CHAPTER 5

What Did They Do?: Web Analytics

Think of two web pages you've created in the past. Which of them will work best?

Before you answer, you should know that you're not entitled to an opinion. The only people who *are* entitled to one are your visitors, and they convey their opinions through what they do on your site. If they do the things you want them to, in increasingly large numbers, that's a good sign; if they don't, that's a bad one.

Analytics is that simple. Everything else is just details.

Before you go further in this chapter, we want you to try an exercise. Find a volunteer, and ask him or her to think of something he or she wants. Then draw a picture on a piece of paper, and get him or her to reply "warmer" or "colder" depending on how close the thing you drew was to the thing he or she was thinking. Repeat until you know what they're thinking of.

Go ahead. We'll wait.

How long did it take for you to figure out what he or she wanted? How many iterations did you go through?

Notice that you had plenty of creative input into the process: you came up with the ideas of what might work, and you deduced what your volunteer wanted through small "improvements" to your picture based on their feedback.

The same process takes place as you optimize a website, with some important differences:

- You're listening to hundreds of visitors rather than a single volunteer, so you have to measure things in the aggregate.
- What you define as "warmer" or "colder" will depend on your business model.
- If you know your audience well, or have done your research, your initial picture will be close to what your web audience wants.
- You can ask visitors what they think they want, through surveys, and they may even tell you.

Your visitors are telling you what they want. You just have to know how to listen by watching what they do on your site, and react accordingly. Web analytics is how you listen. Use it wisely, and eventually the website you present to visitors will be the one they were craving.

For a long time, web traffic analysis was the domain of technologists. Marketers were slow to recognize the power of tracking users and tying online activity to business outcomes. Web activity was mainly used for error detection and capacity planning.

The growth of the Web as a mainstream channel for commerce, and the emergence of hosted analytics services—from Omniture, Urchin, WebTrends, CoreMetrics, and others—finally convinced marketing teams to get involved. Today, free services like Google Analytics mean that everyone with a website can get an idea of what visitors did.

There's no excuse not to listen.

Dealing with Popularity and Distance

Imagine that you own a small store in the country. You don't need much market analysis. You can see what people are doing. You know their individual buying preferences, their names, and their browsing habits. You can stock their favorite items, predicting what they'll buy with surprising accuracy.

Now suppose business picks up and the number of customers grows. You slowly lose the ability to keep track of it all. Clients want more merchandise, which in turn requires more floor space. You have to make compromises. The sheer volume of customers makes it impossible to recognize them all. You lose track of their buying habits. Your customers become anonymous, unrecognizable. You have to start dealing in patterns, trends, and segments. *You need to generalize*.

Websites face similar issues. The Web pushes both visitor anonymity and traffic volume to the extreme. You can't know your visitors—in fact, you can't even tell if they're male or female, or how old they are, the way you can in a store. You're also opening the floodgates, letting in millions of complete strangers without any limits on their geographic origins or the times of their visits.

We're going to provide a basic overview of analytics. Analytics is the cornerstone of a complete web monitoring strategy, and many other sources of monitoring—from performance monitoring, to customer surveys, to usability, to community management—all build upon it. Only by consolidating analytics data with these other information sources can you form a truly complete picture. Only through an integrated view can you return to the confidence of someone who's intimately familiar with her market and customers.

There are many excellent books and websites that deal with analytics in much greater depth than we'll attempt—in particular, with the need to focus on outcomes, to experiment, and to focus on data and facts rather than on opinions and intuition. We can recommend Jim Sterne's *Web Metrics: Proven Methods for Measuring Web Site Success* (Wiley), Avinash Kaushik's *Web Analytics: An Hour a Day* (Sybex), and Eric Peterson's *Web Analytics Demystified* (Celilo Group Media). You can find Avinash Kaushik's blog at *www.kaushik.net/avinash*.

The Core of Web Visibility

Web analytics is the piece of a web visibility strategy that's most tightly linked to the business outcome of your website. It captures your users' sessions, segments them in meaningful ways, and shows you how their visits contributed to your business. All of your other monitoring must tie back to it:

- Web performance and availability ensures that visitors can do what they want, when they want.
- Surveys ensure that you understand visitors' needs and hear their voices.
- Usability and interaction analysis measures how easily visitors can achieve the goals you've set for them.
- Community monitoring links what visitors do elsewhere to your site and your brand.

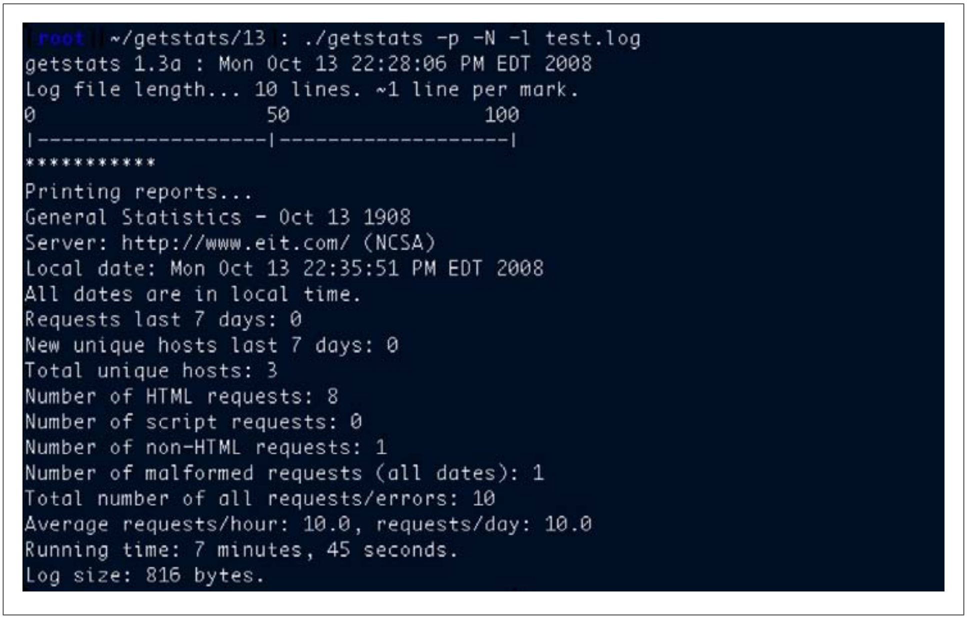
If you're not making decisions about your web presence based on what your analytics tells you, you're making bad decisions. If you're not augmenting analytics with other data, you're making decisions without all the facts.

A Quick History of Analytics

For as long as servers have existed, they've generated logfiles. Early on, these logs were just another source of diagnostic data for someone in IT. Each time a server handled a request, it wrote a single line of text to disk. This line contained only a few details about the request, and it followed the Common Log Format, or CLF. It included information about the user (where she connected from) and about the request (the date and time that it occurred, the request itself, the returned HTTP status code, and the byte length of the document or page transferred).

It was only in the mid-1990s that information such as user agents (the browser) and referrer (where a user came from) was added to logfiles. A slightly more detailed version of HTTP records, known as Extended Log Format (ELF), followed in early 1996. ELF added more client and server information.

ELF gave many companies their first glimpse of what was happening on their websites. Web logs were sparse, and focused on the technical side of the web server: which objects were requested, which clients requested them, when they were retrieved, and the HTTP status codes in response to those requests.



```
root  ~/getstats/13 : ./getstats -p -N -l test.log
getstats 1.3a : Mon Oct 13 22:28:06 PM EDT 2008
Log file length... 10 lines. ~1 line per mark.
0                   50                100
|-------------------|-------------------|
***********
Printing reports...
General Statistics - Oct 13 1908
Server: http://www.eit.com/ (NCSA)
Local date: Mon Oct 13 22:35:51 PM EDT 2008
All dates are in local time.
Requests last 7 days: 0
New unique hosts last 7 days: 0
Total unique hosts: 3
Number of HTML requests: 8
Number of script requests: 0
Number of non-HTML requests: 1
Number of malformed requests (all dates): 1
Total number of all requests/errors: 10
Average requests/hour: 10.0, requests/day: 10.0
Running time: 7 minutes, 45 seconds.
Log size: 816 bytes.
```

Figure 5-1. GetStats v1.3 output, coded in 1993

At first, web operators parsed these logfiles to find problems, searching for a specific error such as a "404 not found," which indicated a missing file. They quickly realized, however, that they also wanted aggregate data from the logs, such as how many requests the servers had handled that day.

So coders like Kevin Hughes developed applications like GetStats, shown in Figure 5-1, that would "crunch" logs and display the results in a more consumable format.

What Was the First Analytics Tool?

GetStats wasn't the first web server log analysis tool, but it was very influential in terms of the way the data was presented and summarized.

Roy Fielding with wwwstat was the first, as far as I can recall, to present statistics in an easy-to-read paragraph summary form that I think was written in Perl. I also took ideas from Thomas Boutell (wusage) and Eric Katz (WebReport).

—Kevin Hughes,
author of GetStats

Before web analytics became interesting to marketers, however, several things had to happen:

- The Web had to become *mainstream* enough for marketers to care, requiring both a large number of connected consumers and a rich visual experience within web browsers. Clearly, this has happened: there were only 38 million Internet users in 1994, but roughly 1.5 billion by January 2009—a 40-fold increase (*http://www.internetworldstats.com/stats.htm*).
- Analytics had to become *visitor-centric*. To be useful for business, logging had to move from individual requests for pages to user visits so that something a user did on page A could be linked to a purchase on page F. Cookies made this possible and added unprecedented accountability to promotional campaigns.
- Analysts needed ways to *segment* visitors so they could decide which browsers, campaigns, promotions, countries, or referring sites were producing the best business results, and optimize their websites accordingly. Better logging and access to browser data offered good segmentation, which meant that analysts could act on what they saw through experimentation.

By the mid-1990s, established brands were launching their web presence. Secure Sockets Layer (SSL) made it safe to conduct transactions, and companies like Pizza Hut, CDNow, and Amazon were selling real things to real customers through web interfaces. The audiences were there, too: the 1998 World Cup website france98.com served, on average, 180 requests *a second* between April 30 and July 26, for a total of 1.3 *billion* requests.

Web analytics companies like Accrue, NetGenesis, and WebTrends started to process web logs in ways marketing, rather than IT, wanted. This was big-ticket enterprise software, consisting of large up-front licenses and powerful servers on which to run it. This software got its data from logfiles or by sniffing web hits directly from the Internet connection, which meant that IT still had to be involved in the deployment process, and that the company had to maintain servers and storage to collect and analyze all of the data.

From IT to Marketing

Technologists were less and less the audience for the data produced by these tools. Instead, their features focused on marketers eager to embrace their online customers who were clamoring for more insight into what was happening online.

In many companies, something important changes when marketing departments become the customer. Operations is traditionally a cost center—the tools it uses are seen as ways to minimize costs—but the marketing department's tools are about maximizing revenue. Marketing makes it cool.

Three important changes placed analytics firmly in the hands of marketing.

JavaScript collection let marketing bypass IT

As marketers took over web analytics, they wanted ways to monitor their websites without having to ask IT for permission to do so. JavaScript made it possible to collect visitor activity without installing servers or parsing web logs. Marketers were free from the tyranny of IT operators.

In 1996, San Diego-based company Urchin launched a hosted analytics service called Quantified. At the same time, a company in Salt Lake City called Omniture introduced a similar analytics tool, SiteCatalyst, which was focused on large enterprises.

Search engines changed the reports marketers wanted

That year, at Stanford University, two students were working on an algorithm for measuring the importance of pages that would eventually change the way the world found information. Google—and other search engines and directories—spawned an ecosystem of advertisers, buyers, and analysts by selling advertising space alongside search results. The result was a thriving online advertising industry that was cheaper and more accountable than its offline counterpart.

Online ads could be changed at a moment's notice—after all, they were just information. They were also completely trackable. Consequently, marketers could test and adjust campaigns, bidding on the search terms that were most likely to drive buyers to their sites. This changed the kinds of reports analysts wanted. Suddenly, it was all about which campaigns and referring sites were most likely to attract visitors that would complete transactions.

Service models meant pay-as-you-go economics

Third-party analytics services enjoyed better economies of scale than individual enterprises could achieve on their own. An analytics service had enough capacity to handle traffic spikes to one individual site, amortizing the analytics processing across machines shared by all customers. Marketers paid for the traffic they measured, not the up-front cost of servers and software.

Now that analytics belonged to the marketing department, billing was tied to revenue rather than cost. Analytics was paid for as a percentage of web revenues. If traffic increased, a company's monthly analytics bill increased—but so (hopefully) did its revenues.

Around this time, the web analytics industry became much more mature, urged on by people like Jim Sterne (who launched the web analytics conference "eMetrics"), Matt Cutler, Jim Novo, and Bryan Eisenberg (who founded the Web Analytics Association). Other industry thought leaders would later emerge, such as Eric T. Peterson and Avinash Kaushik. We also saw the first steps toward standardization of terms, metrics, and reports.

A Very Short Analytics Glossary

Here's the shortest overview of analytics terms we could come up with. It shows you most of the key terms we'll be looking at in the pages ahead.

A *visitor* arrives at your website, possibly after following a link that *referred* her. She will *land* on a web page, and either *bounce* (leave immediately) or request additional pages.

In time, she may complete a *transaction* that's good for your business, thereby *converting* from a mere visitor into something more—a customer, a user, a member, or a contributor—depending on the kind of site you're running. On the other hand, she may *abandon* that transaction and ultimately exit the website.

Visitors have many external attributes (the browsers they're using, the locations they're surfing from) that let you group them into *segments*. They may also see different *offers* or pages during their visits, which are the basis for further segmentation.

The goal of analytics, then, is to maximize conversions by *optimizing* your website, often by *experimenting* with different content, layout, and campaigns, and analyzing the results of those experiments on various internal and external segments.

From Hits to Pages: Tracking Reach

While early web analytics reports simply counted HTTP requests, or *hits*, marketers quickly learned that hits were misleading for several important reasons:

- The number of hits varies by page. Some pages have dozens of objects, while others have only one object.
- The number of hits varies by visitor. First-time visitors to a site won't have any of the graphical content cached on their browsers, so they may generate more hits.
- It's hard to translate hits to pages. Pages may have JavaScript activity that triggers additional requests. A Google Maps visit, for example, triggers hundreds of hits on a single page as a user drags the map around.

Hits: A Definition

Any server request is considered a hit. For example, when a visitor calls up a web page containing six images, the browser generates seven hits—one for the page and six for the images.

Prior to sophisticated web analytics, this legacy term was used to measure the interaction of the user with the website. Today, hits might be a useful indicator of server load, but are not considered useful for understanding visitor behavior.

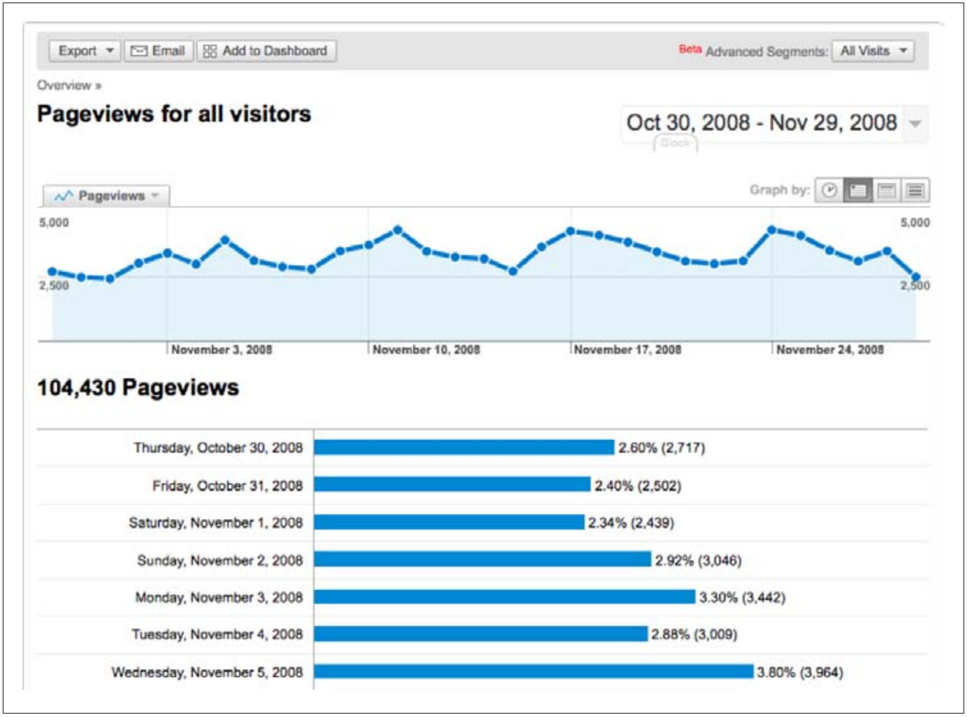

Export
Email
Add to Dashboard
Beta Advanced Segments: All Visits
Overview »
Pageviews for all visitors
Oct 30, 2008 - Nov 29, 2008
Pageviews
Graph by:
5,000
2,500
5,000
2,500
November 3, 2008
November 10, 2008
November 17, 2008
November 24, 2008
104,430 Pageviews
Thursday, October 30, 2008 2.60% (2,717)
Friday, October 31, 2008 2.40% (2,502)
Saturday, November 1, 2008 2.34% (2,439)
Sunday, November 2, 2008 2.92% (3,046)
Monday, November 3, 2008 3.30% (3,442)
Tuesday, November 4, 2008 2.88% (3,009)
Wednesday, November 5, 2008 3.80% (3,964)

Figure 5-2. A Pageview report in Google Analytics

The first useful web analytics metric is the *page view*, which is a measure of how many pages visitors viewed. A page view report like the one in Figure 5-2 shows you how many times a visitor saw a page of your site.

Page views are still misleading. If your site serves 100 pages, you don't know whether the traffic resulted from a single visitor reading a lot of content or a hundred visitors reading one page each. To deal with this, web analytics started to look at user visits with metrics like unique page views, which ignored repeated page views by the same visitor.

From Pages to Visits: The Rise of the Cookie

One of the main ways in which websites distinguish individual visitors from one another is through the use of cookies—small strings of text that are stored on the browser between visits. Cookies are also a major source of privacy concern because they can be used to identify a visitor across visits, and sometimes across websites, without the visitor's explicit approval.

In the very early days of the Web, each request for content stood alone—two requests from the same visitor weren't related. If a user asked for a page about shoes and then moved on to a page about shirts, those appeared to be completely independent requests.

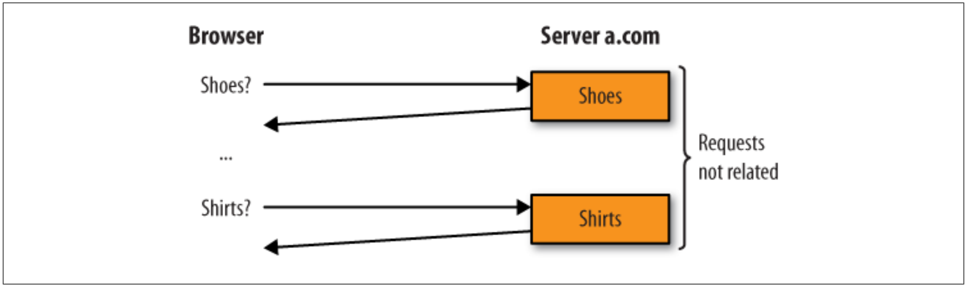


Figure 5-3. A website visit to a server in the early days

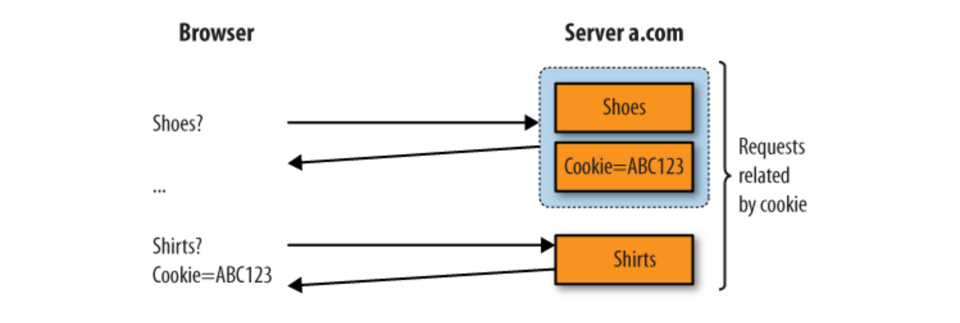


Figure 5-4. A website visit to a server that sends a cookie

This severely limited the usefulness of web applications. You couldn't, for example, put items in a shopping cart; each item looked like an entirely new request, as illustrated in Figure 5-3.

To address this issue, websites started sending cookies to browsers. A *cookie* is a unique text string that a web server sends to the browser when a user first visits. Each time the user returns or requests another page on the site (with the same browser), it passes the cookie to the server, as shown in Figure 5-4. Because of this, *cookies let website operators know which pages they've shown to users in the past*.

A cookie contains long strings of data that aren't supposed to be personally identifiable. (we're using the phrase "supposed to be" because it is technically possible for a website operator to place personally identifiable information inside a cookie—and it happens surprisingly often). In fact, cookies are often encrypted, changed, and updated by the server during a visit. A site cookie also isn't shared: the browser only sends the cookie to the site it got it from in the first place. This means that a cookie can't be used to track a user across several web properties. Figure 5-5 shows the use of a second cookie, this time on site b.com, that is completely unrelated to the cookie sent by site a.com.

Or at least that was the plan.

In Figures 5-4 and 5-5, visitors received two cookies, one from each visit. Consider what happens, however, if those sites serve advertisements from a shared ad network, as

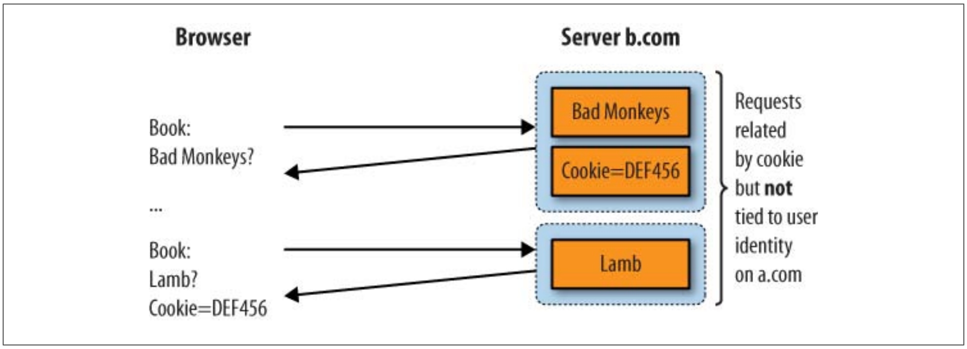


Figure 5-5. The visitor in Figure 5-4 visits a bookstore (b.com) that's unrelated to the previous server, and gets a new—and different—cookie

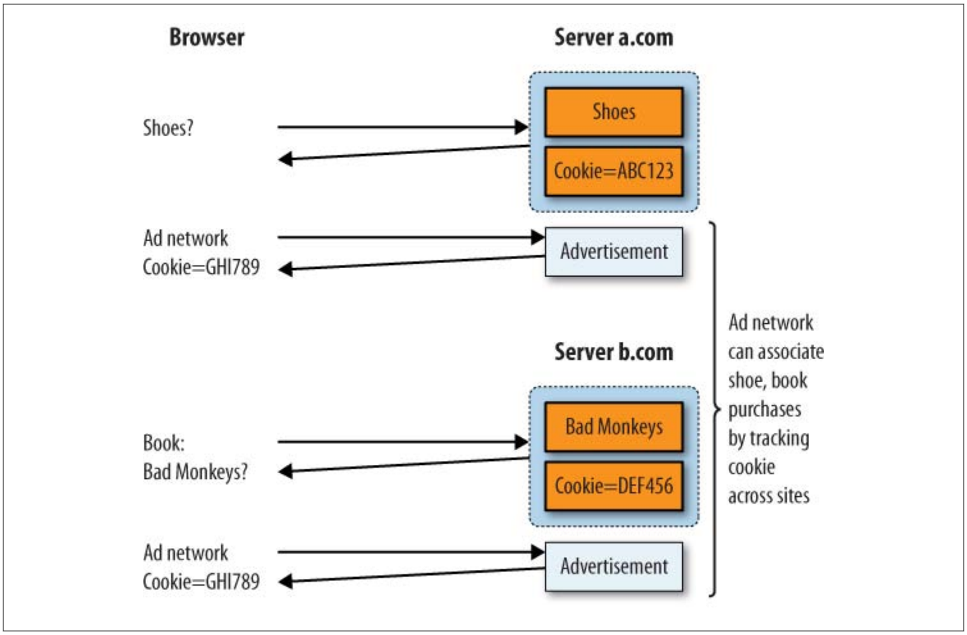
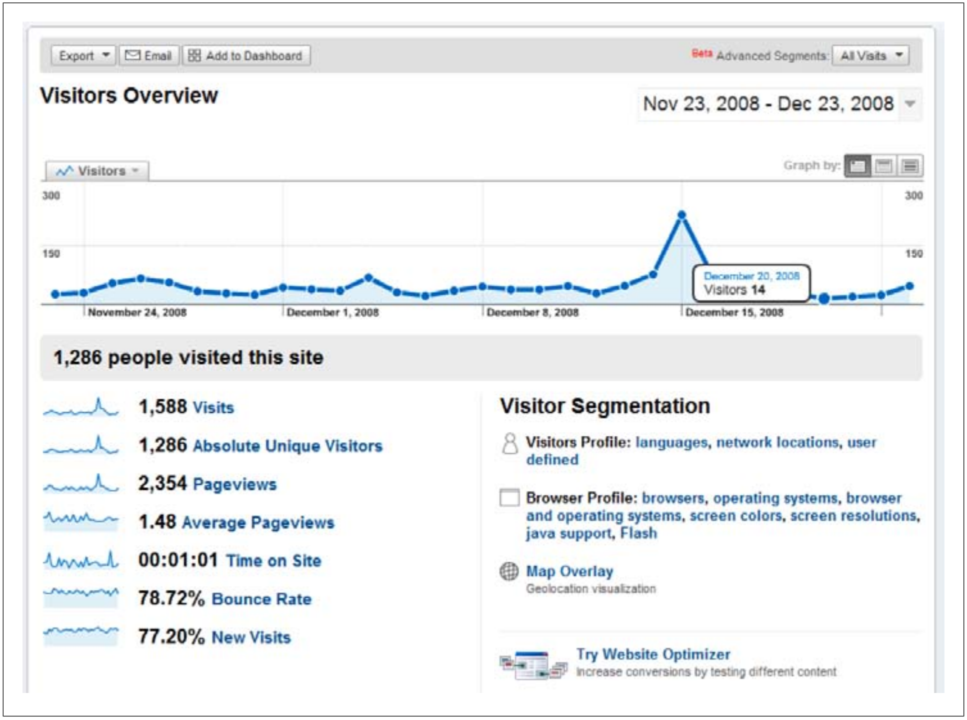


Figure 5-6. Cookies sent by a common ad network

shown in Figure 5-6. When a user retrieves an ad on a.com, then moves to b.com (whose ads are being provided by the same ad network), the ad provider now knows that the two visits are related.

These "tracking cookies" let an advertiser track an individual across many sites, and some consider them a violation of privacy.



Visitors Overview

Nov 23, 2008 - Dec 23, 2008

1,286 people visited this site

1,588 Visits

1,286 Absolute Unique Visitors

2,354 Pageviews

1.48 Average Pageviews

00:01:01 Time on Site

78.72% Bounce Rate

77.20% New Visits

Visitor Segmentation

Visitors Profile: languages, network locations, user defined

Browser Profile: browsers, operating systems, browser and operating systems, screen colors, screen resolutions, java support, Flash

Map Overlay
Geolocation visualization

Try Website Optimizer
Increase conversions by testing different content

Figure 5-7. An overview of site visitors in Google Analytics

Let's be clear about this: the use of cookies means that if you view an ad on a server with adult content on Monday, then visit a bookstore containing ads from the same ad provider on Tuesday, the ad provider could, in theory, tie the two visits together. When tracking cookies are tied to personal information, such as your bookstore account, the risks are even higher.

In practice, this doesn't happen much—perhaps because of strikingly different ad content—but it's the reason for the public outcry when sites like Facebook try to implement visitor tracking as part of their advertising systems.

On the other hand, cookies make it possible to tie together requests for several pages into a web visit from an individual, making them essential for web analytics.

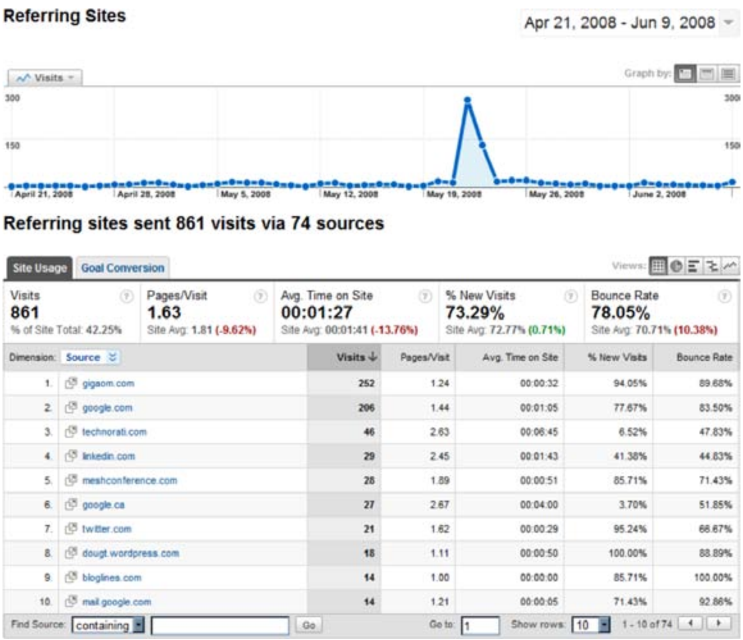
There are other ways to tie together page requests, including Uniform Resource Identifier (URI) parameters and hidden form elements. Some sites resort to these when a visitor's browser blocks cookies. These are less likely to cause privacy violations because they're not stored on a browser across visits while still allowing a website to stitch together the pages of a user's visit and keep track of things like shopping carts.

Visitor analysis, such as that shown in Figure 5-7, makes it possible to calculate other data, such as bounce rate (how many people left after seeing only one page), and how many pages a typical visitor looked at. Tracking visits also means we can now see where visitors came from and where they went.

