

高俊宽◎主编

网络信息检索

WANGGUO XINXI JIANSUO

 合肥工业大学出版社
HEFEI UNIVERSITY OF TECHNOLOGY PRESS

“江苏省‘十二五’高等学校重点专业建设：图书情报与档案管理类”项目资助

网络信息检索

主 编 高俊宽

参 编 (按参编章节顺序排名)

徐 芳 高 峰 任 荣

杨方铭 张 衍

合肥工业大学出版社

图书在版编目(CIP)数据

网络信息检索 / 高俊宽主编. —合肥:合肥工业大学出版社, 2015. 11
ISBN 978-7-5650-2502-0

I. ①网… II. ①高… III. ①网络检索 IV. ①G354.4

中国版本图书馆 CIP 数据核字(2015)第 274752 号

网络信息检索

高俊宽 主编

责任编辑 章建

出版	合肥工业大学出版社	版次	2015年11月第1版
地址	合肥市屯溪路193号	印次	2015年11月第1次印刷
邮编	230009	开本	787毫米×1092毫米 1/16
电话	总编室:0551—62903038 市场营销部:0551—62903198	印张	16.25
网址	www.hfutpress.com.cn	字数	375千字
E-mail	hfutpress@163.com	印刷	合肥学苑印务有限公司
		发行	全国新华书店

ISBN 978-7-5650-2502-0

定价:35.00元

如果有影响阅读的印装质量问题,请与出版社市场营销部联系调换。

目 录

第1章 网络信息检索概述	(001)
第1节 网络信息资源	(001)
1. 信息与信息资源	(001)
2. 网络信息资源	(002)
3. 网络信息资源的特征	(002)
4. 网络信息资源的分类	(003)
第2节 网络信息检索	(004)
1. 网络信息检索的概念	(004)
2. 网络信息检索的分类	(005)
3. 网络信息检索的原理	(005)
4. 网络信息检索的评价	(006)
第3节 网络信息检索的发展趋势	(007)
1. 跨语言信息检索	(008)
2. 语义信息检索	(008)
3. 可视化信息检索	(008)
4. 协同信息检索	(009)
5. 其他	(009)
第2章 搜索引擎	(011)
第1节 搜索引擎概述	(011)
1. 搜索引擎概述	(011)
2. 搜索引擎的常用检索方法	(015)
第2节 搜索引擎的技术原理	(017)
1. 搜索引擎的组成	(017)
2. 搜索引擎的运行	(019)
第3节 常用国内外搜索引擎	(021)
1. Google 搜索	(021)
2. 百度	(028)

3. 国内外其他常用搜索引擎介绍	(033)
第3章 目录型网络信息检索	(041)
第1节 目录型网络信息检索概述	(041)
1. 目录型网络信息检索的概念	(041)
2. 目录型网络信息检索的特征	(041)
3. 目录型网络信息检索工具与搜索引擎的比较	(042)
第2节 目录型网络信息检索原理	(043)
1. 目录型网络信息检索的工作原理	(043)
2. 目录型网络信息检索的工作步骤	(044)
3. 目录型网络信息检索工具的常见类型	(045)
4. 目录型网络信息检索工具的差异	(045)
第3节 国内外常用的目录型网络信息检索工具	(046)
1. 雅虎	(046)
2. 搜狐	(047)
3. 新浪	(048)
4. Dmoz: Open Directory Project (ODP)	(049)
5. The WWW Virtual Library	(050)
第4章 政府信息检索	(052)
第1节 政府信息概述	(052)
1. 政府信息概述	(052)
2. 政府信息公开	(054)
第2节 政府网站信息检索	(056)
1. 政府网站信息的检索方式	(056)
2. 国外政府网站信息检索	(057)
3. 国内政府网站信息检索	(065)
第3节 其他政府信息检索方式及途径	(071)
1. 政务微博	(071)
2. 政务微信	(073)
3. 图书馆或档案馆	(075)
第5章 教育信息检索	(080)
第1节 高校信息公开与教育信息获取	(080)
1. 高校信息公开	(080)

