

图书情报
档案管理

新书
创丛

成颖著

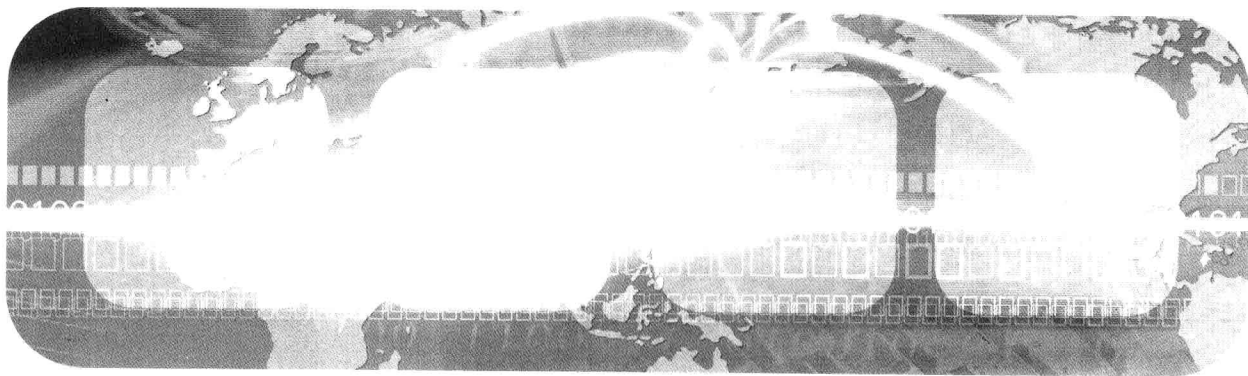
信息检索相关性判据 及应用研究



科学出版社

图书情报与档案管理创新丛书

信息检索相关性判据 及应用研究



成 颖 著

科学出版社

北 京

内 容 简 介

本书通过对 194 名研究对象相关性判据文本的内容分析,解析出了包括传播特征、内容、情境、使用、系统特征、愉悦、质量、总体与文献特征 9 类相关性判据;解析出了包括参考文献、出版单位、出版时间、关键词、基金、篇幅、全文、题名、文献格式、文摘、语种、文献类型、来源期刊、作者和机构 15 个影响信息用户相关性判断的文献特征;研究了任务复杂性以及性别对相关性判据选择的影响。同时,基于相关性判据、价值增值模型、TEDS 模型与信息系统成功模型构建了面向相关性判据的学术信息检索系统成功模型,并证实了模型的有效性。

本书可以作为图书馆学、情报学和信息管理与信息系统专业高年级本科生和研究生的参考用书。

图书在版编目(CIP)数据

信息检索相关性判据及应用研究 / 成颖著. —北京:科学出版社, 2011
(图书情报与档案管理创新丛书)

ISBN 978-7-03-032328-6

I. 信… II. 成… III. 情报检索—研究 IV. G252.7

中国版本图书馆 CIP 数据核字 (2011) 第 185331 号

责任编辑:李 敏 赵 鹏 王晓光 / 责任校对:陈玉凤

责任印制:钱玉芬 / 封面设计:王 浩

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京彩虹伟业印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2011 年 9 月第 一 版 开本:787×1092 1/16

2011 年 9 月第一次印刷 印张:16 1/4 插页:2

印数:1—2 000 字数:362 000

定价:76.00 元

(如有印装质量问题,我社负责调换)

总 序

图书情报与档案管理作为独立的一级学科，如何在激烈的竞争环境中生存、扬弃、发展、创新，探索出一条既符合学科发展规律，又与社会、经济、科技和文化发展与时俱进的学科发展道路，是一代又一代图书情报与档案管理人光荣而神圣的使命。

南京大学信息管理系作为我国图书情报与档案管理学科的重要教学和研究阵地，从20世纪20年代创建伊始就一直以培养人才、创新科研、服务社会为历史使命。80多年来，已毕业的6000多名南京大学信息管理系学子遍及国内外图书情报服务机构，以及文教、传播等其他行业。