From Visits to Outcomes: Tracking Goals

A visitor's page views are simply steps toward a goal or an outcome. It's these outcomes—not page views—that make your business work. After all, a user who comes to the site and buys something quickly generates far fewer page views but is much better for business than someone who browses around for a while before leaving without buying anything.

Tracking visits by outcome lets you make smarter decisions about your website. For example, if you simply analyze visits to your site based on where visitors came from, you'll see which sites are driving traffic to you, as shown in Figure 5-8.



Referring sites sent 861 visits via 74 sources

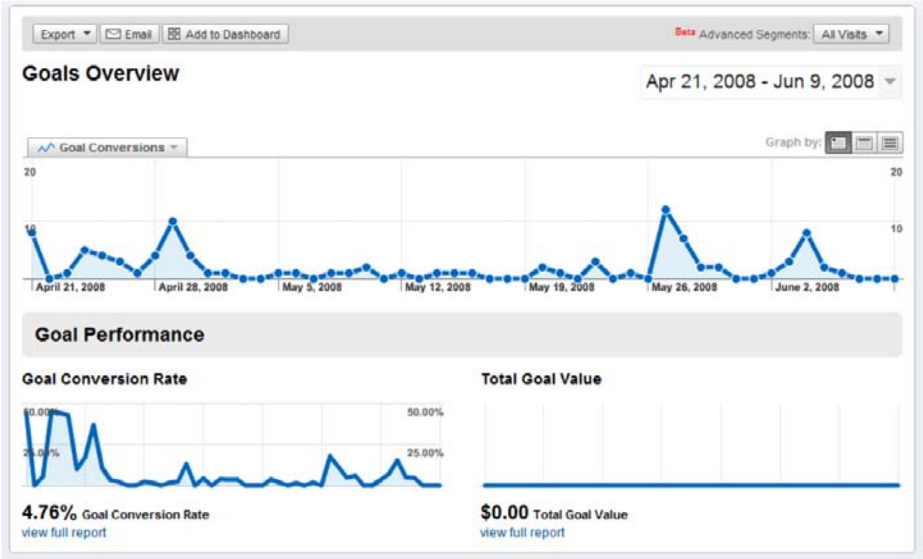
Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
861	1.63	00:01:27	73.29%	78.05%
% of Site Total: 42.25%	Site Avg: 1.81 (-9.62%)	Site Avg: 00:01:41 (-13.76%)	Site Avg: 72.77% (0.71%)	Site Avg: 70.71% (10.38%)

	Dimension: Source	Visits ↓	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	gigaom.com	252	1.24	00:00:32	94.05%	89.68%
2.	google.com	206	1.44	00:01:05	77.67%	83.50%
3.	technorati.com	46	2.63	00:06:45	6.52%	47.83%
4.	linkedin.com	29	2.45	00:01:43	41.38%	44.83%
5.	meshconference.com	28	1.89	00:00:51	85.71%	71.43%
6.	google.ca	27	2.67	00:04:00	3.70%	51.85%
7.	twitter.com	21	1.62	00:00:29	95.24%	66.67%
8.	dougt.wordpress.com	18	1.11	00:00:50	100.00%	88.89%
9.	bloglines.com	14	1.00	00:00:00	85.71%	100.00%
10.	mail.google.com	14	1.21	00:00:05	71.43%	92.86%

Find Source: containing Go Go to: 1 Show rows: 10 1 - 10 of 74

Figure 5-8. A visitor report sorted by number of visits

Now suppose that you have a specific goal, such as having visitors fill out surveys on your website. You can analyze how many people completed that goal over time and measure the success of your business in a report like the one shown in Figure 5-9.



Goals Overview

Apr 21, 2008 - Jun 9, 2008

Goal Performance

Goal Conversion Rate

Total Goal Value

4.76% Goal Conversion Rate

$0.00 Total Goal Value

Figure 5-9. Report of how many visitors attained a goal within a website

Goals and outcomes are essential. If you don't have a goal in mind, you can't optimize your site to encourage behaviors that lead to that goal. We typically express the steps a visitor takes towards a goal with a "funnel" like the one shown in Figure 5-10, indicating how many visitors proceeded through several steps and how many abandoned the process.

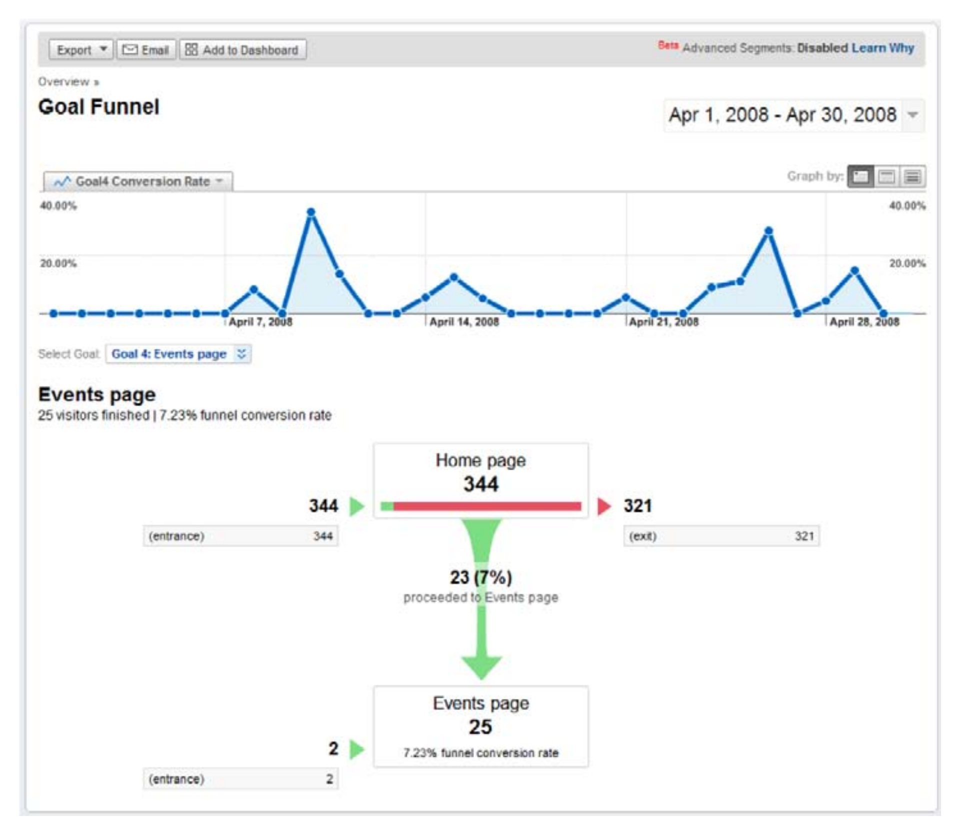
Funnels are a good way to depict where people are coming from and where they're abandoning a process. However, goals provide more than just accounting for conversions; they let you focus on what's working through segmentation.

Not all traffic is the same. Consider the list of sites that sent you traffic (in Figure 5-8). By segmenting those referring websites by goals, rather than just by total visits, you can see which of them sent you *visitors who mattered*—the ones that achieved the goals you hoped they would.

Referring sites that sent you large amounts of traffic may not have contributed to goals, while those who sent only a small fraction may have contributed significantly, as is the case here. From reports like the one shown in Figure 5-11, you can make the right decisions about where to focus your promotions.

Segmentation goes far beyond just measuring the effectiveness of referring sites. You can segment by anything that your analytics tool collects.

Early analytics tools were limited in what they could capture, however. What if you wanted to understand whether people who bought shoes also bought jackets? Or whether platinum members were more likely to comment on a post? Marketers needed



Select Goal: Goal 4: Events page

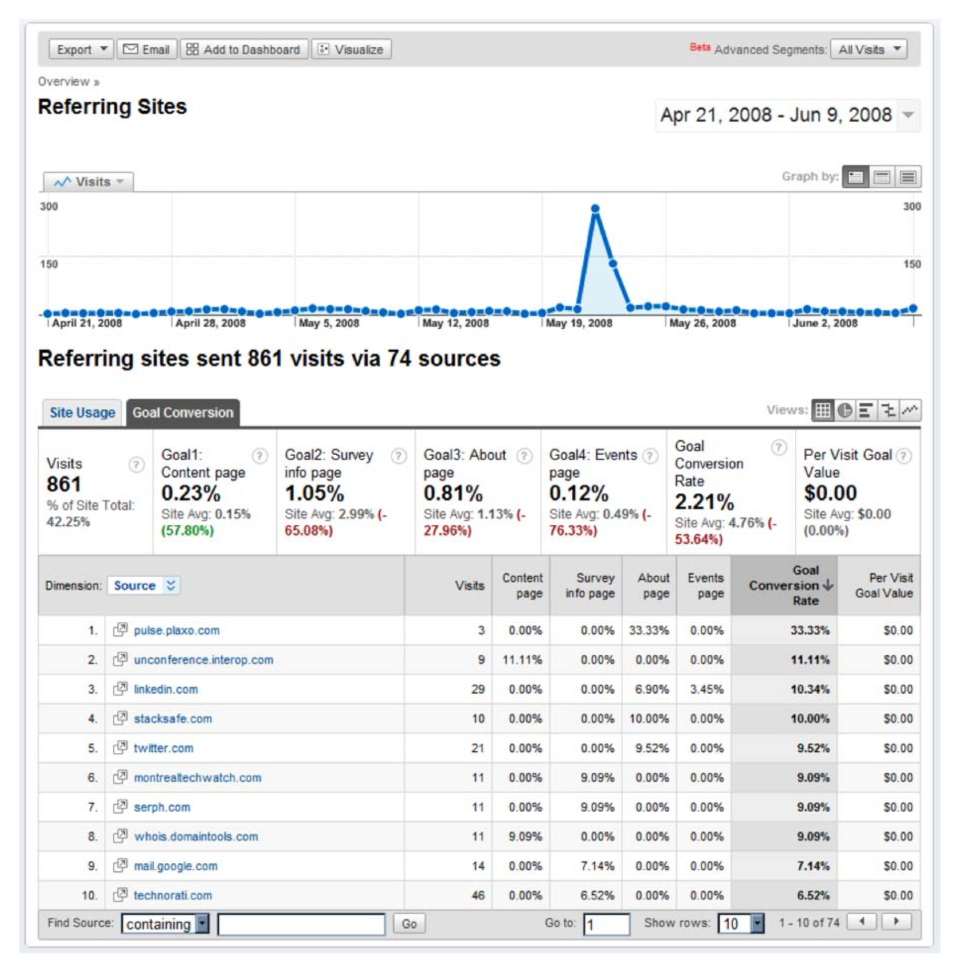
25 visitors finished | 7.23% funnel conversion rate

Figure 5-10. A goal funnel showing conversion and abandonment for a simple goal

a way to embed more context into their pages so they could segment in better ways. They needed page *tagging*.

From Technology to Meaning: Tagging Content

In the previous examples, we knew the goal we wanted visitors to accomplish (the "about" page), and we knew the various sites that were sending us traffic. The referring site is part of the default data an analytics tool collects. To add more context to a visit, we need to embed meaning in the page that the analytics system can later extract. This is called page tagging, and to understand it, we need to look briefly at how browser instrumentation works.



Referring Sites

Apr 21, 2008 - Jun 9, 2008

Referring sites sent 861 visits via 74 sources

Visits	Goal1: Content page	Goal2: Survey info page	Goal3: About page	Goal4: Events page	Goal Conversion Rate	Per Visit Goal Value
861 % of Site Total: 42.25%	0.23% Site Avg: 0.15% (57.80%)	1.05% Site Avg: 2.99% (-65.08%)	0.81% Site Avg: 1.13% (-27.96%)	0.12% Site Avg: 0.49% (-76.33%)	2.21% Site Avg: 4.76% (-53.64%)	$0.00 Site Avg: $0.00 (0.00%)

	Dimension: Source	Visits	Content page	Survey info page	About page	Events page	Goal Conversion ↓ Rate	Per Visit Goal Value
1.	pulse.plaxo.com	3	0.00%	0.00%	33.33%	0.00%	33.33%	$0.00
2.	unconference.interop.com	9	11.11%	0.00%	0.00%	0.00%	11.11%	$0.00
3.	linkedin.com	29	0.00%	0.00%	6.90%	3.45%	10.34%	$0.00
4.	stacksafe.com	10	0.00%	0.00%	10.00%	0.00%	10.00%	$0.00
5.	twitter.com	21	0.00%	0.00%	9.52%	0.00%	9.52%	$0.00
6.	montrealtechwatch.com	11	0.00%	9.09%	0.00%	0.00%	9.09%	$0.00
7.	serph.com	11	0.00%	9.09%	0.00%	0.00%	9.09%	$0.00
8.	whois.domaintools.com	11	9.09%	0.00%	0.00%	0.00%	9.09%	$0.00
9.	mail.google.com	14	0.00%	7.14%	0.00%	0.00%	7.14%	$0.00
10.	technorati.com	46	0.00%	6.52%	0.00%	0.00%	6.52%	$0.00

Figure 5-11. The list of referring sites, this time segmented by goal conversion rate, shows which referring sites contributed to your business the most

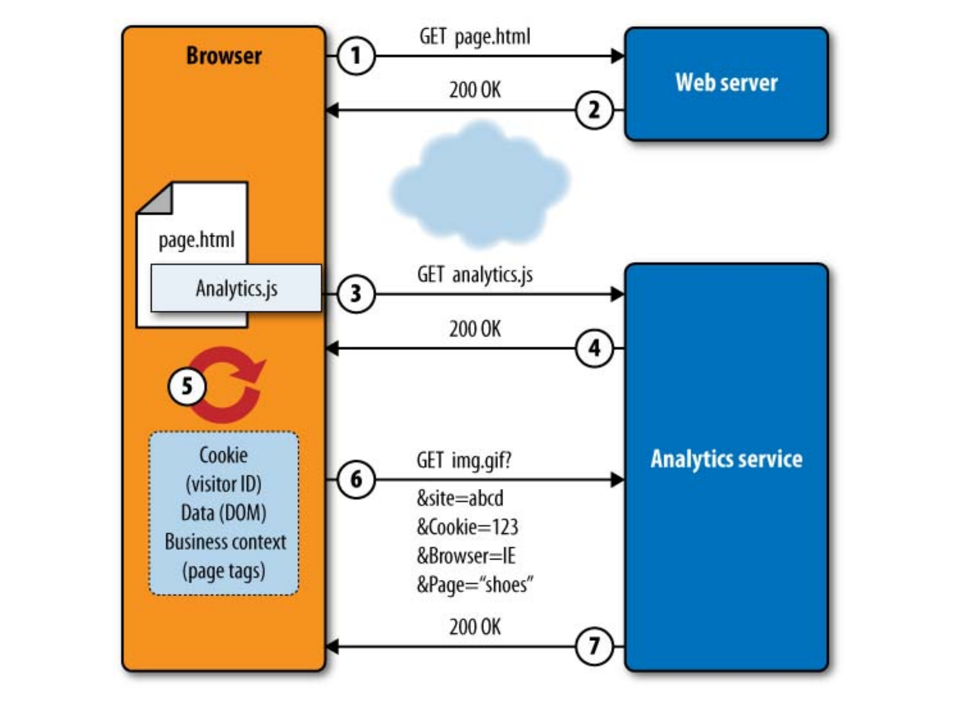
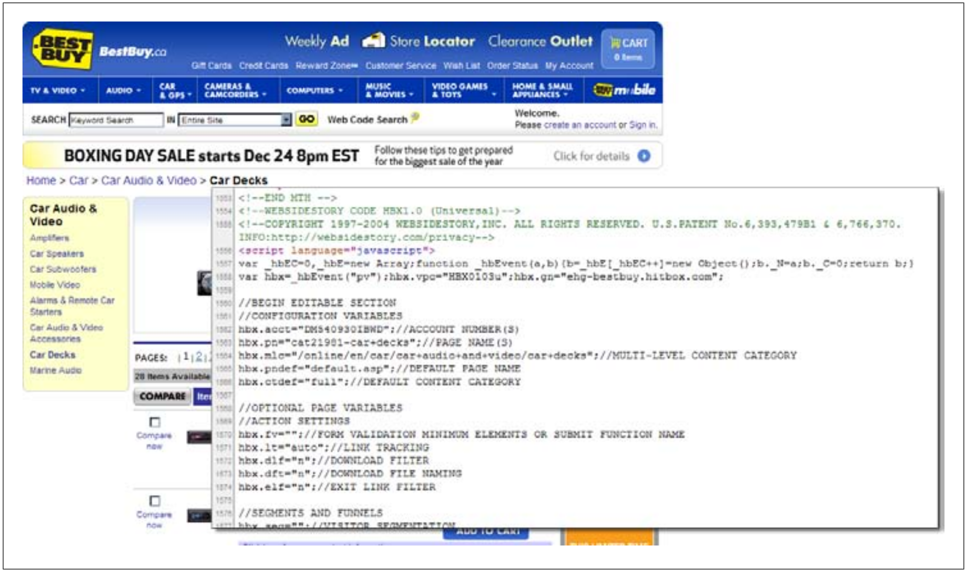


Figure 5-12. How JavaScript-based analytics works

Figure 5-12 illustrates how JavaScript-based analytics systems collect page hits from a visitor.

1. The browser requests a page from the web server.
2. The web server returns the page, which contains a reference to a piece of JavaScript provided by the analytics service.
3. The JavaScript is retrieved from the service. In some cases, it may already be cached on the browser or built into the page itself.
4. The analytics service returns the JavaScript to the browser.
5. The browser executes the JavaScript, which collects information about the page request.
6. The JavaScript puts this information into an HTTP request, often for a tiny image. All of these attributes are sent to the analytics service.
7. The service responds with a small object (which can be ignored).



```
<!--END MTH -->
<!--WEBSIDESTORY CODE HBX1.0 (Universal)-->
<!--COPYRIGHT 1997-2004 WEBSIDESTORY,INC. ALL RIGHTS RESERVED. U.S.PATENT No.6,393,479B1 & 6,766,370.
INFO:http://websidestory.com/privacy-->
<script language="javascript">
var _hbEC=0,_hbE=new Array;function _hbEvent(a,b){b=_hbE[_hbEC++]=new Object();b._N=a;b._C=0;return b;}
var hbx=_hbEvent("pv");hbx.vpc="HBX0103u";hbx.gn="ehg-bestbuy.hitbox.com";

//BEGIN EDITABLE SECTION
//CONFIGURATION VARIABLES
hbx.acct="DM540930IBWD";//ACCOUNT NUMBER(S)
hbx.pn="cat21981-car+decks";//PAGE NAME(S)
hbx.mlc="/online/en/car/car+audio+and+video/car+decks";//MULTI-LEVEL CONTENT CATEGORY
hbx.pndef="default.asp";//DEFAULT PAGE NAME
hbx.ctdef="full";//DEFAULT CONTENT CATEGORY

//OPTIONAL PAGE VARIABLES
//ACTION SETTINGS
hbx.fv="";//FORM VALIDATION MINIMUM ELEMENTS OR SUBMIT FUNCTION NAME
hbx.lt="auto";//LINK TRACKING
hbx.dlf="n";//DOWNLOAD FILTER
hbx.dft="n";//DOWNLOAD FILE NAMING
hbx.elf="n";//EXIT LINK FILTER

//SEGMENTS AND FUNNELS
hbx.seg="";//VISITOR SEGMENTATION
```

Figure 5-13. Page tags within a page at BestBuy.com

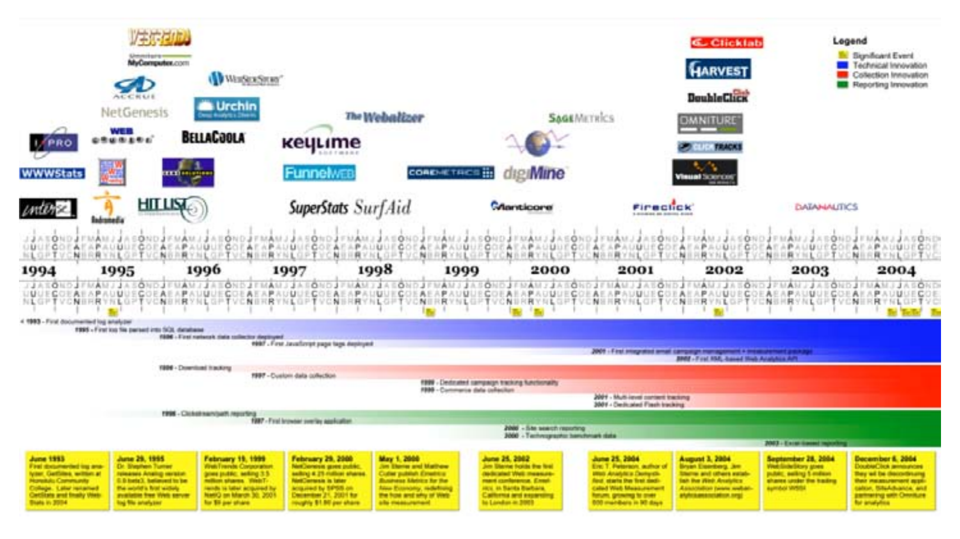
We'll return to this topic later in the chapter when we consider implementation.

The magic of page tagging happens in step 5, in which the JavaScript collects information to send back to the analytics service. This information includes:

- Information about the *technical environment*, such as screen resolution, referring URL, or operating system.
- Information about the *visitor*, such as a cookie that uniquely identifies her and lets the analytics service stitch several pages together into a visit.
- Information within the page itself that provides *business context*. For example, a retailer might tag a page with "shoes" if it offers shoes for sale. Or it might indicate whether the ad the user saw was part of a specific campaign. Figure 5-13 shows an example of page tags on a retail website.

By recording not only technical data on the visit, but also business context, the analyst can then segment outcomes with this context. He can see whether one campaign works better than another, or whether users who buy shoes ultimately buy jackets.

Notice how far we've come from the early days of analytics, in which IT teams looked at hits to measure capacity? Today, web analytics is a marketing discipline used to measure the effectiveness of communications strategies.



June 1993
First documented log analyzer, GetSites, written at Honolulu Community College. Later renamed GetStats and finally Web Stats in 2004

June 29, 1995
Dr. Stephen Turner releases Analog version 0.9 beta3, believed to be the world's first widely available free Web server log file analyzer

February 19, 1999
WebTrends Corporation goes public, selling 3.5 million shares. WebTrends is later acquired by NetIQ on March 30, 2001 for $9 per share

February 29, 2000
NetGenesis goes public, selling 4.25 million shares. NetGenesis is later acquired by SPSS on December 21, 2001 for roughly $1.90 per share

May 1, 2000
Jim Sterne and Matthew Cutler publish *E-metrics Business Metrics for the New Economy*, redefining the how and why of Web site measurement

June 25, 2002
Jim Sterne holds the first dedicated Web measurement conference, Emetrics, in Santa Barbara, California and expanding to London in 2003

June 25, 2004
Eric T. Peterson, author of *Web Analytics Demystified*, starts the first dedicated Web Measurement forum, growing to over 800 members in 90 days

August 3, 2004
Bryan Eisenberg, Jim Sterne and others establish the Web Analytics Association (www.webanalyticsassociation.org)

September 28, 2004
WebSideStory goes public, selling 5 million shares under the trading symbol WSSI

December 6, 2004
DoubleClick announces they will be discontinuing their measurement application, SiteAdvance, and partnering with Omniture for analytics

Figure 5-14. A visual history of web analytics

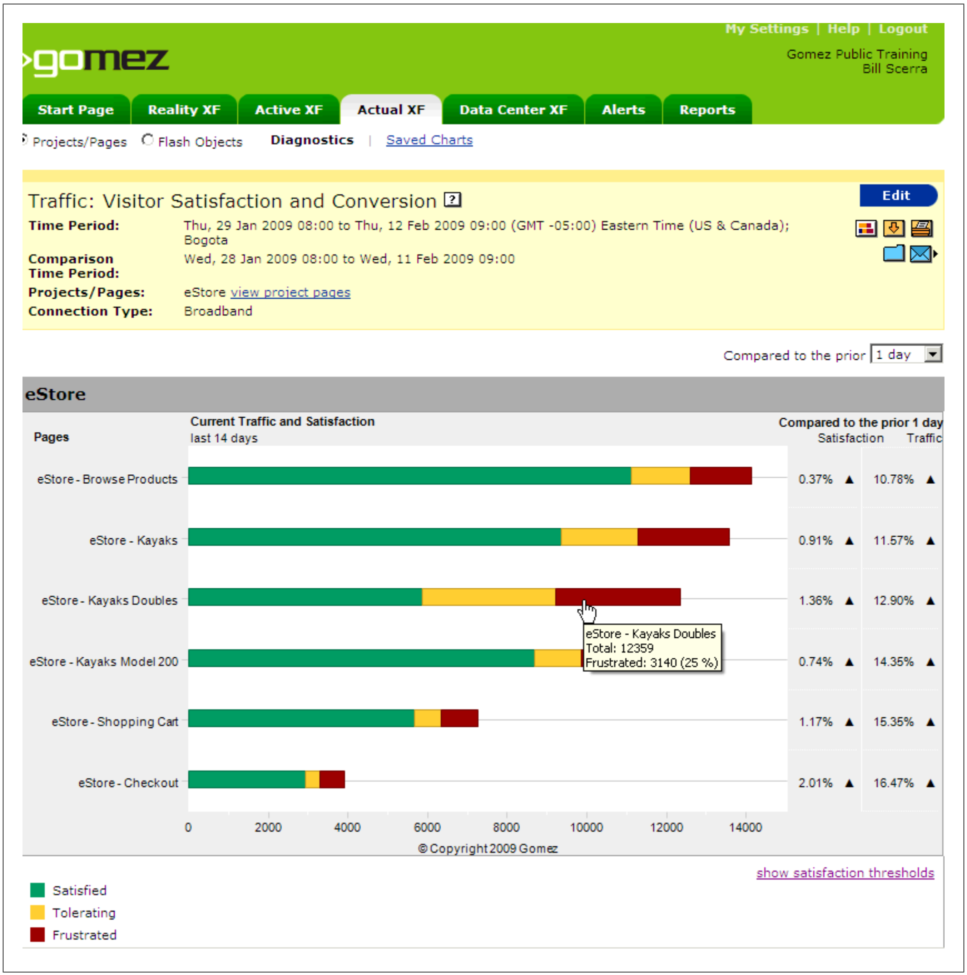
It's also a full-time job, with large organizations often having entire departments devoted to tagging, experimentation, and reporting. A history of web analytics, compiled by the Yahoo! web analytics mailing group as a result of discussions by Eric T. Peterson and John Pestana, is illustrated in Figure 5-14.

Analytics is still changing, particularly around the integration of other data sources and around the move away from page-centric websites.

An Integrated View

Having moved so far from their technical roots, marketers are now realizing that there's more to user experience than just analytics, and that they need to bring other data sources into the fold if they're going to optimize their websites.

- Poor performance undermines user productivity, reduces conversion rates, and discourages visitors from sticking around. Marketers care about the effect performance has on conversion. Figure 5-15 shows an example of a conversion funnel that considers end user performance.
- Visitors have lots on their minds, and marketers ignore it at their own peril. It's easy to survey customers and solicit feedback, so companies are starting to integrate VOC data or page rankings with visits and goals, as shown in Figure 5-16.



