

## 利用链接关系评价网络信息的可行性研究

刘雁书 方 平

(中南大学湘雅医学院信息管理系,长沙 410078)

**摘要** 目的:通过对有代表意义网站的链接特征及站外链接类型及特征的调查分析,评价利用链接关系评价网络信息的可行性。方法:利用搜索引擎 Fast Search 分别检索综合网站与专业网站的被链次数及新浪被不同类型网页链接次数,以分析网站的链接特征,并对新浪站外链接类型及其特征进行分析。结果:综合网站与专业网站被链次数存在明显差异,站外链接数在总链接数中所占比例超过站内链接数。新浪被不同语种及不同国别网页广泛链接。站外链接可分为推荐链接、合作链接、相关链接、资源链接、通讯链接、广告链接 6 种类型。前 3 种类型全部是新浪的主页与频道建立的链接,对评价网络信息价值更高。结论:站外链接关系反映的是被链网页被利用与被推荐的总体情况,与被链网页质量存在正向(肯定)联系,因此利用站外链接评价网络信息是可行的。

**关键词** 链接关系 网络信息评价 链接类型 万维网 引文分析法 搜索引擎 链接分析法

## Study on the Reliability of Link Popularity in Web Information Evaluation

Liu Yanshu and Fang Ping

(Department of Information Management, Xiang Ya College, Central South University, Changsha 410078)

**Abstract** **Objective**: By exploring the feature of link and extra link type of typical Web sites to evaluate the reliability of link analysis method. **Methods**: Using Fast search engine to retrieve link popularity of general and medical Web sites, and Sina's link popularity linked by different types of nodes to analyze the character of link. Sampled and analyzed Sina's extra link to reveal extra link types and its relation with link nodes. **Results**: There were obvious differences between link popularity of general and medical Web sites. And the percent of extra link in total was higher than that of intra link, and Sina was widely linked by Web nodes of different language, country and organization. Extra link types fall into 6 categories: recommend, cooperation, relation, resource, communication and ads, and links with Sina's homepage and channel are all of the first three types, which were much more valuable in evaluating Web information. **Conclusion**: Since extra link relationship reflected the utilized and recommended case of linked nodes, and it has positive relationship with the quality of linked nodes, it is reliable to use extra link to evaluate Web information.

**Keywords** link popularity, link types, Web information evaluation, search engine, World Wide Web, citation analysis method, link analysis method.

互联网出版自由且成本降低,这极大丰富了网络信息来源,但同时加剧了网络信息质量的分化。有必要建立一套较为科学的评价方法以帮助人们更好地获取与利用网络信息。目前网络信息评价模式主要有问卷调查法、专家评议法、在线实时技术监

测、网站访问量统计及用户行为监测等几种。专家评比与问卷调查主观性较大,其代表性与公正性难以得到有效保证。网站流量指标体系较为客观,但目前参加 CNNIC 第三方网站流量认证体系的网站只有 30 多家,对于是否参与排名、是否公开排名结

果完全出于自愿,无法保证网站访问量排名的真实性与全面性。

目前情况下,迫切需要一套客观而公正的方法对网络信息进行评价。可否借鉴情报学领域对传统文献进行定量评价的一些成熟方法从而对网络信息进行评价?引文分析法就是对文献进行定量评价最为有名的方法之一。它的主要依据和研究内容是科学文献之间的引证与被引证关系。

互联网存在着一种与传统文献引证与被引证关系十分相似的情况,就是站点之间的链接与被链接关系。万维网就是由链接点和链组成的资源网络。链接点是超文本表达信息基本单位,表现形式可以是一个网站,网站中的某个频道,某个网页文件,文件中的段落、图像,甚至是电子邮箱。链用来连接相关的链接点,揭示链接点之间的联系,是超文本的灵魂。两个网页之间建立链接,一个是主动实施链接的网页,即施链网页,另一个则为被链接网页。根据施链网页与被链接网页是否处于同一个主机上,可以将某一站点的链接分为站内链接与站外链接两种。

