

学术文献的非主题特征及其意义 *

邹永利

摘 要 主题检索一直是信息检索系统也是信息检索研究的重点,但文献的非主题特征的研究与利用没有受到足够的重视。文献的非主题特征指与文献相关,但与其主题内容或语义没有直接或必然联系的特征。笔者尝试较系统地归纳文献的非主题特征,并基于信息需求研究、信息查找与利用行为研究的有关理论,分析非主题特征对于信息检索理论与实践的意义,并通过基于非主题特征的网络中文学术文献识别系统(NSIRS)的设计与评价,验证其可行性与效果,分析了当前一些信息检索系统的非主题检索功能与特色。图4。表4。参考文献24。

关键词 学术文献 非主题特征 信息检索 信息需求 系统设计

分类号 G350

ABSTRACT Indexing and searching mechanism based mainly upon semantic attributes of documents has been showing certain limitations. The function of the non-topical attributes and their implication to information retrieval and system design are not yet fully explored. By summarizing the non-topical characteristics of documents and by analyzing them in light of information needs, information seeking and use studies, the author attempts to explore their potentialities to information retrieval and information system design. Based on some non-topical attributes, a system, NSIRS, was developed and evaluated in order to test the ideas brought forward in this article. Finally, the non-topical features embedded in some well-known information retrieval systems are discussed as well. 4 figs. 4 tabs. 24 refs.

KEY WORDS Academic documents. Non-topical attributes. Information retrieval. Information need. System design.

CLASS NUMBER G350

文献主题特征的抽取与表示,一直是图书情报学研究的重要课题,而非主题特征由于相对直观、简单、少有歧义,未能引起足够重视,系统的研究很少。反映到检索实践中,则是各类信息检索系统的非主题检索途径稀少单一、缺乏特色。对数据规模有限、受控程度高的传统信息检索系统而言,非主题检索途径的不足也许不会成为用户检索的严重障碍,但对于文献量庞大、受控程度低的网络信息检索来说,同样的问题却会妨碍检索的有效进行,因为非主题特征在排除不相关文献方面,效用至为突出,是主题特征无法取代的。近年,搜索引擎以及网络数据库检索系统都在改进其非主题检索功能,非主题检索途径的多寡、功能的强弱等,正在成为系统评价的重要指标^[1]。本文尝试较系统地考察文献的非主题特征,探讨其对于信息检索理论与实践的意义。

1 研究缘起——主题检索的局限

主题检索是对表示文献主题的标识与表示检索者信息需求主题的标识做相符性比较,两者一致,文献则被认为符合信息需求,检索目的达成。信息检索研究的主要课题也因此是:在文献存储过程中,设法更充分、准确地揭示和表达文献的主题概念以及概念之间的关系,并在用户检索过程中,支持其更准确、方便地表达信息需求概念以及概念之间的关系。无论是人或是机器,准确地标引文献主题是个挑战,用户准确、清晰地表达其信息需求,同样是个挑战。因而主题检索所做的相符性比较,实际上是两种原本就不准确的表达之间的比较。虽然全文检索技术在很大程度上解决了标引过程中主题词选择的难题,但主题检索仍然不能摆脱其固有局限。称其固有是因为它们与人类在知识的理

* 本文系国家社会科学基金项目“网络中文学术文献的自动识别与检索研究——基于学术文献文体、链接及图文相关度的研究与系统开发”(编号:10BTQ049)的研究成果之一。

解与表示方面存在的局限有关。

首先,文献中的用语难以全面、准确地表达相关主题的概念。全文检索方式以作者使用的词语作为衡量文献中所含信息的尺度。但文献中的用语只是作者在表达一种思想、概念、意向,即文献主题时,在多种可能的表示样式、语词符号集合中所做的一种或有限的选择。这种选择既受到诸如篇幅、修辞规律、表述工具等客观因素的制约,同时又受到作者的领域知识与表述能力的制约,若检索系统缺乏良好的补充性标引功能,难免导致检索的盲点。其次,主题概念的表示往往时间特定、地域特定、领域特定,同一概念可能采用不同的表达形式,有些表示可能已过时,不为今天的用户所悉知,以至难以从主题角度找出有关文献。第三,检索词或检索词的组配难以准确、完整地表达检索者的信息需求,尤其是占信息检索大多数的任务型(task-oriented)信息检索需求^[2]。有时检索者不仅不能用系统规定的语言、格式,清晰地予以表示,甚至用自然语言也难以完全表达清楚。然而,主题检索却以检索者能够贴切选择代表其信息需求的检索词为前提。