Traffic: Visitor Satisfaction and Conversion

Time Period: Thu, 29 Jan 2009 08:00 to Thu, 12 Feb 2009 09:00 (GMT -05:00) Eastern Time (US & Canada); Bogota

Comparison Time Period: Wed, 28 Jan 2009 08:00 to Wed, 11 Feb 2009 09:00

Projects/Pages: eStore view project pages

Connection Type: Broadband

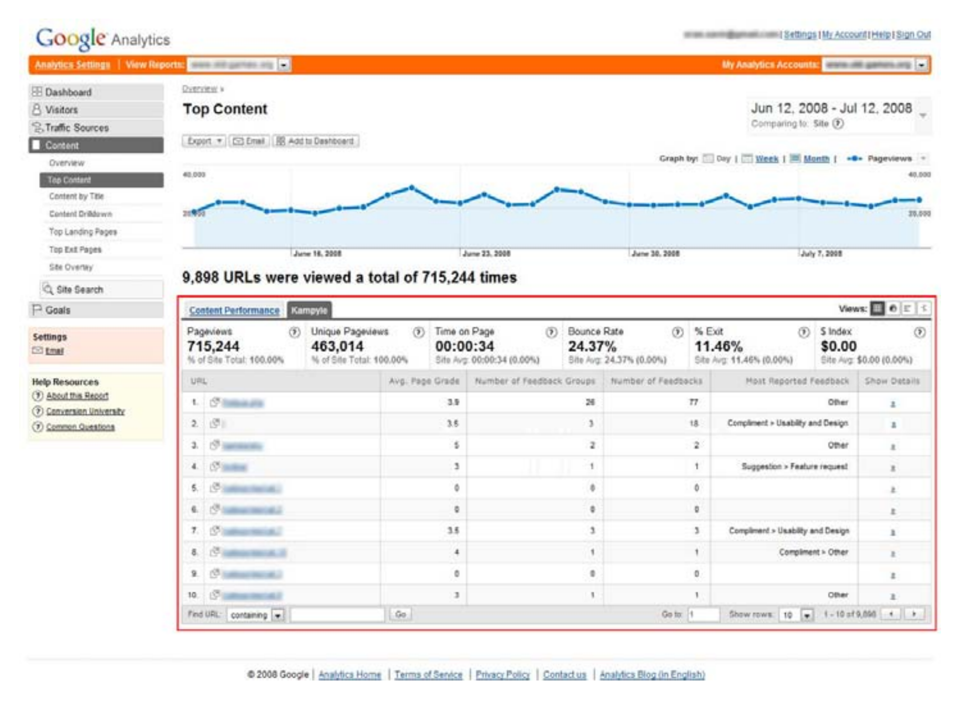
Compared to the prior 1 day

eStore - Kayaks Doubles
Total: 12359
Frustrated: 3140 (25 %)

Pages	Compared to the prior 1 day Satisfaction	 Traffic
eStore - Browse Products	0.37% ▲	10.78% ▲
eStore - Kayaks	0.91% ▲	11.57% ▲
eStore - Kayaks Doubles	1.36% ▲	12.90% ▲
eStore - Kayaks Model 200	0.74% ▲	14.35% ▲
eStore - Shopping Cart	1.17% ▲	15.35% ▲
eStore - Checkout	2.01% ▲	16.47% ▲

🟩 Satisfied
🟨 Tolerating
🟥 Frustrated

Figure 5-15. End user performance correlated with requests for each page in a goal funnel in Gomez



Top Content

Jun 12, 2008 - Jul 12, 2008
Comparing to: Site

9,898 URLs were viewed a total of 715,244 times

Pageviews	Unique Pageviews	Time on Page	Bounce Rate	% Exit	$ Index
715,244	463,014	00:00:34	24.37%	11.46%	$0.00
% of Site Total: 100.00%	% of Site Total: 100.00%	Site Avg: 00:00:34 (0.00%)	Site Avg: 24.37% (0.00%)	Site Avg: 11.46% (0.00%)	Site Avg: $0.00 (0.00%)

	URL	Avg. Page Grade	Number of Feedback Groups	Number of Feedbacks	Most Reported Feedback	Show Details
1.	[redacted]	3.9	26	77	Other	±
2.	[redacted]	3.6	3	18	Compliment > Usability and Design	±
3.	[redacted]	5	2	2	Other	±
4.	[redacted]	3	1	1	Suggestion > Feature request	±
5.	[redacted]	0	0	0		±
6.	[redacted]	0	0	0		±
7.	[redacted]	3.5	3	3	Compliment > Usability and Design	±
8.	[redacted]	4	1	1	Compliment > Other	±
9.	[redacted]	0	0	0		±
10.	[redacted]	3	1	1	Other	±

Find URL: containing ▾ Go   Go to: 1   Show rows: 10 ▾   1 - 10 of 9,898

Figure 5-16. Blending visitor survey and feedback information with analytics data using Kampyle and Google Analytics

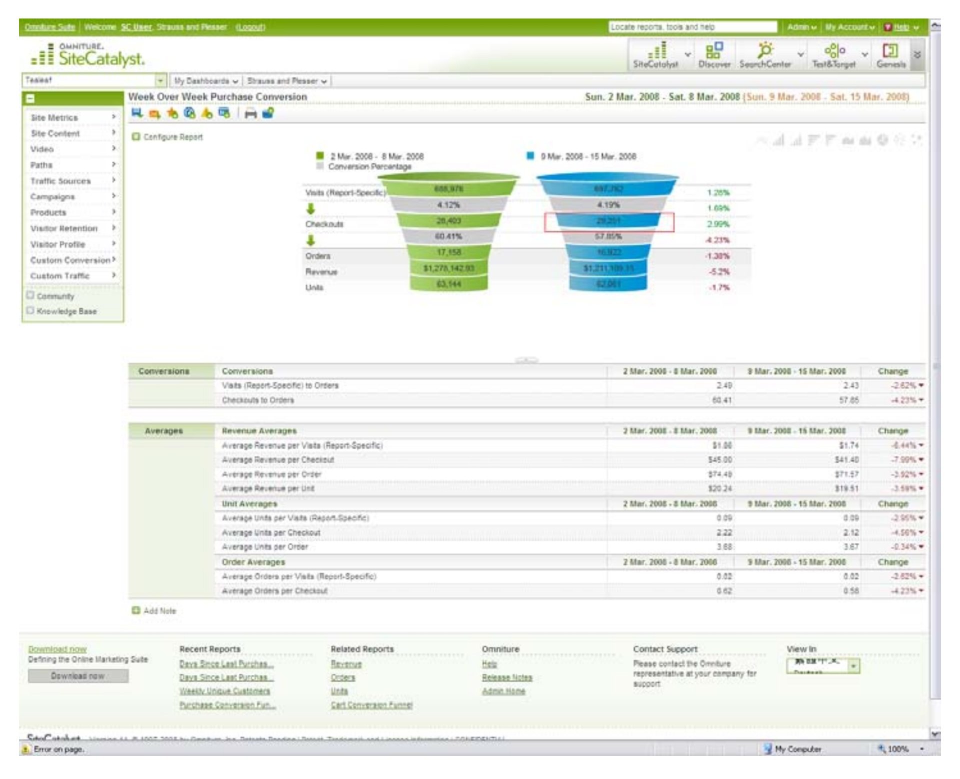
- Usability has a huge impact on conversion rates, so websites are trying to identify bottlenecks and usability issues at each step of a conversion funnel, like the one shown in Figure 5-17, to optimize their sites and improve user experience.

Making sense of all this information means tying it back to the business context of web analytics. As a result, we're seeing significant advances in web visibility, particularly in the area of data integration for web analytics platforms.

Places and Tasks

At the same time that analysts are integrating performance, visitor opinion, and usability data, the nature of websites is changing. Developers are using programming built into the pages themselves in the form of Ajax-based JavaScript, Flash, Java, and Silverlight, resulting in fewer pages but more actions within a page. Some websites take this to the extreme, consisting of just a rich application embedded in a single web page.

For these sites, the old paradigm of a visitor navigating through several pages toward a goal is no longer accurate. Analytics tools need to adjust what they capture and how



Week Over Week Purchase Conversion

Sun. 2 Mar. 2008 - Sat. 8 Mar. 2008 (Sun. 9 Mar. 2008 - Sat. 15 Mar. 2008)

Conversions	Conversions	2 Mar. 2008 - 8 Mar. 2008	9 Mar. 2008 - 15 Mar. 2008	Change
	Visits (Report-Specific) to Orders	2.49	2.43	-2.62% ▾
	Checkouts to Orders	60.41	57.85	-4.23% ▾

Averages	Revenue Averages	2 Mar. 2008 - 8 Mar. 2008	9 Mar. 2008 - 15 Mar. 2008	Change
	Average Revenue per Visits (Report-Specific)	$1.86	$1.74	-6.44% ▾
	Average Revenue per Checkout	$45.00	$41.40	-7.99% ▾
	Average Revenue per Order	$74.49	$71.57	-3.92% ▾
	Average Revenue per Unit	$20.24	$19.51	-3.59% ▾
	Unit Averages	2 Mar. 2008 - 8 Mar. 2008	9 Mar. 2008 - 15 Mar. 2008	Change
	Average Units per Visits (Report-Specific)	0.09	0.09	-2.95% ▾
	Average Units per Checkout	2.22	2.12	-4.56% ▾
	Average Units per Order	3.68	3.67	-0.34% ▾
	Order Averages	2 Mar. 2008 - 8 Mar. 2008	9 Mar. 2008 - 15 Mar. 2008	Change
	Average Orders per Visits (Report-Specific)	0.02	0.02	-2.62% ▾
	Average Orders per Checkout	0.62	0.58	-4.23% ▾

Recent Reports	Related Reports	Omniture	Contact Support	View In
Days Since Last Purchas...	Revenue	Help	Please contact the Omniture representative at your company for support	
Days Since Last Purchas...	Orders	Release Notes		
Weekly Unique Customers	Units	Admin Home		
Purchase Conversion Fun...	Cart Conversion Funnel			

Figure 5-17. A conversion funnel in Omniture's SiteCatalyst linking to Tealeaf in order to analyze and replay visits to understand visitor behavior

they record it accordingly, because not all websites follow the page-centric model on which web analytics was conceived.

Instead of thinking of websites as made up of pages, think of them as a set of "places" and "tasks." Some of the time, users are in a "place"—they're exploring, browsing, and interacting. Occasionally, they'll undertake a "task"—a series of steps toward a goal. They may complete or abandon that task. This is a generalization that works for both page-centric and RIA-based (rich Internet application) websites.

Thinking in terms of places and tasks helps you to understand which outcomes you need to focus on across your entire web business.

Places are where users hang out

On reddit, a list of submissions from others, ranked by popularity, is a place. Visitors can perform a series of small actions like opening linked stories in new tags, or voting things up and down. Similarly, on Wikipedia, a subject entry is a place. In Google Apps, a spreadsheet is a place.

When a user's in a place, you care about his *productivity*—whether the experience was efficient and satisfying. On reddit, are users able to vote submissions up or down smoothly? On Wikipedia, can they scroll through entries easily and find what they're looking for, and do images load? On Google Apps, are they successfully building business projections in that spreadsheet?

You still need to pay attention to abandonment, which happens when a user gets bored doing something and goes elsewhere. But in the case of a place, it's abandonment due to boredom, satisfaction, or disinterest, not due to an inability to accomplish something.

Tasks occur when users have a mission

By contrast, a task is something the user sets out to accomplish. A task comprises several steps, and puts the visitor in a different mental state. Tasks are goal-oriented. On reddit, this is a user creating an account for himself or submitting a new link. On Wikipedia, it's a visitor deciding to edit a page. On Google Apps, it's the employee sharing her spreadsheet with someone.

When a user's trying to accomplish a task, we care about *effectiveness*—whether the task was completed successfully or not. Did the invite result in a new enrollment? Did the user complete the purchase? Was the edit of the article ultimately saved? Was the user able to add the widget to her dashboard? Did the spreadsheet's recipient receive it successfully? When visitors abandon tasks, it's because they ran into a problem. Something was too costly, or confusing, or violated their privacy, or dampened their enthusiasm for the job.

A new way to look at sites

Looking at websites as collections of places and tasks reveals the limitations of page-centric, funnel-minded web analytics.

For one thing, you realize that you need to instrument places and tasks very differently. Places need analysis of actions within the place. How many videos did he watch? How often did he pause them? Did he see the ad? How many entries can she complete an hour? On the other hand, tasks need analysis of progress toward an outcome. Did she send the invite? How far in the form did she get? Which steps took the longest?

To further complicate matters, tasks often involve steps beyond the view of analytics, such as e-mail invitations, instant messages, RSS feeds, and third-party sites. We also want to know other things about tasks. Did the message bounce? Did the invitation's recipient act on it?

Tracking the accomplishment of a task across multiple systems is a challenge, with all manner of tracking cookies, dynamic URLs, and embedded GIFs used to try and follow the task to completion. As outlined in *http://analytics.blogspot.com/2009/02/two-cool-integrations-telephone-leads.html*, some companies coordinate analytics data with offsite goal completion such as phone calls or instant messages. For more information on multichannel monitoring, see *www.kaushik.net/avinash/2008/07/tracking-offline-conversions-hope-seven-best-practices-bonus-tips.html*.

What can you do to get started?

Most web operators have a mental map of their sites. Some even draw it on a wall. You can map out a site, consisting of places and tasks, in this way.

- For each place, make a note of all the events you care about, including timing events ("a video starts playing") and user interactions ("user upvotes and the button's color changes"). Focus on visitor productivity and time spent in the place. Identify the actions that initiate a task (such as "share this spreadsheet"), taking the user out of the place.
- For each task, make a note of all the steps you want to track, including those that aren't on your site. Identify the key metrics you should know (in a mailout, this might be bounce rate, open rate, and click rate.) Focus on conversions, abandonments, and their causes across multiple channels.

The end result is a much more accurate representation of the ebb and flow of your online business. It will probably reveal significant gaps in your web visibility strategy, too, but at least you'll now know where your blind spots are.

The next time you're presenting your web monitoring results, overlay them on the map of places and tasks that you've made. For each place or task, show the analytics (what the users did) and the user experience (whether they could do it). If you have psychographic information (why they did it) such as surveys, or usability metrics (how they did it), include that as well. Also build in any external metrics like mail bounce rates or Facebook group member count.

We'll return to the concept of places and tasks, and web performance considerations for both, when we look at measuring end user experience later in the book.

The Three Stages of Analytics

Any visit can be broken into three distinct stages: finding the website, using the website, and leaving the website. Web analytics is the science of identifying segments of your visitors that go through these three stages in significantly different ways, and then focusing on segments that are the best for your business.

- If a certain marketing campaign helps people find your site, all else being equal, you should put more money into that campaign.
- If a certain kind of site content or layout encourages visitors to use your site in ways that are good for your business, you should focus on that content.

- If certain content, layout, or page load speeds encourage users to stay longer and return more often, you should focus on doing those things more consistently.
- If certain ad locations encourage your visitors to click on those ads, generating revenue for you as they leave the site, you should charge more for those locations and sell more of them.

Finding the Site: The Long Funnel

Getting traffic to your site is often the first thing marketers think about. If all else is equal, more traffic translates into more revenues and a healthy business.

Of course, all else is seldom equal. In addition to search engines, visitors come from a wide range of sources (shown in Figure 5-18), among them hundreds of social networks. Each visit has a different outcome. Some visitors stop by briefly, never to return, while others become lifelong readers or loyal customers.

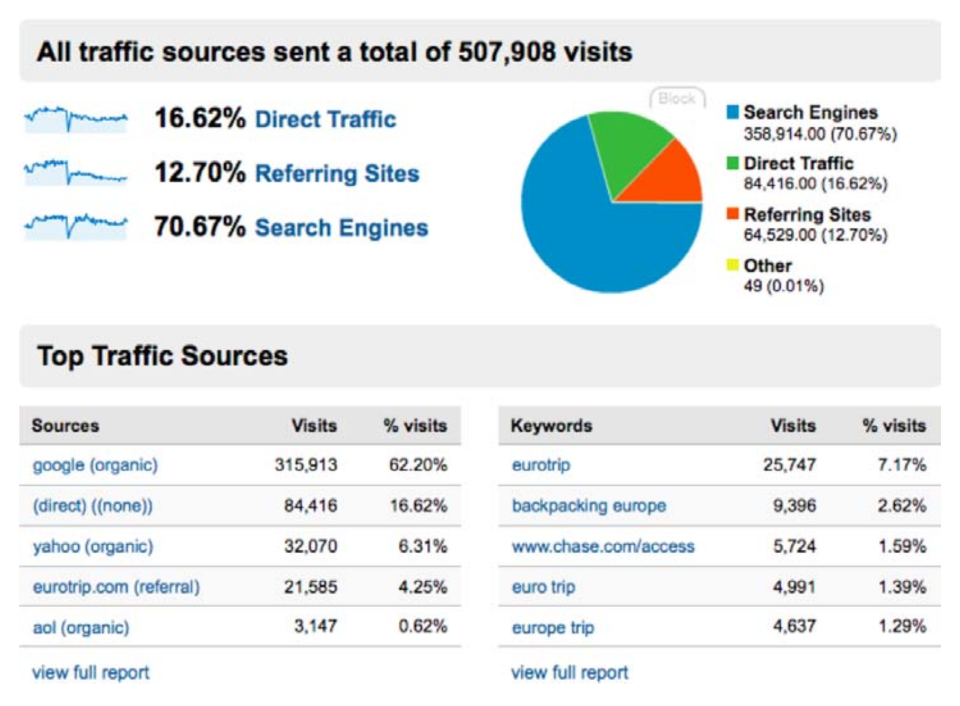
By optimizing the step *before* users hit the entry page, you can increase traffic volumes, which makes for a larger conversion pool. Let's look at the ways users come to your website.

> It's important to make the distinction between an *entry* page and a *landing* page. The entry page refers to the first page a visitor loaded for a visit to your site. However, a visitor might hit multiple landing pages during a single visit. How is this possible? A visitor may navigate into your site initially (making this both the entry page and the landing page), then navigate away to another site, and finally return to your site in a relatively short time frame while their session cookie is still valid, making it seem like the second landing page was part of the visit. In other words, there can be multiple landing pages but only one entry page.

Visitors can find your site in several ways:

- They can type the URL into the browser directly, known as *direct traffic*.
- They can link to you following an *organic search*.
- They can click on an advertisement or sponsorship link, which counts as banner or *paid search* campaign traffic.
- They can follow a link from another site or from a community such as a social network through *referrals*.
- They can respond to an email invitation as part of an *email marketing campaign*.

Each of these sources is treated differently in web analytics.



All traffic sources sent a total of 507,908 visits

16.62% Direct Traffic

12.70% Referring Sites

70.67% Search Engines

Top Traffic Sources

Sources	Visits	% visits
google (organic)	315,913	62.20%
(direct) ((none))	84,416	16.62%
yahoo (organic)	32,070	6.31%
eurotrip.com (referral)	21,585	4.25%
aol (organic)	3,147	0.62%

view full report

Keywords	Visits	% visits
eurotrip	25,747	7.17%
backpacking europe	9,396	2.62%
www.chase.com/access	5,724	1.59%
euro trip	4,991	1.39%
europe trip	4,637	1.29%

view full report

Figure 5-18. Google Analytics Traffic Sources Overview page

Direct traffic

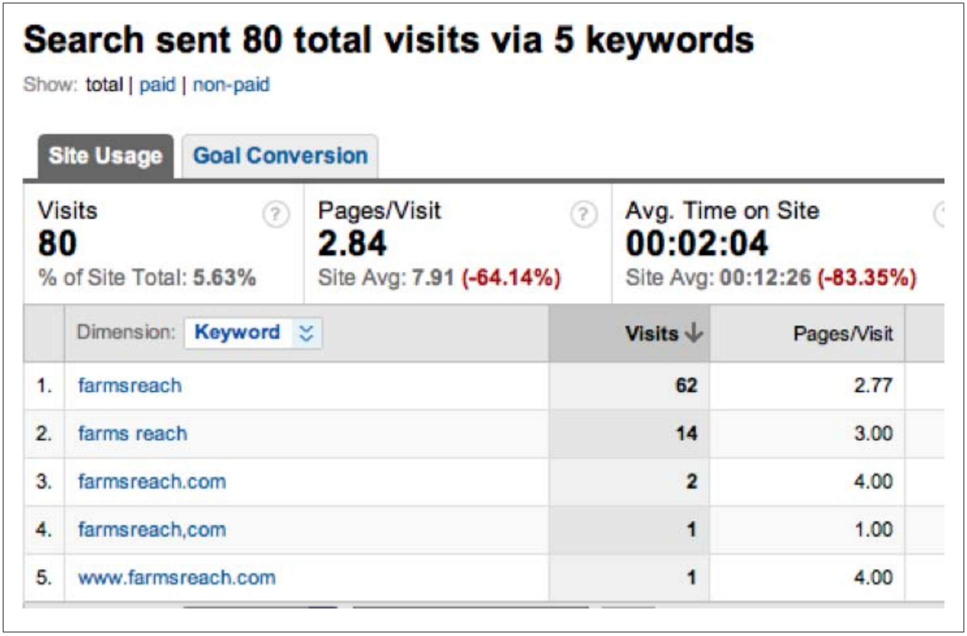
The simplest way for people to visit your website is by typing it into an address bar. They may have heard your brand name and decided to enter it as a URL, or they may have seen the URL in print advertising.

Direct traffic is the least informative of all traffic sources, because you don't know how the visitor discovered you. As a result, it's hard to optimize. If you need to know more about these visitors, intercept them with a survey and ask them how they found out about your site. We'll look at VOC surveys in Chapter 7.

Direct traffic can also be misleading for several reasons:

Navigational search

Between 15 and 20 percent of web searches consist of users who type a URL into the search engine (rather than the browser's address bar). A surprising number of Internet users believe that this is how the Internet works, and only ever use searches to find sites, even those they frequent every day. In these cases, the URL of the website—or its name—appears in keyword reports, as shown in Figure 5-19. This is known as *navigational search*, and will usually be shown as organic search traffic rather than direct traffic.



Search sent 80 total visits via 5 keywords

Show: total | paid | non-paid

Visits	Pages/Visit	Avg. Time on Site
80	2.84	00:02:04
% of Site Total: 5.63%	Site Avg: 7.91 (-64.14%)	Site Avg: 00:12:26 (-83.35%)

	Dimension: Keyword	Visits ↓	Pages/Visit
1.	farmsreach	62	2.77
2.	farms reach	14	3.00
3.	farmsreach.com	2	4.00
4.	farmsreach,com	1	1.00
5.	www.farmsreach.com	1	4.00

Figure 5-19. Navigational search results for www.farmsreach.com; high levels of navigational search for the URL can be an indicator of a nontechnical demographic

Type-in traffic

Type-in traffic occurs when a user types a word—such as "pizza"—into a search engine to see what will happen. Depending on the browser or search engine, this will often take the user to a named site such as "*www.pizza.com*" that makes its money from referring visitors to other sites. Type-in traffic is why popular type-in domains are sometimes sold for millions of dollars.

John Battelle has an excellent write-up on the type-in traffic business at *http://battellemedia.com/archives/002118.php*.

Bookmarking

Users who have visited a site may bookmark it, or the browser may display the most frequently visited sites when it is first launched. Clicking on a bookmark, a "frequently visited" link, or a link in the browser creates a visit with no referring site. The visit is classified as direct traffic by the analytics engine.

Desktop client

Some desktop tools, such as Twitter and email clients, don't provide any information about a followed link—they simply launch a web browser with a URL that the user received through the desktop tool. As a result, these links are classified as direct traffic, even though they were referred through a community.

JavaScript redirection

If a user follows a JavaScript-triggered link to your site, the browser might not pass on referring information and the visit will be mislabeled. Careful scripting can fix this, but few web developers make the extra effort.

Browser inconsistency

Browser behavior can reduce the accuracy of direct visitor count. For example, in certain conditions, Microsoft's Internet Explorer 7 will not pass along referring information if you load a website in another window or tab.

Bots, spiders, and probes

Scripts and automated visitors, such as search engines indexing your website, may not be properly identified and may be counted as direct visits.

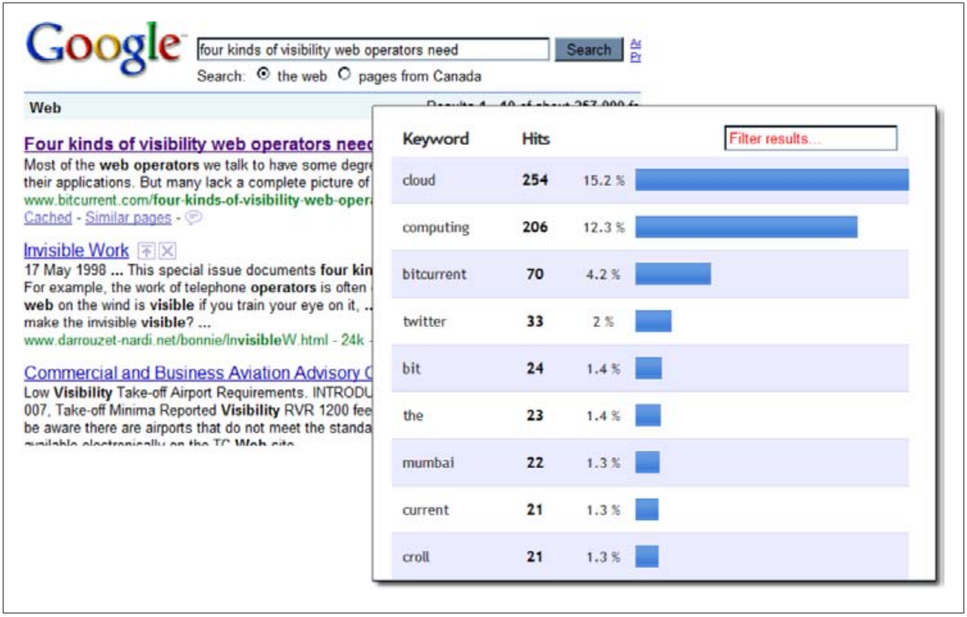
These caveats mean that direct traffic counts can be inaccurate. You can mitigate some of this inaccuracy through the use of custom URLs that make it easier to track the effectiveness of a traffic source. For example, in a print campaign, you can include the name of a specific magazine in the URL that you place in that magazine's ads (i.e., *www.watchingwebsites.com/magazinename*). Then you can track the success of advertising in each magazine by segmenting visits within analytics by the URL.

Organic search

One of the most common ways people find a site is through organic, or unpaid, search. A web user enters a search term and clicks on one of the results, as shown in Figure 5-20. The referring URL includes not only the site that referred the user, but also the search terms that drove the traffic to your site.

Over time, your web analytics package develops a list of the most common keywords driving traffic to your site. Some of the keywords that drive traffic can be quite weird and unexpected, as shown in Figure 5-21, and may be a clue that your visitors' intentions on your site aren't what you think they are, and are worthy of further investigation.

If search engines consider your content more relevant, your site will rank higher than others and appear higher in search results. Many factors influence this, such as the text and metadata on your site and the number of other sites that link to yours.



four kinds of visibility web operators need

Search

Search: ⦿ the web ○ pages from Canada

Web

Results 1 - 10 of about 257,000 fo

Four kinds of visibility web operators need
Most of the **web operators** we talk to have some degr
their applications. But many lack a complete picture of
www.bitcurrent.com/**four-kinds-of-visibility-web-oper**
Cached - Similar pages -

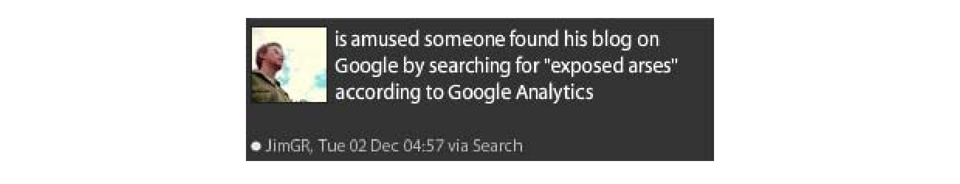
Invisible Work
17 May 1998 ... This special issue documents **four kin**
For example, the work of telephone **operators** is often
web on the wind is **visible** if you train your eye on it, ..
make the invisible **visible**? ...
www.darrouzet-nardi.net/bonnie/**Invisible**W.html - 24k

Commercial and Business Aviation Advisory C
Low **Visibility** Take-off Airport Requirements. INTRODU
007, Take-off Minima Reported **Visibility** RVR 1200 fee
be aware there are airports that do not meet the standa

Filter results...

Keyword	Hits	
cloud	254	15.2 %
computing	206	12.3 %
bitcurrent	70	4.2 %
twitter	33	2 %
bit	24	1.4 %
the	23	1.4 %
mumbai	22	1.3 %
current	21	1.3 %
croll	21	1.3 %

Figure 5-20. Organic search keywords for Bitcurrent.com



is amused someone found his blog on Google by searching for "exposed arses" according to Google Analytics

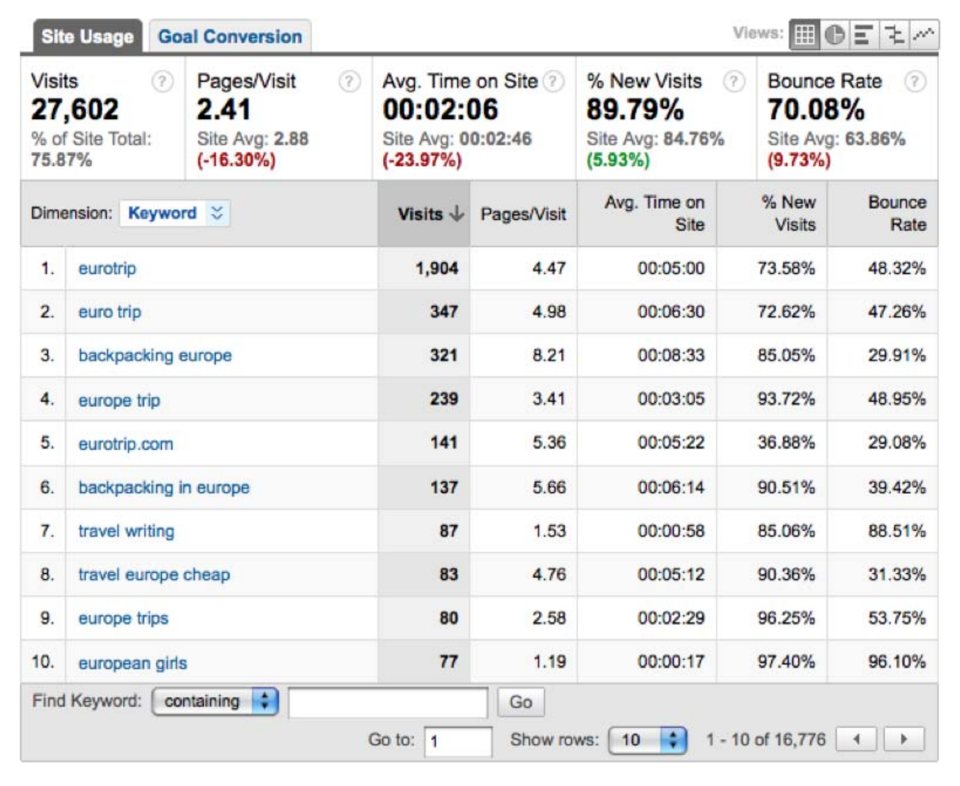
● JimGR, Tue 02 Dec 04:57 via Search

Figure 5-21. Keywords that drive traffic to your site can sometimes be unpredictable

As an online marketer, one of your main jobs is to ensure that your site content is ranked highly and that the right organic search terms send traffic to the site. Web analytics helps you to understand whether your site is properly optimized and which keywords are driving traffic, through reports like the one in Figure 5-22.

> The art of search engine optimization (SEO) is interesting, fun, and not covered in this book. We would rather leave you in the hands of the masters. For a great start in understanding SEO ranking, check out *www.seomoz.org/article/search-ranking-factors*.

Either way, the ultimate goal is to get better visibility and more qualified traffic.



Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
27,602	2.41	00:02:06	89.79%	70.08%
% of Site Total: 75.87%	Site Avg: 2.88 (-16.30%)	Site Avg: 00:02:46 (-23.97%)	Site Avg: 84.76% (5.93%)	Site Avg: 63.86% (9.73%)

Dimension: Keyword

	Keyword	Visits ↓	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	eurotrip	1,904	4.47	00:05:00	73.58%	48.32%
2.	euro trip	347	4.98	00:06:30	72.62%	47.26%
3.	backpacking europe	321	8.21	00:08:33	85.05%	29.91%
4.	europe trip	239	3.41	00:03:05	93.72%	48.95%
5.	eurotrip.com	141	5.36	00:05:22	36.88%	29.08%
6.	backpacking in europe	137	5.66	00:06:14	90.51%	39.42%
7.	travel writing	87	1.53	00:00:58	85.06%	88.51%
8.	travel europe cheap	83	4.76	00:05:12	90.36%	31.33%
9.	europe trips	80	2.58	00:02:29	96.25%	53.75%
10.	european girls	77	1.19	00:00:17	97.40%	96.10%

Find Keyword: containing Go

Go to: 1 Show rows: 10 1 - 10 of 16,776

Figure 5-22. A list of keywords visitors used to find eurotrip.com on search engines

Paid search and online advertising

If you're not getting the attention you want organically, you can always pay for it. According to the Interactive Advertising Bureau, online advertising is a $21 billion-dollar industry. Three kinds of ads, shown in Figure 5-23, make up the bulk of online advertising:

- Pay-per-impression, in which advertisers pay media sites for the number of times they show an ad to a visitor.
- Pay-per-click, in which the media site gets paid each time a visitor clicks on the advertiser's message. This is how paid search ads work.
- Pay-per-acquisition, in which the media site is compensated each time a visitor that they send to an advertiser completes a transaction. Affiliate marketing is an example of a pay-per-acquisition model.

