

“一带一路”
国情文化
丛书

南亚国家 网络信息检索导论

NANYA GUOJIA WANGLUO XINXI JIANSUO DAOLUN

何杰◎主编



世界图书出版公司

“一带一路”
国情文化
丛书

南亚国家 网络信息检索导论

NANYA GUOJIA WANGLUO XINXI JIANSUO DAOLUN

何杰◎主编



世界图书出版公司

广州·上海·西安·北京

图书在版编目 (CIP) 数据

南亚国家网络信息检索导论 / 何杰主编. — 广州:
世界图书出版广东有限公司, 2018.11
ISBN 978-7-5192-5173-4

I. ①南… II. ①何… III. ①网络检索—研究—南亚
IV. ①G254.92

中国版本图书馆 CIP 数据核字 (2018) 第 231551 号

书 名	南亚国家网络信息检索导论
	NANYA GUOJIA WANGLUO XINXI JIANSUO DAOLUN
主 编	何 杰
策划编辑	刘正武
责任编辑	张东文
出版发行	世界图书出版广东有限公司
地 址	广州市海珠区新港西路大江冲 25 号
邮 编	510300
发行电话	020-84451969 84459539
网 址	http://www.gdst.com.cn/
邮 箱	wpc_gdst@163.com
经 销	新华书店
印 刷	广州市迪桦彩印有限公司
开 本	787 mm × 1092 mm 1/16
印 张	23
字 数	495 千字
版 次	2018 年 11 月第 1 版 2018 年 11 月第 1 次印刷
国际书号	ISBN 978-7-5192-5173-4
定 价	65.00 元

版权所有 侵权必究
咨询、投稿: 020-84460251 gzlw@126.com
(如有印装错误, 请与出版社联系)

目 录

CONTENTS

第一章 南亚国家互联网发展概况	1
第一节 互联网概述.....	1
一、Internet 简介	1
二、互联网基础.....	4
第二节 南亚国家互联网发展的历史与现状.....	10
一、印度互联网发展概况	10
二、巴基斯坦互联网发展概况	16
三、孟加拉国互联网发展概况	19
四、斯里兰卡互联网发展概况	25
五、尼泊尔互联网发展概况	33
六、阿富汗互联网发展概况	35
七、不丹互联网发展概况	41
八、马尔代夫互联网发展概况	42
第二章 南亚国家网络信息资源检索与利用	45
第一节 网络信息资源检索与利用.....	45
一、网络信息资源.....	45
二、网络信息资源检索	48
三、网络信息资源利用	55
第二节 南亚国家互联网搜索引擎的使用	65
一、印度互联网搜索引擎及其应用	65
二、巴基斯坦互联网搜索引擎及其应用	80
三、孟加拉国互联网搜索引擎及其应用	85
四、斯里兰卡互联网搜索引擎及其使用	87

五、阿富汗互联网搜索引擎及其应用.....	95
六、马尔代夫互联网搜索引擎及其使用.....	97
第三章 南亚国家门户网站与南盟主要网站.....	99
第一节 南亚国家门户网站	99
一、印度主要门户网站	99
二、巴基斯坦主要门户网站.....	102
三、孟加拉国主要门户网站.....	108
四、斯里兰卡主要门户网站.....	109
五、尼泊尔主要门户网站.....	113
六、阿富汗主要门户网站.....	118
第二节 南盟主要网站	119
一、南盟秘书处网站	120
二、南盟文献中心网站	122
三、南盟信息中心网站	123
四、南盟能源中心网站	124
五、南盟发展基金网站	126
六、南亚大学网站	128
第四章 南亚国家网络媒体信息检索	131
第一节 南亚国家网络新闻媒体信息检索	131
一、印度网络新闻媒体信息检索.....	131
二、巴基斯坦网络新闻媒体信息检索.....	136
三、孟加拉国网络新闻媒体信息检索.....	144
四、斯里兰卡网络新闻媒体信息检索.....	149
五、尼泊尔网络新闻媒体信息检索.....	156
六、阿富汗网络新闻媒体信息检索.....	161
七、不丹网络新闻媒体信息检索.....	165
八、马尔代夫网络新闻媒体信息检索.....	168
第二节 南亚国家广播电视媒体网络信息检索	171
一、印度广播电视媒体网络信息检索.....	171
二、巴基斯坦广播电视媒体网络信息检索.....	176

