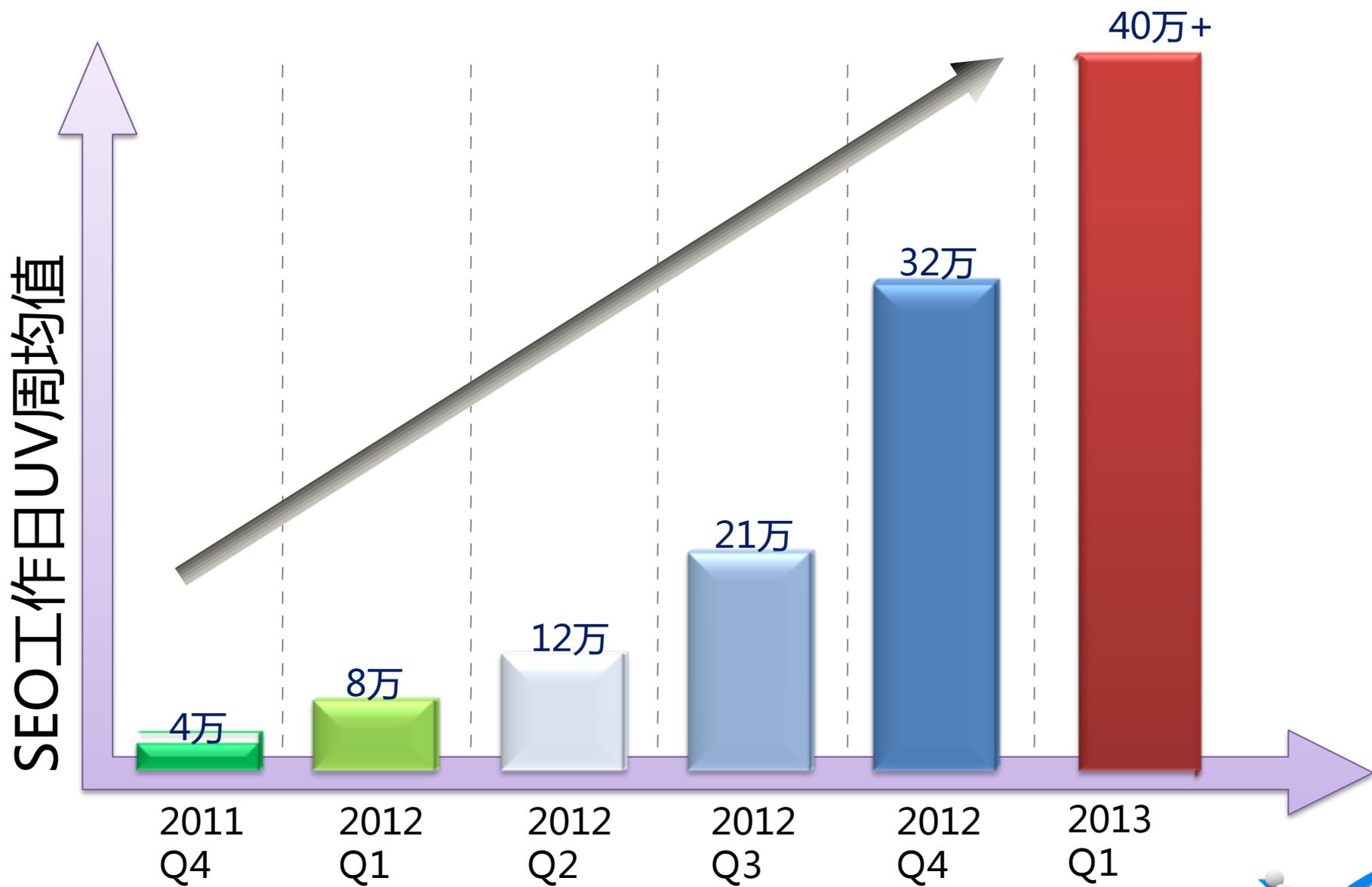


大型网站**SEO**优化实践



Agenda

1 大型网站SEO的特点&优势&挑战

2 Web应用SEO定位

3 大型网站SEO优化基本策略

4 全站SEO效果监控

5 do everything with tools

6 移动 wap seo 优化

7 SEO领域体系化建设

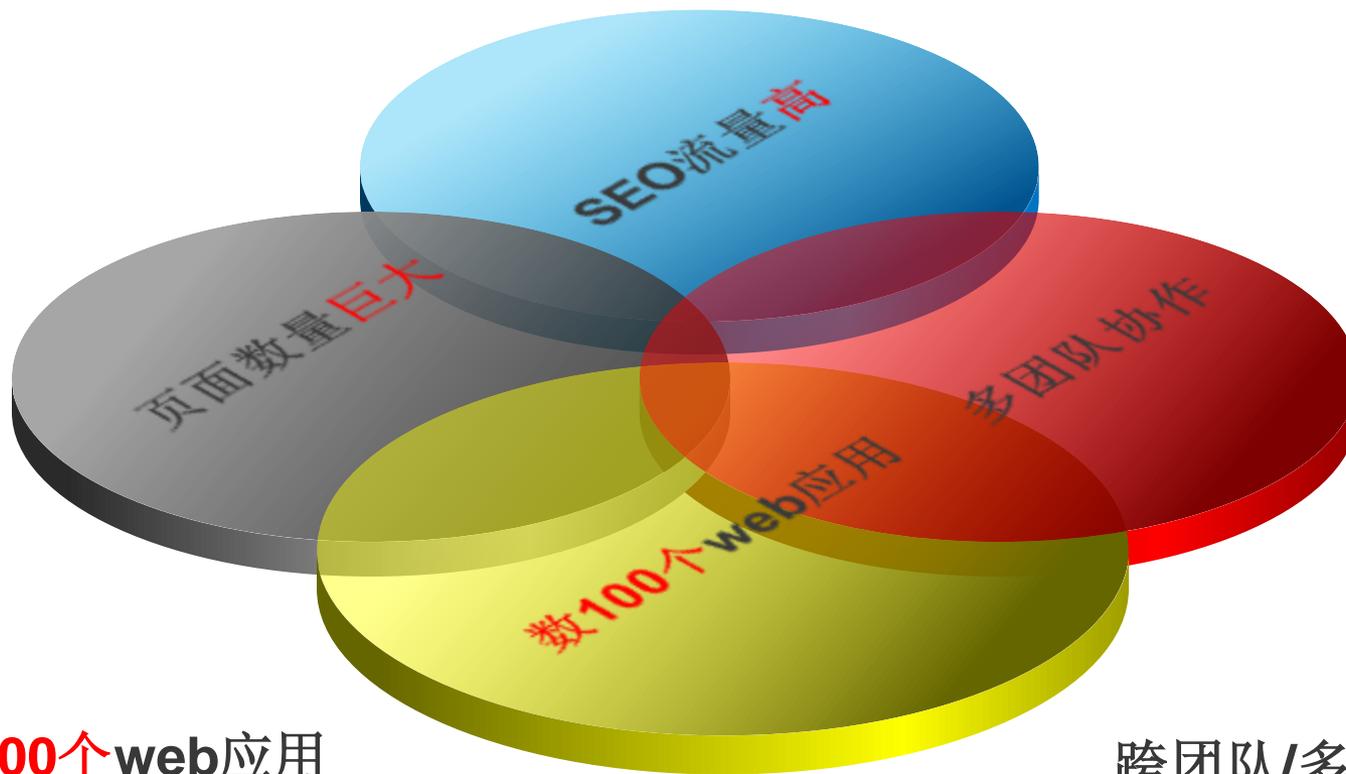
8 大公司SEO项目合作方式及挑战



大型网站SEO的特点

页面数量**巨大**

SEO流量**高**，
长尾流量占比**大**



数**100**个web应用

跨团队/多团队协作

大型网站SEO的优势&挑战

优势

- ✓ 页面内容丰富，数量多 (UGC)。
- ✓ 域名初始权重高 (天生属性)。
- ✓ 无需刻意进行link building，客户会主动帮忙进行链接建设和推广。
- ✓ 网站安全及性能优化由专业团队负责。

挑战

- ✓ 用户内容如何更好地匹配用户搜索query (如：博文，帖子)
- ✓ 如何更好地挖掘，整合用户内容，生成更有价值网页。
- ✓ SEO优化方案经常遭遇用户体验冲突，KPI冲突。

SEO需结合
产品定位及
产品核心价值

B2B的核心价值：批发，采购，产业带，商人社区。SEO需避免获取与产品核心价值无关的流量，否则南辕北辙。

SEO
流量预估

根据产品特点，产品页面数量级，按照同类产品索引量与SEO流量比值来预估流量预期。
如： 页面数*收录率：**SEO流量 = 200:1 ~ 10:1**

注重
品牌效应

某些应用应更注重品牌效应。 如:中国产业带 织里童装 织里童装产业带 南通家纺产业带

营销活动

保证与营销活动相关的关键词都能在百度被搜索到，并排名靠前。(不提供SEO流量预估) 如：312大促，919大促，919中国好货源。 **SEO流量具有时效性的。**

大型网站SEO优化基本策略

SEO技术研究

SEO元素周期表

优化初期三板斧

优化中后期应注重的
三点

SEO技术研究



SEO元素周期表

ON-THE-PAGE FACTORS

These elements are in the direct control of the publisher

CONTENT

Cq	QUALITY	Are pages well written & have substantial quality content?
Cr	RESEARCH	Have you researched the keywords people may use to find your content?
Cw	WORDS	Do pages use words & phrases you hope they'll be found for?
Ce	ENGAGE	Do visitors spend time reading or "bounce" away quickly?
Cf	FRESH	Are pages fresh & about "hot" topics?
Vt	THIN	Is content "thin" or "shallow" & lacking substance?
Va	ADS	Is your content ad-heavy, especially "above the fold"?