目前已有强自力<sup>[1]</sup>、崔雷<sup>[2]</sup>等作者开始利用链接关系对网页进行定量评价。Aguillo<sup>[3]</sup>仿照《科学引文索引》对一些著名网站的被链情况编制了《World Sitation Report》(世界链接报告)。但上述研究都是直接利用链接关系评价网络信息,并未对链接分析法的可行性进行深入探讨。相对于传统文献之间的引用关系,某一网页被链次数远远大于某一篇文章被引用次数,网页之间建立链接的原因也更为复杂。欲利用链接关系对网络信息进行评价,就必须首先对网页之间建立链接的类型与动机进行分析。

目前对链接类型的论述主要集中在站内链接类型上。如 Trigg<sup>[4]</sup>将链接类型分为普通链接与评论性链接两大类, Joslyr<sup>[5]</sup>将链接类型分为实质链接与语义链接两大类;基于图形的超媒体系统(GBH)中的链分为推理链和导航链两大类<sup>[6]</sup>;槐鹤玲<sup>[7]</sup>指出网页之间链接关系包括相似、例如、因果、时间、行为或性质等。

但是网站的站外链接与站内链接的原因是明显不同的。本文主要是利用搜索引擎检索有代表意义的网站被链次数,对网站链接的特征进行调查,并通过对有代表性网站的站外链接的调查,对站外链接类型与特征进行分析,同时与引文分析法进行比较,对链接分析法的可行性进行评价。

## 1 对象与方法

### 1.1 检索工具选择

许多搜索引擎都能检索某一网页的被链次数,可直接用于检索链接关系。表1列出了著名的搜索引擎专业评价网站 Searchenginewatch 推荐的6家能免费检索链接关系的搜索引擎的基本情况。

表1 可检索链接关系的免费搜索引擎基本情况

搜索引擎	收录网页 (百万)	支持 语种	简体中文 检索/显示	区分站 外链接
Fast Search	600	46	能/能	能
Google	1346	26	能/能	不能
Alta Vista	125	25	不能/能	能
Hotbot	110	11	不能/能	能
Excite	50	11	能/能	能
Lycos	38	25	不能/不能	能

注:数据为直接访问上述网站所得,访问时间2001年3月24日。

因为 Google 不能区分站外链接,故排除在外。Fast Search 在其余引擎中收录网页数最多,检索功能最强,其高级检索功能可指定检索全世界46种语种中任一语种编辑的网页,并能将检索结果以相应语种显示;可检字段有全文、题名、URL、链接描述文字、链接到某网页的所有网页(网站)等5种。Fast Search 还可将检索结果限制在某一类型域名内。如“include.edu.cn”,表示将检索结果限制在中国教育科研网内。同样可排除某一类型域名内的检索结果。因此本文选择 Fast Search 作为检索链接关系的工具。

### 1.2 被分析站点的选择

链接特征分析网站:综合网站选择的是在 CNNIC 网站影响力调查中排名居前的新浪、搜狐、网易、首都在线、21世纪5家综合门户网站。专业网站选择的也是在 CNNIC 第六次调查中列出的999健康网、伽玛医生、JK123 医疗网站,并补充了金卫网、37C 医疗网。

站外链接类型分析站点选择新浪。因为新浪是中国最有影响力的综合门户网站,在 CNNIC 的多次评比中都名列前茅。新浪提供的资源丰富,按内容分为新闻中心、科技时代、文化教育、财经纵横等频道,并提供免费邮箱、主题论坛、搜索引擎、软件下载

等各项服务,还有香港、台湾、北美、北京等分站。因此新浪被利用、被链接的几率较大,其站外链接类型也应较为齐全。

### 1.3 测试内容

本课题对链接特征和站外链接类型及特征进行测试。

#### 1.3.1 链接特征

利用 Fast Search 高级检索功能分别检索:

- (1)综合网站与专业网站被链情况;
- (2)新浪被不同语种网页链接情况;
- (3)新浪被不同国家及地区注册网站链接情况。

#### 1.3.2 站外链接类型及特征

利用 Fast Search 检索与含有“sina.com”的 URL 建立了链接的所有网页,并且排除主机为“sina.com”的网页,即将检索结果限制为新浪的站外链接网页。测试结果显示与新浪建立了链接的站外网页总数为 435 095 个。取排列居前的 100 个有效链接网站进行抽样调查。分别进入这 100 个网站主页,利用浏览器查看每一个主页的源代码。在源代码中,找出每一个与新浪建立链接的语句。根据源代码相关语句在主页上找出表示每处链接的词语,观察其所在位置,并打开每处链接,进入被链的新浪网页。分析施链网页与被链网页的关系,对与新浪建立的每一处链接进行归类统计。

某一网站与新浪的连接可能不止一处。如“平阳之窗”网站同时与新浪主页及新浪新闻频道建立链接。因而上述 100 个网站与新浪总共建立了 191 处链接。对这 191 处链接进行以下 3 个方面的统计分析:

(1)站外链接类型分布。网站之间建立链接的原因很多,本文在对预实验结果分析的基础上,根据施链网页与被链网页之间的关系将链接类型分为推荐链接、合作链接、相关链接、资源链接、通讯链接、广告链接 6 种。将上述 191 处链接分别归入这 6 种链接类型中,进行统计。

(2)不同类型链接点被链情况。分别点击上述 191 处链接,进入对应的新浪链接点,观察该链接点是主页、频道、网页、邮箱、BBS(指新浪主题社区、专题论坛、聊天室等栏目)或者 Banner(是一种在页面四周附加的图标或小型的插图,通过这一小部分内容来提供通向其他页面的超链接、发布信息/广告

等)统计新浪不同类型链接点的被链情况。

(3)站外链接类型与链接点类型的关系。在前面调查的基础上,记录上述 191 处链接中每处链接的类型与链接点,分别统计 6 种链接类型出现在主页、频道、网页、邮箱、BBS 或 Banner 中的次数。

### 1.4 测试时间

第一部分测试时间为 2001 年 2 月 27 日,第二部分测试时间为 2001 年 3 月 31 日。

### 1.5 测试说明

(1)利用 Fast Search 检索上述网址的被链情况时,都去除了网址中的前缀“http://www.”。如检索新浪(http://www.sina.com)的被链情况时,检索的是与包含了“sina.com”的 URL 建立的所有链接。根据新浪不同链接点的网址特征(见表 2),我们发现这些链接点包括了新浪各个分站的主页、所有频道(如新浪新闻中心)、频道中的所有文章,还有新浪专题论坛、聊天室或新浪邮箱,即覆盖了与新浪不同类型链接点建立的所有链接,得出的是整个网站的被链情况。其他网址也是如此。

表 2 新浪不同类型链接点 URL 举例

链接点	URL
主页:	http://www.sina.com.cn (北京站)
	http://www.sina.com (北美站)
频道	http://dailynews.sina.com.cn(新闻中心)
	http://edu.sina.com.cn(文化教育)
网页:	http://tech.sina.com.cn/o/58539.shtml
邮箱:	webmaster@sina.com
Banner:	http://ad3.sina.com.cn/ads/landing/game.htm

(2)本试验所得出的某一网站的总链接数、站外链接数、站内链接数是在测试时间点搜索引擎检索得出的数目,可能并非该网站链接数的全部。但是搜索引擎是目前获取网络信息的主要手段,未被搜索引擎收录的网页被访问、被利用的机会极微,因此搜索引擎检索出的某一网站的链接数更能反应该网站被利用的情况。

## 2 结果与讨论

### 2.1 链接特征分析

#### 2.1.1 不同类型网站有不同的链接特征

不同类型网站提供的内容和服务是不同的,因

此被链接的机会也是不同的。本实验中综合网站选择的是中国较为著名的五大门户网站,它们都拥有多个频道,收录新闻、文教、IT、体育、生活、健康、财经等多方面内容,同时提供免费/收费邮箱、聊天室、论坛、网上商城等多种服务,而5个专业网站都是仅提供医学资讯的专业网站,并不提供邮箱、网上商城等其他服务,因此综合网站的站外链接数、站内链接数、总链接数都远远高于专业网站(见图1)。该结果提示我们,在利用链接次数评价网络信息时须选择好比较对象,使其收录内容及提供服务尽量相似,具有可比性。