三、孟加拉国广播电视媒体网络信息检索	182
四、斯里兰卡广播电视媒体网络信息检索	186
五、尼泊尔广播电视媒体网络信息检索	192
六、阿富汗广播电视媒体网络信息检索	195
七、不丹广播电视媒体网络信息检索	202
八、马尔代夫广播电视媒体网络信息检索	203
第五章 南亚国家经济与贸易类网络信息检索	207
第一节 南亚国家经济类网络信息检索	207
一、印度经济类网络信息检索	207
二、巴基斯坦经济类网络信息检索	212
三、孟加拉国经济类网络信息检索	220
四、斯里兰卡经济类网络信息检索	227
五、尼泊尔经济类网络信息检索	234
六、阿富汗经济类网络信息检索	239
七、不丹经济类网络信息检索	245
八、马尔代夫经济类网络信息检索	251
第二节 南亚国家贸易类网络信息检索	255
一、印度贸易类网络信息检索	255
二、巴基斯坦贸易类网络信息检索	260
三、孟加拉国贸易类网络信息检索	266
四、斯里兰卡贸易类网络信息检索	269
五、尼泊尔贸易类网络信息检索	276
六、阿富汗贸易类网络信息检索	278
七、不丹贸易类网络信息检索	280
八、马尔代夫贸易类网络信息检索	280
第六章 南亚国家教育与文化类网络信息检索	283
第一节 南亚国家教育类网络信息检索	283
一、印度教育类网络信息检索	283
二、巴基斯坦教育类网络信息检索	288
三、孟加拉国教育类网络信息检索	295

四、斯里兰卡教育类网络信息检索.....	299
五、尼泊尔教育类网络信息检索.....	308
六、阿富汗教育类网络信息检索.....	313
七、不丹教育类网络信息检索.....	319
八、马尔代夫教育类网络信息检索.....	321
第二节 南亚国家文化类网络信息检索	323
一、印度文化类网络信息检索.....	323
二、巴基斯坦文化类网络信息检索.....	328
三、孟加拉国文化类网络信息检索.....	332
四、斯里兰卡文化类网络信息检索.....	333
五、尼泊尔文化类网络信息检索.....	339
六、阿富汗文化类网络信息检索.....	341
七、不丹文化类网络信息检索.....	343
八、马尔代夫文化类网络信息检索.....	345
附录 南亚国家热门网站	347
参考文献	359
后 记	361

第一章 南亚国家互联网发展概况

第一节 互联网概述

一、Internet 简介

Internet (因特网), 又名互联网, 它是一个开放的、互联的、遍及全球的计算机网络, 是一个能使世界上各种不同类型的计算机之间交换各种数据信息的通信媒介, 是一个国际间的通信网络集合体。

Internet 集现代通信技术和计算机技术于一体, 是计算机之间进行信息交流和实现资源共享的良好手段。通过 Internet, 人们可以实现全球范围的电子邮件、信息查询与浏览、文件传输、电子新闻、语音与图像通信服务等功能。目前, Internet 已经成为覆盖全球的信息服务基础设施之一。

(一) Internet 的发展历程

Internet 从 20 世纪 60 年代末开始研究到现在, 经历了以 ARPAnet 的诞生为标志的起源阶段, 以美国国家科学基金会网络 NFSnet 的建立为标志的实用化阶段, 以美国国内商业互联网 ANSnet 的形成成为标志的商业化阶段, 以及世界各地互联网纷纷加入的公众化阶段。

