

doi:10.3772/j.issn.1000-0135.2011.09.005

面向学术新人的相关性判据研究—基于本科课程 论文的内容分析¹⁾

成颖

(南京大学信息管理系, 南京 210093)

摘要 本文探讨了两个问题:其一是学术新人的相关性判据,其二是影响用户相关性判断的文献特征。在4个年级学生相关性判据文本的基础上,运用内容分析法解析出了包括传播特征、文献内容、情境、文献使用、系统特征、愉悦感知、文献质量、文献总体以及文献特征等9类相关性判据;针对第二个问题,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、作者以及机构等在内的15个文献特征。论文结合每类判据提出了针对学术信息检索系统的改进意见。

关键词 相关性判据 内容分析法 学术新人

On Academic Novice Oriented Relevance Criteria—Based on Content Analysis of Undergraduates' Term Papers

Cheng Ying

(Department of Information Management, Nanjing University, Nanjing 210093)

Abstract Two questions are discussed in this paper, one is on relevance criteria to academic novice, the other is characteristics of document that affects relevance judgement. Nine types of relevance criteria are explored with content analysis to four grades undergraduates' text, respectively communication, documentation content, context, the use of documents, system characteristics, perceived pleasure, quality, document collection feature and the characteristics of document. To the second question, 15 document features are found, including references, publishers, publishing time, keyword, funds, length, full-text, title, document format, abstracts, language, document type, sources, authors and affiliation. Suggestions to improve academic information retrieval system are proposed based on analyzing each type of criteria.

Keywords relevance criteria, content analysis, academic novice

1 引言

在国外, Saracevic^[1-3]早在20世纪六七十年代就提出相关性是情报学基本问题的观点,该观点得到 Schamber^[4]、Barry等^[5]很多学者的认同。马费成^[6]、梁战平^[7]以及张新民等^[8]认为相关性原理是

情报学的基本原理,该观点得到了屈鹏等^[9]实证研究的支持。目前,学术界已经普遍认同将信息检索中的相关性研究分为面向系统与面向用户两个学派。面向系统的相关性研究主要由 Salton^[10]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。

收稿日期:2011年6月27日

作者简介:成颖,男,1971年生,博士,副教授,主要研究方向:信息检索。E-mail:chengy@nju.edu.cn。

1) 本文得到国家社科基金项目“中文学术信息检索系统相关性集成研究”,项目批准号:10CTQ027;教育部人文社会科学规划基金项目“面向用户的相关性标准及其应用研究”,项目批准号:07JA870006。

2 相关研究

相关性判断研究可以追溯到20世纪60年代分别由Cuadra与Katter^[11]以及Rees与Schultz^[12]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。

1967年,Cuadra与Katter以及Rees与Schultz两个小组分别完成的相关性判断实证研究在相关性判断研究的历史上具有重要地位并产生了深远影响,至今仍被频繁引用。Cuadra与Katter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作用。Rees与Schultz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。

1975年Saracevic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断可以归结为五类:文献及文献表示、检索表达式、判断情境、标度以及实验对象。1994年Schamber也对包括1967年Cuadra与Katter和Rees与Schultz的数据进行了分析,总结出了80项相关性判断,并将其分为六类:实验对象、查询请求、文献、信息系统、判断情境以及标度的选择。1994年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断,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”。

20世纪80年代之后的相关性判断研究主要分为两类,第一类是对相关性判断比较全面的研究,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断的某一方面,如文献排序对相关性判断的影响等。其中属于第一类的研究主要有Nilan^[13]、Schamber^[14]、Cool等^[15]、Barry^[5]、Schamber^[16]、Wang等^[17-19]、Tang和Solomon^[20]、Bateman^[21]、Magloughlin^[22]、Crystal和Greenberg^[23]、Xu和Chen^[24]以及Xu^[25]等。属于第二类的研究有Eisenberg与Barry^[26]针对文献排列顺序对相关判断影响的研究、Janes与McKinney^[27]以及Janes^[28]研究了不同类型实验对象对相关性判断的影响、Spink等^[29]的研究主要集中于查询部分相关的判断及其价值、Zhang^[30]研究了相关性判断中的合作行为。

Mizzaro^[31]的研究表明,到1994年为止,有关相关性判断的研究合计有24项;2007年Saracevic将该领域的研究补充到2005年,除了在1997年Mizzaro已经阐明的之外,2007年Saracevic列举了11项1995~2005年的相关性判断研究。作者通过Weily Interscience、Ebsco、Elsevier等数据库的检索发现2005至今的相关性判断研究有12项。综合现有的相关性判断研究可以发现^[32]:①由于学者间研究语境的差异,研究结论存在一定差异。②学者们使用了一组相异的相关性判断术语。③除了相关性的认知本质以及动态性之外,研究证实了相关性判断的主观、情境、多维以及可测度特征。④该领域的研究已经从早先主要探讨学术信息源的相关性判断扩展到不同类型信息源的相关性判断研究,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断研究的开展,研究对象也呈现为多元化,如音乐制作人、儿童、医务人员等。

相关性判断研究中一个无法忽视的现象是,国外已经开展了40多项相关性判断的实证研究,而国内的研究主要是对国外研究成果的综述,如孙绍荣^[33,34]、康耀红^[35]、李国秋^[36]、王家钺^[37]、张新民^[8]以及成颖等^[38,39]的论文;实证研究可以检索到的有于春等^[40]、胡秀梅和邓小昭^[41]等两项研究;第二个现象是相关性判断研究成果的应用很少,上述研究中的作者在结合特定语境完成了相关性判断的实证研究之后,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相关性研究的脱节。本文要进行的研究工作就是针对第一个现象做一些探索,即开展基于本土的相关性判断研究。