有人说,人们要询问的既不是他知道的,又不是他不知道的。如果他知道,则没有必要再问;如果他不知道,则他也不知道如何去问。此话大致道出从主题角度查找信息者的窘境:专业词汇不足(相对于信息需求主题),加上语法知识不足(相对于信息检索系统),以致表述的信息需求(检索式)相对于知识结构的缺陷,即真正的信息需求至多是近似的^[3]。

2 非主题特征

文献非主题特征指与文献相关,但与其主题内容或语义没有直接或必然联系的特征。表1—表3列举了文献的非主题特征,其来源主要有四个:一是创作过程,由著者生成;二是编辑出版过程,由编辑和出版者附加;三是信息服务机构,由信息工作者加工、处理文献时生成;四是用户利用过程,如利用记录、评价等。这样,围绕一份文献,逐渐形成多种类、丰富的非主题标志。

文献非主题特征可以概括为内部与外部两大类,其中内部特征又分为显性和隐性两类。

2.1 文献内部的非主题特征

文献内部的非主题特征指从文献中能够发现、文献自身具有的特征与属性,其中有些特征是显性的,可以直接被观察到,有些是隐性的,需经文献阅读、分析,才能予以发现。比如文献的文体特征、论点取向等,就属于这种性质的非主题特征。表1、表2列举了两类特征的主要类别与内容。

表1 文献的显性特征

类别	内容
文献形态	文本、图像、声音、视听、语种等
物理特征	载体、规格或开本、长度或播放时间
文献类型	综述、试验报告、教科书、会议论文等
出版事项	日期、版本、出版/发行者等
文献构成	篇名、摘要、关键词、目录、正文、图表、引用、注释、索引、附录、参考文献等
逻辑结构	序、先行研究、方法论、调查或推理、结论、谢辞等
外观特征	字号、字体、布局设计、色彩等
特殊符号	标点符号、运算符、方程式等
著述者	著者、译者、编辑、设计者等,音像文献的指挥、导演、表演者、演奏者等
特殊编号	ISBN、ISSN 等

表2 文献的隐性特征

类别	内容
研究类型	基础研究、应用研究、综合研究等
研究方法	实验、实证、观察、逻辑等
论点取向	证伪研究、证实研究等
内容深度	面向读者对象,或理解文献所需知识水平
文体特征	特有表述或典型句型、所含语种数量及比例、平均句长等
引用信息	对特定著者或文献的引用次数、在文献什么部分引用等

2.2 文献外部的非主题特征

文献外部的非主题特征是指文献中可能不存在,或文献本身不具有的特征,具体内容如表3所示。

表3 文献外部的非主题特征

类别	内容
著述者背景	所属机构、职业、职称、学位、学术流派及社会、历史或政治环境等
利用状况	被引用数、借阅次数、点击次数、下载次数等
利用者评价	用户对文献质量或价值的明确评分、评级、评论等
可获得性	费用、传递时间等

3 非主题特征的意义

我们将从信息需求的性质、排除检索、特定信息定位等几个方面,分析非主题特征的理论与实践意义。

3.1 有助于从理论上解释信息需求的矛盾性质

信息需求一方面模糊、笼统,难以用适宜、明确的词语予以描述;另一方面又十分具体、个别,非特定信息不能满足。文献的非主题特征有助于我们理解信息需求这种看似矛盾的两面性。

关于信息需求的模糊性,几种有影响力的理论都有明确论述。Taylor 视信息需求为一种不完整性(incompleteness),存在于提问者的世界映像中,不易明确定义,但阻碍其与周围世界做有目的的互动^[4];Dervin 从“意义构建”理论出发,将信息需求定义为知识结构的缺断(gap),以致用户的行动受阻,产生构建新情景意义(sense)的需要^[5];情报学认知学派的代表人物 Belkin 则认为,信息需求是知识结构不充分、不正常的表现,具有信息需求意味着具有