HTML

Ht	TITLES	Do HTML title tags contain keywords relevant to page topics?
Hd	DESCRIPTION	Do meta description tags describe what pages are about?
Hh	HEADERS	Do headlines & subheads use header tags with relevant keywords?
Hs	STRUCTURE	Do pages use structured data to enhance listings?
Vs	STUFFING	Do you excessively use words you want pages to be found for?
Vh	HIDDEN	Do colors or design "hide" words you want pages to be found for?

ARCHITECTURE

Ac	CRAWL	Can search engines easily "crawl" pages on site?
Ad	DUPLICATE	Does site manage duplicate content issues well?
As	SPEED	Does site load quickly?
Au	URLS	Are URLs short & contain meaningful keywords to page topics?
Vc	CLOAKING	Do you show search engines different pages than humans?

THE PERIODIC TABLE OF SEO SUCCESS FACTORS

Search engine optimization — SEO — seems like alchemy to the uninitiated. But there's a science to it. Below are some important "ranking factors" and best practices that can lead to success with both search engines and searchers.

ON-THE-PAGE SEO

OFF-THE-PAGE SEO

CONTENT	HTML	ARCHITECTURE	LINKS	TRUST	SOCIAL	PERSONAL
Cq ⁺³ Quality	Ht ⁺³ Titles	Ac ⁺³ Crawl	Lq ⁺³ Quality	Ta ⁺³ Authority	Sr ⁺² Reputation	Pc ⁺³ Country
Cr ⁺³ Research	Hd ⁺² Description	Ad ⁺² Duplicate	Lt ⁺² Text	Th ⁺¹ History	Ss ⁺¹ Shares	Pl ⁺³ Locality
Cw ⁺² Words	Hh ⁺¹ Headers	As ⁺¹ Speed	Ln ⁺¹ Numbers	Ti ⁺¹ Identity		Ph ⁺³ History
Ce ⁺² Engage	Hs ⁺¹ Structure	Au ⁺¹ URLs	Vp ⁻³ Paid	Vd ⁻¹ Piracy		Ps ⁺² Social
Cf ⁺² Fresh	Vs ⁻¹ Stuffing	Vc ⁻³ Cloaking	Vi ⁻² Spam			
Vt ⁻² Thin	Vh ⁻¹ Hidden					
Va ⁻¹ Ads						

FACTORS WORK TOGETHER

All factors on the table are important, but those marked 3 carry more weight than 1 or 2. No single factor guarantees top rankings or success, but having several favorable ones increases the odds. Negative "violation" factors shown in red harm your chances.

OFF-THE-PAGE FACTORS

Elements influenced by readers, visitors & other publishers

LINKS

Lq	QUALITY	Are links from trusted, quality or respected web sites?
Lt	TEXT	Do links pointing at pages use words you hope they'll be found for?
Ln	NUMBER	Do many links point at your web pages?
Vp	PAID	Have you purchased links in hopes of better rankings?
Vi	SPAM	Have you created many links by spamming blogs, forums or other places?

TRUST

Ta	AUTHORITY	Do links, shares & other factors make site a trusted authority?
Th	HISTORY	Has site or its domain been around a long time, operating in same way?
Ti	IDENTITY	Does site use means to verify its identity & that of authors?
Vd	PIRACY	Has site been flagged for hosting pirated content?

SOCIAL

Sr	REPUTATION	Do those respected on social networks share your content?
Ss	SHARES	Do many share your content on social networks?

PERSONAL

Pc	COUNTRY	What country is someone located in?
Pl	LOCALITY	What city or local area is someone located in?
Ph	HISTORY	Has someone regularly visited your site or socially favored it?
Ps	SOCIAL	Have your friends socially favored the site?

