

# Using Website Statistics To Determine Newsletter Readership.

**Date:** 3 December 2006

**Requested by:** Jeff Turner (on behalf of the PMI Portland Board of Directors)

**Generated by:** Todd C. Williams

I have been asked to generate statistics on viewing of the newsletter online. But, before I can present those statistics I need to provide some information on website statistics and how they are used.

## Definitions

**Crawlers:** Crawlers (or spiders) are programs (aka robots or bots) that go to a site and read all the pages on the site. They are all designed to look for different things. Some look for items for sale, some are scientific, some grab email addresses for spamming, some look for content. Google has a bot (googlebot) that loves pdfs. To prove that to yourself go to Google and type in "pmi newsletter". If you get the same result I did you will get the VERY first listing as our site. Do you have any clue how much money companies pay for that position. Literally 10's and even 100's of THOUSANDS of dollars. That is because pdfs are considered content with value. No one pdfs something unless they think it is of value. You do not get the same at Yahoo since their crawler is designed differently.

**Entry Page:** These are pages that are the pointed at which a user enters the site. Typing or click a URL that points to <http://pmi-portland.org/events/meetings/roundtables.htm> will generate an entry hit for that page. These numbers are inferred based on the time a user was on the site. Therefore not precise.

**Hits:** Hits are generally raw numbers from the server log file that show requests for an item on the site. This may be only pages, but can also be images, style sheets etc.

**Visits:** Visits are usually calculated numbers based on assumptions of who a user is and based on their IP Address and time between entries in the server log. The biggest problem is that most corporate users show up and the same person. In other words all Nike users show up as one user, hence one visit.

## Tools

I have activated both Awstats<sup>1</sup> and Webalizer<sup>2</sup> on our site to try to get more meaningful statistics. Unfortunately, this is like having two clocks both telling a different time—you no longer know the time. This is because the developer designed the code different to infer results to produce what he or she thought was more "data" data.

For instance Webalizer has the follow rule for Visit<sup>3</sup>:

*"[Visits] occur when some remote site makes a request for a page on your server for the first time. As long as the same site keeps making requests within a given timeout period, they will all be considered part of the same*

---

<sup>1</sup> <http://awstats.sourceforge.net>

<sup>2</sup> <http://www.mrunix.net/webalizer>

<sup>3</sup> [http://www.mrunix.net/webalizer/webalizer\\_help.html](http://www.mrunix.net/webalizer/webalizer_help.html)

**Visit.** *If the site makes a request to your server, and the length of time since the last request is greater than the specified timeout period (default is 30 minutes), a new **Visit** is started and counted, and the sequence repeats. Since only pages will trigger a visit, remotes sites that link to graphic and other non- page URLs will not be counted in the visit totals, reducing the number of false visits."*

and Page as

*"those URLs that would be considered the actual page being requested, and not all of the individual items that make it up (such as graphics and audio clips). Some people call this metric page views or page impressions, and defaults to any URL that has an extension of **.htm**, **.html** or **.cgi**."*

Hence it may not count pdf visits in its "visits". But it does count some since there is a count (see Table 1).

Awstats defines Visit<sup>4</sup> as:

*"Number of visits made by all visitors. Think "session" here, say a unique IP accesses a page, and then requests three others without an hour between any of the requests, all of the "pages" are included in the visit, therefore you should expect multiple pages per visit and multiple visits per unique visitor (assuming that some of the unique IPs are logged with more than an hour between requests)"*

and Page as

*"The number of "pages" logged. Only files that don't match an entry in the NotPageList config parameter (and match an entry of OnlyFiles config parameter if used) are counted as "Pages". Usually pages are reserved for HTML files or CGI files, not images nor other files requested as a result of loading a "Page" (like js,css... files)."*

## **Analysis**

To analyze the statistics you need to look at a variety of reports and make some inferences. Looking at volumes by day and relating that to external events (mailings, maintenance, geo-political events, etc.), entry pages, page hits and number of corporate users will point to some trends. It should go without saying that the aforementioned definitions are also critical.

### **Page Hits**

The problem is that looking at the page hits has a problem. First, it is key to understand that Jani and I manipulate the URLs in the SPORG emails so that we can try to track the emails usage. In November we switched the URL in the newsletter to go to the newsletter page not the newsletter itself. Therefore comparing this to other months would be mostly meaningless. Table 1 shows the results.