3 研究设计

3.1 研究问题

本文围绕相关性判断主要探讨两个问题,第一个问题,探讨学术新人在检索、分析以及利用学术信息时的相关性判断;第二个问题,探讨影响相关性判断的文献特征,即学术新人进行相关性判断时,文献中的哪些特征提供了其进行相关性判断的依据。

3.2 数据搜集

与1991年Schamber、1993年Barry等研究不同的是,本研究中的对象选择没有采用招募的方式,而

是将南京大学信息管理系 2002 级、2003 级、2004 级以及 2008 级本科生、教育科学系 2008 级本科生四个年级的学生作为研究对象。所有学生都被布置一篇“计算机网络”课程论文/作业,让学生在提交课程论文/作业的同时,详细描述检索过程,要求研究对象分别描述检索策略、是否进行二次检索以及进行二次检索的原因、相关文献相关性判断的判据。研究对象中男性比例高于女性,年龄段比较集中,在研究数据搜集的时候,大部分学生的年龄在 18~21 岁,数据的搜集时间分别是 2005 年 12 月、2006 年 11 月、2007 年 10 月以及 2010 年 12 月,分别在秋季学期的中后期进行,学生的人口统计学特征见表 1。

表 1 学生的人口学特征

性别	男	123	教育程度	本科三年级	194
	女	71	年龄	18~21 岁	194
系科	信息管理	188	检索目的	课程论文	187
	教育学	6		课程作业	7

3.3 分析方法—内容分析法

内容分析的各步骤中,最重要的环节之一是编码模式的制定,本研究根据 1993 年 Barry 的编码模式形成过程制定编码模式。

(1)在借鉴 1991 年 Schamber、1993 年 Barry 等研究成果的基础上,编码类别由学生的文本中通过开放的、探索性的、递归的方式归纳产生。

(2)每种编码模式都进行了信度检验。

(3)当信度检验达不到可接受的水平时,则修订编码模式,并且重新随机选择数据进行信度检验,该过程迭代进行,直到达到可接受的信度水平。

(4)编码模式用于编码整个数据集。本研究的类目体系在编码的实践中证明具备互斥性、完备性和信度等特征。

3.4 优势与不足

相较于 1991 年 Schamber 以及 1993 年 Barry 的研究,本文的研究对象为四个年级的三年级本科生,这样做的优点是样本量有保证(共计有 194 名研究对象),样本量显著大于 1991 年 Schamber 以及 1993 年 Barry 的研究(分别只有 30 人与 18 人),大样本对于研究结论信度与效度的意义是显然的。不过,本研究在研究对象的选取方面存在明显的不足之处,表现在样本的同质性比较强,另一个不足之处

是整个研究通过布置课程论文完成,与 1991 年 Schamber 和 1993 年 Barry 等研究完全基于研究对象的自发任务存在一定差别,因此如果要将本研究的结论进行推广,则有必要将本研究与相关研究进行整合分析(meta-analysis)以获得更可靠的结论。

3.5 分析过程

本研究的数据分析工作分两轮进行。

3.5.1 第一轮 数据降维

与先前的研究相比,本研究中的样本量比较大,因此首先需要进行数据简化,从学生的文本中提炼出能够表达相关性判据描述的部分。分析单位的选择并不拘泥于某种特定的形式特征,根据学生的描述提炼出能够表征其相关性判据的单位,因此分析的单位可以是词、句子、段落或者整个篇章。第一轮数据分析工作借助于质性分析软件 Nvivo8.0 完成,由作者独立完成。第一轮分析的结果见表 2。判据数,是指某位同学提到的相关性判据的数量;引用数,是指某位同学引用判据的频次。

表 2 第一轮相关性因素按照年级分析结果

年级	判据数 最大值	引用数 最大值	判据数 最小值	引用数 最小值	判据数 均值	引用数 均值
2002	26	39	1	1	6.88	8.61
2003	32	52	1	1	5.33	6.24
2004	20	28	1	1	4.71	3.56
2008	18	26	1	1	6.54	7.46

第一轮分析的作用包括:其一是数据降维,近 200 名同学用于描述检索过程、检索策略以及相关性的文字篇幅非常大,直接邀请编码员进行编码难度较大,而且达到内容分析所要求的一致性难度则更大,因此这项工作由作者完成比较合适。通过第一轮的工作,原先的大数据集被削减到 471 项影响相关性的因素,通过 Nvivo 从学生的描述中突出显示了 1212 项涉及相关性因素的具体描述,这些结果成为第二轮相关性判据内容分析的依据与基础。其二是作者通过第一轮的预编码工作,对总体的相关性判据情况有了概貌性的、宏观的了解,这样在制定编码模式的时候就有很强的针对性。第三,数据集缩小之后,第二轮的工作量也变小,编码员容易配合,同时也容易达到合乎要求的一致性。

3.5.2 第二轮相关性判据分析

第二轮编码工作由作者以及邀请的一位研究生作为编码员进行,编码员间的信度计算通过霍斯提(Holsti)系数进行,要达到比较高的信度,该值应大于90%^[42]。在第一轮工作的基础上,本研究将相关性判据从宏观上分为九大类,分别是传播特征、文献内容、情境、文献使用、系统特征、愉悦感知、文献质量、文献总体以及文献特征。文献总体以及文献使用判据在1991年Schamber等的研究中没有出现,主要原因可能是研究方法的差异,1991年Schamber等采用的是访谈法,当时他们的研究语境还没有现在那么容易获得大量的全文数据。本研究要求学生就课程论文进行多途径的信息检索,因此容易在检索结果层面获得大的数据集,从而牵涉对于数据集整体的相关性判据。在本研究中,学生要完成课程论文/作业,从而描述相关性判据的同时也阐述了相关文献在其论文中的具体使用,因而和先前的研究相比就多出了文献使用类别的相关性判据。由于技术进步的原因,本研究出现的传播特征也是先前的研究中未曾见过的,该类别的相关性判据主要原因在于因特网繁荣之后,使得转载、点击以及引用变得更加容易而产生。

