What can web analytics do for technical communications?

Web analytics is often used in Internet marketing to understand the success of advertising or determine why customers aren't completing the purchase process on a website. Although the technique is less often used to understand the success of online documentation, I believe that such data could become a powerful tool in developing and maintaining websites containing user assistance and support information.

At Red Gate we've been exploring the value of web analytics in understanding how users interact with our support site. In this article, I'm going to use data from that site to illustrate some potential benefits, and limitations, of using web analytics.

Web analytics basics

'Web analytics' is the use of data such as the number of people viewing pages on a website, how they get tto those pages and what they do next, with a view to improving the website in some way. The main data includes:

Page views (sometimes called 'hits')
The total number of times the page was viewed in a specified period.

Unique page views
The number of visits during which the page was viewed (for example, someone might view the same page several times during a visit to the website; this is recorded as one unique page view).

Exit rate
The percentage of site exits that occurred from this page (that is, visitors went to a different website or closed their web browser).

Time on page
The length of time that a visitor stays on a page during a visit, or the average length of time that a number of visitors spend on a page.

About web analytics tools

The data in my examples has come from Google Analytics (Figure 1) but similar data is available from other web analytics tools. Getting up and running can be as simple as pasting a script into your web page templates (this is all we needed to do with Google Analytics), although server-based tools can be more complex. Wikipedia has a good overview of the available tools and how they work.
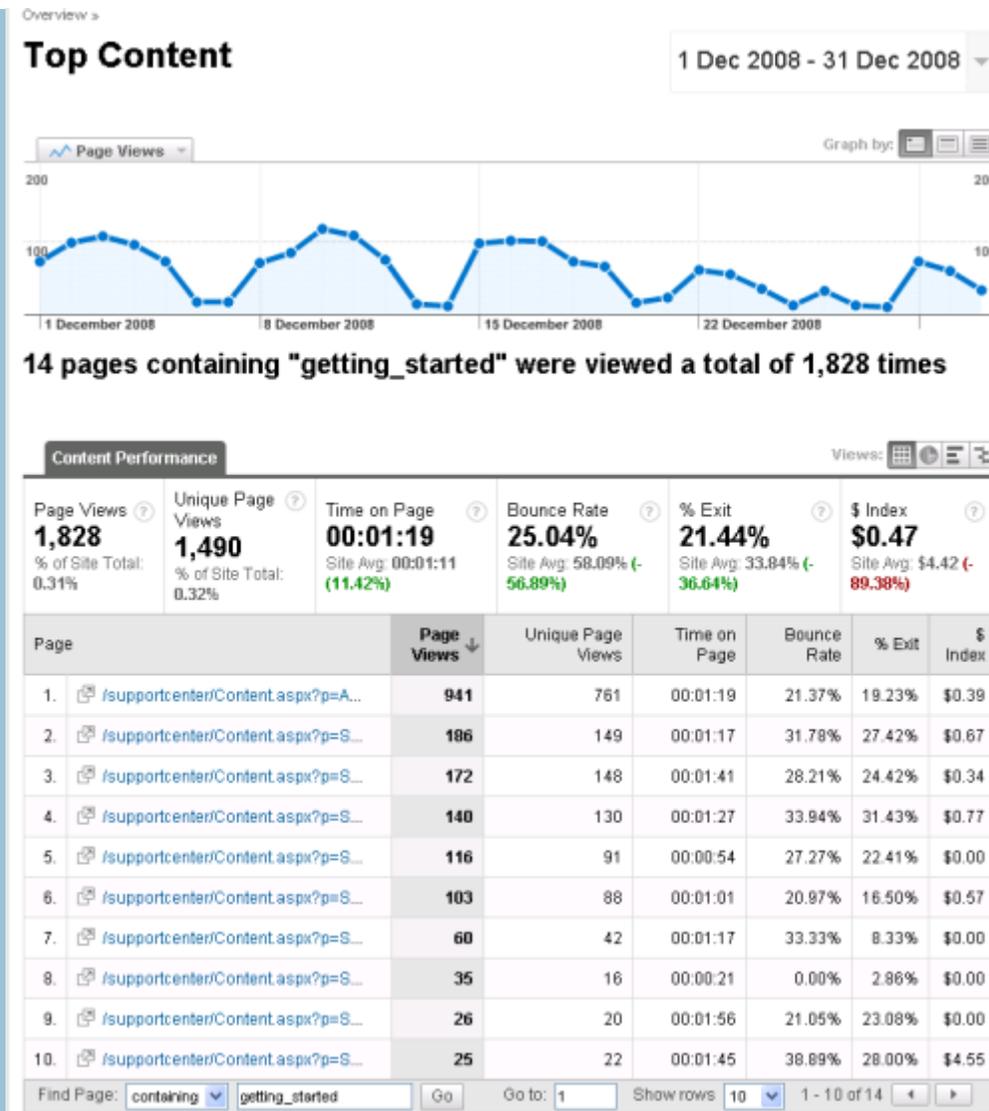
Overview »