“异常知识状态”(Anomalous State of Knowledge)^[6]。概括起来,信息需求是知识结构的一种缺陷,难以清晰、准确地表述。

吊诡的是,当一份文献或信息呈示在用户面前时,用户却能清楚地指出其是否满足所需,显示信息需求又是明确、具体的。在相关性判断研究^[7-12]确认的各种影响因素中,除了主题、论题相关因素外,另有许多因素是非主题性质的(表4)。正是这些要求明确的非主题因素或条件,决定了信息需求的另一面——明确、特定。满足信息需求的文献除须具备特定的主题特征外,还须具备特定的非主题特征,这是由知识异常状态、认知缺断或知识结构缺陷的个别性所决定的。这种个别性也被称作特殊的“个人建构”(personal construct),左右着用户对文献相关性的判断^[13]。

表4 相关性判断中的非主题性条件

序号	内容
1	方法、视点与立场
2	深度、质量、类型或面向知识程度
3	信息源类型与性质
4	地域性
5	是否具体提出了一种有效的理论或方法
6	作为物理实体的文献(媒介)
7	与该领域范式理论的关系
8	用户与著者之间的关系或距离
9	检索目的
10	研究背景
11	发行机构
12	文献长度
13	著者知名度、地位、职业、学位、所属机构
14	著者对论题采取的立场、价值标准、与该领域理论范式的一致性
15	可信度
16	著述目的
17	可获得性
	……

认识信息需求的非主题层面及其性质,改进信息检索系统设计,将有助于用户更准确、清晰地表述信息需求。在信息需求的两个层面中(见图1),主题层面不易被准确表述,而非主题

层面却是用户能够清楚表述,并影响其相关性判断的。在信息检索过程中,向用户提供更丰富的非主题检索手段,无疑有助于其更方便、准确地表述信息需求,获得所需信息。

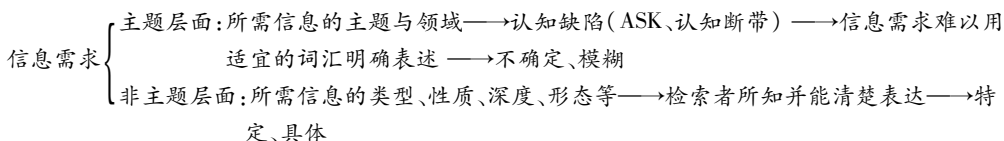


图1 信息需求的两面性

3.2 为系统拓展排除检索功能创造了条件

Marchionini 指出,信息查找与其说是检索相关信息的问题,不如说是排除不相关信息的问题^[14]。Belkin 则称检索和排除是同一硬币的两个不同面,抽象层次上没有什么区别^[15]。Ellis 认为,用户经常是将计算机信息检索系统作为迅速排除或过滤大量无关信息的手段,而不是作为一种能准确找出所需信息的可靠手段^[11]。一些著名的信息查找行为模型^[11,16]显示,人们经常以文献的非主题特征启动检索,并频繁利用这类特征实施对检索的控制,正如 Cool 所说,人们趋近信息依赖的是文本相对信息目标的多种属性^[9]。适宜的排除检索条件(filter)有助于用户结合主题检索,排除大量不相关文献,缩小与目标文献之间的距离。信息需求与文献均具有非主题特征,这为信息系统设计非主题性的排除检索功能提供了依据。

一些结构化数据库检索系统允许用户以文献类型、研究类型、研究深度等作为排除检索的条件,排除或限定检索对象,提高检索的专指度,但总体而言,对非主题特征的利用还不够充分。如前所述,文献的非主题特征至少有 20 个种类 80 余项内容,不少特征可以用来增强系统的非主题搜索功能,排除不相关文献。比如文献的性质,便可作如下区分:是对现有理论方法的证明、

评价,还是具体理论方法的提出;研究方法是定性还是定量;研究目的是证实还是证伪等等。文献的性质既可由人工以定性方法标明,亦可由系统根据文献的一些非主题特征等,以定量方法确定。比如引用文献数与被引用次数可以用来判断文献的某些性质(基础研究、应用研究、研究深度等),图表、算式数可用于判断文献采用的研究方法,利用状况可用于判断文献质量等。此外,作者的背景,如所属机构、学术团体、职业、职称、学位等,亦可用于区分不同研究性质与质量的文献、跟踪同行研究机构的研究成果与研究动向,获取具有某种学术倾向、立场或观点、学术流派的文献等,Ingwersen 将这样的检索称作对一个学科领域“垂直知识结构”的获取^[17]。

