

The Web's Missing Links: Search Engines and Portals

Thomas Haigh – thaigh@computer.org, www.tomandmaria.com/tom

For inclusion in “The Internet and American Business” eds. William Aspray and Paul Ceruzzi, MIT Press, 2007. This version will differ in minor respects from the published version. I have updated my submitted draft to fix many small errors and clumsy expressions marked up on the galley copied by the copyeditor. On the other hand, I have defied some elements of the MIT style guide that seem perverse and avoided introducing a few errors created by the copy editor in the editing process. And of course the pagination will be quite different. If you wish to quote or cite this paper, please use the official published version.

As a dense but disorganized jungle of information, the Web has always relied on automated search engines, human-compiled directory services, and what came to be known as Web portals to steer its users toward the material they seek. Unlike other electronic publishing systems the Web had no central directory or integrated search function, so these services played an integral role in establishing the Web as a useful publishing medium. Search engine firms such as Excite and Lycos were among the first “Internet companies” to make initial public offerings of stock, fueling the boom for dot com stocks during the late 1990s. Although the Web has changed greatly since the early-1990s, throughout its evolution the most visited Web sites have been navigation services. Today the world’s four most visited websites are Yahoo, Microsoft’s MSN portal site, Google, and the Chinese search engine Baidu. Search sites and portals command the lion’s share of the world’s Internet advertising revenue, making them the most successful of all Internet information businesses. Yahoo and Google together receive more

advertising dollars than the combined prime time offerings of the traditional big-three U.S. television networks.¹

Yet contrary to this brief and happy summary, the development of the Web navigation industry has been anything but straightforward. This chapter tells its tempestuous history, beginning with the special characteristics of the Web, and of the Internet itself, when compared with earlier electronic publishing systems. The very features of the Web that brought it instant success also created an urgent need for navigation services. I contrast the hypertext approach to electronic publishing, used by the Web, with the information retrieval approach taken by earlier commercial online publishing systems. Early firms took one of two approaches. Web directories, such as Yahoo, exploited the hypertext nature of the Web and the power of human labor to create online guides combining features of traditional Yellow Pages business directories and library catalog files. Search engines, such as Lycos, Excite, and AltaVista, adapted traditional information retrieval techniques to the new and fragmented world of the Web to create huge, automatically generated, searchable indexes of all the text held on Web pages.

By the late 1990s both groups of firms had evolved into so-called Web portals, competing with each other as well as competitors such as AOL, Microsoft, Netscape and Disney to create one-stop Web sites stuffed with so many different kinds of attractions (among them news, weather, email, music and shopping) that users would browse there for hours at a time. The rush to create these “full-service” portals, their apparent success in the late 1990s, and their subsequent collapse can be explained only by reference to the exceptional economic conditions that warped normal business logic during the dot com boom. Finally, I turn to the success of

¹ Kris Oser, "New ad kings: Yahoo, Google", *Advertising Age*, April 25 2005, 1 forecast that this would take place during 2005, since which online advertising revenues have risen sharply.

Google, which rose to dominate the Internet search field by focusing on excellence in search just as its established rivals began to downplay search in their bid to become portals. Google's focus on improving the user's search experience, and refusal to emulate the crass and intrusive advertising practices of its rivals has, paradoxically, made it the most successful seller of online advertising in the history of the Internet.

The Web as an Electronic Publishing System

Within five years of its 1991 introduction the World Wide Web was already the biggest, most heavily invested in, most publicized, most widely used, and most geographically distributed online publishing system in the history of the world. It was also the first major online publishing system to be built with no centralized index or built-in method to search its contents for items of interest. These two accomplishments are not unrelated. The appealing characteristics of the early Web were achieved only because the thorny issue of search was initially ignored: its simplicity, its flexibility, and the ease with which new websites could be created.

The Web was by no means the first online publishing system. In fact, by the time it made its debut the industry was around twenty years old. The foundations for these online systems were laid in the 1950s, when "information retrieval" first surfaced as a field of study among academic and corporate researchers. Information retrieval combined the use of new technologies (including punched cards, microfilm, and specialist electronic devices as well as early computers) with information theory and the application of techniques of classifying and structuring information drawn from library science.² During the 1950s and 1960s, information

² For an early and enthusiastic report on the application of information retrieval to business, see Francis Bello, "How to Cope with Information", *Fortune* 62, no. 3 (September 1960):162-67, 80-82, 87-89, 92.

retrieval was a central concern of the nascent community of “information scientists,” who were concerned particularly with managing academic journal articles and other scientific information they believed was increasing exponentially in an “information explosion.”³ As the name information retrieval suggests, specialists in this area believed that *retrieving* information was the key challenge, working on methods of organizing, selecting, and displaying information to maximize the effectiveness of this process. As the field developed, its focus was on methods of selecting the relevant results from large collections of electronic records, such as tagging documents with metadata (keywords, date, author, and so on), indexing them, abstracting them, finding the most effective search algorithms, and analyzing users’ searching patterns.

Companies first began to use computers to provide online searching and retrieval from textual databases during the 1960s. In 1969 the *New York Times* announced its Information Bank service, a publicly accessible online database including abstracts from many other magazines, newspapers and magazines, although the system only became fully operational in 1973.⁴ By the start of the 1970s, Lexis, one of the first successful online information services, was created to provide access to legal text databases. Dialog, one of the first online search packages, was

³ The origins of the concept of “information science” are discussed, rather critically, in Hans Wellisch, “From Information Science to Informatics: A Terminological Investigation”, *Journal of Librarianship* 4, no. 3 (July 1972):157-87 and Mark D Bowles, “The Information Wars: Two Cultures and the Conflict in Information Retrieval, 1945-1999”, in *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, ed. Mary Ellen Bowden, Trudi Bellardo Hahn, and Robert V. Williams (Medford, NJ: Information Today, Inc., 1999), 156-66.

⁴ Charles P Bourne and Trudi Bellardo Hahn, *A History of Online Information Services: 1963-1976* (Cambridge, MA: MIT Press, 2003), 322-28.

developed by Lockheed in the mid-1960s, with NASA as its first customer.⁵ From the early 1970s onward, Lockheed used Dialog as the basis for a publicly accessible online service.⁶ Now owned by specialist publishing giant Thomson, Dialog is still a hugely successful service providing access to databases of journals, newspapers, magazines, analyst notes, patents, government regulations and other sources.

But Dialog, and its competitors such as Lexis-Nexis, evolved quite separately from the Internet and have remained distinct from the Web even though most subscribers today use a Web browser to access them. They are sometimes called the “deep Web,” and provide their subscribers with a much more orderly world of professional and technical information inaccessible through Web search engines. Unlike the Web, online information services are centralized, with all documents indexed and held in a central repository. Because documents are tagged with metadata, users can choose to search only for results written during a particular month, published in a certain source, or written by a specific reporter.

The Web took a fundamentally different approach to publishing and finding electronic documents. Previous electronic publishing systems involved a number of users logged into a centralized server, on which documents were stored and indexed. The Web had no central server, and thus no central directory or index. The peer-to-peer structure of the Internet meant that any computer on the network could publish Web pages. If you already had access to a computer with a direct and reasonably fast Internet connection, which in the early 1990s was a common occurrence at universities and in computing research labs and a rather uncommon one elsewhere, then all you needed to do was install the Web server program and create a directory holding a

⁵ Ibid, 141-83.

⁶ Ibid, 280-86.

Web page or two. There were no forms to fill out, no licenses to apply for, no royalties to pay, and no special fees to negotiate. The Web made online publishing almost ludicrously easy, when compared with the enormous amount of work involved in setting up a traditional online information retrieval service or arranging to publish material through a commercial online service such as AOL or CompuServe.

The very ease with which a Web page could be published created a new issue: how would anyone ever find it? If one had the address of a Web page, one could jump directly to it. But Tim Berners-Lee expected users to glide move effortlessly around the Web as they followed links from one document to another. The connections between Berners-Lee's design for the World Wide Web and earlier work on hypertext are well known. By 1991 hypertext had already been incorporated into some very widely used computer systems. Since 1987 the Hypercard system bundled with each Macintosh system had allowed people with no programming knowledge to create complex "stacks" incorporating sounds and animations as well as text and links. In today's terms, Hypercard was half way between a Web browser and PowerPoint. Hypertext was also an integral part of the standard help system used by Microsoft Windows itself and by Windows application programs.

Hypertext represented a fundamentally different paradigm for electronic publishing from traditional information retrieval. Information Retrieval systems searched through a huge body of independent documents to find the best matches for a users' query. Hypertext allowed readers to browse from one page to another, following cross references, jumping around tables of contents, and meandering into footnotes and away along digressions into other works. Long before the Web existed, hypertext pioneer Ted Nelson was popularizing the idea of a world wide hypertext

network knitting together independent documents within what he called the “docuverse.”⁷ Even Vannevar Bush, whose 1945 speculative essay "As We May Think" is often claimed to have invented the hyperlink, imagined a miniaturized library in which users could create and share their own “trails” linking material found in many different books and papers.⁸

The idea of a hypertext system or full text database spanning material held on multiple, independently administered servers was nothing new. Indeed, hypertext specialists initially felt that Berners-Lee had done nothing to advance their research, rejecting his paper describing the Web when he submitted it for consideration at the prestigious *Hypertext'91* conference.⁹ But while the concept of a world wide web of hypertext was nothing new, the reality most certainly was. The actual hypertext systems of the late 1980s were limited in scope. Users could browse about within a single hypertext document, such as a book or a reference manual, but there was no obvious way to create a link from one author’s work to another’s. Establishing a distributed, public hypertext network posed some very challenging problems. How to establish a central database so that all links to a particular document would automatically be preserved when its physical location on the network shifted, or the document itself was revised? How to see every document that had been linked to the document one was currently reading? Nelson himself was never able to fully answer these riddles in practice despite decades of work toward a system he

⁷ Nelson’s fullest explanation of his vision of a worldwide electronic hypertext publishing network open to all was given in Theodore H Nelson, Literary Machines (Swarthmore, PA: Mindful Press, 1982).

⁸ Vannevar Bush, "As We May Think", The Atlantic Monthly 176, no. 1 (July 1945):101-08.

⁹ Tim Berners-Lee and Mark Fischetti, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor (San Francisco: Harper, 1999), 50.

called Xanadu, and they preoccupied an established community of hypertext researchers during the late 1980s and early 1990s.¹⁰

Berners Lee, who was a practicing programmer rather than an academic researcher, dealt with these thorny, fundamental problems by ignoring them. His concerns were more practical: he had to struggle to win permission from his managers at CERN to spend time on his pet project and was obliged to justify the early Web as a useful document sharing system for use within the lab rather than as a long-term research project to create fundamental advances in hypertext.

The Web offered neither an information retrieval search capability nor the full richness of Nelson's original conception of hypertext. Published material was often deleted or changed, meaning that links pointed to missing or irrelevant pages. Links between documents could be followed forward but not backward. Nelson was outraged: "The World Wide Web was not what we were working toward, it was what we were trying to *prevent*," he wrote a few years later, complaining that it had "displaced our principled model with something far more raw, chaotic and short-sighted."¹¹ Famous computing researcher Alan Kay likened the Web's HTML coding system to Microsoft's crude but ubiquitous MS-DOS operating system, which in computer science circles is one of the nastiest things you can say about a design.¹²

¹⁰ A readable overview of the Xanadu project was given in Gary Wolf, "The Curse of Xanadu", Wired Magazine, June 1995, 137-202.

¹¹ Theodore Holm Nelson, "Xanalogical Structure, Needed Now More Than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Reuse", ACM Computing Surveys 31, no. 4es (December 1999).