Top Content

1 Dec 2008 - 31 Dec 2008

Page Views

Graph by:

| 200 | | | | 200 |

1 December 2008   8 December 2008   15 December 2008   22 December 2008

14 pages containing "getting_started" were viewed a total of 1,828 times

Content Performance

Views:

Page Views ?	Unique Page Views ?	Time on Page ?	Bounce Rate ?	% Exit ?	$ Index ?
1,828	**1,490**	**00:01:19**	**25.04%**	**21.44%**	**$0.47**
% of Site Total: 0.31%	% of Site Total: 0.32%	Site Avg: 00:01:11 (11.42%)	Site Avg: 58.09% (-56.89%)	Site Avg: 33.84% (-36.64%)	Site Avg: $4.42 (-89.38%)

Page	Page Views ↓	Unique Page Views	Time on Page	Bounce Rate	% Exit	$ Index
1. /supportcenter/Content.aspx?p=A...	941	761	00:01:19	21.37%	19.23%	$0.39
2. /supportcenter/Content.aspx?p=S...	186	149	00:01:17	31.78%	27.42%	$0.67
3. /supportcenter/Content.aspx?p=S...	172	148	00:01:41	28.21%	24.42%	$0.34
4. /supportcenter/Content.aspx?p=S...	140	130	00:01:27	33.94%	31.43%	$0.77
5. /supportcenter/Content.aspx?p=S...	116	91	00:00:54	27.27%	22.41%	$0.00
6. /supportcenter/Content.aspx?p=S...	103	88	00:01:01	20.97%	16.50%	$0.57
7. /supportcenter/Content.aspx?p=S...	60	42	00:01:17	33.33%	8.33%	$0.00
8. /supportcenter/Content.aspx?p=S...	35	16	00:00:21	0.00%	2.86%	$0.00
9. /supportcenter/Content.aspx?p=S...	26	20	00:01:56	21.05%	23.08%	$0.00
10. /supportcenter/Content.aspx?p=S...	25	22	00:01:45	38.89%	28.00%	$4.55

Find Page: containing | getting_started | Go   Go to: 1   Show rows 10   1 - 10 of 14

Figure 1. Screenshot of Red Gate support site data in Google Analytics

Web analytics at Red Gate

The Red Gate support site is a help and support portal that comprises content such as product help, a knowledge base, marketing videos and public forums. When we moved to this primarily web-based approach, we decided to take advantage of the Google Analytics tool that was already in use within the business.

Our main reason for enabling web analytics in our online help was to find out whether anyone was actually reading it. We quickly confirmed that our help topics were viewed around 14,000 times a month in total. This was interesting, but we felt that it must be possible to get more value from the newly available web analytics data.

We are still in the early stages of exploring the possibilities of web analytics, but we use the data regularly to help us understand and improve our site. In the following sections, I describe the main areas of information that web analytics data enables us to access.

Turning data into understanding

General information such as terminology and levels of usage is useful, but clearly there's more to a web analytics strategy than just looking at the raw data.

For example, it was interesting to discover that the Red Gate help pages are viewed 14,000 times per month, but we had no way of knowing whether that was a good number or not: do we actually want users to view our help more or less frequently?

