

Internet 上的科技信息检索

沈艺 (南京师范大学 210097)

摘要:本文首先介绍 Internet 信息检索的特点和常用的信息检索工具,然后分析 Internet 上带有科技信息指南性质的资源,并通过几个科技资源实例来说明如何从 Internet 上获取有用的科技信息。

1. 引言

Internet 上的信息资源极其丰富,在如此众多的信息资源里,寻找对自己有用的信息,不是一件简单的事情。需要对 Internet 上科技资源的分布状况和检索工具有较深入的认识。文中的资源地址以 URL 的形式给出。

2. Internet 信息检索特点

Internet 上的信息分布在世界各地的主机上,一般的查询,可以查到 400000 个主页、近百种报刊全文、出版社和新闻机构信息、用户新闻和原始技术报告等。传统的联机检索系统,由于技术上的原因,使检索只能局限于某一主机的特定数据库。Internet 信息检索可以同时使用多个主机甚至所有主机的某种资源而用户不必知道它们的具体地址。

Internet 发展十分迅速,其中的信息多而广。在目前谁都可以提供信息服务的情况下,其准确性和完整性难以保证。此外,随着网络的不断发展,许多信息在定期更新,因此在检索某个数据库以前,最好研究一下它的来源,或利用 Internet 信息检索工具系统地浏览所要检索的目标,经过多次选择才会发现最佳的信息源。

由于网络检索工具大多采用 Client/Server 结构,交互式 Windows 界面,其检索途径多,可进行“书签标记”,保留检索历史。在检索窗口按一定规则输入检索式,发出检索命令后即可得检索结果。检索结果通常包括题目、内容简介、http 地址等,由此可获得更详细的内容。

Internet 一个最重要的特点是它的动态性。一方面 Internet 的资源是一个动态系统,许多信息随时间的推移不断被更新;另一方面 Internet 提供了快速访问其资源的多种工具。从理论上讲,Internet 上一旦有某个信息存在,任何一个 Internet 用户就可以立即访问这个信息,Internet 因此成为当今世界获取最新信息的重要途径。

3. Internet 信息检索工具

网络信息检索工具大致分为交互式信息服务软件、名录服务和索引服务软件这三类。目前在 Internet 上运行的交互式信息服务软件有基于 Gopher 和 WWW,名录服务软件有基于 WHOIS、NETFIND 和 X.500,索引服务软件有基于 Archie、Veronica、Jughead 和 WAIS。其中 WWW 是 Internet 上最先进的网络信息检索系统,它把超文本技术、网络技术和多媒体技术融为一体,并把

Internet 上的信息按一定的规则组织起来, 提供查询。Archie 是为 FTP 资源服务, Veronica 和 Jughead 是为 Gopher 资源服务, WAIS 可作为整个 Internet 的文本信息资源服务的检索工具。

一个 Internet 检索工具应该具备以下功能: 首先应有一个简单易用、功能强大的界面; 其次是查询方式容易设定, 可以按用户所需的方式进行查询, 检索结果具有较好的准确性和可读性。否则就不是一个好的检索工具。下面简要介绍当前常用的几种检索工具。

(1) Infoseek. Infoseek 是商业性软件 Netscape 公司提供的检索工具, 分为特殊查询和一般查询两种, 其中特殊查询是有偿服务, 每月 9.95 美元, 按检索式查找全文文献中的前 30 个词, 并以相关程度输出查询结果, 结果的准确性较好。

(2) Yahoo. Yahoo 提供 WWW 服务的节点按主题建立分类索引, 并且有功能完备的检索方式, 非常有效, 但也异常拥挤。目前每天约有 400 万人次访问该节点。Microsoft 公司也选用 Yahoo 作为它的传播媒介。

(3) Lycos. Lycos 是卡耐基、梅隆大学的著名检索工具, 非常庞大, 即使有几台服务器同时提供服务, 有时负担仍然较重。《PC World》最近把 Lycos 列为最好的网络信息检索工具之一, 因为无论从信息质量, 还是从检索结果的相关性而言, Lycos 都是比较优秀的。