¹² Niklas Rudemo, "Beyond HTML: Web issues aired in Darmstadt." Seybold Report on Desktop Publishing 9, no. 9 (May 8 1995):10-11. An insider's discussion of the early Web's hypertext functions and their evolution is given in Robert Cailliau and Helen Ashman, "Hypertext in the Web — A History", ACM Computing

But the Web's spectacular success during 1993 and 1994 shows that a distributed hypertext system did not actually have to deal with these underlying problems to find an enthusiastic group of users. The Web grew fast, and the ambitions of its early users grew with it, until they began to write of it as a new Alexandrian library, tying together the sum of human knowledge. The problem was that the Web was a massive and ever-growing library without reference librarians, quality control, or a card catalog. Any foray onto the Web was likely to bombard the explorer with trivia on any number of esoteric topics. But finding the answer to a specific question could be hard indeed. As one well known science writer reported in mid-1995, "I have briefly signed up with a number of Internet providers, only to become exasperated by the maddening randomness of the Net.... [T]he complicated searches... feel like a waste of time."¹³

As the commercialization of the Web began in earnest in 1994, it was already obvious that creating fast and easy ways of searching or navigating its contents was a necessary precondition for the success of Web based businesses. One early user reported that "the Web is anarchy right now, and the librarians among us will quickly learn there is no easy way to search the unruly masses of home pages and Internet information." Yet, as she correctly suggested, "that's the next step and it will happen quickly, just as rudimentary tools for searching Internet gophers have developed."¹⁴

The problem was particularly pressing for online businesses. Putting up a website was easy. But how would a customer ever find it? A host of new websites and businesses sprang up

Surveys 31, no. 4es (December 1999). Other papers in the same electronic volume give a valuable discussion of the relationship of the Web to hypertext research.

¹³ Charles C Mann, "Is the Internet Doomed?" Inc 17, no. 9 (June 13 1995):47-50, 52, 54.

¹⁴ Nancy Garman, "A New Online World", Online 19, no. 2 (March-April 1995):6-7

to address the problem. Many of them had off-line parallels. In the physical world, for example, indexing and searching happens manually. A great deal of work goes into gathering information to create telephone books, encyclopedias, and Yellow Pages business directories. Businesses advertise their existence in specialist publications and trade directories. But the need for such mechanisms was much greater on the Internet. There was no such thing as foot traffic, and the Web was so new that its users lacked the purchasing habits, established routines, or brand recognition that most businesses rely on. Yet the payoff has also been huge. Web searching is, in retrospect, one of the few truly great business opportunities created by the Internet: a pure information business to supply a missing but vital function of the Web itself.

Clearly some new kind of company was going to get rich bringing Web sites and their potential visitors together to the benefit of both. But because the Web was so new there was no established way of doing this. In 1994 the commercial viability of Web publishing, Web searching and Web indexing was an article of faith rather than an observed fact. While it was certainly possible to spend a lot of money to create a Web directory or search engine, or to fill a website with witty and entertaining writing, nobody knew what kinds of things ordinary people would use the Internet for, what they might be willing to pay for, and what new kinds of online publishing businesses or genres would prove viable. Publishing genres such as newspapers, trade magazines, and directories had evolved over many decades to create well defined business models. In the virtual world of the Internet, a catalog could also be a store, or a trade publication could run an auction. A newspaper could be sold one article at a time, and an encyclopedia could be supported by advertising revenue.

One might expect this to give rise to a period of cautious experimentation on a limited scale. Instead a frenzy ensued, as corporate giants and well-funded startup companies rushed to

colonize every conceivable niche within the emerging ecosystem of Internet business. This was era of day trading, stock splits, investment clubs, financial news channels in bars, companies hoping to make money by giving away computers, and the best selling book *Dow 30,000*. Much has been written about the follies of the dot com era, so there is no need here to recount its sorry history in full.¹⁵ But to understand the story that follows a few aspects of this era must be remembered.

First, politicians, corporate leaders and investment advisers all endorsed the idea of the Internet as a technological force destined to sweep away every aspect of the existing business world. Any idea or business, no matter how silly or staid, could instantly legitimate itself through association with the Internet.

Second, although the huge importance of the Internet was universally recognized, the precise means by which it was to translate into great business success was not. Fads swept the world of Internet business every few months, each one bringing a new rash of start-up companies and promoting existing businesses to make wrenching strategic shifts. As conventional wisdom shifted again and again the early search companies grappled with advertising, rushed to provide original content, explored “push” technologies, tried to build personalization capabilities, and started to call themselves portals.¹⁶

Third, from 1994 to 2000 the ordinary laws of nature governing management and corporate finance were suspended. Excite, Lycos, and other search firms were generously funded

¹⁵ Readable overviews of the dot com era can be found in _____ and John Cassidy, *Dot.Con: How America Lost Its Mind and Money in the Internet Era* (New York: HarperCollins, 2002).

¹⁶ The difficulties of running a company in a business swept by such fads is vividly captured in Michael Wolff, *Burn Rate: How I Survived the Gold Rush Years on the Internet* (New York: Simon & Schuster, 1998).

by venture capitalists, went public, saw their stock prices soar and were able to issue more stock, hire thousands of employees, make expensive acquisitions and branch out into many new areas of business. Through all of this, their losses mounted ceaselessly. Indeed, the rate at which they lost money only increased as their businesses grew. But investors believed that growth, market share and the number of users were more important gages of an Internet company's worth than whether it made a profit. In the final years of the boom, this idea was cynically promoted by a powerful alliance of brokerage houses, investment banks and venture capitalists. Because an Internet-related firm could make an initial public offering of its stock without having a clear prospect of profitability, or more than a token number of customers, venture capitalists funded startup companies secure in the knowledge that a bank could reap huge fees by taking them public long before their viability had been established.¹⁷ Dot com firms may have been no more likely to fail than the average business (a point made by David Kitsch elsewhere in this volume), but they were able to fail much more expensively and spectacularly. When the bubble finally burst in 2000, all the major search and portal firms faced an abrupt, and usually fatal, encounter with reality. Search firms were a crucial part of the dot com business world, and despite the ultimate success of Yahoo and Google the story of their industry cannot be separated from the broader bubble.

Web Directories

Because it was made of hypertext the Web could, with a little work, become its own index. As in a paper book, there is nothing magic about an index page: it is just a page full of

¹⁷ The operation of the IPO feeding chain is explored in Roger Lowenstein, Origins of the Crash: The Great Bubble and its Undoing (New York: Penguin, 2004), 108-26.

references to other pages. Anyone who wanted to could set up a directory page full of links to other websites addressing a particular topic. For the first few years of the Web this was not particularly challenging – when CERN stopped updating its first master list of Web servers in late 1992 it held just twenty-six links.¹⁸ In the early days of the Web thousands of visitors were drawn daily to now forgotten directory sites such as EINet Galaxy, GNN, and CERN's own World Wide Web Virtual Library Subject Catalog, founded by Berners-Lee himself in 1991.

Some early directory sites depicted the Web visually. Flaunting the geographic reach of the Web, each known Web server in a country or region was shown as a labeled point on the map. Clicking on the point opened up the website in question. But the rapid spread of the Web made it impossible to track the location of every server, still less squeeze their names on to a map. According to an MIT survey, there were 130 Web servers on the Internet in mid-1993, 623 by the end of the year, and more than 10,000 by the end of the 1994.¹⁹ Even the most determined part-time indexer would probably have succumbed when the number of Web server doubled in the next six months, reaching 23,5000 in June 1995. The millionth server came online around March 1997, and the 10 millionth in early 2000.

As Web directories struggled to deal with the proliferation of sites, they adopted a hierarchical form, adding new levels whenever individual pages threatened to become unmanageably large. For example, a page of links to music websites might be replaced with multiple pages dealing with different musical genres, which in turn might eventually consist of

¹⁸ Tim Berners-Lee, W3 Servers (CERN, 1992 [cited August 15 2006]); available from <http://www.w3.org/History/19921103-hypertext/hypertext/DataSources/WWW/Servers.html>.

¹⁹ Matthew Gray, Web Growth Summary (1996 [cited September 03 2006]); available from <http://www.mit.edu:8001/people/mkgray/net/Web-growth-summary.html>.

little more than a set of links to pages cataloging websites dealing with particular bands. This made it possible to organize a list of hundreds of thousands of websites, but creating and maintaining such a list required a large and well-organized team of indexers.

Running a popular index could also be expensive because it consumed a lot of server capacity and network bandwidth. By far the most successful of the general-purpose Web directories, and the first to switch to a commercial mode of operation, was Yahoo (or, as it likes to call itself, “Yahoo!”). Yahoo began in early 1994 as “Jerry and David’s Guide to the World Wide Web,” a typical amateur directory, created by two electrical engineering students at Stanford University. The service quickly won a following, but keeping it up to date proved ever more time consuming. In early 1995 its founders, David Filo and Jerry Yang, incorporated the business and, following a path well beaten by Stanford spin-off companies before them, won startup funds from the venture capital companies lining Sand Hill road on the edge of campus.²⁰

Yahoo benefited from a great deal of free publicity. During 1994, publications such as *Business Week*, *Time*, *Newsweek* and hundreds of more specialized titles began to run stories about the exciting new World Wide Web. Over the next few years untold tutorials were published telling readers what the Web was, how to access it, and what to do with it. Web directories were the obvious place to send new Web surfers. These tutorials often included the addresses of Yahoo, and other directory sites. For example, Bill Gate’s notorious 1995 “Internet Tidal Wave” memo alerting Microsoft executives to the strategic power of the Internet included

²⁰ The early history of Yahoo is recounted in Karen Angel, *Inside Yahoo! Reinvention and the Road Ahead* (New York: John Wiley & Sons, 2002) and Robert H. Reid, *Architects of the Web: 1,000 Days that Built the Future of Business* (New York: John Wiley & Sons, 1997), 241-79.

a link to Yahoo under the heading “Cool, Cool, Cool.”²¹ Yahoo was also linked to from home pages created by universities, internet service providers, and other institutions.

Yahoo displayed its first advertisement in August 1994, which at least proved that a source of revenue existed to support the new business. It began to add other kinds of information such as news, weather and stock quotes and affixed its name to a print magazine, *Yahoo! Internet Life*. In April 1996, Yahoo made an initial public offering of stock, providing a rapid payback for its initial investors and cementing its position as a leading Internet firm. Like Netscape the previous year, its shares more than doubled on the first day of trading. At this point Yahoo employed fifty full time people to surf the Web, evaluating Web sites for inclusion in the directory.

The ever-increasing size of the Web made it ever more expensive for competitors to enter the directory business, because of the amount of human effort necessary to match Yahoo’s success. This was an expensive business to do well. Yahoo’s biggest early competition in the Web directory business came from McKinley with its Magellan directory, but the start-up faltered after running out of cash before it was able to sell its stock to the public.²² LookSmart, launched in 1996 by *Reader’s Digest*, survived but never enjoyed huge success. Some search engines and portal sites continued to offer their own directories, but Yahoo dominated the

²¹ Bill Gates, *The Internet Tidal Wave* (US Department of Justice, May 26 1995 [cited May 20 2006]); available from <http://www.usdoj.gov/atr/cases/exhibits/20.pdf>.

²² Julia Angwin, "Excite Will Buy Magellan Search Engine" *San Francisco Chronicle*, June 28 1996, C1. Journalist and former internet executive Michael Wolf gave a memorable portrayal of his experiences with Magellan’s leaders in Wolff, *Burn Rate: How I Survived the Gold Rush Years on the Internet*, 69-104.

market. Likewise, the human generated directory listings remained at the heart of Yahoo's business throughout the 1990s.

The commercial importance of Web directories has dwindled steadily since the mid-1990s. In recent years Yahoo's only main competition in the Web directory field came from the Open Directory Project. This was founded in 1998, to compete with Yahoo using volunteer labor to produce a directory.²³ In 1999 the effort was acquired by Netscape, and at some point in 2000 the size of its Web directory (around 1.6 million entries) is estimated to have overtaken Yahoo's.²⁴ While the project itself is not a household name, its directory listings were made freely available and have been used as the basis of Web directories offered by many popular and once popular Web navigation services, including AltaVista, Netscape and Lycos. By incorporating the Open Directory Project's results these firms could offer a similar capability to Yahoo's directory without having to spend a fortune to create and maintain it.