图书情报与档案管理学科的发展承载着先辈的学术寄托，从创建图书馆学科肇始，学科先辈就在为争取独立学科地位、构建自身核心理论体系及扩大教育规模而努力。图书情报与档案管理获得一级学科地位后，仍然面临“大学科观”与“小学科观”、“图书馆业务中心论”与“情报分析与服务中心论”等不同的学科发展路径争论。20世纪90年代以来，又面临紧跟计算机科学、互联网技术发展，适应工商管理需求的巨大挑战。

走向国际化、建成世界一流大学是南京大学的既定战略目标。南京大学信息管理系近日被批准加入国际 iSchools 图书情报教育联盟，这是我系推进国际化战略所迈出的坚实一步。从国际上看，2000年美国图书情报教育联合会就指出，图书情报教育的服务场景和核心技能正发生深刻蜕变，以数字环境和学科融通为特征的创新更为显著；美国图书馆学会也发现，美国所有图书情报学院开设的课程越来越多地围绕信息科学的前沿展开，充分体现了图书情报与档案管理学科的创新特征；2002年在意大利帕尔玛召开的“图书馆和情报研究国际化”研讨会和在赛萨洛尼基达成的“欧洲图书情报学教育重整和适应标准”，以及2003年在波茨坦召开的“应对变化——图书情报学教育管理变革”和随后推进的“博洛尼亚进程”均反映了学科内涵的跨学科发展与融合发展趋势。

国内外同行皆认为图书情报与档案管理学科目前仍然处于学科范式急剧“转型”阶段。因此，在学科范式和教育模式的探索中，必须以学科创新为前提和己任。南京大学信息管理系历来致力于学科前沿与社会服务的有机融合，其学科与期刊评价、数字图书馆技术、数字出版教育、数字人文阅读、信息系统采纳、信息用户行为、保密科技、档案信息资源建设等专业方向在国内外相关领域具有重要影响力。

咀英嚼华悟真知，南京大学信息管理系汇集了全系最新、最具影响力的一批科研成果，通过出版《图书情报与档案管理创新丛书》，既向国内外学者汇报南京大学信息管理

系的科研进展，也想与国内外同行相互切磋，共同为图书情报与档案管理学科的发展贡献绵薄之力。创新是一个持续的过程，我们也希望通过国家“985”工程等平台的支持，以《图书情报与档案管理创新丛书》为载体，催生更多的学术成果，将图书情报与档案管理学科的学术创新精神延续传承。

古语曰：“周虽旧邦，其命维新”^①。仅以创新为旨向，以《图书情报与档案管理创新丛书》为形式，诉予大家同仁，是为志，亦以为序。

孙建军

2011年8月于南京大学

^① 引自《诗经·大雅·文王》

前 言

相关性是情报学基本问题的观点得到了国内外图书情报学界的普遍认同。国外的相关性研究无论在理论方面,还是实证方面都取得了丰硕的成果,相关成果可参见 Saracevic (1975)、Mizzaro (1997)、Schamber (1994) 和 Saracevic (2007) 等综述。相关性研究虽然也得到国内学界的重视,但是只能检索到为数不多的国外研究成果的综述与介绍。鉴于此,作者开展了基于本土的面向用户的相关性判据研究以及相关判据应用研究。本研究的主要成果与结论有 4 个方面。

(1) 相关性判据集。本研究以南京大学信息管理系 4 个年级与南京大学教育科学系 1 个年级的相关性判据文本数据为信息源,采用内容分析法解析出了包括传播特征、内容、情境、使用、系统特征、愉悦、质量、总体与文献特征 9 类相关性判据。通过频次分析发现,文献特征、质量判据、内容判据、总体和使用占据了相关性判据的主要位置,而愉悦、情境、传播特征和系统特征则相对没有那么重要。与 Schamber (1991) 和 Barry (1993) 研究相比,本研究拓展了面向用户的相关性判据集,提炼出了包括文献总体、文献使用与传播特征等三个新的相关性判据类别。

(2) 影响信息用户相关性判断的文献特征。通过对相关性判据文本数据的内容分析,解析出了包括参考文献、出版单位、出版时间、关键词、基金、篇幅、全文、题名、文献格式、文摘、语种、文献类型、来源期刊、作者与机构在内的 15 个影响信息用户相关性判断的文献特征。

