

WWW

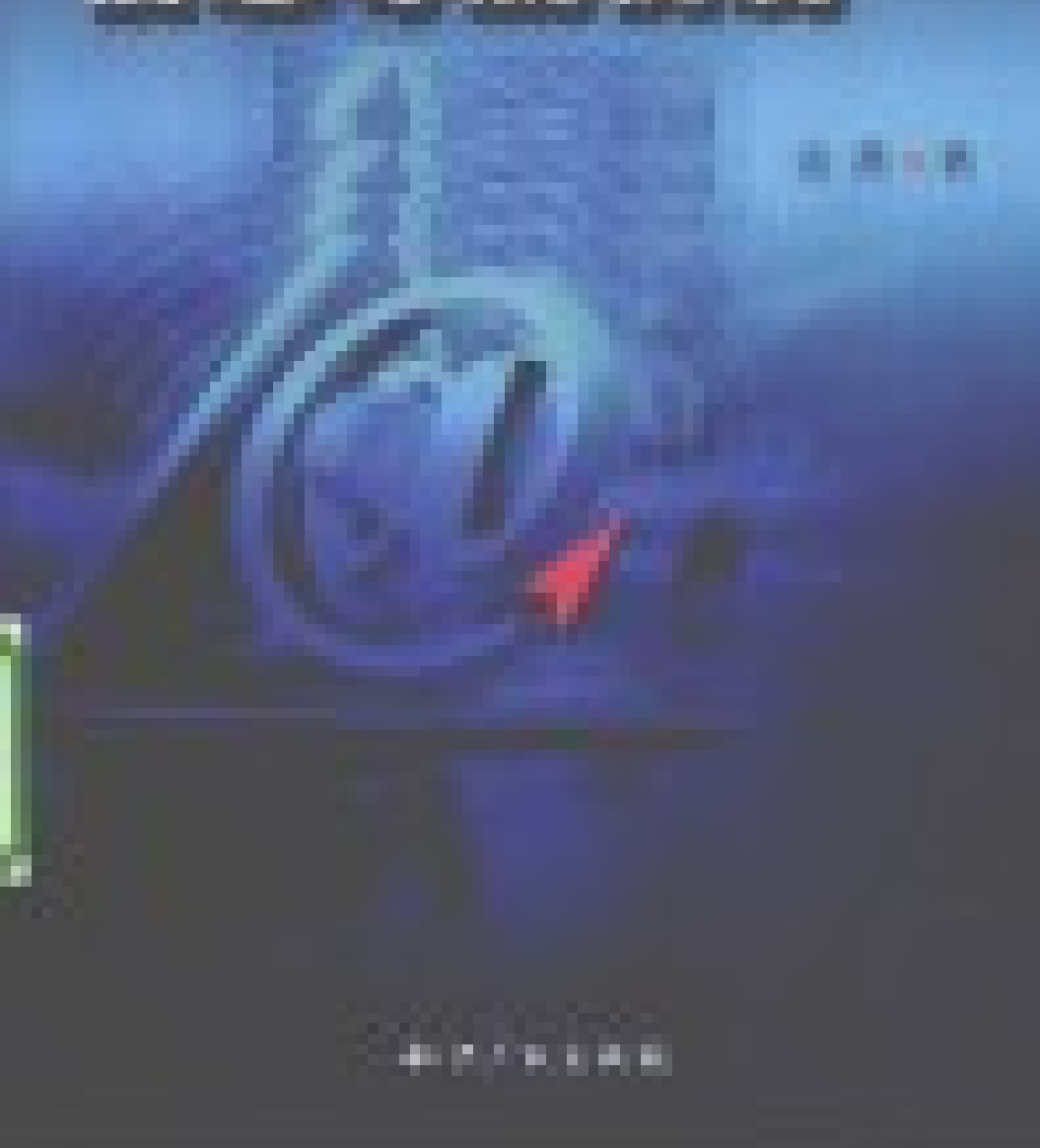
信息导航机制

金燕 ■ 著

知识产权出版社

WWW

信息导航机制



本书获河南省重点学科图书馆学学术著作出版资助

WWW 信息导航机制

知识产权出版社

内容提要

WWW 的混沌、无序以及人们在 WWW 上遇到的迷航与认知过载等种种现状,迫切需要信息导航来辅助人们的信息访问与获取活动。本书以 WWW 信息导航机制为研究对象,探索网络信息导航的内在机理问题,介绍了 WWW 信息导航的产生、内涵及其认知基础,梳理、归纳并提出现有 WWW 信息导航的 5 种主要导航机制:基于超链接的导航机制、基于内容的导航机制、时序机制、空间机制和社会导航机制;同时研究了 WWW 信息导航的可视化辅助机制,以及现有的 WWW 信息导航策略,认为应当从用户、信息结构的设计、信息内容的深层揭示等方面改善现有的 WWW 信息导航工具。

