

# A Brief History of the Internet

By: Marnie Webb

When I was a kid, we had to use Archie, WAIS to find things. We had to use gopher. We were on a first name basis with Unix commands. And we liked it. Hell, we loved it.

The Internet is not just Microsoft Internet Explorer, point-and-click shopping, B2B buzzwords, and AOL. The Internet's been around a lot longer than that. It's earliest incarnation, the Department of Defense's ARPANET, pre-dates Al Gore's first run for Congress.

## Sputnik Led To More Than Neil Armstrong's Famous Step

Dwight Eisenhower created Advanced Research Projects Agency (ARPA) in response to the Russian's 1957 Sputnik launch. By the early 60s, ARPA turned to computer networking and communications. In 1962, Dr. J. C. R. Licklider was put in charge of improving ARPA communication and brought the idea of creating a communication network of linked computers to academia. The idea was to create a decentralized communication system capable of surviving a nuclear attack.

In 1969 a group of UCLA grad students and their professor, Leonard Kleinrock, tried to log on to a computer at Stanford. A telephone in one hand, a computer keyboard in the other, they began to type the word "login". They typed and verified the appearance of each letter with the group at Stanford.

"Do you see the L?"

"Yes, we see the L," came the response.

We typed the O, and we asked, "Do you see the O?"

"Yes we see the O."

Then we typed the G, and the system crashed.

-Sacramento Bee, September 26, 1999.

In 1972, there was a public demonstration of ARPANET; the first email program was written; and, the first chat took place. In 1973, the first international connections - University College of London (England) and NORSAR (Norway) - were added.

The Internet was going global.

## When We Say Archie, We Don't Mean The Guy Standing Between Jughead And Veronica

Imagine trying to find an article on diabetes located on some unknown computer in an unknown file. Imagine that you have to find it by entering the name of the computer and the name of the file rather than by entering in a keyword and choosing from a list of possible articles. That's what it was like in the beginning. You had to know where you were going and how to get there without any directional hints. And, after you found it, you had to download it to your computer so you could look at it.

Then, Peter Deustsch developed Archie, a method of archiving and then searching for files on FTP servers. Brewster Kahle created WAIS, which allowed for the indexing and searching the full text of files. Those helped the finding process. The viewing process was still a lot of work. And then there was Gopher.

Gopher--named for the mascot at the University of Minnesota where Gopher was developed--is a way to allow remote files to be viewed on a computer. This is when the Internet began to get friendly. Look at the list and choose the number. The Gopher interface made life easier.

Veronica (Very Easy, Rodent-Oriented, Net-Wide Index to Computerized Archives) and Jughead (Jonzy's Universal Gopher Hierarchy Excavation And Display) did come along. Both of these tools were developed to index and search sites on Gopher servers. While Gopher servers, as well as Archie, Jughead, and Veronica, are still around, what came next put an end to their widespread use.

## Now It's Time For Those Famous, Three Ws

Technology was moving along. Vint Cerf sketched the basic gateway architecture that would make the Internet possible. TCP/IP - what would become the protocol for sending and receiving data on the Internet was developed; remote computers were searchable. Queen Elizabeth II sent an email message. The stage, as they say, was set for Tim Berners-Lee.

At CERN - the European Laboratory for Particle Physics, Tim Berners-Lee had the idea of "distributed computing" and the need to easily browse papers and documents located on other computers without needing to know anything about the other computer. Hypertext--an idea already in use within documents and computer programs--became hyperlinks, allowing a user to jump from one document to another effortlessly. Berners-Lee's original vision of the Web had to do with collaboration and creating a cooperative work environment.

It's a standard for thirty-something geeks. In the where-were-you-when-Kennedy-was-shot vein, those first heady days of Mosaic bring on fits of nostalgia worthy of WWII veterans sitting at the front table at a Big Band retrospective. It was pictures. It was the ability to click on an underlined word or phrase and go to an entirely different document on an entirely new computer. You didn't have to know the file structure, the name of the host computer. You didn't have to know any arcane commands. You only had to know that those underlined, blue words meant, "Click here".