Written By: [Search Engine Land](#)

Design By: [COLUMN FIVE](#)

Learn More: <http://setmd.com/seotable>

Copyright Third Door Media

大型网站SEO优化初期三板斧

URL规范化

URL规范化是大型网站SEO优化的根本，解决得是搜索引擎网页库，索引库，URL库中的Key问题。URL规范化影响着网站核心资源被抓取及索引效率，权重等。

页面元素优化

- ✓ 页面元素优化决定着未来流量上升的高度。
- ✓ Title&Description，不同的场景，采用不同的模板。
- ✓ H,P,B,Strong标签，图片alt属性，页面无效链接添加rel="nofollow"等。

提升核心资源被抓取效率

- ✓ 优化初期应优先关注收录率，而非排名及流量。(70%比较合理)。
- ✓ 相比在老的URL结构+301+robot禁止策略+nofollow+noindex+cannonical组合策略来提升网站网站核心资源的收录效率，新URL规范化+robots禁止策略反而显得简单粗暴，但非常高效。

URL规范旧状及带来的问题与挑战

URL规范 旧状

- ✓ 多域名；
- ✓ 全静态化，全动态化，半静态化等多套URL规范。
- ✓ URL层级太深；
- ✓ URL解析代码复杂冗余，性能低，关键是维护成本高。

问题与 挑战

- ✓ 参数全静态化，本质还是动态化，SEO效果差。
- ✓ 跟前端工程师合作效率低。如：JS拼接静态URL问题。
- ✓ URL中特殊字符解析问题。如：%2F (/)，%4E (N)。
- ✓ form表单提交产生多余的301跳转，增加前端性能负担(20ms左右)。

URL规范化问题对策

搜索引擎
如何
面对？

- ✓ noindex, nofollow 标签
- ✓ 301 跳转
- ✓ canonical 标签
- ✓ webmaster 可设置动态参数处理方式

我们如
何解决？

How ?

如何提升网站核心资源被爬取效

■ 编入索引的网页总数 ?

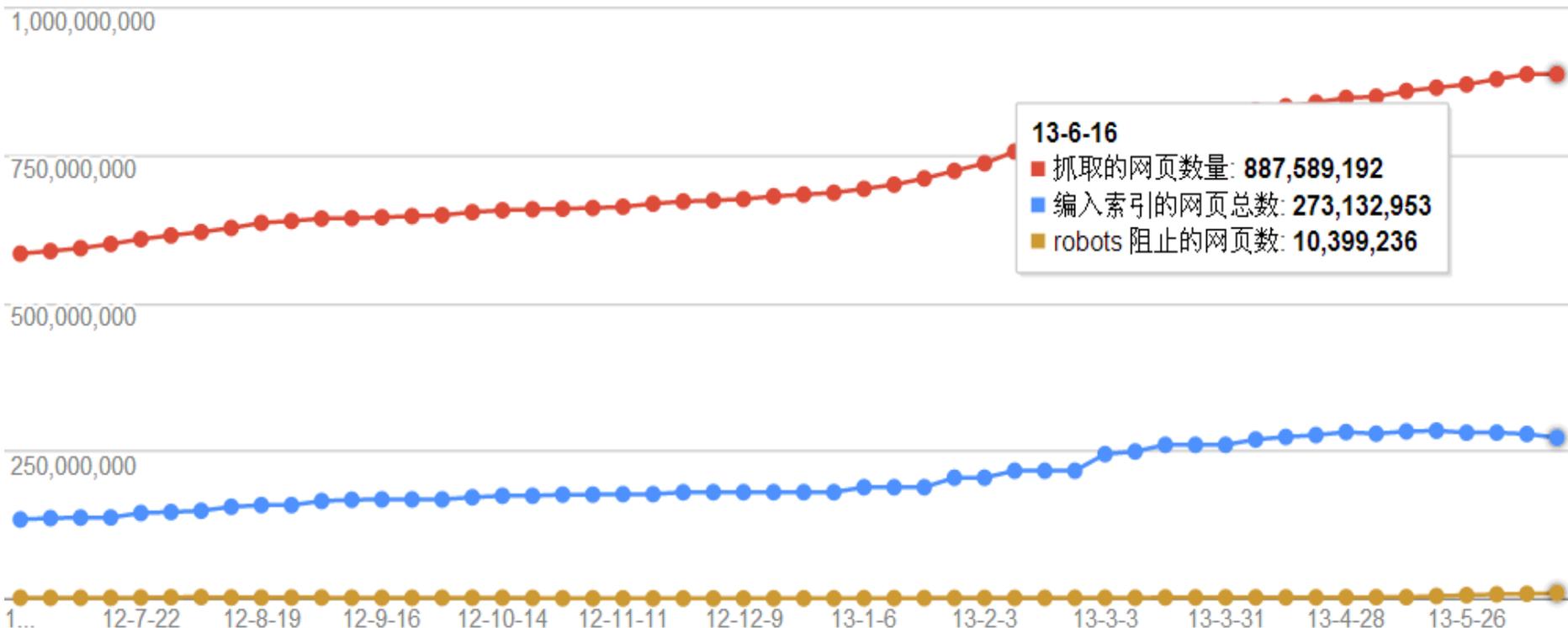
273,132,953

■ 抓取的网页数量 ?