责任编辑:荆成恭

封面设计:智兴设计室

图书在版编目(CIP)数据

WWW 信息导航机制/金燕著. —北京:知识产权出版社,2009. 9

ISBN 978-7-80247-812-1

I. W… II. 金… III. 因特网—情报检索—研究 IV. G354. 4

中国版本图书馆 CIP 数据核字(2009)第 142477 号

WWW 信息导航机制

WWW Xinxi Daohang Jizhi

金燕 著

出版发行:知识产权出版社

社 址:北京市海淀区马甸南村 1 号

网 址:<http://www.ipph.cn>

发行电话:010-82000893 82000860 转 8101

责编电话:010-82000860-8341

印 刷:知识产权出版社电子制印中心

开 本:880 mm×1230 mm 1/32

版 次:2009 年 9 月第 1 版

字 数:200 千字

ISBN 978-7-80247-812-1/G · 307

邮 编:100088

邮 箱:hjb@cnipr.com

传 真:010-82000893

责编邮箱:jcgxj@163.com

经 销:新华书店及相关销售网点

印 张:8

印 次:2009 年 9 月第 1 次印刷

定 价:28.00 元

版权所有 侵权必究

如有印装质量问题,本社负责调换。

前 言

WWW 的混沌、无序以及人们在 WWW 上遇到的迷航与认知过载等种种现状,迫切需要信息导航来辅助人们的信息访问与获取活动。WWW 信息导航是 WWW 信息资源有效获取和利用的一种重要方式,就其本质而言,是一种大型信息资源的访问与获取策略,能够帮助人们在信息搜寻的过程中时时定位,根据用户需求引领用户准确、迅速地实现目标的查询;就其外在形式而言,是一种有序地展示 Web 信息资源内容与结构的手段,也是正确引导 Web 用户顺利实现信息搜寻目标的一种信息服务。

本书以 WWW 信息导航机制为研究对象,试图探索网络信息导航的内在机理问题,介绍了 WWW 信息导航的产生、内涵及其认知基础,梳理、归纳并提取出现有 WWW 信息导航的 5 种主要机制:基于超链接的导航机制、基于内容的导航机制、时序机制、空间机制和社会导航机制。在探讨这 5 种主要导航机制的内涵、基本原理、作用机理的基础上,研究了 WWW 信息导航的可视化辅助机制,以及现有的 WWW 信息导航策略,认为应当从用户、信息结构的设计、信息内容的深层揭示等方面改善现有的 WWW 信息导航。

与同领域的其他研究成果相比,本书的创新点主要体现在 3 个方面:第一,对现有的 WWW 信息导航方法、导航策略等进行系统研究,从中归纳并提取出 WWW 信息导航的 5 种基本机

制和可视化辅助机制,并对各种导航机制的原理、作用机理、特点等进行了深入地分析和比较,认为 5 种导航机制各有长短,适用于不同的情景,导航设计应当根据要实现的目标,合理地选择和应用多种导航机制;第二,引入信息构建思想,探讨了基于信息构建的导航机制的优化问题并认为信息构建是 WWW 信息导航的基础理论之一,通过信息内容和信息结构的构建,可以有效地改善现有 WWW 信息导航存在的问题;第三,引入 ontology 思想,提出了一个基于 ontology 的语义导航模型,认为为导航添加深层语义是从内容上优化现有 WWW 信息导航机制的有效措施之一。

WWW 信息导航的理论研究与实践在不断发展,本书需要补充和完善的内容还有很多,加之作者的水平有限,书中难免有疏漏和错误,恳请专家与读者批评指正。期待与各位同仁就该领域的研究进行交流与合作。

