

●焦玉英

从文献检索到信息检索

——网络环境下检索课教学内容与方法的调整

摘 要 从文献检索到信息检索最大的变化:一是由文献单元向信息单元为基础的组织方式的改变;二是由手工分类、主题标引、著者标引经过机器的主题词、自由词抽取、标引发展到全文标引乃至超文本检索。网络技术、超媒体技术和智能技术等是促其变化的关键。作为一门学科的教学必须创建以CAI课件为主导的实践教学方法和建立信息检索课程的基本框架体系。参考文献4。

关键词 文献检索 信息检索 网络信息 教学内容与方法

分类号 G354

ABSTRACT Among the major changes from document retrieval to information retrieval are: 1) from document-unit-based organization pattern to information-unit-based one, 2) from manual classification, subject heading and author heading to machine-indexed subject terms and free words, full-text indexing and hyper-text retrieval. Network and intelligence technologies are catalysts for the changes. For the reform of the education in this field, we should create a CAI-based practical teaching method and establish a framework for the courses of information retrieval. 4 refs.

KEY WORDS Document retrieval. Information retrieval. Network information. Teaching methods.

CLASS NUMBER G354

在信息处理与传递的电子化与网络化进程中,浩瀚的知识信息海洋,快速、便捷、高功能、简便易用的多元化网络信息查寻工具,形成了世界范围内真正意义上的资源共享格局。据称,世界上最大的因特网能为1亿多用户提供50多万个科研机构、学校、图书馆和信息系统的网络资源;WWW覆盖了2000万个大小不等的信息源;预计下个世纪初将有100万个网络,1亿台计算机和10亿用户入网。全球网络化的大趋势使信息系统及服务方式面临严峻的挑战:网上“浏览—检索”对传统的“提问—检索”服务模式的替代使专职检索代理机构的业务逐步缩小;终端用户对系统的直接介入带来用户成分的巨大变化;系统为用户提供的在线即时获取方式,自组织信息功能使一种以“人为中心”的信息服务成为主流模式等。在这种环境下,以文献单元描述体结构为基础,手工检索方式为主导的传统文献检索向以信息单元组织结构为基础、网上浏览式信息查询方向发展是必然的选择。

本文主要针对网络信息资源的特点,在对传统文献检索教学模式总结的基础上,探讨作为一门学科的信息检索教学内容和方法。

1 文献检索教学模式回顾

传统文献检索系统由存贮与检索两个核心部分组成。存贮是选择文献、按规范化语言文本描述文献内外特征并使其有序化;检索是系统根据用户提问按规范化语言文本进行概念转换,经逻辑匹配输出与提问相关的文献。这种传统的“提问—检索”模式服务系统在运行的几十年中,一直影响着文献检索这门课并构成其教学的基本框架。其中,基本理论要素、结构要素与经验要素是这一框架的最主要内容。这3个基本要素构成了传统文献检索系统“提问—检索”的基本模式。随着数字化、网络化技术的迅速发展,这种教学模式已经基本上完成了历史使命。

1.1 基本理论要素

以文献单元为核心的描述体、规范化文本检索语言、检索工具及检索策略等是文献检索课程基本理论要素的重要组成部分。将其分解如下:(1)以知识为核心的文献生产域到知识的替代、重组、综合为目标的文献加工服务域,进而到文献利用域的发展过程中,文献情报流的动态特征及其内在规律;(2)以文献线索为核心的文本检索工具的基本组成结构、载体类型、出版形式、职能及评价指标体系;(3)检索语言典据文本的结构、原理及其功能;(4)各种类型索引的生成原理、特征及使用方法;(5)以经验为基础的手工检索策略分析、检索效果评价准则。

1.2 结构要素

结构要素侧重于文献检索实践教学部分。具体内容包括:(1)检索系统及其各子系统的结构要素及功能,数据库文档格式、结构及分析与评价;(2)索引系统款目的结构与评价;(3)国内外各种类型文献检索工具的结构体例、组成要素及各要素之间的相互作用关系分析。