887,589,192

■ robots 阻止的网页数 ?

10,399,236



如何提升网站核心资源被爬取效

引导爬虫 抓取网站 核心资源

- ✓ 站点扁平结构，实施比较难，重点优化面包屑。
- ✓ sitemap每日全量更新，补充spider URL库。
- ✓ 页面部分推荐模块异步改同步加载，核心资源URL部署到权重高的页面。

减少 无效URL 的抓取

- ✓ robotst.txt禁止无效URL被抓取及被索引
- ✓ 通过noindex meta标签禁止URL被索引 (百度暂不支持noindex)
- ✓ nofollow的使用

提升 网站抓取量

- ✓ 避免搜索引擎官方爬虫被安全软件拦截
- ✓ 逐步上调站长平台抓取压力
- ✓ 及时更新死链sitemap

大型网站SEO优化中后期应注重的三点

专注Query 分析

- ✓ Search Query = Search Requirement
- ✓ 用户搜索的query是否match页面Title中的关键词？
- ✓ 词库分库分档

调排名

准则：对排在第二及第三页的关键词，我们会重点分析及调优。优化第三页以后的关键词，投入产出比极低。

研究 anti-spam

- ✓ 坚持白帽，杜绝黑帽。(大型网站进行黑帽犹如饮鸩止渴)
- ✓ 研究anti-spam主要防止不正确优化被判作弊

全站SEO效果监控



日志分析脚本

- ✓ 定时分析
- ✓ Email分析报告
- ✓ 图形化报表
- ✓ 统计SEO基本指标数据
(分产品线, 各大搜索引擎)



SEO统计报表

- 傻瓜数据平台
- ✓ 全站SEO效果监控
 - ✓ SEO爬虫分析
 - ✓ SEO收录及反链分析
 - ✓ SEO基本指标统计



排名监控

- ✓ 核心关键词
核心应用每日排名
- ✓ 竞争对手排名监控

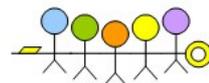


第三方数据

- 1 搜索引擎开放的监控数据
 - ✓ 百度站长平台
 - ✓ Google, Bing webmaster
- 2 量子统计

do everything with tools





想及时了解网站的异常信息? 百度网站异常提醒上线啦!

完善联系方式后可随时随地接收异常信息

马上进入

我的网站

站点管理

消息提醒

数据提交

网站配置

网站分析

网站安全

其他工具

站点管理 > 站点信息

站点信息

info.1688.com

重要的新消息

近期没有重要的新消息

[查看全部消息](#)

当前状态

抓取异常

[查看详情>>](#)

百度索引量

[查看详情>>](#)

外链分析

[查看详情>>](#)

错误类型

数量

日期

索引量

日期

外链总数

404页面

150

2013.06.24

28060

2013-06-13

3860647

控制台

重要的新消息

邮件

http://info.1688.com/: 服务可能出现过中断

Jun 9, 2013

[查看全部](#)

配置

运行状况

流量

优化

实验室

当前状态

抓取错误



搜索查询



站点地图

网站错误

DNS



服务器连接



robots.txt 抓取



网址错误

7 服务器错误

406 找不到

4,718 查询数

1,800,000 展示次数
40,000 点击次数

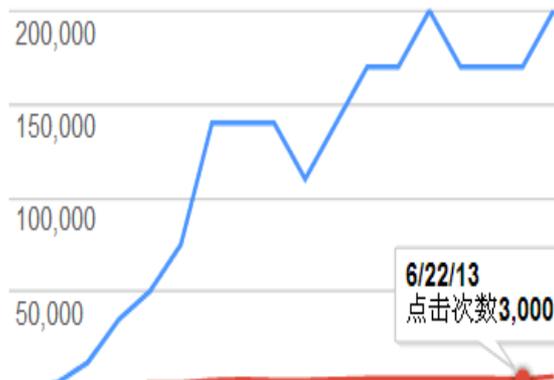
9,797,393 已提交的网:
7,710,448 已编入索引

200,000

150,000

100,000

50,000

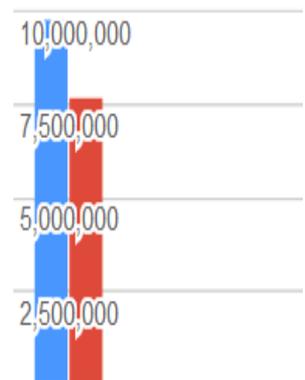


10,000,000

7,500,000

5,000,000

2,500,000



- 我的网站
- 仪表板
- ▶ 配置“我的网站”
- ▼ 报告与数据
 - 页面流量
 - 索引资源管理器**
 - 搜索关键字
 - SEO 报告
 - 入站链接
 - 抓取信息
 - 恶意软件
- ▶ 诊断和工具
- ▶ 消息
- 网站管理员 API