Web Search

One way to find something on the Web was to use a Web directory and click down through the subject headings to find a selection of websites dealing with the topic of interest. The other way, of course, was to search for Web pages that contained a particular word or phrase.

²³ Chris Sherman, "Humans Do It Better: Inside the Open Directory Project", Online 24, no. 4 (July 2000):43-44, 46, 48-50.

²⁴ The date at which ODP overtook Yahoo and its volume of pages are taken from the Wikipedia page on the project. I have been unable to verify this from a more stable source, though by 2001 the New York Times was referring the ODP as "the largest directory of the Web." Pamela Licalzi O'Connell, "Mining the Minds of the Masses" New York Times, March 8 2001, G1.

Yahoo offered a search facility, but until October 2002 this defaulted to searching the keywords and headings in the Yahoo directory, rather than the contents of the websites themselves.

The advantage of this approach was that the necessary indexing could be done automatically by a computer, eliminating the need to hire huge teams of Web surfers and allowing more thorough coverage of the Web. The disadvantage was that creating an automatic index of this kind was a much knottier programming challenge, and required a very fast network connection and a powerful collection of servers.

In the early days of the Web these two approaches were expected to coexist. When thousands of Web pages contained a popular phrase it was unrealistic to expect a user to visit more than the first few dozen in search of enlightenment. While a human surfing the Web for Yahoo could exercise a measure of quality control and editorial judgment, it was hard for automated systems to make an informed judgment as to which of the many Web pages about a popular subject, such as Britney Spears, deserved to be highlighted. So search engines appeared to have an edge in looking for every instance of an obscure phrase, whereas Web directories were better suited to recommending the best material on popular subjects.

The very ease with which material could be published on the Web meant that indexing and searching it posed a huge challenge. Searching the Web was actually a harder problem than designing the Web in the first place. Neither Berners-Lee nor any of his handful of volunteer collaborators and student interns at CERN were specialist hypertext or database researchers, but they managed to design the Web and create the first servers and browser programs in a matter of months. Neither did they have access to powerful hardware: the only special equipment used for the project was two NeXT desktop computers, whose purchase was authorized for other

reasons.²⁵ By necessity, the Web was designed around the existing infrastructure and capabilities of the Internet and so inherited both the strengths and the weaknesses of the underlying network. Search, on the other hand, went against the grain of the disorganized, decentralized, and unregulated Internet. Creating a search engine needed specialist technical expertise and esoteric algorithms, large powerful servers, fast network connections and teams of people ready to constantly tweak its hardware and software configuration.

The concept of an Internet search engine was already established before the creation of the first Web search engines. File transfer between computers was one of the original applications of the Internet, inherited from the early ARPANET.²⁶ All it took to share files with others was to run a piece of software called an FTP (File Transfer Protocol) server, allowing any computer on the Internet to connect, procure a list of available files, and download any of interest.²⁷ The facility was used in many ways during the 1980s and the early 1990s: by computer science departments to publish technical reports electronically, by public domain software archives to disseminate programs, and by the administrators of the Internet itself to share new technical documents.

But the ease of setting up an FTP server created a new problem. Because it required nothing more than downloading and installing a piece of software, there was no central index of

²⁵ Berners-Lee discusses the creation of the first Web software in Berners-Lee and Fischetti, Weaving the Web, 12-50.

²⁶ A Bhushan, RFC 114: A File Transfer Protocol (Network Working Group, 1971 [cited October 20 2006]); available from <http://tools.ietf.org/html/rfc114>.

²⁷ The standard for Internet file transfer was formalized in J Postel and J Reynolds, RFC 959: File Transfer Protocol (Network Working Group, October 1985 [cited October 25 2006]), which extended an earlier TCP/IP based FTP standard published in 1980.

files shared. Someone had to know the Internet address of the server they were looking for and the path to the directory holding the file they needed (perhaps by reading it in a newsletter or an online newsgroup). FTP servers hosting large or popular files consumed large amounts of network bandwidth, forcing the universities and research institutes hosting them to limit the number of simultaneous downloads allowed. This encouraged others to share the burden by establishing “mirror sites.” But spreading multiple copies of files around the network just made things worse if nobody knew where they could be found. It was to address this problem that a team of McGill University students introduced the popular utility Archie in 1990. Users could enter all or part of the name of the file they were interested in acquiring, and Archie would scour the net in search of it. It relied on a “spider” program to automatically query popular servers and then create a centralized, searchable database of all known files and locations. Archie, however, did not search the content of text files, merely the names of files and of the directories in which they were held.²⁸

In 1992 the success of Gopher inspired the creation of Veronica, a searchable database of the contents of thousands of Gopher servers.²⁹ Veronica searched Gopher index headings, but did not access text within the files that these headings pointed to. During the same era WAIS (Wide Area Information Servers) gave users a simple and consistent user interface with which to search the actual contents of text files held on servers all over the Internet. WAIS was promoted by supercomputer company Thinking Machines, which advertised its own computers as WAIS

²⁸ Archie and its operation are profiled in Ed Krol, The Whole Internet User's Guide and Catalog (Sebastapol, CA: O'Reilly & Associates, 1992), 155-68.

²⁹ Veronica is described, and its creators interviewed, in Billy Barron, "Tricks of the Internet Gurus", in Tricks of the Internet Gurus, ed. Anonymous (Indianapolis: SAMS Press, 1994), 519-38.

servers. Yet WAIS was very limited compared with later, Web-oriented, search systems. Rather than hosting a single massive central index, it relied on the people and organizations publishing text documents on the Internet taking the trouble to create and share their own indexes covering this material. Searchers then had to specify which of the hundreds of indexes they would like their search to include. This built on a long-discussed but seldom-implemented concept known as the “federated database.” The search capabilities of the widely used public domain versions were crude, always providing all the documents including a particular set of words regardless of whether they occurred as a phrase or were scattered around different sections of the text.³⁰

The sudden proliferation of Web sites from 1993 onward posed essentially the same problem, but on a bigger scale and with a more difficult target. The challenge was to find and index these pockets of information scattered randomly across the Internet. The basic technology behind automated Web indexing systems is known as a “Web crawler” or “Web spider.” In the absence of any official directory of Web servers, the Web crawler starts with a small set of Web pages and analyzes them to extract the addresses of all the other websites they link to. The program thus crawls from one site to another, discovering new destinations as it goes. Each page discovered by the crawler is saved to a special database, and the words occurring on it are indexed. Then when a user submits a query to the search engine it checks in the index to find

³⁰ WAIS built on an early version of the Z39.50 protocol, intended for use in searching library catalogs. This protocol gained widespread adoption during the 1990s. WAIS is discussed in Larry Press, "Collective Dynabases", Communications of the ACM 35, no. 6 (June 1992):26-32 and a lengthy introduction and tutorial can be found in Krol, The Whole Internet User's Guide and Catalog, 211-26. Brewster Kahle, one of the developers of WAIS, formed a company around the system and sold it to AOL in 1995 for \$15 million. WAIS then vanished without trace, but Kahle used the money to start non-profit Internet Archive, which continues to perform an important role by preserving the long-vanished content of old Websites.

pages in which all the specified terms occur. Techniques to search large bodies of text were already well developed prior to the creation of the Web, for online technical and legal databases. The Web crawler, however, was a new invention.

The first Web crawlers appeared in 1993, as the new Mosaic browser inspired the creation of thousands of Web sites. Washington University's WebCrawler was the best known of the early crawlers, and the first to allow users to search the full contents of the pages it indexed. Over the next few years, dozens of automated Web search services were launched. Running a successful service posed two main challenges: mustering sufficiently powerful hardware and efficient software to crawl a reasonable proportion of the Web on a regular basis, and creating a search system powerful enough to hunt through the resulting database of millions of saved pages in a reasonable time to present the most relevant results.

The most successful search engines built on existing expertise in these areas. AltaVista, launched in December 1995, was created by Digital Equipment Corporation and promoted as a tool to demonstrate the power of servers based on its new Alpha processor architecture.³¹ AltaVista won an enthusiastic following by delivering rapid searches of a comprehensive database via a powerful but uncluttered user interface, which included some of the advanced search features common in commercial information retrieval systems. The Excite search engine emerged from Architext, a firm created by several Stanford students with an interest in automatic textual analysis. After receiving funding from Kleiner Perkins Caulfield and Byers, one of Silicon Valley's leading venture capital firms, it quickly became one of the leading Web search

³¹ The launching of AltaVista is reported in Peter H Lewis, "Digital Equipment Offers Web Browsers its 'Super Spider'" New York Times, December 18 1995, D4.

sites.³² Lycos, another popular search engine, was developed from a Carnegie Mellon University research project and commercialized after being sold to a venture capital fund based (for a change) in Delaware.³³ Lycos boasted the biggest of the early Web databases, with a claimed 11.5 million pages indexed by November 1995.³⁴ Both Excite and Lycos supplemented their automatically generated search results with website reviews. Colorful banners signifying inclusion in Lycos' "Top 5%" directory were proudly displayed on many early websites. Inktomi, a 1996 start-up, was based on technology developed at the University of California at Berkeley. Inktomi's technology was licensed to other sites, and powered the popular HotBot search service offered by Web publisher HotWired. Ask Jeeves joined the competition late, in April 1997. Its gimmick was a friendly butler displayed on the home page to whom users were encouraged to type questions in full English sentences rather than brusque keywords.

Web Advertising

None of the major Web directories or search engines had any income during their first months in operation. This was perfectly normal for early Internet businesses, but obviously could not go on forever. Once an online business had a functional Web site and a stream of visitors, its next challenge was to make some money.

Nelson's concept for the Xanadu global hypertext network involved the use of so-called micropayments, where the creator of a document would automatically be credited with a tiny

³² Laurie Flynn, "Making Searches Easier In the Web's Sea of Data" New York Times, October 2 1995, D5

³³ Jon Auerbach, "In Search Of. Lycos Overhauls Product to Attract Users" Boston Globe, October 6 1995, 57.

³⁴ The figure on the size of the Lycos index is from Mike Holderness, "Online Searches: Where Confusion is Still Free" Guardian, November 30 1995, 6.

payment every time somebody read it.³⁵ Something similar had been realized in traditional online communities such as AOL it had been easy to charge customers for online transactions: additional fees were just added to the user's monthly statement. These were split between the online service and the creator of the content or service – such as the popular Motley Fool personal finance guide. This was similar to the arrangements used to charge users of premium telephone services. Only a small number of companies published information on AOL, but one could imagine a similar mechanism where small businesses and individuals could publish documents and be paid for them.

The economics of providing information on the Internet were different, and rather discouraging. The distinctive technological choices made during the design of the Internet and of the Web had profound implications for the commercial development of the Internet. The early Web was primarily an electronic publishing system, used to browse static pages. Because the Internet had not been designed for commercial use there was no way for a website or network owner to collect money from an ISP in return for services consumed by the user. That seemed to pose an economic problem, since websites could not easily charge their readers for access. The economics of off-line publishing were fairly straightforward: publishers made money on each book or record sold. Selling more copies mean more money, and one hit could underwrite the cost of several flops. In contrast, a popular website would run up huge bills for network bandwidth and servers, without receiving any income to cover this expense. Grabbing more readers meant bigger losses, not bigger profits.

