an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Working Title:

Analysis of Selected Web Site Statistics for OceanWorld (1998-2001)

Author:

Jon Reisch

Date:

01/18/2002
02/25/2002 (Revised)

Abstract:

Research into server log statistics for a given website can provide valuable information and insight into the site's relative strengths and weaknesses. In addition, these same log files contain detailed information about the site's user population. By analyzing this information, site creators can better focus their efforts in providing the various kinds of content which their specific user population demands.

an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Section 1: Technical Background

Our site, http://oceanworld.tamu.edu (henceforth referred to as OceanWorld), is housed on the Texas A&M Department of Oceanography server in College Station, TX. The server is a Compaq Tru64 running the Digital UNIX 5.0 operating system. Apache Web Server software, version 1.2.6, provides the public with a means to access OceanWorld, and records the log files which we will be discussing. All site statistics mentioned in the following pages were generated by Analog 5.1, a freeware log file analysis tool (http://www.analog.cx). Statistical charts and graphs were generated by Analog and ReportMagic for Analog v2.12, a freeware helper application providing graphic output of statistic reports (http://www.reportmagic.org).

Section 2: General Statistics

Table 2-1: General Statistics for OceanWorld

1.	Host name	**OceanWorld**
2.	Host URL	**http://oceanworld.tamu.edu**
3.	Time of first request	**Aug 13, 1998 10:49**
4.	Time of last request	**Dec 31, 2001 23:57**
5.	Successful server requests	**2,161,089 Requests**
6.	Distinct files requested	**3,913 Files**
7.	Successful requests for pages	**385,295 Requests**
8.	Distinct hosts served	**66,241 Hosts**
9.	Total data transferred	**17.90 GBytes**
10.	Total lines	**35 Lines**

an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Section 2: General Statistics (cont.)

As a means of introduction, we will first look at the general statistics for OceanWorld from its first recorded visit on Aug.13, 1998 to the last recorded visit on Dec. 31, 2001. This information can be easily referred to in Table 2-1. To understand fully what these numbers represent it is necessary to distinguish between several key terms. First, a request is any piece of data which the user's browser asks the server to send. This could take the form of visible elements such as web pages and images, or data which is often transferred unbeknownst to the user such as external JavaScript files or Cascading Style Sheets. Similar to a request is a page request or page hit. This kind of request is tallied specifically when the HTML documents themselves (those files with .htm and .html extensions) are demanded by the user's browser. Why then is there such a large difference in the total number of requests versus page requests? Each page on OceanWorld contains an average of 5-10 images, in addition to several "hidden" documents necessary for formatting the content, etc. Thus each page request generates further requests for every image and helper document used in presenting the page. One last bit of terminology which may be unfamiliar is "distinct host." A distinct host refers to a unique visitor to the site, which has not been logged before. Formerly, web statistics reports would generally include a "total number of hits." This highly inflated number did not accurately portray the number of unique users who had visited the site, since it would count the same visitor on as many occasions as he or she accessed the site. With the "distinct host" count, visitors IP addresses are logged, and counted only the first time he or she visits the site.

From the data provided, it is clear that OceanWorld is an active site. In slightly over three and a half years of operation, OceanWorld has grown from little more than a handful of pages to a significant site serving almost 4,000 unique files to nearly 70,000 visitors. In total, 17.9 gigabytes of information have been transferred from over 2 million separate requests for pages, images, and other data related to the site.

As a final note, regarding the accuracy of the log data, only 35 lines of log entries, roughly 4KB of data, were unreadable by our log analysis software. With the total log file size at 250MB, the amount of data contained in these entries is insignificant, as it represents only 1x10-5 of the total data available.

an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Section 3: Year by Year Analysis