3.3 有助于系统更精确地定位特定类型信息

Kircz 将学术文献的读者划分为四种主要类型:非读者(non-readers)、专家读者(informed readers)、半专家读者(partially informed readers)和非专家读者(uninformed readers)^[18]。四类读者事实上都仅利用一份文献极有限和很特定的内容。每一个用户依据其研究进展阶段、涉及领域或教学需要,在不同时候会扮演不同的角色^①,印刷型文献的结构和阅读训练使他们能迅速确定所需信息的所在。关于文献结构的调查

① 非读者多从事管理工作,通常只浏览书目信息;专家读者与文献著者同属一个研究领域,且熟悉这一领域,一般只翻阅有限的几种杂志,且仅利用其中极核心的信息,几乎无人通读论文全文;半专家读者,或是新近涉足该领域的研究者,或是周边学科的研究者,通常只阅读文献的特定部分,重在了解总体观点、结论,或建立与已知文献的联系;非专家读者只阅读文献的绪论、总体性图表、结论以及参考文献部分,以一般性获知为目的。

也显示,文献创作者虽然由于知识背景的不同,在遣词造句方面存在程度不同的差异,但共通的教育、科研训练等,使他们在介绍、描述一个试验,或建立、评论一种理论与方法时,所采用的逻辑结构、展开方式、叙述风格等非常相似,标准化、模式化程度极高。具备一定教育背景的读者之所以无须通阅全文,就能从长篇大论中迅速获取所需信息或高速浏览文献,正是熟悉和利用了文献的这种固定结构与修辞特征。Allen 基于问题解决模式对信息需求的研究揭示,不同类型、不同问题阶段的信息需求,往往需要用文献不同结构部分所含的信息予以满足^[19]。这意味着,检索式是用户针对特定文献结构部分,而非文献整体所具有的语义特征的提问,系统应该提供相应的支持,提高检索的针对性与深度。比如支持用户将检索词出现的位置限定在具有某种特征、标志的文本中,如脚注、图表、算式、结构图、流程图等;其次,用户应能将检索词出现的位置限定在文献的特定结构部分,如背景、假设、试验环境、方法、结论等部分。

应该指出的是,虽然典型科技文献的结构(如导论、问题提出、研究方法、试验或实证、结果分析、研究结论等)较为固定,但这种结构不一定总是显性的,在有些文献中,它可能以隐性方式存在。这种隐性结构,也称作修辞或论辩性(rhetorical or argumentational)结构,亦可用来支持检索,比如允许用户将检索词出现的位置限定于文献中具有特定修辞性特征的文本。Kircz 对科技文献中的修辞性结构作了分类,希望通过对各类别下的典型、惯用的修辞性结构的搜集、确认,使系统能够以之为基础,自动识别和标引相应的文本^[18,20]。例如“new approach/method/methodology”、“advocated here”、“solution to the problem”、“contrary to conventional wisdom”、“in this paper……different approach”等表述,将归入“研究方法”修辞性结构类下。当用户针对于“研究方法”类下的修辞性表述,信息定位能力藉此深入到文献的隐性结构部分。

4 当前信息检索工具的非主题检索功能

虽说文献的非主题特征尚未受到足够重视,但仍有一些信息检索系统提供了富有创意的非主题检索功能。

在网页搜索方面,不少搜索引擎近年都在强化非主题检索功能方面下了功夫,可供用户利用的非主题检索途径亦逐渐丰富起来。目前各主要搜索引擎基本上都允许使用日期、文献类型、语种、域名、国家或地区、搜索词出现位置、相关链接等非主题搜索途径,对搜索进行所需的限定。有些搜索引擎还支持更复杂、高级的非主题途径限定,比如目录深度、网页深度、文件大小、搜索词大小写、搜索词相邻度、搜索词之间相对位置等等。AltaVista 曾是这一领域的佼佼者,几乎支持这里罗列的所有功能,被认为是搜索网上科技文章的最佳通用搜索引擎。遗憾的是,它于 2004 年 5 月底被 Yahoo 吞并,其网络数据库也被 Yahoo 的数据库取而代之,包括搜索词相邻度、相对位置、相似性检索(anchor、applet、like)等在内的许多实用的非主题搜索限定功能不见了,不能不说是广大网民,尤其是科技网民的一大损失。

