



数字化环境下 交互式信息检索

Interactive Information Retrieval in
Digital Environments

著 解 虹
译 夏立新 黄科舫 等



科学出版社
www.sciencep.com

数字化环境下交互式信息检索

Interactive Information Retrieval In Digital Environments

解 虹 著

夏立新 黄科舫 等 译

科 学 出 版 社

北 京

图字：01-2010-2335

Interactive Information Retrieval in Digital Environments by Iris Xie
Copyright © IGI Global 2008

All rights reserved.

Simplified Chinese Edition, published by Science Press, 2010,
which is the authorized translation of the English edition, 2008
IGI Global, the owner of all rights to publish and sell the same.

图书在版编目 (CIP) 数据

数字化环境下交互式信息检索/解虹著;夏立新,黄科舫等译. —北京:
科学出版社, 2010. 8

书名原文: Interactive Information Retrieval in Digital Environments

ISBN 978-7-03-028660-4

I. ①数… II. ①解… ②夏… ③黄… III. ①计算机网络-情报检索
IV. ①G354.4

中国版本图书馆 CIP 数据核字 (2010) 第 161914 号

责任编辑:王京苏 雪 暘 / 责任校对:朱光兰

责任印制:张克忠 / 封面设计:耕者设计工作室

版权所有, 违者必究。未经本社许可, 数字图书馆不得使用

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>

魏立印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2010 年 8 月 第 一 版 开本: 787×1092 1/16

2010 年 8 月第一次印刷 印张: 15 3/4

印数: 1—2 000 字数: 310 000

定价: 49.00 元

(如有印装质量问题, 我社负责调换)

作者简介

解虹 现为美国威斯康星大学密尔沃基分校信息研究学院教授。她的研究领域主要集中于信息查询与检索，特别是用户与信息检索系统之间的交互式信息检索，以及数字时代下交互式检索对各种信息检索系统的设计和评价的影响。她在1999年赢得了美国图书馆和信息科学教育协会（ALISE）论文奖。同时，她还是许多研究项目的主要研究人员，这些项目由不同的机构拨款，包括博物馆和图书馆服务机构（IMLS）、联机计算机图书馆中心（OCLC）、ALISE、国家科学基金会（NSF）等。Iris Xie 还发表了相当数量的期刊论文，以及图书馆和信息科学领域的国际会议论文。

译者序

很高兴能有机会翻译出版《数字化环境下交互式信息检索》这部著作。在我看来,能够将国外优秀华人学者信息检索方面的研究成果引进国内是一件有深远意义的事情,它不仅可以使国内一些学者和学生接触到最新的研究成果,也可以推进我国在交互式检索研究方面的长远发展。

本书的作者 Iris Xie(解虹)作为一名优秀的海外华人学者,早年在华东师范大学经过大学本科的系统学习后考入上海社会科学情报研究所攻读硕士学位。由此开始对以用户为中心的信息检索有了很深的兴趣,尤其是将系统研究学说和用户研究学说结合起来。并最终凭借其扎实的功底进入了美国图书情报专业最优秀的大学之一——Rutgers University 继续深造,师从在互动式信息检索领域的领军者和开拓者 Nicolas Belkin 教授,获得美国一流专家学者在此方向的首肯,可见她在此研究领域有着很深的造诣。在获得博士学位以后,她进入威斯康星大学密沃瓦基分校从事研究和教学工作,和与我熟识的优秀学者张进教授成为同事。解虹博士在 1999 年获得了由图书情报教育协会(ALISE)颁发的最佳论文奖。同时也是许多机构补助研究的主要研究者,其中包括博物馆图书馆服务协会(IMLS)、联机计算机图书馆中心(OCLC)、图书馆与信息科学教育协会(ALISE)等。她著有许多优秀的期刊论文并在情报学、图书馆与信息科学这一领域的国际会议中发表。解虹的研究兴趣和专长是信息查询和信息检索,特别是用户与信息检索系统之间的交互式信息检索及其在数字化时代对各种信息检索系统的设计和评估中的应用。她的一项获美国政府(Institute of Museum and Library Services)资助的科研项目着重于比较初级用户在使用数字图书馆时使用系统帮助和人工帮助的不同,目的是为检索系统设计更好的类似人工帮助的系统帮助机制。她的研究重心主要放在检索策略的转换规律和影响这些转换的因素,尤其是与计划和情境有关的因素,这成为此书翻译出版的契机。虽然身在海外,解虹教授却心系国内的学科发展与专业建设,与我有了许多交流。同时,她也为国内学者到美国去进行学术交流与合作提供了便利。作为译者,我希望能将这样一位学者的全新的观点和理论介绍给国内同仁,加强国内外学者在信息检索方向的交流与合作。

本书开篇简述了对研究图书情报学做出了重大的贡献六种著名的以用户为中心的学说,为了解本书的具体内容搭建了基础平台。然后在全书的前半部分针对不同数字环境下的交互式信息检索,例如,联机公众检索目录、联机数据库、网络搜索引擎、数字图书馆和文本检索会议,详尽地介绍了其概况和交互式研究类型,并对每一种数字环境的交互式检索都做了归纳与总结。本书还讨论了交互式信息检索模型:Ingwersen 的认知模型、Belkin 的文

