

Internet (网)

用户接口

计算机网络

⑥

计算机科学 1996 Vol. 23 No. 6

Internet 用户接口

卢宏恩 周龙骥

(中国科学院数学所 北京 100080)

26-29, 42

TP393

A 摘要 Internet 已成为人类知识海洋的全球最大计算机网络。本文介绍了 Internet 的沿革及其用户接口,并展望了今年的工作。

一、Internet

Internet 这个被称为人类知识海洋的全球最大计算机网络,其产生多少出乎人们的意料,它不是一个组织严密的研究工作成果,而是由需求驱动开发,其成功不得不说是人类有史以来最伟大成就之一。从 4 个网点互连的 ARPANET,到目前遍布全球 90 多个国家和地区的 Internet,几乎所有的网络思想都从这里产生,Internet 的历史就是一部计算机网络史。Internet 的发展基本上经历了三个阶段:

1. 实验室阶段(1969~1985) Internet 由其前身 ARPANET 演变而来。60 年代中期,美国政府意识到计算机即将对教育、军事等各个领域带来巨大冲击,决定资助一个实验性的网络,以试验远程计算机之间的信息交换,这就是 ARPANET 计划。ARPANET 因其资助单位美国高级研究计划局(Advanced Research Projects Agency)而得名。

1969 年 9 月 2 日,ARPANET 正式开通,当时只有 4 个接口消息处理器(IMP),有幸成为这 4 个 ARPANET 首批节点的机构为:斯坦福研究所、加州大学圣巴巴拉分校、加州大学洛杉矶分校、犹他大学。这 4 个 IMP 的开通与各网点间的信息交换,标志着 ARPANET 的诞生。

1972 年 10 月在首届计算机与通信国际会议上 ARPANET 公开亮相,当时有 1,000 多名与会者在华盛顿哥伦比亚特区观看了 40 台左右的终端访问 ARPANET 上位于不同地点的大型计算机的情况,演示的高质量使得当时的与会者相信这一“完美无缺”的技术很快就会成功地运作,但实际上某些计算机和终端仅在演示前才入网,并未经过周密的测试。

到 1975 年,ARPANET 承担了大部分的美国国防部网络通信量,并移交给美国国防部,置于美国国防通信局管辖之下。

随着规模的不断增长,网络的功能也不断地被开发出来,除了被广泛应用的电子邮件 email 之外,如消息广播,ftp, telnet 及 USENET 的新闻系统也相继投入使用,网络的应用不再仅仅限于计算机的研究,已开始步入人们的日常生活。

1983 年 ARPANET 被分割成两个独立的部分:用于军事目的的 MILNET 和后来演变为 Internet 的民用网。

2. 基本建设阶段(1985~1990) 80 年代中期,随着需求的不断增长,美国国家科学基金会 NSF 在全美建立了 6 大超级计算中心,并资助了一个直接连接这些中心的主干网络 NSFNET,允许研究人员访问 Internet 进行信息共享。这一举措极大地推动了 Internet 的发展。这 6 大超级计算中心是:John von Neumann 国家超级计算机中心、圣地亚哥超级计算机中心、美国国立超级计算应用中心、康奈尔国家超级计算机研究室、匹兹堡超级计算机中心、国立大气研究中心的科学计算部分。连接 6 大超级计算中心的 NSFNET 极大地促进了 Internet 主干的发展,研究人员不但能很方便地利用那些超级计算机资源,而且可以得到 Internet 上的所有服务。为了使更多的机构入网,美国国家科学基金会 NSF 还为许多地区提供了启动基金。

这期间,美国全国范围内的高速网络主干快速形成,为 Internet 在世界范围内的飞速发展奠定了坚实的基础。

3. 飞速发展时期(1990~) 进入 90 年代,Internet 开始在世界范围内飞速发展,据有关资料统计,Internet 以每月 50 万台的人网量在增长,以这样的速度,到 2010 年前后世界上任何两个人都可以通过 Internet 进行交流,这不得不说是惊人的。

现在 Internet 已遍布全球包括我国在内的 90 多个国家和地区,国内连入 Internet 的网络有包括北大、清华和中科院三个子网、中国教育科研网 Cer-

net 和邮电部的 ChinaNET,以及最早入网的中科院高能所的结点。

随着 Internet 规模的增长,它的发展也面临着许多难题,其中安全性和检索能力是用户最为关切的问题。先前我们对网络安全认识不足,现已有迹象表明我们入网的机器也面临着非法入侵的危险,这一点希望能够引起大家的足够重视;检索问题关系到我们如何用好 Internet,Internet 好比人类知识的海洋,如果我们不能为用户提供迅速准确的信息检索能力,则用户要查找信息,就好比大海捞针,这势必极大地阻碍 Internet 的普及和发展。