网站 ▼
info.1688.com

索引资源管理器

筛选器: [搜索登录页](#) | [301 次重定向](#) | [404 个错误](#) | [检测到恶意软件](#) | [已由 robots.txt 排除](#) | [重置筛选器](#) | [显示自定义选项](#) ▼

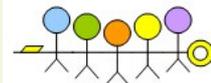
info.1688.com ▶

- detail
- tags_list
- news
- subject
- nlist
- wc-crossdomain.html?
- list
- opw
- biznews
- html
- ?
- subcategory

info.1688.com

包含 193,144 个文件夹/URL
搜索结果: 11,579
搜索点击: 163
入站链接计数: 1,912,883

- detail
- tags_list
- news
- subject
- nlist
- wc-crossdomain.html?
- list
- opw



问题及建议

未通过的规则

[查看页面源码](#)

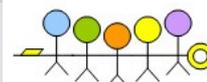
<p>1 H1标签检测</p> <p>严重程度：高</p>	<p>缺少h1标签。</p> <p>请在页面中设置h1标签。h1标签可增强标题、说明和正文副本中的核心关键字。我们建议：</p> <p>(1) h1标签包含的是内容页标题或者可反映页面内容的关键字。</p> <p>(2) 关键字的长度不能超过约 150 个字符。</p> <p>(3) 页面中必须存在，且只能存在唯一的h1标签。</p> <p>更多信息请参考文档：阿里巴巴SEO优化指南</p>	✘
<p>2 图片Alt信息检测</p> <p>严重程度：中</p>	<p>图片缺少说明信息：页面img标签缺乏alt属性或alt属性为空。</p> <p>请对页面每张图片链接都尽量添加Alt属性。Alt属性中的内容主要是对图片内容的说明解释，内容必须跟图片主题相关，可适当包含关键字。这不仅能提升页面关键字的相关性，还能提升页面上的图片被收录的几率以及图片在各大搜索引擎图片频道排名情况。但请不要在ALT属性中堆积关键词。如何更好更合理地设置Alt属性，请参考文档：阿里巴巴SEO优化指南</p> <p>查看问题代码</p>	✘
<p>3 Frame信息检测</p> <p>严重程度：低</p>	<p>检测到页面中存在frame/frameset/iframe标签。</p> <p>页面中使用frame/frameset/iframe标签，会导致百度spider的抓取困难，如无特殊需求，建议您尽量不要使用frame/frameset/iframe等Frame。</p>	✘
<p>4 robots.txt文件检测</p> <p>严重程度：高</p>	<p>检测到该域名下robots.txt文件不存在。</p> <p>域名下robots.txt文件不存在，则允许爬虫抓取网站上任何URL，这将会导致网站的核心资源抓取效率及被索引效率过低。请根据网站自身情况，正确编写robots.txt文件，禁止无效URL的抓取。更多信息请参考文档：阿里巴巴SEO优化指南</p>	✘

已通过的规则

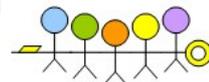
<p>1 URL规则检测</p> <p>重要程度：高</p>	<p>已符合程序对 URL的基本检测，基本符合我们的SEO规范。</p> <p>如需要了解更多的信息，请参考：</p> <p>(1) 来自百度的建议:什么是好的URL? (2) google提供的规范化建议 (3) 阿里巴巴SEO优化指南</p>	✔
---------------------------------------	--	---