本交互式系列性模型、Saracevic 的交互式层次模型,同时,研究了这三种信息检索模型的应用、影响及其局限。除此之外,还研究了其他交互式信息检索模型或方法:Ellis 的信息查询行为模型,Bates 的浆果择取方法,Vakkari 任务基础上的信息检索过程理论,Spink 交互式反馈模型,Hert 的交互式信息检索,Wang、Hawk 和 Tenopir 的用户网络交互多维模型、Pharo 的搜索情境及转换方法,并以此说明或强调信息搜集与检索的交互作用。本书的目的在于通过整合相关理论和实证研究,提出交互式信息检索的理论框架,然而对交互式信息检索研究的意义不仅在于提供了理论框架,还在于为信息检索的设计和评估提供了参考意见。特别在计划-情境模型探讨对数字环境下交互式信息检索理论框架的发展以及跨越多重原则把现有理论框架整合到面向用户的信息检索,这是本书的精华之处。最后本书对将来有待解决的问题和研究方向给出了一些建议。

近年来,动态、复杂、多样的信息海量快速增长导致用户获取所需信息的难度增大,以系统为中心的设计已经无法满足用户的需求,这对传统的信息检索系统带来挑战。因此如何更好地支持用户的信息查询活动也就成为现在的一个研究重点。交互式信息检索关键就是把信息检索过程看做是用户和系统的交流过程,其最大意义在于把信息检索系统看做用户的合作伙伴并以用户为中心,这正能应对当前信息检索所面临的挑战。然而,现有的研究并没有在系统设计中得以实施。以系统为中心的方法对用户和用户行为给予太少的关注,而以用户为中心的方法研究却无法给出实质性的解决方案。同时,采用系统为中心方法的设计者也不关心用户研究及其在信息检索系统设计中的结果。本书就是应对当前学科发展的需求和科学研究的热点而生。从理论和实践上系统、全面地阐述了交互式信息检索最新研究进展。通过本书,读者可以深入了解数字环境下交互式信息检索的理论框架和交互式信息检索的宏观、微观模型以及著者对于信息检索的设计和评估的参考意见。本书在理论与技术上都为交互式信息检索的深入发展做了充分的准备,将成为此研究领域值得关注的一颗新星。而作者指出的现存问题更应该得到广大研究者和专业人员的深入挖掘。本书的面世,必然会引起更多人对交互式信息检索的兴趣与关注,推进信息检索方法的发展,也为人们更加便捷地获取所需信息提供更加先进的环境。

本书的内容涉及五个方面:系统中心论和用户中心论的两大学派;不同网络环境下交互式信息检索研究的精华,这些网络环境包括联机公众检索目录、联机数据库、网络搜索引擎、数字图书馆和文本检索会议(TREC);交互式信息检索模式和框架的探索,主要包含 Ingwersen 的认知模型、Belkin 的文本交互式系列性模型、Saracevic 的层次模型;计划-情境交互式信息检索模型;交互式信息检索模式的意义和应用以及有待解决的问题与研究方向。

本书是作者经历十年潜心研究的心血之作。从内容上说,本书是作者站在巨人们的肩膀上的最新力作。除了包含前辈学者已经出版的关于交互式信息检索方面的研究外,对数字环境下的信息检索作了具有针对性的归纳总结,对宏观和微观信息检索模型作出了详尽的优缺点分析。从学术思想上来说,综合了作者自己和其他学者在交互式信息检索上的最新研究;从理论上和实践上都提出自己在信息检索的设计和评估方面的参考意见。另外,作者对于以前的一些研究成果做了创新,提出了自己的计划-情境交互式信息检索模型,具有很强的实践价值,这是在现有同类书籍中非常少见的。从结构上来说,本书在结构的编排上逐层深入,把系统中心论和用户中心论的两大学派作为切入点,重点论述了用户中心论,从而过渡到各种数字环境的针对用户的交互式信息检

索,再介绍模型,合理地进行实证研究,最终提出自己的观点。重难点突出,层次分明,具有很强的可读性。

将这样一本内容翔实、观点新颖的英文著作翻译成中文,是对我们专业背景和知识积累的考验。本书涉及的领域众多,包括信息科学、信息检索、计算机、数学、心理学、行为学、社会学等。我们组建了优秀的翻译团队,并且积极请教了所涉领域的专家学者,力求做到精准、通达、完整地将作者的本意传递给广大读者,做好作者和读者之间良好沟通的桥梁。

在翻译的过程中有几点需要特别指明:首先在人名的翻译上,为了避免中文名和英文名对应上的失误,我们直接使用外国作者的原英文名而未作翻译。其次,文中的脚注,对于脚注中援引判例、文献等部分保留原文不动。再次,参考文献原样保留,便于对原著所引文献、数据等的查询。最后,在附录中,将专业术语都附注英文,以便于读者对照、理解和把握。

