

文章编号:1004-1478(2011)03-0081-04

基于网站群或多网站的 网站流量统计分析系统的设计

史先进

(河南省教育信息中心, 河南 郑州 450008)

摘要:基于网站群或多网站的网络流量统计原理,设计了网站群或多网站的流量统计分析系统:通过三层设计模式(实体、数据访问、业务逻辑)对系统进行了分层架构,实现了网站群或多网站的流量统计功能,网站的承载量、运行效率和稳定性有了较大的提高.该设计可为用户和管理部门修正或优化网络营销策略提供参考依据.

关键词:网站群;流量统计;网站流量

中图分类号:TP393.4

文献标志码:A

Design of the system of network traffic capacity statistics based on website cluster or multi-website

SHI Xian-jin

(Edu. Infor. Centre of He'nan Province, Zhengzhou 450008, China)

Abstract:Based on the principle of network traffic capacity statistics of website cluster or multi-website, network traffic capacity statistics of website cluster or multi-website analysis system was designed: by means of three-layer design patterns (entity, data accessing, service logic), a layered architecture of the system is presented which realizes traffic capacity statistics of website cluster or multi-website and improves distinctly the performance of the system in load capacity, efficiency and stability. The system can provide a reference for revising or optimizing the strategy of network marketing for user and administration authority.

Key words: website cluster; traffic capacity statistic; network traffic

0 引言

随着互联网的普及和发展,电子商务、电子政务越来越多地通过单位门户网站的形式,向公众提供便捷、个性化和差异化的服务,门户网站承载的业务和应用越来越多.网络营销已日益成为外向型企业取得商务订单的重要途径,网络也越来越成为政府为公众提供服务的重要渠道.及时对网站流量

和用户访问信息进行跟踪、分析、研究,并根据来访者访问网站的浏览习惯,优化本部门 and 行业的网络营销、推广策略,对提高电子商务和电子政务的效率有积极作用.

中国网民的数量正在迅猛地增长,如何招揽更多的用户、获取点击率和更多访问量,如何及时调整网站内容,目前已经成为广大网站管理者最为关心的问题.由此可以想到,如果能够及时有效地了

收稿日期:2011-03-07

作者简介:史先进(1973—),男,河南省商水县人,河南省教育信息中心工程师,硕士,主要研究方向为网络应用、信息系统及信息安全.

解网站运营情况,统计访客使用了哪些关键词、通过什么搜索引擎来到网站的,浏览的页面是哪些,以及访问网站的客户端的电脑使用情况等,就可以有针对性地对网站进行改版,以此来迎合来访者,使网站的有效利用最大化。

设计良好的网站流量分析系统,能够让网站的管理者及时了解网站流量和来访者信息。这样可以充分发挥网络流量统计的重要作用^[1],同时也可以简化表示层的页面复杂度,让程序员可以专心设计流量系统的业务逻辑而不必在代码维护方面花费太多的精力^[2]。目前已有的这方面的研究和实现,多是针对某一特定目的,或是局限于单个网站^[3-4],同时会受到采用的开发技术的制约^[5-6],其功能定位和应用效果不够理想。这些网站流量统计数据多半是存放在本地服务器端,不能和其他网站流量进行分析比对。目前所能看到的可做多网站的统计系统,也多是针对多个单一网站,只是实现了把单一网站的统计数据,放在一个数据库的一个表中,其目标用户还是分散的终端用户,关联性较小,不适宜同行业内、同系统内网站群或多网站的大规模使用。随着管理部门业务的开展,业务内容和服务项目的增多,限于单个网站的网络流量分析系统功能不完善,其功能的可扩展性和所分析的结果已远远满足不了行业(系统)对网站群或多网站网络流量分析、管理的需要,满足不了高并发、高访问量的业务支撑和拓展的需要。专门对网站群、多网站的网络流量综合分析系统的应用研究,尤其是行业(系统)网站群、跨系统(行业)网站的访问追踪、统计、分析研究的报道并不多见。