Figure 3-1: Yearly Report of Total Requests

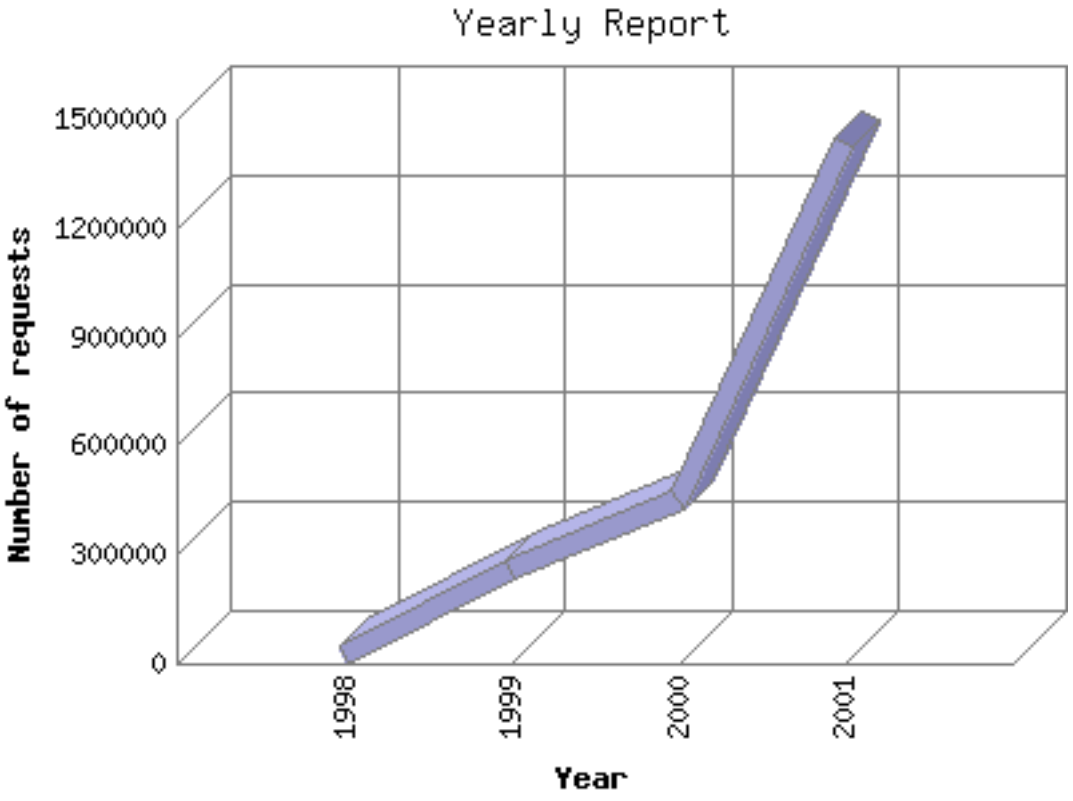


Table 3-2

	Year	Number of requests	Number of page requests
1.	1998	24,960	6,495
2.	1999	254,856	48,475
3.	2000	448,816	79,574
4.	2001	1,432,457	250,751

Most active year: 2001. 250,751 pages sent. **Yearly average:** 96,323 pages

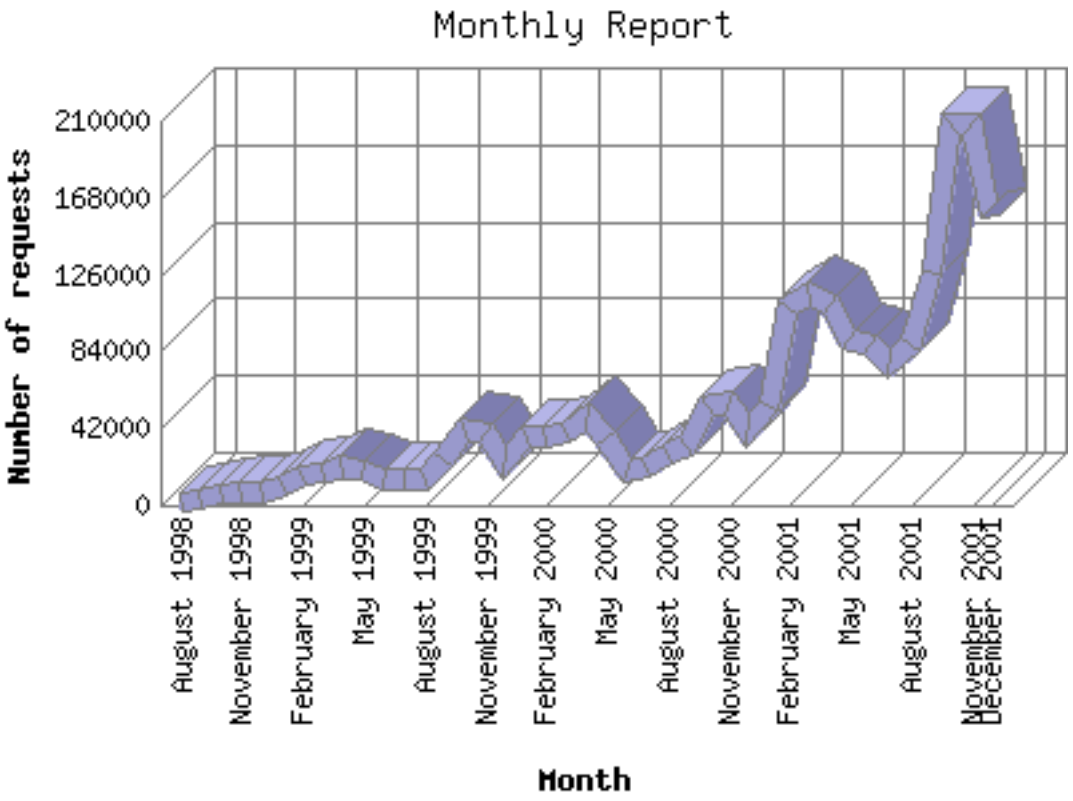
Section 3: Year by Year Analysis (cont.)