1. Internet 的起源

20 世纪 60 年代末, 美国国防部考虑到, 对于传统的用于军事指挥的计算机网络, 一旦遭受到敌国的核打击, 摧毁军事指挥中枢中的某一台关键计算机, 或切断这台计算机与其他部分的联系, 整个网络就会瘫痪, 后果不堪设想。美国军方认为, 有必要开发一个新型的计算机网络, 当其受到攻击时, 即使部分网络被摧毁, 其余部分可以自动建立起新的联系, 保持正常工作。1969 年, 美国国防部高级研究计划局 (Defense Advanced Research Projects Agency, DARPA) 将美国的几个军事及研究中心的计算机主机连接起来, 从此诞生了以军事为目的的 ARPAnet, 这就是当今 Internet 的雏形。

美国国防部高级研究计划局从一开始建网,就考虑到在这个网络上可以使用不同类型的计算机和不同类型的操作系统,这样可使每个原有的计算机不必更换操作系统就可以加入 ARPAnet 运行。为此,他们开发研制了至今仍在使用的 TCP/IP(传输控制协议/网际协议)通信协议。因此,人们通常认为 ARPAnet 是 Internet 的起源。1990 年,ARPAnet 在完成了历史使命后停止运作。

2. Internet 的实用化

ARPAnet 的成功,使许多人清楚地意识到网络的巨大潜力。20 世纪 80 年代中期,美国国家科学基金会(National Science Foundation, NSF)建立了一个支持科研和教育的全国性计算机网络 NSFnet。NSFnet 使用 TCP/IP 协议,把美国的五个超级计算机中心连接起来。随后,许多大学、政府科研机构以及私营科研机构都纷纷将自己的局域网并入 NSFnet。这一时期的 NSFnet 以惊人的速度增长。到 1988 年,NSFnet 已经接替原有的 ARPAnet 而成为这一时期的 Internet 主干网,是 Internet 发展的又一个里程碑。

3. Internet 的商业化

20 世纪 90 年代以前,Internet 的使用一直仅限于教育、军事和科学研究领域,商业性机构一直受到限制。1992 年,专门为 NSFnet 建立高速通信线路的高级网络和服务公司(Advanced Networks and Services)建立了一个传输速率为 NSFnet 30 倍的商业化 Internet 骨干通道——ANSnet。Internet 的主干网由 NSFnet 更新换代为 ANSnet,这是 Internet 向商业化过渡的关键一步。一些公司开始利用 Internet 提供商业服务、收集资料、发布商业广告等,甚至还出现了专为个人或单位接入 Internet 提供产品和服务的公司——互联网服务供应商(Internet Service Provider, ISP)。随着 ANSnet 的不断发展,曾在 Internet 发展中起过重要作用的 NSFnet 于 1995 年 4 月正式关闭。

4. Internet 的公众化

随着世界各地无数的企业和个人纷纷涌入 Internet,Internet 在通信、信息检索和客户服务等方面的巨大潜力不断被挖掘出来,从而给 Internet 带来了一个新的飞跃。

Internet 的应用范围从过去单纯的通信、教育和信息查询等领域向更具效益的商业领域扩张,许多公司和信息服务商在 Internet 上建立了自己的网站。人们可以通过 Internet 进行网上购物、网上就诊、网上教育、网上求职、网上休闲等。可以说,Internet 已经深入到人们日常生活的各个角落。

到 2014 年 8 月,全球活跃社交用户突破了 20 亿人;到 2014 年 11 月,全球活跃

互联网用户突破了 30 亿人。脸谱 (Facebook) 是全球最受欢迎的社交平台, 拥有超过 13 亿活跃用户; 腾讯 QQ 紧随其后, 拥有活跃用户 8.2 亿; 微信 (WeChat) 排在第 6 位, 拥有活跃用户 4 亿。