The ads may be shown to every visitor or they may be selectively displayed according to keywords and their relevance to search engines.

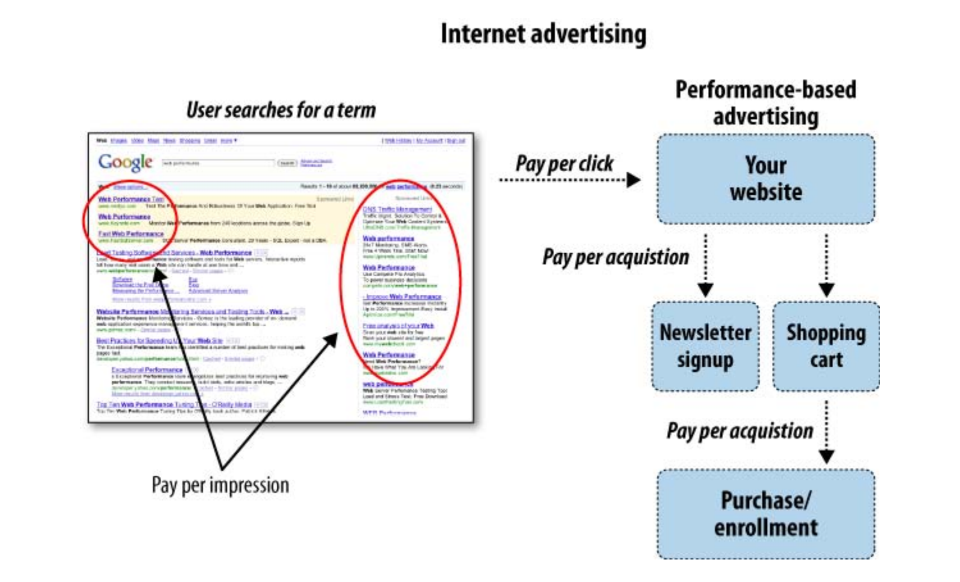


Figure 5-23. The three models of web advertising–pay-per-impression, pay-per-click and pay-per-acquisition

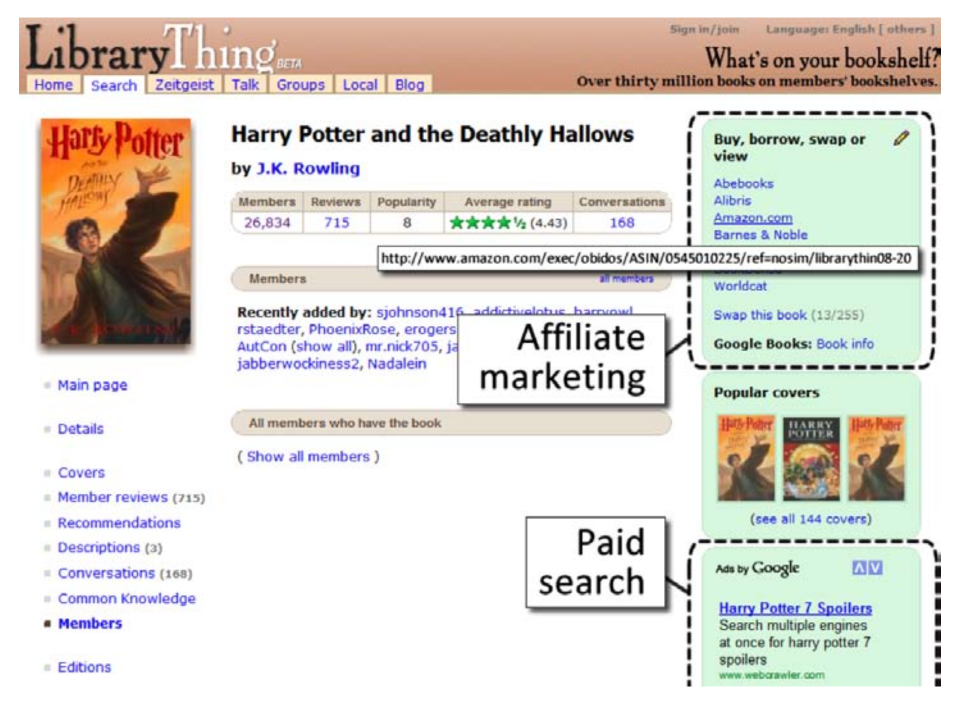
Pay-per-impression advertising is declining, with the possible exception of video advertising. One of the main reasons for this is fraud: marketers prefer to pay for outcomes, and pay-per-click reduces the amount of fraud that they have to deal with. There is also good evidence that picture advertising doesn't encourage users to click on it, so paid search is the most popular form of online advertising today.

Ad payment models vary by advertiser, but are often built around keyword "auctions" and a daily budget for each ad buyer. Ad prices are established by advertisers competing for a keyword.

Google's AdWords, Microsoft's adCenter, and Yahoo!'s Search Marketing display paid ads alongside organic search results. So if you can't optimize your site well enough to qualify for a spot on the first page of their search results, you can still get some real estate on a page of organic results by paying them.

Referrals

A referral occurs when someone mentions your website, which may happen in a blog posting, a mention on a social news aggregator like Digg, a Facebook status message, and so on.



Harry Potter and the Deathly Hallows

by J.K. Rowling

Members	Reviews	Popularity	Average rating	Conversations
26,834	715	8	★★★★½ (4.43)	168

http://www.amazon.com/exec/obidos/ASIN/0545010225/ref=nosim/librarythin08-20

Recently added by: sjohnson416, addictivelotus, harrowl rstaedter, PhoenixRose, erogers AutCon (show all), mr.nick705, ja jabberwockiness2, Nadalein

Harry Potter 7 Spoilers
Search multiple engines at once for harry potter 7 spoilers
www.webcrawler.com

Affiliate marketing

Paid search

Figure 5-24. Affiliate marketing and paid search on book sharing site LibraryThing showing the unique affiliate URL that identifies the referring site

Many online retailers offer affiliate marketing programs that compensate others for sending them traffic. For example, a blogger might recommend books to her readers, and will be compensated by Amazon.com every time someone follows one of her links and buys a book, as shown in Figure 5-24.

For this to work, the referring link includes an affiliate ID—a code that uniquely identifies the referrer so that she can be credited with the referral.

While referrals do come from a somewhat trusted source, paid referrals have a different URL, and many web users can differentiate between a free referral and a paid affiliate marketing link. Paid referrals work well when the source is trusted and users don't mind that the referrer is being compensated, but they're less genuine than simple, unpaid word of mouth.

Linkbacks

There's a form of referral that is specific to blogging. When one blogger references another, the referencing blog alerts the referenced blog that a linkback (or trackback) has occurred. This is a form of attribution that's built into most blogging systems.

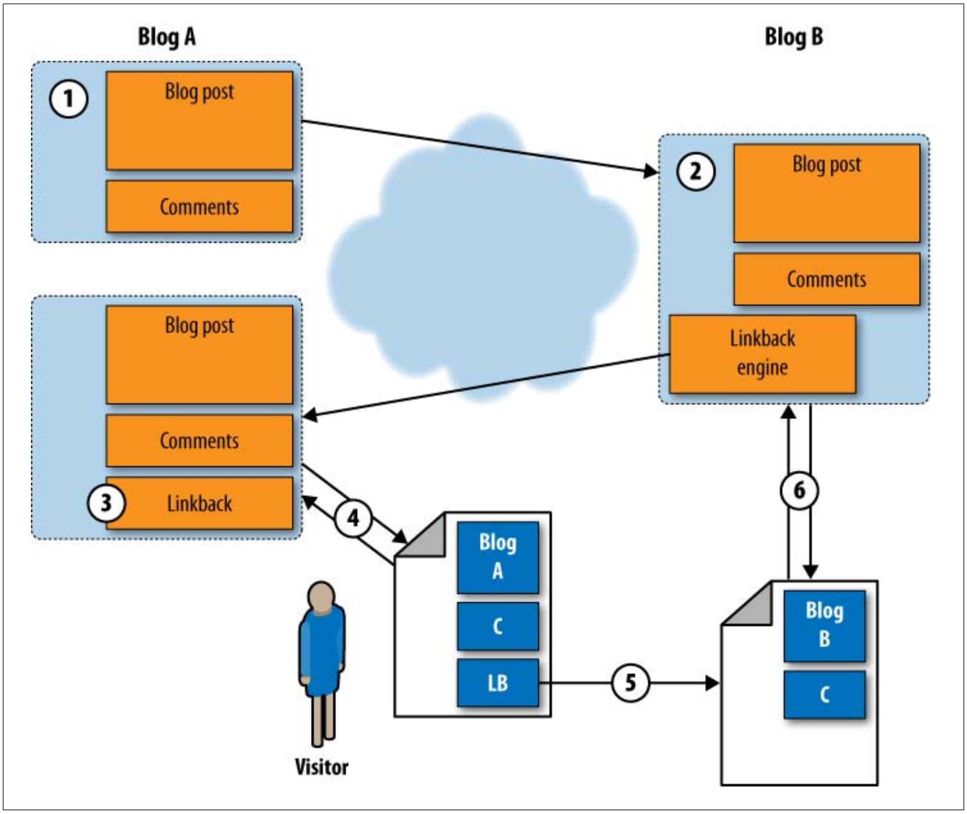
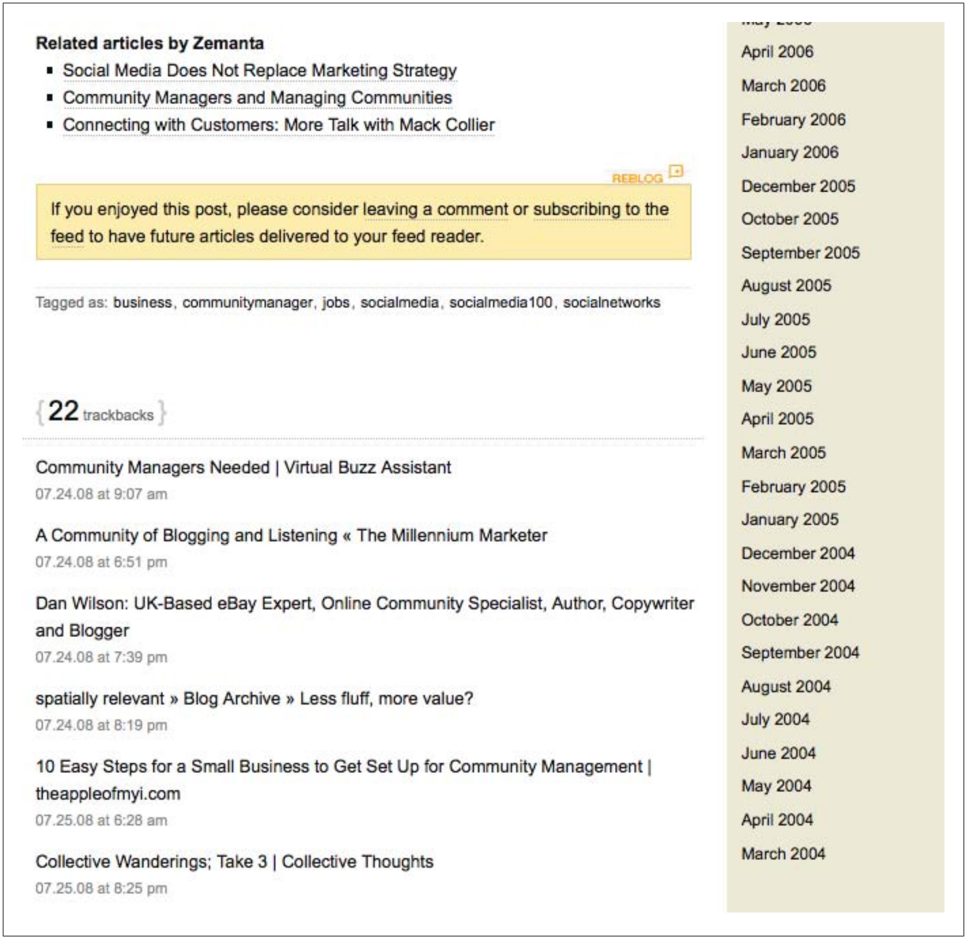


Figure 5-25. How a linkback works

Figure 5-25 shows how a linkback works.

1. The author of Blog A writes a blog post.
2. The author of Blog B mentions A's post in a blog entry he writes, triggering a linkback mechanism on his blogging platform that informs Blog A of the inbound link.
3. Blog A's blogging platform updates the linkback section of the post to include the inbound link from Blog B.
4. A visitor to Blog A reads the post, including the linkback section.
5. The visitor clicks the linkback from Blog B.
6. The visitor sees the post on Blog B that referenced Blog A.

Linkbacks help make blogging dynamic and encourage cross-referencing and attribution of ideas, as shown in Figure 5-26. They allow readers to follow extended conversations between several bloggers.



Related articles by Zemanta

- Social Media Does Not Replace Marketing Strategy
- Community Managers and Managing Communities
- Connecting with Customers: More Talk with Mack Collier

If you enjoyed this post, please consider leaving a comment or subscribing to the feed to have future articles delivered to your feed reader.

Tagged as: business, communitymanager, jobs, socialmedia, socialmedia100, socialnetworks

{ 22 trackbacks }

Community Managers Needed | Virtual Buzz Assistant
07.24.08 at 9:07 am

A Community of Blogging and Listening « The Millennium Marketer
07.24.08 at 6:51 pm

Dan Wilson: UK-Based eBay Expert, Online Community Specialist, Author, Copywriter and Blogger
07.24.08 at 7:39 pm

spatially relevant » Blog Archive » Less fluff, more value?
07.24.08 at 8:19 pm

10 Easy Steps for a Small Business to Get Set Up for Community Management | theappleofmyi.com
07.25.08 at 6:28 am

Collective Wanderings; Take 3 | Collective Thoughts
07.25.08 at 8:25 pm

April 2006
March 2006
February 2006
January 2006
December 2005
October 2005
September 2005
August 2005
July 2005
June 2005
May 2005
April 2005
March 2005
February 2005
January 2005
December 2004
November 2004
October 2004
September 2004
August 2004
July 2004
June 2004
May 2004
April 2004
March 2004

Figure 5-26. 22 linkbacks to an article on chrisbrogan.com

If you're running a blog, linkback analysis is another form of referral analytics you need to track.

The long funnel

Once you've identified visits that came from organic and paid search, advertising campaigns, affiliate referrals, and linkbacks, you're left with direct traffic. Much of this comes from word-of-mouth mentions and social networks—after all, visitors had to hear about you somewhere. However, you can't tell where their first impressions came from, because their visits are the result of several influences you weren't monitoring.

The conversion funnels we've seen so far begin when a visitor first arrives at your site. This isn't an accurate model for the modern Web. Social networks and online communities mean conversion starts long before a visitor reaches your site. Integrating all of these into your understanding of the conversion process is a major challenge for web analysts.

All of these traffic sources represent your "long funnel"—the steps your visitors undergo from initial awareness to conversion—which begins at the conversation prism (by Brian Solis and JESS3, found at *http://theconversationprism.com*).

What happens when users first become aware of your brand on another site affects their conversion once they reach your site. Consider, for example, two visitors from social news aggregator reddit. One visitor may follow a link from the site's home page, while another may read critical comments about your brand within the submission's discussion. Their outcomes on your site are affected by experiences that happened long before they ever saw your home page.

The start of the long funnel is the subject of community monitoring, which we'll look at in Chapters 11 through 14. It's hard to tie community activity to web funnels, in part because of privacy concerns, but long funnel analysis is the holy grail of web analytics and a critical factor to consider in organic traffic acquisition.

Now that we know where users come from, let's figure out what they're doing on your site.

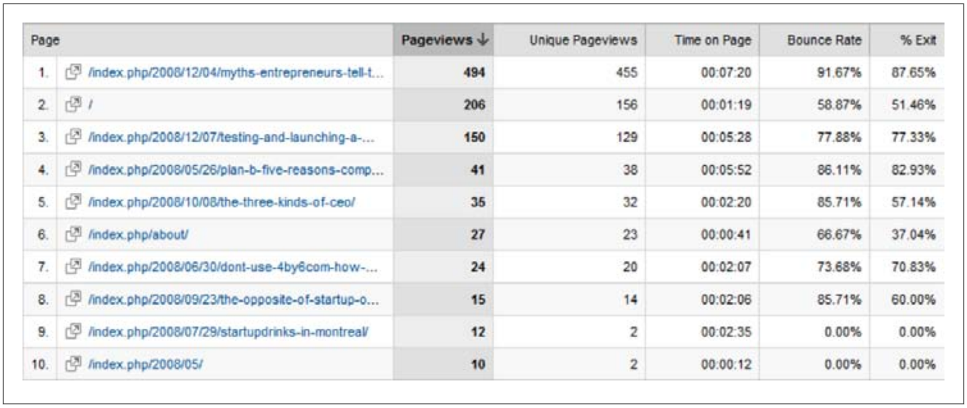
Using the Site: Tracking Your Visitors

You track what your visitors do so that you can identify patterns and improve the site in ways that increase the outcomes you want (such as inviting friends, creating content, or subscription renewals) while reducing those you don't (such as departures or calls to customer support).

Where did they come in?

The first page in a visit is particularly important. Different visits have different first pages, and you can segment outcomes based on the pages at which visitors arrive. You may even have different landing pages for different traffic sources, which lets you tailor your site's message to different segments.

For example, a visitor from reddit might see a message or layout that's designed to appeal to him, while a visitor from Twitter might see a different layout entirely.



	Page	Pageviews ↓	Unique Pageviews	Time on Page	Bounce Rate	% Exit
1.	/index.php/2008/12/04/myths-entrepreneurs-tell-t...	494	455	00:07:20	91.67%	87.65%
2.	/	206	156	00:01:19	58.87%	51.46%
3.	/index.php/2008/12/07/testing-and-launching-a-...	150	129	00:05:28	77.88%	77.33%
4.	/index.php/2008/05/26/plan-b-five-reasons-comp...	41	38	00:05:52	86.11%	82.93%
5.	/index.php/2008/10/08/the-three-kinds-of-ceo/	35	32	00:02:20	85.71%	57.14%
6.	/index.php/about/	27	23	00:00:41	66.67%	37.04%
7.	/index.php/2008/06/30/dont-use-4by6com-how-...	24	20	00:02:07	73.68%	70.83%
8.	/index.php/2008/09/23/the-opposite-of-startup-o...	15	14	00:02:06	85.71%	60.00%
9.	/index.php/2008/07/29/startupdrinks-in-montreal/	12	2	00:02:35	0.00%	0.00%
10.	/index.php/2008/05/	10	2	00:00:12	0.00%	0.00%

Figure 5-27. Top entry pages in Google Analytics, with corresponding time on page and bounce rate

To understand what visitors first saw, run a Top-N report on entry pages, such as the one shown in Figure 5-27, to see some fundamental facts about your website:

- The total number of page views for the page.
- The number of *unique* page views for the page. Subtracting total page views from unique page views shows you how many visitors viewed a page more than once.
- The time visitors spent on the page before moving on.
- How many visitors arrived at the page and left immediately.
- How many visitors ended their visits on the page.

Armed with just this information, you already know a great deal about which parts of your site are working.

- If your objective is to engage visitors, you want create more pages like those that make visitors linger (lingering indicates visitors found the content useful) and that encourage them to go elsewhere within the site.
- If the page has a high bounce rate, you may need to redesign the page to convince visitors to stay.
- If users exit from this page, you'll need content that draws them back into the site for more than one page by suggesting they enroll or view additional content.

Places and tasks revisited

To understand visitor activity beyond the first page, start thinking in terms of places and tasks. Your visitors have arrived at a place on your site, such as the landing page, and can do various things in that place. They can remain there and interact, they can move to another place, or they can complete some kind of task. Figure 5-28 shows an example of a places and tasks diagram for a social news aggregator.

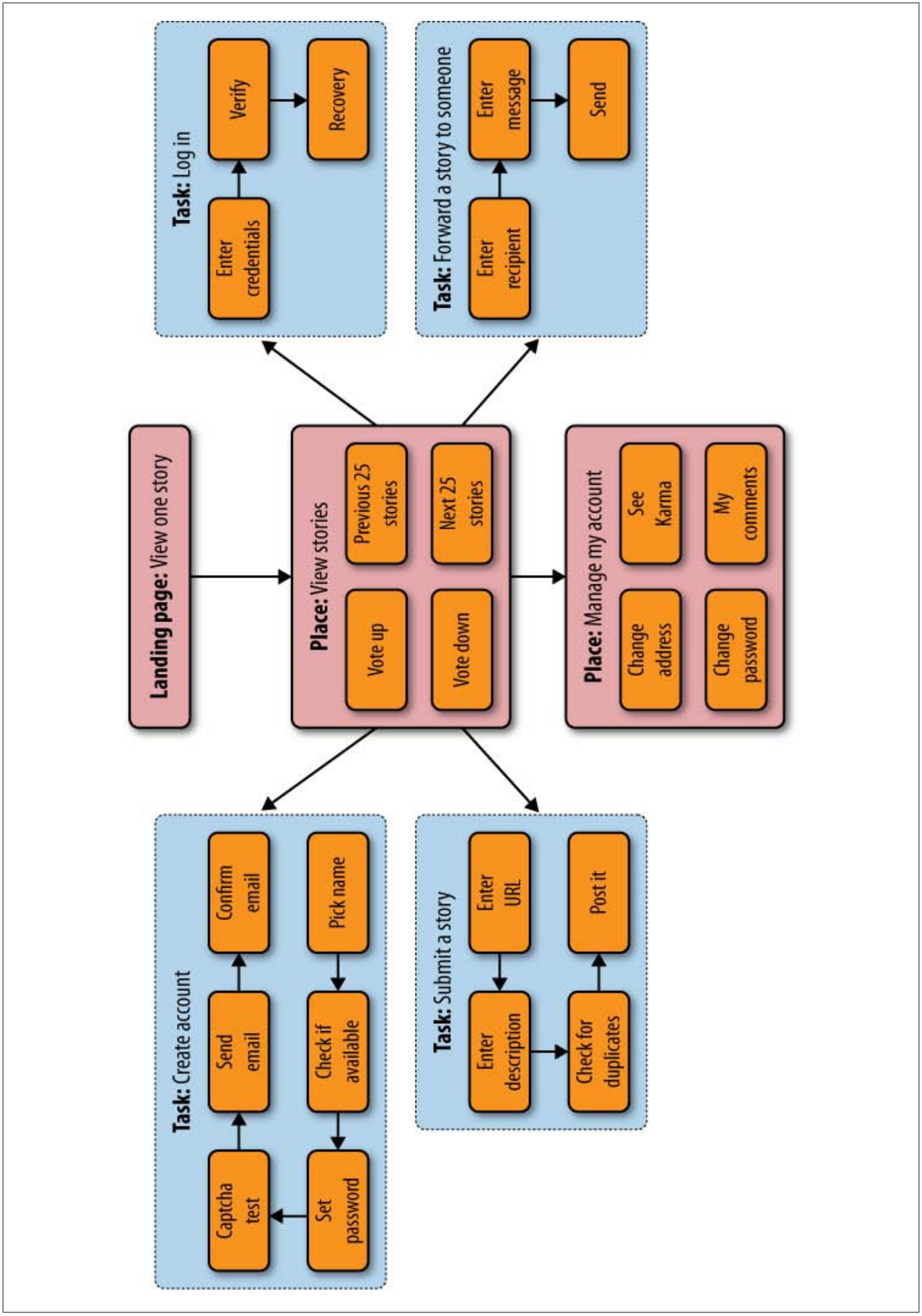


Figure 5-28. An example of places and tasks in a social news aggregator site such as reddit or Digg

The places and tasks model lets you define all the outcomes you want on your site so that you can collect data on them.

For *places*, you care about all-purpose metrics such as exit rate and time spent in the place, as well as where visitors went after each place. You may also have site-specific metrics (for example, number of story votes a visitor placed) that measure productivity, engagement, or value derived from the user's visit.

For *tasks*, you care about progress toward an outcome, and abandonment at each stage. If, for example, a visitor sets out to forward a story to someone, you need to track the many ways in which the task can go awry: the entry of an invalid email address, the visitor abandoning the process, the mail not being delivered, or invitees not acting on mail they've received.

Try to assign each task a dollar value. For e-commerce transactions, this may be the money the visitor spent, but for other kinds of task it may vary. If you have an estimate of the lifetime value of a customer, a completed invitation may have a specific dollar value based on that estimate.

Assigning a dollar value is good for three reasons. First, it makes the somewhat esoteric analytics data concrete: everyone knows what a dollar is. Second, it forces you to quantify how the task outcome helps the business. Third, it makes executives pay attention.

Other examples of tasks you want to track include:

- Adding a contact to a CRM database
- Commenting on a blog post
- Subscribing to an RSS feed or mailing list
- Getting directions or contact information
- Downloading product literature

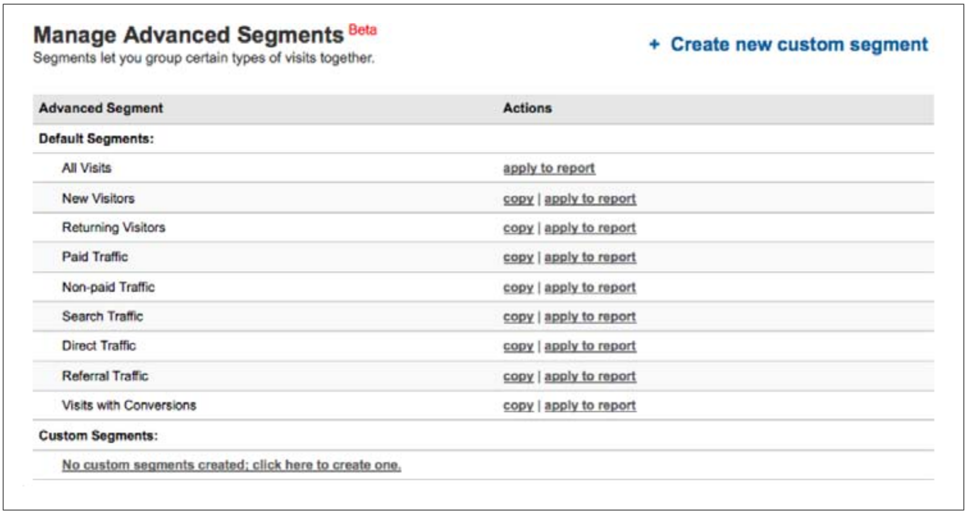
You need a map of the places and tasks, with key performance indicators (KPIs) for each action. Once you know what these are, it's time to test and segment your website to maximize those KPIs.

Segmentation

All visits are not equal. You need to compare different visitor groups to see which do better or worse than others. Table 5-1 shows some examples of segments by which you may want to analyze your traffic.

Table 5-1. Different ways to segment data

Segment	Example
Demographic segments	The country from which the visitor arrived
Customer segments	First-time versus returning buyers
Technographic segments	Visitors using Macintosh versus Windows operating systems
“Surfographic” segments	Those visitors who surf daily versus those who only do so occasionally
Campaign segments	Those visitors who saw one proposition or offer versus another
Promotion types	Those who saw a banner ad versus a paid search
Referral segments	Those visitors who came from one blog versus those who came from another
Content segments	Those visitors who saw one page layout versus another



Manage Advanced Segments Beta

Segments let you group certain types of visits together.

+ Create new custom segment

Advanced Segment	Actions
Default Segments:	
All Visits	apply to report
New Visitors	copy | apply to report
Returning Visitors	copy | apply to report
Paid Traffic	copy | apply to report
Non-paid Traffic	copy | apply to report
Search Traffic	copy | apply to report
Direct Traffic	copy | apply to report
Referral Traffic	copy | apply to report
Visits with Conversions	copy | apply to report
Custom Segments:	
No custom segments created; click here to create one.	

Figure 5-29. Managing advanced segments in Google Analytics

> Some segmentation data isn’t available directly through analytics, and you may need to use customer surveys or enrollment questions to fit visitors into a particular segment—such as those over the age of 40—so that you can analyze a particular KPI or outcome by that segment.

While analytics tools offer many built-in segments, you can create custom segments according to user-defined data or other fields within an analytics tool to slice up traffic in ways that matter to your business, as shown in Figure 5-29.

By comparing KPIs across segments, you can identify what’s working and what’s not. You might decide that one advertisement works while another doesn’t, or that one offer encourages visitors to linger while another makes them leave.