金燕

2009 年 6 月

目 录

第一章 绪 论	(1)
一、研究的背景和意义	(1)
(一)研究的缘起	(1)
(二)研究的意义	(5)
二、国内外相关研究现状分析	(6)
(一)国外 WWW 信息导航研究概况	(6)
(二)国内 WWW 信息导航研究现状	(15)
(三)国内外 WWW 信息导航研究存在的不足和问题	(18)
三、本书的研究内容、研究方法与创新	(19)
(一)研究内容	(19)
(二)研究方法	(20)
第二章 WWW 信息导航概述	(23)
一、导航的概念溯源	(23)
(一)物理世界的导航	(23)
(二)WWW 信息导航	(24)
二、WWW 信息导航的认知基础	(32)
(一)WWW 信息导航中的认知地图	(33)
(二)Spence 导航框架中的认知活动	(37)
(三)WWW 信息导航的认知模型	(39)
三、信息构建与 WWW 信息导航	(41)
四、WWW 信息检索与 WWW 信息导航	(42)



第三章 WWW 信息导航的主要机制	(45)
一、基于超链接的导航机制	(45)
(一) 基于超链接的导航机制的基础——超文本技术	(46)
(二) 基于超链接的导航机制的原理	(51)
(三) 链接的设立	(53)
二、基于内容的导航机制	(54)
(一) 基于关键词查询的导航机制	(54)
(二) 基于语义的导航机制	(62)
三、时序机制	(74)
(一) 时序机制的内涵	(74)
(二) 时序机制的导航机理	(74)
(三) 时序机制在 WWW 信息导航中的应用	(75)
四、空间机制	(79)
(一) 空间机制的内涵	(79)
(二) 相关概念	(80)
(三) 空间知识及其获取	(84)
(四) 空间机制的导航机理	(89)
(五) 空间机制在 WWW 信息导航中的应用	(91)
五、社会机制	(93)
(一) 社会机制的内涵	(93)
(二) 社会导航的类型	(96)
(三) 基于多 Agent 的社会导航模型	(105)
(四) 实例分析——Footprints	(108)
六、五种导航机制的比较	(111)
第四章 WWW 信息导航的可视化辅助机制	(117)
一、可视化与 WWW 信息导航	(117)
(一) 可视化增强用户的认知能力	(118)
(二) 可视化辅助 WWW 信息导航	(120)
二、WWW 信息导航可视化	(124)



(一) 导航元素可视化	(124)
(二) 导航结构可视化	(129)
三、支持导航的可视化技术	(133)
(一) 鱼眼图技术	(133)
(二) “Focus+Context”技术	(134)
(三) 视点控制技术	(136)
(四) 突显技术	(136)
四、可视化导航实例分析	(137)
(一) 实例 1: MAPA TM ——站点结构可视化	(137)
(二) 实例 2: PadPrints——历史记录可视化	(140)
(三) 实例 3: Cybermap——概览图导航 WWW	(142)
(四) 分析	(146)
第五章 WWW 信息导航策略	(149)
一、浏览器导航策略	(149)
(一) 直接搜索策略	(150)
(二) 页面重访策略	(151)
(三) 浏览器导航策略的改进	(154)
二、网站导航策略	(155)
(一) 结构导航策略	(155)
(二) 内容导航策略	(163)
(三) 图式化导航策略	(168)
三、面向用户的个性化导航策略	(170)
(一) 用户个性化信息的获取	(173)
(二) 用户模型的构建	(175)
(三) 个性化导航策略	(176)
四、基于用户查询行为的导航策略选择	(178)
(一) 用户信息查询行为分析	(178)
(二) 基于用户查询行为的导航策略选择	(181)



第六章 WWW 信息导航机制的优化	(185)
一、WWW 信息导航的优化策略	(185)
二、基于信息构建的 WWW 导航设计的优化	(186)
(一) 基于信息构建的导航设计思想	(187)
(二) 基于信息构建的导航设计方法	(191)
(三) 实证分析: 新浪(北京)网站的导航设计	(201)
三、基于 ontology 的语义导航优化	(207)
(一) ontology 相关知识	(207)
(二) 基于 ontology 的语义知识表示	(209)
(三) ontology 对语义导航的优化	(214)
(四) 一个基于 ontology 的语义导航模型	(217)
第七章 结束语	(223)
一、总结	(223)
二、展望	(225)
参考文献	(227)
(一) 中文参考文献	(227)
(二) 外文参考文献	(235)
后 记	(247)

第一章 绪 论

一、研究的背景和意义

(一)研究的缘起

WWW(World Wide Web,简称 WWW 或 Web)系统的前身是 1991 年欧洲量子研究中心在因特网上推出的科研信息交流系统,由该中心的 Tim Berners-lee 和其同事开发。该系统最初仅仅用于学术交流领域。1993 年,伊利诺斯大学的 Marc Andreessen 开发的用户界面 Mosaic 使它得到普及。而随着微软的 IE 和网景公司的 Netscape 等用户界面的广泛应用,WWW 成为因特网上的主流技术,而 WWW 也几乎成了 Internet 的代称。