During 1994 and 1995, many people believed that the missing tools to process micropayments could easily be retrofitted to the Web to provide a sound foundation for

³⁵ Nelson, Literary Machines.

commercial Web publishing. Numerous companies were launched to create “electronic cash” systems for online use, using advanced encryption techniques to create secure and efficient billing systems to support very small transactions. In reality they were not widely adopted by websites or users, and were never integrated into Web browsers. By the end of the 1990s, a host of well-funded, high-profile micropayment startup firms such as FirstVirtual, CyberCash and Pay2See had thrown hundreds of millions of dollars of investors' money at the problem with no success. The difficulty of simultaneously signing up a critical mass of users, enhancing browser and server software to make payments easy, and enlisting Web publishers proved intractable.³⁶

The one major commercial success among internet payment firms was Silicon Valley start-up PayPal, founded in 1998 and acquired by eBay for around \$1.5 billion in 2002. PayPal’s eventual business model was as a hybrid between a credit card processor and an unlicensed online bank rather than a true micropayment company, but it found a profitable niche among smalltime online auction sellers for whom the monthly fees demanded by banks to provide credit card processing services would be uneconomic.³⁷

Instead, the economics of Web publishing developed around advertising. Rather than harvesting micropayments from users, it was easier to leave browser technology untouched and instead deploy new “ad server” programs on websites to incorporate banner and pop-up advertisements into each Web page delivered. The publisher was still receiving a tiny sum for

³⁶ For a summary of the optimism surrounding micropayment firms during the mid-1990s and their rather limited practical accomplishments see Tom Seinert-Threlkeld, "The Buck Starts Here: Will Nanobucks be the Next Big Thing, Or Are We Just Talking Pocket Change?" Wired, August 1996, 133-35, 94-97.

³⁷ Eric M. Jackson, The PayPal Wars: Battles with eBay, the Media, the Mafia, and the Rest of Planet Earth (Torrance, CA: World Ahead Publishing, 2004).

each page displayed, but the sum was coming from an advertiser rather than the reader. Clicking on the advertisement took viewers straight to the advertiser's own website – a crucial advantage over traditional advertising media. HotWired, one of the first commercial Web publishers, pioneered the large-scale sale of Web advertising in 1994.³⁸ The largest Web publishing firms employed their own advertising sales teams, on the model of traditional magazine and newspaper publishers (both of which relied primarily on advertising to cover their costs).

The shift to advertising favored large, commercial Web sites able to attract the interest of advertisers and deploy ad server technologies. In the absence of micropayments amateur Web publishers had no easy way to profit from their work, however popular their sites. This shifted the balance of power toward popular Web directories and search engines, and away from small publishers. By the late 1990s just three firms were estimated to be receiving 43 percent of all online advertising revenues: AOL, Yahoo and Microsoft.³⁹

Early Web advertising was fairly unsophisticated. In theory, the Web allowed companies to track their customers and profile not only their purchasing history but also their browsing patterns: which topics they searched on, which products they lingered over, and what kinds of advertisements they had proved likely to click on in the past. This, according to Internet advertising specialists such as Doubleclick.com, would allow fundamentally new kinds of interactions with customers, showing them exactly the right advertisement or offer at the right

³⁸ HotWired faded quite quickly, but its glory days were captured in Reid, Architects of the Web: 1,000 Days that Built the Future of Business, 280-320.

³⁹ Angel, Inside Yahoo! Reinvention and the Road Ahead, 140.

time.⁴⁰ But in practice, most Web advertising brokers just placed ads on to websites indiscriminately and charged the advertiser ten or twenty dollars per thousand views. Just as with print magazine advertising, this relied on a neutral auditor to certify the actual number of readers exposed to the advertisement. Advertisements might be targeted at a general level (advertisements for video games on websites about video games) but were rarely turned to the content of a particular article, still less to the profile of a particular viewer.

Internet navigation companies quickly realized that they had a significant commercial advantage: they knew what their visitors were thinking about at any given moment. This did not require any elaborate data mining techniques. All they had to do was to see what words the user entered into the search box. Clearly somebody who has just searched for “BMW 3 series” is a much more valuable viewer of an advertisement for a luxury car than someone who happens to be browsing political news on the *New York Times* website. In 1995, Infoseek had already begun to sell advertising on with each search keyword, with popular words such as “Music” sold for forty dollars per thousand searches.⁴¹ These premium rates made advertising an appealing option for search businesses.

Several alternative sources of revenues were explored by search sites. One was imposing paid subscriptions or usage charges. Before the Web, this had been the dominant business model for online information retrieval. Services such as Lexis-Nexis and Dialog charged both

⁴⁰ Doubleclick's plans to profile Web users were reported in Hiawatha Bray, "For Advertisers, Web Offers Wide Audience, Pinpoint Accuracy" Boston Globe, May 5 1996, 41, 45. Its use of cookies to track use across different Websites led to considerable controversy.

⁴¹ Anonymous, "Internet Advertisers Can 'Buy' Key Words" Plain Dealer (Cleveland) December 26 1995, 3C.

subscription fees and usage charges to search databases of newspaper articles, scientific journals, patents, and so on. These services made a successful transition to the Web, but remained almost entirely separate from Web search services. In contrast, subscription-based Web search services had little appeal, because the same service was always available free from other search engines. Before shifting to its advertising supported business model, Infoseek briefly tried to justify subscription and usage fees for its search engine by offering extra features such as the ability to save searches and an index of Usenet newsgroups.⁴² Another company, Northern Light, tried to bridge the gap between the separate worlds of premium online information services and Web search engines by combining Internet search with paid access to proprietary content from academic journals and government sources.⁴³

Another other potential source of revenue was payment from companies in exchange for preferential treatment. This model had long been used by the Yellow Pages and other business directories. Yahoo and other Web directory services began to charge businesses for rapid or premium listings in the directory, supplementing their advertising income. Preferential treatment in search results was more complicated and controversial. It was pioneered by Overture, formerly known as GoTo.com. Overture focused on providing search services to other companies, rather than building traffic to its own websites. During the early 2000s, Overture ran the search functions on both Yahoo and Microsoft's MSN site, making it one of the biggest

⁴² Infoseek offered several plans for heavy or light users, with fees of 10 to 20 cents per search. Trial accounts and certain limited capabilities were free. Greg R Notess, "The InfoSeek Databases", Database Magazine 18, no. 4 (August/September 1995):85-87. Infoseek's subscription model was also reported in Margot Williams, "Getting Around the World Wide Web With the Help of a 'Search Engine'" Washington Post, June 26 1995, F19.

⁴³ Ronald Rosenberg, "Godsent - and a Threat" Boston Globe, June 30 1999, F4.

search specialists on the Web. Overture's technology, introduced in 1998, allowed firms to bid for specific keywords. Selling advertising on keywords wasn't new, but Overture wasn't just showing an advert together with the results. It was selling top place in the search results themselves. Its other innovation was to use an automated auction system, so that prices fluctuated constantly with supply and demand. This prompted protests from some Internet users, who felt that paid search results were a breach of the trust they placed in search engines to provide accurate answers.⁴⁴ Overture's creator, Bill Gross, defended its honor by arguing that a company willing to pay to be listed was bound to be relevant to the query. This, he believed, made Overture's results more useful than those of traditional search engines.⁴⁵

Overture's other key innovation was what has been called the "pay-per-click" model for Web advertising. By the end of the 1990s users had largely stopped clicking on traditional banner advertisements, thus deterring advertisers from paying merely to have their advertisement displayed. With pay-per-click advertising, the advertiser paid only when somebody clicked on the link to visit their website. This innovation revitalized the economics of internet advertising in general, and search advertising in particular.

The third and final possible source of revenue for search technology companies was to sell their technology to other companies. This could be offered as a service (the model of Inktomi and Overture) or by selling search engines as software packages. Big websites generally

⁴⁴ Charles Cooper, Perspective: Paid Search? It Stinks (News.com, 2006 [cited September 16 2006]); available from http://news.com.com/2102-1071_3-281615.htm.

⁴⁵ Overture's initial success was reported in Bob Tedeschi, "Striving to Top the Search Lists" New York Times, 10 December 2001, C7. Overture's story is told in John Battelle, The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed Our Culture (New York: Portfolio, 2005), 104-21.

offered search capabilities to help users navigate through their depths. For this market, Excite provided Excite for Web Servers and Alta Vista offered an intranet version of its search engine, with licenses starting at sixteen thousand dollars.⁴⁶ Indeed, Excite was originally created as a search product for website owners, and was launched as an online service only after a change of business strategy.⁴⁷ Microsoft bundled Microsoft Index Server with its Web server to handle website searches.⁴⁸

But as Internet use spread rapidly during the mid-1990s, many companies began to use the same technologies to create "intranets" full of private information.⁴⁹ These used Web servers and browsers to publish documents internally, creating a market for the use of Internet search technology. For obvious reasons, companies could not rely on public search engines to index their confidential information. A new wave of text searching companies appeared, focused on providing search engine capabilities for intranets. One leader in this field, Open Text, was founded in 1991 to commercialize search technology developed by Waterloo University during

⁴⁶ Pricing for the intranet AltaVista is reported in Eric Convey, "DEC Unveils Corporate AltaVista" New York Times, September 19 1996, 35.

⁴⁷ Flynn, "Making Searches Easier In the Web's Sea of Data".

⁴⁸ A description of some of the search packages available to Website designers is given in Jon Udell, "Search Again", Byte, January 1997, 123-4, 26.

⁴⁹ The market for corporate record indexing and search systems is actually much older than that for Web search systems. IBM offered a software package called STAIRS, developed to manage the mountain of documents IBM's legal team gathered for its defense against the federal antitrust suit launched in 1969. Indeed, lawyers provided an important market for this technology. Most organizations still kept their unstructured documents on paper, and even electronic documents had to be specially tagged and loaded into the index.

the creation of the second edition of the *Oxford English Dictionary*.⁵⁰ Early in the development of the Web it was the operator of a popular public search engine (used at one point by Yahoo to supplement its own directory). Indeed, a 1995 roundup of search tools reported that Open Text had “shot to the forefront of WWW databases with powerful search and display features” and the largest of all Web indexes.⁵¹ Despite this strong start, Open Text soon withdrew from the Web search field to focus on systems for searching internal corporate documents.

Web Portals

During the late 1990s the major Web search companies were all seeking to achieve rapid growth by broadening their operations. As one 1997 report put it, “Analysts have been hounding search companies to break away from the advertising-only revenue model and into a more diversified business.”⁵² As public companies they were vulnerable to such pressure. Despite a conspicuous lack of revenues, Excite, Infoseek and Lycos had all staged initial public offerings in 1996, bringing rapid rewards to their original investors. Ask Jeeves, Overture (or GoTo.com as it was then known) and Inktomi followed suit later in the dot com boom. Excite used some of the money to acquire other navigation services, including Webcrawler and the once-popular Magellan directory. AltaVista was the odd one out. The business was brought and sold several times, but never managed to stage the initial public offering for which it was being groomed.

⁵⁰ Gaston Gonnet, Oral History Interview with Thomas Haigh, 16-18 March 2005. To be published in the oral history collection of the Charles Babbage Institute, University of Minnesota).

⁵¹ Martin P Courtois, William M Baer, and M Stark, "Cool Tools for Searching the Web", Online 19, no. 2 (November/December 1995):14-27.

⁵² Cooper, Perspective: Paid Search? It Stinks .

Along with the rest of DEC it was purchased by Compaq in 1998, and eventually disposed of to Overture.⁵³

During the late 1990s the leading Internet search services Excite, Lycos, and AltaVista followed the same business strategy, which in retrospect was clearly flawed. Like other advertising supported sites they were trying to achieve what was then called “stickiness,” meaning that customers would visit often and remain within the site for a long time, clicking from one item to another. The problem with building a business around a search service was that a good search engine would come up results very quickly. Users would arrive, enter a search term and click on one of the results. At that point they would leave the site, and might not return until the next time they needed to look something up. During this process they would read very little advertising. Search company strategists also worried that it was hard to differentiate search engines or build brand loyalty, since their results were likely to be pretty interchangeable. Even if search results could be improved, doing so would just send visitors on their way even faster. The solution, it seemed, was to make search into just one feature among many and to remake their websites as Web portals.