Marc Andreessen the brain of Mosaic went on to found a little company called Netscape: dominant in the browser wars until Microsoft grew up with Internet Explorer 4.5.

And now it's a point-and-click world, baby, and you can access it without knowing any of that pesky command-line nonsense. All you need to know is the dot-com address and you're surfing like a pro.

Article date: May 26, 2000 Copyright ©2000 CompuMentor. All Rights Reserved.

Resource: <http://www.techsoup.org>

## **How Does the Internet Work?**

### ***A behind-the-scenes look at the Internet***

**By:** Marnie Webb

#### **It's All About Protocol**

Probably, you most commonly interface with the Internet via e-mail or the World Wide Web. That's not the entire Internet. It's just the popular and, relatively speaking, the new part. The Internet has a much longer history.

The Internet is based on a series of standard technical protocols which allow various computers located around the world to access specified files on other computers and then view those files. Specifically, the protocol in question is TCP/IP. TCP/IP allows computers to describe data to one another over a network. Every computer hooked to the Internet understands these two protocols and so can communicate well.

TCP/IP is, as that fancy little slash mark suggests, actually two separate things that work together. TCP - transmission control protocol - takes the information you want to send over the Internet and breaks it down into small chunks of data called "packets." IP - Internet Protocol - takes over and routes those packets through computers to get them to their destination. When the packets arrive at the destination computer, TCP reassembles them into something recognizable.

These two protocols allow information to be addressed, routed, and reassembled. You use this technology every single time you use the web.

There are other protocols involved too. STMP - simple, text mail protocol - works with e-mail. Ftp - file transfer protocol - is essential for uploading and downloading files to and from other computers. That familiar http in your browser's location or address bar stands for hypertext transfer protocol.

All of these protocols ensure that the computers attempting to communicate with each other - through e-mail or web pages or any other mechanism - understand each other.

And that provides the technology necessary for your computer to hook into the Internet.

#### **An ISP is Required**

All those protocols allow computers to communicate. You have a computer. You want to play along. Typically, this is how it works.

Your computer connects to an Internet Service Provider. You may be dialing up or you may be using high bandwidth method, such as DSL or cable, to connect. Your computer connects to your ISP's server. Once there, your ISP provides you with the gateway to connect to any other computer that has opened itself up to the world.

When you type in a domain - such as [www.techsoup.org](http://www.techsoup.org) - that domain is translated into a number - the IP address - and you are taken to that specific computer. Once there, your web browser allows you to look at specific files. These files can include programming, text, pictures, sound, or video in various combinations.

## And E-mail?

You write a letter and take it to the mailbox. It sits in the mailbox until the mail carrier picks it up. The address on the front of the envelope directs the letter to a specific person at a specific location. It's taken to the addressee's mailbox and put inside. It sits there until the addressee walks out to the mailbox and gets the letter.

This is basically how e-mail works. You address the email - [myfriend@domainname.org](mailto:myfriend@domainname.org) - the part before the "@" signifies the person; the part after the "@" the server. Your e-mail sits in a queue on a computer - called a server - at your ISP. Your ISP sends out all the queued up e-mails. Those e-mails go to the specified servers and then are routed to the specified users. The e-mail sits there until the user goes online to pick up her mail.

Article date: June 16, 2000 Copyright ©2000 CompuMentor. All Rights Reserved.

Resource: <http://www.techsoup.org>

## How Domain Name Servers Work

When you use the Web or send an e-mail message, you use a **domain name** to do it. For example, the URL "http://www.howstuffworks.com" contains the domain name **howstuffworks.com**. So does the e-mail address "iknow@howstuffworks.com."

Human-readable names like "howstuffworks.com" are easy for people to remember, but they don't do machines any good. All of the machines use names called **IP addresses** to refer to one another. For example, the machine that humans refer to as "www.howstuffworks.com" has the IP address **216.183.103.150**. Every time you use a domain name, you use the Internet's domain name servers (DNS) to translate the human-readable domain name into the machine-readable IP address. During a day of browsing and e-mailing, you might access the domain name servers hundreds of times!