由于本书的专业性较强,我们建议读者可以先阅读第11章。由此部分对全书有一个提纲挈领的把握,便于在接下来的部分进行深入细致的阅读。此外书中出现的专业词组也是读者深入理解本书的前提,如知识异构状态(ASK)假设(anomalous state of knowledge (ASK) hypothesis)、Belkin的系列性模型(Belkin's episode model)、认知工作背景分析(cognitive work analysis, CWA)、跨语言信息检索(cross-language information retrieval, CLIR)、Dervin的意会法(Dervin's sense-making approach)、计划多维性(dimensions of plans)、情境多维性(dimensions of situations)、Ellis信息查询行为模型(Ellis' model of information-seeking behavior)、精确反馈和被动/隐性反馈(explicit feedback vs. passive/implicit feedback)、影响当前搜索目标转换的因素(factors affecting shifts in current search goals)、信息查询策略(information-seeking strategies)、Ingwersen的认知模型(Ingwersen's cognitive model)、集成信息搜索与信息检索的研究框架(integrated IS&R research framework)、交互动机维(interactive intentions, dimensions)、交互意图转换类型(interactive intention shifts, types)、Kuhlthau提出的信息搜索过程方法(Kuhlthau's information search process(ISP) approach)、宏观层次交互式信息检索模型(macro-level of interactive IR models)、方法实体转换(method-entity shifts)、微观层次交互式信息检索模型(micro-level of interactive IR models)、全适用型/通用性模型(one-size-fits-all model)、联机产业要素(online industry, major elements)、Pharo的搜索情境转换法(Pharo's search situation and transition method)、计划-情境交互式信息检索模型(planned-situational interactive IR model)、Saracevic提出的交互式层次模型(Saracevic's stratified interaction model)、Taylor的信息需求层次法(Taylor's levels of information need approach)、文本流(streaming text)、Vakkari的基于任务的信息检索过程理论(Vakkari's theory of the task-based IR process)等。

夏立新主持了本书的翻译工作。各章的分工分别是:夏立新、刘为翻译了第1、2章;翟姗姗、刘枫翻译了第3、7章;夏立新、李楠翻译了第4章;黄科舫、金晶翻译了第5、11章;宋敏霞、刘为翻译了第6章;徐晨琛、夏立新翻译了第8、9章;李冠楠、黄科舫翻译了第10章。刘为翻译了序、前言。夏立新翻译整理了中英文术语对照表。译稿的第一轮校对工作由各位译者进行交叉校对,第二轮的文字校对工作主要由翟姗姗、刘枫、刘为等承担,翟姗姗承担

了第三轮的统稿工作。解虹博士审阅了译稿,并为本著作的出版撰写了中文版前言,她的工作使本书在学术性、准确性和可读性上增色不少。最后,夏立新对全书进行了修改和统稿,并完成终校。没有我们这个团队的精诚合作,本书的翻译工作也难以顺利完成。

由于译者水平有限,虽然经过反复修改但仍然无法完全避免疏漏之处,希望各位读者能够不吝赐教,指出错误与疏漏,我们将不胜感激。

夏立新

2010年7月于武汉,桂子山



环境的不同方面是理解信息查询和检索的核心 (Cool and Spink, 2002)。互联网的出现已经创造了不同的数字环境,使得千万的用户能够在世界的任何地点任何时间自己来搜索信息。一方面,用户有着不同的知识和技能背景,他们也因为不同的目标而搜索信息。另一方面,不同类型的联机信息检索系统为不同的信息集设计了不同的接口。因此,在数字环境中为了完成自己的目标而找到有效的信息对用户来说是一个挑战。该序言描述了在数字环境下关于信息查询和检索的背景,并解释了编著本书的需要。

1. 信息检索系统和不同的数字环境

信息检索从来都不是一件简单的事情,其问题在于文本的呈现,无论是索引还是全文检索,都不能满足用户对信息呈现形式多样复杂的需求。此外,传统的信息检索系统设计只被用来支持单一的、由用户参与的信息查询策略——检索式构造。新的数字环境在设计 and 检索上都重新定义了联机检索系统。

1.1 信息检索和信息检索系统

什么是信息检索? Meadow 等 (1999) 认为信息检索是“在存储的信息或数据库中找到一些想要的信息”。选择是信息检索的关键。信息检索不仅仅是一个系统的活动,相反,它是一个用户和系统之间进行交流的过程。信息检索的核心问题是如何将用户需要的信息与存储在数据库中的信息进行匹配、比较或相关联。信息检索也可以被看做是信息查询、信息搜索和信息获取。这些词都可以被看做是信息检索的同义词,但他们的侧重点有所不同 (Chu, 2003)。Wilson (2000) 定义了信息查询行为和搜索行为的不同点。信息查询是一种有目的包含了用户和人工信息系统或者计算机信息系统的交互来实现其信息目标的行为。信息搜索行为是指与各种信息系统进行交互时的微观层面上的行为。从文字上来说,学者在表达信息检索这个概念的时候都使用了这些词。在本书中,我们遵循了 Wilson 的定义和其他学者的表达,所以信息查询、信息搜索和信息检索在被引用的时候被交替使用。

信息检索主要可以分为以下几种类型:

- 主题搜索: 找寻拥有共同属性的项。
- 已知项搜索: 找寻一个在用户知道关于其特定信息如作者、标题等的项。

- 专用特定信息搜索：找寻具体的数据或事实。
- 更新信息：为巩固某一领域知识结构而进行的浏览。

什么是信息检索系统？信息检索系统是用来让用户能够在数据库存储的信息中找到相关的信息。一个信息检索系统的典型组成部分包括：

- 用户检索式输入机制
- 用户检索式分析机制
- 文本选择/更新机制
- 文本分析机制
- 文本存储机制
- 文本和检索式匹配机制
- 为用户输入和系统输出的接口

为什么找到想要的信息这么困难？在信息检索领域最主要的问题是对数据库中文本的呈现不能和用户需要的呈现相匹配。用户的知识异构状态（ASK）创造了认知的不确定性，从而阻碍了用户恰当地表达自己的信息需求，而且用户不同层次的需求要求他们必须逐渐对自己所需的信息有更具体的概念（Belkin, 1977, 1978, 1980; Taylor, 1968）。用户的信息需求只能在与系统存储的信息交互时发生，并在与信息检索系统交互的过程中被分类。呈现信息需求这个动态的过程不能和文本的静态的呈现相提并论。

1.2 联机信息检索系统和不同的数字环境

互联网的发展为现存的联机信息检索系统带来了改变，例如联机公共检索目录（OPAC）和联机数据库；同时，互联网也使得新的联机信息检索系统诞生，例如网络搜索引擎和数字图书馆。那么如何来定义联机检索系统呢？联系检索系统和非联机系统不同，联机检索系统有自己的特点。Walker 和 Janes (1999) 指出了联机检索系统的独特之处：首先，联机搜索是实时进行的，用户在进行搜索时几乎可以立刻得到结果；其次，联机信息检索系统提供远程入口，用户只要有互联网连接就可以在任何地方进行搜索。典型的联机信息检索系统可以分为以下四种类型：①联机公共检索目录（OPAC），②联机数据库，③万维网搜索引擎，④数字图书馆。

这些联机信息检索系统的特点是什么？

OPAC 收录了一个图书馆馆藏中相关联的书目数据，更重要的是，终端用户可以对其进行搜索。OPAC 在 20 世纪 80 年代得到了完善，它们开始取代卡片目录。OPAC 成为第一个为终端用户建造的信息检索系统，而且联机费用已不是一个问题（Armstrong and Large, 2001; Chu, 2003）。第一代的 OPAC 沿袭了联机卡片目录模型和布尔搜索模型，它模仿了人们熟悉的卡片目录和像 DIALOG 或 MEDLINE 这样的联机数据库。第二代的 OPAC 融合了这两种设计模型并且为搜索、浏览和显示选项加入高级特性。第三代 OPAC 不仅加强了高级搜索特性还提供分级检索结果（Borgman, 1996; Hildreth, 1985, 1997）。新一代的 Web OPAC 允许用户获取图书馆、出版社和联机供应商的资源（Guha and Saraf, 2005）。今天，用户可以在世界上的任何一个地方甚至是手上进入 OPAC。新一代的 OPAC 也吸取了其他类型的检索系统的优点把高级搜索特性和新的设计融合在了一起，如让用户可以通过 Web OPAC 接口搜索到 OPAC 和联机数据库。

联机数据库在 20 世纪 60 年代开始发展。第一个主要的联机拨号服务是 1968 年的 MEDLINE 和 MEDLARS 的联机版。1972 年, DIALOG (Lockheed) 和 ORBIT (SDC) 提供商业联机服务 (Walker and Janes, 1999)。在 1972 年由现行的 LEXIS/NEXIS 系统的先驱 (Meadow et al., 1999) ——数据中心公司 Data Central Corporation 推出的第一个商业系统允许全文搜索。传统的联机搜索者是作为用户和联机数据库中间人的信息专业人才。20 世纪 90 年代, 联系供应商开始把他们的服务移向万维网, 从而终端用户成为联机数据库的搜索者。在过去的 30 年里, 联机产业经历了巨大的变化。数据库、出版商、制造商和供应商还有更重要的搜索者的数量急剧上升。文本数据库中的全文数据库的增加和多媒体导向的数据库的增加是近年来的两个特点 (Williams, 2006)。新的联机数据库服务更注意个性化和交互性, 而且提供联机数据库服务的专业系统。

1991 年通过使用超文本模型而产生的万维网使得成千上万的用户来搜索联机信息。网络搜索引擎是为用户在网络里导航的最重要工具。直到 2005 年 10 月, Nielsen//NetRatings (Sullivan, 2006) 认为, 搜索请求已达到了 5 100 000。有四种类型的搜索引擎使用户可以实现不同的目标:

- 辅助用户浏览信息的有着分层组织目录的网络目录
- 协助用户搜索信息的有着站点数据库的搜索引擎
- 允许用户同时搜索多搜索引擎的元搜索引擎
- 创造特定主题搜索站点的数据库的专业搜索引擎

网络搜索引擎的一个特别之处是它们呈现搜索结果的分级能力, 这种能力是基于语词出现频率、语词位置链接分析、普及度、人们的记载日期、检索式语词长度和近似度和恰当名词的特性实现的 (Liddy, 2001)。网络搜索引擎的新设计考虑了交互性、人性化和可视性。新的社区搜索引擎得以发展, 使用户之间可以分享搜索结果。一些网络搜索引擎把它们的服务从网络搜索扩展到桌面和其他类型的搜索应用中。