(3) 任务复杂性以及性别影响相关性判据的选择。研究结果显示,用户在面对不同复杂性的任务时,在文献内容、文献使用、文献特征、作者与文献类型等相关性判据的选择方面存在差异;而性别维度仅影响作者、文献类型与文献总体层面的相关性判据选择。由于数据量的原因,本研究没能证实任务复杂性与性别对信息用户在选择来源期刊、机构、系统特征、传播特征与愉悦感知 5 个类别相关性判据的影响。

(4) 基于相关性判据、价值增值模型、TEDS 模型以及信息系统成功模型构建了面向相关性判据的学术信息检索系统成功模型,并证实了模型的有效性。根据模型设计了问卷,问卷共发放 1114 份,回收 1054 份,其中有效问卷 929 份。通过对数据的结构方程分析证实:①系统因子 1 和系统因子 2 对选择性存在正负不同方向的影响,路径系数分别为 0.59 和 -0.39;②系统因子 2、系统因子 3 和选择性对系统效能存在正性的影响,路径系数分别是 0.11、0.15 和 0.50;③系统因子 1 和灵活性对自适应性存在正性的影响,路径系数分别达到 0.14 和 0.70;④系统因子 5、灵活性、时间认知和可靠性对系统性能存在

正性的影响, 路径系数分别为 0.20, 0.29, 0.10 与 0.38; ⑤美感和娱乐体验对情感认知存在正性的影响, 路径系数分别为 0.54, 0.27; ⑥系统因子 4 和系统因子 6 对易用认知存在正性的影响, 路径系数分别为 0.23, 0.21; ⑦完整性、实时性、权威性、有效性对于信息质量存在正性的影响, 路径系数分别为 0.26, 0.26, 0.38 与 0.25; ⑧系统因子 1、灵活性、隐私和自适应性对服务质量存在正性的影响, 路径系数分别为 0.10, 0.15, 0.10, 0.52; ⑨效能认知、性能认知、选择性、易用认知、情感认知与自适应性对系统质量存在正性的影响, 路径系数分别为 0.21, 0.30, 0.29, 0.08, 0.32, 0.23; ⑩系统质量、信息质量对满意度存在正性的影响, 路径系数分别为 0.56, 0.29。最后, 结构方程分析的结果表明, 信息质量、系统质量与满意度对使用意图存在正性的影响, 路径系数分别为 0.35, 0.55 和 0.11, 而服务质量对于使用意图存在负性的影响, 路径系数为 -0.16。

本书共分六章, 主要内容如下。

第 1 章, 绪论。主要包括选题背景、研究内容与意义、研究思路、方法和创新点。本章还对本研究的核心概念相关性从多角度进行了探讨。

第 2 章, 文献综述。全面系统地对国内外相关性理论以及相关性判据研究进行了综述。本章首先通过 3 个阶段阐述了相关性的理论研究; 然后, 将相关性判据的实证研究也分为 3 个阶段进行了详实地综述, 并进行了归纳与总结。文献综述为本研究奠定了坚实的基础, 既阐述了相关性判据研究领域的现状, 同时也为本研究指明了方向。

第 3 章, 相关性判据研究。根据从南京大学信息管理系以及教育科学系 4 个年级学生中采集的相关性判据与行为文本数据, 采用内容分析法, 分析出了包括传播特征、内容、情境、使用、系统特征、愉悦、质量、总体与文献特征 9 个类别的相关性判据。

第 4 章, 相关性判据应用研究——建模与问卷设计。在对南京大学图书馆电子资源中呈现的中外文学学术数据库进行了系统调研的基础上, 结合第三章的研究成果、Taylor 的价值增值模型、Eisenberg 衍生的 TEDS 模型与信息系统成功模型构建了面向相关性判据的学术信息检索系统成功模型, 并根据变量之间的关系提出了研究假设。根据模型完成了问卷设计并对问卷进行了项目分析以及描述性统计分析。

第 5 章, 相关性判据应用研究——分析结果。报告了实证分析的数据分析结果, 分别采用探索性因子方法完成了问卷结构效度的分析、采用 Cronbach's α 系数分析了各构念以及问卷各个部分的信度、采用验证性因子方法分析了问卷的收敛效度和区分度, 最后通过结构方程分析检验并修正了本研究第 4 章提出的理论模型。

