Growth of the Web and Internet from 1993 to 1996

Matthew Gray

December 9, 1996

Abstract

The World Wide Web has grown substantially in the the past three years. This paper quantifies this growth based on data collected by the World Wide Web Wanderer between June, 1993 and June, 1996. The Wanderer employed a comprehensive breadth first search of the web, counting sites it encountered. The number of sites discovered increased from 130 in 6/93 to over 230,000 in 6/96. The Internet at large has grown substantially in this period as well, however the web fraction has increased even more drastically in this time.

Introduction

The World Wide Web was started originally proposed in 1989 and the first implementation appeared in 1990. The Web however, did not gain any widespread popular use until NCSA Mosaic was released in the fall of 1993.

This paper only addresses the subset of the web composed of HTTP servers and the documents (largely, but not exclusively formatted in HTML) served by those servers. Fortunately, this represents the majority of web by any reasonable counting.

This paper identifies a method and terminology for measuring the growth of the web. The measurements are primarily accomplished by "The World Wide Web Wanderer" (hereafter, the Wanderer), written by the author in May, 1993. The Wanderer retrieves a document, and identifies all the links it contains to other documents. It puts these other documents on a queue, and repeats this process, retreiving the document at the top of the queue. The primary measurement taken by the wanderer is the number of "web sites" serving documents.

This paper also contrasts of the growth of the web relative to the growth of the Internet in general. Typically, the growth of the Internet is measured in terms of growth of the number of hosts on the net (hard to define) and the number of users on the net (hard to define and hard to measure). For this reason, comparisons will be made between the growth of the number of web sites and the growth of the number of hosts on the Internet.

Definitions

While most of terms used and defined here are in common usage, they are typically poorly defined and ambiguous. The definitions here are used throughout this paper. It is acknowledged that many of these definitions are overly restrictive for general use. Terms with accepted existing technical definition, such as "HTTP" and "URI" are not redefined here.

Document A document is a single item accessible via HTTP, as identified by a single URI. Typically, this will be an HTML document with links to a number of other documents.

Host A host is a computer connected to the Internet. A distinction is made between **named hosts** which are hosts that are listed on a name server, but need not actually exist or be connected to the net at any given time or could be behind a restrictive firewall, and **pingable hosts** which are constantly on the net and not behind an overly restrictive firewall.

Internet The Internet is the worldwide connected network of computers, formerly connected by the NSFNET backbone. The issue of semi-private networks, such as those behind firewalls, is not relevant to this paper, except as specified in the definitions of *Host* and *Web Site*.

Web Site A web site is an HTTP server with a unique hostname. A single piece of hardware may have multiple hostnames, and even multiple IP addresses. This is counted as multiple hostnames, and hence multiple web sites. A single hostname may also map to multiple machines and IP addresses for load balancing purposes. This counts as a single web site. Additionally, there are certainly a large number of web sites on so-called "Intranets", which are not publicly accessible. It should be obvious that these are not included in this survey.

Methodology

The Wanderer was run on a monthly basis starting in June of 1993. A breadth first search was employed and the search was terminated after an extended series of retrievals which discovered no new sites. A subset of this list would then be used as a starting point for the next survey.

While this clearly has the potential to miss certain well hidden sites, it produces a consistent representative sample. Even if the absolute site counts are necassarily lower than the true count, the rate of growth is believed to be accurate. Additionally, near the end of the survey, as other counts appeared such as those by AltaVista[1] and the Netcraft Web Server Survey[2], it was possible to compare these with the Wanderer's results. The comparisons indicate that these results tended to err on the side of counting fewer sites than others, but only by a few percent.

Additionally, near the end of the survey, *hundreds* of new web sites were appearing daily, so maintaining the previous precision of the monthly counts was abandoned.

Corresponding data from the Net Wizards DNS Walk[3] and the Merit NFSNET backbone statistics[4] were obtained and collated.

Results

As mentioned above, the web has existed since 1990. However, it was initially confined to CERN, where it was developed and did not gain any substantial popularity until 1993. Unfortunately, no well accounted data exists for the number of web sites before June 1993, when the Wanderer was run. In fact, this was the original, and maintained, motivation for the Wanderer at all. The number of web sites discovered is listed in the second column of the following table.