数字图书馆的出现为用户提供了更多的机会来获取各种信息资源。从字面来说, 对于是什么组成了一个可用的数字图书馆这个问题有不同的定义。Chowdhury 和 Chowdhury (2003) 基于 Borgman (1999) 的关于数字图书馆竞争视角的讨论把它们分为了两个主要类型。一种方法着重于数字化内容的获取和检索上; 另一种着重于数字化资源的搜集、组织和服务方面。数字图书馆吸收了信息检索系统的优点, 尽管它们不能等同于那些提供像保存、社区建筑和学习中心这样的附加服务的数字图书馆。Saracevic (2000) 指出一些被纳入信息检索系统设计和评估的方法在数字图书馆中的应用已经很充分了。网络数字图书馆的前期发展于 20 世纪 80 年代末期至 20 世纪 90 年代早期开始 (Fox and Urs, 2002)。由美国国家科学基金会 (NSF)、美国国防高级研究计划局 (DARPA)、美国国家航空航天局 (NASA) 等机构提供基金的数字图书馆创始工程 1 和 2 (the Digital Library Initiative 1 & 2) 在美国的关于数字图书馆技术、社会和行为方面的研究和发展起到领头的作用。各种组织和机构可以为大众也可以为一个特定的用户群体来建立一个数字图书馆。针对不同类型用户群体的交互性、人性化、可视化设计是数字化图书馆的发展趋势。

数字环境中不同类型的信息检索系统是相互关联的。联机数据库被命名为“初始搜索引擎”, 现行的搜索引擎被联机数据库所影响着 (Garman, 1999)。同时, 网络搜索引擎比网页提供的内容要多 (Hock, 2002)。Wolfram 和 Xie (2002) 指出了两个与联机数据

库和网络搜索引擎相关的信息检索环境：传统的信息检索和流行的信息检索。传统的信息检索有着对已经出版了的和未出版的资源的选择性内容概括等特点。此外，传统信息检索被广泛的适用于搜索一个非个人性质的主题。相反，流行的信息检索创造了一个语境使简单的用户入口和对各种全文信息资源的使用得以实现。流行的信息检索环境已经被指出它的内容缺乏可信度而且其资源管理和检索不够精细。数字化图书馆呈现了一个由使用之前和联机数据库提供的信息储存相类似的馆藏传统信息检索，和以网络搜索引擎为实例呈现的流行的信息检索的混合体。数字环境下的信息检索对信息检索系统、用户、信息和环境都有很深的影响。

此外，信息检索实验是一项正在进行的研究活动。近几年，美国国家标准技术研究院（NIST）、美国国防部、美国国防高级研究计划局、美国情报部的高级研究和发展活动（ARDA）等部门赞助了从1992年起每年举办一次的文本检索会议（TREC），为了评估参与者自己的实验所做出共同的努力。TREC的交互式专题为交互式信息检索的研究和交互式检索系统状况的评估和比较提供了一个基本的框架（Dumais and Belkin, 2005）。然而，场所的制约因素、被指派的任务、方便获取的样本、数据收集方法、TREC评估者和短周期都限制了TREC的结果。

2. 数字环境的影响和信息检索的挑战

在过去，搜索信息是信息专业人员的特权。现在，普通的人也成为终端用户。数字环境的出现使信息检索系统、用户、信息和用户与系统交互的环境发生了变化，同时也为用户在为了实现他们的目标想要有效检索信息带来了挑战。

2.1 对信息检索系统的影响和对用户的挑战

在数字环境中，用户不得不面对各种各样的联机信息检索系统。但是不是所有的人都被列入设计时考虑的用户中，这就影响了用户和系统交互的有效性（Dillon, 2004）。从系统的角度来说，传统的信息检索由这样两个核心过程支持：呈现和比较。信息检索的核心是对文本的呈现和用户需求的呈现进行比较（Salton and McGill, 1983；van Rijsbergen, 1979）。这样，只有一种查询策略得到了支持：检索式构造。在数字环境下，语词匹配而不是概念匹配或者问题匹配，尽管搜索机制已经被改进，但这仍然是一个问题。数字环境下的信息检索系统确实为用户查询信息提供了浏览机制，但是检索式输入框仍然是用户表达信息需求的最主要渠道。用户被搜索输入框限制了，而且大部分的搜索都只包含一个或者两个语词（Jansen and Pooch, 2001）。当用户在数字环境下进行多种信息查询策略时（Fidel et al., 1999；Marchionini, 1995；Vakkari et al., 2003；Wang et al., 2000），联机信息检索系统在提供一些浏览帮助的同时依然关注与支持搜索相关的策略。

交互性是数字环境下搜索的最重要特征。用户可以与联机信息检索系统进行交互，也可以通过多种渠道与其储存的信息进行交互。基于网络的信息检索系统的内在的交互特性对用户提出了挑战。当用户在褒扬联机信息检索系统接口的易用性的同时，他们同样关心这些系统进行交互时所缺少的用户控制。网络搜索引擎的简化版设计已经被转化为其他类型的信息检索系统设计。研究者们对接口设计的易用性给予了很多关注而很少考虑到用户