本文拟分析网站群或多网站的网络流量统计原理,并对网站群或多网站的流量统计分析进行研究,以期提供一个完善的解决方案。

1 系统运行原理及架构

1.1 设计原理及需求

1) 点击量。记录每 1 h 的 IP 数和 PV 数,提供各种形式供用户对任意时间段进行查询。IP 数完全基于 24 h IP 防刷。不同于其他系统的把所有点击量都加起来,这也是统计的重要参数之一。

2) 客户端。记录来访者所处的地区、访问者的浏览器、操作系统、语言、时区、屏幕尺寸、屏幕色彩、IP 地址及 Alexa 安装情况,并可对这些数据按任意时间段查询。

3) 流量源。记录点击来源,并根据来源对来路

域名及关键词和搜索引擎进行分析。可对来路信息按时间段和特征字查询,提供多种查询条件进行分析。

4) 关键词。精确的辨别并记录各大搜索引擎搜索进入时用户所搜索的关键词,兼容各种编码格式,无乱码,可按时间段和特征字查询分析,提供多种排序方式。

5) 被访页。记录用户进入网站时的网页被进入的次数(入口网址)和每个网页被浏览的次数。可按时间段和特征字查询,提供多种排序方式。

6) 明细。访问明细和在线用户栏目细致到用户的上网环境的全部信息,并可追踪任一用户的浏览记录。

7) SEO 数据。可以查看百度、谷歌、雅虎、搜搜、Bing、搜狗、有道等收录情况,方向链接情况以及 Alexa 排名和链接数。并且可以随时更新所有数据信息。这是本系统的一项新技术,是其他系统没有的功能模块。

1.2 系统框架的设计

系统体系结构按照符合系统管理员的操作习惯等特性进行设计。所有的模块都采用独立封装,方便系统扩展,提高系统的灵活性,减少开发成本。本系统分为系统管理端和网站用户管理端 2 个部分:系统管理端主要由系统配置、网站管理、账户管理、信息管理、单位管理几个模块构成;网站用户管理端主要由统计概况、SEO 数据、在线访问者、访问明细、流量分析、搜索引擎、访问流分析、访问者信息等几大模块构成,主要具体分析流量的分布情况。

1.3 数据库的设计

数据是网站的核心,所以数据库的设计是否合理,直接影响到整个程序的运行效率,本系统使用的数据库经过精心的设计和分析,达到关系模式的第三范式。系统以用户为中心,建立起所有关系,建立关系的宗旨是把数据冗余减到最小,每一个都是最小元素。共分为 8 个表:访问详细信息表(Details)、数据库字典表(Dic)、保存搜索引擎收录量表(Seo)、系统信息表(System)、用户表(User)、每 d 访问的统计表(ViewDay)、每 h 访问的统计表(ViewHours)、统计站点表(WebSites)。

2 关键技术实现

网站流量分析系统功能应包括:1)能及时了解网站流量和来访者信息;2)表示层页面直观,用户体验良好。能让程序员专心设计流量系统的业务逻

辑而不必在代码维护方面花费太多的精力.若要实现以上功能,基于网站群或多网站的网站流量分析系统实现必须解决好关键技术中的6个环节:

1)数据源的获取.当被统计网站安装好统计调用代码后,即可实时采集访问环境的数据,通过数据采集模块向服务器数据库写入.数据源的获取采用JS+CS结合获取.代码简单明了:

```
<script type="text/javascript" src="http://domian/js.aspx?sid=X&cid=Y&type=none"></script>
```

2)采用Jquery框架实现同步和异步的数据读写操作.与传统的设计相比,采用此技术可以同步或异步调用数据,也可以局部刷新,而不是刷新整个页面,及时生成页面数据以及图表.采用主流AJAX技术,实现异步JavaScript和XML多种技术的综合,包括JavaScript, XHTML, CSS, DOM, XMLXSL和XMLHttpRequest.其中使用XHTML和CSS作为标准化的呈现,使用XMLHttpRequest作为异步数据的读取,使用JavaScript绑定和处理所有数据.