近几年来,Internet 的惊人增长因素是多方面的,除了政治上克林顿政府信息高速公路计划的促进作用之外,环球网 WWW(World Wide Web)的兴起和发展也是不可忽视的。

二、WWW

一提及 Internet,我们不得不谈到 WWW(或 3W,Web),因为 WWW 已成为目前 Internet 上最受欢迎、最为流行、最新的信息检索服务程序。WWW 是在 Internet 的硬件联接基础之上,通过软件上的超链接(Hyperlink)将网络上的各种信息资源连接成一个虚拟的网,Web 上的信息可以说是包罗万象无所不有,如图形、文本、声音、动态图像等等,这些多媒体信息通过超链接组成超文本格式,用户可以通过 WWW 上相应的服务程序应用统一的图形用户界面(GUI),网上的信息资源形象生动地展现给用户,极大地方便了用户对网络上信息资源的查找。这也是 Web 以每年 20 倍的速度增长的重要原因。WWW 是第一个真正的全球性超媒体网络。

1. WWW 发展简史

WWW 这一项目最初是为了开发一个跨计算机平台、跨地域、全球范围内易于访问的分布式超媒体系统,最初设计并无过多的计算机专业人员参与,但却取得了惊人的成功。有关 WWW 发展的一些重要事件如下:

• 1989 年 3 月 欧洲粒子物理实验室(CERN)的科学家 Tim Berners-Lee 首先提出了 WWW 这一新概念,目的是为了更方便 CERN 的科学家进行学术交流。原先 CERN 的研究多在相距很远的异地进行,要进行一次简单的学术活动,通常得运行很多不同的软件,对于科学家们来说非常不方便,于是寻求一种方法让世界各地的研究小组能够很快很方便地共享最新的研究信息。

• 1990 年 10 月 项目的建议得到了 CERN 管理部门的鼓励并进行了重新规划,Robert Cailliau 成为另一个合著者,WWW(World Wide Web)这一名称被正式确定下来。

• 1990 年 11 月 最初的 WWW 原型在 Next 平台上开发成功。

• 1993 年 9 月 WWW 的通信量达到 NSF 主干网的 1%,NCSA 发布通用平台如 X,PC/Windows 和 Macintosh 上的 Mosaic 浏览器工作版本。

• 1993 年 12 月 WWW 获 IMA 奖,许多新闻媒体对 Internet 和 WWW 进行了分析和评述。

• 1994 年 1 月 O'Reilly, Spry 等公司宣布 "Internet in a box" 产品,将 Web 引入家庭。

• 1994 年 3 月 Web 上著名的浏览程序 Mosaic 的创作者 Marc Andreessen 与其同事离开 NCSA 组建 Mosaic 通讯公司(现在的 Netscape 公司),

• 1994 年 5 月 25~27 日 第一届 WWW 国际会议在日内瓦 CERN 召开,主题为 "Woodstock of the Web"

• 1994 年 8 月 在波士顿由 CERN 和 NCSA 创建了国际 WWW 会议委员会 IW3C2。

• 1994 年 10 月 16~20 日 第二届 WWW 国际会议在芝加哥召开,主题为 "Mosaic and the web"

• 1995 年 4 月 第三届 WWW 国际会议在德国的达特茅斯召开,主题为 "Tools and Applications"。

WWW 的历史虽短,但其发展却非常迅猛,近两年来 WWW 已成为访问 Internet 资源最为流行的手段,WWW 服务器也以每年 20 倍的速度增长,据统计目前总数已超过 1.5 万台。同时一股巨大的 WWW 研究开发热潮也正在兴起,每年两次的世界 WWW 大会即可窥见一斑。

2. WWW 的现状与未来

WWW 通过超链接将 Internet 上的各种信息资源有效地组织起来,用户能够在统一的界面上对各种资源进行访问。正是这样,WWW 才得以迅猛的发展。各个层次的用户把包罗万象、形形色色的信息链入 WWW,WWW 很快成为浩瀚的知识和信息的海洋。正如 Internet 一样,Web 的发展也处于无政府状态,虽然有一些组织在管理和协调,但巨大的信息增长,特别是多媒体信息的引入,使得对 Web 的访问有时并不像人们想象的那样便捷,有时甚至会迷失方向。

