

● 刘 辉

我国互联网络信息资源数量分布统计研究^{*}

摘 要 我国互联网信息资源明显集中于北京、上海、江苏等几个沿海省市又高于其他省市,表现出网络环境下信息资源集中模式和分层次模式的双重特征。表2。图1。参考文献6。

关键词 互联网 信息资源 网络资源分布 信息资源开发 信息资源管理 统计分析
分类号 G253

ABSTRACT In China, Internet information resources are mainly located in Beijing, Shanghai, Jiangsu Province and some coastal cities. The author discusses the dual characteristics of concentration and distribution of information resources in the networked environment. 2 tabs. 1 fig. 6 refs.

KEY WORDS Internet. Information resource. Network information distribution. Information resource development. Information resource management. Statistical analysis.

CLASS NUMBER G253

1 前言

互联网信息资源是社会信息资源的重要组成部分。对我国互联网络信息资源数量的分地区分布情况有一个全面、深入了解和掌握,是建设与开发网络信息资源的重要战略基础条件之一。近年来一些机构对我国网络信息资源的宏观分布态势,进行了一些全国性调查^[1]。例如,中国互联网络信息中心(CNNIC)等单位在2001、2002、2003和2004年分别进行了4次数量调查,对互联网络信息资源的数量及其分布状况做了探索性的摸底工作,成为我们研究网络资源分布情况的重要支撑条件。本文主要选用2004年第四次调查的最新资料,对其地区分布特征进行深入分析,旨在探索我国互联网络信息资源

存在的问题和发展趋势,为制定全国和地区性互联网络信息资源发展政策和措施提供参考,促进我国互联网络信息资源的科学开发与利用。

2 数据与分析

CNNIC的该次调查对象包括全国(不包括香港、澳门、台湾地区)所有已注册域名的网站,包括.com,.net,.org和所有.cn域名下的所有网站。调查确定的网络信息资源定义为:国内互联网络上公开发布的网页和在线数据库的总和,其数据截止日期为2004年底,其中有4项关于各省(自治区、市)互联网络信息资源数量的基本数据(见表1)^[2]。

表1 2004年全国互联网络信息资源数量分省(自治区、市)基本数据

编号	地区	域名地域分布比例 (%)	分地区网站比例 (%)	分省网页数 (万)	分省网页字节数 (万 KB)
1	北京	12.9	18.7	30700	795300
2	天津	1.4	1.0	511	12060
3	河北	1.7	2.5	1210	24670
4	山西	0.6	0.7	481	9972
5	内蒙古	0.4	0.4	26.3	576.4
6	辽宁	3.1	3.3	1820	40950
7	吉林	0.8	0.7	353	7156

^{*} 本文系国家社会科学基金资助项目(项目编号03AJY003)。

续表

编号	地区	域名地域分布比例 (%)	分地区网站比例 (%)	分省网页数 (万)	分省网页字节数 (万 KB)
8	黑龙江	1.2	0.9	1240	24250
9	上海	10.5	8.7	12000	269800
10	江苏	8.3	7.5	3390	73710
11	浙江	8.5	11.5	5900	126700
12	安徽	2.0	1.7	2160	61630
13	福建	11.2	5.7	2750	59740
14	江西	0.7	1.1	659	12060
15	山东	4.6	4.0	2700	58040
16	河南	2.1	2.0	2970	66010
17	湖北	1.9	2.3	1860	45860
18	湖南	1.5	1.2	1500	36200
19	广东	16.3	18.2	5160	130200
20	广西	0.8	1.2	696	15560
21	海南	0.4	0.4	697	10970
22	重庆	1.0	1.2	1300	31950
23	四川	4.5	1.9	3700	82970
24	贵州	0.3	0.4	263	4852
25	云南	0.8	0.7	581	11730
26	西藏	0.1	0.3	3.84	38.8718
27	陕西	1.3	0.8	534	10460
28	甘肃	0.3	0.4	1280	24660
29	青海	0.1	0.1	25.5	513.2
30	宁夏	0.2	0.2	110	1974
31	新疆	0.5	0.3	177	3159
	总计	100	100	86700	2053721.4718

资料来源:中国互联网络信息中心.2004年中国互联网络信息资源数量调查报告,2005年2月

本文采用 SPSS 进行分析,对不同数量级的变量作标准化处理,具体利用层次聚类分析方法(Hierarchical Cluster Analysis)中的 Q 类聚类,即根据观察值或变量之间的亲疏程度,使具有共同特点的样本聚集在一起,以便对不同类的样本进行分析(表2)^[3]。从表2观察到,北京在聚类时与其他地区差异最大,在划分成2、3或4类时都单独自成一类,经济相对发达