4 分析结果

4.1 文献特征

文献特征由于主要涉及文献的一些著录信息,比较客观,编码员间的一致性为99%。学生能够非常清晰地表达这些明确的特征,基本上不需要向编码员做更多的解释。分析结果显示共有15个文献特征方面的相关性判据,分别是参考文献、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、机构、来源期刊、文献类型以及作者等,其中机构、来源期刊、文献类型以及作者又区分出不同的子类别。

(1)机构子类别衍生出了几个次一级判据,包括客观、声誉以及多产等。有部分学生只是不带倾向性地提到了机构,本文把第一类归为客观判据,第二类学生主要提到了机构的声誉,因此归为声誉判据,第三类考虑机构是否多产或者高产,归为多产判据。

(2)来源期刊的相关性判据,尽管学生使用的术语有差异,不过其中最基本的相关性判据就是期

刊的声誉和排序。

(3)文献类型判据可以细分为:标准、博士论文、访谈、官方网站、会议文献、技术白皮书、讲座、教材、硕士论文、网络百科、文献类型、新闻以及综述等。因为每种文献都有其自身的特点,因此不同文献类型能够左右其相关性的选择与判断。

(4)作者相关性判据衍生出了几个子判据,包括高产、客观、群体、声誉、研究领域以及职称。有学生比较客观地提到了作者,这一类归为客观判据;第二类提到了作者的声誉,因此归为声誉判据;第三类考虑作者是否高产归为高产判据;第四类学生考察了个体作者还是群体作者,归为群体作者;有学生还考虑了文献主题与作者研究领域的相关性,从而归结为研究领域判据;最后一类是作者的职称。

4.1.1 文献特征相关性判据的频次分析

在文献特征的相关性判据类别中,可以发现文献类型、文摘、题名、作者、来源期刊以及机构等文献特征成为学生判断文献相关的主要依据,相对而言,篇幅、参考文献、出版单位、基金以及文献格式则不那么重要,见表4。表4可以发现文献类型在文献特征的相关性判据中占据了突出的位置,该相关性判据在信息检索系统设计以及学生信息素质的培养与教育中都具有非常重要的价值。首先对于信息检索系统的设计来说,需要充分考虑不同文献类型的价值,针对不同类型的信息用户,宜优先提供相应类型的文献供其选择,对于信息素质的培养与教育来说,则应该让学生充分了解不同文献类型的特点,从而指导学生根据自己的信息需求特点,选择更加合适的文献类型,进而减少不必要的时间与认知浪费。

4.1.2 文献类型相关性判据

表5没有体现出期刊论文的位置,这是因为编码员在编码的时候首先列出了学生阐述中的文献类型,而期刊论文在学生的描述中,没有体现出来,但是通过学生的参考文献可以发现,几乎每名同学都引用了期刊论文,因此作为“大众”的期刊论文需要从中挖掘隐含的非明示信息,而学生在描述中主要列出了作为“小众”的会议文献、学位论文、新闻等文献类型。通过进一步的分析发现,除了2008级有9名同学没有引用期刊论文之外,其他同学都至少引用了一篇期刊论文,每位同学不重复累计,只记一次,这样表5和表6分别增加一行。

表4 文献特征相关性判据频次分析

相关性判据	学生数	百分比%	频次	相关性判据	学生数	百分比%	频次
文献类型	103	29.51	132	语种	15	4.30	17
文摘	43	12.32	44	出版时间	13	3.72	14
题名	35	10.03	36	篇幅	11	3.15	11
作者	28	8.02	34	参考文献	7	2.01	7
来源期刊	26	7.45	31	出版单位	4	1.15	4
机构	23	6.59	33	基金	3	0.86	4
全文	20	5.73	21	文献格式	2	0.57	2
关键词	16	4.58	16	合计	349	100	406

表5 文献类型相关性判据

文献类型	学生数	百分比%	频次	文献类型	学生数	百分比%	频次
访谈	1	0.35	1	新闻	8	2.77	10
讲座	1	0.35	1	网络百科	5	1.73	12
文献类型	2	0.69	2	综述	20	6.92	20
学位论文	3	1.04	3	标准	16	5.54	23
官方网站	4	1.38	5	教材	33	11.42	45
会议文献	5	1.73	5	期刊论文	186	64.36	186
技术白皮书	5	1.73	5	合计	289	100.00	318

表6 文献类型相关性判据的年级分布

相关性判据\年级	02	%	03	%	04	%	08	%
标准	2	1.92	2	1.98	2	14.29	10	14.29
访谈	1	0.96	1	0.99		0.00		0.00
官方网站	1	0.96	1	0.99	1	7.14	1	1.43
会议文献	3	2.88	4	3.96		0.00	1	1.43
技术白皮书	1	0.96		0.00		0.00		0.00
讲座	1	0.96		0.00		0.00		0.00
教材	4	3.85	7	6.93	3	21.43	19	27.14
网络百科		0.00		0.00		0.00	5	7.14
文献类型	2	1.92		0.00		0.00		0.00
学位论文		0.00		0.00		0.00	3	4.29
新闻	7	6.73		0.00	1	7.14		0.00
综述	6	5.77	11	10.89		0.00	3	4.29
期刊论文	76	73.08	75	74.26	7	50.00	28	40.00
合计	104	100.00	101	100.00	14	100.00	70	100.00

从表5和表6可以发现,学生在使用相关文献解决问题时,主要还是采用期刊论文作为信息源,但是年级之间存在很大差异。比如,2002级与2003级学生除了期刊论文之外,其他类型的文献基本上很少,虽然也提及新闻,但是该类型主要是作为排除的对象纳入考虑的,因此,核心的文献类型都是期刊论文。2003级学生期刊论文的使用比例与2002级相似,但是作为期刊论文之一的综述比例在上升。2004级学生解析出的相关性判断非常少,大多数学生采用了教材、标准、官方网站等信息源解决其手头的问题。而2008级学生,在期刊论文的使用方面比例显著下降,主要使用教材、标准、网络百科、官方网站等作为重要的信息源,根据文献学的知识可知,这些文献类型具有系统性、入门性、成熟性等特征。

