



Web Analytics Tool Evaluation and Selection

Writer: Gary Angel, President, Semphonic
Published: February 19th, 2007

Summary: The things that turn out to matter after you've bought and installed enterprise software are almost never the things you evaluated the software for in the first place. In web analytics, the feature checklists used by most organizations provide little or no real differentiation between vendors. And the features vendors tout are often of little or no value. This white paper – written based on years of hands-on experience with web analytics and with all of the leading web analytics solutions – will provide real guidance about the core features that actually matter in web analytics. It will help potential buyers understand where to look for differentiation, how to decide if the features in question will matter to them, and how to better classify web analytics tools and decide which ones fit their “sweet-spot.”

For More Information Contact: Brandon Brock (bbrock@semphonic.com)

Copyright: Semphonic, 2007



Table of Contents

Introduction.....	3
Key Features.....	4
Visitor Segmentation.....	6
Dimensional Reporting.....	11
Management Reporting.....	15
Setup.....	18
Page Hierarchy.....	20
Summary.....	24

Introduction

Web Analytics software has changed dramatically in the past three years. New vendors have emerged. Software solutions have improved dramatically. Companies have grown and evolved. As an industry, web analytics tools have grown up. The tools are uniformly better (by a dramatic margin) in almost every respect. From usability to performance to reliability, the available tools have gotten significantly better. But nowhere has the evolution of enterprise web analytics software been more dramatic than in the feature sets most packages support.

Enterprise web analytic packages have been transformed from skeletal reporting systems whose sole virtue was that they handled web specific data into powerful reporting and analysis packages. Of course, vendors constantly leap-frog one another and race to fill-out feature lists. Inevitably, the solutions come to seem remarkably similar when compared at the check-list level. But this check-list level can hide dramatic differences in approach and philosophy (as well as performance and reliability) that make all the difference in the real world.

In this White Paper, there will be no attempt to produce an exhaustive checklist of features than any client might consider. Instead, the goal will be to look at some of the most important features in any web analytics package and try and help the reader understand how to really differentiate them.

Finally, a word of warning – don't expect this white paper to end with a neat and tidy vendor classification along with a set of grades or recommendations. On the whole, any of the Enterprise-class solutions on the market today are acceptable. Each is more appropriate to certain kinds of sites, business cultures, and analytic and reporting needs. And, given the pace of change, any vendor classification may be obsolete within days of publication. But knowing what features are really important and how to evaluate the depth and quality of their implementation will always be valuable.

Key Features

There are so many different reports and capabilities in web analytics tools these days that it's vitally important to understand which ones are essential, which are useful and which just help flesh out the menu structure.

When you look at an industry check-list, chances are you'll see lots of reports and tools listed. Most of them are useless. So the fact that a vendor does or does not support the report isn't especially interesting. And since most vendors will support all of these reports, you'll see long rows of checks that may make it seem like two very different products are almost identical.

Based on years of experience doing real-world analytics and helping some of the largest web sites in the world implement and use web analytic tools, here are the key features that really seem to make a difference in what you can do with a product:

Visitor Segmentation: Nearly every analysis (and every report) you might ever want to do is going to involve some kind of visitor segmentation. There is no web analytics tool capability that is both so ubiquitous and yet implemented in so many different fashions. The way a tool chooses (or can) implement visitor segmentation is a defining factor in its power and usefulness.

Dimensional Reporting: Most web analytic tools are traditional OLAP-cube tools. Their core functional capability is to provide frequency and cross-tabulations of various web behavioral variables. But web analytics presents some unique challenges to traditional OLAP designs – and how well a tool can surmount or avoid those challenges has a dramatic impact on how flexible and rich the reporting interface turns out to be.

Management Reporting: Much of what vendors say about the ease-of-use of their tools is, in fact, quite true. But ease of navigation doesn't translate into ease of application. Web analytics is hard. The terminology is often confusing. The work required to produce interesting results takes more time and more tool knowledge than most potential users ever have. The idea that you can just hand the keys to your enterprise tool to your marketing managers has been thoroughly discredited. This means that you will almost always need to move data from the core application into either Excel or custom dashboards. The choices a product makes here

Web Analytics Tool Evaluation and Selection

dramatically influence how well the tool will support real-world management reporting.

Setup: Most enterprise systems these days are tag-based (but there are important exceptions). Even within tagging systems, however, there are some vitally important differences in the weight of the tag that translate into critical advantages in ease-of-implementation, analytic flexibility and analytic sophistication. Understanding the implications of what goes into the tag will help many a company choose the solution most appropriate to their needs.

Page Hierarchy: The ability to group pages together, analyze them and work with them as a unit is a little understood but core capability in web analytics. With so much attention to Web 2.0 and driving analysis below the page level, it is ironic that one of the biggest weaknesses in many analytics tools is the ability to look at things ABOVE the page level. Depending on your site and CMS system, the capabilities of a package to group pages, name them and analyze behaviors related to them may be one of the real differentiators between solutions.