的几个省市相互之间较相近,而落后省区相互之间也较相近,如划分为三类时可较明显体现它们的相同与差异之处。而从图1也可看出第2类与第3类间的距离还不太大,而这两类与第3类间的距离则是非常大的。说明北京是全国网络信息资源数量最多的地区,即使与上海、江苏、浙江、福建、广东等发达地区比,也有较大的数量领先优势。

表 2 层次聚类分析中的类成员 (Cluster Membership)

编号	地区	4 类	3 类	2 类	编号	地区	4 类	3 类	2 类
1	北京	1	1	1	17	湖北	2	2	2
2	天津	2	2	2	18	湖南	2	2	2
3	河北	2	2	2	19	广东	4	3	2
4	山西	2	2	2	20	广西	2	2	2
5	内蒙古	2	2	2	21	海南	2	2	2
6	辽宁	2	2	2	22	重庆	2	2	2
7	吉林	2	2	2	23	四川	2	2	2
8	黑龙江	2	2	2	24	贵州	2	2	2
9	上海	3	3	2	25	云南	2	2	2
10	江苏	3	3	2	26	西藏	2	2	2
11	浙江	3	3	2	27	陕西	2	2	2
12	安徽	2	2	2	28	甘肃	2	2	2
13	福建	3	3	2	29	青海	2	2	2
14	江西	2	2	2	30	宁夏	2	2	2
15	山东	2	2	2	31	新疆	2	2	2
16	河南	2	2	2					

Dendrogram using Average Linkage (Between Groups)

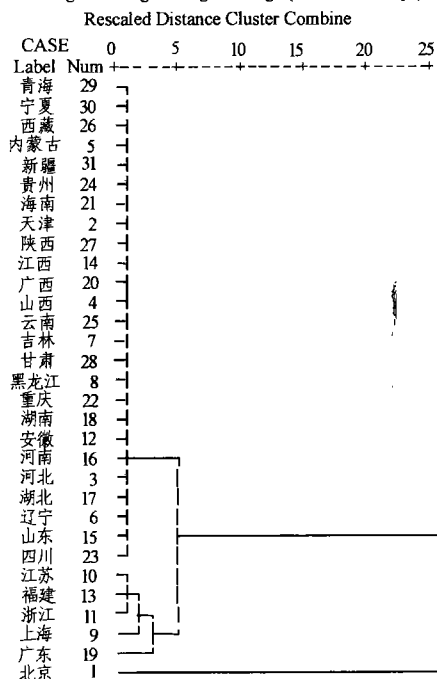


图 1 层次聚类分析的龙骨图

3 结论及讨论

根据表 2 和图 1, 可以将各省市自治区市按互联网络信息资源数量进行大致分类, 共分为三大类。第一

大类, 数量水平明显领先类: 北京。第二大类, 数量发展水平领先类: 上海、江苏、浙江、福建、广东。第三大类, 数量发展水平相对滞后类: 天津、河北、山西、内蒙古、辽宁、吉林、黑龙江、安徽、江西、山东、河南、湖北、湖南、广西、海南、重庆、四川、贵州、云南、西藏、陕西、甘肃、青海、宁夏、新疆。

在对表 1 中“域名地域分布”、“分地区网站比例”、“分省网页数”和“分省网页字节数”四项指标的峰值计算中, 峰值都大于 0, 分别为 2. 103、4. 896、20. 258 和 22. 234, 尤其在网页数和字节数上凸显了向少数地区集中的趋势; 而在这 4 个指标中, 前两个更多地反映了网络信息资源的种类分布, 后两个更多地反映了网络信息资源的内容深度。这说明在网络信息资源集中的地方, 不仅信息种类较齐全, 而且信息内容更详细、更有深度; 在网络信息资源较为稀缺的欠发达地区, 不仅信息种类较少, 而且各信息源在内容含量上也较薄弱, 在系统性和深度上也较欠缺。北京作为首都, 云集了大量的国家机关、全国性媒体机构、知名网络企业、国内知名大学、中科院等国家级专业院所、高新企业等, 为网络信息的生产、收集、加工、整理、存贮和传递提供了相对雄厚的基础, 有利于保证信息的内容权威、种类齐全和更新迅速。据表 1 中数据, 在网页数上, 北京大约占全国总数的 35%, 网页字节数约占 39%, 在网络信息资源的开发上已