Burby and Atchison suggest a promising strategy for getting from raw web analytics data to implementing improvements to a website, but — as with using data in any technical communications strategy — the most difficult step is identifying how successful user behaviour will be represented by the data.
Getting from data to understanding often requires additional information from users.The Conversion Rate Experts website has a useful list of 14 free tools that can enable you to make contact with users.
Understanding users' language
By examining the terms users enter when they perform a search, we get a useful insight into users' world-view, particularly their terminology.

Table 1 shows data for the top ten searches on the Red Gate support site over three months at the end of 2008. We can use this report to ensure that topics use the same terminology as users, by altering titles, tagging topics with additional keywords, or defining synonyms within the search engine.

The data for the top 10 searches is in table 1.

Search term (product page searched from)	Page views	Unique page views	Exit rate	Most common next page, other than site exit (% of visitors to the search page who looked at this page next)
Snapshot (SQL Compare)	54	44	17%	Forum post: Free schema snapshot utility: SQL Snapper (37%)
Command line (SQL Data Compare)	44	26	7%	Knowledge base: SQL Data Compare command-line XML argument file examples (37%)
Asp.net (ANTS Profiler)	30	21	10%	Forum post: How to profile ASP.Net memory leak (15%)
Command line (SQL Compare)	30	21	10%	Forum post: help file for command-line execution (22%)
Thread blocked (ANTS Profiler)	34	20	9%	Forum post: Thread blocked (38%)
Source code (ANTS Profiler)	30	19	7%	Article: Why can't ANTS Profiler find my source code? (64%)

Download (no product selected)	23	19	17%	Red Gate website home page (22%)
Inconsistent API use: generate script rewrite encountered a create action. (no product selected)	18	18	94%	Forum home page (6%)
18210 (SQL Backup)	17	17	53%	Knowledge base: VDI error 1010: Failed to get configuration from server (21%)
Restore (SQL Backup)	21	16	14%	Product walkthrough (8%)

Table 1: Top 10 unique searches on support site October – December 2008

Understanding users' experiences with site navigation
The main purpose of navigation pages is to take visitors to the information they need as quickly as possible. On the Red Gate support site, this includes search, index or 'home' pages, and getting started or 'landing' pages. These pages generally don't have a lot of information on them; instead they consist mainly of a number of hyperlinks.
The usage data for these pages enables us to understand how well these pages support users in trying to navigate around the site.

I'm going to take search pages as an example here, because Wiggins and Rosenfeldbelieve this is an area where web analytics can be particularly useful, but similar principles apply to other navigation pages. Figure 2 shows an example of one of our search pages.

Figure 2. Example search results page

The search experience we want visitors to our site to have is that they perform a search, click on a result, and are so happy with what they see there that they leave the site. They don't come back to the list of search results to see if there's anything better (they don't need to: they've already seen something that answers their question).

If users behave in this way, the data should show the following:

- The exit rate from search pages is low (because visitors don't leave the site on a search page).
- The next page viewed should be a relevant content page (if visitors follow links to other navigation pages or pages for the wrong product, the search probably hasn't been successful; we can also apply a more subjective view of relevance by looking at the content on the page).

Table 1 shows this data for the top ten searches. There isn't space here to look at all the searches, but I'll examine two contrasting examples.

Search example A
'Inconsistent API use: generate script rewrite encountered a create action'
The exit rate on the page is 94%, so most people searching on this term leave the site rather than following any links. The few people who do follow a link, go on to look at the forum home page.

This search doesn't look successful by either of our criteria, then: it has a high exit rate and the next page viewed is not a relevant content page (it's another navigation page). The reason for this becomes obvious when we perform the search on the site for ourselves. The search returns with:

'There were 0 results.'

Are there results we'd expect to see? Is this a bug or a misunderstanding about how to enter search terms?

We would need to look beyond the web analytics data to find this out, but what is very clear here is that web analytics has indicated a genuine problem: people are searching on a term and we don't give them any information about it.