1.3 经验要素

文献检索的实质是用户提问与系统文献标识之间的“匹配”。其检索效果不仅仅取决于系统本身的文献覆盖面、文献摘贮率、索引系统的完备程度等因素,也与检索提问的策略密切相关。在传统的手工检索条件下经验要素占主导地位。内容包括:(1)对检索课题的分析理解;(2)对检索语言语义确切性及相关性的判断及灵活运用;(3)对各类检索工具的专业适用范围、特点、检索途径与检索方法的灵活掌握;(4)对各种检索策略的了解及其配合使用等。

2 网络信息资源特点及其对信息检索的要求

2.1 网络信息资源特点

以因特网为核心连接起来的全球计算机网络是对传统的相对集中和规范为基准的文献数据库及其检索系统的严峻挑战。主要体现在下列几个方面:(1)内容覆盖面广:涉及科技、经济、军事、政治、贸易、商业、金融、教育、娱乐、人文社会科学各个领域;(2)信息浩瀚几乎没有什么范围限制;(3)多媒体化:多媒体技术的应用使静、动态图形(像)、音频及超声频特征的信息得以经过数字化处理后以位串形式存放,采用点、线、区域、关联、形状等多种方式再现出来;(4)个性化:网络技术赋予了每一个用户“我为人人,人人为我”的得天独厚的环境。网上客户通讯、聊天、个人简言、时尚、即时接触、即时援助、个人邮件传递等成

为网上信息资源包罗万象的表征;(5)超时空性:网络化与数字化技术使分布在世界各地主机上的信息资源联为一体,实现了跨时空、跨行业的高效快速传递;(6)动态性与开放性:以因特网为主流的国际信息资源网将数以千万计的主页以超文本链接方式构成一个国际化知识信息平台,为即时的节点加入或删除敞开方便之门;(7)可获取性:网络信息查寻工具有数百个,不少工具拥有百万乃至上千万网页数据,完备的索引系统,提供在线帮助和多功能的指引方式,使用户甚至不进行专门培训也可以从网上获取各种信息。

2.2 网络环境对作为一门学科的信息检索的要求

网络技术、超媒体技术甚至智能技术,不仅创新了计算机的硬件环境,而且改变了传统的信息服务方式。但从信息查寻的角度看,它在词语的选择、查寻的策略,甚至索引工具的类型等方面与传统检索方式既有差异,也存在某种程度的共性。因此,笔者认为,对作为一门学科的信息检索教学内容与方法的改革而言,必须对下列一些关键问题作出新的选择。

2.2.1 支持提问式及浏览式并重的检索语言

检索实践表明,规范化文本检索语言由于建立在统计基础上,以关键词匹配为核心,因而存在专业性强、标检不一致、检索提问不确切、语义逻辑歧义等问题。但它在支持提问式检索服务方面发挥了重要作用。网上浏览式查寻中以词语或分类作为入口仍十分畅行,至少在未出现理想化的后控语言之前,“提问—检索”和“浏览—检索”并用的检索语言仍十分必要。目前主要采用:

(1)改进型的检索语言。即一种主题语言与分类语言兼容互换的主题—分类一体化系统。如英国《科学文摘》主题指南,美国《生物学文摘》的主题—分类表,《分面叙词表:工程及相关学科叙词表》、《中国图书分类主题词表》等,今后仍应作为信息检索语言教学的一个重要方面。

(2)混合型检索语言。即受控语言和自然语言兼容并用的语言。这种语言从系统内部而言是高度规范化的,对系统外部用户接口而言又是高度自由化的。网络环境下这两方面检索语言的结合是发展方向。这是因为,虽然受控语言规范程度高,有片面强调专业用语而忽视人的因素的弊端,但在以提问为出发点,按用户特定检索命令检索方面的作用却不可忽视。而自然语言以计算机推理生成人们日常用语为基础。它不依赖于人对规范化语言的标引技巧或掌握的熟练程度,是支持浏览式检索的重要举措。目

前,自然语言处理技术已大量用于全文检索系统,在自动标引、理解检索要求、概念检索等方面效果显著。但作为检索语言,它在句法、语义分析、段落及篇章的深层次揭示方面存在的问题仍在进一步研究。