Visitor Segmentation

What is visitor segmentation? It's the ability to subset the visitors (or visits) on your web site and look at just the behaviors of that subset. Typically, vendors implement visitor segmentation as a meta-cut on the data. You apply a visitor segment and then all your base reports and analytic tools reflect just the data from the selected group.

Naturally, this leads to a core set of necessary capabilities. These include the types of segments that can be built, the kind of criteria that can be used to create segments, the range of analysis available on a subset, and the speed of creation and total number of subsets the analyst has available.

Here's a more detailed break-out of the essential visitor segmentation capabilities.

Segment Creation

- ✓ Can segments be created without tags
- ✓ Segmentation Logic: can full logical operators be used to define segments
- ✓ Segmentation Variables: what data can be used to define segments
- ✓ Can external data be used natively and combined with web data in segment creation
- ✓ Can segments be created via data-driven techniques like neural networks
- ✓ Can segments focus on visit or visitor behavior
- ✓ Can segments be defined based on time and event sequences
- ✓ Can distributions be produced on key behaviors to assist in segment creation

Segment Methodology

- ✓ Are segments samples or against all data
- ✓ Are segments created in real-time or delayed

Segment Creation in Web Analytics

The ability to create segments without tags is a fundamental capability and no one should consider a system that doesn't provide this. Segments are the analyst's primary tool. But they change for almost every single analysis. You won't use the same segments over and over – you'll need new segments every time with highly specific behavioral definitions. Tagged segmentations are inherently static (as well as limited to the simplest of behaviors). And the cycle time – even in the rare case where a tag can be changed – makes this method completely impractical.

Segmentation Logic

Segmentation logic has less obvious impacts than tagless segmentation. It is, nevertheless, the area where most of even the best analytic tools have serious shortcomings. In an effort to protect you from your presumed stupidity (or their performance problems) most analytic solutions have a simple way to build segmentation logic – the set of behaviors that defines a visitor segment. By far the most common method is this: you can define multiple conditions; each condition is an *AND* that must be met for the visit/visitor to be included; within each condition you can set multiple values that meet the criteria. These values are implicitly *OR* values – if the visit/visitor meets any one of them then the condition is satisfied. This sounds pretty powerful. You can build logic like this: I want any visitor who viewed 5 or more page views in the Product section of the site *AND* started and order *AND* didn't place an order. But as powerful as this seems, there are many, many conditions that can't be created. Here's one common one: I'd like to create a Visit based Prospect segment defined as anyone who has 0 Purchases *OR* who became a customer in this session. Oops. That *OR* is between two types of conditions. Can't be done with this type of criteria definition. Or suppose I'd like to say that my Engaged Prospects are any visitor with 5 or more Page Views *OR* 2+ Visits. That's an *OR* across conditions. Again, can't be done. There are a range of different criteria builders out there, but there's only one that would really make sense – full logical control including *AND* and *OR* plus parenthetical grouping. No single deficiency in web analytic tools is as consistently frustrating as this seemingly small feature – a feature that isn't even on most checklists out there.

Segmentation Variables

A related but less ubiquitous source of frustration is the types of data that can be used for segmentation. Most tools that provide segmentation do a fairly reasonable job of opening up variables. But many hidden limitations exist – particularly if you want to use segments built from external data as part of an additional subset.

Data Driven Segmentation

This is more a wish-list item than check list item because nobody actually supports it. But any modern data analysis system should provide data-driven visitor segments using tools like neural segmentation. Any rule-based segmentation is deeply flawed when it comes to combining complex constellations of behaviors. Someday, this will be an important capability. But for now, just assume that this is an empty check box on every tool in the marketplace.

Visit and Visitor Behavioral Segmentation

Web Analytics Tool Evaluation and Selection

Web behavioral analysis is somewhat unusual because it does often focus on two very different levels of analysis – the visitor and the visit. Of these, visitor segmentation is probably more important. Segmentation over time across sessions is absolutely essential to many, many types of analysis – and any tool that doesn't support this capability shouldn't be your short-list. That being said, the visit is often an interesting unit of analysis in its own right. Segmentation criteria should be definable for both Visitors and Visits – and ideally should be able to isolate segments like this: I want all visits without a purchase by Visitors who have made a purchase.

Time-Based Segmentation

Virtually every single web analysis you might actually need involves a time component. What did visitors do in the week after they signed up for a class? What did visitors do in between registration and purchase? What did visitors do in the three days after downloading a trial? What did visitors do in the two sessions prior to buying a video? What was all of the behavior of a Prospect up to an including a Purchase but not after? All of these questions are basic, simple and obvious. But answering them is nearly impossible unless a web analytic tool lets you define segments based on a time component that allows you specify criteria relative to dates and or specific actions on the web. This capability is hardly supported in the existing marketplace.

