

基于链接分析的中国体育协会网站 信息资源特征的比较

司虎克¹, 蔡犁¹, 王 兴¹, 李建方², 肖 欣¹

摘 要: 依据网络信息计量学理论和运用网络链接分析方法, 通过相关搜索引擎对我国体育协会网站及其页面的信息资源链接进行了收集与解析。主要分析了网站的文件数、页面数、链接数、链接密度, 扩散系数以及网络影响因子等指标的特性及它们的相互关系, 揭示了它们的分布特征, 提出了网络信息资源与其他形式信息资源的关系。

关键词: 链接分析; 网站; 信息资源; 体育协会

中图分类号: G80-05 文献标识码: A 文章编号: 1006-1207(2010)05-0031-05

Comparison between the Information Resources Characteristics of the Chinese Sports Associations' Websites Based on Link Analysis

SI Hu-ke, CAI Li, WANG Xing, et al.

(Shanghai Institute of P.E., Shanghai 200438 China)

Abstract: Based on the methods of network information metrological theory and network link analysis, the article analyses the statistics collected from the Chinese sports associations' websites and the information resources links of the web pages through the relative search engines. It makes analysis of the quantity of the websites' documents, pages and links, link density, spread coefficient and the characteristics of the indices such as web impact factors and their correlations. It reveals the distribution characteristics of those indices and sets forth the relations between the web information resources and the information resources of other types.

Key words: link analysis; website; information resources; sports association

20世纪90年代中期以来, 计算机互联网技术得到了飞速的发展, 网络信息已逐步进入社会的每个领域, 对推动社会、经济、文化、体育等方面的发展起着越来越重要的作用。我国体育事业的发展与科技进步、创新是分不开的, 其中一个很重要的基础环节就是信息构建与交流, 而体育网络信息资源构建与发展则是“加强体育信息工作, 组织信息化建设重点工程, 积极推进运动项目管理的信息化建设”^[1]的重要任务之一。体育网络信息资源是指通过计算机网络储存和传播的各种体育信息资源的总和, 包括体育行政管理、体育新闻传播、体育赛事转播与评论、体育史料记录查询、体育学相关学科专业文献数据库、体育项目技战术视频数据库与应用软件以及体育科学数据等等。其具有广泛性、开放性、及时性、数量多、增长快等特点, 因此也越来越被各类体育从业人员所接受, 并成为获取体育信息的主要来源之一。

十多年来, 我国各级政府体育部门、各类体育院校与科研院所、国家体育总局及其所属运动项目管理中心与相关各个体育协会等机构, 在体育网络信息资源建设方面取得了飞速发展, 建立了各类体育网站, 发布了各类体育相关信息。但同时我们应看到由于网络信息资源具有结构复杂、来源广泛的特点, 许多问题或矛盾必然会应运而生, 如网

络信息特点与用户不同需求之间的矛盾, 各类体育网站建设规模、风格与用户所需要的查全率、准确性之间的矛盾, 以及网站建设基础性问题等等, 一直在困扰着体育网络信息资源建设与发展。因此加强对相关体育网站的研究, 力求探索网站发展的内部规律及其影响因素, 从网站层面探索体育网络信息资源建设状况与特征, 这正是本课题研究的主要任务与目的。

1 研究现状

近年来我国一些学者分别从现代信息技术在体育中的应用与发展、网络环境下的信息资源管理、体育信息管理的理论构建及应用系统开发研究、体育信息网络的建设与发展前景、全国高等体育教育系统文献信息资源共享模式研究、奥运信息门户的设计与实现、体育信息整合暨区域间信息共享的研究、我国体育管理信息化的研究、北京奥运会对我国体育信息化建设的影响等不同角度和层面, 对我国体育信息化建设过程中的热点和难点问题进行了有益的探索与研究^[2-12], 并认为我国的网络环境下体育信息资源的建设还很落后, 一些发达国家在体育领域的信息应用系统建设和体育信息资源的开发利用方面已经取得了领先地位, 形成了比较健全的信息采集、处理、保存和使用体系, 拥有一批综合性数据库, 能够为国民提供