是什么原因导致不同年级的学生在文献类型的使用方面产生如此大的差异呢?这四届学生都是在三年级的时候参与了本研究,因此人口学方面的特征几乎没有差异,因此单纯从年龄等角度考虑不能给出答案,那么只有从不同的任务角度去解析、寻找问题的答案。通过任务的分析可以发现,2002级、2003级以及2008级的学生都被要求完成一篇综述性的论文,而2004级学生则完成一道作业题,因此学生通过自己的思考,再参考一下教材、标准等信息源就可以找到答案,这也是在2004级学生分析出的判断比较少的原因。其他三个年级的任务虽然相似,但是任务难度存在差异,2002级的问题在当时还是学界研究热点,综述、教材等比较综合性的文献类型基本没有这些内容,因此学生更多地寻找原始期刊论文作为参考文献进行文献综述,2003级的题目有多种选择,部分题目相对比较成熟,因此可以发现教材与综述的比例在上升,但是依然依靠期刊论文作为主要信息源。而2008级的综述题目已经退出了研究热点的行列,大量教材中已经出现了相关问题比较系统的阐述,并且已经制定了相关的标准,所以教材、标准等非常成熟的文献成为学生的主要信息源,表5中期刊论文的比例偏高,主要是因为只要出现该类型即计算一次所致,如果计算绝对数,则教材的比例将明显上升。

从上述分析可以发现,不同的任务类型对相关性的影响非常深远,对于简单问题(2004级问题),学生大多可以不需要借助于相关的信息源,主要根据自己现有的知识,加上相关思考即可以解决问题,很多学生写出相关性判断主要是因为作业中

有该要求,从而列出了教材、标准等参考文献。而相对复杂的任务,如2008级的问题,相对比较成熟,学生可以从标准、教材等成熟的文献中找到完成课程论文的信息源。2002级与2003级的任务则比较复杂,从而学生需要大量运用原始期刊论文作为信息源才能完成课程论文。

从表6还可以发现,文献类型也体现了时代特性,如网络百科这一新型的信息源,在2005~2007年刚出现,学生在引用文献类型的时候,没有涉及该文献类型,但是时间推移到2010年,网络百科的发展已经形成了一定的规模,诸如百度百科、维基百科等面向大众的普及型百科网站已经成为一种影响学生的相关性判断与使用的重要文献类型。

4.1.3 机构相关性判断

机构相关性判断中影响最大的是声誉,共有18位同学提及;有4位同学比较客观地提到了机构;提到高产判断的只有一位同学。机构判断在信息检索系统的分析与设计中可以发挥更重要的作用,如大部分信息检索系统中,已经将机构作为检索入口,但是在检索结果输出设计时将机构纳入考虑的还不是很,CNKI出现了有关机构的分组信息,但其依据的依然是机构的字顺,没有体现机构的声誉以及高产等判断特征。

4.1.4 作者相关性判断

作者相关性判断中影响最大的也是声誉,共有10位同学提及;而后,有7位同学比较客观地提到了作者,不过没有明显的倾向性,后面依次是职称、研究领域、群体作者以及高产。通常,根据职称评定的规则,职称高的作者其成果质量应该高于低职称作者,但是该因素和机构的声誉之间存在很大的交叉关系,现实中存在着这样的现象,即声誉高的机构,其低职称作者的研究成果可能比声誉一般机构的高职称作者的研究成果要好。作者的研究领域与成果主题是否吻合也成了一个重要的相关性判断,如果作者在某个研究领域进行了长时间的研究,其研究成果的质量应该有保证,反之则质量难以保证。学生提到群体作者比单一作者的成果质量要高一些,从总体而言也许如此,但实践中该结论是否成立有待进一步探索。根据洛特卡定律,某一领域的高产作者会成为该领域的核心作者,从而显现出高相关性。

4.1.5 来源期刊相关性判据

共有 26 位同学的结果中有“核心期刊”以及“来源期刊”等描述,本文将其归结为来源期刊的声誉,根据布拉德福定律,该判据可以应用于检索系统的相关性排序等领域。

4.2 内容特征的相关性判据

分析结果显示,信息内容的相关性判据包括:比较分析方法、技术实现、类别、实践、实验、数据、图解、专题性以及综合性。这些相关性判据从文献内容的两个不同维度对信息内容进行编码,一个维度涉及文献的部分内容,另一维度涉及文献的总体内容。第一个维度解析出的判据有:比较分析、方法、技术实现、类别、实践、实验、数据、图解;第二个维度从一位同学的描述中获得了借鉴,即将文献内容从总体上划分为专题性与综合性两个判据。专题性就是文献仅阐述某一项技术、协议或者标准,内容的内聚性比较强。综合性则是阐述的内容比较广泛,通常包括不同技术的比较与分析等内容。通过这种比较简明的相关性判据的编码,编码员间的一致性很快提高到 95%。从表 7 可以发现,由于课程论文属于综述性质,学生在文献内容的选择方面主要倾向于综合性、有图解的以及专题性文献,文献内容中的方法、数据、所属类别、技术实现以及实验等对学生的相关性判断影响相对要小得多。

不同学科有其不同的文献内容特征,比如文史哲等学科,可能没有或者很少实验、数据等内容,因此要将文献内容的相关性判据应用到信息检索系统的分析与设计,难度高于其他几个相关性判据类别。