Distribution Reporting

The ability to view distributions of behavior (as in how many visitors viewed this page 1 time, how many 2 times, how many 3 times, etc.) is both a key reporting and segmentation capability. Why is this a segmentation capability? The web analyst needs to supply the rules to build segments – and discovering those rules without distributions is much, much harder. Imagine an operations site with a key page – submit a form. What's a heavy user of the form? You can look at the average page views and see that's its 3.5 views per visitor. But what's the distribution? It makes a huge difference. On many customer operations sites, the distribution may be a radical U – with high spikes at 0 and 1 and a shocking spike somewhere out in the 100+ range. But if you don't know what the distribution is like, you have no way to build appropriate segment definitions. So you have to experiment – how many visitors do I get if I say 5+ form submits. Too many? Let's try 8. And so on. The distribution of visitors (and visits too for that matter) around a behavior is an essential descriptive view for an analyst. And unlike time based segmentation criteria it is not totally without support in the tool world!

Segment Methodology

There are two main divisions in segmentation methodology. The first main divide is between sampling-based segmentation systems and those that

Web Analytics Tool Evaluation and Selection

use all of the data (comprehensive). The second issue is whether segments are created in real-time or delayed.

These aren't necessarily either/or decisions. Some tool sets have various components and approaches that span pretty much every combination of these alternatives – and there are some pretty good reasons why each has a place.

Sampled or Comprehensive Data

Let's start with sampled data vs. comprehensive data. Vendors provide sampled data for one simple reason – performance. It's much easier to deliver fast answers against sampled data than it is against comprehensive data – especially if you're talking about a large web site with hundreds of millions of requests monthly (or more).

In a perfect world, you'd like to have near instantaneous analysis against comprehensive data. But this isn't, of course, always possible. So the real question is how much you lose when you employ a sampling methodology.

On the whole, sampling solutions are very viable. Sampling, done correctly, can almost always provide answers that are near-enough - especially given the built-in slop factor inherent in web analysis. New users of web analytic solutions are frequently (and rightly) put off by the fact that “nothing ever ties!” Rightly or wrongly, though, you get used to some level of imperfection. Indeed, sampling has the virtue that it puts your expectations about the data in a firmly reasonable place.

Compared to some other methods of data trimming (like dropping infrequent paths), sampling is very much to be preferred. Sampling rarely distorts the data into unrecognizable forms – whereas data trimming will frequently do just that in situations where the data has a very long tail.

A great deal of customer segmentation is for purely analytic purposes – not to support management reporting. And for analytic purposes, the difference between sampled data and comprehensive data is often unimportant.

It is also convenient to be able to check samples against comprehensive data – to either validate conclusions or spot-check for cases where your sampled data is returning suspect answers.

Real-Time vs. Delayed Segmentation

There is a deep relationship between sampling and this issue. Real-time segmentation may be impossible without sampling – so one of the biggest

Web Analytics Tool Evaluation and Selection

potential benefits to sampling is enabling the analyst to make and report on segments without having to wait hours or days.

How big a deal is this? It's important. If you are using segments to support management reporting, you probably won't care much about this. If you're going to be using a segment for the next couple of years, it doesn't much matter if it takes a day or two to create. But most segmentation is for analytic purposes – and needs to be responsive to changing needs. What's more, an analyst often doesn't know if a segment is going to be useful. So if you have to wait a long time to view the results of a segment, it can make the cycle times on analysis frustratingly long. This is especially problematic if your system places caps on how many segments you can create (this is pretty common when segments are being built on a vendor's data warehouse). Use up your quota of segments because of segment definition errors, mistakes in judgment and just plain wrong guesses about what might prove interesting and you are out of luck.

Dimensional Reporting

If you don't have a technical background, it can often be hard to understand why web analytics tools often have strange and apparently unreasonable restrictions on the types of reports or analysis you can do. But as arbitrary as these restrictions often seem, they are borne of a very real and fundamental set of problems that all arise from a few simple facts about web measurement that may actually have considerable influence on your tool selection.

Fact number one is that web analytic solutions intended for big enterprises often have to deal with almost unprecedented amounts of detail data. Big web sites generate billions of rows of data in a year. It's a lot. Even the amount of data for a single day can be enormous. Because the raw data set CAN be so large, tools often can't take the approach of using the low-level data to answer every question. Instead, many web analytic solutions rely on a technique developed well before web analytics to handle very large data sets in other fields. This approach (OLAP is the techie term) involves the creation of data cubes that pre-summarize specific relationships in the data.

But a cube can't capture every possible relationship in the data – or it gets to be bigger than the original raw data. So the cube builder (in this case your web analytics vendor) has to make some tough decisions about what data to include. In some cases, these decisions are driven by basic analysis of what the vendors think is likely to be important: screen size by referring site – no; conversion by campaign – yes. But there are other factors to consider and the biggest is something called cardinality. Cardinality is a measure of how many different values a variable can have. OLAP always worked great for basic customer analysis because most of the variables had very low cardinality (gender – two values; age – no more than about hundred values and often reduced into four or five categories; income – usually reduced to three or four categories). The lower the cardinality, the more data compression you get with a cube and the more variables you can cross-tabulate and make available in the reporting. The higher the cardinality, the less performance benefits a cube will provide unless lots of variable cross-tabulations are eliminated.