2. 教育信息获取	(088)
第2节 开放教育资源信息检索	(099)
1. 开放教育资源	(099)
2. 国际开放课件联盟/开放教育联盟	(103)
3. 我国台湾地区的开放式课程计划	(108)
4. 中国内地的国家精品课程	(110)
5. 视频公开课	(117)
第3节 MOOC 信息检索	(123)
1. MOOC (慕课) 概述	(123)
2. 三大 MOOC 平台	(125)
3. 国内 MOOC 平台	(130)
4. 中文 MOOC 学习社区	(134)
5. 总结: 如何上好一门 MOOC	(137)
第6章 学术信息检索	(139)
第1节 学术信息概述	(139)
1. 学术信息概述	(139)
2. 开放获取概述	(142)
第2节 开放获取期刊信息检索	(145)
1. 开放获取期刊概述	(145)
2. 国内著名的开放获取期刊检索平台	(147)
3. 国外著名的开放获取期刊检索实例	(153)
第3节 机构知识库信息检索	(159)
1. 机构知识库概述	(159)
2. 国内著名的机构知识库检索平台	(163)
3. 国外主要机构知识库检索平台	(167)
第7章 文化信息检索	(172)
第1节 文化信息概述	(172)
1. 文化和文化信息	(172)
2. 文化信息的类型	(173)
第2节 公共文化信息检索	(174)
1. 公共文化信息检索	(174)
2. 公共图书馆信息检索	(176)
3. 博物馆信息检索	(190)

第3节 文化产业信息的检索	(194)
1. 文化产业信息检索	(194)
2. 新闻出版信息检索	(197)
3. 广播电视信息检索	(203)
第8章 休闲购物信息检索	(206)
第1节 休闲购物信息概述	(206)
1. 休闲信息概述	(206)
2. 购物信息概述	(207)
第2节 休闲休息检索	(209)
1. 娱乐休闲信息检索	(209)
2. 旅游休闲信息检索	(224)
3. 体育休闲信息检索	(228)
第3节 购物信息检索	(229)
1. 综合购物信息检索	(229)
2. 比较购物信息检索	(231)
3. 专业购物信息检索	(233)
4. 物流与支付	(237)
第9章 学术规范与论文写作	(240)
第1节 学术规范	(240)
1. 学术规范概述	(240)
2. 参考文献使用规则	(241)
第2节 论文写作	(244)
1. 论文概述	(244)
2. 学术论文写作程序	(246)
3. 论文投稿与发表	(247)
后 记	(253)

第1章 网络信息检索概述

1952年,美国学者穆尔斯(C. W. Mooers)提出信息检索(Information Retrieval, IR)的概念。网络信息检索(Network Information Retrieval, NIR)是随着计算机、网络及通信等信息技术(Information Technology, IT)的发展与普及的产物。中国互联网络信息中心(CNNIC)的统计数据显示:截至2015年6月30日,中国网民数量达到了6.68亿,网站数357万,手机网民数量达5.94亿且仍在持续不断增长。可见,在信息高速发展的今天,通过网络来获取信息的用户正在不断增加,网络信息资源正日益成为人们工作、学习和生活中不可或缺的组成部分。

需要注意的是,互联网的出现使得信息的采集、传播的速度和规模达到空前的水平,出现了“信息爆炸”等现象。面对浩瀚无边的“信息海洋”,人们会出现“信息焦虑综合征”等负面情绪与问题。另外,由于网络的开放性、自由性、便捷性等特点,网络上充斥着各色的信息资源,这些网络信息资源在质量方面往往参差不齐,如何在包罗万象的网络信息资源中利用最少的时间检索出对自己最有用的信息就显得尤为重要。于是,各种网络信息检索技术、方法与工具应运而生。本书首先介绍网络信息检索的基本问题;然后对网络信息检索的两种工具:搜索引擎和目录型网络信息检索进行深入、系统的分析;最后,从政府、教育、学术、文化、休闲、网购等方面对网络信息检索的应用进行论述。

