

开源软件 Piwik 的门户网站统计技术和方法分析

温明章

(华南农业大学图书馆 广州 510642)

【摘要】探讨 Piwik 开源软件在图书馆门户网站系统中的应用,在此基础上扩展收集真实 IP 的功能,并重点介绍系统的参数指标间的关系和可靠性,包括退出率、页面停留时间、转化率、访问来源和关键词等参数指标。

【关键词】Piwik 网站统计技术 统计方法分析

【分类号】G250.7

Analysis of the Portal Statistical Techniques and Methods Based on Source Software Piwik

Wen Mingzhang

(South China Agricultural University Library, Guangzhou 510642, China)

【Abstract】Open source software Piwik is applied in the Web portal system of library. On the basis of the Piwik, the paper extends the function of collection of real IP, and puts emphasis on the relationship and reliability of the parameters in the system, including the bounce rate, page retention time, conversion, referer and keywords etc.

【Keywords】Piwik Technology of website statistics Analysis of statistical method

1 引言

随着 Cookie 和 Ajax 技术的迅猛发展并且成熟地被应用到各领域,在线分析工具也得到快速发展。在线分析工具对网站访问信息进行记录和归类,在此基础上进行统计分析,如网站访问量的增长趋势图、用户访问最高的时段、访问最多的网页、停留时间、用户使用的搜索引擎和主要关键词等。这些数据可以分析门户网站各页面中栏目的布置和可到达率问题,提升用户操作满意度。建立网站的统计分析系统,可以有效地评估图书馆门户网站中资源栏目的设计使用、部署及用户使用的效果,对于合理改善图书馆网站发布资源的页面布局和提升用户体验都有较好的推动作用。

2 在线统计分析工具分析

目前,在线统计分析工具有 SiteCatalyst^[1-3]、WebTrends^[4-7]、Google Analytics^[8-10]、百度统计^[11,12]以及 Piwik^[13]等。从表 1 分析可知,免费的在线分析工具可满足图书馆的一般需求。在免费的在线统计分析工具中,Piwik 软件在开放和开源方面表现比较突出。

收稿日期:2012-06-07

收修改稿日期:2012-08-13

表1 商业模式比较

工具	性能比较	适应用户群	商业模式
SiteCatalyst/ WebTrends	产品功能全面 强大	大型用户	收费服务,服务费用高
Google Analytics	功能全面,设置 复杂	大中型用户	免费使用,部分 API 开 放使用
百度统计	功能较全面	中型用户	免费使用
Intelli- ture ^[14-16]	功能全面可靠	中型用户	提供免费的和收费的 两种商业模式
Piwik	功能全面可靠	中型用户	开源免费使用,支持 GPLv3 授权协议

在应用技术方面,分析技术主要有 HTTP 协议头信息技术^[17]、第一方 Cookie^[18]、第三方 Cookie^[18]、P3P 协议 Cookie^[19]、LSO 技术^[20] 以及 Web Beacons 和 JavaScript 技术,笔者实验得出现有工具使用的各项技术基本一致,应用技术的情况如表 2 所示:

表2 技术对比

工具	日志 统计 型	脚本 统计 型	HTTP 协议 头信 息	第一 方 Cookie	第三 方 Cookie	P3P 协议	LSO 技术	Web Bea- cons	JavaScript (Ajax)
Google Analytics	×	√	√	√	×	√	-	√	√
百度	×	√	√	√	×	√	-	√	√
WebTrends	√	√	√	×	×	√	-	√	√
SiteCatalyst	×	√	√	√	×	√	-	√	√
Piwik	○	√	√	√	×	√	○	√	√
Intelliture	×	√	√	√	×	√	-	√	√
其他的在线 统计工具	-	√	√	√	×	√	-	√	√

(说明:“√”表示“支持”,“×”表示“不支持”,“-”表示“不定项”,“○”表示“可扩展”。)