4.3 质量类别相关性判据

本研究将质量定义为学生的主观判断或者影响

学生对文献信息提供的情境因素。文献特征与质量判据的主要区别在于这些特征能够被实际用户以外的用户评判的程度。文献特征是文献的内在特征,可以通过其是否出现在文献中进行客观的判断。质量类别则不能作为文献的内在特征,主要依赖于研究对象进行相关性判断时的情境、知识水平以及感知,如文献是否“通俗易懂”完全依赖于研究对象对文字的感知及其先前的知识水平。

质量类别判据的制定极具挑战性,不仅仅是形成类别判据,同时也体现在编码员间达到可接受一致性水平的难度。主要原因在于质量类别判据的形成仅仅依赖于学生对潜在内容的认识。虽然一组关键词与短语用于每种质量判据,但是实际的编码工作中还需要考虑这些关键词或者短语使用的具体语境。换句话说,编码过程中涉及大量的编码员的自我阐释,从而导致了编码员间的不一致。这种不一致最终通过不断地修订编码手册得以消解,如参照 1993 年 Barry 的编码规则,有一个编码为深度/广度,通过实际的编码发现,采用这种大粒度的编码判据,有大量的实例可以编码为该类,从中文的习惯可知,深度和广度实际上是两类不同的规范,因此将 Barry 的判据分拆为深度与广度两个独立的判据。有些类别编码员之间总是存在分歧,因此部分类别是通过讨论确定编码,最终编码员间的一致性提高到 92%。最终的分析结果显示共有 19 个信息质量方面的相关性判据,分别是广度、简明性、结构清晰、可靠性、可理解性、美感、权威性、深度、实时性、系统性、先进性、详细、新颖性、学术性、知识水平、直觉、主题、专指性以及准确性。

表 8 描述了相关性质量类别的使用频次等信息。表 8 的数据表明,学生的相关性质量判据存在很大的差异,其中占据比较重要位置的是主题、详细、可理解性以及权威性判据,而结构清晰、系统性、知识水平、美感以及直觉等判据位居其次。

表 7 内容类别的相关性判据

内容判据	学生数	百分比%	频次	内容判据	学生数	百分比%	频次
方法	2	1.87	2	实践	9	8.41	11
数据	2	1.87	2	比较分析	10	9.35	10
类别	3	2.80	3	专题性	15	14.02	17
技术实现	4	3.74	4	图解	18	16.82	19
实验	4	3.74	5	综合性	40	37.38	51
				合计	107	100.00	124

表 8 质量类别相关性判据

相关性判据	学生数	百分比%	频次	相关性判据	学生数	百分比%	频次
主题	79	17.40	81	实时性	13	2.86	17
详细	64	14.10	75	可靠性	12	2.64	12
权威性	44	9.69	52	先进性	11	2.42	12
可理解性	43	9.47	54	准确性	8	1.76	9
广度	40	8.81	51	系统性	7	1.54	7
新颖性	35	7.71	38	结构清晰	7	1.54	7
简明性	27	5.95	39	知识水平	2	0.44	2
深度	24	5.29	27	直觉	1	0.22	1
学术性	20	4.41	24	美感	1	0.22	1
专指性	16	3.52	17	合计	454	100.00	526

表 9 总体类别相关性判据

相关性判据	学生数	百分比%	频次	相关性判据	学生数	百分比%	频次
顺序	2	2.90	2	文献计量	3	4.35	4
外部证实	2	2.90	2	查全率	5	7.25	5
质量	2	2.90	2	学术数据库	7	10.14	7
搜索引擎	3	4.35	3	同质性	14	20.29	20
存取	3	4.35	3	数量	24	34.78	26
查准率	4	3.50	4	合计	69	100.00	78

4.4 总体类别相关性判据

1991年Schamber等的研究更多地关注文献个体的相关性判据,本研究解析出了除个体相关性判据之外的总体相关性判据。所谓的总体相关性判据就是在考虑相关性判据时不再单纯地从每篇文献的个体角度去考虑,而是从检索结果集的整体去宏观地判断是否相关,另一种情况就是检索者在检索之初就从宏观的角度考虑问题,比如优先查全,然后查准,而这些思考就会带来检索策略的调整与变化。由于该组相关性判据基本上通过明示信息即可编码,从而编码员之间的一致性很快达到93%。编码结果显示总体类别的相关性判据包括:查全率、查准率、存取、数量、顺序、搜索引擎、同质性、外部证实、

文献计量、学术数据库、质量等具体判据。

表9表明,在总体类别的相关性判据中,数量、同质性等占据了主要的位置,而顺序、外部证实以及质量则不太重要。检索结果集的数量对相关性的影响排在了总体相关性判据的首位,通过对学生文字描述的分析发现如果文献结果集比较小,则学生倾向于将其列为相关,如“本文对此书的借鉴意见说实在不是太多,借此书主要是因为我去图书馆时,有关蓝牙书籍均被我们系同学‘抢’完了,我就没有挑剔地借鉴了这本书籍。该书对蓝牙技术的知识介绍比较细致和全面,我仅参考了其很小一部分知识”。而当结果集比较大时,由于存在着比较大的选择余地,需要继续运用诸如是否存在同质性等判据继续选择,从而意味着当检索结果集比较大时,合理的结

果集排序就会显得非常重要。

对于学术数据库和搜索引擎而言,搜索引擎检索出来的信息的学术性差异非常大,有部分检索结果本身就是正式发表的学术论文,但是也存在大量的简介以及新闻等学术性有限的信息,而学术数据库中虽然也存在学术论文良莠不齐的现象,但是总的来说,其学术的严谨性、科学性、可靠性等指标显然远远高于搜索引擎,因此学生在谈到搜索引擎的结果时,基本以拒绝为主,对于学术数据库则总体以采纳和接受为主。