本章主要介绍网络信息资源的含义、特征、分类等基本问题,以及网络信息检索的含义、类型、原理及检索评价等基本问题,为后续章节的学习奠定基础。

第1节 网络信息资源

1. 信息与信息资源

1.1 信息

一位美国科学家在诗句中写道:没有物质的世界是虚无的世界;没有能源的世界是死寂的世界;没有信息的世界是混乱的世界。

美国哈佛大学的研究小组给出了著名的资源三角形:没有物质,什么都不存在;没有能量,什么都不会发生;没有信息,任何事物都没有意义。

信息、物质和能源一起构成当今人类社会的三大资源。信息是一个内涵与外延都十分广泛的概念,从不同的角度和目的可以对信息有不同的看法和理解,目前关于信息的定义有若干种。代表性的定义有:信息论的创始人申农在《通信的数学理论》中首次提出“信

息是用来消除不确定性的东西”；控制论的创始人维纳认为，信息是人们在适应外部世界并且使这种反应作用于外部世界的过程中同外部世界进行交换的内容的名称；我国学者钟义信认为，信息是事物运动的状态与方式，是事物的一种属性。总之，信息是在自然界和人类社会生活中普遍存在的一种现象，是事物存在方式和运动状态的属性。

1.2 信息资源

文献调查表明：信息资源的概念在学术界仍没有统一的定论。但是，目前有两种具有代表性的观点：从狭义上说，信息资源是指信息本身或信息的集合，主要指信息内容；从广义上说，信息资源是指信息及其相关因素的集合，涉及信息的生产、分配、交换、消费等整个过程。

2. 网络信息资源

简单来说，网络信息资源是互联网络上公开发布的网页和在线数据库的总和。网络信息资源又称虚拟资源、电子信息资源、数字资源、联机资源等，是指通过计算机网络，利用计算机技术、通信技术、多媒体技术等相关技术，以数字化形式存在于互联网上的各种类型的信息资源的总和。

与传统信息资源相比，网络信息资源的信息内容的表述与组织方式不同，信息表述所依附的媒介物也不同，因而其检索的技术、方法与工具也与传统的信息检索有所区别。

3. 网络信息资源的特征

网络信息资源是依托于互联网的一种方便快捷的新型数字化资源，它的出现改变了传统意义上的信息资源的检索方式，产生了传统意义上的信息资源所不具备的许多特点，使信息资源更加丰富。与传统的信息资源相比，网络信息资源有如下特征。

3.1 数量繁多，分布广泛

网络的开放性和自由性使得网络信息资源正在飞速增长，网络信息资源犹如夜空中的繁星一样众多。网络在全世界范围内的普及打破了传统信息资源在时间和空间方面的限制，网络信息资源分布呈现出全球性的特点。

3.2 类型众多，内容丰富

网络信息资源的类型多种多样，除了传统模式的文字外，还包括图像、声音、动画、视频、软件等其他类型的信息资源。网络信息资源的内容也十分丰富，涵盖了大部分的领域和学科内容，能够基本满足绝大部分用户的需求。

3.3 时效性强，更新及时

网络为人们的生活提供了便捷的交流方式，用户总能第一时间在网络上查看有关的信息资源，它具有很强的时效性和动态性，通信技术的发展更为网络的更新提供了可能。

3.4 成本低廉，获取方便

在网络上获取信息资源的成本一般都比较低廉，所花的时间成本和经济成本都要比获取传统的信息资源要低得多，且网络信息资源更易于用户方便快捷的获取，而不再受制于时间和空间的限制。

3.5 分布随意, 质量不一

网络的开放自由给用户提供了发布信息资源的广阔空间, 而资源的多样性和缺乏专门的管理使网络信息资源在分布上呈现出无序的特点, 从整体上看, 资源的分布较为随意, 由于用户的水平和能力等原因, 信息资源的质量也显得参差不齐。

3.6 交互性强, 方便共享

在网络的环境下, 每个用户既充当着信息资源拥有者的角色, 又充当着信息资源生产者的角色, 用户可以在网络上享受相对的自由, 分享自己已有的信息资源, 从而达到和他人资源共享的目的。