经成为全国的龙头,是全国的网络信息资源中心。由于信息基础设施的发达,网络系统技术的进步,时间、空间以及其他一切信息传输的物理障碍越来越容易被打破^[4]。在现代高速信息网络环境,集中在北京的网络信息资源已经成为全国各地共享的战略资源。

上海等5个发达省市由于在经济、科技等方面的优势,在网络信息资源开发方面也走在全国的前列,甚至在网络域名分布比例上与北京不分伯仲。如果说北京遥遥领先是因为国家级信息资源密集,并负有向全国提供网络信息服务的职责,这五省市排在前面则基本上是依靠自身的优势,体现出与本地经济社会发展更紧密结合的特征。

各国实践表明,网络信息资源的开发利用也依赖于相关技术和设备,这势必会导致信息资源向某一局部区域过度富集,容易产生网络信息贫富的两极分化现象^[5]。在我国由于地区间发展水平的不平衡,反映到网络信息资源开发上也呈极度的不均衡,欠发达省市的网页数分布之低让人担忧,其网络信息资源开发程度太低会在很大程度上影响当地社会进步的整体进程。

我国各地区网络信息资源数量的分布比较悬殊,远较之我们通常研究到的各省科技资源、文献信息资源等分布更倾斜,表现了网络状态下的特征。欲缩小信息贫富差距,就要实现信息资源共享。学界认为信息资源共享模式主要有3种:集中型模式、网络型模式和层次型模式^[6]。信息借助互联网而跨越地理局限,同一信息资源可被不同地区的千万人利用,体现了共建共享的思想,是网络状态下集中型模式的体现;各省对本地网络信息资源的开发则是自身融入开

放发展战略的必要组成部分,这是网络状态下的层次型模式的代表。

在探讨我国网络信息资源建设与利用战略时,在进行集中规模化开发和分散式开发相结合方面,在运用行政力量和市场力量双重推动方面,在保障信息公开平等共享与有偿合理使用方面,在努力开发地方特色信息资源与充分利用外部信息资源方面,都值得进一步研究和探索。

参考文献

- 1 第15次中国互联网络发展状况统计报告(2005/1). http://news.xinhuanet.com/it/2005-01/19/content_2480485.htm [2005-06-15 查阅]
- 2 中国互联网络信息中心. 2004年中国互联网络信息资源数量调查报告. 2005年2月. http://tech.sina.com.cn/focus/04_net_rep/index.shtml [2005-06-15 查阅]
- 3 余建英,何旭宏. 数据统计分析与SPSS应用. 北京:人民邮电出版社,2003
- 4 王慧博. 试析网络环境下信息资源的有效配置. 情报杂志,2004(1)
- 5 赵金保,王晓红. 论信息网络资源配置. 邯郸医学高等专科学校学报,2004(3)
- 6 马费成,裴雷. 我国信息资源共享实践及理论研究进展. 情报学报,2005(3)

刘 辉 浙江财经学院工商管理学院副教授,武汉大学信息管理学院博士生. 通信地址:杭州市文华路浙江财经学院工商管理学院. 邮编310012。

(来稿时间:2005-10-26)

(上接第31页)

- 3 Vasilis Theoharakis, Veronica Wong. Marking high-technology market evolution through the foci of market stories: the case of local area networks. The Journal of Product Innovation Management, 2002(19)
- 4 Michael J. Gallivan; Gordon Depledge. Trust, control and the role of inter organizational systems in electronic partnerships. Information Systems Journal, 2003, 13 (2)
- 5 Charles D. Bodkin, Monica Perry. Goods retailers and service providers: comparative analysis of website marketing communications. Journal of Retailing and Consumer Services, 2004(11)

- 6 Altinkemer, A Chaturvedi, S Kondareddy. Business Process Reengineering and Organizational Performance: An Exploration of Issues. International Journal of Information Management, 1998, 18(6)
- 7 包昌火. 情报研究方法论. 北京:科技文献出版社,1990
- 8 Bi ju Paul Abraham, Soumyo D. Moitra. Innovation assessment through patent analysis. Technovation, 2001(21)

黄晓斌 中山大学资讯管理系副教授、情报学博士。通信地址:广州市. 邮编510275。

成 波 中山大学资讯管理系情报学硕士。通信地址同上。
(来稿时间:2005-10-18)