同质性判据对于检索系统的分析与设计有着非常重要的启示(占到20.29%)。根据本文作者的检索体验,有时检索结果集虽然较大,但是文献的同质性非常强,很多文献的新意非常有限,因此检索系统如果能过滤掉一些十分相似的文献,对于减少信息用户的认知努力意义很大。这个问题可以从不同的角度解决。一是从信息生产的源头控制同质性非常强的文献进入学术文献系统;二是信息检索领域加强基于内容的鉴别研究与应用,不过难度很大;三是提供优秀的排序算法,利用期刊、作者或者机构的声誉将相关度比较低的文献排在后面,从而减少信息用户的认知努力,提高整个系统的效率。

4.5 使用相关性判据

本文的研究对象——学生在描述相关性判据的同时也阐明了相关文献的使用,同时使用也成为学生进行相关性判断的重要依据。通过探索性的分析,本文解析出的有关信息使用的相关性判据包括:读者对象、潜在利用、架构借鉴、有效性以及提供背景知识等。该组相关性判据也是通过明示信息获得,编码者间的一致性为91%。对于学生这样的领域新人首先面对的就是如何入门,因此提供背景知识成为首选的使用需要,共有23位同学提到了该判据;其次有21位同学提到有效性作为重要的使用需求,但是该判据在本研究中被严重低估,因为所有被列为参考文献的相关文献都是有用的,而本研究仅仅从学生的阐述中解析出了该判据,因此该判据没有排在第一位,但也说明该判据非常重要。13位同学提到架构借鉴,在使用类别中占到了21.31%,该判据说明作为初学者的学生还不太了解综述的写法,所以需要从综述性论文中借鉴综述的架构、行文风格等写作层面的内容。2位同学提到潜在利用判据,虽然频次比较低,但是就作者来说,也经常会将可能有用的文献先行下载。虽然也只有2位同学,

不过文献的读者对象判据对于检索系统的分析和设计有潜在的价值,如果系统能够提供对读者对象的处理,则信息用户可以根据该判据选择符合自己需求的文献,从而可以在很大程度上减少信息用户的认知努力,提高整个系统的效率。

4.6 系统特征相关性判据

与Taylor^[43]的研究结果相似,学生的相关性判据描述中也提到了系统特征,分析结果显示涉及系统特征的相关性判据有响应速度、著录详细以及存取。该组相关性判据通过明示信息即可获得,编码者间的一致性为90%。学生提及的系统特征判据数量比较少,其中有21位同学提到存取,该判据是影响学生相关性判断最主要的系统特征。对于信息用户来说,从需求分析、形成检索策略、提交检索提问、到返回检索结果,当需要点击链接获取文献的时候,被告知不能获取,那种沮丧的心情可想而知。因此对于信息检索系统来说,提供清晰的是否可以存取的标志就显得非常重要,如在CNKI中不论相关文献是否提供下载,都有PDF以及CAJ的下载标记,对于那些不提供下载的文献,在信息用户点击之后,系统才提示因为保密等原因不提供下载。实际上,对于是否能够下载提供不同的标记在系统实现层面是非常简单的,最简单的做法只要让该标签成为disable即可。有1位同学提到著录详细程度判据,从现有的学术数据库的输出结果来看,著录详细程度差异不大,都提供了基本的著录内容,大多提供了详细格式和/或简单格式的选择,因而著录层面的界面设计如何向着更符合信息用户的偏好是改进方向,属于人机交互等领域的研究,不再深入探讨。1位同学提到响应速度,该判据对信息用户来说是一个非常重要的系统特征,由于学生主要在校园网上使用这些数据源,总体响应速度可以满足。

4.7 传播特征相关性判据

本研究与1991年Schamber等研究相比,技术环境已经发生了变化,因此衍生出了先前研究中没有出现的一些相关性判据,本研究解析出了与信息传播相关的相关性判据,分别是引用、点击率以及转载,分别有6位、2位和1位同学提到这些判据。该组相关性判据涉及的同学很少,表述也很清楚,编码者间的一致性为96%。转载是随着网络技术发展之后衍生出来的相关性判据,在先前主要为纸质载

体的信息传播机制中,虽然也存在类似于《人大复印资料》等转载的方式,不过量是非常有限的。网络技术的不断进步,带来了转载的便利,转载量也成为评价文献质量的非常重要的判据,相应地也成为信息用户进行相关性判断的判据;相较于转载,点击率则更成为网络时代的特征,信息资源的点击率在在一定程度上可以反映信息的质量;引用在传统的文献传播中本身就占据了很重要的位置,文献计量学中非常重要的部分就是文献引用的相关度量,包括引用、被引、同被引等概念与理论已经比较成熟,已经有大量的学术论文以及教科书围绕该主题完成。在网络环境下,引文得到了拓展,已经发展成为非常重要的多学科协同的网络计量学,其中有代表性的成果如基于网络引证关系的 PageRank 算法等。引用、点击率以及转载这三个传播特征的相关性判据在信息检索系统的分析与设计方面在部分系统中得到了应用,如 CNKI 就提供了下载频次和引用频次的排序指标,信息用户可以根据这些指标排序检索结果。

4.8 情境相关性判据

除了文献总体特征、文献个体特征、系统特征等类别的相关性判据之外,信息用户的情境特征也在一定程度上影响信息用户的相关性判断。本研究中主要涉及时间限制判据。该判据有代表性的例子有“但是时间也有限”、“由于时间仓促”等。提到该判据的学生数量不多,仅有 5 位同学提到了该判据,所占比例不大,编码者间的一致性高达 98%。

4.9 愉悦相关性判据

本研究最后解析出来的相关性类别是愉悦,主要解析出了情感认知子类,有代表性的实例包括“但由于在这篇文献字体清晰,看起来舒服,所以对这篇文献认真读了一下,当然还是主要阅读了它的前两个方案,来作为我的解决方案的参考的”、“ProQuest 的一个优点是可以选择语言,从而可以将文章翻译为中文,然后看起来就比较舒服”。与情境类别的相关性判据相似,提到该类别的学生只有 2 位同学,编码者间的一致性高达 100%。该类别的相关性判据在商业性网站的分析设计中得到了广泛的重视,如近些年的信息系统采纳研究,围绕 TAM