第 6 章, 研究结论与进一步研究方向。根据前面章节的研究, 围绕构建的模型, 探讨了选择性、系统效能、自适应性、系统性能、情感认知、易用认知、信息质量、服务质量、系统质量、满意度与使用意图等的影响因素, 并分别根据研究的结果对学术信息检索系统的分析、设计与完善提出了具体意见。

本书的完成得到了南京大学孙建军教授、沈固朝教授、苏新宁教授、叶继元教授、朱庆华教授、郑建明教授、朱学芳教授与华薇娜教授的指导和帮助; 书中的诸多观点得益于

巢乃鹏副教授、李刚教授、汪传雷教授、柯青副教授、李明副教授、李君君博士、李宝强博士、苏君华博士与朱惠博士的有益建议，特此致谢。

感谢汪传雷教授、李刚教授、郭爱民副教授、巢乃鹏副教授、王水良研究员、陈先来副教授、张雪峰副教授、柯青副教授、闵红平老师、苏君华老师、李宝强博士、张俊松老师、王琛老师、徐春蕾老师、丁海燕老师、王云峰老师、沈奎林老师、李琳博士、程慧平博士、陈彬、孙海霞、史文轩、任静、黄贺方、杨石山、张露、王青、吴夙慧、龚思婷和张敏等同学在问卷调查方面所给予的帮助；对吴夙慧与张敏同学所付出的劳动致以诚挚的谢意。

本书得到南京大学“985”三期工程的资助。特此致谢！

成 颖

2011 年 7 月于南京大学

目 录

总序

前言

第1章 绪论	1
1.1 研究背景	1
1.2 研究内容和意义	3
1.2.1 研究内容	3
1.2.2 研究意义	3
1.3 研究思路、方法与创新	3
1.3.1 研究思路	3
1.3.2 研究方法	4
1.3.3 创新之处	5
1.4 术语定义——相关性	6
1.4.1 面向系统的相关性	7
1.4.2 面向用户的相关性	9
第2章 文献综述	14
2.1 相关性理论研究综述	14
2.1.1 第一阶段的研究	15
2.1.2 第二阶段的研究	15
2.1.3 第三阶段的研究	17
2.1.4 小结	23
2.2 相关性判据研究综述	23
2.2.1 第一阶段的研究	24
2.2.2 第二阶段的研究	26
2.2.3 第三阶段的研究	39
2.2.4 相关性判据的比较研究	56
2.2.5 小结	60
第3章 相关性判据研究	64
3.1 研究问题	64
3.2 数据搜集	64
3.2.1 研究对象的选择	64
3.2.2 数据搜集的基本要求	64

3.2.3 数据搜集方法	65
3.2.4 数据搜集方法的优势与不足	66
3.3 数据分析方法	66
3.3.1 内容分析法	66
3.3.2 方法的优势与不足	67
3.4 相关性分析结果	68
3.4.1 数据特征	68
3.4.2 数据集	69
3.4.3 数据分析过程	69
3.5 相关性判据	74
3.5.1 文献特征相关性判据	74
3.5.2 内容特征相关性判据	83
3.5.3 质量类别相关性判据	86
3.5.4 总体类别相关性判据	92
3.5.5 使用相关性判据	96
3.5.6 系统特征相关性判据	98
3.5.7 传播特征相关性判据	99
3.5.8 情境相关性判据	100
3.5.9 愉悦相关性判据	101
3.6 相关性判据——性别与任务复杂度分析	101
3.6.1 检验方法	101
3.6.2 大类相关性判据	101
3.6.3 子类别相关性判据分析	103
3.7 小结	116
3.7.1 相关性判据小结	117
3.7.2 影响相关性判断的文献特征小结	119
3.7.3 相关性判断的任务复杂性与性别分析	120
第4章 相关性判据应用研究——建模与问卷设计	122
4.1 引言	122
4.2 学术信息检索系统调研	124
4.2.1 调查对象	124
4.2.2 调查内容	125
4.2.3 调查结果	125
4.3 模型建构	139
4.3.1 VAM 与 TEDS 模型	139
4.3.2 信息系统成功模型	142
4.3.3 面向相关性判据的学术信息检索系统成功模型	143