控制的问题。现存的联机信息检索系统不同时支持易用性和用户控制 (Xie and Cool, 2000; Xie, 2003)。相应的, 联机信息检索系统的设计需要清楚用户的参与情况和系统的作用以使用户与系统的交互更加方便 (Bates, 1990; White and Ruthven, 2006; Xie, 2003)。

所有的联机信息检索系统在它们的设计上都有一些共同点, 例如搜索输入框。但是联系信息检索系统的设计并没有一个标准。不同类型的信息检索系统有着不同的接口设计和不同的搜索机制。就算是同一个类型的信息检索系统内, 接口的设计和搜索机制也不相同。更糟糕的是, 在不同的信息检索系统种对搜索的命令也不相同。这就限制了用户与这些系统及储存信息之间的交互能力。在过去, 用户在图书馆或信息中心搜索信息, 而数字环境为用户提供了在他们自己的环境中搜索信息的机会, 比如在家里和工作场所。他们组织性的/机构型的工作任务和家中环境可能会影响他们的信息检索过程 (Cool and Spink, 2002)。最重要的是, 当用户享受着这种随时可以找到自己想要的信息的同时, 他们也失去了当遇到困难时从中间人那里得到帮助的好处。

再者, 数字环境缩短了系统和用户之间的距离。同时, 数字环境也使用户得到任何训练这种事情变得困难或者根本不可能。用户只能通过每个系统的帮助功能寻求帮助。但是, 由于信息检索系统里面模糊的和明确的帮助功能设计的不足, 用户几乎不使用帮助功能。此外, 用户在他们进行信息检索过程中的每一个步骤中都需要帮助, 但是他们并不总是清楚自己寻求帮助的情况和需求 (Cool and Xie, 2004; Trenner, 1989; Xie and Cool, 2006)。最后, 正如 Jansen (2005) 指出的那样, 在很多情况下, 帮助机制是为了辅助检索式构成过程为创建出来的, 而不是为了作为一种在信息检索全过程都在的伙伴而创建的。

2.2 对用户的影响和对用户的挑战

在数字环境中, 任何人都是一个潜在的终端用户。对于任何在数字环境下的已知信息检索系统来说, 他们的目标都是通用入口。用户可能代表了有着不同背景的各种用户群体。他们可能在年龄、语言、文化、目标知识、系统知识和信息查询技巧上都不相同。有的用户可能对信息检索系统方面没有任何知识或者甚至有可能有计算机恐惧症; 还有的用户也许能够构建复杂的检索式并且适应系统。有的用户可能对他要搜索什么都一无所知, 而有的用户却在这个领域是专家。用户可能有不同的搜索目标, 例如, 查找事实性信息、查找有着相同特性的项、更新信息等。就算他们可能在查找相同的信息, 不同的搜索问题也可能导致他们为解决自己的问题要检索出不同的检索结果。用户也可能有不同的查询策略和行为, 比如搜索、浏览等。问题是如何支持这些有着对系统环境不同熟悉程度、不同的信息查询技巧、不同专业知识、不同搜索目标和不同信息查询策略的终端用户。总而言之, 问题就是如何建立一个能够支持不同用户群体的不同需求的联机信息检索系统。

在数字环境下, 用户能够为了不同类型的信息而使用 OPAC、网络搜索引擎和数字图书馆。他们过去的经验和背景影响了他们与不同类型信息检索系统交互的方式。他们可能是某一种信息检索系统的专家用户但是却是另一个信息检索系统的初级用户。他们把自己在使用一种信息检索系统的构思模式和查询策略带入到另一个信息检索系统 (Wang et al., 2000)。而且, 新一代的网络用户希望 OPAC 和其他类型的信息检索系统有着和网络

搜索引擎一样的设计和特性 (Novotny, 2004; Yu and Young, 2004)。有经验的联机搜索者已经对某一程度的搜索复杂性适应了传统的联机数据库, 同时又对基于网络的信息检索系统的无效率感到不满 (van Brakel, 1997; Bates, 1997)。

另外一个对用户的改变和他们的期望有关。互联网的出现制造了一种假象, 让用户以为他们可以在很短时间内找到他们需要的信息。于是人们在搜索信息的时候失去了耐心。研究者们已经开始比较网络搜索和传统信息检索的异同。这些研究一经发现, 网络搜索引擎遵循信息检索系统的基本原则, 研究发现网络用户在使用例如联机数据库和 OPACs 时表现出非常不同的模式。比如说, 大部分的网络用户在一个搜索阶段中没有很多检索式, 而且每一个检索式都很短; 布尔算法很少被使用; 许多用户只提交一个检索式而不用连续的检索式继续进行检索 (Jasen and Pooch, 2001; Silverstein et al., 1999; Spink et al., 2001)。

2.3 对信息的影响和对用户的挑战

一般来说, 相关性一直是用户在评价检索到的信息时主要关心的问题。在互联网出现之前, 用户对权威还有从传统信息检索系统中检索到的信息的质量没有任何怀疑。在数字环境中, 与检索结果的交互已经成为了信息检索交互中的重要组成部分。用户与结果进行交互从而找到解决其问题的信息; 这些结果引导他们为所需要的信息进行搜索或者如果搜索结果没有提供相关的信息这些结果会引导他们重构检索式。