Domain name servers translate domain names to IP addresses. That sounds like a simple task, and it would be -- except for five things:

- There are billions of IP addresses currently in use, and most machines have a human-readable name as well.
- There are many billions of DNS requests made every day. A single person can easily make a hundred or more DNS requests a day, and there are hundreds of millions of people and machines using the Internet daily.
- Domain names and IP addresses change daily.
- New domain names get created daily.
- Millions of people do the work to change and add domain names and IP addresses every day.

The DNS system is a **database**, and no other database on the planet gets this many requests. No other database on the planet has millions of people changing it every day, either. That is what makes the DNS system so unique!

## IP Addresses

To keep all of the machines on the Internet straight, each machine is assigned a unique address called an **IP address**. IP stands for **Internet protocol**, and these addresses are 32-bit numbers normally expressed as four "octets" in a "dotted decimal number." A typical IP address looks like this: 216.183.103.150

The four numbers in an IP address are called **octets** because they can have values between 0 and 255 ( $2^8$  possibilities per octet).

Every machine on the Internet has its own IP address. A server has a static IP address that does not change very often. A home machine that is dialing up through a modem often has an IP address that is assigned by the ISP when you dial in. That IP address is unique for your session and may be different the next time you dial in. In this way, an ISP only needs one IP address for each modem it supports, rather than for every customer.

If you are working on a Windows machine, you can view your current IP address with the command **IPCONFIG.EXE**

As far as the Internet's machines are concerned, an IP address is all that you need to talk to a server. For example, you can type in your browser the URL **http://216.183.103.150** and you will arrive at the machine that contains the Web server for HowStuffWorks. Domain names are strictly a human convenience.

## Domain Names

If we had to remember the IP addresses of all of the Web sites we visit every day, we would all go nuts. Human beings just are not that good at remembering strings of numbers. We are good at remembering words, however, and that is where domain names come in. You probably have hundreds of domain names stored in your head. For example:

- www.howstuffworks.com - a typical name
- www.google.com - the world's best-known name
- www.mit.edu - a popular EDU name
- encarta.msn.com - a Web server that does not start with www
- www.bbc.co.uk - a name using four parts rather than three
- ftp.microsoft.com - an FTP server rather than a Web server

The COM, EDU and UK portions of these domain names are called the **top-level domain** or **first-level domain**. There are several hundred top-level domain names, including COM, EDU, GOV, MIL, NET, ORG and INT, as well as unique two-letter combinations for every country.

.com – commercial	.edu – educational	.gov – government
.mil – military	.net – network organization	.org – organizations (non-profits)

For example, in the COM first-level domain, you've got names like:

- google
- msn
- microsoft

Every name in the COM top-level domain must be **unique**, but there can be duplication across domains. For example, **howstuffworks.com** and **howstuffworks.org** are completely different machines.

In the case of **bbc.co.uk**, it is a third-level domain. Up to **127 levels** are possible, although more than four is rare.

http: //www.newton.k12.in.us/ art/galleries/cggallery6.html  
| protocol | Host | Path & File details |

## Naming Your Files

All file names should be saved in lowercase.

When you SAVE AS, you are saving the code as an HTML file, which is a web page. Every web page has a name that ends in .htm or .html. This is how the browser knows it is a web page. The very first page of your web site must be named index.html or index.htm in all lowercase. (sometimes default.html depending on the server)

When saving file names, use only letters or numbers – no characters. You can use the tilde (~), underscore (\_), hyphen (-) or period (.) Never use a space in a file name. Keep file names as short as possible. It makes for fewer typos, less information to load, and shorter URLs.

## Distributing Domain Names

Because all of the names in a given domain need to be unique, there has to be a single entity that controls the list and makes sure no duplicates arise. For example, the COM domain cannot contain any duplicate names, and a company called Network Solutions is in charge of maintaining this list. When you register a domain name, it goes through one of several dozen registrars who work with Network Solutions to add names to the list. Network Solutions, in turn, keeps a central database known as the whois database that contains information about the owner and name servers for each domain. If you go to the whois form, you can find information about any domain currently in existence.