表 2 可以说明所有的在线分析工具都采用了 JavaScript、Web Beacons、第一方 Cookie 技术和 HTTP 协议头信息以及 P3P 协议 Cookie 技术等。从在线分析工具的商业模式和应用技术上分析,Piwik 开源软件比较适合图书馆使用,所以笔者使用 Piwik 进行图书馆门户网站的用户行为研究。

3 系统实践思路

目前,有关网站统计技术应用方面的文献较多,有部分文献只是使用 Web 服务器日志来分析访问者行为^[21,22],服务器日志收集到的用户信息比较少,无法收集访问者使用环境情况和停留时间等一些信息,而且只是论述整个的网站访问量或是单个页面的访问量情况,而对访问的来源、访问者的行为路径和停留时间等访问行为参数指标并没有论述。虽然有些文献论述过访问行为的一些参数指标^[23,24],但是没有对这些参数指标间的关系和可靠性进行分析。笔者利用 Piwik 系

统的前端运行 JavaScript 技术来收集访问者的大量访问行为数据信息,包括访问者使用环境、停留时间和访问上下文等信息,其分析原理和技术已经非常成熟,也在互联网上各商业网站得到了广泛的应用,但在图书馆门户网站的应用较少。笔者并没有对具体的 Cookie 和 JavaScript 等技术和原理进行剖析,而是在 Piwik 系统实际应用的基础上,重点对 Piwik 系统收集的页面停留时间、转化率、访问来源分类和关键词等数据参数指标进行分析,并在实践中对这些指标间的关系或可靠性进行考量,从而辨别出页面或转化目标是否需要优化。

基于 Piwik 的系统分两个部分实现,即 Piwik 开源软件安装与改造和图书馆门户页面的改造。Piwik 开源软件安装与改造是在图书馆的独立服务器上安装 Piwik 系统(具体安装过程可访问 <http://piwik.org/docs/installation/>),可以响应校园网和公网读者或服务的请求,并按 Piwik 开发规则,增加收集真实 IP 信息的功能。图书馆门户页面的改造是在图书馆门户目标页面上加入 Piwik 的 JavaScript 跟踪代码,目标页面就是网站管理员认为需要进行统计分析的页面以及转换目标。

4 基于 Piwik 软件的系统开发

4.1 Piwik 基本分析

Piwik 是一个 PHP 和 MySQL 架构的开放源代码的 Web 统计软件,使用 JavaScript 技术跟踪所有 PC 机用户,而且使用 Web Beacons 技术来支持非 JavaScript 环境。Piwik 的安装非常简单,使用者只需要按系统提示即可完成。Piwik 可以提供退出页面、网页逗留时间、关键字分析、访问路径、外部网站和最佳搜索引擎等功能。系统的核心工作流程如下:

(1) 访问者通过浏览器向门户服务器发送请求。

(2) 门户服务器响应请求,返回预先设置了 Piwik 的跟踪代码的页面。Piwik 的跟踪代码部分如下:

```
<! -- Piwik -- >
<script type = "text/javascript" >
.....
try {
var piwikTracker = Piwik.getTracker( pkBaseURL + " piwik.
php", 1);
//请求 Piwik 服务器端进行开始跟踪
.....
} catch(err) {}
```

```

</script >
//TAG begin
< noscript > < p > < img src = " http://localhost/piwik/piwik.
php? idsite = 1" style = " border:0" alt = " /> </p > </
noscript >
//idsite 为跟踪页面在 Piwik 系统的编码,每个目标页面有一个
唯一号码
<! -- End Piwik Tracking Code -- >
//TAG end

```

(3) 客户端浏览器执行 TAG 标记,向 Piwik 请求数据获取的 JavaScript 代码。

(4) Piwik 服务器响应请求,返回用于收集网页、浏览器和 Cookie 信息的 JavaScript 代码。

(5) 客户端浏览器执行 JavaScript 代码,发送相关网页被访问的行为数据给 Piwik 服务器。

(6) Piwik 服务器产生点击流日志,保存于后台数据库中,以便日后统计分析。

Piwik 收集的信息有:被系统屏蔽后的 IP 地址信息、访问来源链接、第一次访问时间、累计网页浏览数和第几次来访数据等。由于 Piwik 比较注重用户隐私,对 IP 信息进行了屏蔽,这也是 Piwik 的不足之处,为了能够准确定位读者群,笔者需要对访问来源的 IP 信息进行分析,并对 Piwik 进行必要的二次开发。