等模型以电子商务为主要研究对象,大量的实证研究都将愉悦认知作为其重要的构念。在学术信息检索系统中,该概念还没有得到应有的重视。在充分注重信息内容的基础上,尽可能地提供能够给信息用户带来愉悦感觉的系统特征,显然是必要的。

5 研究结论

针对相关性判据研究的两个问题,下面分别讨论研究的结论。

5.1 相关性判据小结

针对第一个研究问题:相关性的判据问题,本文通过内容分析法,在学生数据的基础上分析出了传播特征、内容、情境、使用、系统特征、愉悦、质量、总体以及文献特征等 9 类相关性判据,表 10 概括了不同类别相关性判据的分布情况。由表 10 可见,文献特征、质量判据以及内容判据、总体和使用占据了相关性判据的主要位置,相对不是太重要的则是愉悦、情境、传播特征以及系统特征。这些提示在检索系统的分析与设计中,对于文献特征以及质量判据需要给予突出的关注。对于其他维度的相关性判据也需要更多地从发展的眼光做更多的思考,比如愉悦认知相关性判据,在目前的学术信息检索系统中还没有得到应有的重视,但是在商务型网站的分析与设计中就已经占有非常重要的位置。

5.2 影响相关性判断的文献特征小结

针对相关性判据研究的第二个问题,影响用户相关性判断的文献特征,本研究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、作者以及机构等在内的 15 个文献特征相关性判据子类(表 4、表 5、表 11)。其中文献类型、来源期刊、作者以及机构又包括更加具体的类别(表 5、表 11)。表 4、表 5 和表 11 的数据表明,来源期刊、作者以及机构的声誉显著地影响信息用户的相关性判断,因此检索系统的分析与设计者应该充分将这些因素纳入检索结果的排序、查询扩展以及相关反馈等有利于改进检索系统质量的功能中。

表10 总体类别相关性判据

相关性判据大类	学生数	相关性判据子类	频次	相关性判据大类	学生数	相关性判据子类	频次
内容	10	比较分析	10	总体	5	查全率	5
	2	方法	2		4	查准率	4
	4	技术实现	4		3	存取	3
	3	类别	3		24	数量	26
	9	实践	11		2	顺序	2
	4	实验	5		3	搜索引擎	3
	2	数据	2		14	同质性	20
	18	图解	19		2	外部证实	2
	15	专题性	17		3	文献计量	4
	40	综合性	51		7	学术数据库	7
	107	小计	124		2	质量	2
情境	5	时间限制	5	文献特征	69	小计	78
	5	小计	5		7	参考文献	7
系统特征	21	存取	24		4	出版单位	4
	1	响应速度	1		13	出版时间	14
	1	著录详细	1		16	关键词	16
	23	小计	26		3	基金	4
愉悦	2	情感	3		11	篇幅	11
	2	小计	3		20	全文	21
质量	40	广度	51		35	题名	36
	27	简明性	39		2	文献格式	2
	7	结构清晰	7		43	文摘	44
	12	可靠性	12		15	语种	17
	43	可理解性	54		103	文献类型	132
	1	美感	1		26	来源期刊	31
	44	权威性	52		28	作者	34
	24	深度	27		23	机构	33
	13	实时性	17		349	小计	406
	7	系统性	7	传播特征	2	点击率	2
	11	先进性	12		6	引用	6
	64	详细	75		1	转载	1
	35	新颖性	38		9	小计	9
	20	学术性	24	使用	2	读者对象	2
	2	知识水平	2		13	架构借鉴	22
	1	直觉	1		2	潜在利用	4
	79	主题	81		23	提供背景知识	34
	16	专指性	17		21	有效性	24
	8	准确性	9		61	小计	86
	454	小计	526				

表 11 文献特征相关性判据的频次分析

相关性判据	学生数	相关性判据子类	频次	相关性判据	学生数	相关性判据子类	频次
作者	1	高产	2	机构	1	高产	1
	7	客观	7		4	客观	4
	2	群体	2		18	声誉	28
	10	声誉	13		23	小计	33
	3	研究领域	4	来源期刊	26	声誉	31
	5	职称	6		26	小计	31
	28	小计	34				