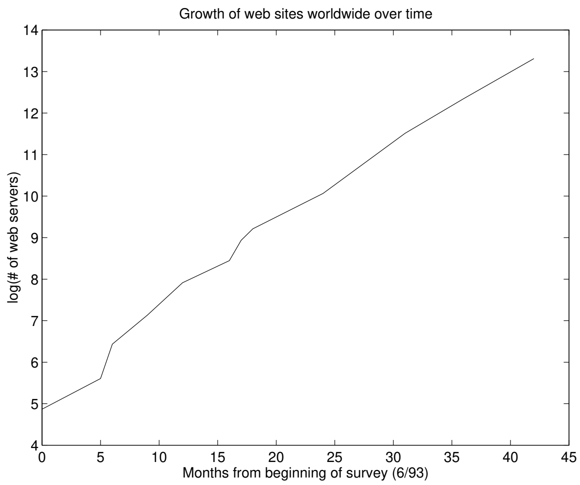
Additionally, the web started out as a vastly academic project. In June of 1993, when there were only 130 servers worldwide, only a half dozen sites existed outside of acadmia. Today, nearly two thirds of web sites are in the `.com` domain, which does not include commercial sites in `.net`, `.co.uk` or other international commercial domains. The percentage of sites in the `.com` domain is listed in the third column of the following table.

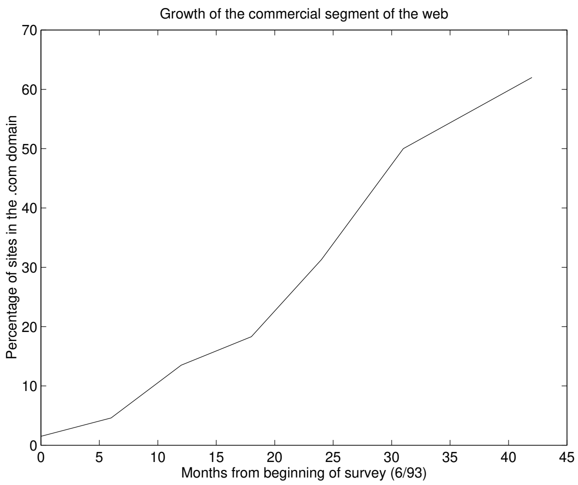
[1] http://altavista.digital.com/
[2] http://www.netcraft.co.uk/survey/
[3] http://www.nw.com/
[4] ftp://nic.merit.edu/statistics/nsfnet/

The web has also become a vastly more significant on the Internet than it was early in its evolution. In fact, the web is popularly credited with driving the explosive growth of the Internet. The number of hosts on the internet per web server has steadily decresead (1 host per web server would imply that every host was a web server). In the following table, the fourth column contains this data. The number listed is for named hosts, followed by the number for pingable hosts, in parentheses.

Month	Number of Web Sites	Percent .com sites	Net hosts per web server
6/93	130	1.5	13,000 (3,846)
12/93	623	4.6	3,475 (963)
6/94	2,738	13.5	1,095 (255)
12/94	10,022	18.3	451 (99)
6/95	23,500	31.3	270 (46)
1/96	100,000	50.0	94 (17)
6/96	230,000	NA	41
12/96[5]	600,000	61.7	NA

The log plot on the left of the increase in number of web servers worldwide clearly shows a more or less constant *exponential* growth rate. The right plot shows the increase in the commercial segment of the web.





[5]This number is from the Netcraft Server Survey

This gives us an average doubling period of 100 days. In fact, there has been a small slowing of the growth, as the doubling period was roughly 90 days for 1993 and 1994, and has increased to just under 115 days for 1995 and 1996. This however, is not substantial slowing, and is still conspicuously exponential.

Another way to measure growth of the web would be to measure usage, rather than number of web servers. Unfortunately, this is, in general, a very hard quantity to measure. Prior to the dismantling of the NSFNET, however, it was possible to get an approximate guage of usage of different protocols by analyzing traffic on the NSFNET backbone. For the early days of the web, in 1993 and early 1994, the web was a small fraction, comparable to gopher, and other even less popular protocols, and therefore often dismissed as yet another fad. Gopher experienced similar initial exponential growth, but leveled off at the end of 1993.

By the time the NSFNET was dismantled in spring of 1995, the web had far outrun all other protocols, except ftp, which it finally overcame in April, the final month of the NSFNET.

The following data is summarized from the monthly NSFNET backbone usage reports from Merit.[6]

Date	% ftp	% telnet	% netnews	% gopher	% email	% web
6/93	42.9	5.6	9.3	1.6	6.4	0.5
12/93	40.9	5.3	9.7	3.0	6.0	2.2
6/94	35.2	4.8	10.9	3.7	6.4	6.1
12/94	31.7	3.9	10.9	3.6	5.6	16.0
3/95	24.2	2.9	8.3	2.5	4.9	23.9
4/95	21.5	2.5	8.7	1.8	4.6	26.3

Interpretation

Conclusion

[6] ftp://nic.merit.edu/statistics/nsfnet/