4.2 Piwik 收集 IP 信息的二次开发

(1) Piwik 后台数据库改造

Piwik 将用户的访问时间信息、访问事件和浏览器的环境等信息记录在 MySQL 数据库的 Piwik_log_visit 表中,笔者在此表上新增加一个 ipinfo 字段以便可以存放用户的真实 IP 信息。在 MySQL 的命令窗口执行如下命令语句完成信息表的建设:

```

“alter table Piwik. 'Piwik_log_visit' Add column ipinfo varchar(20)
not null default 0 AFTER 'location_provider'”

```

(2) 新建一个获取 IP 信息的类

Piwik 提供插件扩展功能的方法^[25],所有的二次开发类必须遵守此方法。具体实现步骤为:在 Piwik 的 plugins 目录中建立 ipinfo. php 扩展类继承 Piwik_plugin 类,重写 Piwik_plugin 中 getListHooksRegistered 方法获取已经注册的“Tracker. Visit. setVisitorIp”钩子;将此钩子绑定一个方法来获取 IP 信息;将 IP 信息存入 Piwik_log_visit 表的 ipinfo 字段中。ipinfo. php 部分代码如下:

```

class ipinfo extends Piwik_plugin { //继承 Piwik_plugin 类

```

```

.....
Public function getListHooksRegistered() {
    return array( 'Tracker. Visit. setVisitorIp' => 'setVisitorIpAd-
dress'
//注册一个“Tracker. Visit. setVisitorIp”钩子来获取 IP 信息。
);
}
Static public function applyIPMask( $ip, $maskLength ) {
    //解开屏蔽 IP 信息
    $i = Piwik_Common::strlen( $ip );
    If( $maskLength > $i ) {
        $maskLength = $i;
    }
    While( $maskLength -- > 0 ) {
        $ip[ -- i ] = chr(0);
    }
    Return $ip;
}
function setVisitorIpAddress( $notification ) {
// 建立与“Tracker. Visit. setVisitorIp”钩子绑定的方法
    $ip = & $notification -> getNotificationObject();
//得到的参数引用相关联的钩子
    $ip = self::: applyIPMask ( $ip, Piwik_Tracker_Config::: getIn-
stance() -> Tracker[ 'ip_address_mask_length' ] );
//获取真实 IP 地址
}..... }

```

5 Piwik 统计参数指标在实际应用中的利用与分析

同一个读者每次访问图书馆网站,在图书馆网站或者某个页面产生的停留时间、访问来源(如搜索引擎、搜索关键字和网站来源)、离开页面链接和目标转化率等都有可能不一样,分析这些信息才能使网站统计分析有意义。笔者重点结合实际使用情况对 Piwik 的跳出率、目标转化率、页面停留时间和访问来源分类等指标进行利用和分析。

5.1 跳出率^[12]与停留时间的数据关系分析

x 是跳出网站/网页的访问次数,y 是访问网站/网页的停留时间,B 是跳出率,V 是访问次数,A 是访问者在网站/网页平均停留时间,T 是所有访问网站的总时间。跳出率与停留时间关系如下^[26]:

$$A = T/V \quad (1)$$

$$T = 0x + (V - x)y \quad (2)$$

$$x = VB \quad (3)$$

公式(1)和(3)代入公式(2)得出:

$$A = [0x + (V - x)y] / V = (V - VB)y / V = y(1 - B) \quad (4)$$

由公式(4)得出:

$$y = A / (1 - B) \quad (5)$$

公式(5)^[26]可以证明:跳出率越低,网站的平均停留时间就越接近实际停留时间。所以在优化网站布置和设计时,不仅要有平均停留时间的统计还要参考退出率统计的大小。

如果仅分析单纯的页面/网站停留时间意义不大,主要原因有:

(1)这些指标是战术层面的,分析员无法认识到这些数据对页面统计的影响,单纯的页面停留时间或网站停留时间并不能反映出具体的页面对读者的影响力;

(2)这些指标需要很多推断,这些数据无法直接推断出某页面的好坏,简单的逻辑依据数据量越大越好;

(3)这些指标主要是短期数据,随着互联网的发展,依照访问会话来度量指标远远不能满足需求,长期的访客行为、读者生命周期等更加能反映问题,不同的停留时间需要用不同的逻辑去判断。

笔者进行的实验如表3所示:

表3 平均访问时间的实验

网站	平均访问页数	平均访问时间	跳出率
华南农业大学图书馆门户网站	35	35分钟	23%
网络数据库网站	15	15分钟	50%

出现上述不同数据的主要原因有:华南农业大学读者想获取本馆资源;第一时间可能是访问华南农业大学图书馆的门户网站;因为目标位置不确定,门户网站内容丰富,所以读者要多花费时间;网络数据库网站是一个对电子资源进行分类的导航网站,提供各种资源的访问入口,因此读者进入网络数据库一般都有明确的目标,快速找到目标、搜索,然后快速离开,也可能是网站界面简洁或是读者很容易就找到想要的东西,所以网络数据库网站的跳出率高,停留时间自然就少,而网络数据库的读者使用量很高,这时单纯的跳出率数据分析就失去了意义。

笔者在网站分析的初期,发现直接使用单一停留时间数据,在有些情况下不能找出数据变化的原因,而且通常会耗费大量的精力。经过多次实验后,笔者理清停留时间与跳出率的关系,基本可以判断出准确的统计分析结果。笔者并没有列出具体停留时间与跳出

率对应的关系值,是因为统计分析是一个比较综合的分析过程,不同的网站和服务对象等因素都会影响它们的关系,所以还要依靠分析者在应用中不断总结。

5.2 转化率和跳出率的分析

转化率是读者访问图书馆网站,到达目标页面(链接、下载图片、文章等)的人次与网站总人次之比,比如所有到达淘宝店铺并产生购买行为的人数和所有到达店铺的人数的比率就是淘宝店铺的转化率。

对于华南农业大学图书馆网站,笔者将如下信息定为转化目标,也是重点跟踪对象,按触发动作划分目标:

(1)链接到外部网站:网络数据库栏目下的所有资源导航链接(如CNKI、维普等)、CALIS的外文期刊导航平台、广东省高校联盟平台、珠三角数图联盟和天河九校资源等;

(2)访问特定网站:馆藏书目检索、读秀、馆内通知、借阅规则、党建园地和机构组织等;

(3)下载文件:随书光盘、表格和一些常用软件。

它们的转化率统计如表4所示:

表4 2月份转化率统计

转化目标	转化率
外文期刊导航平台(一级栏目)	16%
二级栏目下的外文期刊导航	7%
老版的CALIS第三方统一检索平台	几乎为0%
珠三角数图联盟(一级栏目)	9%
广东省高校联盟平台(一级栏目)	15%
馆藏书目检索(一级栏目)	60%
CNKI	37%
维普	15%
借阅规则	2%
党建园地	5%
机构组织	5%
随书光盘	39%
常用软件	4%
读秀	38%

从表4可以看出,转化率高说明跟踪对象主体设置和布置有明显效果,内容更加符合读者使用;CNKI、维普和老版的CALIS第三方统一检索平台都是部署在同一个级别栏目上,但CALIS第三方统一检索平台的转化率几乎为零,笔者发现此平台的建设速度慢,资源配置技术要求高,而且整合的资源过少,检索命中率低,不能满足读者要求,所以这个平台已被屏蔽。而外文期刊导航在2011年11月至2012年2月期间,一级栏目上就有16%的转化率,二级栏目只有7%,这体现出栏目部署很重要。外文期刊导航平台、广东省高校联盟平台、珠三角数图联盟和天河九校资源有部分资