The portal concept, which first appeared around 1997, suddenly became ubiquitous in 1998. The first use of the term Web portal in a major US newspaper was by the *New York Times* in an October 1997 article which stated that “Yahoo has been the most successful so far in establish an identity as a hip, if quirky, Web portal.”⁵⁴ The term fitted with Yahoo’s

⁵³ AltaVista’s sorry history is recounted in Jim Hu, [AltaVista: In Search of a Turning point](http://news.com.com/2102-1023_3-270869.html) (News.com, July 31 2001 [cited September 12 2005]); available from http://news.com.com/2102-1023_3-270869.html.

⁵⁴ Tim Race, "Infoseek Revises Its Internet Search Engine" [New York Times](#), October 20 1997, D.15 The term “portal” had occasionally been used in a different sense earlier the 1990s as an alternative to the more common

ostentatiously wacky corporate culture, complementing irreverent job titles such as the “Chief Yahoo” tag given to one of its cofounders. Portal, literally just another word for an entrance gate, suggested a device through which one might travel to distant lands in an instant. Magic portals were a cliché of fantasy fiction, and featured prominently in the plotline of the Dungeons & Dragons children’s cartoon of the 1980s as well as in various spells and magical items within the game it was based on.

Portals aimed to provide something akin to AOL’s well integrated user experience for people confused by the variety and complexity of the Web. They combined search capabilities with real-time chat functions, online games, weather, news, stock information, television listings, shopping opportunities, discussion areas, maps, and other services. All of them made frantically licensed technologies and acquired businesses to merge new features into their portals. Portals offered customization capabilities, so that users could adjust the topics displayed on the main page to suit their own interests. The idea was that with all these different services available in one place, users would set the portal as their home page, visit frequently, and spend a lot of time skipping happily from one area to another while reading advertisement after advertisement.⁵⁵

The conventional wisdom was that portals represented the most valuable part of the entire Internet and would be the launching point for every exploration of the Web. And, it must be admitted, early Web users really did rely on search and portal sites. According to a list published

“gateway” to describe an interconnection point between two networks for the exchange of email or, in a few cases, to allow users of proprietary online services like AOL to browse the Web.

⁵⁵ For a contemporary description of the excitement surrounding the portal concept, see Rajiv Chandrasekaran, "One-Stop Surfing; Today's Hot Web Concept Is 'Portals.' Tomorrow, Who Knows?" Washington Post, October 11 1998, H.01.

in June 1996 of the most visited websites the twelve most popular included six search specialists: Yahoo, WebCrawler, AltaVista, Lycos, InfoSeek, and Excite. The other sites (AOL, Netscape, Prodigy, GNN, CompuServe and MSN) were all Internet access or technology companies whose websites functioned as portals. So-called destination sites such as Amazon, EBay, the *New York Times*, and CNN had yet to break into the top twenty-five.⁵⁶

The portal field was increasingly crowded, as online services and browser companies realized that they could turn their popular websites into Web portals. In 1997 Netscape, which then enjoyed a steady stream of visitors to its website, decided to remake its homepage as a portal called Netcenter. Netscape.com had long been one of the most popular sites on the Internet. As a supplier of browser software Netscape had a particular advantage. Each Web browser opens a particular Web page automatically when it is started, known as the home page.⁵⁷ It is not obvious how to change this, and people do so slowly or not at all (according to one 2001 report, 59% of Web users still had the initial setting).⁵⁸ Realizing that all this Web traffic was more attractive to investors than the company's struggling attempts to sell its Web browsers, the

⁵⁶ Anonymous, "Top 25 Web Sites" USA Today, June 28 1996, 4D.

⁵⁷ The idea of a home page went back to Tim Berners-Lee and the origin of the Web. Berners-Lee had imagined that browsers would include integrated editing capabilities, so that each user would have a personal home page that he or she could edit to include links to pages of interest as well as public messages for other visitors. (Something rather like a blog). This explains the dual meaning of the term home page as both "the default start page for someone's browser" and "the main page holding information about a person or company." James Gillies and Robert Cailliau, How the Web Was Born: The Story of the World Wide Web (Oxford: Oxford University Press, 2000), 193-4.

⁵⁸ David Lake, "Microsoft Dominates the World Wide Web" The Industry Standard, August 23 2001.

firm created a new division to build up Netcenter into a fully fledged portal.⁵⁹ When Netscape Navigator slipped below 50% market share for the first time since its launch a company spokesman shrugged off the news, saying “It’s not the browser wars anymore, it’s the portal wars... We are getting huge growth in Netcenter users, and that’s what counts...”⁶⁰

Both AOL and its new rival MSN (the MicroSoft Network) realized that they could complement the proprietary material offered to their online service providers with public websites holding the typical portal content such as news and email. Microsoft launched msn.com as a public portal in 1998, consolidating material created in its existing websites such as Hotmail, MSNBC, and Expedia.⁶¹ Microsoft made msn.com the default home page for its increasingly dominant Internet Explorer Web browser and later unleashed an additional piece of software, MSN Explorer, intended to give Web surfers a nearly packaged client similar to that offered by AOL to its users.⁶²

The line between Web publisher and portal began to blur, as search engines, browser firms and online services morphed into portals and began to fill their site with information and services, as well as links to other websites,. Specialist Web publishing companies such as CNET began to call themselves portals as they added new websites (which were sometimes, on the model of television, called channels) to their networks. This followed the model attempted by

⁵⁹Suzanne Galante, Netscape Outlines Web Strategy (News.com, March 25 1998 [cited September 11 2006]); available from <http://news.com.com/2100-1001-209497.html>.

⁶⁰ Paul Festa, Study: Netscape Share Below 50% (News.com, September 29 1998 [cited September 15 2006]); available from http://news.com.com/2102-1023_3-216043.html?tag=st.util.print.

⁶¹ Saul Hansell, "Where Does Microsoft Want You To Go Today. The New Strategy: Keep Web Surfers Busy With a Series Of MSN Sites" New York Times, November 16 1998, C.1.

⁶² David Pogue, "The Web Gets a New Dashboard" New York Times, October 26 2000, G.1.

media giant Time Warner early in the history of the Web. In 1994, it became the first major media company to invest heavily in the Web when it created Pathfinder.com, a website full of electronic content from all Time Warner's subsidiaries.⁶³ This single domain was shared between operations as diverse as *Time*, *Fortune*, special topic Web publication such as O.J. Central, Warner Brothers movie *Batman Forever*, and Warner Books publication *The Bridges of Madison County*. The site eventually included discussion and chat areas, free email accounts, and a personalized news service. Individual Time Warner publications were forbidden from using their own domain names to build independent websites. The attempt to rival AOL by creating a huge mass of professionally produced media content had an obvious appeal to the managers of an unwieldy conglomerate whose entire existence was predicated on a faith in the existence of "synergies" between unrelated businesses. But the practical benefits of this integration were limited. And burying popular brands such as *Fortune* within Pathfinder was perverse. Although much visited domains such as www.time.com merely forwarded users to the appropriate area of Pathfinder, users continued to identify with the individual publication brands rather than their corporate parent. A company spokesperson reportedly said that 98% of visitors to the site were forwarded from other domains rather than going directly to Pathfinder.⁶⁴ As Web users proved unwilling to pay subscriptions for online publications Pathfinder never came close to covering its

⁶³ A lively, well informed account of Pathfinder's creation and early life is given in Wolff, [Burn Rate: How I Survived the Gold Rush Years on the Internet](#), 109-38. The trouble history of Time Warner's involvement with the Web is explored in Frank Rose, "Reminder to Steve Case: Confiscate the Long Knives", [Wired Magazine](#), September 2000, 156-72.

⁶⁴ Jim Hu, [Time Warner to Shutter Pathfinder](#) (CNET News.com, April 26 1999 [cited August 29 2006]); available from <http://news.com.com/2100-1023-224939.html>.

costs through advertising. In early 1999, the Pathfinder strategy was officially abandoned and Time Warner's surviving websites were freed to go their own separate ways.

Despite Pathfinder's problems, the idea of building a self-contained Web world around the different media brands owned by one huge corporation still appealed in late 1998 when Disney and Infoseek (then the eighth-busiest site on the Web) announced the merger of their Web efforts to create the Go.com Web portal.⁶⁵ Disney's media holding included the ABC television network and ESPN as well as its own brand of sugary family entertainment. Disney launched a heavy campaign to plaster Go.com promotions across its television channels, books, theme parks, and other outposts of its corporate empire. While intended to compete with Yahoo and the other major portals, Go.com would give preferential treatment to search results and links featuring the firm's own offerings. Disney aimed to spin off Go as a separate company, offering a special class of stock representing its operations to take advantage of the enormous sums of money that investors were funneling into Internet stocks. Unfortunately, Web surfers continued to think of, and visit, Disney's online sites such as ESPN.com and ABC.com as separate entities. As an unknown brand, Go.com struggled to make an impression. It lost more than a billion dollars in 1999, most of it spent to acquire Infoseek.⁶⁶

. In the late 1990s, the portal companies looked up to AOL as a model not only for its click integration of different services but also for its clout with advertisers. AOL came late to advertising. In 1996, it was facing stiff competition from Internet services offering unlimited

⁶⁵ Roger Taylor, "Disney and InfoSeek to Launch New Web Portal" Financial Times, December 14 1998, 23.

⁶⁶ Keith L. Alexander, "Despite Setbacks, Go's Chairman Sees Green Light Ahead" USA Today, December 27 1999, 6B.

usage for a flat monthly fee of around \$20, while it continued to charge several dollars per hour. When AOL followed suit, this irreversibly changed the economics of its business. That same year it also acquired CompuServe, cementing its dominance of the industry. Previously, AOL had made money out of keeping users online longer, but now it faced ruin if they spent the whole day online. Its modem pools were massively overloaded, and for a while it was jokingly known as “America On Hold.” In October 1996 the firm reported a \$353 million quarterly loss, reflecting previous customer acquisition costs it could no longer justify as an investment to produce future usage revenue. Instead, AOL increasingly relied on advertising to make up for the costs incurred as its users spent more time online. AOL users spent a great deal of time in AOL’s own proprietary areas (particularly its chat rooms) and on its website, proving plenty of opportunities to display advertisements to them.

By 1999, 16% of AOL’s revenue was coming from advertising.⁶⁷ It was selling more online advertising than anyone, but it was also extracting huge payments from “partners” to feature their products prominently on its system. AOL had a dominant position as the largest Internet service provider. During the late 1990s, thousands of well funded start-up firms were desperate to attract visitors to their sites. The conventional wisdom was that the first popular site in a particular market, such as online pet food, would enjoy a huge advantage. Companies could therefore go public or sell themselves for hundreds of millions of dollars purely on the basis of a healthy stream of visitors to their websites, regardless of their financial position. As a result, startups were taking in millions of dollars from venture capitalists and throwing much of this money at AOL with the idea that a prominent spot would be a quick way of brining visitors and

⁶⁷ Danny Sullivan, comScore Media Metrix Search Engine Ratings (SearchEngineWatch.com, August 21 2006 [cited August 27 2006]); available from <http://searchenginewatch.com/showPage.html?page=2156431>.

boosting their profile in the media. The trend started in 1996 when short-lived telephone firm Tel-Save offered a cash payment of \$100 million to be the exclusive long distance telephone provider advertising on AOL. Many other deals followed. Such a deal, usually talked up as a “strategic partnership” might give the company enough credibility to make its own initial public offering.⁶⁸

Following the model of AOL, the Web portals were able to bring in money by signing partnership deals with other Internet companies. How much sense this made is unclear. Even after AOL connected its service to the Internet, its users continued to rely on special AOL software that gave pride of place to AOL’s own offerings and sponsors. In contrast, portals such as Excite and Lycos had no special grip on their users, meaning that being the “exclusive” provider of online books on Lycos (as BarnesandNoble.com did) or the exclusive CD retailer on Yahoo (as CDNOW did) was unlikely to do much to justify the millions of dollars it cost.⁶⁹

During the boom years, deals of this kind appeared to confirm the strategic power of portals as the hubs of the Internet. They continued at a dizzying rate. Between August and October 1999, Lycos (now billing itself as “the world’s largest online community”) claimed to have created “the Internet's first true full service ecommerce portal with the launch of LYCOShop, the Web's most complete integrated shopping destination,” to have launched more than a dozen new “localized” versions of the portal for different countries, to have formed or extended “strategic alliances” with IBM, Fidelity, American Greeting Cards, and AOL, and to

⁶⁸ The importance of these deals to AOL is explained in Nina Munk, Fools Rush In: Steve Case, Jerry Levin, and the Unmaking of AOL Time Warner (New York: HarperCollins, 2004), 100-08.