3)采用伪JS技术,快速采集数据和生成图表,执行速度非常快,效率非常高.在1个页面里所有的数据只获取1次,获取完数据后立即关闭连接.获取分页数据的部分实现代码:

```
public static DataSet ExecuteDataset ( string connectionString, CommandType commandType, string commandText, int startRecord, int maxRecords, string srcTable, params SqlParameter[] commandParameters)
{
    if ( connectionString == "" || connectionString == null ) connectionString = SQL.connectionStr;
    if ( connectionString == null || connectionString.Length == 0 ) throw new ArgumentNullException ( " connectionString" );
    using ( SqlConnection connection = new SqlConnection ( connectionString ) )
    {
        connection.Open();
        return ExecuteDataset ( connection, commandType, commandText, startRecord, maxRecords, srcTable, commandParameters );
    }
}
```

4)采集的每项数据都经过分析和合理验证,能过滤掉一些没用的数据,有效数据才写入数据库,数据处理完毕就可直接生成页面数据并显示.这些技术可避免有些人通过一些软件或方法来刷流量,有别于其他同类应用产品——先把所有数据都写

入数据库后,再用人工分析去掉一些没用的信息.如在线统计可以精确到每1min的访问情况.其片段代码如下:

```
sp. TableName = "Haedu_Details";
sp. ReturnFields = " id, SiteId, ComeTime, Internet, Area, ComeUrl, ViewUrl, System, Language, EngineName, Keywords, Resolution, Browser, TimeZone, UserAgent ";
sp. PageIndex = this. CutePager1. CurrentPageIndex;
sp. PageSize = this. CutePager1. PageSize;
sp. Orderfld = " id ";
sp. Where = " where siteid = " + id + " and datediff ( n, ComeTime, getdate() ) < = " + sb. ToString();
```

5)按照主流的统计系统综合设计理念,采用左右结构设计,左边是导航菜单,右边是左边选择后显示的结果.结构清晰、操作简便、调用灵活.

6)所有统计数据由统一数据库管理,每个注册用户独享私有数据.成果的推广对象是所有互联网用户,每个注册用户都有自己的网站,并且是1个用户多个被统计的网站.所以在数据分析上,做出独立的详细分析,针对每一个被统计站点都有一个独立的数据,站点之间数据不会混乱,并且用户自能看到自己的统计站点,并对站点的流量进行分析.

3 实际效果展示

网站的访问量是实时变化的,图1是某被统计站点的一个时间段的IP和PV访问曲线图.可以看出从8:00到13:00访问人数稳定在250人左右,而人均访问页面数量是独立访问用户的2倍,说明每一个独立访问用户在访问网站时继续访问了该网站的另一个页面,如果IP和PV量一样的话,说明该网站不具有用户再向下一个页面访问的兴趣,这说明网站质量相对较差,弹出率太高.

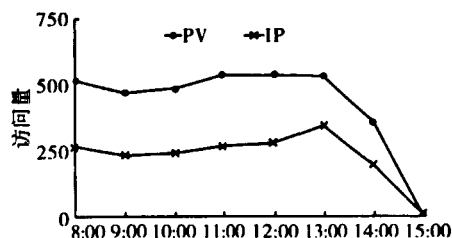


图1 访问量曲线图

用户的上网环境是需要获取的重要信息.表1为某网站来访用户的操作系统分布表.