收稿日期: 2010-07-29

基金项目: 2007年上海市教委科研创新重点项目(07ZS122); 上海市重点学科(体育人文社会学)建设资助项目(Y0901)

第一作者简介: 司虎克, 男, 教授, 博士生导师。主要研究方向: 网络环境下的体育信息建设的理论与实践。

作者单位: 1. 上海体育学院, 上海 200438; 2. 石家庄东方美术学院 影视动画系, 河北 石家庄 050700



体育资源和相关服务；同时还分析了我国的体育信息网络化建设与国外的主要差距，表现在：

(1) 由于我国人口众多，网络普及并不高，获得信息也就不平衡，反映出国家整体信息化基础薄弱。

(2) 政府体育部门信息化程度不高，网站水平较低。

(3) 我国体育信息资源开发利用不够，数据库建设跟不上网络的发展速度。

(4) 网络建设业务系统水平低，应用和服务领域狭窄，信息资源相互关联度较低。

(5) 体育信息资源开发利用滞后，标准不统一，信息安全保障存在隐患。

(6) 体育信息化人才缺失，资金投入不足。

通过对相关文献的分析与归纳发现，我国的体育网络信息资源建设中的一个重要环节值得我们关注，即各类体育网站建设状况不如人意，发展不平衡，网站互相链接不合理，还缺乏对体育网站的系统研究及其相应评价，反映出我国各类体育网站仍处于发展的初步阶段，其辐射影响力还很有限。

2 研究对象与方法

2.1 对象

我国各级政府体育部门、各类体育院校与科研院所、国家体育总局及其所属运动项目管理中心与相关各个体育协会等机构都建立了各自的体育网站。根据研究目的，在众多体育网站中笔者以中国体育协会所属各个项目协会网站为研究总体，但是如果把所有项目协会网站数据都下载下来，则是一件非常繁重而复杂的工作。为此以简单随机抽样方法从中抽取了19个项目协会网站样本作为研究对象，超出了在最大容许相对误差小于0.5，置信度为95%，所估计比例取值50%条件下所对应的最小样本含量，符合统计学样本含量确定的要求^[13-14]，见表1。

表1 体育协会网站网址
Table 1 Websites of the Sports Associations

项目协会网站名称	网址
乒乓球协会网	http://tabletennis.sport.org.cn/
游泳协会网	http://swimming.sport.org.cn/
网球协会网	http://tennis.sport.org.cn/
棒球协会网	http://baseball.sport.org.cn/
冰球协会网	http://icehockey.sport.org.cn/
帆船帆板协会网	http://sailboarding.sport.org.cn/
滑冰协会网	http://skating.sport.org.cn/
击剑协会网	http://fencing.sport.org.cn/
健美操协会网	http://www.caa.net.cn/
马术协会网	http://equestrian.sport.org.cn/
体操协会网	http://www.cga.net.cn/
武术协会网	http://wushu.sport.org.cn/
射击协会网	http://shooting.sport.org.cn/
高尔夫协会网	http://www.golf.org.cn/
曲棍球协会网	http://hockey.sport.org.cn/
手球协会网	http://handball.sport.org.cn/
台球协会网	http://billiards.sport.org.cn/
田径协会网	http://www.athletics.org.cn/
足球协会网	http://www.fa.org.cn/

2.2 链接分析方法

应用链接分析方法，选择网络链接特征作为衡量中国体育协会网站资源利用情况的评价指标，利用链接分析中的搜索引擎作为获取研究所需数据的来源工具，在此基础上采用统计分析方法对中国体育协会网站链接指标所反映的特征进行分析比较，以进一步探索各相关体育协会网站在链接特征上的规律性及与其它指标之间的联系。

下载和获取网站原始数据的工具有很多种，比较流行的是WebZip、Offline Explorer (OE)、Teleport Pro、WebCopier Pro等。我们选用OE作为获取各相关体育协会网站原始数据的下载工具，是考虑到OE获取的数据更全面，速度较快，在动态网页上的数据获取能力强的特点。