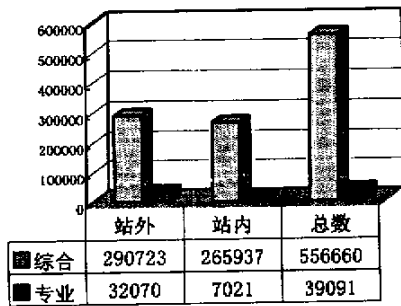


图1 综合网站与专业网站被链次数比较

### 2.1.2 站外链接在总链接次数中比例极大

站内链接数与站外链接数表达的意义是不一样的。站内链接主要用来组织内部网页,站内链接数越多,反映该网站收录网页多,导航机制健全。而站外链接数反映的是该网页的外界影响力。表3显示综合网站的站外链接数占总链接数的52%,高于站内链接数所占比例。而专业网站站外链接数所占比例更是高达82%!因此通过站外链接数对站点特别是专业站点的外界影响力进行评价是有意义的,是可行的。

表3 站外链接与站内链接所占比例

	站外链接		站内链接	
	次数	百分比	次数	百分比
综合网站	1453614	52%	1329687	48%
专业网站	160352	82%	35104	18%

### 2.1.3 覆盖面广

互联网是一个跨越地域、跨越国家的全球性网络,从理论上讲,全球的任一网站、网页之间都可以建立链接关系。链接关系反映的是网页在全世界范围内的影响力,这样广的覆盖范围是传统的问卷调查法望尘莫及的。表4和表5从不同侧面反应了新

浪被不同语种和不同国家网页链接的情况。这些数字反映了新浪在世界各地的影响力,同时也反映了国际之间网络信息交流的情况。

表4 新浪被不同语种网页链接情况(次数)

简体中文	英文	繁体中文	日文	朝鲜文	俄文
480762	230916	192267	485	402	65

表5 新浪被不同国家及地区注册网站网页链接情况(次数)

台湾	香港	日本	英国	俄罗斯	韩国
1124	81563	1176	199	170	107

## 2.2 站外链接类型及特征分析

### 2.2.1 链接类型分析

(1)推荐链接。推荐链接较特殊,施链网页与被链网页之间不需要存在某种相关性,施链网页多在“精彩网站链接”、“热站推介”、“推荐网站”、“网络导航”、“首页链接”、“热门网站”等标题下列出它认为质量好或者热门的站点。文献[8]列出的15种传统文献引文动机有5种是反面引用,是对参考文献的批评、修改或否定;文献[9]列出了6种不恰当或不必要的参考文献引用情况。而推荐链接的动机单纯而明确,是施链网页对被链网页的直接肯定。利用推荐链接评价网络信息质量比利用参考文献评价印刷型文献质量更为有效。

(2)合作链接。调查发现合作链接的原因主要为:①使用了被链网页的服务功能。如“广润在线”网站使用了新浪的全套搜索目录。②主办单位之一。如网上期刊《中国新闻传播学评论》在主页建立了与其合作单位“新浪网”的链接。③信息来源。如“江门国际网”在“更多新闻”标题下建立与新浪新闻频道的链接,将新浪新闻直接作为自己网站内容的补充。④内容相关。“二十一世纪体育”网站与新浪球报频道建立的链接就属于这种类型。

(3)相关链接。施链网页与被链网页内容上关联程度最高。因新浪为综合门户网站,内容包罗万象,施链网页多是与新浪某一频道建立相关链接。本次调查中的5处相关链接中,有4处都是与新浪相关频道建立的链接。如“驱动之家 Mydrivers”与新浪龚师傅硬件频道建立的链接。