源可能重复,读者在某个平台上获取资源的转化率高,说明读者比较喜欢这个平台,利用率自然会升高。对一些常用的重点资源,比如 CNKI、维普和超星电子书,转化率自然就高,在网站布置时可以考虑重点突出这些高转化率的平台和资源,但是对一些刚推广资源和稀缺资源,使用对象本身很少,分析员也要兼顾考虑。

以上跟踪对象的跳出率统计如表 5 所示:

表 5 2 月份跳出率统计

跳出率为 0-10%	11% - 20%	21% - 40%	41% - 70%	71% - 90%	90% - 100%
馆藏书目 查询	读秀数据库	图书馆门 户网站	老版网站	党建园地	主页流动 通知
随书光盘	外文期刊 导航	网上展厅	软件下载	机构组织	培训通知
CNKI, 维普	广东省高校 联盟平台/珠 三角数图联 盟/天河九校 资源		网络数据 库网站	借阅规则	错误页面 跟踪

从表 5 可以看出,跳出率分 6 等数量级别,分级的数据只是笔者经实验做出的判断,并得出如下结论:同类型的导航库或栏目的跳出率数量级别基本一样;对于那些能帮助用户很快完成任务(如通知类)的页面,跳出率就会高;对于分类页或导航用于引领用户深入发掘的页面,跳出率就会比较低。

如何分析才有利于提高转化率和降低退出率,关于这方面论述有不同的观点和建议,不过有一个共同的观点即优化网站和页面的关键词。Google 公司的联合创始人 Kaushik^[27] 提出几个提高转化率和降低退出率的观察方法,笔者认为其中三点比较符合优化图书馆网站:

(1)分析网站流量的来源:特别是关于关键词和 URL 的来源,确定网站被访问者搜索的命中率;分析访问者的访问途径是否达到网站设计的预期目标和效果。分析来源是一个最基本的观察方法,网站管理者应该清楚转化目标是如何被读者获取、通过什么路径到达目标,只有这样才可以认清网站和页面优化目标。

(2)页面触发动作是否明显突出和有效:这是离开链接、目标和页面来源整合考量项。使用此项观察方法,必须结合 Piwik 统计项的离开链接、目标和页面来源的数据。

(3)关注访问者反馈信息:这是最直接和最好的方法,也是必要的。只是反馈信息还要依赖读者的参与度和反馈的质量,如果反馈的信息少或是反馈信息

质量差,分析员也无从下手。

除了以上三种方法,笔者还总结出以下分析方法:

(1)栏目(资源等)归类分析。从表 5 可以得出同类型的资源应被归类统计,分析员定出同类资源参考值范围,然后观察在某个时期内的数值变化情况,如果变化值超出参考值范围,应该立即深入调研。

(2)同一个资源在不同级别栏目下同时考察。有很多资源内容会涉及到多个类型的栏目,栏目分类的合理性将体现为相应转化率。如果某资源具有高转化率说明读者接受此资源在网站的分类情况,低转化率则说明此资源分类不合理,需要调整和优化分类。

(3)分时段考察。有些资源是时段性的,比如每年 9 月至 10 月新生入学,借阅规则、培训通知等新生急需资源的转化率自然就高,而在其他时间就会很低,分析员不能因此否定这些资源的合理性,因此要合理分时段统计分析资源。

5.3 访问来源的分类应用

笔者重点要统计出读者是如何到达被统计页面的,读者的访问路径是怎样的。这里的访问来源是指一次访问或一个网页浏览的真实 IP 来源(简称推荐来源)。访问来源的统计是提高转化率和降低退出率的一个重要分析方法,Piwik 提供最佳搜索引擎、外来网站以及来源国家/地区的统计功能,这些不能满足笔者的要求,而且来源国家/地区的统计功能没有实际意义,笔者将访问来源从不同的角度进行分类。

(1)按访问来源网站的性质分类

①搜索引擎。如果访问起始于搜索引擎的搜索结果,它的访问来源就可以归类为搜索引擎。这类访问来源还可以进一步细分到具体的搜索引擎(百度、Google 等),然后还可以利用搜索的关键词进行细分。