(二) Internet 的结构与组成

1. Internet 的结构

Internet 是覆盖全球范围的一个大型网络, 由无数个分布在各个国家、各个地区的各个政府、企业、学校、社区等局域网, 通过路由器而连接成的国际互联网络。从逻辑结构来看, 同一般的计算机网络一样, Internet 也可以将整个网络划分为通信子网和资源子网。通信子网包括全球范围的广域网通信线路、分布在世界各地的大小不等的局域网中的通信设备和传输介质以及用于将这些网络连接起来的路由器; Internet 的资源子网则是由世界各地局域网中的服务器、客户机、信息资源以及其他相关的网络资源等组成。从 Internet 使用者的角度来看, Internet 是一个由分布在世界各地的无数个计算机连接而成的巨大信息服务平台。Internet 的内部结构对于 Internet 用户来说是完全透明的。接入 Internet 的主机既可以是信息资源的使用者, 也可以是信息资源的提供者。

2. Internet 的组成

Internet 主要是由通信线路、路由器、局域网以及信息资源组成。

(1) 通信线路

通信线路是连接各个局域网和路由器, 从而使它们连成一个更大的覆盖全球范围的国际互联网络的基础通信设施。通信线路一般分为有线和无线两种, 有线如光纤、双绞线等, 无线如卫星通信等。

目前, Internet 使用的广域网通信线路主要是光纤和卫星通信。光纤的数据信号传输距离达几百甚至几千千米, 且具有带宽较大、误码率较低、安全性较好、抗干扰性较强等特点; 卫星通信的覆盖范围较大, 一颗同步卫星可以覆盖地球 1/3 面积。局域网中一般采用光纤和双绞线。

(2) 路由器

路由器是 Internet 中最重要的设备之一, 其主要作用是将各个局域网或广域网连接

起来,从而形成一个覆盖全球的信息网络。

路由器的功能是在不同的网络之间存储、转发、分组数据,并进行路由选择、协议转换、多路重发和错误检测等。当路由器接收到从某一个网络传输来的数据包时,它根据数据包的包头信息中所指示的目的地址,通过路由选择算法为数据包选择一条最佳的输出路径,并负责存储和转发数据包。一般情况下,数据从源主机到达目的主机,中间要经过多个路由器和多个网络。

(3) 局域网

采用 TCP/IP 通信协议的局域网是 Internet 的基本组成细胞。每个局域网通过通信线路和路由器连接在一起就成了 Internet 的一部分,千万个局域网互联就构成了 Internet。在这些局域网中除了包含有必备的网络通信设备及传输介质外,还有成千上万台主机。根据在网络中扮演的角色不同,这些主机又可分为服务器和客户机。

服务器是信息资源与网络服务的提供者,它一般是性能较好、存储容量较大的计算机。服务器中可以安装各种提供一定服务和信息资源的服务器程序。根据所提供的服务不同,服务器又可分为 WWW 服务器、FTP 服务器、Mail 服务器、文件服务器、域名服务器等。

客户机是信息资源和服务的受用者,它可以是普通的微机或移动终端。用户通过使用客户机中的各种客户端软件,来使用服务器所提供的相应资源和服务。

(4) 信息资源

在 Internet 中,信息资源是最宝贵的资源。信息资源一般都存储于分布在世界各地的局域网的服务器中,人们可以通过访问这些服务器来得到他们想获取的各种信息。在这样一个庞大的信息海洋中,如何能方便快捷地获得自己所关心的信息呢? WWW 服务的出现使得信息资源的组织方式更加合理,而搜索引擎的出现使信息的检索更加快捷。目前,全球有许多优秀的门户网站和搜索引擎,如:雅虎(www.yahoo.com)、谷歌(www.google.com)、百度(www.baidu.com)等。

二、互联网基础

(一) TCP/IP 协议

为了保证计算机网络中每台计算机都能正常通信,就必须有一套网络中各个节点共同遵守的规程和约定,这些规程和约定就是网络协议。

TCP/IP 协议是在 Internet 中使用的通信协议，是以传输控制协议 TCP（Transmission Control Protocol）和网际协议 IP（Internet Protocol）为核心的一组协议。其作用是连接在 Internet 中的每台计算机，不论是否属于同一类型，也不论是否使用相同的操作系统，都能方便地进行数据传输和实现资源共享。

(二) IP 地址和域名系统

为了在网络环境下实现计算机之间的通信，网络中的任何一台计算机必须有一个不允许重复的地址，这个唯一的地址将保证数据传输准确无误。Internet 用两种方法来标识网络上的计算机，即 Internet 地址（IP 地址）和域名系统。