⁶⁹ Suzanne Galante and Paul Festa, Lycos Up on Smaller Losses (News.com, August 27 1997 [cited 2006 August 20]); available from http://news.com.com/2100-1001_3-202723.html.

have acquired or invested in half a dozen smaller firms including a maker of MP3 audio players. None of these deals amounted to much in the end, but they certainly provided the impression of a dynamic and ambitious firm.⁷⁰

The major portals were joined by a wealth of specialist portals. Portals sprang up in China, India, Europe and elsewhere. Although the Web itself was world wide, details such as shipping costs and customs regulations meant that most Internet businesses functioned within national boundaries. Even purely information sites, such as online magazines, were of interest only to people who could read the language involved and had some familiarity with the culture of the country involved. National portals made a lot of sense. By the same logic, some websites (particularly those operated by local newspapers) began to sell themselves as regional portals, integrating news, weather, business listings, cultural information, restaurant reviews and the like for a particular area. The *Boston Globe's* site, Boston.com, was one of the first and most successful of these, including material from local television and radio stations and magazines as well as the paper's own stories.⁷¹ Newspapers saw the regional portal model as a means of safeguarding their local stranglehold on classified advertising, traditionally a major source of revenue, from predation by specialist online marketplaces.

Meanwhile, so-called industry portals were set up as online hybrids of trade publications and business-to-business marketplaces. In the mid-1990s many assumed that the Web would

⁷⁰ Anonymous, [Lycos Reports 126% Increase in Revenues](#) (findarticles.com, November 22 1999 [cited September 04 2006]); available from http://www.findarticles.com/p/articles/mi_m0WUB/is_1999_Nov_22/ai_57758941/print

⁷¹ David Carlson, "Media Giants Create Web Gateways", [American Journalism Review](#), September 1999, 88.

replace traditional relationships between industrial customers and suppliers with specialist commodity markets, in which spot prices for all kinds of goods and services fluctuated constantly as electronic bids were placed and met. Few companies were rushing to adopt this new model for their purchasing needs, and so the more ambitious industry portals collapsed rapidly. For example, three big metal industry portals, Aluminium.com, MetalSpectrum and MetalSite all folded within a two week period in June 2001.⁷² Another flurry of doomed firms tried to set themselves up as “eGovernment Portals” to bring citizens together with government institutions. Among these was govWorks, featured in the memorable documentary film *Startup.com*.

Large organizations set up portals for their employees and customers. Portals were intended to provide a single, easy to use entry point for all the different offices, services and sources of information spread out among the constituent parts of large bureaucratic organizations such as city governments, universities and major corporations. Perhaps the most ambitious in scope were the national and local government portals announced by politicians eager to show that their governments were accessible, forward looking and in touch with their citizens. The United States offered not just “FirstGov.gov, The U.S. Government’s Official Web Portal” but also the cartoon-like “FirstGov for Kids,” complete with a Web “treasure hunt” to reward children for finding the correct answers to questions such as “What federal agency provides citizens with information about the Air Quality Index?”⁷³ So popular were portals that a new industry grew up to supply portal software for use by corporate customers. Portal packages were

⁷² Tom Stundza, "Dot.coms Implode, Survivors Seek New Focus", *Purchasing*, August 09 2001, 16B10.

⁷³ US General Services Administration, *FirstGov for Kids* (Spring 2003 [cited August 26 2006]); available from <http://www.kids.gov/activity.htm>.

sold by major corporate information technology providers such as Sun Microsystems, IBM, and Oracle, as well as by specialists such as Plumtree and Hummingbird Communications, and by Yahoo itself. Over time, many portal packages evolved to incorporate the capabilities of content management systems, generating Web content dynamically from databases and templates.

Enthusiasm for portals grew along with the bubble in dot com stocks. Excite, Lycos, Netscape and Yahoo were much better known as stocks than as companies, and so their success was gauged on their soaring share price rather than their actual prospects. Stock tips were everywhere in the late 1990s, as financial television channels grew, investment clubs thrived and newspapers profiled the sensational initial public offering of the day. YHOO was perhaps the greatest of the pure Internet stocks. Adjusted for splits, it rose from an offering price of \$24.50 in 1996 to a peak of \$2,850 at its all time high in January 2000. The other portal firms enjoyed similar success. In May 2000, Terra Networks offered stock then worth a staggering \$12.5 billion in a successful bid for Lycos, then the third most visited portal site. But by then the dot com crash had already begun.

The Crash

Between March 2000 and October 2002 the NASDAQ composite, a measure of share prices on the high-technology oriented NASDAQ exchange, fell from 5,047 to 1,114. Internet stocks fared far worse. Portals saw their advertising revenues dry up, as most online advertising came from other online businesses. The plunge in the value of existing Internet stocks meant that no more Internet stock offerings were possible, which meant that venture capitalists were no longer handing millions to doomed or speculative startups, which meant that there was no flow of easy money to portals from startup companies rushing to buy visitors to their websites. The sudden drop in online advertising was a particular challenge to the portals, whose existence was

premised on the idea that it would rise rapidly for many years to come. When the flow of new money ceased they faced a crisis.

Often this was terminal. Excite, for example, had merged with high-speed internet pioneer @Home and squandered its cash on dubious acquisitions, such as the payment of \$780 million for an online greeting card company.⁷⁴ It ran out of money soon afterwards and was liquidated. Go.com, in contrast, was clearly doomed even before the crash having failed to attract advertisers or visitors. Disney tried to salvage the portal by refocusing it on entertainment and leisure, but it continued to hemorrhage money and slip down in the Internet rankings.⁷⁵ In early 2001 Disney closed the division, shut down the portal and disposed of its assets.⁷⁶ The once-mighty Lycos withered away more slowly. In 2004 what remained of Lycos was sold again, this time for less than 1 percent of the price that Terra Networks had paid four years earlier. Its original business was essentially destroyed, though efforts are now under way to revive the brand.⁷⁷

AOL began to fall apart in 2001, just a few months after using its stock to buy Time Warner to form AOL Time Warner. Time Warner was the world's biggest media company, but the failure of Pathfinder and a number of other Internet initiatives had left its stock stagnant.

⁷⁴ Ben Heskett and Jeff Peline, Why Excite@Home failed: A Postmortem (News.com, 28 September 2001 [cited 2 October 2001]); available from <http://news.cnet.com/news/0-1014-201-7340505-0.html>.

⁷⁵ Verne Kopytroff, "Disney's Go.com Changes Direction with Web Site" San Francisco Chronicle, September 15 2000, B1.

⁷⁶ Saul Hansell, "Disney, in Retreat from Internet, to Abandon Go.com Portal Site" New York Times, January 30 2001, C.1.

⁷⁷ David Shabelman, Can Lycos Learn New Tricks (News.com, September 8 2006 [cited September 10 2006]); available from http://news.com.com/2102-1032_3-6113790.html.

AOL, on the other hand, had an enormously high share price because (inasmuch as any rational reason can be provided) it appeared to have been growing rapidly for a long time and was expected to continue to do so indefinitely. Facing pressure to show continued growth, AOL began to claim revenues based on exchanges of advertisements with other companies in which no money really changed hands.⁷⁸ (It eventually agreed to pay \$510 million to settle an investigation by the Securities and Exchange Commission into these practices).⁷⁹ But such tricks could work only for a short time, and within eighteen months of the merger almost all the senior AOL executives had been purged from the firm. Far from being the engine of growth for the merged company, AOL was dragging down the performance of successful parts of the conglomerate, such as the HBO subscription television channel. Its base of loyal dial-up subscribers saved AOL from total annihilation, but its diminishment was clearly signaled in 2003 when the AOL Time Warner board decreed that, hence forth, plain old Time Warner would be a better name for the firm.

Its main competitor, MSN, never lived up to expectations. By 2001 it had become the second most visited portal in the United States (after Yahoo), though by that point the portal business was collapsing.⁸⁰ It did become the second biggest provider of dial-up Internet access in the United States by the end of the 1990s, and according to Web traffic metering specialist

⁷⁸ Alec Klein, "Unconventional Transactions Boosted Sales" Washington Post, July 18 2002, A.01.

⁷⁹ David A Vise, "Time Warner Settles AOL Cases For \$510 Million " Washington Post, December 16 2004, A.01.

⁸⁰ Jane Black, Is That MSN Breathing Down AOL's Neck? (BusinessWeek Online, May 8 2001 [cited September 12 2006]); available from http://www.businessweek.com/print/technology/content/may2001/tc2001058_449.htm?chan=tc.

Alex.com the MSN website is currently the fourth most visited in the country. For any other company, those results would have reflected a triumph. But there was really little to cheer about, given all the money that Microsoft pumped into MSN, the \$400 rebates it handed out liberally to win three year subscriptions from new computer purchasers, the special placement it received within Windows, and its huge advantage as the default home and search pages of Internet Explorer. Fortunately for Microsoft, its Windows and Office monopolies continued to generate money much faster than the rest of the company could spend it.

Beyond the immediate crisis affecting all firms reliant on online advertising, the portal industry faced a particular problem: as the Internet matured, most people did not really want or need portals. As users grew accustomed to the Web they created their own lists of links and bookmarks, visiting different sites for different kinds of information. People no longer needed a portal to tell them that they could buy a book from Amazon, read the *New York Times* online, or purchase a plane ticket from Expedia. While a search engine was still a good place to go to investigate a new topic, the rush to remake search engines as portals had drawn resources away from the development of their search technologies and filled their front pages with slow-loading and distracting clutter. Furthermore, as portals sought to provide every possible service they had no real way to differentiate themselves from their competitors. As *Washington Post* reporter Rob Pegorano noted in 2000, “These sites are all unique in pretty much the same way. Most of their content even comes from the same third-party sources--news from the Associated Press and Reuters, forecasts from the Weather Channel, shortcuts to buy books through Amazon.com, and so on.”⁸¹ Personalization features were supposed to be the compelling benefit of portals, but few users even bothered to configure their preferences, and still fewer kept them updated.⁸²

⁸¹ Rob Pegorano, "Any Portal in a Storm" *Washington Post*, November 3 2000, E.01.

Of the independent would-be portal sites, only Yahoo survived the crash. Its luster as a stock vanished quickly, as YHOO lost 97% of its early 2000 peak value in less than eighteen months.⁸³ But Yahoo the business retrenched, downsized, brought in a new CEO and redoubled its efforts to drum up advertising revenues and created premium services that users were willing to pay for. Since 2002 Yahoo has been consistently profitable, dominating what was left of the portal field and enjoying some success with features such as online video sharing, social networking, auctions and job listings.

Google and the Resurgence of Search

By the end of the 1990s the Internet search pioneers Lycos, Infoseek, Excite and AltaVista were no longer particularly interested in the search business. Their technical, financial and managerial resources were aimed squarely at the Web portal business. Their Web search capabilities were just one of many items, packed into home pages full of eye-grabbing attractions. In 2002, a Yahoo spokesman explained that his firm was “first and foremost a media company” and that “[s]earch and directory is an increasingly small part of what we do.”⁸⁴

Starved of resources, the search engines were actually becoming less useful. As the Web got bigger and bigger, the issue was no longer finding a sufficient number of pages containing the search term but ranking the results so that useful pages came first. Studies found that only about 20 percent of users would go beyond the first page of search results in search of relevant pages.⁸⁵ But search engines might find many thousands of matching pages, and used quite simple

⁸² Black, Is That MSN Breathing Down AOL's Neck? .