4.3.4 研究假设	144
4.4 问卷设计	149
4.4.1 问卷第一部分	149
4.4.2 问卷第二部分	154
4.4.3 问卷第三部分	159
4.5 问卷前测	159
4.6 研究对象	161
4.7 描述性统计分析	161
4.7.1 缺失值的处理	161
4.7.2 描述性分析结果	161
第5章 相关性判据应用研究——分析结果	169
5.1 数据分析方法	169
5.2 探索性因子分析——结构效度	169
5.2.1 检索系统的浏览功能	169
5.2.2 检索结果显示	170
5.2.3 检索系统的导航功能	171
5.2.4 检索结果的排序功能	171
5.2.5 检索结果的标记功能	172
5.2.6 检索结果的分类/分组功能	173
5.2.7 查询扩展功能	174
5.2.8 检索方式选择	175
5.2.9 检索字段与限定条件选择	176
5.2.10 检索技术选择	178
5.2.11 相关反馈功能	178
5.2.12 检索结果输出	179
5.2.13 交互性	180
5.2.14 个性化服务	181
5.2.15 帮助服务	181
5.2.16 激励机制	182
5.2.17 链接功能	183
5.2.18 存取方式	183
5.2.19 问卷第一部分	184
5.2.20 问卷第二部分外生潜变量	188
5.2.21 问卷第二部分内生潜变量	195
5.2.22 信息质量	201
5.2.23 自适应性	201
5.2.24 易用认知	202

5.2.25 选择性	203
5.2.26 系统性能认知	203
5.2.27 情感认知	204
5.2.28 系统质量	205
5.2.29 服务质量	205
5.2.30 满意度	206
5.2.31 使用意图	207
5.3 信度分析	207
5.3.1 问卷第一部分	207
5.3.2 问卷第二部分的外生潜变量	212
5.3.3 问卷第二部分的内生潜变量	214
5.4 验证性因子分析	215
5.4.1 收敛效度	215
5.4.2 区分效度	220
5.5 结构方程分析	222
5.5.1 测量模型的拟合标准	222
5.5.2 结构模型的拟合标准	222
5.5.3 研究假设的结果解释	224
第6章 研究结论与进一步研究方向	226
6.1 研究结论	226
6.1.1 选择性的影响因素	226
6.1.2 系统效能的影响因素	226
6.1.3 自适应性的影响因素	227
6.1.4 系统性能的影响因素	227
6.1.5 情感认知的影响因素	228
6.1.6 易用认知的影响因素	228
6.1.7 信息质量的影响因素	229
6.1.8 服务质量的影响因素	230
6.1.9 系统质量的影响因素	231
6.1.10 满意度的影响因素	233
6.1.11 使用意图的影响因素	235
6.2 进一步的研究方向	237
6.2.1 多视角的相关性排序方案	237
6.2.2 检索系统的界面改进研究	237
参考文献	238

第1章 绪 论

相关性是情报学基本问题的观点得到了国内外图书情报学领域主流学者的普遍认同。

国内,马费成教授和梁战平教授高屋建瓴地提出了情报学的相关性原理。马费成(2007)教授认为“研究和揭示情报相互关联(即相关性)的规律和规则,是有效组织和检索信息、知识、情报的基础”;梁战平(2007)教授提出“相关性原理应用非常普遍,如情报检索、引文分析、数据库知识发现、知识地图、数据挖掘、市场分析研究、专利图分析、证券分析等等”;张新民等(2008)总结了相关性类型及其关系的研究成果,进一步证实了马费成教授的观点;屈鹏等(2007)的研究提供了数据方面的支持,其在对国际情报学研究主题进行聚类分析的基础上,提出的研究结论认为“一级类目情报学研究下面,……,这些主题主要归属于检索和计量两个方面。其中检索主要包括文本检索会议、相关性与相关反馈、模型和用户行为等这些信息检索基础理论(相关性、评价、模型)方面的内容,它们应该是情报检索理论研究的核心部分,是情报学对于检索领域的主要贡献所在”。