Unfortunately, and here's key fact number two, web analytics is filled with key variables of very high cardinality. These include page names, content sections, search terms (internal and external), referring sites, paths, and unique visitors. Most of the things you actually care about on the web have thousands, tens of thousands, hundreds of thousands or even millions of different values.

Web Analytics Tool Evaluation and Selection

So one of the biggest differences between web analytic solutions is the approach they've taken to solving the "big" data issue and how clever they've been at getting around the inherent limitations of their approach.

There are solutions that try to solve the "big" data issue by tackling the raw data directly. This approach has lots of advantages – because it means they don't ever have to say "no" to a question or report. The interface can give you almost unlimited access to the data. Indeed, if the data is stored in a truly open system like Oracle or SQL-Server then even when the interface can't do something the data is still readily available. This is a fantastic approach and very attractive. Until, that is, the performance isn't good or the hardware to get good performance is more than you can afford. If you don't have enormous quantities of data or you are prepared to invest in sufficient hardware, then direct to data solutions will almost always be significantly better than OLAP solutions.

If you do have lots of data and more limited budgets, however, then a big part of how attractive a web analytics solution is going to be is how good a job they've done of hiding the OLAP limitations. If your organization is unusual in size or structure or analytic requirements, then it is vital that you insure that the variables important to you are in the data in their entirety.

Here are some of the common short-cuts analytics vendors take that can damage your ability to do analysis:

Cropping paths

Sites can generate a staggering number of unique paths. Does the solution crop paths (show only top x thousand)? This can be crippling for sites with lots and lots of low volume but important pages – publishing sites being the most common case. If you are a large site and your vendor is cropping paths you will routinely find that your pathing tool is completely useless.

Cropping pages

Most sites have thousands of pages. No problem. But if you have hundreds of thousands of pages your vendor may not report data on low-volume pages.

Cropping search terms

As with pages, the thresholds here tend to be in the hundreds of thousands of different search terms. If your site is heavily dependent on search AND has very high spread in term usage (neither of which is actually all that unusual) then a vendor may not show you data for all your terms.

Dimensional Reports

Many vendors restrict which variables you can cross-tabulate. Not including reporting cuts for numeric variables like avg. page time, for example. This isn't often an issue but if your data needs are unusual, then these kinds of limitations can be fatal.

For visitors and visitor segmentation, here are two of the most common short-cuts vendors take:

Uniques Reporting

Daily uniques are almost useless. Vendors often don't de-dupe visitors except at the daily level – and almost never at custom levels. What's more, they don't always provide de-duped visitor counts at weekly and monthly levels. And when they do, they often don't do it consistently across all reports. This creates more errors, havoc, and user confusion than almost any other area in web analytics. It is inherently confusing and the vendors almost never tell you in any given report what level of de-duping has been applied.

Segment Building

You don't always have access to every variable you'd like when it comes to creating segments. In most cases, you'll get all the things you need but there will be exceptions. Make sure, if you're a commerce site, that you can build segments from e-Commerce variables – both on an individual order basis and in terms of total customer spend. Make sure you can build segments from custom variables. And make sure you can build segments based on both individual session variables (like referring site) and lifetime customer variables like original referring site or original campaign).

Vendors who rely primarily on cubes will sometimes provide a tiered approach to data access. The cube will support basic reporting and analytic requirements with interactive report generation, while the user will also have the ability to submit longer running queries to a data warehouse. This can be a very effective approach, but it, too, raises important issues. Here are the biggest "gotchas" with tiered implementations:

Limited Data Access

Just because you're getting access to the "data warehouse" doesn't necessarily mean you're getting access to the raw data. Sometimes, you're just getting query access to the cube. That has uses, but one of them isn't answering any new questions. This really isn't a tiering strategy at all – just a way of providing automated programmatic access to the cube.

Limited Query Access

Remember that in most cases, you're running your query on a vendor's data warehouse. So vendors protect themselves against overuse or bad queries by providing either restricted SQL – with key capabilities removed – or a restricted interface through which queries must be constructed. This means that while you may have significantly improved data access, there are probably lots of queries you really won't be able to make.

Queuing

Long queries are going to take a long time. Fair enough. But don't assume that just because queries "usually process over night" that your queries will. If this is an important part of your overall analytic needs, then make sure you get Service Levels that match. For larger sites, service levels are not always what the vendor implied during the sales process.

Management Reporting

Few areas of web analytics are as consistently misunderstood as management reporting. Unfortunately, these deep misunderstandings translate into errant views about what matters in a product and some very bad decisions about how to roll-out the tool they select.

Here are some of the most common mistakes made by companies:

- Thinking that every stake-holder should have training and access to the tool
- Thinking that in-product dashboards are a complete solution to management reporting
- Thinking that metrics need to be actionable
- Thinking that reports and analysis are the same thing

Each of these very basic mis-conceptions can have dire consequences for appropriate tool selection and implementation.