无论如何, 在数字环境中, 任何人都可以在网络中通过上传信息而成为信息的制造者。没有人会去审视或者更正网上信息的内容。因此, 用户不得不自行判断网络信息的质量和权威性。此外, 网络还为用户提供了不同的搜索环境; 它包括不同内容、不同格式和不同组织方式的各种信息 (Fidel et al., 1999; Jansen et al., 2000; Wang et al., 2000)。当用户和检索结果进行交互时, 他们不仅要做出相关性的判断, 还要做出对权威性和质量的判断。但是, 用户只愿意投入很少时间来评估这些结果。在 Xie (2006) 的关于用户对数字图书馆的评价的研究和 Rieh (2002) 的网络的评估研究中, 大部分用户都认为由于没有一个控制网络信息质量的机制, 他们要自己对质量和权威性进行判断, 这对他们来说是一个挑战。虽然有些信息检索系统有权威性控制系统, 但是用户还是想要有一种方法来自己做判断。

对于用户的另一个挑战是在电子环境中有着庞大的信息量, 这会导致认知超载 (Bilal, 2000)。这个问题有两面: 一方面, 虽然大部分信息检索系统努力增大它们的信息储存量, 但是它们只为很少的可用信息进行了编目; 另一方面, 信息检索算法是为了少量的连贯的储存信息创造的, 而基于网络的信息检索系统拥有的数字化信息是动态和多样的 (Arasu et al., 2001)。更糟的是, 很多电子资料是多媒体的而且使用不同语言。这种信息检索的多媒体、多语言的不确定性和复杂性为用户在有效检索多媒体和外语信息的时候带来了更多的挑战, 特别是在他们与信息检索系统交互过程中评价和理解信息的时候 (De Vries, 2001; Downie, 2003; Gey et al., 2005; Goodrum and Spink, 2001; Oard, 2001; Peters, 2005; Smeaton, 2004)。此外, 电子资料是从它们的印刷或者实体格式转化过来的, 在转化过程中, 这些资料的内容和语境可能会丢失 (Mi and Nesta, 2005)。

3. 对交互式信息检索模型的需要

应对数字环境中信息检索的挑战,其中一种方法就是构建一个交互式信息检索模型。Matchionini (1995) 认为,人类的存在就是一系列和环境的交互。交互性是人类基本的特性,而且现代社会的复杂性也迫使人们更多地和机构还有系统进行交互。无论如何,数字系统正在取代人类来进行交互。数字环境下不断发展的交互带来了更多的挑战和问题。

3.1 信息检索作为交互的实质

因为信息检索是个交互的过程,所以不确定性和交互性是信息检索的两个最重要特征。Taylor (1968) 的关于问题折中的经典著作提出在用户完成信息查询这个过程中有四个层次的信息需求。基于一个用户的不确定性,信息需求来自于为用公式表示的问题。Taylor 的著作的贡献在于它假定了一种可能导致一个表达出来的请求的用户思想中的某一特性的心理状态。Wersig (1979) 使用了“疑问状态”的概念,在这种问题情况中用知识和经验可能解决这种疑问,他根据一个个体的知识、信仰和情境来作为判断信息查询行为的前期条件。Belkin 的知识异构状态 (ASK) 假设 (1977, 1978, 1980) 是 Taylor 的模型的一个延伸。知识异构状态 (ASK) 与 Taylor 的“内在需求”和 Wersig 的“疑问状态”相似,都说明了一种在特定情况下用户的知识不够用的状态。知识异构状态 (ASK) 给出了一个框架,在这个框架中用户查询信息的原因可以被明确地显示出来并用于信息检索中。根据 Taylor 的“内在需求”、Wersig 的“疑问状态”和 Belkin 的知识异构状态 (ASK),如果用户不能认识到他们的知识/问题空间的状态,他们就会最终进入一个不确定的状态。他们需要与信息、系统和环境进行交互来明确他们的信息问题。

Ingwersen (1992, 1996) 的认知模型、Belkin (1996) 的文本交互式阶段性模型和 Saracevic (1997) 的层次模型是被引用最高的交互式信息检索模型;它们都描述了普遍的交互式信息检索及其主要组成部分。当 Ingwersen 的模型更着重于研究交互式信息检索的认知方面时, Belkin 的模型强调用户与文本的交互 (信息查询过程); Saracevic 的模型注重于理解用户和系统的不同层次间的相互作用。这三种模型都认同的是:①信息问题/需求是动态的,并会在信息查询和检索的过程中变化;②信息问题/需求可以通过交互行为而被明确。

3.2 对交互式信息检索机制的需要

这三个交互式信息检索模型为研究在数字环境下的交互式信息检索提供了基础。Ingweren 和 Järvelin (2005) 进一步提出了一个集成交互式信息查询、检索和行为过程的 IS&R 研究框架。但是这些模型只在宏观层面上阐述了交互式信息检索,它们并不能说明在交互式信息检索过程中出现的特殊的过程或问题,也不能把影响信息检索交互的因素与在数字环境中用户的信息查询策略或行为联系起来。最为重要的是,在数字环境下,一个交互式信息检索框架需要从有着不同目标的不同用户与不同信息检索系统交互的实证研究中总结出来。正如 Saracevic (1996) 指出的,“信息检索交互行为是一个复杂的过程,它很大程度取决于情境或者语境:一方面它的起点是用户,并且与用户、用户目标或问题、