---

<sup>4</sup> [http://awstats.sourceforge.net/docs/awstats\\_glossary.html](http://awstats.sourceforge.net/docs/awstats_glossary.html)

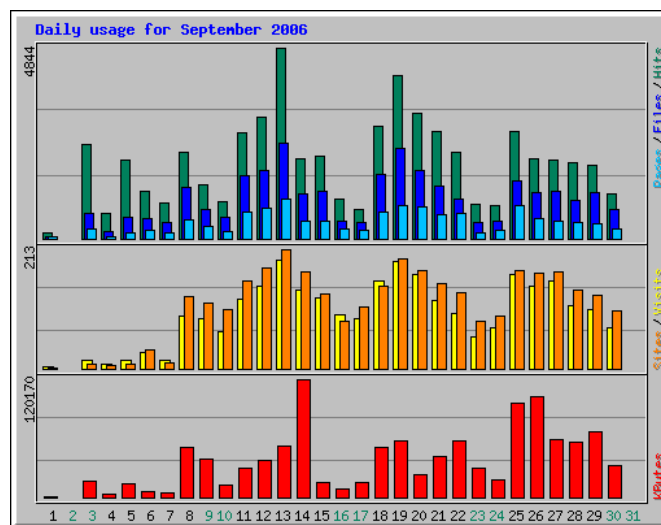
URL	Description	Views	
		Awstats	Webalizer
/Home.php	Home page	1483	1496
/Calendar/month.php	Event Calendar	1030	1233
/events/meetings/chpt_meeting.php	Chapter meeting page	683	691
/news/newsletter.htm	List of all newsletters	410	562
/resources/job_postings.htm	Job Posting description page	333	359
/events/meetings/roundtables.htm	Roundtable description page	237	276
/DocumentLibrary/Newsletter/2006-11.pdf	The November Newsletter	133	1002

**Table 1 - Top Page Views (entire month)**

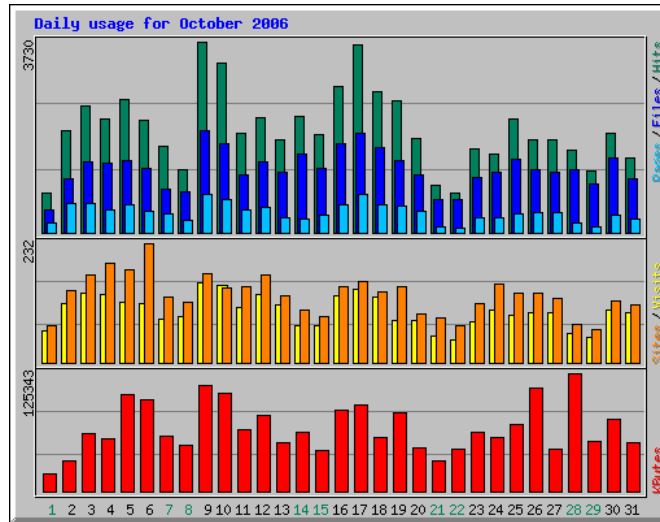
The disparity in the number of views of the newsletter between analysis tools is very significant and indicates these values cannot be used to determine the actual viewing number for the November newsletter. We either have nobody (figuratively) reading the newsletter or we have nearly every member. Neither number appears a rational answer. So we need to look at other data.

### **Correlating Events**

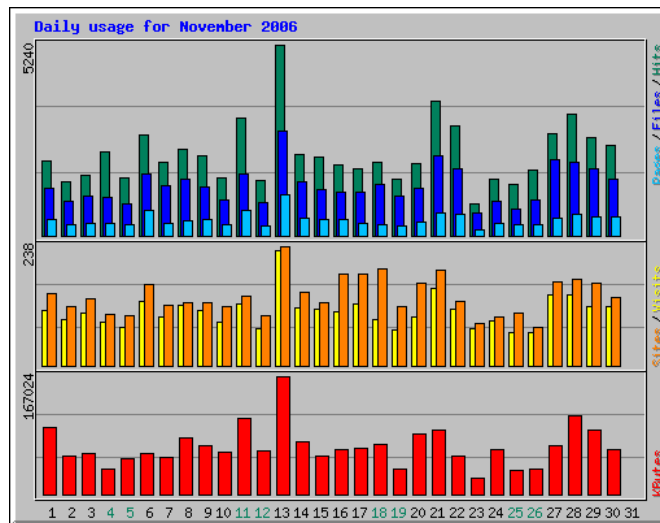
Statistics are available for peak usage of the website. These are accurate when discussing hits and based on the assumptions the files and pages. Again the data is relative between tools. Below I have attached three months of graphics for site hits. These happen to be from Webalizer but it really does not mater since we are looking for peaks.