Training and Tool Access

Have you ever heard a vendor say “You know, our tool is pretty hard to use”? Probably not. The analytics vendors have done a very nice job of software engineering. But they have not (and probably never will) produced a system that is really appropriate for most managers. Web Analytics tools are for analysts. There are a variety of reasons for this – including difficult terminology, the complexity and richness of reports, and the time it takes to do real analysis. But the essential point is simple – most stake-holders in an organization need reports not tool access. That means that management reporting functionality is one of the essential features of a web analytics tool. Yes, we call them web analytics tools. But for many organizations, it would be much more accurate to say that they are web reporting tools. If you under-value this function because you believe that your Marketing Manager can just log-in to the tool and get data you are going to seriously misjudge your needs.

Dashboards are all you need for Management Reporting

In theory, this might in fact be true. In practice, it never is. There are lots of reasons why dashboards are very limiting. First, you often need to add and tune data to your reports. Dashboards are poor for this. Second, historical data is absolutely essential for almost every management report. Dashboards are notoriously limited in handling trending well. Third, most reports end up containing quite a bit of information. That often means excruciating performance if the report has to be generated each time it is accessed. Fourth, people have other uses for the data – and need to be able to move it as easily as possible into their other reports.

For all of these reasons, and more, we generally favor Excel-based reporting and consider Excel Integration as a feature of very high importance when evaluating solutions. If your needs are reporting focused, then a high-quality Excel Integration should be one of the most important items on your list.

Metrics must be Actionable

No single metric is ever actionable. Believe it or not, there is no such thing as a “good” conversion rate. Trust us. It’s true. A complete discussion of this issue is far outside the scope of vendor selection. However, what does impact tool evaluation is a mistaken focus on reporting as the ability to display a bunch of metrics in a single report. Instead, report-building should be focused on providing a story about a business problem this is rich in highly relevant information.

We like to describe reporting as a process of gradually accumulating more and more context around a problem. And what we’ve found is that some kinds of information are almost always essential to that exercise. Trended data, baselines, competitive information and cost data are among the critical reporting values that are often essential and either missing or poorly supported in Dashboards.

Reports and Analysis are the Same Thing

This is often a corollary to the “Metrics must be actionable” mistake – and tends to confuse people during both tool selection and implementation. During tool selection, this mistake can lead to very basic errors in tool selection. Some web analytics tools are very strong for analysts but very poor for managers. Others are very much the reverse. When you’ve become confused about who you are trying to serve, it’s easy to get a tool that’s totally unsuitable to your real needs. If you don’t have a web measurement group (and aren’t planning on hiring someone like us), then you don’t need a deep analytic solution. You need a good reporting solution. Similarly, if you are trying to support an aggressive analytics team, don’t expect that they are going to be happy using a tool whose focus is on nicely formatted reports and Excel Integration.

This basic mistake is often even more disastrous after vendor selection. Many, many companies have implemented a good enterprise web analytics solution, set up a report set, and then sat back and waited. A few months later when nothing has changed and nobody has used the data in any meaningful way, they begin to wonder what’s happened and blame the tool. Wrong. Reports are not analysis. Reports may drive analysis – but they don’t drive action.

Web Analytics Tool Evaluation and Selection

Given these basic points, here are some of the key features you do need think about when it comes to Management Reporting:

All Report Items in Dashboard

Even though Dashboards aren't the be all and end all of management reporting, they are still an essential piece of the puzzle. And one of the most important tool capabilities when it comes to Dashboards is the ability to use any report in them. This should include the ability to apply filters (like Search Filters or Segment Filters). Without this capability for near-universal inclusion, you'll almost certainly find that the things you really want to capture in a Dashboard aren't somehow supported.

Separate Dashboards into Slides for Presentation

When you need to deliver a fair amount of information in a report, you don't want to have to jam it all into a single dashboard – and yet you want the report consumer to be able to move through the Dashboard logically and consistently. The ability to carve up Dashboards into online slides is critical in this regard. It can also help address a persistent problem with most dashboards – very slow loading.

Ability to Save Dashboards and Deliver by Schedule eMail

It's a mistake to let most senior managers near a web analytics tool – even the dashboards. That – and sheer convenience – are the two reasons why it's important to be able to deliver dashboard views outside the tool. In addition, the ability to schedule and eMail deliver dashboards make it possible to insure that appropriate information is PUSHED to the right places. It's a lot easier, otherwise, for it to be forgotten.

Rich Trending in Dashboards

Historical data is at the heart of nearly every decent report set. And historical views are rarely as simple as you might think. It's essential to be able to trend historically the way you want – including a flexible way to specify periods (last x plus year ago), ability to mix periods in a single dashboard, ability to specify side-by-side period views, and the ability to calculate and show change rates dynamically.

Fast Performance on Load

No matter how wonderful your dashboards it will all be for naught if they take too long to load. Because dashboards often combine four or more different reports, they can be painfully slow to refresh. This is a basic usability issue – and because it is a prevalent problem you need to be aware of it when evaluating tools.