3.7 关联度高, 链接开放

网络信息资源利用超文本技术在相关的资源和资源之间建立起某种联系, 再将不同的资源通过节点链接在一起, 形成一个巨型的链接图。这种链接关系是对传统信息资源的一大挑战, 用户可以很容易地从一個链接资源转到另一个相关链接资源内容上。

3.8 版权纠纷, 难于管理

和传统的纸质信息资源的版权保护体系相比较, 网络信息资源的版权纠纷和不确定性明显更多。网络中包含着很多界定模糊的概念, 这使得相关法律法规的执行变得更加困难和难于规范, 也就使某些侵权行为更加肆无忌惮。保护网络信息资源的版权问题刻不容缓。

综上所述, 网络信息资源和传统的信息资源相比, 有很多明显的优点, 例如数量和类型众多、时效性强、成本低廉、交互性强、关联度高等, 但同时也具有分布随意、难于控制等问题。

4. 网络信息资源的分类

网络信息资源种类繁多, 内容丰富, 按照不同的方式可以将其划分为多个不同的类型。

4.1 按信息来源划分

(1) 政府信息资源: 由政府公开发布在网络上的各种信息, 代表着政府部门意志的信息, 具体包括政府的相关新闻、具体的部门设置、制定的政策法规和规章等。

(2) 社会信息资源: 来自社会各个阶层、各个领域的信息资源, 包括具体的教育信息、科技信息、社交信息和娱乐信息等。

(3) 商业信息资源: 商业内部或商业机构中的信息, 有些是对外公开的, 有些则是商业机密, 主要包括企业的历史、产品、销售渠道、顾客信息等。

4.2 按采用的网络传输协议划分

(1) WWW 信息资源: 是指一种基于超文本传输协议并利用超文本、超媒体和多媒体等技术清晰直观地展现给用户的信息资源。

(2) Telnet 信息资源: 是指利用网络通信协议通过计算机远程登录服务来访问其共享的资源。

(3) FTP 信息资源: 是指利用文件传输协议来实现信息资源的传递和共享的资源。

(4) 用户服务组信息资源: 是一种最受欢迎的信息交流方式, 包括新闻组、邮件列

表、专题讨论组、兴趣组和辩论会等形式。

(5) Gopher 信息资源：是一种基于菜单的网络服务，用户可以在菜单选项中选择相对应的资源即可与远程服务器建立联系。

4.3 按网络信息资源的组织形式划分

(1) 文本信息资源：是指按一定的文本顺序排列的信息资源。

(2) 超文本、超媒体、多媒体信息资源：它可以将各种形式的资源如文本、图像、声音等很好地结合起来，从而达到更加丰富的效果。

(3) 数据库信息资源：它可以将大量的信息资源以一定的方式整理在一起，用户可以直接在数据库中方便地查找检索。

(4) 网站信息资源：是指将各种方式的信息资源经过一系列的整合，组织在一个网站内的资源。

4.4 按信息交流方式划分

(1) 非正式出版信息：是指流动性、随意性较强，信息量大、信息质量难以保证和控制的信息。它在网络信息资源中占很大的比重。

(2) 半正式出版信息：是指受到一定的产权保护，但是没有纳入正式出版信息系统的信息。

(3) 正式出版信息：是指受到一定产权保护，信息质量可靠、利用率较高的知识性、分析性信息。

4.5 按信息的载体和传播范围划分

(1) 光盘信息资源：主要指以光盘的载体形式展现出的信息资源。

(2) 联机信息资源：是由计算机联机服务系统提供的集中式的网络信息资源。

(3) 互联网信息资源：是指分布在互联网上的各种包括文字、声音、图形、图像、视频等在内的资源。

第 2 节 网络信息检索

1. 网络信息检索的概念

目前关于网络信息检索的概念尚无统一的定义，一般来说，网络信息检索是指运用特定的网络信息检索技术和策略，利用相关的网络信息检索工具，对互联网上的信息资源进行查找的一种检索方式。