(4) WebCrawler. WebCrawler 收集了 35 万个以关键词检索为索引的 WWW 文档, 各个领域的内容由不同的专业人员维护。他可限时间检索, 限价格检索, 累积检索结果建立自己的数据库。

Internet 信息检索工具的特点在于他可让用户通过关键字来查询网上的信息, 给用户带来了极大的方便, 由于不同的检索工具由不同的组织分类归总, 检索结果的表达方法各不相同。一个好的检索工具应提供以下两种功能: 一种是检索元页, 它允许用户在一个节点访问多个不同的检索工具; 另一种是多线索检索页, 他不仅允许你指定使用哪些检索工具, 还可以设置你准备等候的时间, 因为一个复杂的多线索查询可能会持续几小时。

4. Internet 上科技信息指南

Internet 上的科技资源无论是内容还是形式都非常广泛, 用户要获得自己需要的信息, 常要花费相当多时间和精力。为此, 一些人或机构按照一定的方式, 对 Internet 上的信息进行组织和分类, 建立了科技信息指南, 方便信息的存取。

(1) Clearing house 的 Internet 学科资源指南。Clearing house 的 Internet 学科资源指南是美国密执根大学图书馆和信息与图书馆研究学校联合提供的服务。他收集和整理了 Internet 上资源信息, 按学科分类向用户提供各种资源的指南, 访问该指南的 URL 地址是:

Gopher://una.hh.lib.umich.edu:70/1/metdirs

路径为 Guides on the science

WWW://http2.sils.umich.edu/lou/chome.html

该指南所包括的学科分为人文学 (Humanities)、社会科学 (Social Sciences) 和自然科学 (Sciences) 三大类, 指南信息来源于一些 Internet 用户以及参加“发掘 Internet 资源”研究项目的学生。

(2) WWW 虚拟图书馆。WWW 虚拟图书馆是一个分布式资源系统, 内容包罗万象, 基本按照学科分类并依字母顺序排列, 大类中包括子类, 是一个树状系统。他们大都提供某一领域的多种资源, 并提供良好的工具和环境, 用户以访问这些“虚拟图书馆”为起点, 通过他所提供的与 Internet 上资源的连接, 可以较方便地了解并获得自己感兴趣的信息。访问 WWW 虚拟图书馆的 URL 地址为:

http://www.w3.org/hypertext/Data Sources/
bySubject/Overview.html

用户也可直接访问某个学科的虚拟图书馆, 如化学虚拟图书馆, 其 URL 地址为:

http://www.chem.ecla.edu/chempointer.html

另外, 一些 Internet 通用检索工具都不同程度地带有指南信息, 如 Yahoo, 其化学资源指南的 URL 地址为:

http://www.yahoo.com/Science/Chemistry

5. Internet 重要科技信息资源

Internet 上的科技资源非常丰富, 这里仅选取一些具有代表性的资源进行简单介绍, 以期对 Internet 上的科技资源有所认识。

(1) 出版物目录和文档拷贝。Titlenet 是由 Inforonics 公司在 Internet 上建立和维护的商业性图书目录服务。出版商通过它可以介绍自己的图书、杂志、光盘及数据库产品等。该目录是一个 Gopher 服务器, URL 地址为:

hopher://info.info.com

用户可以利用 Gopher 对图书目录进行浏览, 也可以通过作者名、书名和关键词三种方式对图书目录进行检索。所有书名都与一份订书单相连, 用户可以把它下载到自己的机器上, 打印后通过传真向出版商订购, 也可通

过 Internet 用 E-mail 订购。

除大出版商有自己独立的目录外, Titlenet 还提供一个小型出版商的联合目录, 用户可以对一些小型出版商的图书进行查询。出版商的情况也可以从 WWW 上查到, 其主页地址为:

<http://www.lights.com/publisher/>

除订购图书外, Internet 用户还可通过网络来获得 UnCover 的文章拷贝服务, 用户联机订阅杂志中的文章, 在 24 小时内将收到所订文章的传真。目前 UnCover 可进行联机服务的杂志约有 17000 种, 文章约 600 万篇, 其 URL 地址为:

<http://www.carl.org/uncover/unhome.html>

(2) Internet 上的专利服务。美国的 SPO 维护着一个专利全文数据库, 库中存放美国专利与商标局自 1972 年以来的专利约 170 万条。该数据库每周维护一次, 囊括所有最新的美国专利。使用前, 首先要注册, 即向 Spo-putent@spo.eds.com 发一封内容仅为“help”的电子邮件, 用户将收到答复, 当用户按照要求返回所签署的服务协议和收费单后, 就可以获得 SPO 的专利服务了, 并有三次免费服务。其 URL 地址为:

<http://www.spo.eds.com/patent.html>

此外, Internet 上收发电子邮件的用户可以通过订阅相应的通信组免费获得美国专利局最新专利信息。用户可按以下地址发出一封电子邮件: patent@world.std.com, 信件内容应包括一系列关键字, 这样, 用户即可获得自己感兴趣的内容。

(3) 联机检索系统。

·STN 系统

STN 是著名的国际科学技术信息网络系统, 他由三个著名的科技信息中心组成, 他们分别是:

①德国专业信息中心 FIZ, 地址为:

<http://www.fiz-karlsruhe.de>

②美国化学学会的化学文摘 CAS, 地址为:

<http://info.cas.org/ONLINE/online.html>

③日本科技信息中心 JICST, 地址为:

<http://www.jicst.go.jp/www/ServiceGuide95-e/stn.html>

三家信息中心通过海底电缆互相连接, 组成覆盖全球的信息网络系统。STN 主要提供科技信息的联机检索服

务, 用户入网时, 需先建立一个合法帐号。目前 STN 大约可提供 190 个数据库, 内容涉及几乎所有的科学技术领域, 还包括一些与商业有关的内容。

·DIALOG 系统

DIALOG 是世界上最大的联机信息服务系统, 所收集的信息包括科学、技术、医学、商业、贸易等。通过 DIALOG, 还可以联机阅读 100 多种报纸及数千种杂志, 找到有关化学物质、财政统计、环境问题、商标、版权等各方面的信息。Data-Star 是 DIALOG 系统的一个全作伙伴, 拥有 300 多个数据库, 并着重于欧洲的商业、医疗、药物等方面的信息服务。Data-Star 还在欧洲的报纸、公司目录、欧洲联盟的法律文件等方面有很强的服务能力。DIALOG 需要建立合法帐号才能联机, 其 URL 地址为:

<http://www.dialog.com/dialog/dialog.html>

此外, DIALOG 还向用户提供其他多种形式的服务, 如发行数据库软盘、专利、杂志、会议文集、标准以及其他资料的原始文献拷贝。并通过 Internet 的电子邮件向用户发布主题目录。

·美国国家标准与技术研究院 NIST

从 NIST 的 WWW 服务器上可以浏览到各个实验室的年度技术报告, 以及 NIST 实施的先进技术计划的介绍, 其地址为:

<gopher://zserve.nist.gov>

<http://www.nist.gov>

·美国国家宇航局 NASA SCAN 数据库

该数据库是一个研究论文和研究报告的数据库, 库中收集了航空和航天研究中最新的研究报告以及在杂志上最新发表的研究论文, 内容每月增加两次, 库中的研究报告及论文按学科分类存放。如果查到自己感兴趣的文章, 可以把全文传回来, 其 WWW 和 Gopher 地址分别是:

<http://www.sti.nasa.gov/STI-homepage.html>

<gopher://gopher.sti.nasa.gov>

用户通过匿名 FTP 服务器也可获得 NASA SCAN 的服务, 地址为:

<ftp.sti.nasa.gov>

注册时用“anonymous”作帐号, 保密字使用用户自己的 E-mail 地址。

(来稿时间: 1996 年 11 月)