用户的能力、状态和意向相关, 另一方面它也含有系统的特性和能力、信息资源及接口这些方面的内容”。Mantovani (1996) 认为理解交互是很困难的, 因为在交互行为中那些一直在变化的不仅仅是这些因素本身或交互者的因素, 而是它们之间的联系的结构。为了构建数字环境下的交互式信息检索框架, 我们需要研究以用户为中心的方法、不同信息检索数字化环境的特征和数字环境下的交互式信息检索的实证研究, 还有现有的交互式信息检索模型和方法。

4. 本书概述

4.1 本书目标

本书的目标是为信息检索交互发展一个理论模型、进一步讨论在数字环境下信息检索系统的设计和评价的应用。本书的来源是 1999 年情报学科学教育联盟 (ALISE) 颁奖的作者获奖专题论文: Planned and Situated Aspects in Interactive IR: Patterns of User Interactive Intentions and Information Seeking Strategies。本书给了作者一个契机来整理她 10 年来在这一重要而独特的领域中的研究和其他研究者的著作。

4.2 本书结构

本书分为四个部分。

第一部分即第 1 章, 是对全书的概述, 是全书的基础。序言部分介绍了本书的背景并回答了为什么这本书要出版的原因。第 1 章以面向系统和面向用户方法间的区别展开讨论, 并进一步呈现出对理解交互式信息检索来说很重要的各种面向用户的方法。

第二部分包括第 2~6 章, 概述了各种信息检索的环境并深入回顾了关于在这些环境中交互式信息检索的实证研究。第 2 章到第 5 章主要阐述了 OPAC 中的交互式信息检索、联机数据库、网络搜索引擎和数字图书馆环境。对信息检索环境的概况部分为读者呈现了信息检索系统的历史和背景、各种类型信息检索系统的定义、各种类型信息检索系统的现状和对用户的挑战。对交互式信息检索实证研究的回顾部分按照这些实证研究中的议题来进行划分, 其中包括任务/目标和它们的影响、信息查询策略的层次、用户的知识结构、联机帮助、可用性研究、交互式信息检索系统的评价等。另外, 第 6 章总结了 TREC 环境的交互式专题和不同类型的交互式专题研究, 特别是交互式专题的贡献及其局限。这个章节同样也讨论了关于主要在跨语言评估论坛 (iCLEF) 中, 基于交互式专题的跨语言信息检索研究的著作。

第三部分包括第 7~9 章, 重点阐述了交互式信息检索模型的发展。第 7 章回顾了在这个领域中发展的交互式信息检索模型的宏观和微观层面, 并进一步讨论了这些模型的优点和局限性。第 8 章是这本书的核心部分, 为读者呈现了作者的信息检索模型——计划-情境交互式信息检索模型。关于该模型的讨论包括: 对这个模型的概述, 对用户目标、任务和它们呈现的不同层次的讨论, 工作任务和搜索任务的多维性, 用户的个人信息结构, 社会组织环境, 信息检索系统, 信息查询策略的多维性, 目前搜索目标和信息查询策略的转换及这些转换的影响因素。第 9 章通过研究报告和讨论, 致力于研究人们如何在其中

研究提案的写作过程中查询和检索信息，这一大规模导航研究得出的结果，阐述和证明了计划—情境交互式信息检索模型。

第四部分包括第 10 章，讨论了交互式信息检索模型在交互式信息检索系统的设计和评估中的应用。第 10 章讨论了在交互式检索系统的设计和评估中该模型的理论和实际应用情况，特别为解决以下问题提出了建议：如何支持多类型的信息查询策略，如何在系统作用 and 用户参与情况这一方面平衡易用性和用户控制的关系，如何构建交互式帮助机制，如何构建一个多维度的评价机制来评价交互式信息检索系统。

最后，第 11 章对本书的贡献作了总结，讨论了未来研究的方向并为交互式信息检索的进一步研究提出了问题。

4.3 目标读者

本书为研究者、设计者、教师、本科生和研究生，以及对数字环境下交互式信息检索、信息检索系统设计和信息检索系统评估感兴趣的专业人士设计。理论框架和对理论与实证的综述将会为交互式信息检索的创新研究提供基础，而且可以作为与信息检索和信息检索系统设计相关学科课程的教学素材。关于应用的讨论部分将为设计者、其他为大众及特定用户人群进行设计和评估新的交互式信息系统专业人员提供引导。

以下机构的成员将会是本书的主要读者：①美国信息科学与技术学会（ASIST）；②美国计算机协会（ACM）；③美国电气和电子工程师协会（IEEE）；④情报学科学教育联盟（ALISE）；⑤各种图书馆机构，如美国图书馆协会（ALA）、专业图书馆协会（SLA）等。其次是其他相关领域（如物理学、通讯学、计算机科学、工程学、健康学、教育学等）中对交互式信息检索和信息检索系统的设计及评估感兴趣的研究者及工作者。