2.3 获取与解析网站数据的步骤

2.3.1 建立下载的层数和下载方式。在下拉菜单“新建”中选择“项目”，在项目名称中为项目命名，在“地址”中输入所要下载的网站网址，除去下载级别的限制，在文件修改检查中选择“不下载已有文件”。

2.3.2 设置过滤器 (URL)。在对特定服务器中的数据进行下载的时候，在“协议”中选择“装入全部协议”，在“服务器”中选择“装入文件若该文件位于初始服务器”，在“目录”中选择“装入全部目录中的文件”，在“文件名”中选择“装入全部文件名”。

2.3.3 设置链接转换。OE可以将下载文件中的链接转换成可以离线浏览的相对地址，但转换后将无法确定原始网站中的链接信息，因此在下载中把该项设置为“不转换”。以上设置结束后，其它设置多采用默认设置。然后点击工具栏中的“下载”按钮就可以把指定网站的数据下载到指定的文件夹中。

2.3.4 本研究获取网络数据的时间设置范围为2007年4月至9月，在这6个月中连续每月从相关体育项目协会网站获取及下载相关网络数据，并对相关数据作统计学处理与分析。

2.3.5 本文的数据解析工具为webStat，webStat需要在Java环境中运行。程序运行以后，把分析目标指向下载的指定文件夹，然后点击“解析链接”，程序就会自行对指定的文件进行解析，解析结果会显示在下边的空框中。

3 结果与分析

3.1 相关体育项目协会网站建设的基本特征

3.1.1 网站总链接数是网站内部所含有的网络链接总量，它是衡量网站链接数量特征最重要的指标之一。一般来说，网站的总链接数越多，网络信息的组织体系就越完备，信息揭示程度也就越高。

从表2中我们可以看出，总链接数达到60 000以上的有5家，分别是乒乓球、游泳、网球、曲棍球和田径协会网站，占所选取样本的26.3%；总链接数在10 000~60 000之间的有12家，分别是棒球、冰球、帆船板、滑冰、击剑、马术、武术、射击、高尔夫、手球、台球和足球协会网站，占所选取样本的63.2%；总链接数在10 000以下的仅有健美操、体操2家网站，占所选取样本的10.5%。这提示我们相关体育项目协会网站建设与发展不平衡，在一定程度上反映了各个体育协会网站之间的发展规模与信息容量等方面存在着较大差距。



表2 相关体育项目协会网站6个月数据(均值)一览表

Table II Six-month Statistics (Average Value) of the Relative Sports Association Websites

协会	总链接数	自链接数	出链接数	被链接数	页面平均链接数	链接密度
乒乓球	65904.3	30816	35088.3	31.5	23	725.4
游泳	83104.3	40923	42181.3	37	22.7	819.4
网球	62771.2	19025.3	43745.8	63.5	16.6	872.1
棒球	38662.3	19480	19182.3	29.8	23.9	436.6
冰球	25495.8	12824.2	12671.7	28.7	23.5	489.9
帆船板	24878.8	12978.8	11900	21.3	25.7	514.4
滑冰	53498	22190.2	31307.8	50.3	26.2	683.4
击剑	24484.8	10669.7	13815.2	28.5	20.5	763.6
健美操	812.12	623	189.2	11.5	15.5	38786.8
马术	10961.8	4802.7	6159.2	23.3	21.4	765.9
体操	3975.8	3650.3	325.5	15.5	8.6	62467.3
武术	25296	11575.7	13720.3	13.8	21	631.5
射击	27322.2	8370	18952.2	24	16.7	658.3
高尔夫	27991.5	16412.5	11579	29.3	28.2	272.5
曲棍球	69615.5	33687	35928.5	32.7	24.2	499.6
手球	47010.2	29257.3	17752.8	28	27.6	733.5
台球	33866	16766.8	17099.2	34.7	21.2	695.4
田径	64410.2	20487.3	43922.8	44.3	36.2	265.7
足球	32153.7	19806.7	12356	24.5	28.6	390.5
总均值	38011.3	17597.2	20414.6	30.1	22.7	5866.9