除了应对网上的信息提出一个新的分类标准之

外,如何开发出更能满足用户需求的功能更强大的信息浏览和检索工具,也是 Web 发展所面临的难题。

Internet 为我们提供的服务是多种多样的,如 email, telnet, ftp, news, USENET 等等。其中两类基于 WWW 的软件为我们更好地利用 Internet 提供了强有力的工具,这就是浏览器和检索工具。

三、浏览器

浏览器是这样一类接口程序,它们将用户给出的 URL (Universal Resource Locator) 上的信息以相应的格式展示给用户,这些信息多半是由超文本标记语言 HTML (Hyper Text Markup Language) 编写的超文本信息,用户可以通过这些超文本上的超链接访问其它相关的信息。超文本标记语言可以将文本、图形、图像、声音等多媒体信息编辑在一起,用户得到的就是一个丰富多采的多媒体界面。

浏览器也是由字符型转向超媒体形式的,下面介绍现在较为流行并且比较有特色的几种浏览器:

1. Mosaic 是 NCSA 的 Marc Andreessen 等人开发的一个超媒体的浏览程序,已成为事实上的浏览器标准,现在已超过 2,000 万份的拷贝。

NCSA Mosaic 这一软件是为在 Internet 上查寻和检索信息而设计的一个分布式超媒体系统,为不同的协议、数据格式以及遍布 Internet 的信息服务器提供了统一的界面。用户键入的 URL 包括三个部分:读取相应数据所用的协议和方法、服务器的地址及文件所处的路径,如 <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MetaIndex.html>。其中 http 标明读取这一数据是应用超文本传输协议 http (Hyper Text Transport Protocol); www.ncsa.uiuc.edu 指文件所处服务器的地址是 NCSA 的 Web 服务器;而其余的部分给出了所要读取的文件在服务器上的路径。Mosaic 不仅支持 http 协议,而且同样支持 ftp, telnet, gopher 等其它协议,这样用户就可以在统一的界面下,对各种服务器上采用不同协议存取的数据进行访问。Mosaic 的主要特性有:

- 可显示纯文本和超媒体图形、图像信息、友好的图形用户界面;

- 可记录并标记用户已访问过的信息;

- 对用户重要的或经常用到的文档可通过热点表进行快速的存取;

- 提供了同一文档中和遍布 Internet 的搜索能力;

- 可以对 Internet 上任何地点的文档用文字或声音进行标记;

- 完全支持基于 TCP/IP 的通讯;

- 对任意的 Viewer 和其它数据格式可容易地扩展。

Mosaic 的开发成功从根本上改变了用户访问 Internet 的方式,先前用户要应用不同的软件和繁杂的命令实现所需的功能,现在只需在统一的图形用户界面下,用鼠标来敲击相应的按钮即可完成,一切都变得那么简单,正因为这样 Mosaic 取得了极大的成功并成为事实上的标准。

2. Netscape Navigator 是建立在 Mosaic 模型上的 Web 浏览器,由 Netscape 通讯公司开发,该公司由 Mosaic 的原开发者 Marc Andreessen 同 SGI 公司的创造人 James Clark 于 1994 年 5 月共同创建的,今年该公司的股票一上市就创造了一个奇迹,当天就使资金只有 1,700 万美元的小公司一夜间便成为一个 20 亿美元的新巨人,价值足足增加 100 多倍,年仅 24 岁的 Andreessen 本人也成为继 Bill Gates 之后,美国青少年的又一新偶像。

Navigator 是 Andreessen 离开伊利诺依大学后,对 Mosaic 进行进一步改进后而推出的更为精致的 Web 浏览程序。它除了保留 Mosaic 原有的特点之外,还克服了 Mosaic 的一些不足。Navigator 的一大优势就是加快了信息传输的速度,首先传送文本信息,推迟图形、图像等多媒体信息的传输,在有图形的地方留出一个标记,文本传送完毕后就可根据已看到的信息去选择所需的链接。这样用户就不必象 Mosaic 那样非要等到整个超媒体文件都传送完毕,才能看到文本和图形信息,然后再根据用户意愿去访问其他的超链接。这不但加快了用户的检索速度而且减少了不必要的信息传送。这一改进使 Navigator 在现有的浏览器中表现出了最好的速度优势。据测试 Navigator 的平均下载速度要比 Mosaic 快一倍以上,这同图形等多媒体信息与整个文件的比例有一定关系, Mosaic 在这方面已有一定的改进。

Netscape Navigator 的另一聪明之举在于它将 Web 上的另一类强有力的检索工具的 URL 列入它的按钮之内,用户可以很方便地通过 Navigator 去利用这些工具进行信息搜索。