保存

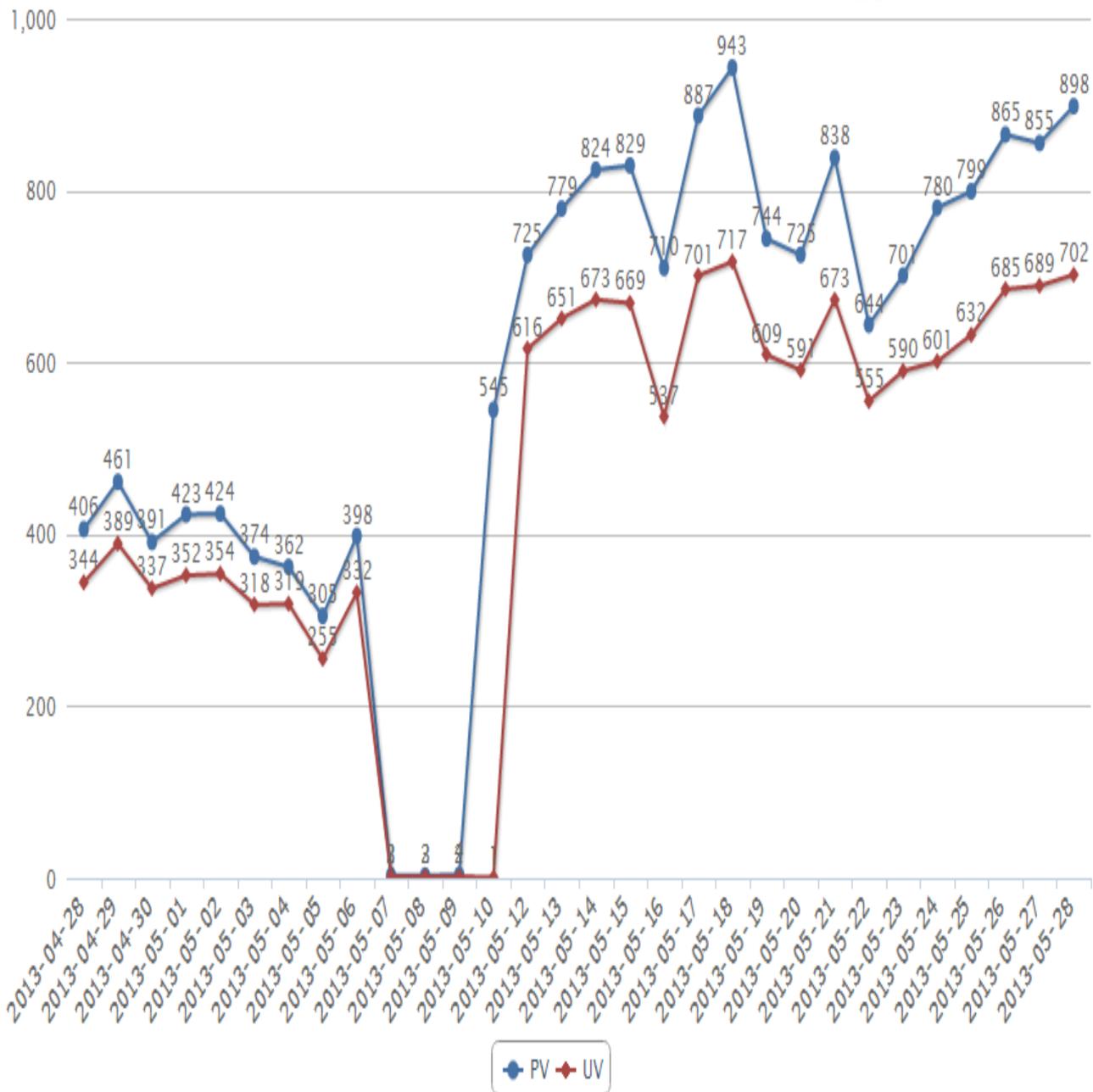
查找



类目ID	关键词	标题	摘要	操作	百度抓取日期 百度是否更新 百度更新日期
	救灾物资 (流量)	救灾物资_救灾物资价格_优质救灾物资批发/采购 - 阿里巴巴		编辑 / 删除 已生效 2013-05-13 13:21:10.0 steven.cheng	2013-05-28 未更新 暂无
	利趣拿铁 (流量)	利趣拿铁_利趣拿铁价格_优质利趣拿铁批发/采购 - 阿里巴巴		编辑 / 删除 已生效 2013-05-13 00:00:00.0 steven.cheng	2013-05-28 未更新 暂无
	移动电源 (流量)	#_品牌移动电源_移动电源批发_移动电源代理 - 阿里巴巴	阿里巴巴为您搜索到了@条#产品的最新批发信息, 其中包括#产品的配置、批发价格、供应商联系方式、采购渠道等。您还可在阿里巴巴上找到更多#周边产品的信息。	编辑 / 删除 已生效 2013-03-05 17:00:01.0 3cshuma	2013-05-28 已更新 2013-02-17
	九阳豆浆机 (流量)	#_#最低价_#批发/串货-全国- 阿里巴巴	阿里巴巴为您找到@款#, 2013最新款#批发, 找海量优质#卖家, 厂家直销#货源, 上阿里巴巴批发市场。\$	编辑 / 删除 已生效 2013-03-05 17:00:01.0 jiaju	2013-05-22 未更新 暂无
	美的 (流量)	#_#最低价_#批发/串货-全国- 阿里巴巴	阿里巴巴为您找到@款#, 2013最新款#批发, 找海量优质#卖家, 厂家直销#货源, 上阿里巴巴批发市场。\$	编辑 / 删除 已生效 2013-03-05 17:00:01.0 jiaju	2013-05-15 未更新 暂无
	对联2013批发 (流量)	#_对联2013批发价格_最火对联2013批发_厂家直销_优质对联2013批发 - 阿里巴巴	阿里巴巴厂家直销对联2013批发, 阿里巴巴为您精心挑选了几千种对联, 对联2013批发, 几百家供应商任你挑选, 可混批, 春节用品一站备齐, 进货好去处。\$	编辑 / 删除 已生效 2013-03-05 17:00:01.0 3cshuma	2013-05-28 未更新 暂无
	十字绣 (流量)	#_十字绣价格_热销十字绣厂家直销热卖厂家价_源头十字绣批发底价 - 阿里巴巴	阿里巴巴厂家直销对联批发, 阿里巴巴为您精心挑选了几千种对联春联, 对联批发, 几百家供应商任你挑选, 厂家直销, 实地认证, 可混批, 春节用品一站备齐, 进货好去处。\$	编辑 / 删除 已生效 2013-03-05 17:00:01.0 3cshuma	2013-05-28 未更新 暂无
	泰国佛牌 (流量)	#_泰国佛牌价格_泰国佛牌_泰国佛牌好货源 - 阿里巴巴	阿里巴巴上有您想要的泰国佛牌, 海量批发信息, 挑选余地广, 价格全国最低, 工艺精美, 让你真正做到批发无忧。\$	编辑 / 删除 已生效 2013-03-05 17:00:01.0 3cshuma	2013-05-28 未更新 暂无