(4)资源链接。指链接了被链网页的某种资源,如新闻报道、网页文章,或获取了软件、音频、视频等非文件型文件。这是6种链接中数量最多的。上述

91 处资源链接有 90 处是引用新浪的新闻报道或网页文章,只有一处是利用新浪的软件资源。

(5) 通讯链接。新浪提供的通讯服务很多,如免费邮箱、主题社区、论坛、聊天室等,这也是目前中国网站吸引用户的重要手段。上述 191 处链接中有 22 处链接使用的是新浪的通讯功能(见图 2),主要是使用新浪邮箱,作为网站管理者联系信箱、新闻媒体接受投稿、用户联系方式等。

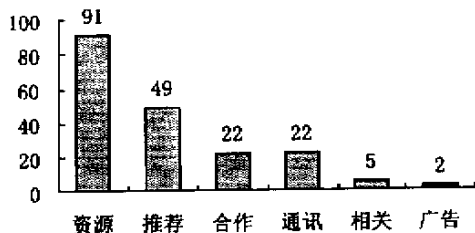


图 2 新浪站外链接类型分布

(6) 广告链接。互联网上的广告通常是在主页、访问量多的频道和网页上设一个 Banner,再由此链接到相关广告上去。主要分为 4 种类型:①为其他企业所作广告;②为自己的商业性服务所作广告;③商务咨询公司建立的与电子商务网站的链接,本次调查中的两处广告即属于这种情况;④个人网站资助性广告。①、②类广告多以站内链接的形式出现,是目前互联网广告的主要形式。从本次调查结果来看,广告在站外链接中所占比例不大,不会对利用站外链接评价网络信息质量造成干扰。另外,一个网站的广告投放量大,也能从侧面证明该网站访问量、质量高。

### 2.2.2 站外链接特征分析

表 6 链接类型与新浪链接点关系分析

	主页	频道	网页	邮箱、BBS	Banner	合计
推荐链接	41	8	0	0	0	49
合作链接	14	8	0	0	0	22
相关链接	1	4	0	0	0	5
资源链接	0	0	91	0	0	91
通讯链接	0	0	0	22	0	22
广告链接	0	0	0	0	2	2
合计	56	20	91	22	2	191

(1) 与主页及频道建立的链接价值更高

推荐链接直接对被链网页质量作出了肯定,在评价网络信息上是极为有效的。合作链接与相关链

接中施链网页与被链网页密切相关。施链网页在众多类似站点中选择 A 站点而不是选择 B 站点作为其合作伙伴(或相关网站),表示施链网页认为 A 站

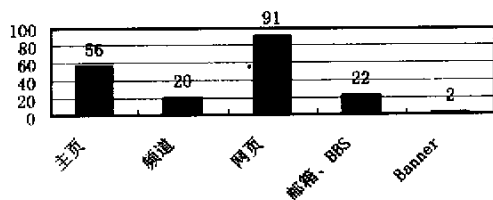


图 3 新浪不同类型链接点被链次数

点是最适于利用的,实质是对 A 站点质量的肯定,即 A 站点作为“合作伙伴”或“相关站点”被链接与 A 站点的质量有某种正向(肯定)的联系。相对瞬息万变的网络信息来说,以这几种形式出现的链接较少更改,稳定性较好。分析表 6 我们发现这 3 种链接类型共有 76 处,全部出现在与主页和频道建立的链接中,因此与主页和频道建立的站外链接对评价该网站质量价值最高。

(2) 网页链接数量上最多,但价值略低且变化大

表 6 显示与网页建立的链接共 91 处,全部为资源链接。虽然与新浪网页建立的链接是其站外链接中最多的,但存在一个网站链接新浪多处网页的情况。如 SOYOU 新闻频道 2001 年 2 月 27 日一次性引用了新浪 32 处网页的新闻报道。虽然引用新浪网页也能说明该网页质量高,但新浪 32 处网页被同一个网站引用,与新浪主页被 32 个网站作为推荐网站链接在评价新浪质量上是有显著差距的。另外网页链接是变化最大的,如 3 月 17 日 SOYOU 新闻频道引用新浪的网页就变为 20 处了。因此在利用网页链接评价网络信息质量时须慎重。

值得注意的是某一网站如果大量引用新浪网页,有时也会将新浪主页或频道作为推荐站点、合作伙伴或相关站点建立链接。如 SOYOU 新闻频道同时在“好站 HOT”栏目下列出了“新浪新闻中心”。另外一般网站为减少服务器负担,都要求其他站点与该网站主页而不是某一网页建立链接,通过主页再访问相关网页。