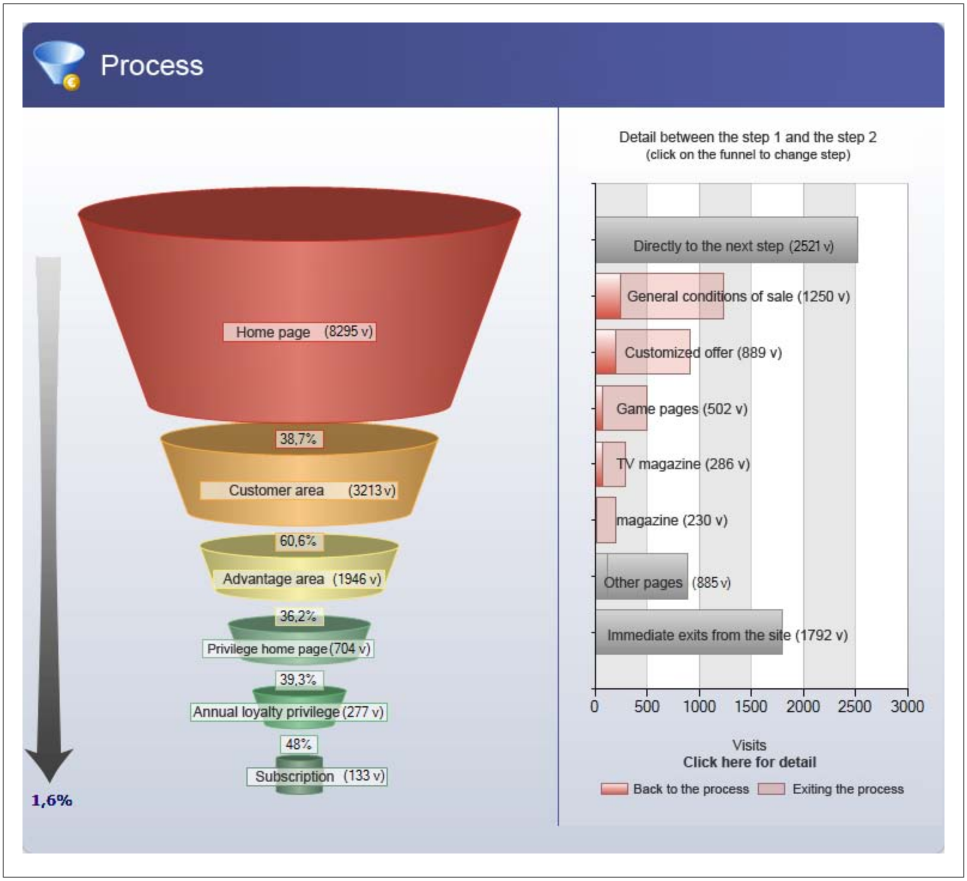


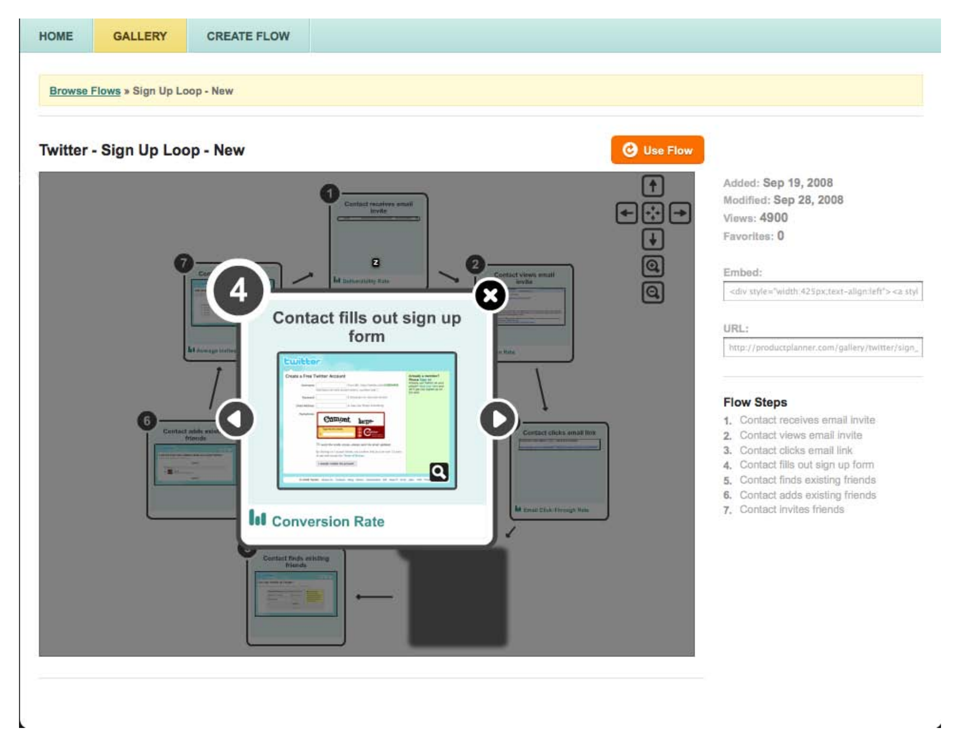
Figure 5-30. A subscription funnel shown in XiTi Analyzer

Goals

The ultimate KPI is the goal, or outcome, of a visit. Within your analytics package, you identify the outcomes you want, such as a purchase confirmation page or an enrollment screen, as well as any steps leading up to that outcome. This is often visualized using a funnel graph (Figure 5-30).

Funnels work well, but they're myopic. They make it hard to identify where users are going, and don't take into account reentry into the process. They also focus on web activity alone, while many websites have goals that include email messages, subscriptions, and other steps that can't easily be captured by simple web requests.

As web analytics tools adapt to today's more distributed conversion processes and usage patterns, we'll likely see places-and-tasks models that track KPIs for each step in a process. The KISSMetrics ProductPlanner (shown in Figure 5-31) is a great resource for determining which metrics to capture for many popular web design patterns.


HOME
GALLERY
CREATE FLOW
Browse Flows » Sign Up Loop - New
Use Flow
4
Contact fills out sign up form
Conversion Rate
1
Contact receives email invite
2
Contact views email invite
Contact clicks email link
Email Click-Through Rate
Contact finds existing friends
6
Contact adds existing friends
7

Twitter - Sign Up Loop - New

Added: Sep 19, 2008
Modified: Sep 28, 2008
Views: 4900
Favorites: 0

Embed: <div style="width:425px;text-align:left"><a styl

URL: http://productplanner.com/gallery/twitter/sign_

Flow Steps

1. Contact receives email invite
2. Contact views email invite
3. Contact clicks email link
4. Contact fills out sign up form
5. Contact finds existing friends
6. Contact adds existing friends
7. Contact invites friends

Figure 5-31. Twitter's sign-up process with the associated metric important for this step in the funnel on ProductPlanner

Putting it all together

The core of web analytics is knowing what outcomes you want, describing KPIs that represent those outcomes, scoring various visitor segments against those KPIs, and testing various combinations of content and segments to maximize those KPIs. Despite your best designs and offers, however, visitors often won't do what you wanted—they'll leave. Knowing why they left—and where they went—is equally important.

Leaving the Site: Parting Is Such Sweet Sorrow

While you can't see a visitor's activity once she's left your site, you can tell a lot about her motivations by looking at the last things she saw. You can also intercept parting visitors with a feedback form to try to learn more about their departure, which we'll look at more closely in Chapter 7.

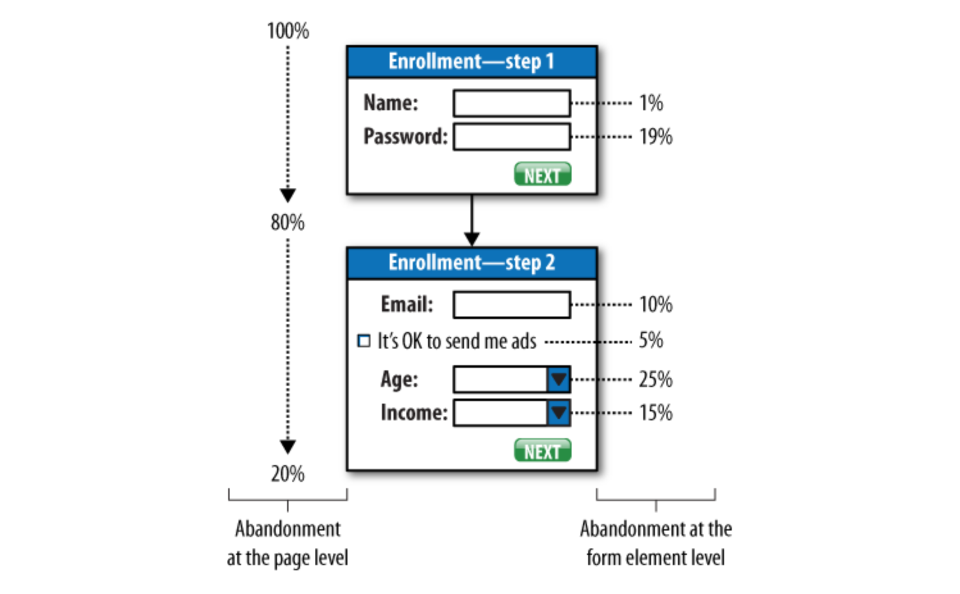


Figure 5-32. While web analytics looks at page-by-page abandonment, visitors usually abandon a process at a specific form element

Abandonment and bounce rate

Abandonment happens when people leave your site before doing something you wanted them to. They may simply leave out of boredom, because of an error or performance issue, or because they changed their minds about completing a particular task.

In recent years, analytics tools have started to look within a single page, to the form elements on that page, in an effort to identify which components of a page drive visitors away.

Consider, for example, a form that asks for personal information such as age. If you analyze the abandoned page at the page level, you won't know that it's a specific form element that's the problem. However, if you use web interaction analytics (WIA) tools to do form-level analytics, you'll have a much better understanding of how visitors are interacting with the site, as shown in Figure 5-32.

We'll look at WIA in more detail in Chapter 6. For now, remember that abandonment is a coarse measurement and there are ways to perform more detailed analyses of where you're losing visitors.

Bounce rate is a close cousin to abandonment. Visitors "bounce" when they enter your site and leave immediately. Bounce rate can mean that visitors didn't find what they expected to find, which may be a sign that search engines aren't properly indexing your content or that you're buying the wrong search keywords. For news sites, blogs, and

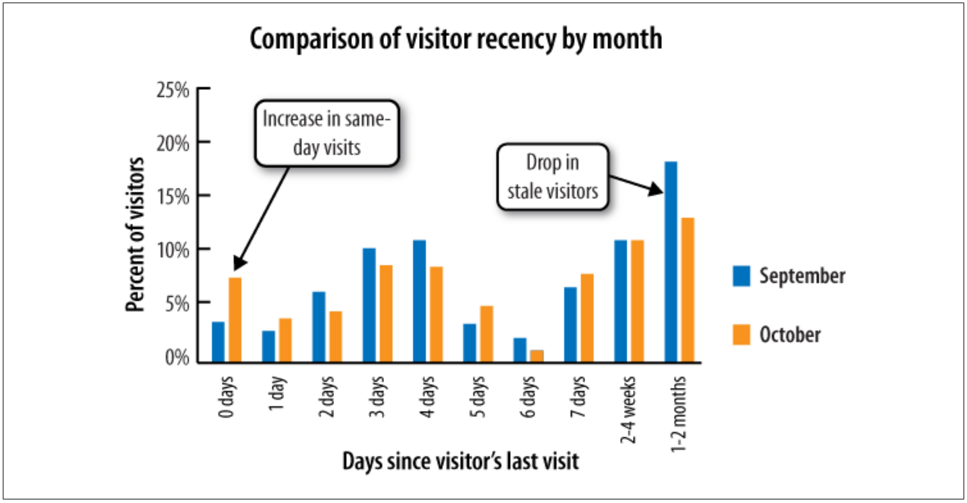


Figure 5-33. A comparison of last visits by user population, showing a reduction in visitor disengagement and attrition from one month to the next

certain content sites, a high bounce rate may be normal—visitors come to read one new article, then leave. Focus on encouraging visitors to bookmark your site, return, share content with others, and subscribe. You should care more about lifetime visitor loyalty than about one-page visits.

Attrition

Unlike bounces and exits, which happen at the end of a visit, attrition happens when your relationship with a visitor grows stale. There are many ways to measure attrition. You might look at the percentage of returning visitors, but if your site is acquiring new visitors, a declining percentage of returning visitors might simply indicate rapid growth.

A better measure of attrition is the number of users that haven't returned in a given period of time. By comparing attrition for two time periods, as shown in Figure 5-33, you can tell whether things are getting better or worse, and whether you're successfully retaining visitor attention.

Many community sites celebrate the number of visitors they have without considering the engagement of those visitors. Any site that depends on user activity and returning visitors must look at attrition carefully and make concerted efforts to reengage users who haven't visited the site in a long time.

Desirable Outcomes

While abandonment is a bad thing, not all departures are bad.

Visitors might leave your site once they've done what they wanted. Perhaps they've successfully completed a purchase, at which point they've finished their transactions.

Your work's not done: you need to learn more about why they bought what they did through surveys, and to encourage them to enroll and return.

Visitors may also have enrolled or subscribed, preferring to hear from you via RSS feed or email subscription. If this is consistent with your business model, it's a good thing. You do, however, still need to monitor whether they're receiving and acting on email messages that you send them or returning when you add new content to your RSS feed to ensure that the enrollment "stuck."

Tracking referrals and ad clicks

You may have sent the visitor elsewhere through a paid referral or an ad, in which case you're making money from the departure. An ad referral is a complex dance between your media website, the ad serving network, and the advertiser to whom a visitor eventually connects. Figure 5-34 shows how this works.

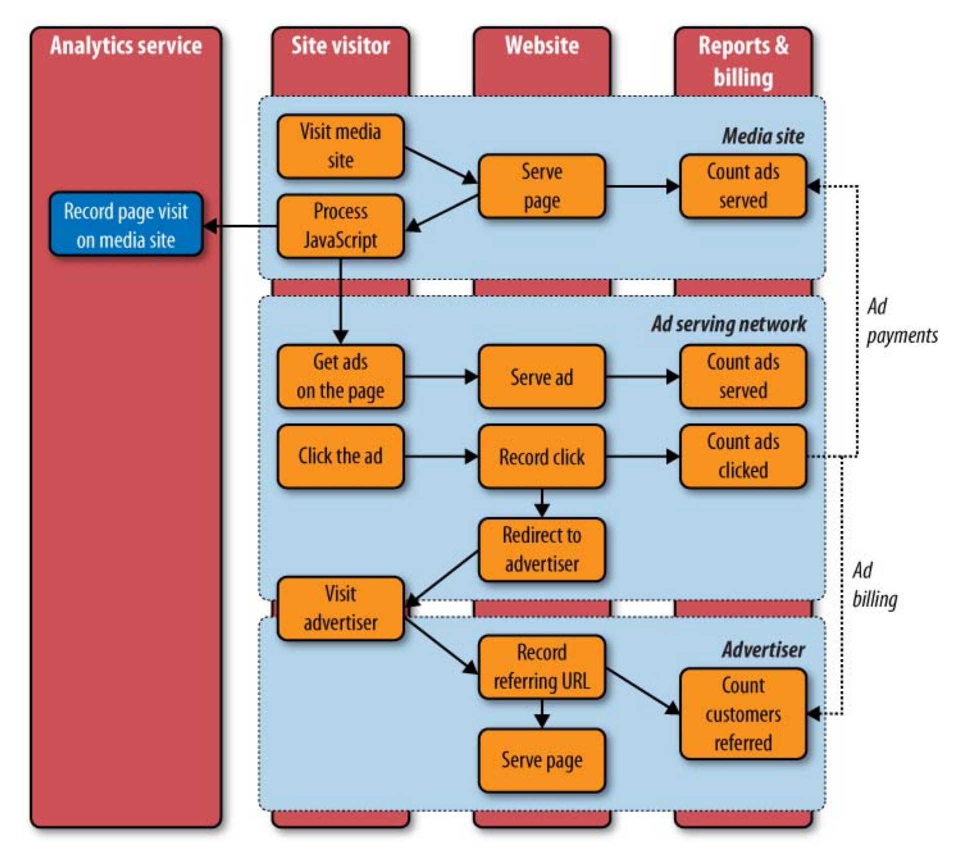


Figure 5-34. How an ad network works with a media site to serve ads and count clicks

Your web page contains a reference to content from an ad inventory provider. This content includes a link to the ad network (such as *http://pagead2.googlesyndication.com/pagead/iclk*) that contains a unique identifier for the ad and the destination site's URL.

The visitor's browser loads the ads that are on the page from an ad serving network, and if the visitor clicks the ad, the network records the click before sending the visitor to the destination (the advertiser's site).

This model doesn't let you count how many ads visitors clicked. As a result, you don't have your own numbers to compare to the ones the ad network tells you—in other words, you're letting the company that's paying the bills say how much it owes you. Not good.

To resolve this issue, you can add JavaScript to your pages to let you know when a visitor clicks an ad, then compare your measurements to the advertiser's numbers in order to resolve any disputes. This additional step is reflected in Figure 5-35 (visit *www.google.com/support/analytics/bin/answer.py?answer=55527&ctx=sibling* for details on using this for the Google Analytics JavaScript).

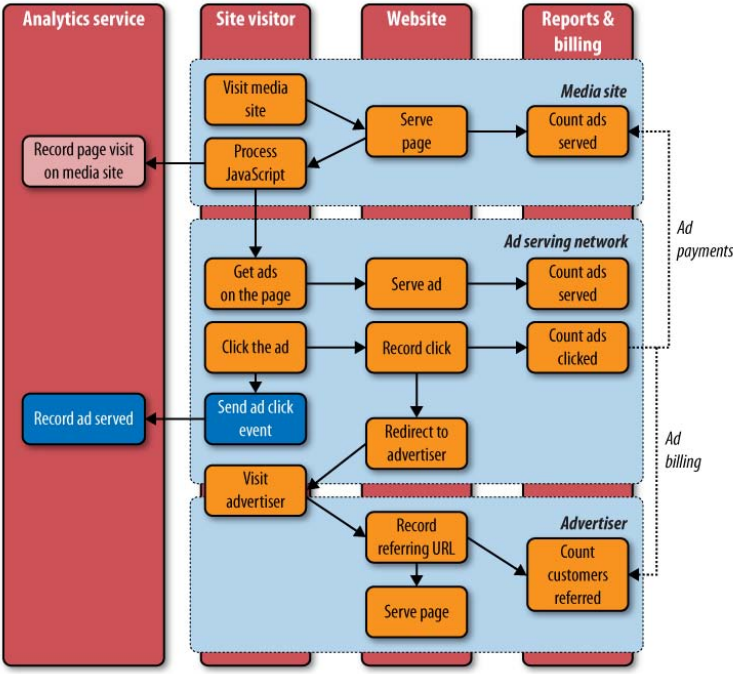


Figure 5-35. Capturing ad click events and sending them to your analytics service lets you count clicks on embedded ads

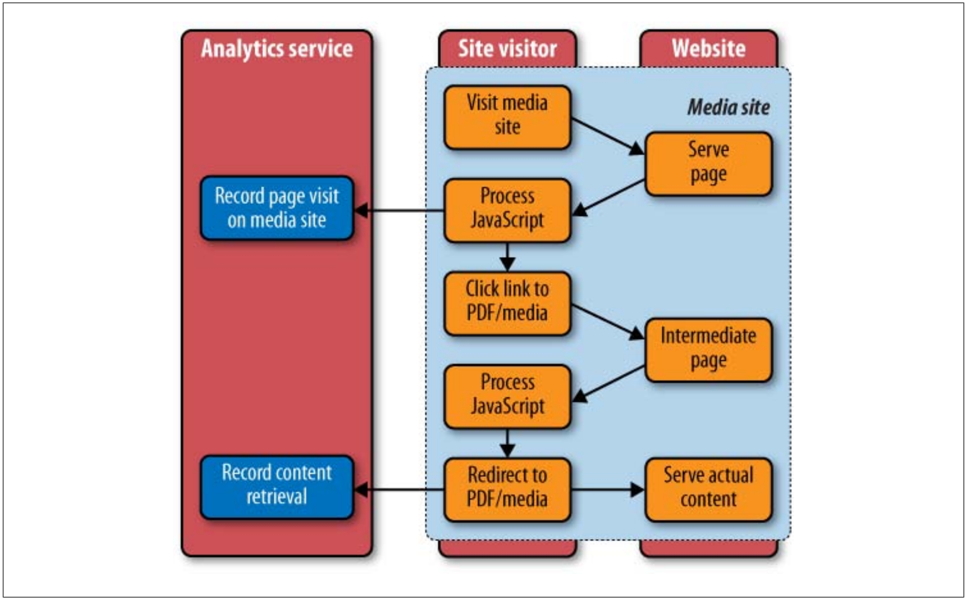


Figure 5-36. Using an intermediate page with JavaScript to capture clicks for non-HTML content

If you're running a sponsored site where advertisers pay for a time period or banner placement, you may be inserting advertising or banners on behalf of your sponsors. To track clicks you can either first redirect the visitor to an intermediate page that records the action or you can use JavaScript to send a message to your analytics system.

You can also do this if you're letting visitors download non-HTML content, such as Adobe PDF documents, since the document you're sending them can't execute JavaScript to let you know it was downloaded. Analytics tools can track requests for this intermediate page, giving you ad-tracking data as shown in Figure 5-36 (visit *http://www.google.com/support/analytics/bin/answer.py?answer=55529&ctx=sibling* for information on using this for Google Analytics).

Implementing Web Analytics

Measuring arrivals, visitor activity, and departures is the lifeblood of a web business. Without this information, you can't make smart decisions about your content, marketing, or business model. You can collect basic analytics data, such as bounce rate and visits, with just a few minutes' work.

More advanced deployment—tracking goals, building custom segments, and tagging content—takes work, however. And if you want to use analytics for accounting data (such as daily orders) as well as for site optimization, you'll probably have to work with the development team to extract additional information from backend databases.

There are free (or cheap) analytics solutions from the big advertising vendors, and Google Analytics has done a tremendous amount to make web operators aware of analytics.

Whatever you're doing, your implementation will have six basic steps:

1. Defining your site's goals
2. Setting up data capture
3. Setting up filtering
4. Identifying segments by which to analyze goals
5. Tagging page content
6. Verifying that everything is working

Let's look at the steps you'll need to take to implement web analytics.

Define Your Site's Goals

Your first step is to understand and map out your web business.

That might sound like a platitude, but it's an essential step in the process. Grab a whiteboard and some markers, and draw out your site using the places-and-tasks model outlined above. In each place, list what makes a visitor "productive." For each task, identify the steps a visitor needs to take to accomplish the task, and the metrics you need to collect in order to track those steps.

Now go through the tasks and see which ones drive your business (this will depend heavily on which of the four kinds of sites you're running). Each of these tasks will become a goal funnel in your analytics system, and you should try to assign a monetary value to each goal, even if it's something as simple as "getting contact information."

Once you've identified the important places and tasks, list the assumptions you're making. Somewhere in your business model, you're assuming that a certain number of visitors perform a given task. Write this number down; if you don't know what it is, guess. Do the same thing for every assumption you're making—how many people will exit from a particular place, how many will be referred by a particular kind of site, and so on.

You'll quickly realize that there are other things you'd like to know. How long does a place take to refresh? What do visitors think of this content? Where in the form do visitors leave? Did the email invitation bounce? You can collect all of these metrics—though sometimes you'll need third-party tools or coding to collect them. In the long term, you should try to integrate these metrics into a single, comprehensive view. For now, however, let's stick with simple web analytics. We'll come back to other kinds of monitoring later in the book.

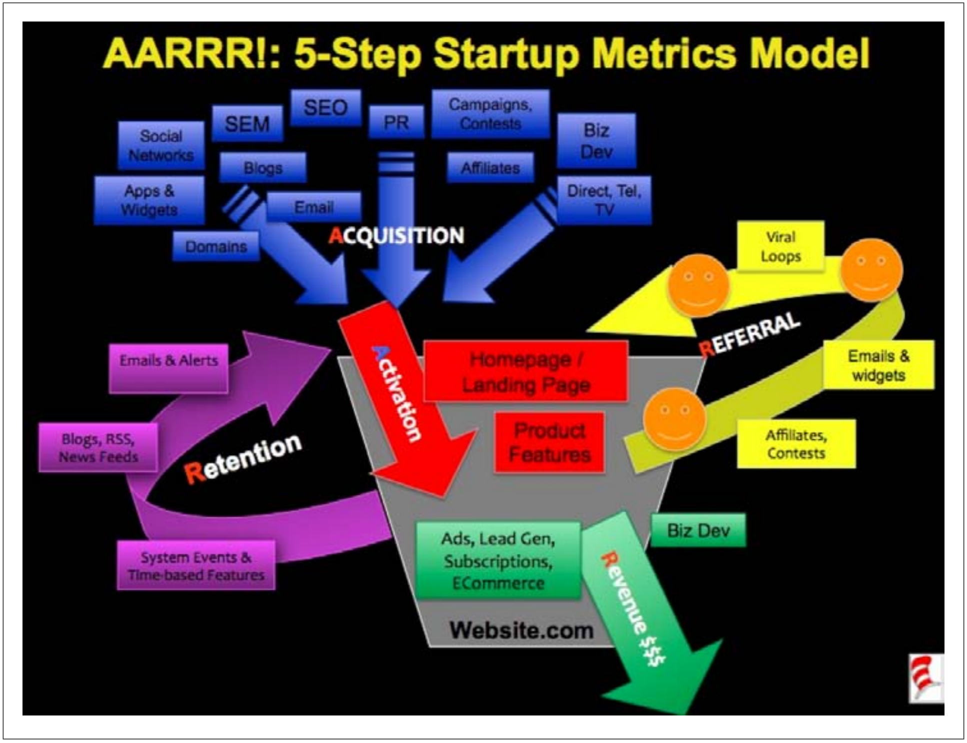


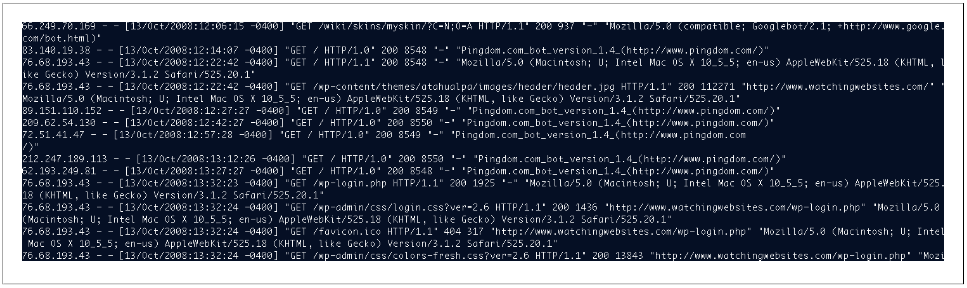
Figure 5-37. Dave McClure's 5-Step Startup Metrics Model

If you need some inspiration for your places and tasks, particularly if you're running a web startup, check out Dave McClure's "Startup Metrics for Pirates – AARRR!" at *www.slideshare.net/Startonomics/startup-metrics-for-pirates-presentation*.

McClure's model, shown in Figure 5-37, shows the many sources from which traffic can arrive (acquisition, retention, and referral), the places within the site (activation), and the tasks visitors can perform, such as enrollment (retention), inviting others (referral), and spending money (revenue). It provides a valuable starting point for thinking about your web business.

Notice that you don't need a web analytics account yet—you're still nailing down the really important question: *what is my site trying to encourage people to do?* Set aside a day for this process—it's a lot of work that often involves a surprising number of people in the company. Mapping your site is a time-consuming and iterative process; if it seems too daunting, pick a subset of the site at first, such as a particular application or a certain domain.

Once you know which metrics you want to collect for each place and task on your site, you need to define them within the analytics package you're using.



```
66.249.70.169 - - [13/Oct/2008:12:06:15 -0400] "GET /wiki/skins/myskin/?C=N;O=A HTTP/1.1" 200 937 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.
com/bot.html)"
83.140.19.38 - - [13/Oct/2008:12:14:07 -0400] "GET / HTTP/1.0" 200 8548 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com/)"
76.68.193.43 - - [13/Oct/2008:12:22:42 -0400] "GET / HTTP/1.1" 200 8548 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_5_5; en-us) AppleWebKit/525.18 (KHTML, l
ike Gecko) Version/3.1.2 Safari/525.20.1"
76.68.193.43 - - [13/Oct/2008:12:22:42 -0400] "GET /wp-content/themes/atahualpa/images/header/header.jpg HTTP/1.1" 200 112271 "http://www.watchingwebsites.com/" "
Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_5_5; en-us) AppleWebKit/525.18 (KHTML, like Gecko) Version/3.1.2 Safari/525.20.1"
89.151.110.152 - - [13/Oct/2008:12:27:27 -0400] "GET / HTTP/1.0" 200 8549 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com/)"
209.62.54.130 - - [13/Oct/2008:12:42:27 -0400] "GET / HTTP/1.0" 200 8550 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com/)"
72.51.41.47 - - [13/Oct/2008:12:57:28 -0400] "GET / HTTP/1.0" 200 8549 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com
/)"
212.247.189.113 - - [13/Oct/2008:13:12:26 -0400] "GET / HTTP/1.0" 200 8550 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com/)"
62.193.249.81 - - [13/Oct/2008:13:27:27 -0400] "GET / HTTP/1.0" 200 8548 "-" "Pingdom.com_bot_version_1.4_(http://www.pingdom.com/)"
76.68.193.43 - - [13/Oct/2008:13:32:23 -0400] "GET /wp-login.php HTTP/1.1" 200 1925 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_5_5; en-us) AppleWebKit/525.
18 (KHTML, like Gecko) Version/3.1.2 Safari/525.20.1"
76.68.193.43 - - [13/Oct/2008:13:32:24 -0400] "GET /wp-admin/css/login.css?ver=2.6 HTTP/1.1" 200 1436 "http://www.watchingwebsites.com/wp-login.php" "Mozilla/5.0
(Macintosh; U; Intel Mac OS X 10_5_5; en-us) AppleWebKit/525.18 (KHTML, like Gecko) Version/3.1.2 Safari/525.20.1"
76.68.193.43 - - [13/Oct/2008:13:32:24 -0400] "GET /favicon.ico HTTP/1.1" 404 317 "http://www.watchingwebsites.com/wp-login.php" "Mozilla/5.0 (Macintosh; U; Intel
 Mac OS X 10_5_5; en-us) AppleWebKit/525.18 (KHTML, like Gecko) Version/3.1.2 Safari/525.20.1"
76.68.193.43 - - [13/Oct/2008:13:32:24 -0400] "GET /wp-admin/css/colors-fresh.css?ver=2.6 HTTP/1.1" 200 13843 "http://www.watchingwebsites.com/wp-login.php" "Mozi
```

Figure 5-38. A standard httpd logfile

You don't need to configure monitoring of places much, since the default analytics metrics, like bounce rate, time on page, and page views tell you most of what you need to know. If users can perform simple on-page actions, such as upvoting submissions or paginating through content, you may need to record these in-page actions with your analytics tool.

When it comes to monitoring tasks, simply tell the analytics tool what the "goal" page is—for example, a payment confirmation screen—and the pages leading up to that goal.

Set Up Data Capture

Now that you know what you want to capture, it's time to set up the analytics system. There are several ways to deploy web analytics, from server logs to JavaScript-based collection.

Server logs

ELF files that web servers generate contain a limited amount of data, but if you're running a website behind a firewall or have specific privacy restrictions, they may be your best and only source of visitor information.

Popular web servers like Apache httpd and Microsoft IIS generate ELF-formatted logs (shown in Figure 5-38). These logs contain text strings, separated by spaces, that analytics tools import and reassemble into page requests and user visits.

Here's an example of a single hit recorded in an Apache server log:

```
10.100.3.200 - - [28/Mar/2009:11:50:55 -0400] "GET
  / HTTP/1.1" 200 53785 "http://twitter.com/seanpower"
  "Mozilla/5.0 (Windows; U; Windows NT 6.0; en-US; rv:1.9.0.7)
  Gecko/2009021910 Firefox/3.0.7" 125763
```

Here's a similar hit recorded in a Microsoft IIS server log:

```
2008-08-12 20:05:34 W3SVC216049304 10.0.0.1 GET /
WebUISupportFiles/images/bg_page.gif - 80 - 10.0.0.1 Mozilla/4.0+
(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+2.0.50727;+InfoPath.2) 200 2921
```

Despite some differences in formatting and extra characters, server logs from various web servers are essentially the same. Each string in these logs tells you something about the visitor. For example:

- The *client-ip* field is the IP address of the visitor. You may be able to resolve the address to the organization to which it is registered; you can also look it up using a geolocation database to find out where the IP address is physically located.
- The *cs-referrer* field shows the site that has referred the traffic to you.
- The URI may have special information, such as a campaign identifier (for example, *www.watchingwebsites.com/alnk&campaign_id=130*) or other information against which you can segment visitors.
- The *user agent* field shows information on the browser and operating system the user is running.
- The *HTTP status code* shows errors that may have occurred in the request. However, you can't rely too heavily on HTTP status codes—a "200 OK" message may simply indicate that an apology page was delivered correctly. We'll look at performance and availability monitoring in Chapters 8 through 10.