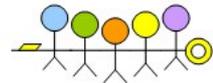


弓箭



查找

日期	关键词	UV	趋势 / 搜百度 / 搜倾国
2013-06-17	弓箭	577	趋势 / 搜百度 / 搜倾国
2013-06-17	竹篮	253	趋势 / 搜百度 / 搜倾国
2013-06-17	广场舞舞蹈服装	248	趋势 / 搜百度 / 搜倾国
2013-06-17	3d打印机	203	趋势 / 搜百度 / 搜倾国
2013-06-17	手枪	190	趋势 / 搜百度 / 搜倾国
2013-06-17	联想g400笔记本	162	趋势 / 搜百度 / 搜倾国
2013-06-17	四轮电动车	162	趋势 / 搜百度 / 搜倾国
2013-06-17	宗申三轮摩托车	153	趋势 / 搜百度 / 搜倾国
2013-06-17	pvc板	138	趋势 / 搜百度 / 搜倾国
2013-06-17	久保田688收割机	136	趋势 / 搜百度 / 搜倾国



移动 wap seo 优化

手机搜索 2012Q3 市场监测数据

2012Q3 中国手机搜索市场用户搜索量分布状况调查



艾媒咨询 (iiMedia Research) 数据显示, 2012Q3 中国手机搜索用户搜索量占比分布方面, 百度上升明显, 以 37.3% 占比位列第一, 360 占比达到 21.7%, 较上季度略有下降。腾讯搜搜和搜狗较上一季度占比有所提升, 分别达到 20.9% 和 6.9%。谷歌占比明显下滑, 仅有 4.1%。

艾媒咨询 (iiMedia Research) 分析认为, 百度品牌借助 PC 互联网用户搜索习惯及近期不断发力移动搜索市场, 仍然有很大机会领跑移动搜索市场。但面对 360、搜搜等强有力的竞争对手, 百度移动搜索发展之路仍具风险。

艾媒情报库: www.iimedia.cn

本报告于 2012-12 由艾媒咨询提供



移动 wap seo 优化

自主适配

在apache服务器上根据user-agent 自主适配到wap页

标注Meta
声明

```
<meta name="mobile-agent" content="format=html5;url=http://m.1688.com">  
<meta name="mobile-agent" content="format=xhtml;url=http://m.1688.com/std/">
```

开放适配
协议

百度站长平台提交开放适配协议，如适配协议与页面标注Meta声明冲突，以适配协议为准。

DOCTYPE
声明

DOCTYPE声明有助于搜索引擎识别该页面是否适合手机浏览。wap页协议：xhtml协议/wml协议/HTML5协议

移动标准
sitemap

google有特定的移动标准sitemap

SEO 领域体系化建设



大公司SEO项目合作方式及挑战



挑战：

- 1 SEO项目见效周期相对较长，如何保持耐心？
- 2 SEO优化项目考核目标是否一致？
- 3 SEO优化与用户体验冲突时，该如何权衡？



THANK YOU!