⁸³ Bill Breen, "She's Helping Yahoo Act Normal", Fast Company.com, April 2003, 92.

⁸⁴ Ben Hammersley, "Is Yahoo Losing the Plot?" The Guardian, May 2 2002.

⁸⁵ Amanda Spink and Bernard J. Jansen, Web Search: Public Searching of the Web Kluwer, 2004), 104-17.

methods to rank results, looking particularly at the number of times the keyword appeared. As website operators became more sophisticated, they found it easy to manipulate these results. For example, some search engines could be manipulated into displaying a page in response to a query on a popular term such as “antivirus software” just by inserting that term into the raw code for the page thousands of times in such a manner that it would never actually be displayed. The top results for many queries were hijacked in this manner by operators of pornographic websites. A small industry grew up promoting these techniques to website operators as a cost-effective alternative to paid advertising. Its apologists dubbed the practice “search engine optimization”, but critics preferred the less flattering “search engine spamming.”⁸⁶

It seemed that search engines had reached the limits of their capabilities. In 1999 Danny Sullivan, whose blog-like site Search Engine Watch is a major source of search engine news and comment, boldly announced that “This was the year that humans won. [Earlier] you had one major search service, Yahoo, that used human beings to categorize sites while the others were trying to use technology to do the same thing. But now with six out of the top 10 services, the main results you get will be by people.”⁸⁷ He was apparently referring to the incorporation of Open Directory results by the major portals. Sullivan was not alone in this belief. The next year Chris Sherman, a search engine consultant, insisted that the project was “leading a resurgence of

⁸⁶ For an early report on search engine spam and efforts to combat it, see Jeff Evans, “Power Searching” Toronto Star, December 9 1999, 1.

⁸⁷ Paul Festa, Web Search Results Still Have Human Touch (News.com, December 27 1999 [cited September 17 2006]); available from <http://news.com.com/2100-1023-234893.html>.

human-compiled Web directories, toppling spider-complied search engines from their dominant positions as principal gateway to the Internet.”⁸⁸

Sullivan and Sherman were quite wrong. As the established search companies marched in unison over the cliff in pursuit of portal status they had left behind a niche for a company focused on providing the best search experience to users. A new search company, Google, seized this opportunity and proved that Web search was one of the Internet’s most profitable business opportunities. Like Excite and Yahoo a few years before, Google began as the personal project of some Stanford University computer science students. Larry Page and Sergey Brin launched their service as google.stanford.com. They made the most of Stanford’s cultural and technical resources, using huge quantities to network bandwidth and storage space to build a fully functional version of the service with a massive database before seeking funding for their idea. In 1998 they founded Google, Inc. to turn the service into a business. Drawing on Stanford’s well established connections with Silicon Valley businesses the pair won initial investments from various technology industry entrepreneurs, followed in 1999 by a generous injection of \$25 million, of which most came from the valley’s two most storied venture capital firms: Kleiner Perkin and Sequoia Capital.

But while Google enjoyed privileged access to the sea of easy money floating around Silicon Valley toward the end of the boom years, its founders resisted many of the ideas imposed on other search firms by the experienced managers brought in to steer their evolution. In contrast with the ever-more crowded front pages of the portals, Google’s Web page consisted of nothing more than a white space in which floated a simple logo, a box to type search terms into, and a button labeled “Google Search.” The results pages were similarly Spartan. With no

⁸⁸ Sherman, "Humans Do It Better: Inside the Open Directory Project".

advertisements, Google pages loaded almost instantly even over a dial-up connection. Indeed, its founders were initially quite dismissive of the idea of advertising:

The goals of the advertising business model do not always correspond to providing quality search to users... [T]he better the search engine is, the fewer advertisements will be needed for the consumer to find what they want. This of course erodes the advertising supported business model of the existing search engines... [T]he issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm.⁸⁹

Behind this simple interface lay a search engine of exceptional power. Google's coverage of the Web was soon unsurpassed, as its crawler (dubbed "the googlebot") inched its way into the mustier fringes of the Web. But its biggest attraction was the consistently high relevance of its results.

Without human intervention, Google somehow pushed high quality, relevant sites toward the top of its search results. This has been attributed to its much-discussed "PageRank" algorithm, patented by Stanford. Unlike the crude techniques used by early search engines, this method consistently put the most important and widely linked to sites on a particular topic within the first page or two of its search results. PageRank looked not just at the page itself, but also scoured its database to assign the page a rank according to the links created to it from other sites. This idea was inspired by the long established practice of ranking the importance of scientific papers according to the numbers of citations they received. (Citations were indexed by the

⁸⁹ Sergey Brin and Lawrence Page, "The Anatomy of a Large Scale Hypertext Web Search Engine", in Proceedings of the seventh international conference on World Wide Web (1998), 107-17. This article includes a good description of the initial structure and operation of Google in its Stanford days.

Institute for Scientific Information, giving raw data for researchers in the field of bibliometrics and for the institute's own annual rankings of science's greatest hits.) Google extended the idea so that links from highly ranked sites were themselves accorded more weight than links from poorly ranked ones. Google also incorporated some established information retrieval principles, considering the print size of terms, their closeness to each other and their position on the page in determining ranking. While this method proved to have its own vulnerabilities to exploitation, it continues to produce high-quality results.

Google's founders tried to license their invention to AltaVista, Excite, Infoseek, and Yahoo before making their commitment to commercialize it themselves. All these firms turned it down – search was no longer a priority for them. Google's advantage came not just from its algorithm but also from its unwavering focus on providing an effective search service. Its staff constantly tweaked the internals and user interface of its search engine, making little adjustments and adding features. As the service became more popular, they worked hard to eliminate search spam and keep results relevant.

Google's success hinged on technical feats in operating systems and parallel computing as well as the original cleverness of its ranking algorithm. Google searched more of the Web than its competitors, and gave more useful answers, faster, and to more users. This was possible only by applying unprecedented volumes of computer power to the problem. Google accomplished this without running out of money only by finding innovative ways to combine many thousands (and eventually hundreds of thousands) of cheap computers based on commodity personal computer hardware running the free Linux operating system. Its competitors

often relied on small numbers of expensive, multi-processor systems using proprietary hardware and operating systems from firms like IBM and Sun.⁹⁰

Having created the Web's best and fastest-growing search service, Google still needed a way to make money. In 2000 it began selling advertisements, with an approach it called AdWords. Like Overture, Google accepted bids from potential advertisers to have their ads appear when users searched on particular terms. It also copied Overture's "click-per-pay" model. However, Google realized that users disliked large, distracting advertisements and would not trust a search service in which the top results appeared only because users had paid for them. Google instead presented a single paid result, clearly labeled as a "sponsored link," and a handful of simple advertisements, each consisting of three short lines of text, grouped in a separate part of the screen (initially at the top of the screen, and later in a column at the right). Google added a twist of its own to Overture's model, by factoring in how often advertisements were clicked on as well as the amount bid by the advertiser in deciding how to order them. Even the highest bids brought little revenue to Google unless people were prepared to click on the adverts once they were shown, so this tweak simultaneously improved the relevance of the adverts displayed and boosted profits.⁹¹ 2000 was not a good year for the online advertising industry, but Google's increasing popularity and ability to present users with advertisements directly related to whatever they were researching soon allowed it to charge a premium. It turned out that sticking to search could be exceptionally lucrative. Google's advertising is among the world's least obtrusive but

⁹⁰ Mitch Wagner, Google Bets The Farm On Linux (InternetWeek, June 1 2000 [cited September 20 2006]); available from <http://internetweek.cmp.com/lead/lead060100.htm>.

⁹¹ Google's adoption of advertising is discussed in David A Vise and Mark Malseed, The Google Story (New York: Delacorte, 2005), 89-102.

most profitable. Indeed, the very sparseness of the ads raised their value by decreasing the supply.⁹²

As well as its own popular search service, google.com, Google also won contracts to provide search services to many of the Web's most popular sites. It shared its advertising revenues with the sites concerned. By 2002 these included the MSN, AOL and Yahoo Web portals. Google's superior information retrieval capabilities and more efficient advertising technologies allowed it to edge out Overture and Inktomi for these crucial accounts. The portals continued to think of search as a peripheral feature best outsourced to a specialist firm. Google was happy to support this misapprehension. A 2001 article stated that "Google vehemently denies that it has designs on its portal customers' turf," and quoted a company spokesman as saying that "we have 130 customers... they don't feel we're competing with them, and we're comfortable with that model."⁹³

Google's success even revived some of the old motifs of the .com era. Google's leaders pride themselves on its distinctive corporate culture, including its aggressively uncomplicated motto "Don't be evil," the prominent display of pianos, video games, and gourmet food in its offices, and a corporate mandate to make work fun. In 2004 it went public, and while the stock broke with tradition by failing to rise on the first day of trading, its value quadrupled over the next eighteen months. Page and Brin achieved prominent spots on the *Fortune* list of the world's most wealthy people while still in their early 30s, and stock options made its early employees

⁹² A good summary of the current state of Internet advertising was given in Anonymous, "The Ultimate Marketing Machine", *The Economist*, July 8 2006.

⁹³ Paul Festa, *Is Google Ogling Yahoo's crown?* (News.com, 2001 [cited September 25 2006]); available from http://news.com.com/2102-1023_3-255601.html.

millionaires overnight. The determination of Google's founders to impose their culture of technical tinkering on the world of business extended even to the initial public offering, which was conducted using a unique auction process to set the share price and structured in such a way as to leave control of the company in their hands via a special class of stock with superior voting rights.

THE WEB NAVIGATION BUSINESS TODAY

After the success of Google became apparent, Yahoo changed course to invest heavily in search technology (as did Microsoft, Amazon and many other firms). Yahoo brought Inktomi in 2002 and Overture in 2003. As Inktomi had already purchased what was left of AltaVista, that left Yahoo with a powerful array of search technologies. In 2004 it switched its main Web search over to its own technologies, marking the first time that Yahoo had operated its own public search engine rather than outsourcing the work to others.⁹⁴ This investment has failed to stem Google's rise. According to the comScore Media Metrix rankings of Internet search traffic, Google held a 44.1% share of the 6.5 billion Web searches in the United States during August 2006, versus 28.7% for Yahoo (now the owner of Overture) and 12.5% for Microsoft.⁹⁵ Other estimates give Google more than 60% of the search market.⁹⁶

⁹⁴ Jim Hu and Stefanie Olsen, Yahoo Dumps Google Search Technology (News.com, 2004 [cited September 17 2006]); available from http://news.com.com/2102-1024_3-5160710.htm.

⁹⁵ comScore Networks, Google Regains Some Ground from Previous Month's Share Decline with 0.4 Share Point Increase in August Rankings (comScore, 2006 [cited September 15 2006]); available from <http://www.comscore.com/press/release.asp?press=1006>.

⁹⁶ Catherine Holahan, "Yahoo's Lost Bid Doesn't Spell Doom", Business Week, August 9 2006.

Driven in large part by the success of search advertising, Internet advertising on Yahoo and other leading sites rebounded and has shown steady growth.⁹⁷ In 2005 online advertising revenues reached a new high of 12.5 billion dollars, according to the most widely accepted estimate. Of this moment, 43 percent came from search advertisements.⁹⁸ By 2006 Google was selling far more advertising than any other Internet firm, and had overtaken traditional media powerhouses such as the broadcast television networks and Gannett Co. (owner USA Today and more than a hundred other daily newspapers including, 23 television stations, and more than a thousand other periodicals).⁹⁹ It achieved this by displaying advertisements to the users most likely to click on them. According to one newspaper report, “For every page that Google shows, more than 100 computers evaluate more than a million variables to choose the advertisements in its database to display - and they do it in milliseconds”¹⁰⁰ These variables are said to include the time of day, the location of the user, and type the of internet connection, but not personal information. Google has finally delivered on the idea, much discussed during the .com era, that

⁹⁷ Ben Elgin, "Google and Yahoo! Rolling In It", Business Week, October 21 2005.