When using web logs, the analytics tool retrieves logfiles from servers at regular intervals and processes the data they contain. The system usually deletes the logfiles after parsing them. Figure 5-39 shows how logfile-based analytics work.

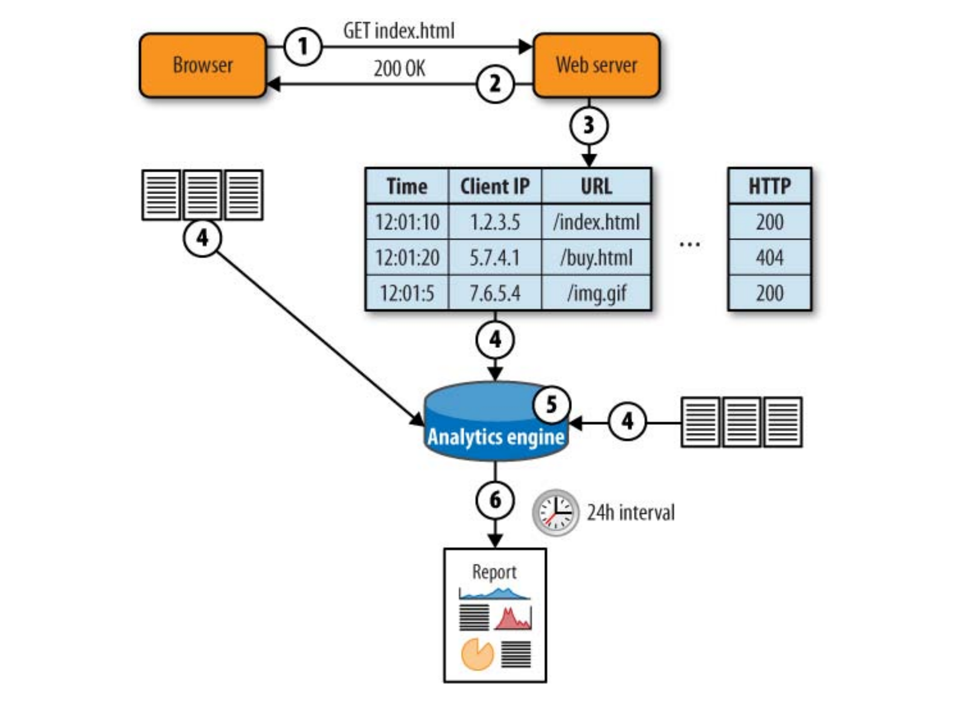
1. A visitor's browser requests an object from the server. The browser provides a variety of data, such as referring URL and browser type, along with the request.
2. The web server responds with the requested object, as well as information such as object size and compression.
3. The web server writes a line to the logfile describing the request and response.
4. Logfiles from all of the web servers are copied to the analytics tool at regular intervals.
5. The analytics engine stores and aggregates all of the logfile data across all of the servers.
6. At regular intervals—often daily—the analytics engine generates a series of reports.

In more modern logfile analytics, the operator can request reports dynamically from the engine without waiting until a reporting period is complete.

Here are some of the advantages of a server log capture approach:

Better data privacy
: You own the data. You don't need a third party, such as an analytics service, to generate reports. This lets you control where the data goes and who can see it.

Useful for internal websites
: To use a third-party analytics service, you need to be able to send data outside of your organization. If your users' browsers can't reach the Internet, you can't use



Time	Client IP	URL		HTTP
12:01:10	1.2.3.5	/index.html		200
12:01:20	5.7.4.1	/buy.html	...	404
12:01:5	7.6.5.4	/img.gif		200

Figure 5-39. Web analytics implementation using logfiles as a data source

these services. This is particularly common for intranets and other in-house environments.

Works when JavaScript won't
Logs may also be useful if your visitor community doesn't use JavaScript, since other analytics collection approaches that rely on client JavaScript won't work. This is common with some mobile devices.

Pages load faster
If you rely on client-side JavaScript to monitor pages, your pages will take longer to load. By analyzing logs instead, you won't add to page delay.

Easier to merge with other data
The server may generate additional logs that contain useful data such as CPU load or custom metrics. This data can be interleaved with HTTP requests according to timestamps. If you don't have web server logfiles, you can't merge these other logs with them.

There are some disadvantages, however, to relying on server logs for capture:

Implementing them will require the efforts of more people

IT and marketing departments need to work together to deploy log-based analytics. This can mean implementation delays if analytics isn't IT's top priority (and it seldom is!).

Log analysis is a lot of work

As traffic grows, the amount of processing needed to parse web logs grows alongside it. On large sites, a day's web logs can consume hundreds of megabytes. Logging every request can also consume precious server resources, slowing down the site even further when it's busy. And if you don't plan carefully, it can take more than a day to process a day's web activity, meaning you're constantly playing catch-up while your analytics get increasingly out of date.

It may not be possible for you

Depending on your hosting environment, you may not be able to get the web server logs in order to process them.

Limited visibility

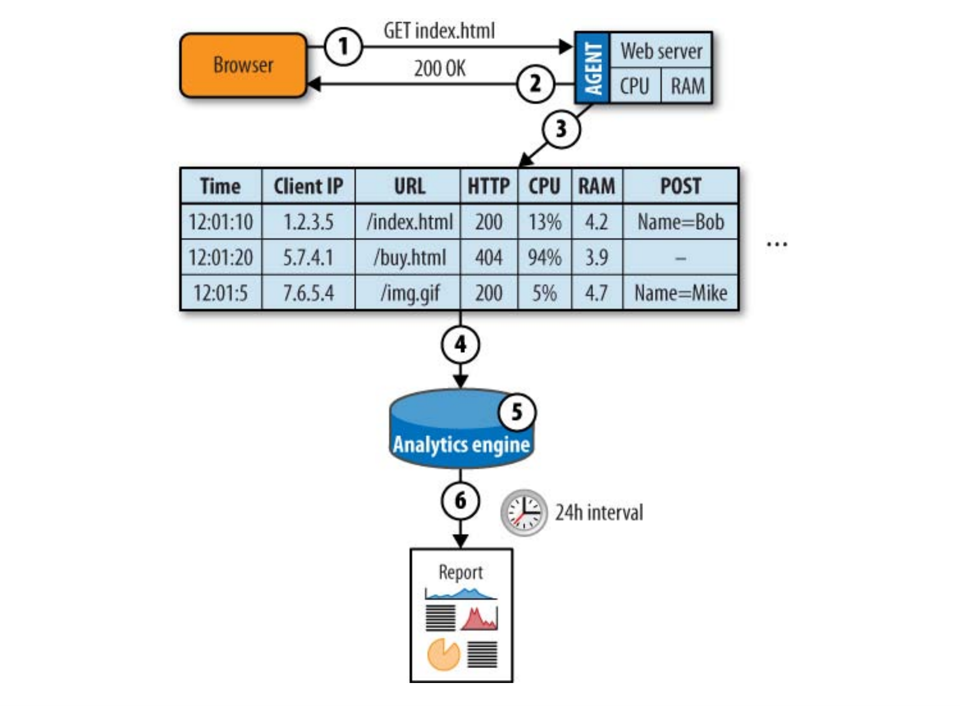
Log-based analytics can only report what's in the log files. Browser-side actions and information about the browser itself, such as cookies, screen resolutions, coordinates of user clicks, and so on, can't be collected from logs. Furthermore, logfiles don't collect POST parameter information unless it's stored within the URI stem as parameters, which can pose a major security risk when users forward URIs to others.

Most of the web analytics industry has moved beyond web logs to client-side collection models unless they have no alternative but to use web log analysis.

Server agents

If you need more data than a web log contains, but absolutely *must* rely on the server itself to collect that data, you can use server agents. These are programs that run on the server and record transaction information, often with additional application context. These agents may also help with logfile collection by pushing logs to a centralized platform. Figure 5-40 shows an example of this.

1. A visitor's browser requests an object from the server. The browser provides a variety of data, including referring URL and browser type, along with the request.
2. The web server responds with the requested object, as well as information such as object size and compression.
3. The agent assembles the request and response data, as well as system health metrics and other data about the environment at the time of the request, and writes a line to the logfile.
4. The agent massages the data it's collected, compresses it, and passes it to a central engine for analysis. The agent may also proactively alert when it detects problems.



Time	Client IP	URL	HTTP	CPU	RAM	POST
12:01:10	1.2.3.5	/index.html	200	13%	4.2	Name=Bob
12:01:20	5.7.4.1	/buy.html	404	94%	3.9	–
12:01:5	7.6.5.4	/img.gif	200	5%	4.7	Name=Mike

Figure 5-40. Server-agent-enhanced web analytics deployment

5. The analytics engine stores and aggregates all of the logfile data across all of the servers.
6. At regular intervals—often daily—the analytics engine generates a series of reports.

Popular web servers have well-defined interfaces to which an agent process can connect. On Apache servers, the module model (*http://modules.apache.org/*) allows third-party code to see each web request and the corresponding response. On Microsoft's servers, ISAPI (the Internet Server API) provides this connectivity. Companies like Symphoniq also use agent models to capture end user and platform health simultaneously.

Modules and ISAPI extensions may not just be passive listeners in a web conversation between web clients and a server—they may also change it. Many websites use modules for content caching, filtering, and authentication. These modules can get a better understanding of a web visit and can create more verbose logs than what's included in basic ELF.

See *http://www.port80software.com/products/* for some examples of server agents that provide additional visibility and control over web traffic on the server itself.

Server agents give you more insight into transactions, at the expense of more computing power and additional components you have to manage. If you collect analytics data with server agents, there are some advantages:

Additional information
: Server agents may record additional information about the health of servers (such as CPU, memory, and I/O) and request data (such as POST parameters) that is typically not collected by web servers

Lower data collection overhead
: Server agents can handle some of the aggregation, logfile compression, transfer, and decompression needed to get logs from the servers to the analytics tool. This reduces the amount of work the analytics engine needs to do in crunching log data, as well as the total volume of logfiles that need to be collected from each server.

SSL visibility
: Agents on servers are "inside" the SSL encryption boundary and can see every part of a web transaction without additional work managing SSL keys.

Turnkey solution
: For a small site running on only one machine, a server agent may act as both a collector and a reporting interface.

Data is available, even when the service dies
: Server agents may still collect data when the HTTP service itself isn't functioning properly.

Relying on a server agent to collect data has some downsides, however.

Use of server resources
: Agents consume resources on the servers, which are better used handling requests and serving web pages.

More things to break
: Agents add another point of failure to the web servers. For sites of any size, web servers are stripped-down, single-purpose machines that have been optimized to do a few things quickly and reliably.

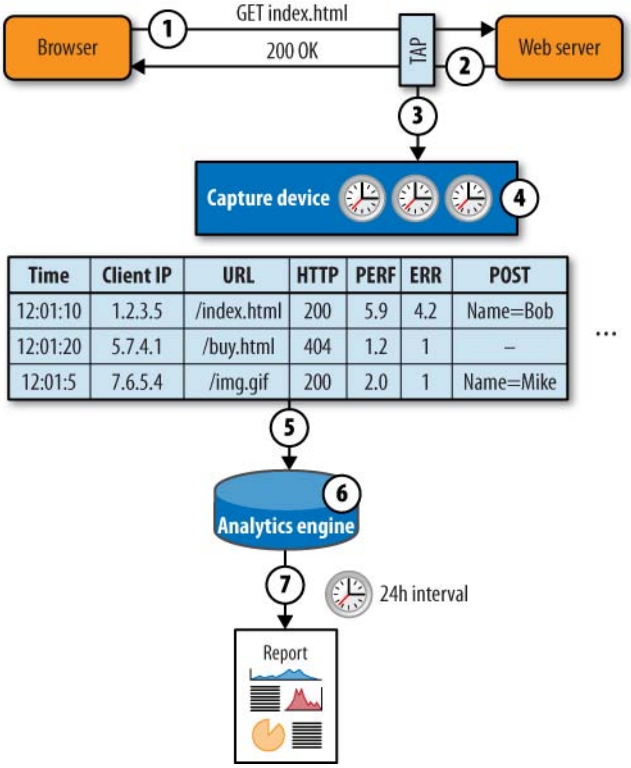
No client visibility
: Server agents share many of the limitations of web logs, because they don't have visibility into the client's environment and can't capture important segmentation data about what happens on the client.

Dependent on a working server
: If the server stops working, so does the server agent, leaving you guessing about what went wrong.

Man in the middle: Traffic capture

A third approach to collection is to capture a copy of the data flowing between clients and the web server, either through the mirroring functions of a switch, load balancer, or network tap, and to use this data to re-create the HTTP requests and responses.

To do this, a network device in front of the web server cluster makes a copy of every packet, which is then reassembled into TCP sessions consisting of HTTP requests and responses. The device is usually deployed just in front of or just behind the load balancer, or on the load balancer itself. Figure 5-41 illustrates how this model works.



Time	Client IP	URL	HTTP	PERF	ERR	POST
12:01:10	1.2.3.5	/index.html	200	5.9	4.2	Name=Bob
12:01:20	5.7.4.1	/buy.html	404	1.2	1	–
12:01:5	7.6.5.4	/img.gif	200	2.0	1	Name=Mike

Figure 5-41. Web analytics deployment using passive sniffing technology

1. A visitor's browser requests an object from the server. The browser provides a variety of data, such as referring URL and browser type, along with the request.
2. The web server responds with the requested object, as well as information such as object size and compression.

3. A passive monitoring device sends a copy of the network traffic to a collector device.
4. The collector device reassembles the packets into the HTTP request and response, and may also detect errors, record timing information, and extract snippets of text from the content of the response.
5. The collector stores this data as an augmented ELF containing all the information recorded about the transaction.
6. The analytics engine stores and aggregates all of the logfile data across all of the servers.
7. At regular intervals—often daily—the analytics engine generates a series of reports.

Passive traffic capture was an early favorite for analytics companies like Accrue, who wanted to offload work from web servers, analyze traffic destined for several servers, and get more visibility.

Capturing analytics data through a passive sniffing approach has some advantages:

Aggregates data from many servers
Passive capture can collect many servers' requests at once, provided they cross a single network segment.

Captures network health information
The passive device may record information such as network timings or packet loss.

Works when servers break
If the server dies, the device still records the circumstances surrounding that failure, as well as requests from browsers that went unanswered during an outage.

Custom data collection
Passive capture can collect data from not only default headers, but also from the payload of a transaction (such as the dollar value of a checkout). This can be used as a substitute for some kinds of page tagging.

No network or server delay
This kind of collection introduces no delay to the network and is quick to deploy.

Sees all requests
Passive capture can see requests for content that doesn't execute JavaScript, such as an RSS feed, a media object, or an Acrobat PDF.

Disadvantages of passive capture include:

Requires data center access
A passive capture approach involves network equipment—you need physical access to the data center to deploy this kind of system and you need help from IT to deploy it.

Requires SSL keys

Passive capture can't sniff encrypted traffic without help. You need to install SSL keys on the sniffing equipment, which may pose a security risk.

Problems with IP addresses

If the capture device is in front of a load balancer, it won't see the IP addresses of the individual servers. Instead, it sees only the address of the load balancer, making it hard to tell which server handled which request without additional configuration.

Problems with IP addresses, part two

Conversely, if the passive capture device is installed behind the load balancer, it doesn't see the true IP address of each request, so it can't always determine the source of traffic, making it impossible to segment by region or by domain name.

Lack of client visibility

A passive capture device can't see what's happening on the browser itself, which is an increasingly important part of any application.

Because of these limitations, most passive capture tools focus on monitoring web performance and end user experience, where their insight into network performance is invaluable. Logfile generation from such devices is a side effect of how they capture data, but web analytics is seldom the main purpose of these tools.

Static image request

Early attempts at hosted analytics were simple: the web operator would embed a third-party object, such as a small banner or one-pixel image, in each page. Each time the page was loaded, the browser would request the small object from the third party, which would keep track of how many web requests it had seen.

Figure 5-42 illustrates this approach to collection.

1. The site operator embeds a reference to an image on the counter's site. The link includes the site operator's unique ID.
2. A visitor's browser requests an object from the server. Along with that request, the browser provides a variety of data such as referring URL and browser type.
3. The web server responds with the page containing the image reference to the site counter.
4. The browser parses the page it receives and sees that it has to ask the counter site (*counter.com*) for an image (*display.gif*).
5. The browser requests *display.gif* from the third-party counter site, and includes the unique ID of the site that was embedded in the page.
6. The counter site looks up the number of hits that the site with that ID had and increments it.
7. The counter site sends a rendered image containing the hit count back to the browser.

8. The browser displays the page containing the hit count.
9. The site engine may also offer reports on traffic volumes, like the one in Figure 5-43.

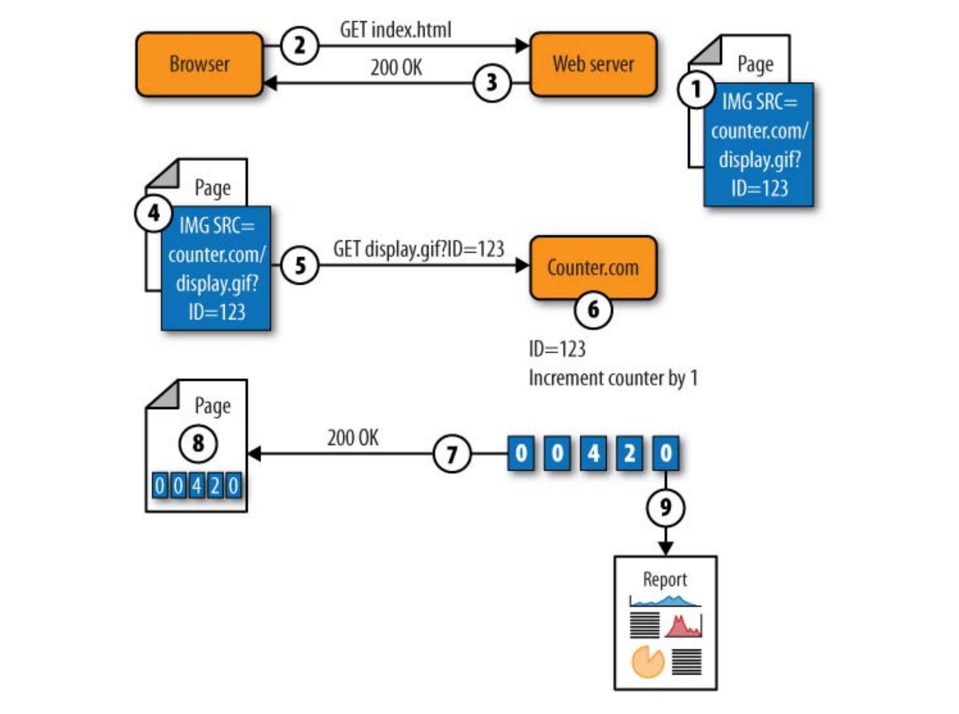
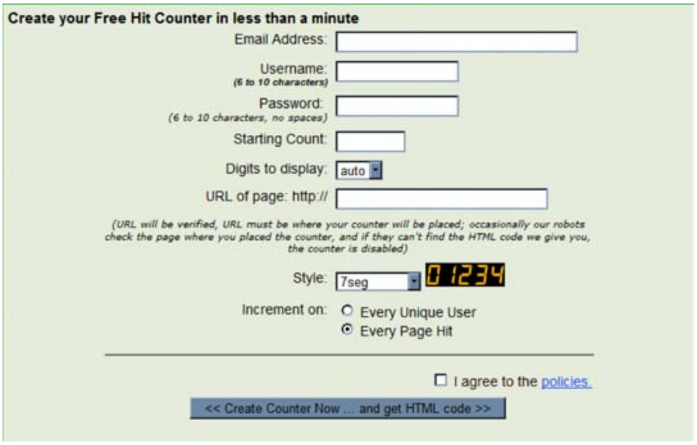


Figure 5-42. Web analytics deployment using an image-based counter



Create your Free Hit Counter in less than a minute

Email Address:

Username:
(6 to 10 characters)

Password:
(6 to 10 characters, no spaces)

Starting Count:

Digits to display: auto

URL of page: http://

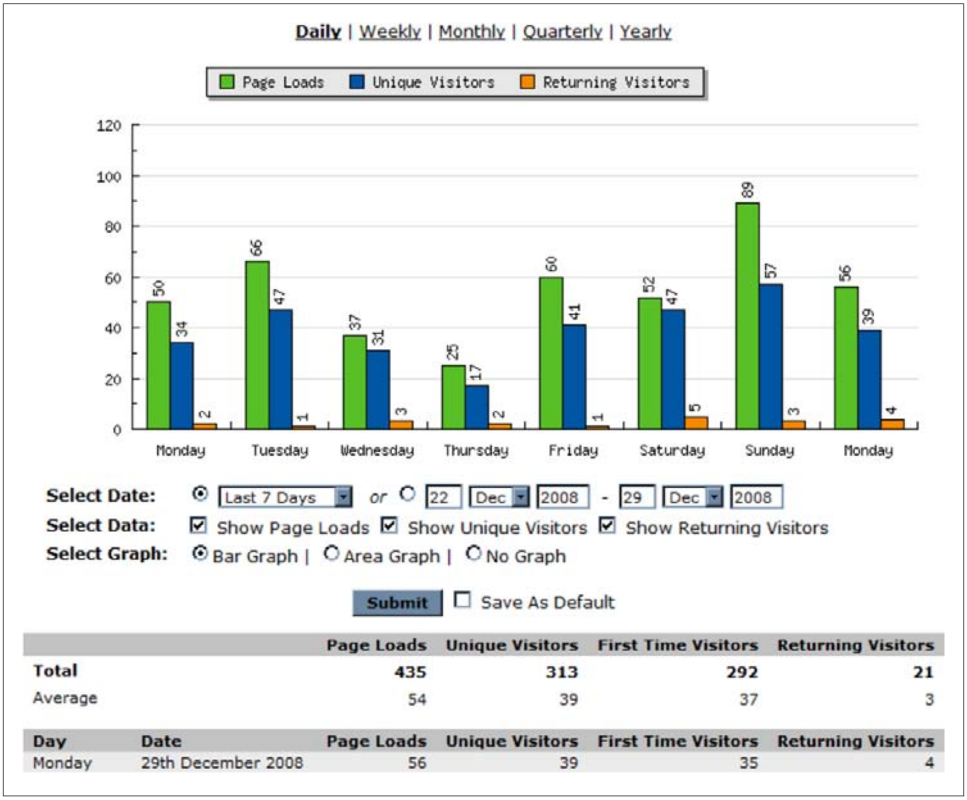
(URL will be verified, URL must be where your counter will be placed; occasionally our robots check the page where you placed the counter, and if they can't find the HTML code we give you, the counter is disabled)

Style: 7seg

Increment on: ○ Every Unique User
● Every Page Hit

□ I agree to the policies.

Figure 5-43. A report from Statcounter.com



Select Date: ● Last 7 Days or ○ 22 Dec 2008 - 29 Dec 2008

Select Data: ☑ Show Page Loads ☑ Show Unique Visitors ☑ Show Returning Visitors

Select Graph: ● Bar Graph | ○ Area Graph | ○ No Graph

□ Save As Default

	Page Loads	Unique Visitors	First Time Visitors	Returning Visitors
Total	**435**	**313**	**292**	**21**
Average	54	39	37	3

Day	Date	Page Loads	Unique Visitors	First Time Visitors	Returning Visitors
Monday	29th December 2008	56	39	35	4

Figure 5-44. VisibleCounter hit counter configuration

Initially, services like VisibleCounter, shown in Figure 5-44, simply embedded the number of page hits in an image that was displayed to each visitor.

Companies that created these services quickly realized that they could glean more from each request:

- Several requests from the same browser are probably the same user; this allows the company to estimate the number of visitors.
- Requests that include a cookie from the service are returning visitors who have been to the site before.
- The identity of the page being visited can be found from the referring URL of the retrieved counter image, since that image is a component of the parent page.

As a result, static image requests became a way to get rudimentary traffic statistics.

Static image models aren't really analytics. They only capture page traffic information, and as we've seen, this is misleading data that's hard to act upon. We mention them here because this approach is unfortunately still used by certain media sites that rely on these statistics to share with advertisers to get premium rates.

Static image requests don't carry any custom information about the visitor or the browser environment. Modern JavaScript is superior to this approach. Stat counters are used by casual website creators who aren't interested in understanding their visitors' behaviors or improving their sites. Now that there are free web analytics packages available, there's no excuse to use a static image model for anything other than a lazy way of showing your audience how many visitors you've had.

JavaScript

The first three methods we looked at are all forms of server-side collection. In other words, they're installed near the web server itself. JavaScript (and the static image model) is a client-side collection model. This is by far the dominant model on the Internet today. Browsers are able to collect and send a great deal of information about a user's visit to a third-party system such as an analytics service. Figure 5-45 shows how this works.

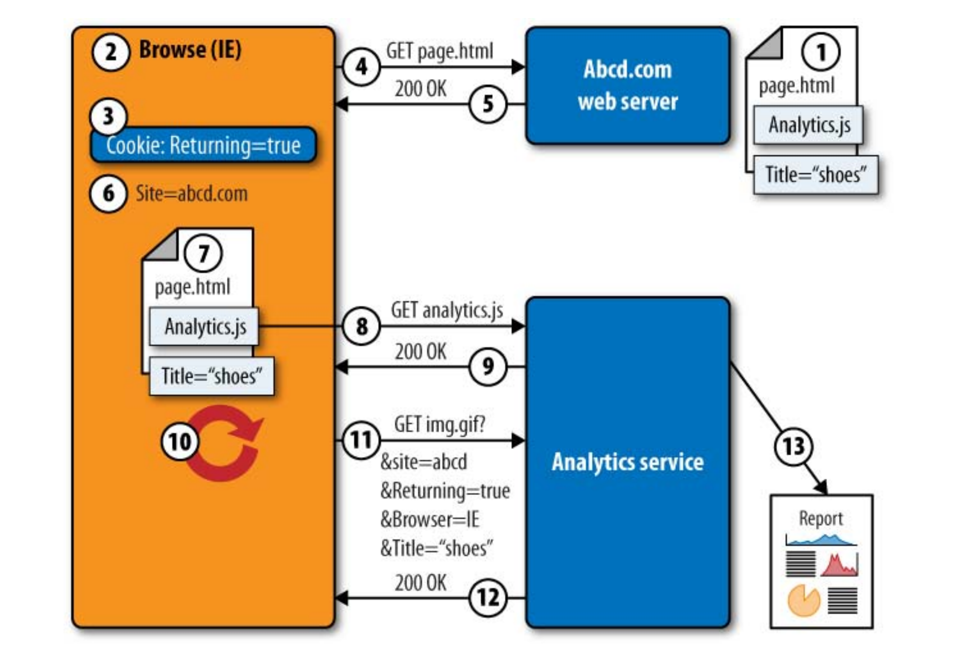


Figure 5-45. JavaScript-based collection of web analytics data

1. The site deploying analytics (Abcd.com) inserts a snippet of JavaScript code into its pages. The code includes a reference to a file, *Analytics.js*, that's stored on the analytics service's servers. The site also includes tags in the page that explain what it's about (in this case, "shoes").

2. A visitor launches his browser. Data about the browser is stored within the Document Object Model (DOM) of the browser.
3. The browser also has site-specific information, such as a cookie associated with Abcd.com. This cookie was created during a previous visit to the site.
4. The browser requests a page from Abcd.com.
5. The web server responds with a page containing the analytics script reference.
6. The browser's DOM also records the fact that the site being visited is abcd.com.
7. As the page is loaded, the browser realizes it has to retrieve *analytics.js* from the analytics service.
8. The browser requests the *analytics.js* code from the service.
9. The analytics service responds with the analytics JavaScript. This may be cached on the browser for future use.
10. When the browser receives the *analytics.js* JavaScript, it executes it. The script collects data from the DOM (browser type, site name), site-specific cookies ("returning=true") and tags within the page ("Title=shoes").
11. The analytics script appends all the data it has collected to a request for a tiny (1×1 pixel) image from the analytics service.
12. The analytics service returns the tiny image, storing the information it received about the visit.
13. The analytics service generates reports based on this data.

For some web platforms, such as blogs, analytics vendors provide plug-ins to make installation even easier. Most of the time, however, implementing JavaScript-based analytics is as easy as entering a few lines of code in a *footer.php* file or changing some file templates.

If you were installing Google Analytics, the JavaScript code would look something like this:

```
<script type="text/javascript">
  var gaJsHost = (("https:" == document.location.protocol) ?
  "https://ssl." : "http://www.");
  document.write(unescape("%3Cscript src='" + gaJsHost +
  "google-analytics.com/ga.js' type='text/javascript'%3E%3C/script%3E"));
</script>
  <script type="text/javascript">
  var pageTracker = _gat._getTracker("UA-xxxxxx-x");
  pageTracker._trackPageview();
</script>
```

Every time a browser loads a page with this script, it will make a request for *http://google-analytics.com/ga.js* and run the script it receives.

JavaScript is a popular collection approach, offering many advantages:

Sees what the visitor sees

JavaScript sees what the user's environment is like. This includes DOM information such as resolution and activity within the page, or the page's title, or events such as Onload that mark important milestones in a page's delivery.

Works with SSL

Because it's on the client, it can collect data within SSL connections.

Augments analytics with page context

It can read tags from the page (identifying things like the products a user is browsing) and send them to the analytics platform, allowing you to more easily segment visits.

Scales with number of users

In a JavaScript collection model, the browsers do some of the work of collecting user data. They then generate a single hit per page, rather than the hit-per-object rates of server-side collection. This reduces the load on servers somewhat.

Opens the door to hosted services

Using JavaScript means you can send information to someone else via the visitor's browser. Most companies' IT organizations would be uncomfortable sending logfiles to a third party, but when the visitors' browsers send that data, it's acceptable.

Keeps reporting with RIAs and "long" pages

JavaScript can manage communications with the analytics service beyond the initial page load. If a page has small subevents, such as a user clicking "play" on an embedded object, JavaScript can capture this client-side activity by sending additional information to the analytics service. JavaScript can also measure things like cursor and mouse movement and even send messages when a user closes a web page.