3. Hot Java Mosaic 和 Navigator 所代表的一类浏览器,展示给用户的信息多是静态的不可交互的,并且要求用户要安装一些其他的应用程序来观

看相应的图形和图像数据。

Sun 公司推出的 HotJava(热咖啡,据说因开发人员在工作中多饮用热咖啡而得名)使 Internet“活起来”,它在 Mosaic 建立的 Internet 浏览技术基础上进行了扩展,将静态的数据转换为动态的应用以实现浏览器任意动作的能力,用户就可以实现交互式的动态应用,如交互式广告和个性化的报纸等。

另外,HotJava 为用户提供了存取这些应用的一种新方式,软件在网络上透明地移动,无需“安装”软件,用户所需要的软件会自动装入你的系统,这种透明地获取应用程序的能力大大解除了固定的媒体类型,如图像、文本的束缚,使开发者可以自由发挥。

HotJava 拥有这些动态特性源于它是用一种叫做 Java 的新语言编写而成。概括地说,Java 是一种简化的、安全的可移植的 C++ 语言,该语言是分布式的、可解释的、动态的,并且具有神经元的体系结构,用 Java 编写的代码就可运行于不同的 CPU 平台上。

HotJava 与其他的浏览器的本质区别在于,HotJava 本质上无需另外一些浏览器拥有许多具体的固定装配的知识,不必利用这些知识来识别数据类型、网络协议和网络导航的行为。它可以在应用中不断地自动升级,不但可以使静态的数据动起来,而且为添加新的功能提供了极大的灵活性。HotJava 这一优势是因为 Java 语言将要浏览和演示的数据与运行它们的小应用程序 Applet 关联起来,当 HotJava 发现用户的系统没有这些 Applet 时,就会自动地从远程服务器上相应的 Applet 自动装配到用户的系统上,用户不必担心没有安装必要的软件。为了加快访问速度,HotJava 可以推迟图像和 Applet 的读取。

HotJava 在人机交互方面进行了十分有益的改进。

4. WebFORCE 在各大计算机公司相继推出自己的 Web 产品之时,以图形功能见长的 SGI 公司也不示弱,推出了 WebFORCE 的 WWW 解决方案,它包括 SGI 的高性能图形工作站等硬件产品,以及 Netscape Navigator 的辅助创作工具 WebMagic,它们为用户制作 Web 上的主页(Home Page)提供了一个功能强大、易于使用的动态 WYSIWYG HTML 编辑器。先前用户制作 Home Page 只能用文本编辑器来制作,而这种工具使用户以所见即所得的方式直观形象地编辑 HomePage,并可很容易地加入声音、图像等多媒体信息。

同样 SGI 公司也在 HomePage 的三维立体化方面进行了尝试,用户可以看到富于真实感的三维的 HomePage,其中的一些三维实体可做成超链接,用户可以通过敲击相应的实体来访问其他相关的信息。

除了浏览器以外,WWW 上另一类有力的工具就是搜索/检索工具。

四、检索工具及其它工具

为了更好地帮助用户在浩如烟海的 Internet 上查找所需的信息,开发出了很多索引和检索工具,其中比较著名的有:

- Lycos 由卡内基梅隆大学开发,据称已对 Internet 的 91%进行了索引,称为“Internet 的目录”
- WebCrawler 由华盛顿大学的 Brian Pinkerton 开发。

还有 World-Wide Web Worm、Harvest Broker、Discovery 等。

这些工具的基本构成有:搜索器(search engine),负责决定整个工具的行为,如决定下一步搜索那一个文件,并激活文件检索;数据库(database)对搜索到的有关信息做出索引,并记录有关的链接,将它们以一定的顺序记录下来;Agent 负责在搜索器指引的方向下从网络上检索相应的文件;查询服务器 Query Server 负责处理用户以关键字等方式要求的查询请求。

这类工具对用户给出的以关键字方式的查询请求,提供相应的信息地址,并对它们以某种方式进行排序。

由于 Internet 上信息的爆炸性增长,尽管这些工具不断地对网上的信息进行搜寻,以保证它们数据库中的信息的新鲜和可靠,但由于信息的庞杂和飞速增长,这一点是很难做到的。