The clearest indication of the growth of the site in the last three years can be found in Figure 3-1 and correspondingly in Table 3-2. The figure indicates the total number of requests, while the table further provides the yearly page requests. OceanWorld's yearly growth rate is clearly astonishing. In the last year alone, site traffic increased 300% from the previous year of 2000. Several major site changes can help account for a least part of the increase. By the summer of 2000, the graphic layout of the site was completely redone. The newer, more user-friendly interface and professional look no doubt have enticed a large number of visitors. In the fall of 2001, Dr. Stewart's undergraduate and graduate oceanography class sites were added as subdirectories to OceanWorld. The response to these smaller sites has been tremendous as will be discussed later. In addition to site changes, Dr. Stewart, Dr. James, and Dr. Margaret Hammer have have promoted the site at several state and national gatherings of educators, oceanographers, and other scientists. An increase in this type of promotion over the last two years could also have contributed to the remarkable increase in traffic. However, none of these suggestions, even taken together, seem to account for such numbers. One assumption is that OceanWorld has established itself as a premiere, reliable source for information regarding oceanographic topics. Word of mouth promotion by satisfied users is the hallmark of a successful, memorable site. Finally, the yearly average number of page requests provides another indication of the success of 2001. With the average at almost 100,000 page requests a year, the number in 2001 exceeded this figure by two and a half times.

an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Section 3: Monthly Analysis

Figure 3-1: Yearly Report of Total Requests



With the fluctuation of site traffic plotted on a monthly basis, we can understand when the busiest times of year for the site are, as well as begin to infer some information about the user population. As an educational site, our primary stated target audience consists of students and teachers from kindergarten to high school grade levels. The selection of this broad ranging group influences the choices made in both the graphic design of the site and the reading and language level of the content. Examining the data in the monthly report, we have out first indication that our site is primarily viewed by the students we intended it to be. As reflected in the graph, there is a major drop in site activity during the summer months as well as a slight dip toward the end of November through early January. This is consistent with the pattern of grade school and high school summer vacation and Christmas break. While traffic decreases significantly during both of these periods, it does not all together halt. Last year for example, traffic during the summer exceed that of all the other years' summer figures combined. One may assume then that the site is attracting an increasingly non-grade school/non-student population as well. With the addition of Dr. Stewart's online textbook, audio lectures, and course notes last fall, as well as the completion of many new teacher resources in the spring of 2001, OceanWorld offers a diverse selection of material for a more diverse user population than ever before.