1. IP 地址

IP 地址是指互联网协议地址（Internet Protocol Address），又译为网际协议地址。IP 地址是 IP 协议提供的一种统一的地址格式，它为互联网上的每一个网络和每一台主机分配一个逻辑地址，以此来屏蔽物理地址的差异。

IP 地址是一个 32 位二进制数字，通常表示成 4 组十进制数字，每个数字可取值 0—255，其间用“.”号间隔。如百度首页的 IP 地址是 202.108.22.5。

IP 地址现由互联网名字与号码指派公司 ICANN（Internet Corporation for Assigned Names and Numbers）分配。

IP 地址分为 A、B、C、D、E 五类，它们适用的类型分别为：大型网络、中型网络、小型网络、多目地址、备用。其中 A、B、C 较为常用，D、E 类为特殊地址。

表 1-1 A、B、C 三类 IP 地址

类别	最大网络数	IP 地址范围	最大主机数	私有 IP 地址范围
A	126 (2 ⁷ -2)	0.0.0.0—127.255.255.255	16777214	10.0.0.0—10.255.255.255
B	16384 (2 ¹⁴)	128.0.0.0—191.255.255.255	65534	172.16.0.0—172.31.255.255
C	2097152 (2 ²¹)	192.0.0.0—223.255.255.255	254	192.168.0.0—192.168.255.255

2.域名系统

Internet 的域名结构由 TCP/IP 协议中的域名系统（DNS）进行定义。Internet 使用一种称为域名的层次命名方式。整个 Internet 划分成多个顶级域，每个顶级域都有规定了的国际通用域名。顶级域名的划分有两种方式：一是根据所从事的行业领域；二是以国家或地区代号作为顶级域名。

表 1-2 通用顶级域名

机构	商业机构	网络服务机构	组织机构	政府机构	教育机构
域名	.com	.net	.org	.gov	.edu

表 1-3 南亚国家顶级域名

国家	阿富汗	孟加拉国	不丹	印度	斯里兰卡	马尔代夫	尼泊尔	巴基斯坦
域名	.af	.bd	.bt	.in	.lk	.mv	.np	.pk

域名的一般结构如下：“计算机主机名.机构名.网络名.最高层域名”，每个层次为域，各个层次之间用“.”号分隔，从左向右域名层级逐级上升。如，新浪网域名为：www.sina.com.cn，其中 www 表示这台主机的名称，sina 表示新浪网，com 表示商业，cn 表示中国。在域名中不区分大小写字母，域名在整个 Internet 中是唯一的，当高级域名相同时，低级子域名不允许重复。

Internet 实际使用的是 IP 地址，一个 IP 地址可以对应多个域名或无域名，但一个域名只能对应一个 IP 地址。例如：中央电视台的 Web 服务器 IP 地址是 202.108.249.206，但它有 2 个域名，分别是 www.cctv.com 和 www.cctv.com.cn。

3. DNS

DNS（Domain Name System）域名系统，是一种组织成域层次结构的计算机和网络服务命名系统。它用于 TCP/IP 网络，最主要用于 Internet 中通过用户的友好名称来定位计算机和服务。当用户在应用程序中输入 DNS 名称时，DNS 服务可以将此名称解析为相关的其他信息。

比如，用户在访问北京大学网站时，在浏览器地址栏中敲入一个很容易记忆的友好名称（域名）：www.pku.edu.cn 即可。但是计算机在网络上通信时，只使用数字的 IP 地址：162.105.129.12。显然，IP 地址要比域名难记得多，人们更喜欢使用容易记忆的域名来访问网站。DNS 可以使 IP 地址与域名之间建立起映射关系。

（三）统一资源定位器

为了确定具体信息资源在网络中的位置，Internet 使用统一资源定位器（Universal Resource Locator，URL）来描述具体资源和位置。URL 的组成结构为：应用协议、主机地址以及文件名。其中，主机地址可以是域名，也可以是 IP 地址。URL 的一般格式如下：