国外,Saracevic(1975,2007)更是早在20世纪70年代就提出相关性是情报学基本问题的观点,其他众多的学者如Schamber(1994)、Barry(1994)等也持相同观点。目前,信息检索中的相关性研究在国外已经取得了丰硕的成果,在相关性本质的认识、相关性模型、相关性判据以及相关性行为研究等方面都取得了一系列成果。

1.1 研究背景

学术界已经普遍认同将信息检索中的相关性研究分为两个学派,即面向系统与面向用户两个类别。面向系统的相关性研究主要由Salton(1989)等为代表的计算机界进行,面向用户的相关性研究则主要由图书情报学界进行,其中相关性判据研究一直是面向用户相关性研究中的一个核心课题。

相关性判据研究可以追溯到20世纪60年代分别由Cuadra和Katter(1967)和Rees和Schultz(1967)完成的两项相关性实验,70年代以及80年代前期的实证研究甚少,进入80年代中期之后,实证研究再次引起了学者的重视。迄今已经有多项实证研究相继展开,从研究手段上看60年代的研究主要基于实验室完成,80年代之后的研究则主要基于自然主义方法。

Cuadra和Katter(1967a,b)与Rees和Schultz(1967a,b)两个小组分别完成的相关性判据的实证研究在相关性研究历史上具有重要的地位并产生了深远的影响,至今仍被频繁引用。Cuadra和Katter(1967a,b)的研究认为焦点变量、界限变量、情境变量、刺激材料变量、个体差异变量以及量化尺度模式在相关性判断中起着重要的作用。Rees和

Schultz (1967a, b) 的研究认为实验对象及其认知在相关性判断中起主要作用。

Saracevic (1975) 在对上述两项研究的数据加以分析后认为, 影响相关性判断的判据可以归结为 5 类: 文献及文献表示、检索表达式、判断情境、标度以及实验对象。Schamber (1994) 也对包括 Cuadra 和 Katter 以及 Rees 和 Schultz 的研究数据进行了分析, 总结出了 80 项影响相关性判断的判据, 并将其分为六类: 实验对象、查询请求、文献、信息系统、判断情境与标度的选择。Schamber 认为“相关性的多维性仅仅是总体上的结论。实际上, 存在更多影响相关性判断的判据, 完整地将其列出是不可能的, 这里列出的 80 项仅仅是一个比较合理的抽样而已”。

80 年代之后的相关性判据研究主要分为两类, 第一类是对相关性判据比较全面的研究, 主要研究特定情境中所有影响相关性判断的因素, 第二类研究侧重于相关性判据的某一方面, 比如文献排序对相关性判断的影响等。

其中属于第一类相关性判据整体研究的主要有 Nilan 等 (1988)、Schamber (1991)、Cool 等 (1993)、Barry (1994)、Schamber (1994, 1996)、Wang (1994, 1998, 1999)、Tang 等 (1998)、Bateman 等 (1998)、Magloughlin 等 (2002)、Crystal 等 (2006)。

属于第二类相关性判断部分因素研究的有 Eisenberg 和 Barry (1988) 针对文献排列顺序对相关判断影响的研究、Janes 和 McKinney (1992) 与 Janes (1994) 研究了不同类型实验对象对相关性判断的影响问题、Spink 等 (1998) 的研究主要集中于探询部分相关的判断判据及其价值、Zhang (2002) 研究了相关性判断中的合作行为。

根据 Mizzaro (1997) 的研究, 到 1994 年为止, 有关相关性判据的研究合计有 24 项; 1994 年之后的相关性判据研究可以从 Saracevic (2007) 得到反映, 除了在 Mizzaro (1997) 已经阐明的之外, Saracevic (2007) 列举了 11 项 1995 ~ 2005 年的相关性判据研究。根据作者从 Weily Interscience、Ebsco、Elsevier 等数据库的检索结果获得了 2005 至今的相关性判据研究 12 篇。