从表1可以看出,来访者用户有大部分用户客户端操作系统为XP,现在市场上安装Windows7操作系统的用户明显增多.在考虑网站的程序等从

表1 某网站来访用户的操作系统分布表

操作系统	占来访者的比例/%
Windows XP	74.74
Windows 7	20.64
Windows Server 2003	0.06
Windows CE	0.07
IOS	0.15
Linux	0.41
Mac	0.42
Windows Vista	0.97
Other	2.48

技术设计上应多考虑 Windows XP 及 Windows 7 平台的兼容性。

对来访用户系统的高兼容性可提供统计数据。现在的浏览器有多个版本和种类,可以让注册用户根据来访者的来源,对自己的网站进行有针对性的测试,这也是网站需要改版的原因之一。

可客观分析网站的被收录量。这些数据可影响到网站的运营情况和决策调整情况,可以调整广告和目标人群的针对性。表2为某网站被搜索引擎检索的情况。

表2 某网站被搜索引擎检索分布表

搜索引擎名称	占来访者的比例/%
百度	90.26
有道	0.03
Bing	0.42
搜搜	0.61
搜狗	2.38
谷歌	6.30

从表2可以看出,网站的用户大多通过百度搜索引擎来访,这说明在各大搜索引擎中,网站的关键词在百度中占有排名优势。每个搜索引擎算法不一样,网站对于百度搜索引擎的友好度比较高。但这也反映出一个问题,直接输入网址的用户几乎没有,用户没能记住网站网址,这样会导致用户粘度不高,可能访问一次后就不再访问。但从广告投放的角度来讲,搜索引擎来访网站的比例也高,说明每次来访者都不一样,这种情况的出现相对商业广告投放客户来说可能针对性更强。

该设计完全实现了网站群或多网站的流量统计和分析,并形成一个完整的系统。实现了以下功能:1)网站访问量信息的统计和一般分析;2)网站访问量趋势分析;3)在可以获得数据的情况下,与竞争者进行对比分析;4)用户访问行为分析;5)网

站流量与网络营销策略关联分析;6)网站访问信息反映出的网站和网络营销策略的问题诊断;7)对网络营销策略的相关建议。真正解决了专业的网站访问统计报告并不只是对网站流量统计数据的简单汇总或列表,分析的内容也不仅仅局限于网站流量数据本身,而是通过对网站流量统计及其他相关数据的系统分析,来发现网站运营中可能存在的问题,并为有效解决这些问题提供决策依据。

4 结论

本研究成果通过三层设计模式(实体、数据访问、业务逻辑)对系统进行了分层架构,完成了对基于组件、模块式扩展与开发,从技术层面解决了网站群或多网站的网络流量分析问题。但随着被统计网站数量的不断增加,系统也将会面临并发、网络环境、硬件环境等方面的考验。目前该研究成果已成功应用到河南省高校网站群的统计分析系统中,系统运行稳定、统计准确、操作方便,得到了用户的好评。其成果可为互联网上任一网站提供统计服务,网站管理用户只需将相应的调用代码在被统计的网站中加载即可使用系统的所有功能,统计数据按网站管理用户分级集中管理,通过数据接口提供网站管理用户网站的调用和管理。设计上力求人性化、结构简单、条理清晰、操作简单,方便随时随地的查看分析报告,并在承载量、运行效率和稳定性上有了较大的提高,解决了局限于单个网站流量分析系统数据无法共享、无法实现多网站综合分析的难点,可为用户和管理部门修正或优化网络营销策略提供参考依据。

参考文献:

- [1] 李航,许春艳.基于ASP.NET Web用户控件站点流量统计系统的设计[J].科技传播,2010(9):127.
- [2] 魏高起,谢晓尧,高仲慧.基于JSP技术的流量统计系统设计与实现[J].贵州师范大学学报:自然科学版,2009;27(3):89.
- [3] 钟克吟.数字资源访问流量统计分析系统设计与应用[J].现代图书情报技术,2008(6):91.
- [4] 焦蓉梅.浅谈网站流量统计分析法[J].科技信息,2007(16):518.
- [5] 李朝阳,唐雨.一种基于Netflow的网络流量统计分析系统[J].计算机时代,2009(9):49.
- [6] 彭仁夔.JS源码分析与开发[M].北京:电子工业出版社,2010.