Section 5: Directory and Request Analysis

Figure 5-1 OceanWorld Directories as a percentage of Total Requests

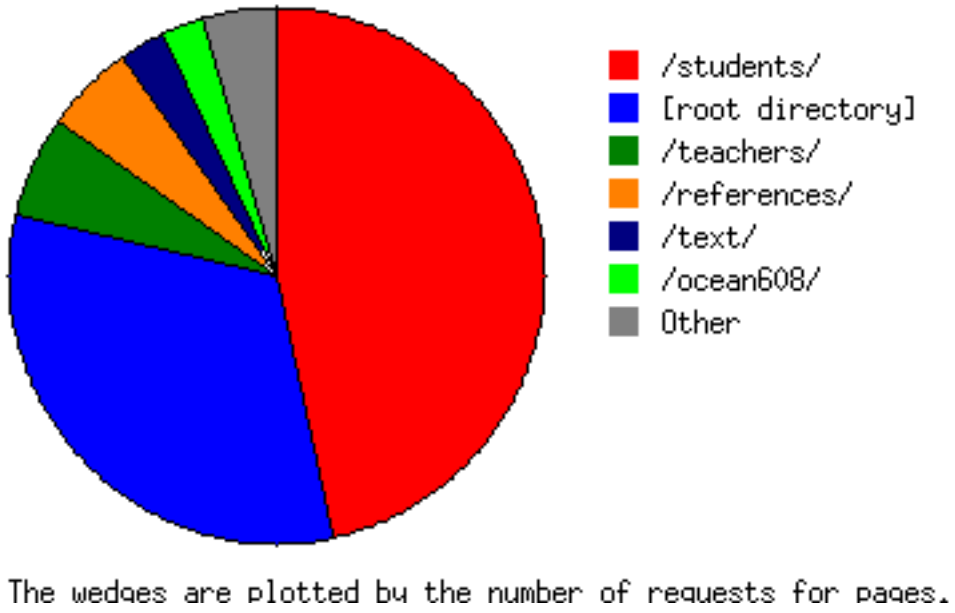


Table 5-2: OceanWorld Directories ranked by Number of Page Requests

	Directory Name	Number of Requests	Number of Page Requests	Percentage of Bytes
1.	/students/	476,257	179,468	26.21%
2.	[root directory]	635,865	123,171	20.44%
3.	/teachers/	60,038	23,814	5.7%
4.	/references/	40,183	21,013	2.50%
5.	/text/	11,263	11,206	0.27%
6.	/ocean608/	73,738	9,722	22.33%
7.	/ocean410/	15,175	3,173	0.51%
8.	/ocean401/	12,593	2,426	0.42%
9.	/education/	9,985	2,388	0.28%
10.	/realdatalinks/	1,794	1,350	0.8%

Section 5: Directory and Request Analysis

Table 5-3: Top 10 OceanWorld Directories with subdirectory listings ranked by Page Requests

Directory Name	% of requests	# of requests	% of page requests	# of page requests	% of the bytes
1 /students/	22.30%	476,257	46.58%	179,468	26.21%
/students/iceberg/	4.70%	87,775	9.49%	36,532	4.64%
/students/coral/	3.90%	84,295	7.70%	27,202	3.60%
/students/currents/	3.10%	66,826	6.84%	26,357	6.63%
/students/fisheries/	1.61%	34,860	4.27%	16,431	1.30%
/students/waves/	1.94%	42,039	4.17%	16,050	2.42%
/students/iceage/	1.36%	29,265	2.10%	11,527	1.58%
/students/satellites/	1.70%	36,869	2.98%	11,460	2.78%
/students/weather/	1.30%	22,292	2.38%	9,160	0.93%
/students/elnino/	0.77%	16,584	1.91%	7,389	0.62%
/students/jason1/	0.54%	11,672	1.48%	5,690	0.53%
/students/realdatalinks/	0.23%	5,105	0.51%	1,986	0.17%
2 [root directory]	29.42%	635,865	31.97%	123,171	20.44%
3 /teachers/	2.78%	60,038	6.19%	23,814	5.70%
/teachers/f_techlab_sta/	0.90%	1,773	0.40%	1,576	0.30%
/teachers/poo_techlab_sta/	0.80%	1,629	0.37%	1,391	0.80%
4 /references/	1.86%	40,183	5.46%	21,013	2.50%
/references/technology/	0.11%	2,510	0.42%	1,617	0.11%
/references/topex_poseidon/	0.60%	1,240	0.27%	1,028	0.40%
5 /text/	0.52%	11,263	2.90%	11,206	0.27%
/text/students/	0.36%	7,701	1.99%	7,644	0.12%
/text/teachers/	0.90%	1,812	0.48%	1,812	0.10%
6 /ocean608/	3.41%	73,738	2.52%	9,722	22.33%
/ocean608/homework/	0.17%	3,614	0.89%	3,389	0.23%
7 /ocean410/	0.70%	15,175	0.82%	3,173	0.51%
/ocean410/homework/	0.60%	1,227	0.29%	1,079	0.60%
8 /ocean401/	0.59%	12,593	0.63%	2,426	0.42%
9 /education/	0.47%	9,985	0.62%	2,388	0.28%
/education/earthscience/	0.43%	9,318	0.58%	2,207	0.24%
10 /realdatalinks/	0.90%	1,794	0.36%	1,350	0.80%
[not listed: 46]	38.13%	824,198	1.97%	7,584	21.90%

an educational website funded by NASA through the JASON-1 Education Project, and produced by the Texas A&M Department of Oceanography

Section 5: Directory and Request Analysis (cont.)