为了理解网络信息检索的概念，我们可以分析一些日常工作、学习和生活中可能遇到的例子。

实例 1：小明通过百度地图查找前往面试公司的交通工具及路线。

实例 2：小赵想买笔记本电脑，通过 QQ 向朋友咨询了相关的信息。

实例 3：小李想出国留学，通过网络查找美国学校与专业的排名情况。

实例 4：小王想去韩国济州岛旅游，通过 Google 查找济州岛的旅游攻略。

实例 5: 小康在图书馆 OPAC 系统里面找到了想要借阅书籍的相关信息。

以上 5 个实例中, 1、3、4 属于网络信息检索的范围, 而 2、5 不属于网络信息检索的范围。分析实例 1、3、4, 我们可以得出网络信息检索的两个充分条件: 以网络为工具和检索网络上的信息资源。

2. 网络信息检索的分类

根据每个人不同的检索习惯和不同的检索目的, 可以将网络信息检索分为三类。

2.1 浏览

用户在互联网上浏览时, 当看到感兴趣或想了解的信息时可以利用文件中的超文本链接从现有网页跳转到另一个相关的网页。在用户的检索目的性不太强的时候, 可以利用浏览的方式进行检索; 但是如果过分地依赖这种检索方式, 容易让人迷失在网络信息资源的大海中。

2.2 利用目录型网络信息检索工具的检索

目录型检索工具是采用人工方式或半自动化方式对信息资源进行采集、评价、组织、过滤和控制等, 按照一定的学科顺序或其他顺序对这些整理过的资源进行分类, 本书第 3 章将详细介绍。用户不需要搜索具体的关键词, 只需要按照分类目录层层挖掘下去即可。由于是人工编排录入, 所以该检索具有较高的准确性, 可以清晰地看出大类和子类之间的关系。但是, 人工的方式必然会花费更多的时间成本和精力成本, 再加上网络信息资源千变万化、更新迅速, 使得这种方式的检索存在着更新速度慢、学科类别分类不统一等问题。目录型检索工具中最具有代表性的网站就是雅虎、新浪、搜狐等。

2.3 利用搜索引擎型检索工具的检索

搜索引擎型检索工具是互联网上较为普遍和使用范围较广的一种检索方式。搜索引擎型检索工具让自动跟踪索引器(如网络蜘蛛、网络机器人等工具)遍历整个互联网, 将各种相关的网络信息资源搜集到一起并建立一个庞大的资源数据库, 当用户发出具体的操作请求后, 它会根据用户的具体需求而调用资源数据库中的相关资源并将其返回给用户, 本书第 2 章将对此进行详细介绍。虽然搜索引擎型检索工具能很人性化地“识别”用户的一系列操作且快速地给用户提供较为准确的信息资源, 但是由于它采用计算机软件自动对信息资源进行加工和检索而不是人工加工检索, 以及存在着检索软件的水平不同等问题, 这种检索方式检索出来的信息资源的质量还有待提高。

3. 网络信息检索的原理

信息检索是将信息按照一定的方式组织和存储起来, 并根据用户的需要找出有关信息的过程。在整个网络信息检索的过程中, 用户通过互联网这个平台, 运用相应的检索技术和检索策略, 在计算机检索系统的数据库中检索自己想获取的信息资源, 数据库会根据用户的检索情况在终端设备上呈现出最终的检索结果。

想要检索网络信息资源, 首先要保证数据库中存储有相应的数据资源, 然后才能对其进行检索。因此, 从广义上说, 网络信息资源包括了信息存储和信息检索这两部分。网络信息检索原理的示意图如图 1-1 所示。