Reads cookies

JavaScript can get details on the visitor from a cookie and send this to the service. So, if a user is a "platinum" subscriber to your website, you can store this in a cookie on the user's browser. When the JavaScript executes, it can read values from the `document.cookie` property of the page and turn them into labels for the visitor's session that you can then use for segmentation.

Analyzes mashups

JavaScript sees requests to third-party sites. Because the JavaScript is executing as part of the container page, it sees all component objects, such as a Google Map or a YouTube video, and can report on them as well. If you're monitoring a mashup, you need JavaScript.

When collecting through JavaScript on a client, you may face the following disadvantages:

Can't see out of the sandbox

JavaScript runs within a browser. Browsers "sandbox" the code they run to prevent malicious websites from gaining control of visitors' computers. The script can't see information such as networking statistics, because it can't communicate directly with the operating system on which it's running.

Requires a loaded page

If the page doesn't load, JavaScript is useless. This is the main reason JavaScript hasn't had the same success for web performance monitoring that it has enjoyed for web analytics. If you want to measure problems, a system that doesn't work when problems occur isn't as useful.

Requires a JavaScript interpreter

Some mobile devices, and some very paranoid visitors, may have disabled JavaScript on their browsers, limiting your ability to collect from them. Users who delete their cookies may also be misreported as new visitors when, in fact, they are returning users.

> According to w3schools (*http://www.w3schools.com/browsers/browsers_stats.asp*), about 5% of users on the Internet today block JavaScript, although this percentage has been in a steady decline since 2005. This study may be misleading, as there is no indication that it takes mobile devices into account.

Page tagging still sucks

JavaScript's real limitation becomes apparent due to the necessity of manually tagging a page to give it meaning. In reports, there is a big difference between seeing a page named "id=3&item=19&size=9&c=5" versus "Big Blue Shoes with Pumps." Consequently, you will need to build analytics into your site from the outset, and it must provide context by individually tagging pages. Many sites have tens of thousands of unique pages. Tagging pages is a web analyst's most hated chore, and a daunting task if done as an afterthought.

Despite the limitations outlined here, JavaScript offers such improved visibility into your visitors that it's by far the leading approach to web analytics collection used online today.

Comparing data capture models

You'll probably use JavaScript for collection unless you have a site that can't take advantage of third-party Internet-connected services. Complementing JavaScript with passive capture for measuring user experience is also increasingly common within large organizations.

Table 5-2 summarizes the trade-offs of each approach.

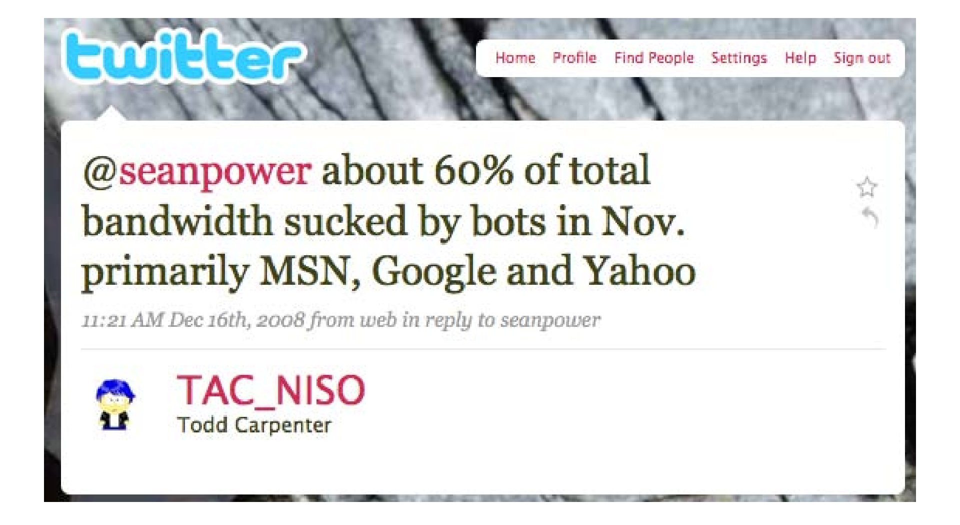
Table 5-2. A comparison of different methods of deploying analytics

	Web logs	Server agents	Passive capture	Image request	JavaScript
Deployment					
Requires physical access?			Y		
Requires server administrator?	Y	Y			
Requires changes to page content?				Y	Y
Requires Internet-connected users?				Y	Y
Data stored by third party?				Y	Y
Requires access to SSL keys?			Y		
Requires logfile collection?	Y	Y			
Data visibility					
Sees POST parameters?		Y	Y		Y
Sees network statistics (packet loss)?		Y	Y		
Sees what happened when web service dies?		Y	Y		
Sees what happened when entire server dies?			Y		
Sees failed requests (404s)?	Y	Y	Y		
Sees requests for non-JavaScript objects (PDFs, RSS feeds)?			Y		
Sees browser information (resolution, DOM, last visit)?					Y
Sees third-party content (mashups)?					Y
Sees client-side activity (mouse movement, etc.)?					Y
Performance impact					
Adds to server CPU load?	Y	Y			
Adds to page load time?				Y	Y

Set Up Filters

Now that you've decided how you want to collect analytics data, it's time to decide what you don't want to keep. Much of your web traffic isn't from the visitors you care about: it includes malicious attackers trying to exploit known weaknesses, spammers posting content you're going to block anyway, and crawlers busily indexing the Web. It may include synthetic tests that your web operations team is running to measure the health of the website, or it may consist of visits from internal users and employees whose traffic shouldn't be counted.

No two analytics tools will calculate the same number of hits for a website. This happens for many reasons: different filters, different definitions of a hit or a page, different



@seanpower about 60% of total bandwidth sucked by bots in Nov. primarily MSN, Google and Yahoo

11:21 AM Dec 16th, 2008 from web in reply to seanpower

TAC_NISO
Todd Carpenter

Figure 5-46. Interesting bot stats for niso.org

JavaScript approaches, the impact of bots (which, as Figure 5-46 shows, can be significant), and different ways of inferring pages from objects.

Some of the work of filtering out the noise may be done for you, depending on your chosen approach. Most analytics packages will block visitors with known user agents (those that self-identify as crawlers). If you're using JavaScript collection, crawlers that don't execute JavaScript won't send tracking hits to the analytics providers, but if you're using server-side collection methods you'll need to exclude this traffic.

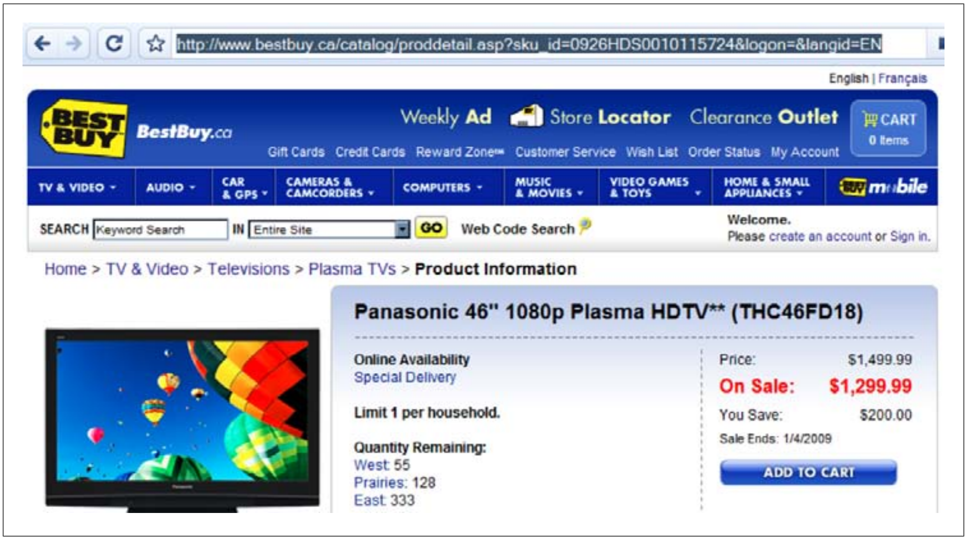
You may not want to block synthetic traffic—it may be good to know who's crawling your site and how often they're updating their indexes—but you need to identify traffic that shouldn't be included in your estimates of real visitors, or you'll artificially lower your conversion rates.

As your site becomes better known, you'll see an increase in the number of machine-driven visitors that generate hits. We've seen sites where more than half of the total web traffic isn't actual visitors.

Identify Segments You Want to Analyze

Now that you're measuring how well visitors are attaining the goals you've set for them, it's time to segment those visitors up into groups. This way, you can see which groups are performing poorly and help them. You can also try changes to content, promotion, and design and see whether they improve things.

All analytics tools have some amount of built-in segmentation—new versus returning visitors, referring site, browser type, country of origin, and so on. You may want to



http://www.bestbuy.ca/catalog/proddetail.asp?sku_id=0926HDS0010115724&logon=&langid=EN

Welcome.
Please create an account or Sign in.

Home > TV & Video > Televisions > Plasma TVs > Product Information

Panasonic 46" 1080p Plasma HDTV** (THC46FD18)

Online Availability
Special Delivery

Limit 1 per household.

Quantity Remaining:
West: 55
Prairies: 128
East: 333

Price:	$1,499.99
On Sale:	$1,299.99
You Save:	$200.00
Sale Ends: 1/4/2009	

Figure 5-47. A page on BestBuy.ca showing a single URL (/catalog/proddetail.asp) for all product views

create new segments on the fly using page tags, which will allow you to group and analyze visitor performance according to new dimensions.

One quick way to do this is to use unique strings in a URL to create custom segments. For example, if you have pages whose URLs contain the strings "/buyer/" and "/seller/" in them, you can ask most analytics tools to use these to generate a custom segment such as "all URLs that contain /seller/" to see KPIs for just that part of the site.

Tag Your Pages to Give Them Meaning

With collection in place, goals identified, and filtering ready to clean up the noise, it's time to tag your pages. Tagging provides context about the page that's passed from your website to the browser, and from the browser to the analytics provider.

Tagging is where JavaScript-based collection gets complicated. To implement basic JavaScript analytics, all you had to do was include a few lines of text on each page. However, as far as the analytics service knows, those pages are identical. The only thing that makes them different is their URLs.

Consider a retail outlet that has a single page name for all products, which is a common characteristic of many retail sites. In Figure 5-47, that URL is */catalog/proddetail.asp*, and it's the same URL regardless of what product the visitor is looking at. The URL doesn't tell us much about what's on the page. Any human can see that the page is categorized as TV & Video→Televisions→Plasma TVs, but without some help, the JavaScript doesn't know what the page is about.

Fortunately, there's a way to take information (such as "Televisions") and augment the visitor's session record with it. Let's look at the script contained in the page in Figure 5-47 and see how it works.

First, the page includes variables that identify content, such as the user's account, the product name and model, and the multilevel content hierarchy of the product.

```
<!-- BEGIN WEBSIDESTORY CODE HITBOX COMMERCE HBX1.3 (cartadd) -->
<!--COPYRIGHT 1997-2004 WEBSIDESTORY,INC. ALL RIGHTS RESERVED.
 U.S.PATENT No.6,393,479B1 & 6,766,370. INFO:http://websidestory.com/privacy-->
<script language="javascript">
var _hbEC=0,_hbE=new Array;function _hbEvent(a,b){b=_hbE[_hbEC++]
=new Object();b._N=a;b._C=0;return b;}
var hbx=_hbEvent("pv");hbx.vpc="HBX0131.01a";hbx.gn="ehg-bestbuy.hitbox.com";

//BEGIN EDITABLE SECTION
//CONFIGURATION VARIABLES
hbx.acct="DM540930IBWD";//ACCOUNT NUMBER(S)
hbx.pn="0926hds0010115724-panasonic+46+1080p+plasma+hdtv+(thc46fd18)";//PAGE NAME(S)
hbx.mlc="/online/en/tv+and+video/televisions
/plasma+tvs/details";//MULTI-LEVEL CONTENT CATEGORY
hbx.pndef="default.asp";//DEFAULT PAGE NAME
hbx.ctdef="full";//DEFAULT CONTENT CATEGORY
```

The page can also include custom variables (which may be used for experimentation, custom error tracking, or segmentation), as well as details such as the price or whether the visitor was told that quantities were limited. All of this data can be used later to segment buyers and see if pricing affected conversion.

```
//CUSTOM VARIABLES
hbx.hc1="panasonic|panasonic+46+1080p+plasma+hdtv+(thc46fd18)";//BRAND|PRODUCT
hbx.hc2="0926HDS0010115724";//SKU
hbx.hc3="";//CUSTOM 3
hbx.hc4="";//CUSTOM 4
hbx.hrf="";//CUSTOM REFERRER
hbx.pec="";//ERROR CODES

//COMMERCE VARIABLES
hbx.cacct="975410043989";
hbx.pr="panasonic+46+1080p+plasma+hdtv+(thc46fd18)";  //comma delimited products
hbx.bd="panasonic";
hbx.ca="televisions";
hbx.pc="1299.99";    //comma delimited prices
hbx.qn=""; //comma delimited quantities
hbx.sr="1";    //store
hbx.cp="null";  //campaign
hbx.cam="0";  //cart add methodology, 0 = highwatermark, 1 = incremental
hbx.pv=1;  //product view flag, 0 = cart add, 1 = product view

//END EDITABLE SECTION
```

Note that these variables are seldom handcoded into the page. Rather, the application inserts them dynamically each time it renders the page. In other words, implementing tagging means working with developers.

Now the analytics script assembles all of these tags and details, along with browser information such as the type of web browser (`navigator.appversion`), the title of the page (`document.title`), and the referring URL (`document.referrer`):

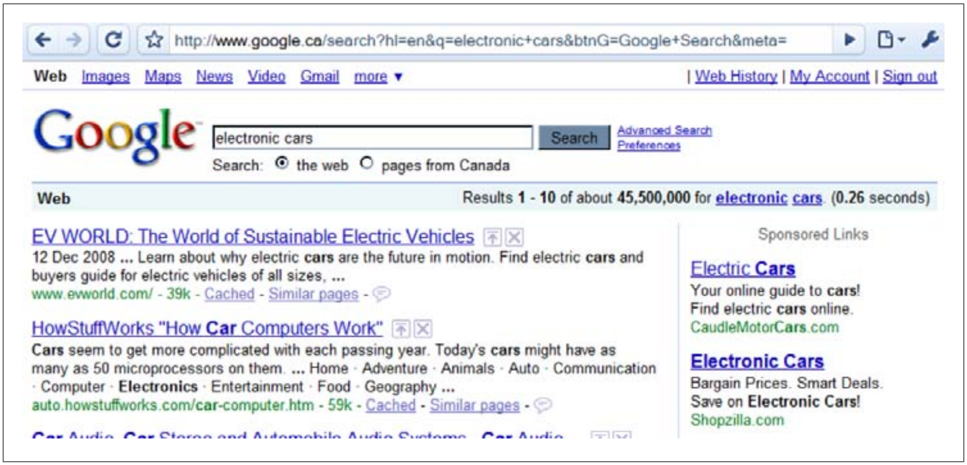
```
function $ii( a, b,c){ return a.indexOf(b, c?c:0)};
function $is(a,b, c){return b>a.length?
"":a.substring(b,c!=null?c:a.length)};function $a(v){ return
 escape(v) }; var _sv=10, _bn=navigator.
appName,_mn="we74",_bv=parseInt(navigator.appVersion),_rf=$a(document.referrer),
_epg="n&cam="+hbx.cam+"&pv="+(hbx.pv?"1":"0&abd_type
=cart_add")+"&product="+$a(hbx.pr)+
"&quantity="+$a(hbx.qn)+"&brand="+$a(hbx.bd)+"&category="+$a(hbx.ca)+"&
price="+$a(hbx.pc)+
"&store="+$a((hbx.sr=="S"+"TORE")?1:hbx.sr)+"&tz=
PST&aid="+hbx.cacct;if(!$ii(_bn,"Micro"+
"soft"))_bn="MSIE";if(_bn=="MSIE"&&_bv==2)_bv=3;function $l(m,l){return m=="/"?
m:(($ii(m,"/")?"/":"")+(m.lastIndexOf("/")==l?m.substring(0,l):m))};function $n(
a,b){return(a==""||a=="/")?"/":$is(a,hbx.ctdef!="full"?a.lastIndexOf("/",b-2): $ii(
a,"/"),b)};function $o(a,b,c){var d=location.pathname,e=$is(d,d.lastIndexOf("/")+
1,d.length);if(a&&b==c){return(hbx.pndef=="title"&&document.title!=""&&document.
title!=location)?document.title:e?e:hbx.pndef}else{return(b==c)?$n(d,d.lastIndexOf(
"/")):$l(b,b.length-1)}};function $p(a,b,c,d){return ""+(c>-1?$o(b,$is(a,0,c),d
)+";"+$p($is(a,c+1),b,$ii($is(a,c+1),";")):$o(b,a,d))};
hbx.mlc=$p( hbx.mlc,0,$ii(hbx.mlc,
";"),"CONTENT+CAT"+"EGORY");hbx.pn=$p(hbx.pn,1,$ii
(hbx.pn,";"),"PUT+"+"PAGE+NAME+HERE" );
</script><script type="text/javascript" src="/javaScript/hitbox/hbx.js"></script>
<noscript>
```

The result of all this work is a reference to an image embedded in the page. The browser requests the image from the analytics service (in this case, *ehg-bestbuy.hitbox.com*).

```
<img src="http://ehg-bestbuy.hitbox.com?hc=none&
cd=1&hv=6&ce=u&hb=DM540930IBWD&aid=975410043989&n=0926hds0010115724-panasonic
+46+1080p+plasma+hdtv+(thc46fd18)&vcon=/online/en/tv+and+video/
televisions/plasma+tvs/details&seg=&cmp=&gp=&cam=0&pv=1
&abd_type=cart_add&product=panasonic+46+1080p+plasma+hdtv+(thc46fd18)&quantity
=&price=1299.99&store=1&tz=PST&vpc=HBX0131.01an"
 border="0" width="1" height="1"></noscript>
<!-- END WEBSIDESTORY CODE  -->
```

All of the metadata about the page is appended to the request in the form of a set of URI parameters. The publisher of the page never intends to display this image (it's 1 × 1 pixel wide with no border). The image doesn't matter, though: it's the request that matters.

If you have a small site that doesn't have dynamic backends, however, you may have to manually edit page content to provide context about the page to whatever analytics package you're using. Some packages, like Google Analytics, make this fairly simple, while enterprise-grade packages are much more demanding, but provide better segmentation and analytics as a result.



http://www.google.ca/search?hl=en&q=electronic+cars&btnG=Google+Search&meta=
Web Images Maps News Video Gmail more ▾
| Web History | My Account | Sign out
Google
electronic cars
Search
Advanced Search
Preferences
Search: ● the web ○ pages from Canada
Web
Sponsored Links

Results **1** - **10** of about **45,500,000** for **electronic cars**. (**0.26** seconds)

EV WORLD: The World of Sustainable Electric Vehicles
12 Dec 2008 ... Learn about why electric **cars** are the future in motion. Find electric **cars** and buyers guide for electric vehicles of all sizes, ...
www.evworld.com/ - 39k - Cached - Similar pages

HowStuffWorks "How **Car** Computers Work"
Cars seem to get more complicated with each passing year. Today's **cars** might have as many as 50 microprocessors on them. ... Home · Adventure · Animals · Auto · Communication · Computer · **Electronics** · Entertainment · Food · Geography ...
auto.howstuffworks.com/car-computer.htm - 59k - Cached - Similar pages

Car Audio - **Car** Stereo and Automobile Audio Systems - **Car** Audio

Electric **Cars**
Your online guide to **cars**!
Find electric **cars** online.
CaudleMotor**Cars**.com

Electronic Cars
Bargain Prices. Smart Deals.
Save on **Electronic Cars**!
Shopzilla.com

Figure 5-48. A Google search showing the referring URL containing query terms

The tagging process varies by analytics platform. There are a few great analytics integration checklists on the Web. For Google Analytics, check out:

- *http://blog.vkistudios.com/index.cfm/2008/12/5/Google-Analytics-Power-User--Tutorials-and-Screencasts--Part-2--Account-setup*
- *http://analytics.mikesukmanowsky.com/analytics/index.php/2007/08/18/google-analytics-installation-guide/*

For larger platforms, like Omniture, check out:

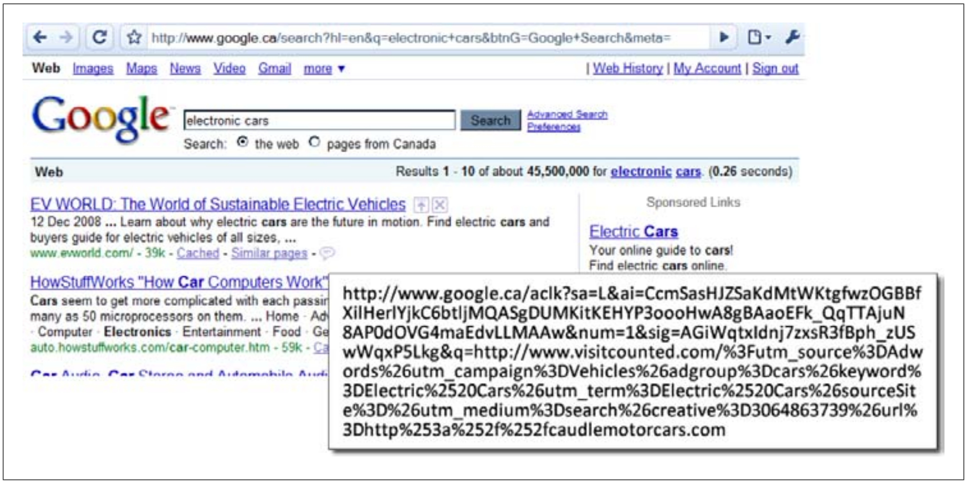
- *www.wickedsciences.com/blogs/?p=6*
- *www.kpilist.com/2008/10/evar-sitecatalyst-custom-conversion.html*

One final note on JavaScript implementation: page tags can bloat pages and cause them to load slowly, so it makes sense to put this kind of content at the bottom of your pages.

Campaign Integration

By defining your goals, you're able to determine where visitors went, and by tagging your pages, you're able to provide context for what they did. There's one more piece missing, however: you need to capture the things that drove people to your site in the first place.

When a user searches for something, the referring URL contains a list of search terms. A search for "electronic cars," for example, has the URI parameters *q=electronic+cars*, as shown in Figure 5-48. When an analytics script sends the referring URI to an analytics service, the service can parse this URI and see that the keywords "electronic" and "cars" were used.


http://www.google.ca/search?hl=en&q=electronic+cars&btnG=Google+Search&meta=
Web
Images
Maps
News
Video
Gmail
more ▾
| Web History | My Account | Sign out
Google
electronic cars
Search
Advanced Search
Preferences
Search: ● the web ○ pages from Canada
Web

Results **1** - **10** of about **45,500,000** for **electronic cars**. (**0.26** seconds)

EV WORLD: The World of Sustainable Electric Vehicles
12 Dec 2008 ... Learn about why electric **cars** are the future in motion. Find electric **cars** and buyers guide for electric vehicles of all sizes, ...
www.evworld.com/ - 39k - Cached - Similar pages -

HowStuffWorks "How **Car** Computers Work"
Cars seem to get more complicated with each passir
many as 50 microprocessors on them. ... Home · Adv
· Computer · **Electronics** · Entertainment · Food · Ge
auto.howstuffworks.com/car-computer.htm - 59k - Ca

Sponsored Links

Electric **Cars**
Your online guide to **cars**!
Find electric **cars** online.

```
http://www.google.ca/aclk?sa=L&ai=CcmSasHJZSaKdMtWKtgfwzOGBBf
XilHerlYjkC6btljMQASgDUMKitKEHYP3oooHwA8gBAaoEFk_QqTTAjuN
8AP0dOVG4maEdvLLMAAw&num=1&sig=AGiWqtxIdnj7zxsR3fBph_zUS
wWqxP5Lkg&q=http://www.visitcounted.com/%3Futm_source%3DAdw
ords%26utm_campaign%3DVehicles%26adgroup%3Dcars%26keyword%
3DElectric%2520Cars%26utm_term%3DElectric%2520Cars%26sourceSit
e%3D%26utm_medium%3Dsearch%26creative%3D3064863739%26url%
3Dhttp%253a%252f%252fcaudlemotorcars.com
```

Figure 5-49. The URI for an ad click leading to an intermediate page that records the click-through

If the link is part of a paid campaign, however, the link doesn't go directly to the destination site. Instead, it first requests an ad click object from the search engine (in this case, `www.google.ca/aclk`) with a number of parameters identifying the campaign, the advertiser, and so on, as shown in Figure 5-49.

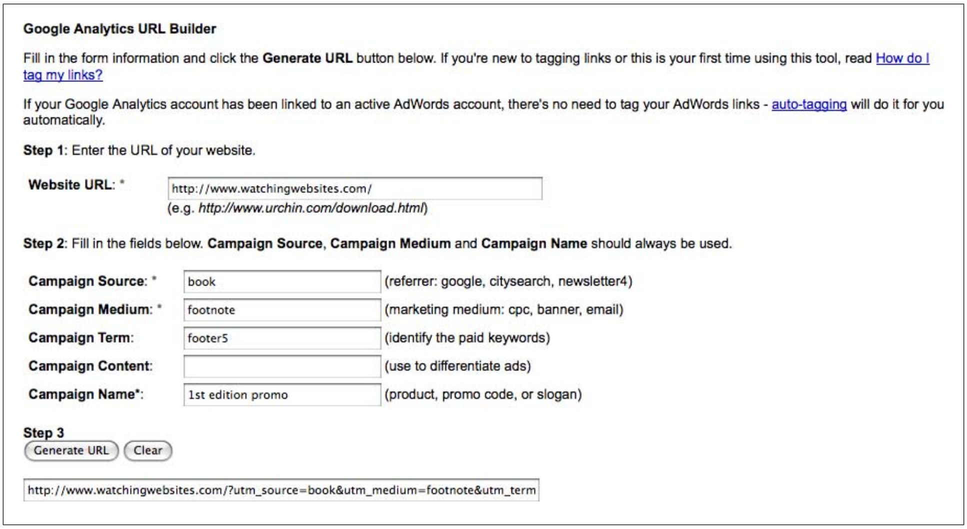
Your analytics tool needs to extract the campaign information from the subsequent referring URI, which you can then use to segment visitors according to campaigns and goal attainment. Google Analytics does this automatically for Google AdWords, but you can manually create a URL using Google's URL Builder (*www.google.com/support/googleanalytics/bin/answer.py?hl=en&answer=55578*), shown in Figure 5-50.

Campaign information can also be embedded in pages. In the Best Buy example cited earlier, the site operator can embed form, segment, and campaign information within the page according to landing page or elements of the referring URL. Some of these attributes may be passed from page to page during a visit.

```
//OPTIONAL PAGE VARIABLES
//ACTION SETTINGS
hbx.fv="";//FORM VALIDATION MINIMUM ELEMENTS OR SUBMIT FUNCTION NAME
hbx.lt="auto";//LINK TRACKING
hbx.dlf="n";//DOWNLOAD FILTER
hbx.dft="n";//DOWNLOAD FILE NAMING
hbx.elf="n";//EXIT LINK FILTER

//SEGMENTS AND FUNNELS
hbx.seg="";//VISITOR SEGMENTATION
hbx.fnl="";//FUNNELS

//CAMPAIGNS
hbx.cmp="";//CAMPAIGN ID
hbx.cmpn="";//CAMPAIGN ID IN QUERY
hbx.dcmp="";//DYNAMIC CAMPAIGN ID
hbx.dcmpn="";//DYNAMIC CAMPAIGN ID IN QUERY
hbx.hra="ATT";//RESPONSE ATTRIBUTE
hbx.hqsr="";//RESPONSE ATTRIBUTE IN REFERRAL QUERY
hbx.hqsp="";//RESPONSE ATTRIBUTE IN QUERY
hbx.hlt="";//LEAD TRACKING
hbx.hla="";//LEAD ATTRIBUTE
hbx.gp="";//CAMPAIGN GOAL
hbx.gpn="";//CAMPAIGN GOAL IN QUERY
hbx.hcn="";//CONVERSION ATTRIBUTE
hbx.hcv="";//CONVERSION VALUE
```



Google Analytics URL Builder

Fill in the form information and click the **Generate URL** button below. If you're new to tagging links or this is your first time using this tool, read How do I tag my links?

If your Google Analytics account has been linked to an active AdWords account, there's no need to tag your AdWords links - auto-tagging will do it for you automatically.

Step 1: Enter the URL of your website.

Website URL: * http://www.watchingwebsites.com/
(e.g. *http://www.urchin.com/download.html*)

Step 2: Fill in the fields below. **Campaign Source**, **Campaign Medium** and **Campaign Name** should always be used.

Campaign Source: *	book	(referrer: google, citysearch, newsletter4)
Campaign Medium: *	footnote	(marketing medium: cpc, banner, email)
Campaign Term:	footer5	(identify the paid keywords)
Campaign Content:		(use to differentiate ads)
Campaign Name*:	1st edition promo	(product, promo code, or slogan)

Step 3

http://www.watchingwebsites.com/?utm_source=book&utm_medium=footnote&utm_term

Figure 5-50. By entering campaign parameters into Google's URL builder, you create a URL that, if followed, will count toward that campaign's visits

Go Live and Verify Everything

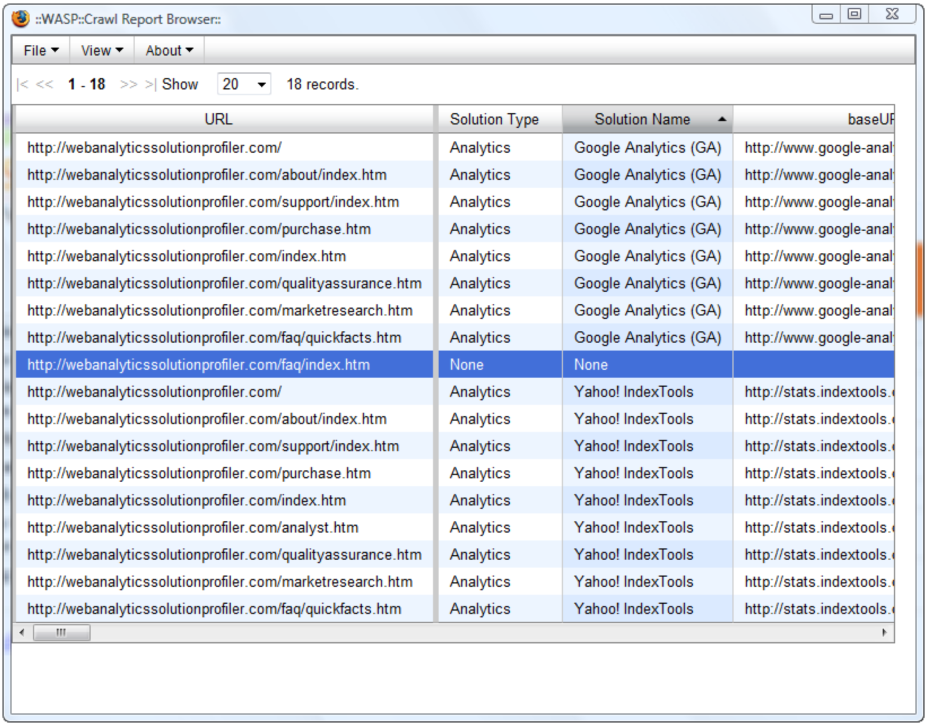
With collection, filtering, goals, tagging, and campaign segmentation in place, it's time to start testing. Here's a quick preflight checklist.

