 2008, 75(1): 163-174
- [13] Jian Cheng Guan, Xia Gao. Exploring the H-Index at Patent Level[J]. Journal of the American Society for Information Science and Technology, 2008, 59(13): 1-6
- [14] 官建成,高 霞,徐念龙. 运用 h-指数评价专利质量与国际比较[J]. 科学学研究, 2008, 26(5): 932-937
- [15] 次仁拉珍,乐思诗,叶 鹰. 世界百强企业 h 指数探析[J]. 大学图书馆学报, 2009(2): 76-79
- [16] Derwent World Patents IndexSM[EB]. 2009-07-30. <http://www.thomsonscientific.com.cn/products/patent/dwpi-ji.htm>
- [17] 徐南平. 面向应用过程的陶瓷膜材料设计、制备与应用[M]. 北京: 科学出版社, 2005
- [18] 金碧辉, Rousseau Ronald. R 指数、AR 指数、h 指数功能扩展的补充指标[J]. 科学观察, 2007, 2(3): 1-8
- (责编:刘影梅)
-
- (上接第 10 页)
- A Review[A]. Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences[C], 1998. Volume 5: 314
- [26] Ron Weber. Conceptual Modelling and Ontology: Possibilities and Pitfalls[J]. Journal of Database Management, 2003, 14(3): 1
- [27] Irwin G, D Turk. An Ontological Analysis of Use Case Modeling Grammar[J]. Journal of the Association for Information Systems, 2005, 6(1): 1-36
- [28] Yair Wand, V C Storey, Ron Weber. An Ontological Analysis of the Relationship Construct in Conceptual Modeling[J]. ACM Transactions on Database Systems, 1999, 24(4): 494-528
- [29] Brian Dobing, J Parsons. How UML is used[J]. Communications of the ACM, 2006, 49(5): 109-113
- [30] Poels G, et al. Experimental Research on Conceptual Modeling: What Should We Be Doing and Why[J]. ER2006, Lecture Notes in Computer Science 4215: 544-547
- [31] Aymeric D, A A Benoit, P Michel. An Evaluation of Inter-Organizational Workflow Modeling Formalisms[J]. Journal of Database Management, 2004, 15(2): 74
- [32] P Green, M Rosemann, M Indulska. Ontological Evaluation of Enterprise Systems Interoperability Using EbXML[J]. IEEE Transactions on Knowledge and Data Engineering, 2005. Volume 17(Issue 5): p. Page(s): 713-725
- [33] Siau K, J Erickson, L Y Lee. Theoretical vs. Practical Complexity: The Case of UML[J]. Journal of Database Management, 2005, 16(3): 40
- [34] Davies I, et al. How do Practitioners Use Conceptual Modeling in Practice? [J]. Data & Knowledge Engineering, 2006, 58(3): 358-380
- [35] Dobing B, J Parsons. Dimensions of UML Diagram Use: A Survey of Practitioners[J]. Journal of Database Management, 2008, 19(1): 1
- [36] Agarwal R, et al. On the Usability of Representations[J]. Communications of the ACM, 2000, 43(10): 83-89
- [37] Weber R. Are Attributes Entities? A Study of Database Designers' Memory Structures[J]. Information Systems Research, 1996, 7(2): 137-162
- [38] Briand L C, Labiche Y, Di Penta M, et al. An Experimental Investigation of Formality in UML-based Development[J]. IEEE Transactions on Software Engineering, 2005, 31(10): 833-849
- [39] Andrew Gemino, Y Wand. Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties[J]. Data & Knowledge Engineering, 2005, 55: 301-326
- [40] I Vessey. Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature[J]. Decision Sciences, 1991, 22(2): 219-240
- [41] Mayer R E. Models for Understanding[J]. Review of Educational Research, 1989, 59(1): 43-64
- (责编:贺晓利)