随着 WWW 信息数量、类型的巨幅增长以及人们信息获取的需要,WWW 上的信息导航问题日益成为一个值得关注的问题。WWW 信息导航^①的产生源自于如下两个方面。

(1) WWW 自身的发展及其引起的“迷航”、“认知过载”等问题

WWW 作为人们信息存取的主要平台之一,允许任何人在

① 作者注:本书中与网络信息导航同义。



任何时间、任何地点自由地发布和传播信息。因而，自诞生之后，WWW 信息的类型、数量成几何级数迅速增长。目前，WWW 集成了大量的文本、声音、图像、动画、多媒体信息，已经成为人们获取网上信息的主要源泉。表 1-1 反映了 WWW 信息在因特网上占据的份额比重^①。而人们常用的搜索工具 Google 则宣称其已经索引了 4 285 199 774 个网页^②。

表 1-1 因特网上各类信息流量的比例

信息类型	Telnet	FTP	E-mail	WWW	其他
流量比例	1.6%	8.4%	10.7%	78.3%	1%

也正是由于 WWW 的自由性，其信息的质量、信息组织的质量难以控制，使之成了一个生态失衡的大信息空间，在给人们提供丰富信息资源的同时也给人们的信息获取活动带来了不便。更重要的是，WWW 采取了超文本组织信息的方式，通过超链把成千上万的网页、网站和信息内容联系在一起。这种非线性的超文本组织机制，一方面改变了传统信息组织的静态线性形式，使 WWW 信息的组织富有灵活性、动态性，能够实现节点间的非线性跳转，允许人们随意地浏览节点信息；另一方面，随着网页数量的巨幅增加，链的数量越来越多，超大量的链的存在会使链间关系越来越复杂，错链、误链和死链大量存在。这些因素使得在多节点、多层链的跳转过程中，用户容易出现“迷航”和“认知过载”现象。

g 迷航，原指探险者在一个陌生的环境中迷失方向的现象。此处的迷航，特指在 WWW 信息查询过程中，由于链间

① 袁中直等.《化学化工信息资源检索和利用》，江苏科学技术出版社，2001。

② <http://www.google.com> [2004-06-29]。



关系的复杂性以及用户信息处理能力的限制,导致用户在多层链的跳转过程中,不知道自己现在在哪里,从哪里来,将要去哪里去的现象。网络用户的迷航有多种表现,不知道自己现在在哪里、曾经去过哪里、将要去哪里是一种迷航;而到达一个页面,却不知自己为何到这里也是一种迷航;跟随链进行多层跳转中偏移了目标或找不到来时的路径也是迷航。

认知过载,是指因同一时间需要集中注意力处理的任务过多引起的额外的心理负担。在面临众多链需要作出一定的选择决策时,用户容易产生认知过载。此外,“back-forward”回溯使总的访问线路相对比较清楚,但在多次回溯某一或某些节点时,也会使用户的认知负担加重。

除了上面提到的 WWW 和超文本信息组织方式自身的原因之外,用户及设链者也是导致产生迷航与认知过载的主要因素。用户背景知识的差异、对 WWW 系统的熟悉程度、先前的经验、链缺乏语义以及设链者和用户对同一概念理解的差异等都可能会引起迷航或认知过载。

大量的迷航和认知过载,会严重影响用户的信息获取效率,甚至会引起信息查询行为的中断。

(2) 用户信息获取需求的发展

作为信息社会的成员,用户信息获取需求日益提升,获取信息是信息社会中的人们生产、生活、学习的基本必要条件。随着信息基础设施的发展以及 PC、Internet 的普及,WWW 因其信息量丰富、信息更新较快等特点逐渐成为人们获取信息的主要渠道之一。然而,WWW 提供的信息获取方式却越来越难以满足人们日益增长的、质量要求越来越高的