②一般网站。访问来自于搜索引擎以外的网站,例如图书馆门户网站友情链接等。大多数工具会将广告来源作为一个独立的分类,通过提取 URL 中预先加入的特殊参数来获得一些更为详细的附加信息。

③没有访问来源。通常是指访问者直接进入网站的情况,例如访问者直接在浏览器里输入网站地址,或者从收藏夹中点击网站地址。但是由于技术上的局限性,造成了一部分访问来源丢失的情况,例如浏览器的弹出窗口和网页的重定向等情况也将统计为无访问来源。

(2)按访问来源的 IP 分类

笔者根据华南农业大学网络的具体情况,按 IP 地

址来源进行统计分析,其 IP 地址分配情况如表 6 所示:

表 6 IP 地址分配表

区域	IP 段	区域	IP 段
教工区	202.116.160.0 - 202.116.191.0	跃进南区	172.17.0.0 - 172.17.63.254
	172.18.0.0/16		
研究生楼	172.16.128.0 - 172.16.191.254	跃进北区	172.17.0.0 - 172.17.63.254
华山区	172.16.0.0 - 172.16.63.254	东区	172.16.65.0 - 172.16.127.254

教工区包括教工办公楼以及教工宿舍,研究生楼包括在校博士以及在校研究生宿舍,东区、跃进南北区和华山区基本是本科生的宿舍区。

5.4 搜索引擎的关键词统计方法的分析

现在读者比较喜欢使用百度、Google 等搜索引擎检索资源,近年来很多图书馆员抱怨读者的流失,其中的原因有很多,但是如果现在流行的搜索引擎找不到图书馆的资源页面或网站,这就是图书馆流失读者的一部分原因;或是读者要经过繁琐的搜索才可以找到图书馆的某些资源,这也是图书馆工作的一种缺陷表现。建立一个用户快捷、方便、高效的搜索目标信息和资源的网站将是图书馆网站管理员的重点工作。

在网站设计开发或是资源发布的时候,关键词的选择非常重要,以下是笔者整理的一些关于关键词确定方法以及参考标准:

(1) 关键词搜索结果数量

搜索结果少于 50 万的属于关注度较小的;搜索结果位于 30 - 100 万的属于中等偏小的;搜索结果位于 100 - 300 万的属于中等;搜索结果位于 300 - 500 万的属于中等偏上;搜索结果达到 500 万以上的属于高难度词。

(2) 关键词搜索次数

它的数值反映了关键词的用户搜索频繁度,日搜索量越大,说明此词关注度越高。关键词搜索次数可以分成以下几个数值范围(以百度指数为参考依据):搜索次数少于 100 的属于关注度较小的;搜索次数位于 100 - 300 的属于中等偏小的;搜索次数位于 300 - 500 的属于中等;搜索次数位于 500 - 1 000 的属于中等偏上;搜索次数达到 1 000 以上的属于高难度词。一般难度太高的关键词,根本就不可能实现。

(3) 关键词的命中率

网站或资源页面众多关键词中高命中率的词被作

为重点关键词。关键词的命中率可以分成以下几个数值范围:命中率 70% 以上的为最高;命中率为 50% - 70% 的为高;命中率为 30% - 50% 的为中等;命中率为 10% - 30% 为低;命中率 10% 以下的为最低。

优化网页设计、网站链接和网页内部结构等才能让搜索引擎更好地抓取信息,对搜索引擎更具有友好性,进而在搜索结果上有更前的排序。要想知道如何优化图书馆网站,管理员就应该知道目前网站被搜索的具体情况,关键词的分析成为关键突破点。关键词的正确选择很重要,还要实际反映图书馆资源内容,导致关键词的分析难度增大。有些图书馆资源根本没有定位自己的关键词,在线统计分析工具无法进行分析,读者通过搜索引擎也无法检索到相关的资源。