Search example B
'Command line' (from SQL Data Compare product page)
The exit rate for this search page is very low, at 7%, so it looks like visitors are finding useful results, rather than leaving the site at this page. 37% of views of this search page go on to look at a knowledge base article next: 'SQL Data Compare command line XML argument file examples'. This looks like a promising result, so it is likely that visitors are successfully finding the information they need.

Once again, though, we'd need to look beyond the web analytics data to confirm that users are finding out all they need to know about the SQL Data Compare 'command line'. For example, we'd compare this term with our data on support calls (assuming that if people don't find what they need on our support site, they'll call our support team instead). In this period there were three SQL Data Compare support calls in which users were asking something about the command line. This is a low number, which confirms that this search has been successful.

Identifying pages that no one reads

Pages such as help topics and knowledge base articles are designed to be read: the effort that technical communicators spend on carefully crafting topics only has value to the business if users read the topic. Web analytics data can give us an understanding of page usage, although this analysis requires a deeper understanding of the purpose of the content than in the previous examples.

We'll look at a particular type of help topic: 'worked examples' (as shown in Figure 3). We can begin our search for pages that no one reads by looking at the number of page views: if there are no views of a page, no one is reading it. As you can see from Table 2, there are no worked example topics with zero page views, which is good news.

If we were to find pages with zero views, we would use additional sources to determine why no one is viewing the page (perhaps users don't need the information on this page or perhaps they need it but are having trouble finding the page).

However, the number of page views only tells us that the page is being viewed, not that it is being read (visitors might view the page and leave, without reading much of the content). Of course, we can't use web analytics data to determine whether visitors are reading entire pages but, as worked example topics are designed to be followed in detail, we do expect visitors to stay on the page for a significant amount of time. A reasonable guess suggests that an average of less than 30 seconds spent on the page will be an indication that visitors are not using the page in the way we had hoped.
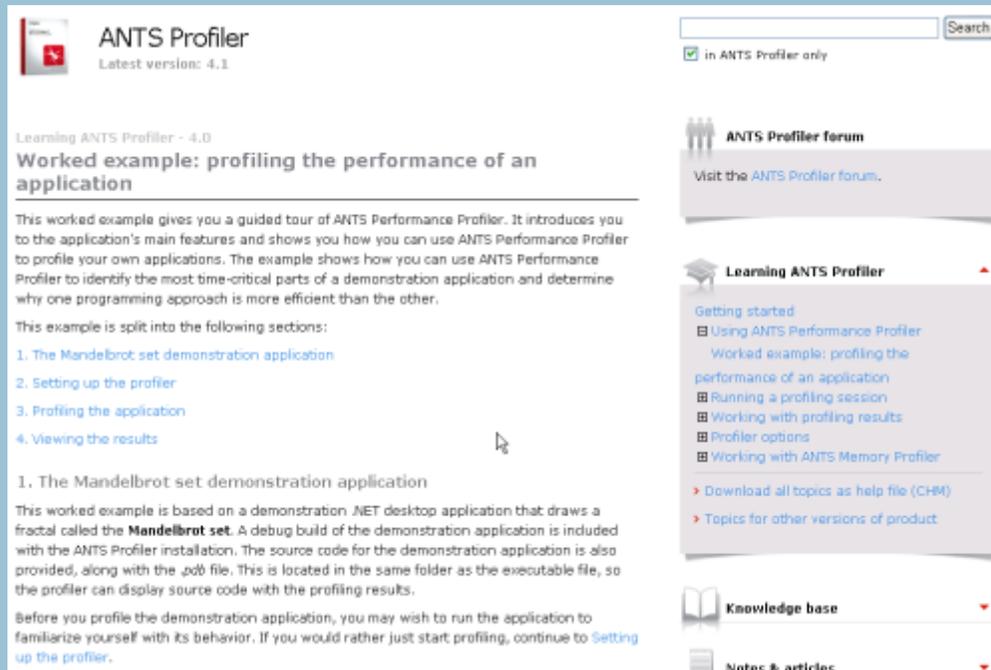

Figure 3. Example of a 'worked example' page

The data in Table 2 shows that the lowest average time spent on a worked example page was two minutes and three seconds. This is well over our 30-second minimum, which we can interpret as an indication that these pages are generally being read.