综合现有的相关性判据研究可以发现 (成颖和孙建军, 2004): ①不同学者的研究结论存在明显的差异。上述众学者的研究在描述相关性判据时都是围绕着一个具体的判据完成的, 而判据在不同的情境中最容易变化, 因而也就导致了各位学者得到的相关性判据差异明显, 这种状况意味着在相关性判据的研究中缺乏一个良好的描述框架。②在相关性判据的研究中, 尽管学者们使用了一组相异的术语, 不过大都认为用户是相关性判断的核心力量, 同时认为用户的评估行为是一种认知现象。③除了相关性的认知本质以及动态性之外, 众研究还证实了相关性判据是主观的、情境的、多维的以及可测度的。④相关性判据研究已经从原先的主要探讨学术信息源的相关性判据问题转变到多种不同类型的相关性判据研究, 比如音乐、链接、医学影像、视频以及互动问答平台等。⑤随着针对不同类型文献的相关性判据研究的开展, 研究对象也呈现多元化, 比如音乐制作人、儿童、医务人员等。

在上述相关性判据研究中, 存在的无法忽视的第一个现象就是虽然国外已经开展了 30 多项相关性判据的实证研究, 但是该领域在国内一直鲜有研究开展, 对国外研究成果的介绍散见于孙绍荣、康耀红、李国秋、王家钺、张新民以及成颖等的论文中; 第二个现象是相关性判据研究成果的应用很少, 上述研究中的作者在结合某语境完成了相关性判据的研究之后, 鲜见继续开展紧随其后的应用研究, 这种状况直接导致了面向系统与面向用户的

相关性研究的脱节。

本研究针对上述两个现象做一些具体的工作，即首先开展基于本土的相关性判据研究，然后基于发掘出的相关性判据开展相关应用研究。

1.2 研究内容和意义

1.2.1 研究内容

1) 用户的相关性判据研究

在相关性判据的研究中，学者们大都认为用户是相关性判据的核心力量，同时认为用户的评估行为是一种认知现象。基于该结论，拟研究个体在相关性判据中的认知、心理以及行为因素对相关性判据的影响。

2) 相关性判据应用研究

面向用户的相关性判据研究应该为面向系统的相关性研究提供用户层面的内容。本研究以 Taylor (1986) 的价值增值模型以及 Scholl 等 (2011) 改进的 TEDS 模型为基础，结合信息系统成功模型，探讨相关性判据在学术信息检索系统改进中的应用研究。研究结论可以从用户的相关性判据的角度为学术信息检索系统的改进提供依据。

1.2.2 研究意义

(1) 相关性研究是情报学的核心，本研究将对情报学的理论体系的完善以及应用研究的深入提供有效的支撑。

(2) 目前，面向系统与面向用户的相关性研究基本上各自独立。本研究认为，面向用户的研究应当为面向系统的研究提供更多用户层面的内容，以利于在检索系统的分析、设计以及利用方面更多地考虑用户的因素。本研究的价值之一就是可以通过面向用户的研究为面向系统的相关性研究提供个性化的支持。

(3) 用户进行相关性判断需要依赖自身的知识储备、社会、心理以及问题情景等多方面因素，本研究力图充分发掘用户的心理与社会层面的因素，从而为用户的相关性判断提供较为有益的指导。

1.3 研究思路、方法与创新

1.3.1 研究思路

本研究首先在文献调研的基础上，结合专家咨询以及访谈等方式拟定研究的理论框架；然后采用文献调研、内容分析以及数理统计等方法完成本研究的基础工作——相关性

判据的综述；再次根据从南京大学信息管理系以及教育科学系 2002 级、2003 级、2004 级和 2008 级获得的相关性判据以及行为方面的原始数据，利用内容分析法进行基于本土的相关性判据的研究工作，进而采用数理统计的相关方法分析不同任务维度以及性别维度的相关性判据的差异；最后根据 Taylor 的价值增值模型、Eisenberg 的 TEDS 模型以及信息系统成功模型开展相关性判据的应用研究，研究中采用结构方程模型研讨学术信息检索系统中的相关性判据所蕴含的各种功能以及功能点对于信息质量、系统质量与服务质量的影响，并进而讨论三个因子对于满意度以及使用意图的影响，根据路径系数总结出对于满意度以及使用意图产生主要影响的各相关性判据及其所涵盖的功能点，从而提出有针对性的改进学术信息检索系统的相关建议，研究路线图如图 1-1 所示。