3.1.2 网站自链数

网站自链数是指网站中的自我链接的总数量。自链数的数量以及其在总链接数中的比例在很大程度上可以反映网站的性质。自链数大并且在总链接数中所占比例高的网站,往往其内部的信息揭示较为充分,自身具有较为完备的体系。

网站自链数在30 000~40 000以上的分别是乒乓球、游泳和曲棍球网站,占样本数的15.8%;自链数在10 000~30 000之间的则占大多数,为网球、棒球、冰球、帆船板、滑冰、击剑、武术、高尔夫、手球、台球、田径和足球协会网站,占所选取样本的63.2%;而在10 000以下的,则分别是健美操、马术、体操和射击协会网站,占21.0%。

3.1.3 网站出链数

出链数是指网站中outlink(该网站链接到其他网站的数量)的总量。出链往往是指向网络用户所需要获取的最终资源,或是带有参考推荐的性质。通过该指标以及其在总链接数中的比例对判断网络信息资源的性质具有非常重要的参考价值。网站出链数在30 000~40 000以上的分别是乒乓球、游泳、网球、滑冰、曲棍球和田径协会网站,占样本数的31.6%;自链数在10 000~30 000之间的则分别是棒球、冰球、帆船板、击剑、武术、射击、高尔夫、手球、台球和足球协会网站,占52.6%;而在10 000以下的,则分别是健美操、马术和体操协会网站,占15.8%。

3.1.4 被链接网站数

被链接网站数是指网站的出链所指向的链宿的数量。它反映了特定网站的辐射力和该网站与其他相关网站的密切程度。被链接网站数越多,网络的开放性就越强,用户通过该网站所能获得的相关信息和信息线索就会越丰富。被链接网站达到40以上的有网球、滑冰和田径3个网站,占样本数

的15.8%;在20~40之间的分别是乒乓球、游泳、棒球、冰球、帆船板、曲棍球、击剑、马术、射击、高尔夫、手球、台球和足球协会网站,占68.4%;在20以下的是健美操、体操和武术,占15.8%。

3.1.5 页面平均链接数

页面平均链接数是度量链接出现频率的指标,其与链接密度不同,它揭示的是平均每个网页中链接的数量,用于网站信息的组织程度和揭示的深度。因此,该指标与网站字节数无关,但会受网站设计风格的影响。页面平均链接数达到20以下的体育项目协会网站分别是网球、健美操、体操与射击,占21.1%;达到20以上的则有乒乓球、游泳、棒球、冰球、帆船板、滑冰、曲棍球、击剑、马术、武术、高尔夫、手球、台球、田径、足球协会网站,占78.9%。

3.1.6 链接密度

链接密度是度量网站内链接频率的指标,与总链接数不同,链接密度消除了网站规模和页面文件大小所造成的影响。根据不同类型的链接,可以得出不同类型链接的分布密度。网站链接密度大的说明网站的信息含量高,对信息的揭示深度大,网站的体系结构严密。因此这一指标对反映网站的信息含量和揭示深度具有重要价值。各体育协会网站链接密度均值大于总均值的仅有2家网站,分别是健美操和体操,占总数的10.5%,其余网站均低于总均值,反映了各体育协会网站所包含信息量及其信息深度存在较大差异,仅有少数网站在网站基本建设中较为注重相关信息量的增加以及揭示信息量的深度,以满足用户的需求。

3.2 体育协会网站指标间的相关分析

体育协会网站各指标之间还存在着某种内在的联系,这



些联系也反映了体育协会网站运作和建设中的内在规律与特点，有助于为我们进行网站评价提供有价值的参考。

从表3可以看出，总链接数、自链接数、出链接数、被链接站数与页面数这5个指标之间的相关系数均为0.70以上（达到了0.01的极显著性水平）。我们认为这主要是因为总链接数是自链接数和出链接数的总和，自链接数和出链

接数的任何一个变量发生变化，总链接数都会发生相应的变化。由此看来，总链接数、自链数和出链数是一个紧密的整体，相互影响。同时我们还看到，被链接网站数与总链接数、出链接数以及自链接数都具有高度相关性，这主要是因为被链接网站数是出链数指向的网站的总数量，所以被链接网站的数量变化自然就带动了。