结构化数据库检索系统对文献非主题特征的利用相对容易,因而提供的相关功能也多一些,一般都支持对文档结构(篇名、关键词、摘要、参考文献、全文等)、创作者属性(作者、第一作者、单位等)、出版事项(刊名、年、期、时间等)、编码(中图分类号、ISSN、统一刊号)等项目的限定检索。大多数系统支持以出版物为单位的浏览、引用跳转浏览,部分系统甚至可以就特定领域推荐核心信息源、著者,排除检索条件也更为多样化,例如 EBSCOhost、ProQuest、Emerald、EI 等都将文献类型等作了进一步细分。ProQuest 以及旧版的 Emerald 在文献非主题特征的利用上尤具创意,各有特色。

ProQuest 设有文献特征(document feature)和文献类型(document type)选项。文献特征包括图、示意图、雕版图、方程式、插图、地图、照片、

彩页插图、表等 9 种类型的图表(见图 2),用户可根据需要,将检索限定在包含或不包含某种特定类型图表的文献;文献类型分为年度报告、图书、个案研究、评论、封面报道、社论、演讲、统计等,总共有 23 个大类、10 个小类,用来限定检索范围。Emerald 的内容指示(content indicator)与文献类型(document type)选项提供了丰富的非主题检索条件。前者允许用户对检索对象的创新性、实践意义、可读性以及研究意义作三个层次或水平的限定,以满足不同层次、不同检索目的用户的多样化需要;后者允许用户将文献类型细分为个案研究、比较/评价、概念解释、一般综述、报道、文献综述、研究论文、述评、调查、技术报告、技术论文、具有实践应用的理论研究、具有可行例子的理论研究、观点、纯理论等 15 个类型(见图 3)。

遗憾的是,新版的 Emerald 仅提供文章类型(Article type)选项,其他特色都不见了。

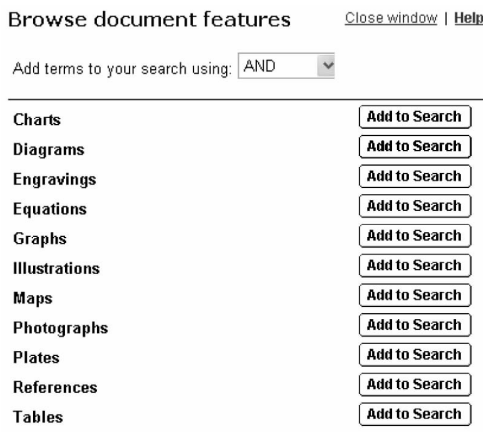


图 2 ProQuest 的文献特征选项

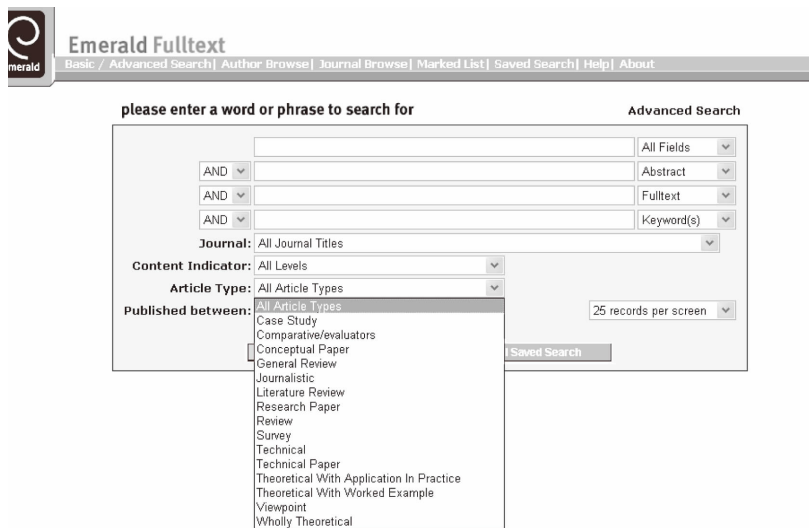


图 3 Emerald 的文献类型复选框

5 实验研究——非主题特征与学术文献的自动识别

为检验非主题特征的实践意义,我们以 VC + 6.0 为开发平台,采用 ACCESS2000 后台数据库技术,设计了一个基于文献非主题特征的网络中文学术文献识别系统(NSIRS)。NSIRS 旨在解决通用搜索引擎在检索网络中文学术文

献时查准率低的问题。虽然 Elsevier Science 公司 2001 年推出 Scirus、Google 公司 2004 年推出 Google Scholar 等学术文献搜索引擎,但在搜索网络中文学术文献方面,均存在局限:Scirus 的中文信息覆盖率低,而 Google Scholar 的搜索范围则主要是结构化数据库中正式发表的文献,对非正式发表的学术文献,覆盖率低,且对文件的大小作出限制,超过限制的文档将被排除在外^[21]。