另外,虽然关键字可以在一定程度上表达用户的查询意愿,但由于每个人的表达方式和知识范围的不同,检索工具给出的信息并不能完全满足用户的需求,有时甚至会有很大的偏差。

另一类工具软件是信息检索软件,它们加入了字典转换成份,可以对用户给出的关键字进行相应的转换,这类软件多应用于电子图书馆。

还有 MIT 媒体实验室的 Hery Lieberman 设计了一个辅助浏览软件,它通过监视用户浏览 Web 的行为,推测出用户的意图,并根据自己的知识给出相应的建议。

(下转第 42 页)

实际工作的需要。数据复杂性使得需要更多的领域知识,巨大的数据库对算法的效率提出更高的要求,不断变化的环境和信息种类(如多媒体信息)需要新的发现方法,复杂的问题可能需要多种发现策略协作,对社会领域、商业领域的数据库的发现要当心可能是违法的。

自主性(Autonomy)和普适性(Versatility)是KDD系统所追求的两个标准,自主性要求更多的领域知识,而普适性则要求相对地领域无关。这种看上去的不和谐可以得到一定程度的改善。首先,一些领域知识,比如数据依赖,是可以从数据库中自动提取的。其次,我们不妨设计更好的算法从用户那里得到知识,这自然需要功能强大的交互式工具。再次,多重发现的各种结果及阶段知识都可用来指导进一步的发现。

数据库中的数据,尤其是非数值型的数据,很难给出一种贴切的表示使之在数据空间中有合适的位置。而且,由于数据的随机性和数据间的相对弱关联,数据空间可能呈现出稀疏状态。这些都影响发现的准确性和效率。

参考文献

[1] Jiawei Han, Yongjian Fu, Discovery of Multiple-Level

Association Rules from Large Databases, In Proc. 21th Intl. Conf. on VLDB Zurich, Switzerland, 1995

[2] Ramakrishnan S. et al., Mining Generalized Association Rules, Same to [1]

[3] Hongjun Lu et al., NeuroRule: A Connectionist Approach to Data Mining, Same to [1]

[4] David K. Y. Chiu et al., Information Discovery through Hierarchical Maximum Entropy Discretization and Synthesis, in G. Piatetsky-Shapiro and W. J. Frawley (eds.), Knowledge Discovery in Database, AAAI/MIT Press, 1991

[5] Jiawei Han et al., Knowledge Discovery in Database: An Attribute-Oriented Approach, In Proc. 18th Intl. Conf. on VLDB Vancouver, Canada, 1992

[6] W. J. Frawley et al., Knowledge Discovery in Databases: An Overview, AI Magazine, Fall, 1992

[7] C. J. Matheus et al., System for Knowledge Discovery in Databases, IEEE Trans. Knowl. Data Eng., 5(6), 1993

[8] Rakesh Agrawal et al., Database Mining: A Performance Perspective, Same to [7]

[9] Kenneth A. Fauflman et al., Mining for Knowledge in Databases: Goals and General Description of the INLEN System, Same to [4]

(上接第29页)

总之,这类帮助用户查询和检索的工具也是当前研究开发的热点,很多软件也正在不断地开发和完善之中。

这里,举一个简单的例子。当用户在Internet上查找信息时,基本上有两种情况:一是已知要找信息的URL,这时比较简单,只需运行相应的浏览器,键入URL即可,如要浏览WWW国际会议的相关信息,只要输入 <http://www.elsevier.nl/cgi-bin/ID/www94>,然后通过超链接进行浏览并可将需要的论文和相关信息拷贝到自己的用户中。

另一种情况是不知道URL地址,这时就需要借助一些检索工具来找出相应的URL或其它线索。上文提到的工具都可做到这一点,用户可以利用这些检索工具,输入相应的关键词,就可得到有关的URL或其他帮助信息。Netscape公司将一些常用的检索工具做成了一个文档,用户可以通过访问这一文档来激发相应的工具,其URL是 <http://www.netscape.com/home/internet-search.html>。

五、今后工作

Internet的用户接口除了应在用户的视觉效果和可用性方面要进行改进之外,关键的是要智能化。因为Internet的巨大的信息量使得一个人穷尽其一生也是无法理解的,这就要求我们开发出具有智能化的接口,为用户用好Internet提供帮助。

自然语言的接口是很有必要的,不仅是对用户查询的理解,而且要对网上的文章进行理解,这样才能更好地同用户的意愿相匹配,这种自然语言理解能力并不需要十分强大,但却是十分必要而有意义的。此外,对网络上新信息的学习和发现能力,也很重要。

总之,我们如何把AI、DAI和相关学科中的一些有用的方法和知识应用到Internet这一飞速发展的领域中,在引进技术、引进人才之后如何更好地利用高科技引进信息为我所用,是我们面临的一个课题和挑战。