致谢:relevance criteria 原先被译为“相关性标准”,根据马费成教授的指导,本文将其翻译为“相关性判据”,特此致谢!文责自负。

参 考 文 献

- [1] Saracevic T. Relevance: A review of and a framework for the thinking on the notion in information science [J]. Journal of the American Society for Information Science, 1975, 26(6): 321-343.
- [2] Saracevic T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance [J]. Journal of the American Society for Information Science and Technology, 2007, 58(3): 1915-1933.
- [3] Saracevic T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance [J]. Journal of the American Society for Information Science and Technology, 2007, 58(13): 2126-2144.
- [4] Schamber L. Relevance and information behavior [J]. Annual Review of Information Science and Technology, 1994, 29: 3-48.
- [5] Barry C L. The identification of user relevance criteria and document characteristics: Beyond the topical approach to information retrieval [D]. Syracuse: Syracuse University, 1993.
- [6] 马费成. 论情报学的基本原理及理论体系构建 [J]. 情报学报, 2007, 26(1): 3-13.
- [7] 梁战平. 我国科技情报研究的探索与发展 [J]. 情报探索, 2007(7): 3-7.
- [8] 张新民, 化柏林, 罗卫东. 相关性研究探析 [J]. 情报学报, 2008, 27(5): 691-697.
- [9] 屈鹏, 吴龙婷, 隆捷, 等. 国际情报学研究主题的聚类分析——基于 1996~2003 年的 LISA 数据库 [J]. 情报学报, 2007, 26(6): 909-917.
- [10] Salton G. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer [M]. Boston: Addison-Wesley, 1989.
- [11] Cuadra C A, Katter R V. Experimental studies of relevance judgements: Final report [R]. Santa Monica: System Development Corp., 1967.
- [12] Rees A M, Schultz D G. A field experimental approach to the study of relevance assessments in relation to document searching (2 vols., NSF Contract No. C-423) [R]. Cleveland: Case Western Reserve University, 1967.
- [13] Nilan M S, Peek R P, Snyder H W. A methodology for tapping user evaluation behaviors: An exploration of users' strategy, source and information evaluating [A] // Proceedings of the American Society for Information Science [C]. Medford: Learned Information, 1988, 152-159.
- [14] Schamber L. Users' criteria for evaluation in a multimedia information seeking and use situation [D]. Syracuse: Syracuse University, 1991.
- [15] Cool C, Belkin N J, Kantor P B. Characteristics of texts affecting relevance judgments [A] // Williams M E. Proceedings of the 14th National Online Meeting [C]. Medford: Learned Information, 1993: 77-84.
- [16] Schamber L, Bateman J. User criteria in relevance evaluation: Toward development of a measurement scale [A] // Proceedings of the American Society for Information Science [C]. Medford: Learned Information, 1996: 218-225.
- [17] Wang P. A cognitive model of document selection of real users of information retrieval systems [D]. College Park: University of Maryland, 1994.
- [18] Wang P, White M D. A Cognitive Model of Document Use during a Research Project. Study II. Decisions at the Reading and Citing Stages [J]. Journal of the American Society for Information Science, 1999, 50(2): 98-114.
- [19] Wang P, Soergel D. A cognitive model of document use

- during a research project. Study I: Document selection [J]. *Journal of the American Society for Information Science*, 1998, 49(2): 115-133.
- [20] Tang R, Solomon P. Toward an understanding of the dynamics of relevance judgment: An analysis of one person's search behavior [J]. *Information Processing & Management*, 1998, 34(2/3): 237-256.
- [21] Bateman J. Modeling changes in end-user relevance criteria: An information seeking study [D]. Denton: University of North Texas, 1998.
- [22] Maglaughlin K L, Sonnenwald D H. User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments [J]. *Journal of the American Society for Information Science and Technology*, 2002, 53(5): 327-342.
- [23] Crystal A, Greenberg J. Relevance Criteria Identified by Health Information Users During Web Searches [J]. *Journal of the American Society for Information Science and Technology*, 2006, 57(10): 1368-1382.
- [24] Xu Y, Chen Z. Relevance judgment: What do information users consider beyond topicality? [J]. *Journal of the American Society for Information Science and Technology*, 2006, 57(7): 961-973.
- [25] Xu Y. Relevance judgment in epistemic and hedonic information searches [J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(2): 179-189.
- [26] Eisenberg M, Barry C. Order effects: A study of the possible influences of presentation order on user judgment of document relevance [J]. *Journal of the American Society for Information Science*, 1988, 39(5): 293-300.
- [27] Janes J W, McKinney R. Relevance judgments of actual users and secondary judges: a comparative study [J]. *Library quarterly*, 1992, 62(2): 150-168.
- [28] Janes J W. Other people's judgments: A comparison of user's and other's judgments of document relevance, topicality, and utility [J]. *Journal of the American Society for Information Science*, 1994, 45(3): 160-171.
- [29] Spink A, Greisdorf H, Bateman J. Examining different regions of relevance: From highly relevant to not relevant [A]. *Proceedings of the American Society for Information Science [C]*. Medford: Learned Information, 1998, 3-12.
- [30] Zhang X M. Collaborative Relevance Judgment: A Group Consensus Method for Evaluating User Search Performance [J]. *Journal of the American Society for Information Science and Technology*, 2002, 53(3): 220-231.
- [31] Mizzaro S. Relevance: The whole history [J]. *Journal of the American Society for Information Science*, 1997, 48(9): 810-832.
- [32] 成颖, 孙建军, 宋玲丽. 相关性标准研究 [J]. *中国图书馆学报*, 2005, 31(3): 55-60.
- [33] 孙绍荣. 美国的文献相关性试验评介 [J]. *情报业务研究*, 1989(2): 137-140.
- [34] 孙绍荣. 论试验方法在情报学理论建设中的作用: 美国的相关性试验的启示 [J]. *情报理论与实践*, 1989, 12(2): 9-10.
- [35] 康耀红. 相关性及排序原则 [J]. *情报理论与实践*, 1990, 13(4): 7-9.
- [36] 李国秋, 吕斌. 检索相关性研究的发展 [J]. *情报理论与实践*, 1996, 19(2): 56-59.
- [37] 王家铨. 信息检索中相关性概念的研究 [J]. *现代外语*, 2001, 24(2): 181-191.
- [38] 成颖, 孙建军, 巢乃鹏. 信息检索中的相关性模型 [J]. *图书情报工作*, 2004, 48(12): 46-50.
- [39] 成颖, 孙建军. 信息检索中的相关性研究 [J]. *情报学报*, 2004, 23(6): 689-6965.
- [40] 于春, 彭爱东, 王波, 等. 信息用户对信息检索相关性判断的因素分析 [J]. *图书情报工作*, 2009, 53(3): 103-107.
- [41] 胡秀梅, 邓小昭. 文献呈现次序对用户相关性判断的影响 [J]. *图书情报工作*, 2010, 54(14): 32-35.
- [42] 李本乾. 描述传播内容特征 检验传播研究假设——内容分析法简介(下) [J]. *当代传播*, 2000(1): 47-49.
- [43] Taylor R S. Value-added processes in information systems [M]. Norwood: Ablex Publishing, 1986.

(责任编辑 化柏林)