Is performance still acceptable?

Compare pages with JavaScript to those without, and ensure that the time it takes to load the page hasn't increased too much. If it has, you'll negatively affect conversion. For more information on how to measure page performance, see Chapters 8 through 10

Are pages instrumented properly?

Go to each section of the site and view the page source to ensure the scripts appear in the correct places. It can be hard to do this manually, but you can use tools like Stephane Hamel's WASP (*http://webanalyticssolutionprofiler.com/*), shown in



URL	Solution Type	Solution Name	baseUR
http://webanalyticssolutionprofiler.com/	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/about/index.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/support/index.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/purchase.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/index.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/qualityassurance.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/marketresearch.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/faq/quickfacts.htm	Analytics	Google Analytics (GA)	http://www.google-anal
http://webanalyticssolutionprofiler.com/faq/index.htm	None	None	
http://webanalyticssolutionprofiler.com/	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/about/index.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/support/index.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/purchase.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/index.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/analyst.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/qualityassurance.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/marketresearch.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.
http://webanalyticssolutionprofiler.com/faq/quickfacts.htm	Analytics	Yahoo! IndexTools	http://stats.indextools.

Figure 5-51. A WASP report showing pages with missing tags

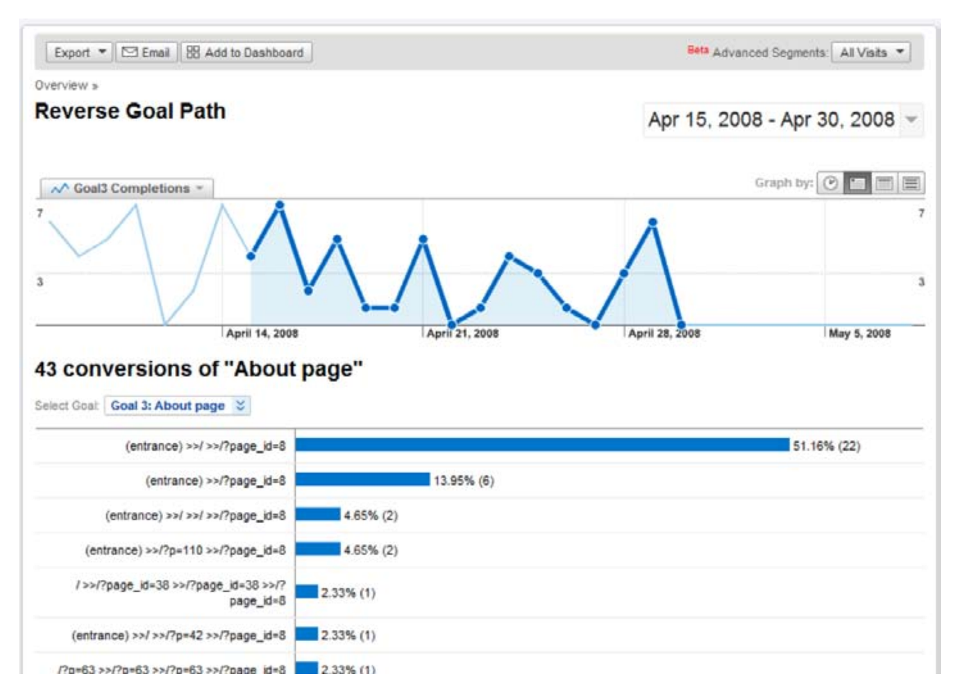
Figure 5-51, to crawl a site and identify improperly instrumented pages, or a tool like Firebug to see what's running when you load a site.

Are goals properly configured to measure conversions?

A common problem when implementing analytics is getting goals wrong. If analytics isn't measuring the right steps in a goal, you can't measure success. Fortunately, there are a variety of ways to check goal outcomes and see if you're tracking them properly.

A reverse goal path, like the one shown in Figure 5-52, shows how visitors reached a goal and may identify other paths toward an outcome that you didn't know about. This is a good way to catch missing steps in a conversion process.

You should also compare your financials to your analytics: if the accounting department says you received 20 orders on Monday, but your system only saw 5, something is definitely amiss.



(entrance) >>/ >>/?page_id=8	51.16% (22)
(entrance) >>/?page_id=8	13.95% (6)
(entrance) >>/ >>/ >>/?page_id=8	4.65% (2)
(entrance) >>/?p=110 >>/?page_id=8	4.65% (2)
/ >>/?page_id=38 >>/?page_id=38 >>/?page_id=8	2.33% (1)
(entrance) >>/ >>/?p=42 >>/?page_id=8	2.33% (1)
/?p=63 >>/?p=63 >>/?p=63 >>/?page_id=8	2.33% (1)

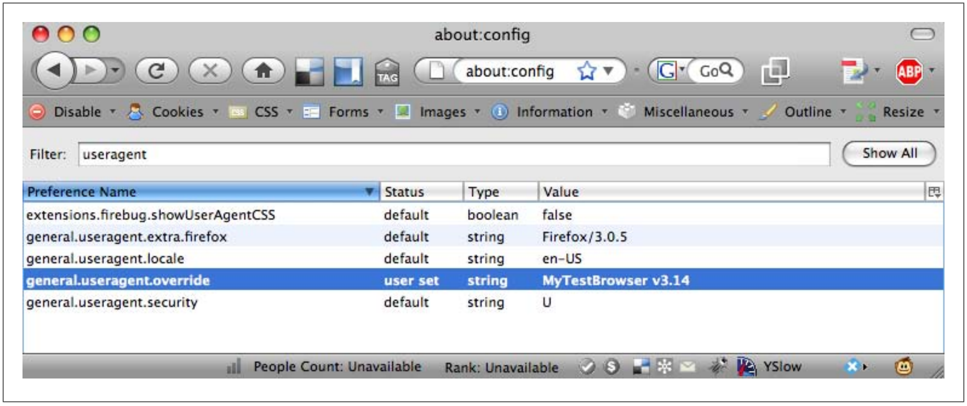
Figure 5-52. Google Analytics' Reverse Goal Path shows how people got to the goal in a funnel

Are tags working correctly?

The easiest way to check page tags is to run reports and see if pages are properly identified. If you want to examine your configuration more closely, change your browser's user agent to something unique, making it easier to identify within your analytics package.

Here's a good way to use Firefox to determine if your tags are working correctly:

1. Open a new browser window.
2. Enter **`about:config`** in the browser's address bar.
3. Right-click the blank area on the screen and select New from the drop-down list.
4. Select String to create a new browser property as a string.
5. Enter **`general.useragent.override`** for the preference name.
6. Enter a user agent string of your choosing for the value.
7. To check that it's properly configured, type **`useragent`** in the Filter field. You should see a screen similar to the one shown in Figure 5-53.



Filter: useragent

Preference Name	Status	Type	Value
extensions.firebug.showUserAgentCSS	default	boolean	false
general.useragent.extra.firefox	default	string	Firefox/3.0.5
general.useragent.locale	default	string	en-US
general.useragent.override	user set	string	MyTestBrowser v3.14
general.useragent.security	default	string	U

Figure 5-53. Editing the useragent setting in Firefox

Now when you surf with this browser, you'll be identified with the new string instead of your standard browser. You can try out the site, then run an analytics report for this user agent or look for it in logfiles to see the results.

Unfortunately, most analytics tools generate reports only once a day, which means that testing configurations can be time-consuming and iterative, with day-long waits before you can see if your latest changes are working. Some analytics packages, such as Clicky, report data more frequently, making them easier to configure and test.

Sharing Analytics Data

Once your analytics is in place and running correctly, you need to share the data with those who can make use of it.

No organization wants to be overwhelmed with dozens of reports at the outset. Giving your stakeholders open access to the entire system may backfire, as they won't understand the various terms and reports without some explanation. It's far better to pick a few reports that are tailored to each recipient and send regular mailouts. For example:

- For *web designers*, provide information on conversion, abandonment, and click heatmaps.
- For *marketers and advertisers*, show which campaigns are working best and which keywords are most successful.
- For *operators*, show information on technical segments such as bandwidth and browser type, as well as countries and service providers from which visitors are arriving. They can then include those regions in their testing and monitoring.
- For *executives*, provide comparative reports of month-over-month and quarterly growth of KPIs like revenue and visitors.

- For *content creators*, show which content has the lowest bounce rates, which content makes people leave quickly, and what visitors are searching for.
- For *community managers*, show which sites are referring the most visitors and which are leading to the most outcomes, as well as which articles have the most content.
- For *support personnel*, show which search terms are most popular on help pages and which URLs are most often exits from the site, as well as which pages immediately precede a click on the Support button.

Include a few KPIs and targets for improvement in your business plans. Revisit those KPIs and see how you're doing against them. If you really want to impress executives, use a *waterfall report* of KPIs to show whether you're making progress against targets. Waterfall reports consist of a forecast (such as monthly revenue) below which actual values are written. They're a popular reporting format for startups, as they show both performance against a plan and changes to estimates at a glance. See *http://redeye.firstround.com/2006/07/one_of_the_toug.html* for some examples of waterfall reports.

Once analytics is gaining acceptance within your organization, find a couple of key assumptions about an upcoming release to the site and instrument them. If the home page is supposed to improve enrollment, set up a test to see if that's the case. If a faster, more lightweight page design is expected to lower bounce rates, see if it's actually having the desired effect. Gradually add KPIs and fine-tune the site. What you choose should be directly related to your business:

- If you're focused on retention, track metrics such as time spent on the site and time between return visits.
- If you're in acquisition mode, measure and segment your viral coefficient (the number of new users who sign up because of an existing user) to discover which groups will get you to critical mass most quickly.
- If you're trying to maximize advertising ROI, compare the cost of advertising keywords to the sales they generate.
- If you're trying to find the optimum price to charge, try different pricing levels for products and determine your price elasticity to set the optimum combination of price and sales volume.

There are many books on KPIs, such as Eric Peterson's *Big Book of Key Performance Indicators* (*www.webanalyticsdemystified.com/*), to get you started.

Repeat Consistently

For your organization to become analytics-driven, web activity needs to be communicated consistently. Companies that don't use analytics as the basis for decision-making become lazy and resort to gut-level decisions. The Web has given us an unprecedented ability to track and improve our businesses, and it shouldn't be squandered.

The best way to do this is to communicate analytics data regularly and to display it prominently. Annotate your analytics reports with key events, such as product launches, marketing campaigns, or online mentions. Build analytics into product requirements. The more your organization understands the direct impact its actions have on web KPIs, the more they will demand analytical feedback for what they do.

If you want to add a new metric to analytics reports you share with others, consider showcasing a particular metric for a month or so, rather than changing the reports constantly. You want people to learn the four or five KPIs on which the business is built.

Start Experimenting

Think of your website as a living organism rather than a finished product. That way, you'll know that constant adjustment is the norm. By providing a few key reports and explaining what changes you'd like to see—lower bounce rates, for example, or fewer abandonments on the checkout page—you can solicit suggestions from your stakeholders, try some of those suggestions, and show them which one(s) worked best. You'll have taught them an important lesson in *experimentation*.

One of the main ways your monitoring data will be used is for experimenting, whether it's simply analytics, or other metrics from performance, usability, and customer feedback. Knowing your site's desired outcomes and the metrics by which you can judge them, you can try new tactics—from design, to campaigns, to content, to pricing—and see which ones improve those metrics. Avoid the temptation to trust your instincts—they're almost certainly wrong. Instead, let your visitors show you what works through the science of A/B testing.

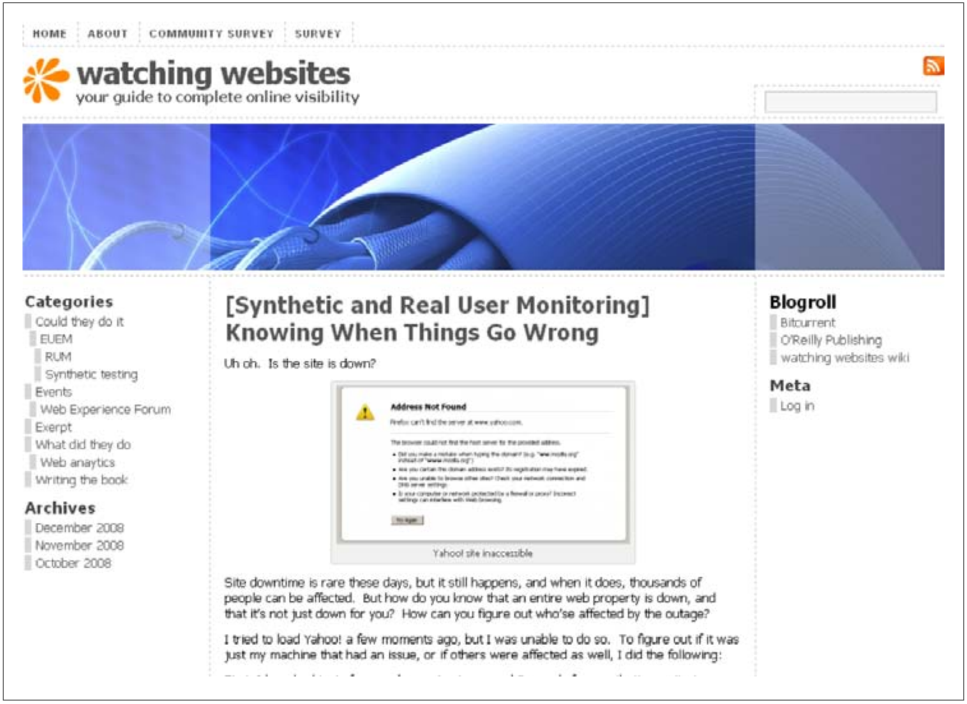
Everything Is an Experiment

You should be implementing A/B testing on some aspect of your site *every day* to maximize its potential.

In fact, any time you find yourself writing one marketing email, laying out one page, preparing one Twitter notice, picking one photo, or drawing one illustration, create two versions instead. Test on a small audience to determine which one works better, then send it to the broader audience.

This simple change will have an immediate, positive impact on all of your online marketing.

A/B testing involves trying two designs, two kinds of content, or two layouts to see which is most effective. Start with a hypothesis. Maybe you think that the current web layout is inferior to a proposed new one. Pit the two designs against one another to see which wins. Of course, "winning" here means improving your KPIs, which is why you've worked so hard to define them. Let's assume that you want to maximize enrollment on the Watching Websites website. Your metric is the number of RSS feed



[Synthetic and Real User Monitoring] Knowing When Things Go Wrong

Uh oh. Is the site is down?

Address Not Found

Firefox can't find the server at www.yahoo.com.

The browser could not find the host server for the provided address.

- Did you make a mistake when typing the domain? (e.g. "ww.mozilla.org" instead of "www.mozilla.org")
- Are you certain the domain address exists? Its registration may have expired.
- Are you unable to browse other sites? Check your network connection and DNS server settings.
- Is your computer or network protected by a firewall or proxy? Incorrect settings can interfere with Web browsing.

Yahoo! site inaccessible

Site downtime is rare these days, but it still happens, and when it does, thousands of people can be affected. But how do you know that an entire web property is down, and that it's not just down for you? How can you figure out who'se affected by the outage?

I tried to load Yahoo! a few moments ago, but I was unable to do so. To figure out if it was just my machine that had an issue, or if others were affected as well, I did the following:

Figure 5-54. The landing page, in its current form, has an RSS icon in the top-right corner

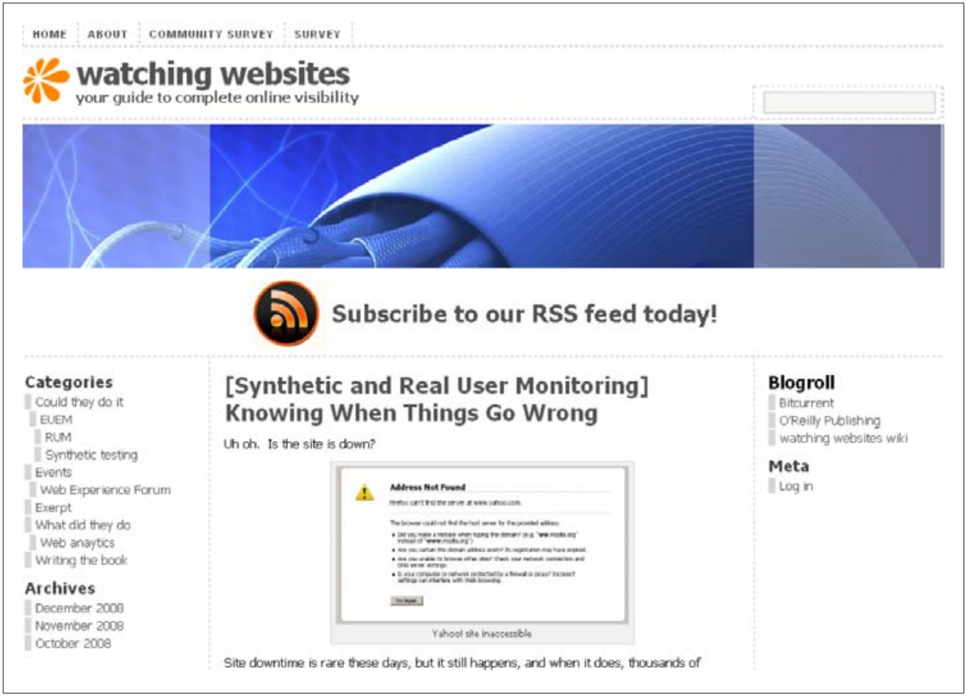
subscribers. The current layout, shown in Figure 5-54, hides a relatively small icon for RSS subscription in the upper-right of the page.

We might design a new layout with a hard-to-miss enrollment icon (called the "treatment" page) like the one shown in Figure 5-55 and test the treatment against the original or "control" design.

RSS subscriptions are our main metric for success, but you should also monitor other KPIs such as bounce rate and time on site to be sure that you have improved enrollment without jeopardizing other important metrics.

Testing just two designs at once can be inefficient—you'd probably like to try several things at once and improve more quickly—so advanced marketers do *multivariate* testing. This involves setting up several hypotheses, then testing them out in various combinations to see which work best together.

It's hard to do multivariate testing by hand, so more advanced analytics packages can automatically try out content across visitors, taking multiple versions of offers and discovering which offers work best for which segments, essentially redesigning portions of your website without your intervention.



HOME
ABOUT
COMMUNITY SURVEY
SURVEY
watching websites
your guide to complete online visibility
Subscribe to our RSS feed today!

Categories
Could they do it
EUEM
RUM
Synthetic testing
Events
Web Experience Forum
Exerpt
What did they do
Web anaytics
Writing the book

Archives
December 2008
November 2008
October 2008

[Synthetic and Real User Monitoring] Knowing When Things Go Wrong

Uh oh. Is the site is down?

Address Not Found

Firefox can't find the server at www.yahoo.com.

The browser could not find the host server for the provided address.

- Did you make a mistake when typing the domain? (e.g. "ww.mozilla.org" instead of "www.mozilla.org")
- Are you certain this domain address exists? Its registration may have expired.
- Are you unable to browse other sites? Check your network connection and DNS server settings.
- Is your computer or network protected by a firewall or proxy? Incorrect settings can interfere with Web browsing.

Try Again

Yahoo! site inaccessible

Site downtime is rare these days, but it still happens, and when it does, thousands of

Blogroll
Bitcurrent
O'Reilly Publishing
watching websites wiki

Meta
Log in

Figure 5-55. The landing page with an alternate RSS feed icon and a call to action

When embarking on testing, keep the following in mind:

Determine your goals, challenge your assumptions
: What you think works well may be horrible for your target audience. You simply don't know. The only thing you know is which KPIs matter to your business and how you'd like to see them change. You're not entitled to an opinion on how you achieve those KPIs. Decide what to track, then test many permutations while ignoring your intuition. Perhaps your designer has a design you hate. Why not try it? Maybe your competitors do something differently. Find out if they have a reason for doing so.

Know what normal is
: You need to establish a baseline before you can measure improvement. Other factors, such as a highly seasonal sales cycle, may muddy your test results. For example, imagine that your site gets more comments on the weekend. You may try an inferior design on a weekend and see the number of comments climb, and think that you've got a winner, but it may just be the weekend bump. Know what normal looks like, and be sure you're testing fairly.

Make a prediction

Before the test, try and guess what will happen to your control (baseline) and treatment (change) version of the site. This will force you to list your assumptions, for example, "The new layout will reduce bounce rate, but also lower time on site." By forcing yourself to make predictions, you're ensuring that you don't engage in what Eric Ries calls "after-the-fact rationalization."

Test against the original

The best way to avoid other factors that can cloud your results is to continuously test the control site against past performance. If you see a sudden improvement in the original site over past performance, you know there's another factor that changed things. We've seen one case where a bandwidth upgrade improved page performance, raising conversion rates in the middle of multivariate testing. It did wonders for sales, but completely invalidated the testing. The analytics team knew something had messed up their tests only because both the baseline and control websites improved.

Run the tests for long enough

Big sites have the luxury of large traffic volumes that let them run tests quickly and get results in hours or days. If you have only a few conversions a day, it may take weeks or months for you to have a statistically significant set of results. Even some of the world's largest site operators told us they generally run tests for a week or more to let things "settle" before drawing any important conclusions.

Make sure both A and B had a fair shake

For the results to be meaningful, all of your segments have to be properly represented. If all of the visitors who saw page A were from Europe, but all of those who saw page B were from North America, you wouldn't know whether it was the page or the continent that made a difference. Compare segments across both the control and the treatment pages to be sure you have a fair comparison.

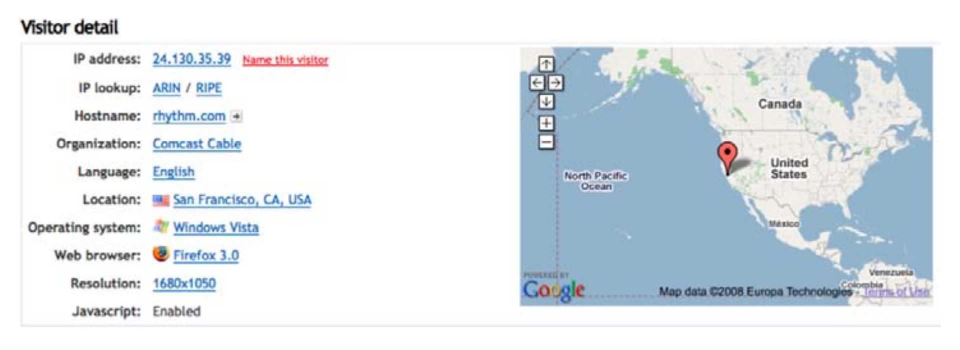
Repeat and ramp

Once you've completed a few simple A/B tests, you can try changing several factors at once and move on to more complex testing. As an organization, you need to be allergic to sameness and addicted to constant improvement.

For more information on testing, we suggest *Always Be Testing: The Complete Guide to Google Website Optimizer* by Bryan Eisenberg et al. (Sybex) and "Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO" (*http://exp-platform.com/hippo.aspx*).

Choosing an Analytics Platform

Once you've factored in the collection methods, kinds of reports, and testing capabilities you'll need for your analytics, it's time to choose a platform.



Visitor detail

IP address:	24.130.35.39 Name this visitor
IP lookup:	ARIN / RIPE
Hostname:	rhythm.com
Organization:	Comcast Cable
Language:	English
Location:	San Francisco, CA, USA
Operating system:	Windows Vista
Web browser:	Firefox 3.0
Resolution:	1680x1050
Javascript:	Enabled

Figure 5-56. Visitor detail in Clicky Analytics

Free Versus Paid

Eric T. Peterson (*http://blog.webanalyticsdemystified.com/weblog/2007/07/the-problem-with-free-analytics.html*) concluded that if you've deployed a free analytics solution, you're probably:

- Only casually making use of web analytics
- Understaffed in the analytics department
- Lacking experience with web analytics in general

Far from criticizing free tools themselves, Peterson makes the point that you're much more likely to take analytics seriously if you've paid for it. What's more, you get what you pay for—free tools provide information, but require much more manual effort on the part of marketers to make and test their changes. In a small organization, analytics often becomes a lower priority than selling or design.

When you pay for something, you'll have access to support, and you may even be able to impact the road map of the product, depending on how common your request is or how much clout you have with the company providing the service.

On the other hand, Brian Eisenberg, CEO at FutureNow, Inc., says:

> My philosophy has always been to "get good at free then pay." There's no sense paying for something until you really operationalize its use. With today's free tools offering 65–85% of the functionality of high-end tools, I am not sure free is only for the causally involved. About 30% of paid implementations also have Google Analytics or Yahoo! analytics installed.

Real-Time Versus Trending

At the entry level, analytics tools can be divided into two broad groups. Real-time tools, like those from Clicky (Figure 5-56), show you what's going on right now, and are the closest you can get to spying on individual users, giving you their locations, IP addresses, and so on.

Real-time tools concentrate less on goals, outcomes, and conversions. They are useful for root-cause analysis and quick answers (who tweeted that message, who blogged about me first, and so on), but are not designed for long-term decision making.

Goal-oriented analytics tools, on the other hand, have less detailed information. They tend to be tied to backend business models (for example, Yahoo!, Google, and Microsoft's tools automatically integrate with their search and paid search keyword systems). This is consistent with their business models. Services like Google Analytics want you to convert more people through AdWords so that you and they can make more money. Therefore, companies who create these tools want to help you optimize conversions without getting down to individual users.

Some solutions may offer both real-time and long-term perspectives.

Hosted Versus In-House

If you need to keep data to yourself for legal or ethical reasons, or if your visitors' browsers don't connect to the public Internet, you may have no choice but to run your own analytics platform.

If you can use a third-party solution, we suggest you do so—you'll get faster development of features and useful functions, like comparative reports that show how you're faring against others. The only exception to this rule is if you have a business that's tied tightly to some custom analytics.

Data Portability

If you can't get a copy of your data from your analytics provider, you can't leave. Having the ability to bring your data in-house—or better yet, to import it into an alternate analytics provider—means you can negotiate better pricing and keep your service providers honest. This alone is a good reason to care about data portability.

There's an important second reason that's central to the theme of this book. As we'll see in the closing chapters, you'll want to combine analytics data with other information in order to make better decisions about your business. Your analytics solution should either be able to import third-party data for analysis, or export analytics data to a data warehouse you run so that you can analyze it yourself.

The Up-Front Work

Analytics can be hard work, but it pays off. How much time and effort you invest will dictate what you get in return.

What You Get for Free

If you simply add JavaScript for analytics to your website, you'll get basic traffic metrics, some fundamental behavioral data (such as bounce rates) and Top-N reports that show you the most common landing pages, exit pages, and traffic sources.

What You Get with a Bit of Work

By adding goals to your analytics system, you can start to sort traffic according to business outcomes. Your top-N reports will now be about goal attainment, rather than being simply about popularity. You'll not only know who's sending you visitors, but who's sending you the *right* ones. You'll also know the most common paths through your site and where people are leaving.

What You Get with a Bit More Work

By creating custom segments and tags that add context, you'll know more about what's happening on the site. You can segment traffic and understand which products, offers, or users behave in ways you want, and which don't. This focuses your marketing efforts, site design, advertising, and keywords selection.

What You Get with a Lot of Work

Using clients with custom event monitoring, you can see how visitors interact with your site. This isn't for the faint of heart—it requires modifications to JavaScript and a lot more testing. You can also automate multivariate testing (or buy analytics tools that automate it for you), so you're constantly testing competing layouts, offers, campaigns, and content. While this is a lot of work, the result is a site that's constantly and automatically adapting to what its visitors want.

Web Analytics Maturity Model

Throughout this book, we're going to look at a maturity model for web visibility. The first of these models is web analytics, shown in Table 5-3. It borrows heavily from work by Bill Gassman of Gartner and Stephane Hamel of Immeria, and shows how companies progress through various levels of maturity with their web analytics.

Table 5-3. The Web Analytics Maturity Model (adapted from Stephane Hamel and Bill Gassman)

Maturity level	Level 1	Level 2	Level 3	Level 4	Level 5
Focus?	Technology: make sure things are alive	Local site: make sure people on my site do what I want them to	Visitor acquisition: make sure the Internet sends people to my site	Systematic engagement: Make sure my relationship with my visitors and the Internet continues to grow	Web strategy: Make sure my business is aligned with the Internet age
Who?	Operations	Merchandising manager	Campaign manager/SEO	Product manager	CEO/GM
Analytics	Page views, visits, visitors, top ten lists, demographics, technographics	Path analysis, funnel reports, A/B testing, KPIs, dashboards	Merchandising, segmentation, SEO, community referrals, campaign optimization, personas, KPI alerts	Multichannel aggregation, cost-shifting analysis, lifetime visitor value, personalization, dynamic content serving	Multichannel sales reporting, activity-based costing, balanced scorecards, strategic planning, predictive analytics, integrated user experience

Most organizations begin by looking only at traffic. They then turn their efforts inward, trying to get their own sites in order and defining KPIs that should guide their improvements. Once the site is converting its visitors and encouraging them toward the goals you want, organizations focus outward to increase the amount of traffic that's coming in. As the site grows, they incorporate additional data from other sources, such as performance, call center traffic, visitor lifetime value, user feedback, and so on.

Truly mature organizations move beyond even this level of integration, making web analytics a part of their strategic planning process and running their businesses through KPIs. Web metrics become a part of performance reviews and business unit goal-setting, and companies start to use analytical tools to look forward as well as back.