NSIRS 实际上是一个学术文献的二次排序系统,由三个功能模块组成:一次检索预处理、非主题特征分析与处理、结果处理等,运行流程如图4所示。用户首先在通用搜索引擎(Google)

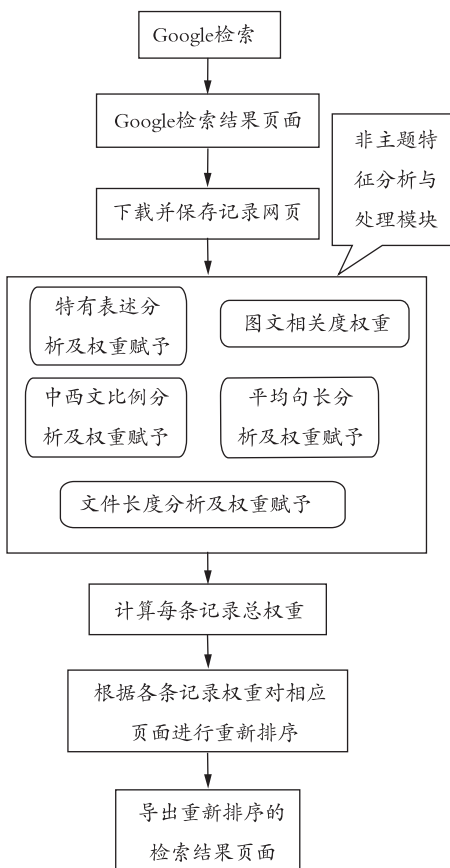


图4 NSIRS 工作流程

上执行检索,返回的结果网页被下载到临时数据库中;非主题特征分析与处理模块对特有表述、平均句长、图文相关度、中西文比例以及文件长度等进行分析,并根据分析结果赋予各记录相应的权重。特有表述是指学术文献中非主题性的高频词语,分为结构性表述与修辞性表述两类。文献的结构性表述如序言、导论、研究

背景、目的、意义、方法(论)、文献综述、结论、研究展望等;修辞性表述如基于、探讨、旨在、分析、系统、研究、模型等。每种特有表述对应系统预设的权重,其出现频次越高,文献的权重也就越高。图文相关度指文献所在网页上的图像是文献的相关图像,还是广告、按钮等不相关图像。借助图像图表说明研究过程及研究结果是学术文献的一大特色,其常用图形就有流程图、曲线图、柱形图等,文中所含图形的类型甚至可以用来确定文献的类型^[22]。因此,网页的图文相关度越高,越有可能是学术文献;同样,文献的平均句长越长、西文字符比例越高、文件越大,越有可能为学术文献,被赋予的相应权重也就越高^①。最后,根据各记录所得到的总权值,对各条记录进行重新排序,并以 Google 网页格式导出重新排序的检索结果。

基于“前 X 命中记录查准率”方法^[23-24]所作评价显示,NSIRS 在识别网络中文学术文献方面,明显优于 Google 通用搜索引擎。我们就 10 个不同主题进行检索实验,计算并比较 Google 与 NSIRS 的前 30 条记录的相对查准率。实验显示,单独使用特有表述、句子平均长度、图文相关度、中西文比例、文件长度等任一指标进行检索,前 30 条记录的查准率基本上都高于 Google 原有的查准率;当综合使用五项指标时,NSIRS 的平均查准率为 48.2%,高于 Google 的 35.7%,显示文体特征能够有效地用于网络中文学术文献的识别与检索。NSIRS 尚不是一个独立的网络中文学术文献自动识别与收集软件,所建文献特征体系、识别参数体系等也远不够完善。

6 结语

以上我们探讨了学术文献的非主题特征及其意义,除了需要在理论上作更深入、系统的研究整理、补充外,其实践意义也有待作进一步、更全面的检验,NSIRS 的应用实践只是一项初步探索。中文学术文献的非主题特征十分丰富,

① 由于学术文献要求准确、严谨地表述概念之间的关系,致使句子的容量大,长句的使用率高于其他类型文献。我们比较了学术文献与新闻报道的平均句长、西文字符比例,发现前者的平均句长为 72.68 字符,后者为 40.90 字符(均包含西文字符在内),西文字符的比例则分别是 25% 与 2%。