Excel Integration

Web Analytics Tool Evaluation and Selection

Excel is the preferred method for delivering many types of reports. It has the advantages of being powerful, integratable and well-known. Having easy Excel Integration (a plug-in that allows direct access to the web analytics data from Excel) is a major plus in reporting capabilities. In addition, it's important to insure that the Excel integration provides access to all of the reports that are available in the interface. Building a good report set will often strain the capabilities of the interface – so it's important that the Excel tool not further trim your data options.

Ability to Schedule and Deliver Excel Reports

As with Dashboard delivery, this is largely a matter of convenience. Many Excel reports need some tweaking anyway, but where you have canned Excel reports it is a considerable convenience to be able to automatically push them to the appropriate audience. So much of an analysts time is often taken up with just refreshing reports - having this capability can dramatically improve the productivity of your analysts.

Setup

Every vendor will tell you how easy it is to implement their solution. “Just slap a tag on and you’re done.” The bad part of this is that it’s actually true. You can just slap a tag on and be done. That’s why there are so many bad tagging implementations from which good measurement is virtually impossible.

Vendors short-sell themselves with this approach. For an up-front win (see how easy it was!) they buy a long-term mess.

This basic simplicity often hides what are fairly dramatic differences in the “Weight” of the tag – how much information needs to be put in the tag and how complicated it is to get it there. Tagging is usually an exercise in simple Javascript coding. So the programming aspects aren’t really that difficult. But to really support measurement on a large or complicated site, you need to make many difficult decisions before building the tag.

Be aware that sometimes differences in tagging complexity reflect real differences in analytic capability. One system may indeed be easier for the simple reason that it doesn’t provide the same capabilities.

Here are some of the most common differentiators between systems that you should evaluate if you are trying to see just how difficult tagging is likely to be.

Hierarchy Specified in Tag

A hierarchy is a grouping of all the pages on the site into a multi-level taxonomy. It is an important capability for analysis. But if you have to specify the hierarchy a page belongs to in the tag you may either be in for lots of work or be unable to use your tool’s hierarchy capabilities.

Campaigns in Tag

This can be quite a complex set of issues. If you manage lots of campaigns, then chances are you will want to look at campaigns rolled-up in a number of different ways. Getting that roll-up information into your web analytics solutions is non-trivial and usually isn’t done through the tag. More commonly, you use a single campaign identifier in the tag. Then you append the information either on the back-end or through the interface. Either way, it’s something you’ll need to spend some time on. In addition, just because a vendor has “tagless” campaigns doesn’t mean you can specify the roll-ups you may need. Nor, in this particular case, is tagless always an advantage. Many measurement organizations struggle to find out about new campaigns. If they need to create a campaign before it goes live and nobody bothers to tell them, then the data is often lost.

Visitor Segmentation in Tag

For the most part, you shouldn't have to build visitor segments into a tag. Segmentation should be a tagless capability. But there are times when tagged visitor segmentation is necessary (or desirable) in a system. Not every attribute of a customer/visitor can be reconstructed from information known by default to the web analytics vendor. If you need to pass information about customer class, customer duration, or any other persistent variable, then you'll have to implement this in the tag. Depending on your vendor, this could be as segments or custom variables.

eCommerce Information

Everything said about campaigns could pretty much be repeated here. Any eCommerce company will have to do a modest amount of custom coding to implement a decent tag for their cart and checkout. This isn't actually a big differentiator between vendors. But if you have a large number of SKUs, then you need to be especially cognizant of this. Like campaigns, products usually need to be analyzed by multiple roll-ups. Most vendors support at least some roll-ups. Unfortunately, you don't always have the roll-up information when the tag fires. That means your information needs to find its way into the system either through the back-end or the interface. Not all vendors support this and, if they don't and it's among your needs, it is a fatal lack.

Custom Variables

Most systems support at least some ability to pass custom variables in the tag. This is generally a good thing – since it implies more analytic richness. However, where a vendor has implemented a tagless solution for a specific problem (like hierarchy) – that is usually to be preferred. Deciding how to use your custom variables is never a slam dunk and most sites dramatically under-use the slots available to them.

Page Hierarchy

For most web sites and web analysts, the basic unit of analysis has always been the page. And with the advent of widespread multivariate testing and Web 2.0, analysts are concerned about how to drive that level down a notch. There's nothing wrong with that – both multivariate testing and Web 2.0 require that approach. But the simple truth is that lots of web site analysis and reporting is actually more interesting when you take the analysis UP one level – and look at behavior by groupings of content and not individual pages.

If your web site is less than a hundred pages, you can ignore this set of issues. For small web sites, Content Groupings just aren't that interesting. But for web sites that contain thousands or tens of thousands of pages, the vast majority of individual pages are not usefully analyzable as separate entities.

Here are some examples of how, when and why content groupings are so important.

- A product manager wants to know how many visitors looked at his product in the last month, how many looked at any of his product's customer support pages and how many looked at both.
- Your marketing manager wants to be able to target visitors whose primary interest (measured by total page views) is in Product X.
- Your marketing manager wants to be able to target visitors whose primary interest (measured by total page views) is in Product Category Y.