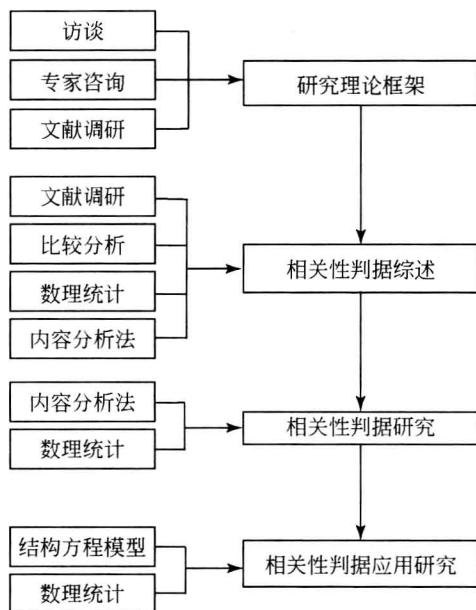


图 1-1 研究路线图

1.3.2 研究方法

本书主要采用文献调研法、专家咨询、访谈、内容分析法、比较研究、数理统计方法以及结构方程模型等多种研究方法。

1) 文献调研法

本研究系统地就相关性理论、相关性判据、相关性行为、相关性评估、Taylor 的价值增值模型、Eisenberg 的 TEDS 模型以及信息系统成功模型等领域进行文献调研，从而充分地掌握这些领域当前的研究现状与进展。同时也就结构方程模型的相关应用进行必要的调研，掌握结构方程模型的分析模式，了解学者们在使用该模型时所遇到的问题以及解决的方案。全面的文献调研是本研究得以开展的基础。

2) 专家咨询与访谈

在文献调研的基础上, 经过充分吸收, 确定初步的研究思路并拟定初步的研究方案, 并就研究方案访谈南京大学以及武汉大学的相关领域专家, 就方案的思路等相关问题开展专家咨询, 根据专家意见优化研究方案, 经过多次交互之后, 形成研究方案。

3) 内容分析法

目前, Kerlinger 给出的内容分析法的定义最具代表性, 即内容分析是以测量变量为目的, 对传播进行系统、客观和定量分析研究的一种方法。内容分析法的特征可以归纳为明显、客观、系统、量化 4 个方面。内容分析法的分析步骤包括提出研究问题或假设、确定研究范围、抽样、选择分析单位、建立分析的类目、建立量化系统、进行内容编码、分析数据资料与解释结论。本研究已经积累了 4 个年级的学生围绕课程作业形成的相关性判据文本数据, 对这些无结构的文本, 本研究将按照内容分析法的分析流程进行分析处理, 从而获得信息用户的相关性判据, 完成主要的一项研究工作。

4) 比较研究

比较研究法是对事物异同关系进行对照、比较, 从而揭示事物本质的思维过程和方法。本研究通过对国外的相关性判据实证研究进行比较分析, 从中发现这些研究的共性与差异, 力图找出研究中所蕴涵的跨平台、跨语境的相关性判据集。

5) 数理统计法

本研究中涉及大量的基础数据的统计分析, 包括相关性判据分析中的求和、均值、 t 检验、秩和检验等, 还包括问卷分析中大量的最大值、最小值、标准差、均值、主成分分析、信度分析以及效度分析等。

6) 结构方程模型

结构方程模型, 也称为结构方程建模 (structural equation modeling, SEM), 是基于变量的协方差矩阵分析变量之间关系的一种统计方法, 也称为协方差结构分析 (侯杰泰等, 2004)。SEM 整合了因素分析和路径分析两种统计方法, 同时检验模型中包含的线性变量、潜在变量、干扰或者误差变量间的关系, 进而获得自变量对因变量影响的直接效应、间接效应或者总效应 (吴明隆, 2009)。本研究通过 TEDS 模型与 ISSM 模型, 结合相关性判据进行 SEM 建模, 分析检索系统的功能、功能点以及元素对相关性判据的影响, 进而分析相关性判据对信息质量、系统质量以及服务质量的影响, 并探讨三者对满意度和使用意图的影响。SEM 分析的结果将对基于相关性判据的检索系统改进提供积极的借鉴。

1.3.3 创新之处

(1) 构建了一套基于本土研究的面向用户的相关性判据集。国内面向系统的相关性研