(3) 调查发现,与邮箱、BBS 链接点建立的链接全部为通讯链接,主要是利用新浪提供的服务。与 Banner 建立的链接全部为广告链接。

## 3 结 论

本文调查发现,新浪站外链接主要可分为推荐

链接、合作链接、相关链接、资源链接、通讯链接及广告链接 6 种类型。新浪不同类型链接点被其他网页链接的主要原因有 (1) 认为新浪质量高, 直接推荐新浪 (2) 与新浪存在某种合作关系 (3) 内容上密切相关 (4) 引用新浪的资源, 如新闻报道、文章、软件资源等 (5) 使用新浪的某种服务, 主要是免费邮箱、专题论坛等 (6) 广告。上述 6 种情况都能从不同侧面不同程度说明新浪质量高。与引文分析法的基本假设相比, 一个网页被其他网页链接的原因有两种: 一是提供的资源或服务被利用, 二是因为该网页质量高而被推荐。这两种原因都能说明一个网页被链接与其质量存在正向(肯定)的关系, 链接分析法符合引文分析法的基本假设<sup>[10]</sup>。一个网页被链接次数越多, 从一个侧面证明该网页质量越高, 可以利用被链次数对该网页信息质量进行评价。而且网站的站外链接数普遍高于站内链接数, 专业网站的站外链接比例更高, 因此利用站外链接评价网络信息是可行的。

互联网是一个跨越地域与国界的全球性网络, 链接关系反映的是网页在全世界范围内的影响力, 这样广的范围是传统问卷调查法所望尘莫及的。能直接利用搜索引擎来检索某一网页被链情况, 这大大加快了数据收集、积累和处理的过程。并且这些搜索引擎绝大部分都是免费的, 其检索结果是一个客观的数字, 易于操作、经济实用, 有效地弥补了同行评议等定性方法的缺陷。

链接分析法借鉴了引文分析法的方法与思路, 是情报学研究方法在网络时代的新应用。但网络信息资源涵盖范围广、变化大, 网页之间的链接关系也极为复杂, 本文是对链接分析法的一个初步探讨, 期待同行专家深入的研究。目前链接分析法的主要实

验工具为搜索引擎, 搜索引擎的发展必将为链接分析法的深入应用提供更为有力的支持。

### 参 考 文 献

- 1 强自力. 利用搜索引擎高级检索功能评价大学图书馆 Web 站点. 大学图书馆学报, 2000(4): 53 ~ 54, 64
- 2 崔雷, 丁雁. 互联网上重要临床医学网址简介. 情报科学, 2000, 18(2): 164, 176
- 3 Aguillo IF. STM information on the Web and development of new Internet databases and Indicators Proc. [online]. 1998. Learned Information. London. Available from :http://www.cindoc.csic.es/cybermetrics/link21.html Visited date :2001-4-30
- 4 Trigg R. A Network-based approach to text handling for the on-line scientific community. Chapter 4 :taxonomy of link types : [ Ph. D. dissertation ]. University of Maryland Technical Technical Report. TR-1346 1983. 11
- 5 Joslyn C, Heylighen F. Links and link types. http://pepsmel.vub.ac.be/linktype.html Last updated :1995. 6. 29
- 6 基于图形的超媒体系统 GBH. 网上周刊, 1997 (682): 119, 1997-11-03
- 7 槐鹤玲. 教育超文本中的概念图解导航. 首都师范大学学报(自然科学版), 1999(3): 66 ~ 70
- 8 Weinstock M. Citation Indexes (part 1). From :Encyclopedia of Library and Information Science. vol 5. New York :Marcel Dekker, 1971
- 9 Jhorne FC. The citation index :author case of spurious validity. Journal of Clinical Psychology, 1977, 33 :1157 ~ 1161
- 10 刘君玉. 引文分析的评价. 情报理论与实践, 1998, 21(2): 87 ~ 89

(责任编辑 许增棋)