其逻辑关系的发现与表述也因而更为复杂,需要不断地去找出并测试新的表述规则。NSIRS的规则扩展及聚类等功能还较薄弱,我们计划在原系统基础上,建立更规范的规则库及聚类功能,比如以正则表达式作为规则表述工具,建立“学术文献非主题特征正则表达式库”。希望通过这样的探索,使文献的非主题特征之于图书情报学知识体系的意义更为明确。

参考文献:

- [1] Notess G R. Search engines by search features [OL]. [2004-04-07]. <http://www.searchengineshowdown.com/features/byfeature.shtml>.
- [2] Vakkari P. Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval[J]. *Information Processing and Management*, 1999 (35), 819-837.
- [3] [日]细野公男, 邹永利. 情报需求及其表述中的知识结构与情报系统设计[J]. *图书情报工作*, 1999, (1): 1-7.
- [4] Taylor R S. Question negotiation and information seeking in libraries[J]. *College and Research Libraries*, 1968 (29): 178-194.
- [5] Dervin B. From the mind's eye of the user: The sense-making qualitative-quantitative methodology [G]//D. Glazier, R. Powell. *Qualitative Research in Information Management*. Englewood Colo: Libraries Unlimited, 1992: 61-84.
- [6] Belkin N J, Oddy R, Brooks H. ASK for information retrieval [J]. *Journal of Documentation*, 1982 (38): 61-71.
- [7] Barry C L. User-defined relevance criteria: An exploratory study[J]. *Journal of the American Society for Information Science*, 1994, 45 (3): 149-159.
- [8] Park T K. The nature of relevance in information retrieval: An empirical study[D]. Bloomington: Indiana University, 1992.
- [9] Cool C, Belkin N, Kantor P. Characteristics of texts affecting relevance judgments [G]//14th National Online Meeting. Medford: Learned Information, 1993: 77-84.
- [10] Howard D L. Pertinence as reflected in personal constructs[J]. *Journal of the American Society for Information Science*, 1994, 45 (3): 172-185.
- [11] Ellis D. A behavioral approach to information retrieval system design[J]. *Journal of Documentation*, 1989, 45 (3): 171-212.
- [12] Wang P L, Soergel D. A cognitive model of document use during a research project: Study 1. Document selection[J]. *Journal of the American Society for Information Science*, 1998, 49 (2): 115-133.
- [13] Kelly G. A theory of personality: The psychology of personal constructs[M]. New York: Norton, 1963.
- [14] Marchionini G. Information seeking in electronic environments[M]. London: Cambridge University Press, 1995.
- [15] Belkin N J, Croft W B. Information filtering and information retrieval: Two sides of the same coin? [J]. *Communications of the ACM*, 1992, 35 (12): 29-37.
- [16] Bates M. The design of browsing and berry-picking techniques for the online search interface [J]. *Online Review*, 1989, 13 (5): 407-424.
- [17] Ingwersen P. Information retrieval interaction[M]. London: Taylor Graham, 1992.
- [18] Kircz J G. Rhetorical structure of scientific articles: The case for argumentational analysis in information retrieval [J]. *Journal of Documentation*, 1991, 47 (4): 354-372.
- [19] Allen B L. Information task: Toward a user-centered approach to information systems[M]. San Diego: Academic Press, 1996.
- [20] Kircz J G. Modularity: The next form of scientific information presentation? [J]. *Journal of Documentation*, 1998, 54 (2): 210 - 235.
- [21] Jacso P. Google scholar: The pros and the cons [J]. *Online Information Review*, 2005, 29 (2): 208-214.
- [22] Nickerson J V. The meaning of arrows: Diagrams and other facets in System Sciences Literature [C]// The Proceedings of the Hawaii International Conference on System Sciences, Big Island, Hawaii, January 3-6, 2005: 321-331.
- [23] 韩圣龙, 赖茂生. 网络信息检索工具评价实验——英文网络检索工具评价实验[J]. *情报科学*, 2001 (3): 293-301.
- [24] Leighton H V, Srivastava J. First 20 Precision among World Wide Web search services (Search engines) [J]. *Journal of the American Society for Information Science*, 1999, 50 (10): 870-881.

邹永利 中山大学资讯管理学院副教授, 博士。
通讯地址: 广州中山大学资讯管理学院。邮编: 510275。

(收稿日期: 2010-12-03; 修回日期: 2010-12-31)