协议+ “://” +主机域名（或 IP 地址）+路径及文件名

如访问百度首页可以输入“http://www.baidu.com”，也可以输入“http://202.108.22.5”。

1. WWW

WWW 即 World Wide Web，又称为万维网、3W 等。WWW 起源于 1989 年 3 月，是由欧洲量子物理实验室所发展出来的主从结构分布式超媒体系统，其特点是可以方便迅速地浏览和传递分布于网络各处的文字、图像、声音和多媒体信息。访问 WWW 信息资源服务器时激活的第一个页面称为主页，每个主页都是超文本的。

2. 超文本

在互联网中每一文件在描述一定主题的同时还涉及许多其他概念，这些“其他概念”又被其他文件作为它们的主题内容予以描述。超文本（Hypertext）就是在这个文件所包含的这些概念与描述这些概念的其他文件之间建立一定的链接关系，并通过一定机制标记这些链接关系，帮助用户在浏览文件时使用简单方法直接调用描述这个概念的另一个文件。

在传统的纯文本中，我们只能看到文字信息本身，不能进行字号的变换、字体的着色，也不能添加图形、声音、动画，更不能链接到其他相关的文本。纯文本的这种信息组织方式，很难适应人们在互联网中交换信息的要求，于是人们研究出了超文本标记语言 HTML（Hyper Text Markup Language）。超文本（Hypertext）和超文本标记语言（HTML）是组织 WWW 信息资源的重要方式和手段。HTTP（Hyper Text Transfer Protocol），即超文本传输协议，是 WWW 客户机与服务器之间的应用层传输协议。

（四）网站

网站（Website）是指在互联网上，根据一定的规则，使用 HTML 等工具制定并展示特定内容的相关网页的集合，这些网页通过超链接联系起来。简单地说，网站就是互联网上一块固定的面向互联网用户发布信息的平台，包括网站地址和网站空间两个部分。

根据网页的性质，可以分为静态网站和动态网站。静态网站是没有采用任何脚本程序开发的网站，其网页完全使用 HTML 语言编写，无法直接在网络上更新内容。动态网站则有一个网站管理后台，以管理员的身份登录后，可以对整个网站的内容进行

修改，并在互联网上即时更新。

1. 网页

网页 (Webpage) 是计算机连接网络时，浏览器窗口中显示的一个页面，是网站的基本组成单位。网页是一个实实在在的文件，存放在服务器中。当用户输入访问命令，网页文件通过互联网进行传输，并显示到用户屏幕上。

2. 主页

网站的首页被称为主页 (Homepage)，一个好的主页是网站成功与否的关键。一般来说，主页包含站点名称、网站标志和主要服务项目等信息。不同类型的网站，其主页风格也不相同。例如导航类网站的首页，按照类别将各种网站的信息整齐排列；门户网站则以丰富的信息量和栏目导航为主；搜索引擎网站的主页则通常简洁明了。

(五) 浏览器

浏览器是浏览 Internet 信息的客户机软件，又称万维网导航工具，是用于检索、查询、采集、获取和管理网络信息资源的一种交互式应用软件。浏览器不只是浏览 Web 页面，它还可以用来收发电子邮件、阅读新闻、下载文件以及播放音乐、动画、视频等。目前比较流行的浏览器有微软公司的 Internet Explorer (IE)、谷歌公司的 Chrome、Mozilla 公司的 Firefox、苹果公司的 Safari 等。美国网站通信流量监测机构 StatCounter 的数据显示，2012 年 7 月，谷歌 Chrome 全球市场份额达到了 33.8%；微软 IE 紧随其后，为 32.04%；Firefox 为 23.7%；苹果 Safari 排名第四，为 7.12%。