信息获取需求。我们知道,搜索引擎检索和内容列表浏览是 WWW 提供给人们的最重要的信息获取方式。但目前的现状却是,在面对海量的网络信息时,众多网络搜索引擎显得被动无力,往往返回大量的超链,需要用户进行二次甚至多次的筛选和定夺,无端地浪费用户的时间、精力和财力。这主要是因为目前的搜索引擎多是采用基于关键词的匹配方式,这种简单的关键词匹配,割裂了信息内容间的语义关系,致使大量的无关链接作为检索结果返回,因而有了“在因特网上您总能找到(甚至只能找到)您不需要的东西”^①一说。与此同时,内容列表因分类体系不唯一、交叉链过多等容易导致“迷航”,若内容列表的层次过多,会增大用户的认知负担。一些基于交互的协作式导航(社会导航)的出现,虽然能够以比较灵活的形式向用户提供信息获取的途径,但尚未得到广泛应用;而能够充分体现信息内容的语义属性的语义导航更是少之又少。必须承认的一个事实就是,WWW 提供的现有导航方法已经难以有效地解决用户信息查询中的迷航和认知过载等问题,已经严重影响了人们对 WWW 的利用以及满足人们信息获取的需求。

而这些,一方面归因于网络信息空间的混沌现状,另一方面也归因于网络信息导航工具功效的不够完善。然而,目前网络信息导航工具的设计者并没有充分、深入地思考 WWW 和 WWW 信息导航的内在机理,没有把网站序化与导航的内在机制、导航工具的设计理念以及 WWW 恰当地结合

① 夏立新:《WWW 环境下学术信息资源及其检索策略研究》,武汉大学博士学位论文,2002。



起来。在这样的背景下,迫切需要引进新理论、新思想、新技术,以改善现有的导航工具设计,帮助用户在多维的网络信息空间内顺利遨游而不致发生“迷航”,使之能够在海量的网络信息资源中快速、方便、准确地获取所需的信息与知识。虽然目前尚未有研究者全面研究 WWW 信息导航的机制,但这个问题却深深地吸引了本书的研究者。

(二) 研究的意义

WWW 混沌、无序以及人们在 WWW 上遇到的迷航与认知过载等种种现状,迫使一种能够有助于解决上述问题的机制出现。WWW 信息导航机制因需而生。就其内在本质而言,WWW 信息导航是一种大型信息资源的浏览与检索策略,能够帮助人们在信息搜寻的过程中,时时定位,根据用户需求引领用户准确、迅速地实现搜寻目标;就其外在形式而言,是一种有序地展示 Web 信息资源内容与结构的手段,也是正确引导 Web 用户顺利实现信息搜寻目标的一种信息服务^①。但只有把有效的导航机制与良好的信息环境结合起来,其导航的效果才能得以更好地发挥。

信息导航问题是网络环境下信息组织管理中的关键性问题^②。但目前的现实却是 WWW 信息导航缺乏理论指导。因此,重视和加强 WWW 信息导航问题的研究,对于有效解决目前 WWW 信息导航缺乏理论指导的现实问题,开辟网络信

① 毕强等:“Web 信息空间导航研究”,《图书情报工作》,2003,(8),第 73-78 页。

② 毕强等:“Web 信息空间导航研究”,《图书情报工作》,2003,(8),第 73-78 页。



息资源开发与利用的新局面,实现 Web 增值服务,具有重大的理论意义与现实意义。大处而言,有助于从宏观上把握 WWW 信息导航的机理,明确影响 WWW 网络信息导航绩效的因素,确定 WWW 信息导航设计的规范、原则,加速网络信息资源有序化的进程;小处而言,有助于改善 WWW 信息导航工具的设计,提高网络导航的效率和效果。

本书以 WWW 信息导航机制为研究对象,试图探索网络信息导航的内在机理问题,为未来网络信息导航工具的设计提供理论依据和框架。本课题的研究,一方面能够在一定程度上丰富和拓展 WWW 信息导航的理论体系,具有一定的理论意义;另一方面,有助于优化和指导网络导航工具的设计,对进一步解决目前网络信息导航中存在的问题,具有重要的现实意义。

二、国内外相关研究现状分析

(一) 国外 WWW 信息导航研究概况

总的来说,国外近年来对信息导航的理论研究和应用研究都十分活跃, HCI (Human-Computer Interactive) 1997 年专门召开了“电子世界的导航 (Navigation in Electronic Worlds)”为主题的会议,讨论了导航的定义、导航与认知、导航的设计等问题。心理学、计算机学、信息检索学等多个学科和领域的研究者都参与到信息导航的研究中。心理学领域的研究者们从认知的角度探讨认知风格等因素对导航的影响;计算机学领域的研究者们多是从试验和实践的角度进行导航工具