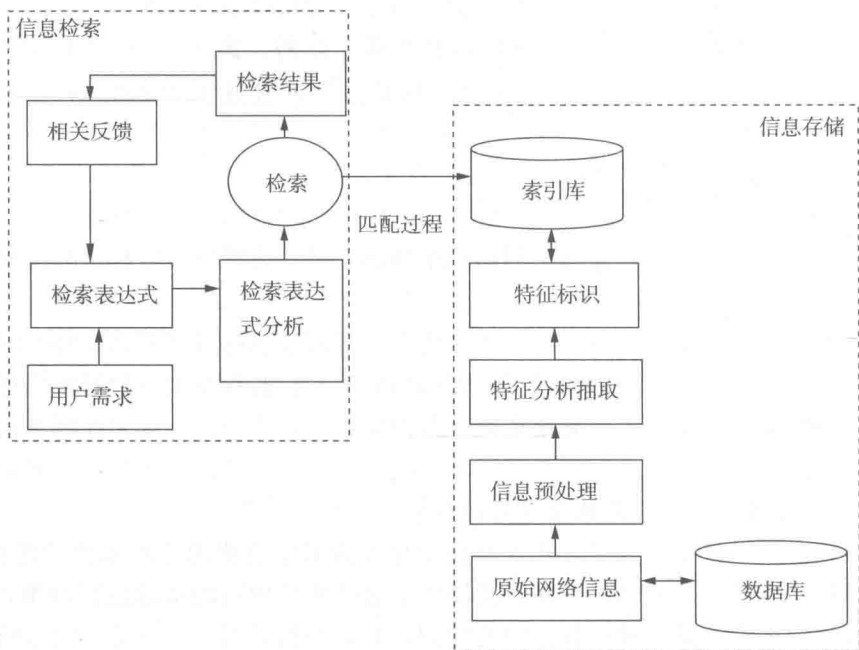


图 1-1 网络信息检索原理

3.1 信息存储

信息存储的过程是以人工或者自动化的形式收集各种类型的原始网络信息资源，将其按照计算机特定的格式和语言存储在数据库中。由于网络信息资源数量庞大，为了提高检索效率，通常还会对网络信息进行进一步的处理，构建索引库，主要包括预处理、特征分析抽取、特征标识等环节。

3.2 信息检索

信息检索是建立在相应的检索基础之上的，用户根据自己的检索需求，通过检索表达式发出相应的检索指令，索引服务器会根据用户的需求进行检索表达式分析并通过与索引库中的信息进行配对。如果有匹配的网络信息资源，索引服务器会向数据库服务器请求输出相关的检索结果，用户还可以根据检索结果进行反馈或者修正检索表达式，直至检索到用户满意的信息资源。

4. 网络信息检索的评价

4.1 网络信息检索评价的概念

网络信息检索评价是指对互联网上的信息资源按照各方面的指标对其质量进行的全面具体的评价。

4.2 网络信息检索评价的指标

网络信息检索评价的指标有很多，可以将它们分为以下几个方面。

(1) 数据库内容

① 收录范围。收录范围对检索结果有直接影响，数据库收录内容的多少是信息检索

的基础。收录范围越大,检索结果就越可能全面完整,查全率就越高。

② 更新频率。互联网信息资源的开放自由和传播迅速使信息资源的更新变得更加频繁,内容更加丰富。网络信息资源的更新周期越短,信息资源检索结果的准确性就越高。

③ 数据库索引。数据库索引对信息检索的结果也十分重要。索引内容的完整性会大大提高信息检索的效率,使检索的结果更加精确,花费的检索时间更加短暂。

(2) 检索功能

① 检索方式。不同的检索方式会对检索结果产生不同的影响。如果能针对具体的检索环境利用合理的检索方式,则可以在一定程度上大大缩短检索的时间成本,提高检索效率。

② 检索技术。检索技术对信息检索提出了更高的要求。运用更多更先进的检索技术,例如布尔检索、截词检索、概念检索、引文检索等方法会使检索的范围更加广泛、结果更加可靠。

③ 检索限定。在网络信息检索中,每个数据库都会根据情况给出相应的限定词,如主题、作者、关键词、题名、刊名等,让用户根据自己的需求自由选择限定词的字段。检索限定越多,越利于检索,就越有可能找到所需要的资源。