表3 体育协会网站链接指标间的相关程度一览表

Table III Correlations between the Link Indices of the Sports Association Websites

		总链接数	自链接数	出链接数	被链接数	页面数	总文件数
Spearman 相关系数	Correlation	1.000	0.942**	0.893**	0.835**	0.916**	0.609**
	总链接数 Sig. (2-tailed)	.	0.000	0.000	0.103	0.000	0.006
	N	19	19	19	19	19	19
	Correlation	0.942**	1.000	0.754**	0.721**	0.798**	0.500*
	自链接数 Sig. (2-tailed)	0.000	.	0.000	0.000	0.000	0.029
	N	19	19	19	19	19	19
	Correlation	0.893**	0.754**	1.000	0.823**	0.960**	0.616**
	出链接数 Sig. (2-tailed)	0.000	0.000	.	0.000	0.000	0.005
	N	19	19	19	19	19	19
	Correlation	0.835**	0.721**	0.823**	1.000	0.793**	0.525*
	被链接数 Sig. (2-tailed)	0.103	0.000	0.000	.	0.000	0.021
	N	19	19	19	19	19	19
	Correlation	0.916**	0.798**	0.960**	0.793**	1.000	0.693**
	页面数 Sig. (2-tailed)	0.000	0.000	0.000	0.000	.	0.001
	N	19	19	19	19	19	19
	Correlation	0.609**	0.500*	0.616**	0.525*	0.693**	1.000
	总文件数 Sig. (2-tailed)	0.006	0.029	0.005	0.021	0.001	.
	N	19	19	19	19	19	19

** P < 0.01 * P < 0.05

出链数的变化，使与出链数关系密切的总链数和自链数也会发生相应的反应。

另外，我们还认为出链数表现的是网站内部的链接量，反应一个网站内部的系统性和规模，是内部资源利用率较高的一种表现。而出链数是网站指向外部的链接数，是网站利用外部资源程度的表现，两者之间关系并不明显。由此可见，网站的规模增大对提高网站资源的利用率是非常有影响的，对内部资源的利用率提高了，就会对信息资源产生更高的需求，那么对外部资源的需求就相应提高了，所以指向外部的链接数量也就会增加。因此，网站的规模大小对本网站利用外部资源的程度会产生很大的影响，出链接数和被链接网站数这两个指标对衡量网站的规模和系统性也是有重大作用的。

通过分析我们还认为本研究样本网站所反映的页面数与总链接数、自链接数、出链接数具有高度相关性；页面数是反映网站规模的最直接的指标，而链接数是反映网站规模的较间接的指标，它们之间的联系应该是很密切的。据此笔者认为，总文件数也是反映网站资源数量的指标，两者具有高度相关性，但这两个指标并不反映网站的规模。可见网站内部资源的多少和网站规模的大小是两个不同的概

念，网站规模更多强调的是网站利用资源的程度，或者说网站内部结构的系统性。所以，各体育协会网站存在两种现象给我们的启发是：一种是网站有很多的资源，但利用率很低，资源没有充分实现或体现自己的价值；而另一种则是内部资源并不高，但由于相互链接引用的程度高，信息资源就可能实现价值的最大化。这对于指导体育协会网站的建设或客观评价都具有积极意义。

另外，从表4我们可以看出体育协会网站的入链接数、网络影响因子、扩散系数之间的相关程度特点。网站的入链接数是指该网站被其他网络信息资源链接的次数，因此是评价网站影响力和价值的重要尺度^[15]。而网络影响因子和扩散系数均可以作为评价网站建设的网络信息资源被外界利用的程度或强度。这3个指标不仅反映了多年来体育协会网站建设的成效，而且其相互之间的关联程度较大，从一个层面综合反映了我国体育协会各网站从总体上对外界的影响力是客观存在的。

通过对我国体育协会网站的链接分析与研究，在一定程度上揭示了体育协会网站建设与发展过程中体现了一定的规模效应，可以说网站所反映出的网页和内部链接体现了网站规模和结构。而各个体育协会网站内部链接数的影响与外部链接