Without rich content grouping capabilities, answering them can range from very difficult to nearly impossible.

The difficulty is that in every case the question necessarily involves a basket of pages – and sometimes the basket of pages is quite large and potentially diverse. Pathing tools (even the best) or a next pages report or a page detail report simply won't get the job done. Not only are they too much work, they won't de-duplicate visit and visitor numbers properly.

Most web analysis tools do have some content analysis grouping and reporting capabilities. As with visitor segmentation, however, there are often crippling limitations on the usage and implementation requirements.

Web Analytics Tool Evaluation and Selection

Here's a list of the key features to consider when it comes to grouping pages and using these content groupings in real-world analysis:

- ✓ Ability to Organize and Group Pages and Entire Sites
- ✓ Ability to use Groups of Pages
- ✓ Ability to View Content Levels Graphically

Organize and Group Pages and Entire Sites

The view we take of how content is related across a web site is fundamental to all analysis. A Functional analysis will group pages by logical type. A visitor segmentation will (typically) group pages by site interest.

What makes problems in taxonomy (the classification and grouping of pages) especially challenging is the fact that no single taxonomy is likely to support a very wide range of analytic problems. If, for example, I am interested in what visitors who trade Stocks are interested in, I might need a taxonomy that classifies pages based on tool-type (portfolio analysis, stock finder, research, etc.). But I might also want a taxonomy that classifies page content based on equity type or market (all pages about IBM, MSFT or all pages about large cap and small cap, or all pages classified by relevant exchange).

Similarly, if I'm interested in the way visitors use on-site customer support, I may want to classify all pages by Customer Support (and Support sub-functions or contact mechanisms) while leaving all other pages in one great bucket.

There is no one single correct taxonomy for a web site – there are only taxonomies appropriate for more or fewer analytic problems.

Creating Page Groups

As with most web measurement setups, page group could be created via the tag or within the software solution. The vast majority of web analytic solutions opt the lesser solution – creation in the tag. Tag-based taxonomies have a number of fundamental limitations and can – in many circumstances – make page grouping essentially impossible.

A tag-based scheme is necessarily limited to a single hierarchy. Whether it is a good hierarchy or not depends largely on the structure of your site or your ability to use a CMS to provide taxonomy. Most tag-based systems will use the directory structure of your site as the default for building a site classification. If your directory structure closely matches your business needs, then you may be satisfied with this approach. If your directory

Web Analytics Tool Evaluation and Selection

structure is flat or organized along technical principles, the hierarchy created this way is likely to be useless.

If your site is highly dynamic or uses Web 2.0, static hierarchy groupings based on the directory structure (or CMS) are especially problematic.

There really is only one logical place to construct a taxonomy – and that is in the web analytics application. It would be nice if the web analytics application could be passed a general purpose taxonomy (via tag or CMS as is now the case). However, simply providing this capability isn't enough. To be useful, it needs to be implemented in a plausible fashion. If your site has 5000 pages, classifying content page-by-page in a dialog box just isn't a viable solution. A taxonomy-builder needs to provide search, copy, paste, and powerful grouping functions to make building site-wide classifications practical.

As important, is the ability to construct multiple “point” taxonomies that can be used for specific analytic purposes. These are not site-wide hierarchies – just the gathering together in a single group some set of specific pages. This capability is best supported by drag-and-drop or even multi-selection from a page-list.

Using Page Groups

Page Groups are important for a number of different kinds of analysis – not least of which is basic reporting. You can easily get Page View counts for any group of pages simply by adding them up. But if you want to know how many visits or visitors included a page view of a page in a set, you can't just add up the numbers. If you do, you'll get vastly misleading numbers because many of the visitors and visits will have touched multiple pages in the group.

So one of the most important uses of Page Groups – something that you really can't live without – is getting de-duped visitor and visit counts.

In addition to basic reporting on Page Group usage, one of the more useful capabilities is pathing based on Group. For this capability, you really need to be able to define site-wide hierarchies. Sometimes, it's okay to see pages in-and-out of a single group. But more often you want to be able to reduce pathing complexity by treating the whole site as a series of modules. The ability to do content group pathing is a major addition to overall path usefulness.

Finally, page groups are extremely useful in building visitor segments. Your analytic solution should let you use visitor usage of a group (views, visits, total time, etc.) as one of the criteria for creating a segment. This is very

Web Analytics Tool Evaluation an Selection

important to building robust segments – since page level behaviors are often much less indicative of overall visitor interest.

Graphical View of Taxonomy

Viewing a site hierarchy is one of the things that simply works much better when done graphically. Presentation style will rarely be a decisive difference between web analytics systems – and this is probably the least important hierarchy capability.

Nevertheless, it is much easier to take in hierarchy when you can view it graphically. Ideally, a system should let you see as much of the hierarchy as you can fit on the screen, should allow you to have multiple nodes open and should allow you to drill all the way to the page level within the hierarchy display.