Name servers do two things all day long:

- They accept requests from programs to convert domain names into IP addresses.
- They accept requests from other name servers to convert domain names into IP addresses.

When a request comes in, the name server can do one of four things with it:

- It can answer the request with an IP address because it already knows the IP address for the domain.
- It can contact another name server and try to find the IP address for the name requested. It may have to do this multiple times.
- It can say, "I don't know the IP address for the domain you requested, but here's the IP address for a name server that knows more than I do."
- It can return an error message because the requested domain name is invalid or does not exist.

When you type a URL into your browser, the browser's first step is to convert the domain name and host name into an IP address so that the browser can go request a Web page from the machine at that IP address. To do this conversion, the browser has a conversation with a name server.

Resource: <http://computer.howstuffworks.com/dns.htm>

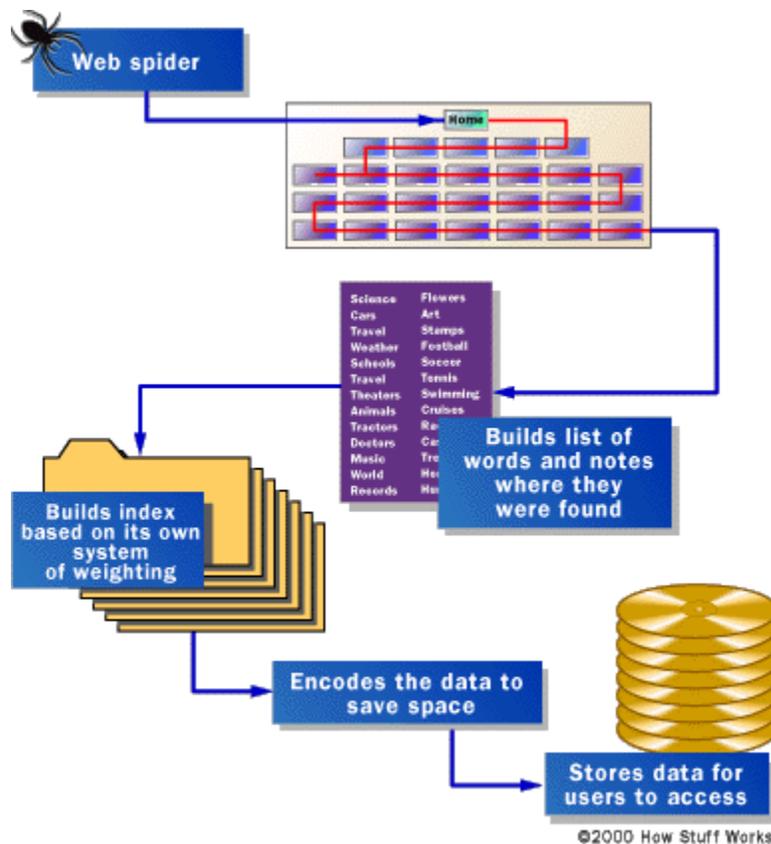
# How Internet Search Engines Work

When most people talk about Internet search engines, they really mean World Wide Web search engines. Before the Web became the most visible part of the Internet, there were already search engines in place to help people find information on the Net. Programs with names like "gopher" and "Archie" kept indexes of files stored on servers connected to the Internet, and dramatically reduced the amount of time required to find programs and documents. In the late 1980s, getting serious value from the Internet meant knowing how to use gopher, Archie, Veronica and the rest. Today, most Internet users limit their searches to the Web, so we'll limit this article to search engines that focus on the contents of Web pages.

## An Itsy-Bitsy Beginning

Before a search engine can tell you where a file or document is, it must be found. To find information on the hundreds of millions of Web pages that exist, a search engine employs special software robots, called **spiders**, to build lists of the words found on Web sites. When a spider is building its lists, the process is called **Web crawling**. In order to build and maintain a useful list of words, a search engine's spiders have to look at a lot of pages.