(3) 检索效果

① 检全率。检全率是指对于一项特定的检索,实际检索出的文件数与所有应该被检出的文件数的比。它是衡量信息检索效率的重要指标,它可以反映出检测信息是否全面。

② 检准率。检准率是指对于一项特定的检索,实际检出的文件之中有多少比例是真正应该被检出的文件。它也是衡量信息检索效率的重要指标,可以反映出所检测信息的准确程度。

③ 响应时间。系统的响应时间对检索结果有一定的影响,当然,响应时间主要和网络的传输速度、通信设备等有关,如果系统的响应时间过长,则会降低用户的检索体验。

④ 检索结果的相关性。信息资源的检索结果特别繁多,如果用户从上至下逐条检索会花费大量的时间且收获甚微。按照相关性从高到低依次排序的检索结果可以让用户花最少的时间检索出最贴切的资源。

(4) 用户界面

用户界面是信息检索评价中很重要的一项指标,它能直观地反映出检索结果的大致信息。用户界面是否明确清晰、简洁明了、交互友好、方便易用等因素,都跟信息检索的评估有很大的关系。

第3节 网络信息检索的发展趋势

信息检索一直都是图书馆学、情报学的重要研究领域之一,随着网络技术的发展和互联网用户需求的不断变化,传统信息检索的缺点日益凸显。为了解决这些问题,学术界出现了一些新的信息检索理念,如跨语言信息检索、语义信息检索、可视化信息检索、协同

信息检索以及基于云计算的信息检索等，这些领域也是网络信息检索发展的趋势。本节将对其进行简单的介绍。

1. 跨语言信息检索

跨语言检索 (Cross Language Information Retrieval, CLIR) 是指可用一种提问语言检索出用另一种语言表达的信息，也就是一种跨越语言界限进行检索的系统。调查数据表明，互联网上 65.6% 的网页为英文，5.85% 的网页为日文，3.87% 为中文。另外，根据 Global Reach 最新统计数据，2007 年 12 月信息用户增长最快的三个语种，均为非英语，其中西班牙语网络用户增长 3.26%，日语用户增长 2.39%，法语用户增长 1.36%；以英语为母语的美国、英国信息网络用户增长也很迅速，但均没有超过 1%。可见，随着非英语互联网用户的大幅度增加，消除语言差异和障碍对信息检索的影响，提高网络信息资源检索的效率已经成为迫切需要解决的问题。

跨语言信息检索研究最早可追溯到 1973 年 G. Salton 的论文 *Experiments in Multilingual Information Retrieval*。当时的研究主要是针对国际联机检索进行的，由于检索系统不普及，因而人们对网络信息的需求并不强烈。20 世纪 90 年代后期，互联网的迅猛发展使得跨语言信息检索研究真正成为热点，互联网的全球化信息结构引发了对跨语言信息检索的迫切需要，这就促使越来越多的研究团体深入研究跨语言信息检索问题，并研制开发跨语言信息检索的不同方法。

跨语言信息检索研究涉及了语言学、情报学、计算机科学等多门学科知识，是一个综合性强、富有挑战性的研究领域。目前，跨语言信息检索中的文档的结构化、文档的语言识别以及文档的标引等基本问题依然是研究的热点问题和趋势。

2. 语义信息检索

传统的基于关键词的检索方式由于字义本身与其概念的延伸不在同一级上，使得查询的结果可能仅是在字面意义或者某层意义上的相匹配，例如：输入“苹果”，检索结果可能会包括表示水果的“苹果”和“苹果”电脑与手机等；输入“拍卖”可能会检索出“羽毛球拍卖完了”等信息，其“检准率”会受到影响。事实上，人们想要的往往是信息的概念及相关成分，语义检索 (Semantic Retrieval) 的出现正是为了解决这一问题。

通俗地讲，语义信息检索是指搜索引擎的工作不再拘泥于用户所输入请求语句的字面本身，而是透过现象看本质，准确地捕捉到用户所输入语句后面的真正意图，并以此来进行搜索，从而更准确地向用户返回最符合其需求的搜索结果。