表4 体育协会网站入链接数、网络影响力、扩散系数之相关程度
Table IV Correlations between the Link Quantity, Web Impact Factors and Spread Coefficient

		入链接数	网络影响力	扩散系数
入链接数	Correlation	1.000	0.774**	0.731**
	Sig. (2-tailed)	.	0.000	0.000
	N	19	19	19
网络影响力	Correlation	0.774**	1.000	0.979**
	Sig. (2-tailed)	0.000	.	0.000
	N	19	19	19
扩散系数	Correlation	0.731**	0.979**	1.000
	Sig. (2-tailed)	0.000	0.000	.
	N	19	19	19

** P<0.01

数均反映了各自的特点,这对体育协会网站在未来建设与发展无疑具有重要参考价值。当然我们也认为我国体育协会网站建设还应注重在保证规模与结构完善的基础上,进一步提高网站所提供的信息质量与数量,以吸引更多的用户点击进入各类体育协会网站,最大程度满足其需要的相关体育信息。

4 结论

4.1 各体育协会网站的总体发展趋势反映出网站内部的系统性和引用外部资源的程度都比较稳定,且逐步提高。但由于体育项目本身的影响力和其它诸多因素的影响,体育协会网站资源被外界引用的程度并不稳定,各网站与其它网站的关联存在较大的差异,不同体育协会网站其影响力差别较大。

4.2 体育协会网站的建设应以技术性发展为主要方向,在增加信息资源数量的同时,应进一步强化网站的内部结构体系,提高网站数据的系统性和关联性,从而提高各体育协会网站的总体影响力。

4.3 网站的建设应紧密联系用户的需求,网站与外界的关联程度与网站的主题内容和运作方式有关系,因此以用户需求为基本目标,提高网站的信息内容质量,增强信息的组织与发布,从而扩大网站影响力应成为各体育协会网站建设的重要任务。

4.4 建立一整套较完善的体育协会网站建设的评价系统,定期对网站建设和信息资源的质量进行评价并以此作为网站建设的重要依据,为网站的建设和信息资源更新提供正确的方向,能有效促进网站不断发展和完善。

参考文献:

- [1] 刘鹏.在全国体育科技工作会议、全国体育反兴奋剂工作会议开幕式上的讲话[J].体育科学,2005,25
- [2] 廖慧平,尧旭华等.我国体育信息网络化发展现状分析[J].山东体育学院学报,2005,21(2)
- [3] 马铁.现代信息技术在体育中的应用与发展[J].体育科技信息,1997(3):56.
- [4] 董小英.网络环境下的信息资源管理[D].北京大学博士学位论文,1997:12.
- [5] 董伦红.体育信息管理的理论构建及应用系统开发研究[D].北京体育大学博士学位论文,2002:4.
- [6] 张江南.体育信息网络的建设与发展前景[J].情报学报,2003,21(1):86-88.
- [7] 洪雁.奥运信息门户的设计与实现[D].武汉大学硕士学位论文,2005:4.
- [8] 胡彪,饶坚等人.体育信息整合暨区域间信息共享的研究[J].武汉体育学院学报,2006,(2):5-8.
- [9] 尹博.我国体育管理信息化的研究[J].北京体育大学学报,2006,29(5):611-613.
- [10] 文红斌,马飞.北京奥运会对我国体育信息化建设的影响[J].商场现代化,2006,469(6):213.
- [11] 周兰君,张天健.国内外体育信息网络化发展现状及对策研究[J].广州体育学院学报,2003,23(6)
- [12] 廖慧平,刘兴杰.我国体育信息网络化建设的指导思想、目标、思路与原则[J].武汉体育学院学报,2005:39(1)
- [13] 卢淑华.社会统计学[M].北京:北京大学出版社,2000:556-560.
- [14] 柯惠新,黄京华等.调查研究中的统计分析法[M].北京:北京广播学院出版社,1999:304-309.
- [15] 段宇峰.网络链接分析与网站评价研究[M].北京:北京图书馆出版社,2005:157-158.

(责任编辑:陈建萍)