How does any spider start its travels over the Web? The usual starting points are lists of heavily used servers and very popular pages. The spider will begin with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel, spreading out across the most widely used portions of the Web.



**"Spiders" take a Web page's content and create key search words that enable online users to find pages they're looking for.**

Google.com began as an academic search engine. In the paper that describes how the system was built, Sergey Brin and Lawrence Page give an example of how quickly their spiders can work. They built their initial system to use multiple spiders, usually three at one time. Each spider could keep about 300 connections to Web pages open at a time. At its peak performance, using four spiders, their system could crawl over 100 pages per second, generating around 600 kilobytes of data each second.

Keeping everything running quickly meant building a system to feed necessary information to the spiders. The early Google system had a server dedicated to providing URLs to the spiders. Rather than depending on an Internet service provider for the domain name server (DNS) that translates a server's name into an address, Google had its own DNS, in order to keep delays to a minimum.

When the Google spider looked at an HTML page, it took note of two things:

- The words within the page
- Where the words were found

Words occurring in the title, subtitles, meta tags and other positions of relative importance were noted for special consideration during a subsequent user search. The Google spider was built to index every significant word on a page, leaving out the articles "a," "an" and "the." Other spiders take different approaches.

These different approaches usually attempt to make the spider operate faster, allow users to search more efficiently, or both. For example, some spiders will keep track of the words in the title, sub-headings and links, along with the 100 most frequently used words on the page and each word in the first 20 lines of text. Lycos is said to use this approach to spidering the Web.

Other systems, such as AltaVista, go in the other direction, indexing every single word on a page, including "a," "an," "the" and other "insignificant" words. The push to completeness in this approach is matched by other systems in the attention given to the unseen portion of the Web page, the meta tags.

Resource <http://computer.howstuffworks.com/search-engine.htm>

## Building a Search

Searching through an index involves a user building a **query** (an instance of questioning) and submitting it through the search engine. The query can be quite simple, a single word at minimum. Building a more complex query requires the use of Boolean operators that allow you to refine and extend the terms of the search.

The Boolean operators most often seen are:

- **AND** - All the terms joined by "AND" must appear in the pages or documents. Some search engines substitute the operator "+" for the word AND.
- **OR** - At least one of the terms joined by "OR" must appear in the pages or documents.
- **NOT** - The term or terms following "NOT" must not appear in the pages or documents. Some search engines substitute the operator "-" for the word NOT.
- **FOLLOWED BY** - One of the terms must be directly followed by the other.
- **NEAR** - One of the terms must be within a specified number of words of the other.
- **Quotation Marks** - The words between the quotation marks are treated as a phrase, and that phrase must be found within the document or file.

## Meta Tags

**Meta tags** allow the owner of a page to specify key words and concepts under which the page will be indexed. This can be helpful, especially in cases in which the words on the page might have double or triple meanings -- the meta tags can guide the search engine in choosing which of the several possible meanings for these words is correct. There is, however, a danger in over-reliance on meta tags, because a careless or unscrupulous page owner might add meta tags that fit very popular topics but have nothing to do with the actual contents of the page. To protect against this, spiders will correlate meta tags with page content, rejecting the meta tags that don't match the words on the page.

The two most common Meta tags let you specify a description and a set of keywords for the page, respectively, and you can also declare all sorts of other things. Meta tags always appear between the <HEAD> and </HEAD> tags of the HTML source. Here are the Meta tags that appear at the top of the main index page for **HowStuffWorks**:

```
<Meta Name="Description" Content="This is a website to over 50 fantastic 'HowStuffWorks' articles!">
```

```
<Meta Name="keywords" Content="how stuff works, how things work, the way things work, information, tutorials, explanation, explains, understand, understanding, tell me, students, teachers, study, educational">
```

Resource: <http://computer.howstuffworks.com/web-page6.htm>