Summary

What really matters when you buy a web analytic solution? The answer will vary from company to company and even from user to user. There is no one “right” solution for every problem and every situation. And if you’re doing a tool evaluation, your job is likely to be complicated by the range of features and reports that vendors have incorporated into their products. Deciding which ones matter to you (or anyone) and which ones don’t is no easy task. Nor is the choice of vendor obvious. There are five or six companies now with enterprise web analytics tools that can compete very reasonably for your business.

By understanding some of the key areas of web analytics that really do differentiate tools, you’ll be in a much better position to understand what’s right for your company and make a good decision. In addition, you’ll be better positioned to understand how the tool choices you made translate into decisions and perceptions about how your implementation and your rollout need to be structured.

There are many, many different features and functions in web analytic tools. But the five discussed here (Visitor Segmentation, Dimensional Reporting, Management Reporting, Setup, and Content Classification) were chosen because they turn out to be important to most companies and because they are points of real differentiation.

You can’t expect to differentiate vendors with a simple checklist approach: pathing: yes or no. Web Analytics tools have gotten much too robust for such an approach to work. You’re better off honing in on the specialized needs of your company. Key factors like having a very large number of SKU’s with complicated roll-ups, an organizational lack of IT resources or a particular integration need are often deciding factors in choosing a vendor. It’s also worth thinking honestly about the culture and expertise within your organization. Some products are easier to implement and more suited to reliable, fast reporting. If you don’t have a lot of analysis expertise or a real analytic culture, such products might suit you much better than a bulkier, harder-to-implement but analytically more robust solution. On the other hand, if you are chomping at the bit for deep analysis, then you’ll need to really pay attention to the differences in visitor segmentation and hierarchy that are outlined here.

Classification	Sub-Classification	Feature	Importance	Appropriate For	Comment
Visitor Segmentation					
	Segment Creation	Tagless Creation	Critical	All	
		Advanced Segmentation Logic	High	Advanced Analytics	Wide range of implementation
		Universal Segmentation Variables	Medium	Advanced Analytics	
		Data Driven Segmentation	Medium	Advanced Analytics	Not Supported in common tools
		Visit and Visitor Segmentation	High	Advanced Analytics	
		Time-Based Segmentation	Very High	Advanced Analytics	Poorly supported in common tools
	Methodology	Distribution Reporting	Very High	All	Poorly supported in common tools
		Comprehensive Data	Critical	All	Even with all data you will always miss some meaningful comprehensive data
		Sample Data	High	Advanced Analytics	Acceptable for analytic solutions
		Real-Time Segment Creation	Very High	Advanced Analytics	
		Unlimited Segmentation	Very High	Advanced Analytics	You will be overwhelmed by how many segments you do receive
Dimensional Reporting					
	Cropping	Un-Cropped Paths	High	Advanced Analytics	While paths are useful they are often not what you think it is and can be very noisy in cropped paths



Web Analytics Tool Evaluation an Selection

		Un-Cropped Pages	Medium	Sites with a very high number (100s of thousands) of different pages (not page views) Sites with very large search campaigns who will rely on WA package for PPC analysis.	Significant is very lar
		Un-Cropped Search Terms	Low	Advanced Analytics	
	Cross-Tabulation	Unlimited Cross-tabulation	High	All	
	Data Access	Monthly, Weekly and Custom Uniques	Very High		
		Limited to "Cube" Variables	Medium	Advanced Analytics	Most com even take of what th for advan this can b
		Restricted SQL or Restricted Interface	Medium	Advanced Analytics	
		Queuing	Low	Advanced Analytics	An SLA c
Management Reporting					
	Dashboards	Any item in Dashboards	High	All	
		Ability to separate Dashboards into a presentation	Medium	All	
		Ability to save Dashboard reports and deliver via email	Medium	All	
		Rich trending in Dashboards	High	All	
		Fast Load Performance	High	All	
	Excel Integration				



Web Analytics Tool Evaluation an Selection

		Ability to Load every report into Excel Ability to schedule and deliver reports	Very High Medium	All All	Some Data capabilities important - but both useful in j environment
Setup	Tagging				
		Hierarchy in Software not Tag	High	All	
		Campaigns outside of tag	High	Less-sophisticated Campaigns	Also useful
		Campaign Roll-ups outside of Tag	High	High-Volume campaign environments	
		Visitor Segments in Software as well as Tag	High	All	
		e-Commerce rollups outside of Tag	Variable	Critical if you have many SKUs and rollups that aren't available when the tag fires.	More is better many companies under-utilize have.
		Custom Variables	Medium	Advanced Analytics	
Page Hierarchy	Creation				
		Organize and Group Pages into a single unit	High	All	Reporting as well as
		Organize entire site into logical structures	Medium	Advanced Analytics	Poorly supported common



Web Analytics Tool Evaluation an Selection

Use	De-Duped Visitor Counts	Very High	Critical for many reports
	Pathing	Medium	Advanced Analytics
Usability	Building Segments	Low	Advanced Analytics
	Graphical View	Medium	All