Table 5-4 Top 20 Most Request HTML Documents

	File Name	Number of requests	Percentage of the bytes	Date and time of last access
1.	/index.html	47,787	10.40%	Dec. 31, 2001 23:04
2.	/references/glossary.htm	8,504	4.60%	Dec. 31, 2001 21:56
3.	/students/students1_sm_table.htm	7,616	1.20%	Dec. 31, 2001 18:47
4.	/students/iceberg/icebergintro_sm_table.htm	7,168	1.62%	Dec. 31, 2001 22:40
5.	/currents_facts.html	4,283	2.30%	Jun. 15, 2000 11:33
6.	/teachers/teachers1_sm_table.htm	3,912	1.20%	Dec. 29, 2001 06:45
7.	/wave_facts.html	3,860	2.49%	Jun. 15, 2000 11:24
8.	/iceberg_facts.html	3,803	2.37%	Jun. 15, 2000 11:09
9.	/ice_age_facts.html	3,454	2.53%	Jun. 15, 2000 11:08
10.	/ocean_role_facts.html	3,254	1.51%	Jun. 15, 2000 11:09
11.	/students/coral/coralintro_sm_table.htm	3,179	0.74%	Dec. 31, 2001 16:06
12.	/students/currents/currentsintro_sm_table.htm	3,176	0.63%	Dec. 31, 2001 18:33
13.	/teachers/currents_oceancircul.htm	3,110	2.62%	Dec. 31, 2001 18:31
14.	/home.html	2,910	0.29%	Jun. 14, 2000 09:09
15.	/side_stuff.html	2,786	0.36%	Nov. 9, 1999 13:41
16.	/classroom_activities.html	2,770	0.20%	Jun. 15, 2000 11:09
17.	/ocean608/index.html	2,764	0.37%	Dec. 31, 2001 18:17
18.	/students/iceberg/iceberg1.htm	2,694	0.49%	Dec. 31, 2001 22:41
19.	/students/currents/currents1_sm_table.htm	2,541	0.38%	Dec. 30, 2001 05:18
20.	/students/currents/currents3_sm_table.htm	2,517	0.58%	Dec. 31, 2001 01:08

Section 5: Directory and Request Analysis (cont.)

The final reports we will analyze are the Directory and Request Reports. These statistics more than any other provide the best view of how the site is being used and which sections are the most popular. The first pie chart and table deal with the requests to a directory on the site. For example, the student content of the site is located in the /students/ folder on the server one level below the OceanWorld root directory. Any page requested from within this folder would be tallied in the log file as a request for the folder. Based on the figures from the preceding reports and tables, the directory reports contains some data which is to be expected. Namely, it is not surprising that the /students/ folder receives almost one half of all the total page requests to the site. However, two other listed directories are a surprising presence on this list. The first, the /text/ directory, contains the text-only version of the site. The Section 508 regulations passed by the US Government demand that government funded and sponsored sites must be accessible to those with physical disabilities. In response to this new regulation OceanWorld offered a text-only version of the site starting in the summer of 2001. The question on the minds of many was how much use would the text-only site receive. As indicated by the statistics here it is clearly a significant amount, ranking fifth by page requests in only eight months of being offered. Another surprising figure seen here is the inclusion of all three oceanography class websites in the top ten directories. Specifically, the use of the OCNG 608 site is staggering. The 72,000 requests made to the sub-site account for 22% of the total bytes of data transferred. This seemingly disproportionate number results from the inclusion of Dr. Stewart's online MP3 format lectures. Primarily serving TAMU students with access to high-bandwidth connections, the .mp3's of each of Dr. Stewart's hour-long lectures are approximately 10MB a piece.

Section 6: Conclusions and Future Areas of Work

Research into OceanWorld's log files clearly indicates that the site is an active, growing resource used by students ranging from elementary to college age levels. The upward trend of the site's exposure and volume of visitors per year is a good indicator that OceanWorld has established a reputation as a reliable, first-rate resource for the latest topics in oceanography. Support for the addition of features such as a new graphic interface and a text-only site is strong, while innovative concepts such as providing distance education courses with MP3 audio lectures is overwhelming.

If there is any one area of improvement to work towards, it is attracting more teachers to the site. While the /teachers/ directory ranks third amongst the most accessed areas of the site, its page request figure is only one-fifth that of the student section's corresponding statistic. In part stems from the fact that there are many more students than teachers to begin with, but already our team has begun outlining additions and revisions to make to the teachers' section. One such idea is the offering of a day-by-day curriculum for grade school teachers who, for example, have only a week to cover some large aspect of oceanography. Another key area to improve in is developing a system for more actively soliciting user feedback and input. As of last fall, a set of form based questionnaires were developed for this purpose. Unfortunately their use has been far less than hoped for. Again, this situation is already being addressed by team members. To start with, the questionnaires might be linked prominently on each page, instead of only to the major site division index pages. With the site gaining users at an enormous rate, the benefits which OceanWorld provides to the online community and the students who take advantage of its content far outweigh these issues which are already on their way to being resolved. The OceanWorld team has worked exceedingly hard to create a superior site serving the needs of both students and teachers interested in the field of oceanography. Clearly, we are on the right track.