下面以 IE 浏览器为例，简单介绍几种浏览器使用技巧，以帮助用户更好地利用网络信息资源。

1. 保存网页

在 Internet Explorer 中，可以通过“文件”下拉菜单的“另存为”一项将当前页面的内容保存到硬盘上，既能以 .HTML 文档 (.HTM/.HTML) 或文本文件 (.TXT) 的格式存盘，又能实现完整网页的保存。在“文件名”框中键入网页的文件名，在“保存类型”下拉列表中选择“Web 网页，全部 (*.htm;*.html)”选项，可将当前 Web 页面中的图像、框架和样式表全部保存，并将所有当前页面显示的图像文件一同下载并保存到一个“文件名.file”目录下。Internet Explorer 还将自动修改 Web 页中的链接，方

便用户进行离线浏览。最后，单击“保存”按钮即可。

2.收藏夹

将网页添加到收藏夹列表。用户需要再次打开该网页时，单击工具栏上的“收藏夹”按钮，然后单击列表中的快捷方式。

将网页添加到收藏夹栏。将经常访问的网站放到收藏夹栏中，单击一下就能访问这些网站。

3.主页

如果用户有一个需要首先访问的入口网页，则可将其设置为主页。如果用户有多个需要经常访问的入口网页，可将这些网页设置为一组主页选项卡，使其在每次启动 Internet Explorer 时显示。

4.历史记录

如果需要访问最近浏览过的某个网页，但是没有保存其链接，可以单击工具栏中的“收藏夹”按钮，然后单击“历史记录”。

(六) 搜索引擎

随着信息时代的到来，Internet 上的信息资源呈爆炸式的增长，并且分散存储在全世界各个地方的服务器中。用户无法获知所有服务器的地址，那么如何迅速方便地找到自己所需要的信息呢？搜索引擎正是为解决这个问题而出现的。

搜索引擎本身也是 Internet 上的一个 Web 站点，它的主要功能是在 Internet 上主动搜索 Web 服务器信息并将其自动索引，其索引内容存储于可供查询的大型数据库中。当用户输入与所需信息相关的关键词时，搜索引擎会为用户提供包含该关键词的所有网页，并提供访问该网页的链接，用户通过这些链接可方便快捷找到自己所要的信息。

目前网络上的搜索引擎一般可分为两类：

1.通用搜索引擎

通用搜索引擎在国外代表有谷歌（Google），国内则有著名的百度，用户可用其查找广泛综合性的信息。这些搜索引擎从互联网提取各个网站的信息（以网页文字为

主),建立起数据库,并能检索与用户查询条件相匹配的记录,按一定的排列顺序返回结果。

谷歌主要收录大级别网站的页面,按照一定程序对输入关键词进行排名,搜索响应时间较快,支持中英文检索,可进行关键词检索、词组检索和高级检索,不区分英文字母大小写,所有字母均当作小写处理。谷歌界面简洁、检索精确度高、质量高,是目前国际使用最广泛的搜索引擎之一。

百度是当前国内最大、使用率最高的中文通用搜索引擎,其功能完备,搜索精度高,在支持中文搜索方面甚至超过了谷歌。

2. 垂直搜索引擎

垂直搜索引擎是专为查询某一学科或主题的信息而产生的查询工具,是对网页库中的某类专门的信息进行一次整合,定向分字段抽取需要的数据进行处理后再以某种形式返回给用户。垂直搜索引擎的特点是“专、精、深”。垂直搜索引擎往往是针对某一个行业的,多为专业化搜索引擎,相比较通用搜索引擎的海量信息无序化,其为用户提供的不是成百上千万个相关网页,而是范围极为缩小、极具针对性的具体信息,如物流搜索、CNKI 文献搜索等。

第二节 南亚国家互联网发展的历史与现状

一、印度互联网发展概况

印度互联网起步早,发展慢,进入 21 世纪,随着移动互联网的出现和普及,印度互联网发展开始提速。

(一) 印度互联网发展历史

1. 印度互联网的起步

(1) 教育与科研计算机网(ERNET)

印度互联网的前身可以追溯到 1986 年投入运营的教育与科研计算机网(Education and Research Network, ERNET)。教育与科研计算机网由印度电子部(Department of Electronics)发起,隶属于印度交通和信息技术部旗下,由印度政府和联合国开发计划