(3)网络环境下检索语言研究的趋向。有关检索语言研究将主要集中于:保证网络状态下多数据库查询中的词语的兼容并用;增加网络状态下自由关键词限定字段,提高检准率;图示词表与词间关系显示的并行化;自然语言检索提问中相关词语下的文献显示;能向用户提供透明窗口,集标引、检索、用户于一体的检索语言。

2.2.2 文本检索工具与网络查寻工具的配合

以因特网为代表的多元化网络信息查寻工具为用户提供高效、快速的全方位信息服务是不争的事实。但在利用网络查寻工具的同时,还不能完全摒弃文本式检索工具,两者在类型划分、特征、功能方面仍可以相互配合。信息检索教学中应注意对它们的综合与比较。如:(1)文本式等级制分类语言的结构与超文本链接式构成的信息链树状等级结构的分析、比较;(2)文本式检索工具中关键词索引系统与网络查寻工具常用的基于自然语言的简单关键词搜索方式的分析、比较;(3)网络字典式查寻工具与传统的事实、数据型检索工具的分析、比较;(4)网络索引式查寻工具与传统索引工具的分析、比较;(5)网上交互式查寻工具的主要特征与功能分析;(6)万维网基于推(push)与拉(pull)技术的多媒体浏览式工具的特征分析及利用策略等。

2.2.3 检索策略的共性与差异分析

传统检索策略的运用主要基于直觉和经验而不是逻辑思维。网络技术发展很快,在人们未开发出真正具有高级智能的网络信息自动搜索工具之前,依靠用户自身的本领提高检索效率仍十分重要。因为,所有的信息检索系统(包括光盘、联机系统在内)在检索策略方面都有不同程度的共性和难以确切区分的差异。如传统检索与网络查寻均采用布尔逻辑策略,字段、范围限定及组配策略,相邻检索与位置检索策略。只是在语义、句法关系规定的符号系统方面各检索系统自成体系。教学中应归纳其共性,强调其差异性。

另外,对于网络查寻中基于菜单式的查寻(如Gophor, wais)策略和基于超文本链式查寻和万维网浏览技术中的推拉技巧,以及扩大网络以外的信息资源关系,疏通信息资源渠道的策略等也是教学内容的重

要组成部分。

3 创建以CAI课件为主导的实践教学方法

传统文本检索工具体积大,更替难,翻阅费时费力。使用光盘技术和联机教学实习又只能解决快速查寻问题,并不能全面帮助学生理解和综合运用各种检索基本技巧。因而,CAI课件应作为信息检索这门课程实践教学的主要方式。根据信息检索课的特点,CAI可以采用表式与图式相结合的方法,配以检索工具、检索途径和检索示例。该课件大致包括5个模块:(1)信息资源模块:以表式方式再现WWW, HTTP, GOPHOR, FTP, WAIL, WAIS等信息资源类型、内容及特征;(2)检索工具模块:以表式方式对信息资源指南类工具,字典式工具、索引型工具、交互式查寻工具的结构、功能、特点及主要查寻方式的比较分析;(3)检索语言及索引系统模块:以图式方式展示分类等级链(含词间等级关系、相关关系),主题语言句法关系;以表式方式列举各种索引系统的名称,索引款目结构、检索功能、特征及评价标准;(4)检索策略模块:以表式方式归纳各类信息系统的策略模型、使用效果及评价准则;(5)辅助检索模块:列举常用数据库名称及主题目录,网上服务域名、所在地理位置及IP地址,中外信息检索系统主要咨询服务内容等。

项目 本文系“面向21世纪信息管理课程体系与教学内容改革”项目研究论文。

参考文献

- 1 曾民族. 网络信息检索现状和性能评价. 情报学报, 1997, 16(2)
- 2 曾蕾. 联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展趋向. 情报学报, 1994, 13(5)
- 3 蔡巍. Internet 网上信息检索新趋势. 情报杂志, 1998, 17(4)
- 4 陈光祚, 焦玉英等. 科技文献检索. 武汉: 武汉大学出版社, 1987

焦玉英 武汉大学情报科学系教授。出版著作2部、译著2部, 发表论文30篇。通讯地址: 武汉大学大众传播与知识信息管理学院。邮编430072。

(来稿时间: 1999-08-27)