6 结 语

在线分析工具使用的技术比较广泛并且非常成熟。Piwik 开源工具为 Web 管理者提供了一个免费的开源环境,而且全面收集了用户信息以便网站管理员进行相关统计分析。统计分析是一个比较综合的分析过程,分析员需要了解各渠道的数据以及统计分析方法的过程,还要知道整个互联网的变化情况,因为网络情况会影响到很多指标,如停留时间、转化率和访问来源等。因此要提供一个网站统计分析的准确数据,分析者就要了解统计分析方法基本技术和网络情况,对参数指标的可靠性和指标之间的关系进行分析。

参考文献:

- [1] SiteCatalyst. Omniture 营销套件[EB/OL]. [2012-02-10]. http://www.sitecatalyst.com.cn/download/form.php?file=09_productoverview_onlinemarketingsuite_LO.pdf. (SiteCatalyst. Marketing Suite of Omniture[EB/OL]. [2012-02-10]. http://www.sitecatalyst.com.cn/download/form.php?file=09_productoverview_onlinemarketingsuite_LO.pdf.)
- [2] SiteCatalyst[EB/OL]. [2012-02-10]. http://www.sitecatalyst.com.cn/download/form.php?file=09_zh_productoverview_sitecatalyst.pdf.
- [3] SiteCatalyst. SiteCatalyst 功能[EB/OL]. [2012-02-10]. <http://www.sitecatalyst.com.cn/function/>. (SiteCatalyst. Function of SiteCatalyst[EB/OL]. [2012-02-10]. <http://www.sitecatalyst.com.cn/function/>.)
- [4] 融海咨询. WebTrends 客户名单[EB/OL]. [2012-02-10].