Product	Worked example topic	Page views	Time on page (mins:secs)
ANTS Profiler	Performance profiling	608	03:01
ANTS Profiler	Memory profiling	452	06:04
SQL Compare	General example	117	03:19
SQL Compare	Scripts folder	45	04:45
SQL Data Compare	Synchronising databases	165	03:21
SQL Data Compare	Restoring from backup	16	03:25
SQL Dependency Tracker	General example	87	05:34
SQL Doc	General example	75	04:17

SQL Multi Script	General example	33	03:24
SQL Packager	Packaging as .exe	35	02:09
SQL Packager	Packaging as C project	30	02:03
SQL Prompt	Examples	137	04:15

Table 2: Usage of worked example topics, October – December 2008

If there had been a page with low average time spent on it, it could have been an indication that the topic isn't needed or that the format of the page or the quality of the content were deterring visitors from reading the page. In both of these cases, web analytics wouldn't be able to offer much insight: instead we'd have to look at other feedback from users, such as support data or usability tests.

This lack of qualitative insight is the weak point of web analytics data. This example reveals other limitations:

- This analysis shows us that the content we're producing is being used. It doesn't tell us what content is missing.
- The data doesn't tell us whether users were successful in finding the information they needed on the page. To find that out, we couldlook at what page they viewed next, or — more commonly — we'd check with our support and sales team, to ensure that users aren't having trouble learning to use these tools.

Planning documentation projects

Documentation projects are often under-resourced, and with little access to user feedback it can bedifficult to identify priorities. Web analytics data can help here by identifying high- and low-use content, as well as giving insight into users' issues.

Pages with zero views are the first target. If the reason the page is not viewed is that no one needs the information it contains, topics can probably be removed from future iterations of the documentation so that no effort has to go into maintaining content that is not used or needed.

There are no examples of pages with zero views in the data that I've included in this article, but Table 2 shows an example of a page with low page views compared to another page for the same product, SQL Data Compare:

- 'Restoring from backup' has 16 views
- 'Synchronising databases' has 165 views.

The page with 16 views is likely to be a lower priority for future work than the other (additional factors may also affect priority, of course

In other analyses at Red Gate, we've also used web analytics data to understand whether there's a need to continue to make the help for older versions of products available. On the Red Gate support site these pages are clearly marked, and it's very likely that visitors arriving at them are doing so deliberately. The numbers of page views on these pages indicate the level of interest in these older software versions.

Search terms also offer interesting potential.

These terms can give an insight into software usage: for example, in Table 1 we can see that 'command line' is a common search term for the SQL Data Compare product. This tells us that we have customers interested in using the command line for this product, so we need to ensure that attention is given to this area in the documentation.

Similarly, we sometimes see error strings occurring frequently as search terms. This indicates that users are coming across these errors, and that the error message within the software probably needs attention, as it has not been sufficient to help the user recover from the error.

Conclusion

The examples in this article demonstrate that web analytics can help to find useful information about a support site. The examples also show that arriving at conclusions is a complex process that often requires the addition of data from outside the web analytics. Some investigation targets, such as determining how useful a page is to its readers, may not be worth this effort, whereas others, such as identifying common search terms that don't return useful results, are likely to be more fruitful.

Web analytics offers a unique insight into users' actual behaviour — rather than their reported behaviour — and as such can be very valuable. The data is divorced from qualitative understanding of users' experiences, though, and this means that web analytics is more suited to identifying likely problem areas than to understanding in depth the nature of the problems. Even in these circumstances, the data is still valuable for narrowing down areas that need attention or improvement.

Technical documentation is very different from the marketing pages with which web analytics is more traditionally used but, provided you have a good understanding of what you are measuring, web analytics can help you discover really valuable information about usage.

Source: http://communicationcloud.wordpress.com/2009/11/19/what-can-web-analytics-do-for-technical-communications/