**Figure 1 – September 2006 - Webalizer**



**Figure 2 - October 2006 - Webalizer**



**Figure 3 - November 2006 – Webalizer**

The key point is there are high hit rates (green bars on top graph) on:

- September 13 and again on September 19
- October 9 and again on October 17 (not quite as pronounced).
- November 13 and small peak on November 21.

Looking at our calendar you will find that November 13 and October 9 are the Monday nine days prior to the Chapter Meeting. September 19, October 17 and November 21 are the days of Chapter Meetings. September 13 is the day we sent out the corrected URL in the meeting announcement.

Looking in my deleted items folder I find that the newsletter was distributed two Fridays prior to the Chapter Meeting and the “Don’t forget to register...” reminder is sent the subsequent Monday. So the hits on that Monday could be from people finally seeing both emails.

### **Entry Pages**

To try to determine which is getting the most attention we need to look at “entry pages”. Table 2 shows the entry points. The key thing to understand is that the

emails that are sent have specific pages they point to in accessing our site. The Newsletter email points to <http://pmi-portland.org/news/newsletter.htm> and the "Don't forget to register..." email points to [http://www.pmi-portland.org/events/meetings/chpt\\_meeting.php](http://www.pmi-portland.org/events/meetings/chpt_meeting.php). Since few people in there right mind (Jani and I excluded) would never type these URLs we can safely assume that the newsletter email was a significant portion of the count. (Note: the skeptical observer would note that the Google reference above points to the newsletter page and might bias the results. In November <http://www.google.com/search> did hit our site 435 times. I have no idea if those people clicked through to our site and what page they went to.)

URL	Description	Awstats		Webalizer	
		Viewed	Entry	Hits	Visits
/	Home Page	1428	1015	2549	1525
/news/newsletter.htm	List of all newsletters	410	210	562	239
/events/meetings/roundtables.htm	Roundtable description page	237	118	276	150
/events/meetings/chpt_meeting.php	Chapter meeting page	683	161	N/A	N/A
/resources/job_postings.htm	Job Posting description page	333	49	359	71
/contact.htm	Contact Page	227	45	279	89

**Table 2 - Top Enter Pages (entire month)**

This data shows that somewhere in the low 200's of users did click through to the Newsletter page, slightly more than did the Chapter Meeting page which has multiple emails sent. This statistic also shows that between 400 and 600 people "viewed" or "hit" the newsletter. Other URLs are presented to give a feeling for the size of that number.

### **Corporations**

Lastly there are big users to throw off the statistics. Everyone in large corporations general show up as one user. As Table 3 shows all Nike employee show up as one user and hit our site 766 times, we have no idea how many people that really represents. The point being that these visits will drive the newsletter access numbers lower.

No, I am not a huge corporation but I do visit the site a lot and could bias the statistics, but I assure you by the time the newsletter is shipped I have seen it enough to never revisit it. (Side note: obviously missing from this list is Con-way, Standard and Freightliner, probably more.)

URL	Hits	User/Company
c-67-171-243-27.hsd1.wa.comcast.net	4396	Todd Williams
mail2.zoey.a-dec.com	856	A-DEC
mail.cadencemc.com	847	Cadence
barrierc241.nike.com	766	Nike
ptldnet.ci.portland.or.us	634	City of Portland
portal.kindercare.com	601	KLC
ptr-198-107-240-51.pgn.com	332	PGN

**Table 3 - Large User Statistics**

## Conclusion

People are reacting to our meeting announcements and the newsletter emails and they do not appear to be being ignored. Only a guess can be placed on the actual readership and that might be in the five hundred range or higher due to the number of users that have corporate addresses (A-DEC, Nike, Cadence, etc.). You can draw your own conclusion.

It is too early to determine if readership is increasing or decreasing we will need a few more months for that. This is the best way to use these statistics since, all other things being equal, trends in numbers will tell us trends in usage.

We have no concept about what the paper readership was and I do not have statistics from the previous site on hits to the Chapter Page.

Hence it is my opinion that no conclusion can be drawn from this data on the effect of moving to an electronic newsletter distribution.