语义信息检索涉及 Web 语义、本体 (Ontology)、人工智能领域的智能代理 (Intelligent Agent) 和机器学习等相关技术，目前这些技术尚处于发展的阶段，真正的语义信息检索尚需要进一步的努力才能实现。

3. 可视化信息检索

互联网和信息检索技术的高速发展，在一定程度上解决了人们查找信息的难题，但依然存在诸多问题，检索式表达、检索结果表达的可视化问题就是其中之一。近年来随着可

可视化技术 (Visualization Technology) 的发展, 网络信息检索的可视化, 特别是搜索引擎的可视化被越来越多的搜索引擎服务提供商所青睐。20 世纪 90 年代以来, 随着可视化技术的进步, 可视化信息检索系统得到了长足的发展。

可视化信息 (Visualization Information) 是指一切反映客观世界的可视信息媒体, 包括图片、动画、影像、视频等, 是一种反映在特定载体上的信息, 是可以加工、存储并被人们利用的信息。可视化信息的检索就是利用可视化技术设法为用户提供一个可视化的环境以支持用户完成信息检索、浏览、挖掘等超出传统的信息系统所能实现的功能。

可视化信息检索 (Visualization Information Retrieval) 受到国内外学者的广泛关注, 目前出现了一些新的研究领域, 比如动态的网络可视化信息检索系统, 就是一种对各可视化检索系统与网络导航系统相互兼容而形成的一种新的检索系统。

4. 协同信息检索

在现实生活中, 人们在很多与他人合作的任务中, 如旅行计划、组织社会活动或者完成家庭作业等, 有着协同检索的需求。Web2.0 及计算机支持的协同工作 (Computer Supported Cooperative Work, CSCW) 等新技术的发展, 使得协同信息检索开始成为信息检索领域新的研究热点。

简单来说, 协同信息检索是为了解决某个具体的协同检索问题, 适应某个具体目标、环境的检索系统。协同信息检索系统 (Cooperative Information Retrieval System) 与群组软件 (Groupware) 或者协同软件 (Collaborative Software), 以及 P2P (Peer-to-Peer) 技术、即时通讯 (Instant Messaging) 技术、聚类 (Clustering) 搜索技术、标引 (Indexing) 技术等有着密切的关系, 有效的协同信息检索系统是这些技术协同工作的结果。

协同信息检索系统在国内是一个新领域, 国外的研究也处于实验室研究阶段。现有的协同信息检索系统从实现功能看, 包括支持实时交流的 Coagmento 系统、CoSearch 系统等; 运用交互桌面实现协同数字资源搜索的 GroupLens 系统等; 支持推荐、评价的 Tapestry 系统等; 支持深度数据挖掘算法的 VisSearch 系统等; 隐性捕捉检索信息的 I-SPY 系统等。

5. 其他

近年来, 随着云计算 (Cloud Computing)、大数据 (Big Data) 等理念和技术的出现, 有人提出了基于云计算的大数据检索、基于云计算的信息检索以及云计算环境下的绿色信息检索等检索理念。另外, 有人对少儿信息检索、少年网络信息检索行为等也进行了研究。

参考文献

1. 隋莉萍. 网络信息检索与利用 [M]. 北京: 清华大学出版社, 2008.
2. 高飞. 网络信息实用检索 [M]. 北京: 中国计量出版社, 2006.

3. 陈泉. 网络信息资源检索与利用 [M]. 北京: 清华大学出版社, 2010.
4. 张洋. 网络信息资源开发与利用 [M]. 北京: 科学出版社, 2010.
5. 朱江岭. 网络信息资源检索 [M]. 北京: 海洋出版社, 2010.
6. 谷琦. 网络信息资源组织管理与利用 [M]. 北京: 科学出版社, 2008.
7. 朱红, 等. 网络信息检索与利用 [M]. 北京: 人民邮电出版社, 2010.
8. 符绍宏, 等. 互联网信息资源检索与利用 [M]. 北京: 清华大学出版社, 2012.