- <http://webtrends.ronghai.com/webtrends.jsp?pageid=cuslist>. (Ronghai Consulting. Client List of Wet Trends[EB/OL]. [2012-02-10]. <http://webtrends.ronghai.com/webtrends.jsp?pageid=cuslist>.)
- [5] 融海咨询. WebTrends 网站运营分析解决方案[EB/OL]. [2012-02-10]. <http://wenku.baidu.com/view/8833802e7375a417866f8f0d.html>. (Ronghai Consulting. Solutions of Web Analytics on WebTrends[EB/OL]. [2012-02-10]. <http://wenku.baidu.com/view/8833802e7375a417866f8f0d.html>.)
- [6] 张学宏. 北京大学图书馆的主页日志分析[J]. 现代图书情报技术, 2005(5):81-83, 70. (Zhang Xuehong. Log File Analysis of Peking University Library Homepage[J]. *New Technology of Library and Information Service*, 2005(5):81-83, 70.)
- [7] WebTrends. Unified Digital Marketing Performance; Maximizing Return on Digital Marketing Campaigns[EB/OL]. [2012-02-10]. <http://understanding.webtrends.com/forms/JAN27112UnifiedWhitepaper>.
- [8] Google Analytics. 免费使用更智能、更友好的企业级网站分析服务[EB/OL]. [2012-02-10]. <http://www.google.com/analytics/>. (Google Analytics. Enterprise - Class Web Analytics[EB/OL]. [2012-02-10]. <http://www.google.com/analytics/>.)
- [9] Google Analytics. Google Analytics 服务条款[EB/OL]. [2012-04-01]. <http://www.google.com/analytics/tos.html>. (Google Analytics. Terms of Service about Google Analytics[EB/OL]. [2012-04-01]. <http://www.google.com/analytics/tos.html>.)
- [10] 王玥, 周玲, 李博, 等. 利用 Google Analytics 追踪科技期刊门户网站流量数据[J]. 中国科技期刊研究, 2011, 22(2):248-252. (Wang Yue, Zhou Ling, Li Bo, et al. Tracking Traffic Data for Journal's Website Using Google Analytics[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2011, 22(2):248-252.)
- [11] Baidu 统计. 成功案例[EB/OL]. [2012-01-10]. <http://tongji.baidu.com/web/welcome/cases>. (Baidu Analytics. Successful Cases[EB/OL]. [2012-01-10]. <http://tongji.baidu.com/web/welcome/cases>.)
- [12] Baidu 营销. 百度统计教程[EB/OL]. [2012-02-05]. <http://yingxiao.baidu.com/support/class/tongji/>. (Baidu Marketing. Tutorials of Baidu Statistical Tool[EB/OL]. [2012-02-05]. <http://yingxiao.baidu.com/support/class/tongji/>.)
- [13] Piwik. Using Piwik[EB/OL]. [2011-11-10]. <http://piwik.org/docs/>.
- [14] Inteliture. Professional Search Engine Optimization Services[EB/OL]. [2012-02-15]. <http://www.inteliture.com/search-engine-optimization-services.html#analysis>.
- [15] Inteliture. Customers[EB/OL]. [2012-03-16]. <http://www.inteliture.com/customers.html>.
- [16] Inteliture. Search Engine Marketing Case Studies[EB/OL]. [2012-04-11]. <http://www.inteliture.com/search-engine-marketing-case-studies.html>.
- [17] 任栋, 刘连忠. HTTP 头信息在 Web 网站开发中的应用研究[J]. 计算机应用, 2002, 22(3):65-67, 70. (Ren Dong, Liu Lianzhong. Research on HTTP Headers for Web Site Development[J]. *Computer Applications*, 2002, 22(3):65-67, 70.)
- [18] 梁雪松. 基于 Cookie 的认证机制及其安全性分析[J]. 通信技术, 2009, 42(6):132-134, 137. (Liang Xuesong. Cookie-based Authentication Mechanism and Its Security Analysis[J]. *Communications Technology*, 2009, 42(6):132-134, 137.)
- [19] Reagle Jr J M, Wenning R. P3P and Privacy on the Web FAQ[EB/OL]. [2011-10-18]. <http://www.w3.org/P3P/P3FAQ.html>.
- [20] Moscaritolo A. Top Websites Using Flash Cookies to Track User Behavior[EB/OL]. [2011-11-09]. <http://www.scmagazine.com/top-websites-using-flash-cookies-to-track-user-behavior/article/141486/>.
- [21] 刘延华, 陈国龙, 高洪垒. 基于增量复制的 Apache 日志管理分析系统设计与实现[J]. 漳州师范学院学报: 自然科学版, 2009, 22(3):43-47. (Liu Yanhua, Chen Guolong, Gao Honglei. Design and Implementation of Apache Log Management Based on Algorithm of Increment Copy[J]. *Journal of Zhangzhou Teachers College: Natural Science Edition*, 2009, 22(3):43-47.)
- [22] 郭振英, 赵文兵, 魏育辉. 电子资源日志统计系统分析与设计[J]. 现代图书情报技术, 2008(9):102-106. (Guo Zhenying, Zhao Wenbing, Wei Yuhui. Analysis and Design on Electronic Resource Log Statistics System[J]. *New Technology of Library and Information Service*, 2008(9):102-106.)
- [23] 耶健. 网络读者访问流量统计分析系统的研究与实现[J]. 现代图书情报技术, 2005(6):54-59. (Ye Jian. Study and Realization of Flux Statistical Analysis System for Library Network Reader[J]. *New Technology of Library and Information Service*, 2005(6):54-59.)
- [24] 曹进军. 构建基于读者 Web 访问行为分析的决策支持系统[J]. 图书馆杂志, 2007(4):49-53. (Cao Jinjun. Construction of Decision Supporting System Based on the Analysis of Web Access Behavior of User[J]. *Library Journal*, 2007(4):49-53.)
- [25] Piwik. FAQ Piwik[EB/OL]. [2011-11-10]. <http://piwik.org/faq/>.
- [26] The Roi Revolution Blog. Time on Page and Time on Site - How Confident Are You? [EB/OL]. [2011-10-10]. http://www.roirevolution.com/blog/2008/05/time_on_page_and_time_on_site_how_confident_are_yo.php.
- [27] Kaushik A. Six Tips for Improving High Bounce Rate / Low Conversion Web Pages [EB/OL]. [2011-11-10]. <http://www.kaushik.net/avinash/tips-for-improving-high-bounce-low-conversion-web-pages/>.

(作者 E-mail:superskill@scau.edu.cn)