⁹⁸ PriceWaterhouseCoopers, IAB Internet Advertising Revenue Report: 2005 Full-Year Results (Internet Advertising Board, 2006 [cited 2006 September 8]); available from http://www.iab.net/resources/adrevenue/pdf/IAB_PwC_2005.pdf.

⁹⁹ Google reported revenue of \$2.69 billion for the third quarter of 2006, up 70% from the previous year. Sara Kehaulani Goo, "Surge in Profit Reflects Google's Widening Lead" Washington Post, October 20 2006, D.01. Gannett reported operating revenues of \$1.9 billion over the same period.

¹⁰⁰ Saul Hansell, "Google Wants to Dominate Madison Avenue Too" International Herald Tribune, October 30 2005. It seems unlikely that Google really evaluates a million variables – Hansell may mean “a million possible combinations of variables.”

Web advertising can be much more efficient than offline advertising because of the potential to target it more precisely.

Yahoo's purchase of Overture and its innovative search advertising technology underpinned its return to financial health. It so far remains less effective than Google in maximizing advertising revenues by displaying the adverts that users are most likely to click on, though by the end of 2006 it aimed to have pulled ahead of Google in this respect with a new systems incorporating information on users' demographics, query history and browsing habits.¹⁰¹ Yahoo's recent progress illustrates the absurdity of stock valuations during the dot com era. Yahoo reported earnings of close to two billion dollars for 2005, around thirty times higher than those for 1999. Yet its stock, which is by no means undervalued, has regained less than 45% of its peak value.¹⁰²

As well as selling advertisements on their own websites, Google and Yahoo have also become brokers of advertising to be displayed on smaller websites. Again this was not entirely new. Since the mid 1990s, advertising firms such as DoubleClick had sold advertisement space in bulk. They controlled which advertisements were shown where, and hosted the ads on their own servers, so that all an Internet publisher had to do was insert a link to the ad server in the appropriate place on their pages. But Google's AdSense system refined the concept, by offering

¹⁰¹ Elinor Mills, Yahoo: Our Ads are Better (News.com, May 18 2006 [cited September 20 2006]); available from http://news.com.com/2102-1024_3-6073504.html and Saul Hansell, "Yahoo Is Unleashing a New Way to Turn Ad Clicks Into Ka-Ching" New York Times, May 8 2006.

¹⁰² Yahoo's share price in the two years to September 2006 varied from \$24.91 to \$43.66. However, because the stock split 2:1 twice between those times, the \$475 peak closing value of January 03, 2000 corresponds to a price of \$118.75 for each of today's Yahoo shares.

the same bid system and pay-per-click model used on its own website.¹⁰³ Revenue is shared between Google and the operators of the websites where advertisements are displayed. This has helped to shift the economics of Web publishing back toward smaller amateur and semi-professional ventures. (It has also created a new and hugely profitable industry of sites holding nothing but Google advertising links, placed on attractive yet vacant domains such as clothes.com or on sites such as nytomes.com, yagoo.com or ebay.com reached by mistyping more popular domain names).¹⁰⁴

For all its success, widely syndicated pay-per-click advertising led inexorably to a new problem: click fraud. Unscrupulous Web site operators arrange for automated programs to repeatedly mimic the effect of Web surfers clicking on the advertisements displayed on their sites. Google then collects money from the advertisers and passes a cut on to the fraudsters.¹⁰⁵ While Google quickly recognized click fraud as a serious challenge to the viability of its business, it has been hard to eliminate. Estimates of the proportion of Google AdSense clicks that are fraudulent have ranged widely, but some are as high as 35 percent.¹⁰⁶ Google itself has refused to release detailed information on the scale of the problem.

So far, Google has succeeded in making itself the world's most successful Internet company without falling victim to the distractions and overreach that destroyed earlier search

¹⁰³ Hansell, "Google Wants to Dominate Madison Avenue Too".

¹⁰⁴ Paul Sloan, "Masters of their Domains", Business 2.0, December 1 2005.

¹⁰⁵ Charles C Mann, "How Click Fraud Could Swallow the Internet", Wired Magazine, January 2006, 138-49. Click fraud is also discussed in Vise and Malseed, The Google Story, 240-49.

¹⁰⁶ Elinor Mills, Google Calls Click Fraud Estimates Overblown (News.com, August 8 2006 [cited September 21 2006]); available from http://news.com.com/2100-1024_3-6103387.html.

companies. Theoretically its position is vulnerable, because search users could switch to a new and better search service much more quickly and easily than, for example, a different operating system. Microsoft, Amazon, Yahoo, and Ask Jeeves have all spent large amounts of money to try to produce a Google-killing search service. In practice, Google continues to give better, faster and more useful search results than its competitors. It has extended its search capabilities to include images and videos, to search inside files other than Web pages (such as Acrobat, Word, and Powerpoint documents), to search for the best prices on specified goods, to search inside the contents of digitized books, to search the Usenet newsgroup archives, to search inside the email messages stored by users of its Gmail service, to search files on its users' personal computers and to incorporate results from closed-access services such as academic journals.

While maintaining its lead in search, Google has taken advantage of its human and financial resources to develop or acquire a large number of other services. Its list of options recalls the plethora of features crammed into the portal sites of the late 1990s. They include the Google Maps cartography and route finding service along with the related Google Earth satellite image browser, the Blogger blog hosting service, the Google Groups discussion group service, an online spreadsheet application, a chat and internet telephony service, and an online calendar system. Gmail and Google Maps both gained large and enthusiastic user bases by offering services far more elegant, powerful and interactive than their established competitors such as Hotmail and MapQuest. But, in contrast to the cluttered, ugly pages of the late 1990s portal sites, Google's homepage remains a pristine sea of white, holding a single simple graphic for the company logo, a single input box for the search, and just two buttons: "Google Search" and "I'm Feeling lucky." The search box can be used to obtain weather information or stock quotes, perform calculations, track UPS packages, convert currencies, or look up telephone numbers but

its unassuming exterior does nothing to frighten the novice user. Meanwhile Google's full, and potentially overwhelming, list of services is displayed only by clicking on an unobtrusive link labeled "more" and choosing the option "even more."

Furthermore, and in contrast to the old portals, Google is not determined to confine users to its own site. Instead, it has been exploiting the original philosophy of the Internet by making its services easy to customize as part of other applications. Google Maps, for example, can be configured by another website, such as a real estate listings service, to plot the location of points of interest. Its search engine can easily be configured to display only results from a particular website, removing the need for websites to install their own inferior search systems for the benefit of visitors. Google's willingness to make itself part of the Web's software infrastructure by unleashing the power of independent developers may make it a crucial part of the emerging market for location-sensitive Internet services.

Google's steady consolidation of power over Internet search and advertising seems destined to involve it in an ever-growing number of controversies. Google has been sued by online businesses whose sales slumped after they fell in its search rankings, by advertisers suspicious of click fraud, and by publishers seeking to prevent it from digitizing their copyright material for its Google Print book indexing project. It has been denounced by a congressional committee for cooperating with the Chinese government in censoring search results, and taken to court by the US Department of Justice to force the release of query records. Privacy activists fear that Google search makes it too easy to invade the privacy of others, and that the company itself has built up an unprecedented database on the interests and activities of its users. Google appears to have acted at least as responsibly as its peers in all these areas, but given the limitless ambition

and self-confidence of its leaders the firm seems unlikely to escape the kinds of resentments that built up against Microsoft in earlier decades.¹⁰⁷

As Google grew, AOL continued to dwindle from its position as the Internet superpower of the late 1990s. By 2006, it was clear that AOL's growing reliance on Internet advertising posed a strategic challenge for the shrunken business. Should AOL's special attractions such as its instant messenger software, news feeds, and email service be used exclusively as a means of enticing and retaining dial-up customers to its Internet access service? Or should it focus on making its public aol.com Web portal as attractive as possible to all Internet users, in the hope of selling more advertising? The problem was that the more features it reserved for its dial-up customers, the less special its public Web portal could be. In recent years, as AOL's customers shifted in ever greater numbers to broadband services (many of them to its sister company Time Warner Cable) this issue became harder to ignore. Opening up its full range of services to the public might revitalize its portal business but hasten the departure of its remaining nineteen million or so dial-up customers many of whom were tied to the service by the difficulties involved in changing their aol.com email addresses. (The service was also notoriously difficult to cancel, as the process involved long, and sometimes repeated, pleading sessions with "customer retention" specialists).¹⁰⁸ AOL packed its extras for users of high-speed Internet connections as AOL for Broadband. Then, in August 2006 it broke with the past and announced that its

¹⁰⁷ For an example of anti-Google backlash, see Gary Rivlin, "Relax Bill Gates; It's Google's Turn as the Villain" New York Times, August 24 2005, A.1.

¹⁰⁸ Randall Stross, "AOL Said, 'If You Leave Me I'll Do Something Crazy'" New York Times, July 2 2006, Business.3.

premium features, including email accounts, would now be available free of charge to all Internet users.¹⁰⁹ In other words, it is now AOL that is trying to be like Yahoo.

Microsoft, meanwhile, is trying to be more like Google. Only two of Microsoft's portal services won any real following: the Hotmail mail service and MSN Messenger instant messaging software. MSN has struggled to create momentum; in the first six months of 2006 its revenue dropped from a year earlier and it reported losses even as its competitors declared record profits.¹¹⁰ Google's success in making capabilities such as Google Maps into a building block for other website developers, has inspired Microsoft to follow suit. Since 2005, Microsoft has been downplaying the MSN service and redesigning popular components such as its Hotmail system as parts of a new "Windows Live" initiative.

The market for corporate intranet search products (often called "enterprise search") has continued to grow but remains quite small. On a business level, these systems are now a crucial part of many firm's efforts to comply with legal requirements to retain documents, improve document work flow and business processes, and back up business data against possible disaster. By 2005 Open Text held the largest single share, an estimated 13 percent, of a market that

¹⁰⁹ Sara Kehaulani Goo, "In Strategy Shift, AOL Makes Most Services Free" Washington Post, August 3 2006, A.1.

¹¹⁰ Paul R. La Monica, The Internet wars: A Report Card (CNNMoney.com, May 4 2006 [cited September 12 2006]); available from http://money.cnn.com/2006/05/04/technology/search_reportcard/ and Microsoft, Microsoft Reports Fourth Quarter Results and Announces Share Repurchase Program (July 20 2006 [cited September 04 2006]); available from http://www.microsoft.com/msft/earnings/FY06/earn_rel_q4_06.msp. MSN was a division of Microsoft, so direct comparison with competitors is difficult. In mid-2006 Microsoft reorganized to merge MSN into a broader "Online Services" division.

remained heavily fragmented among more than a dozen companies.¹¹¹ Autonomy, one of the other main suppliers, claims to provide more useful results than traditional keyword search, by automatically generating taxonomies and associations between terms. In 2005, Autonomy acquired search pioneer Verity, in one of a series of mergers that are consolidating the field.¹¹²

Corporate search products are evolving in a different direction from their Web-based cousins, toward ever closer integration with existing corporate applications. Unlike public search engines, systems of this kind must often tag documents according to security and access levels, and show only the material that a given user is authorized to see. As corporate search, content management, and portal software companies are acquired by larger enterprise software firms such as Computer Associates, IBM, Oracle and BEA it appears that these formerly distinct kinds of software have merged with each other and, increasingly, with the complex systems these major firms sell to help organizations run their core operations. Autonomy's products, for example, are designed to integrate corporate information from sources such as email, Enterprise Resources Planning systems (such as SAP), voice and video into a single searchable corpus.

Enterprise search systems are licensed as software packages, providing their producers with money from sales plus updates, support fees and consulting charges to implement systems. Autonomy reported revenues of \$117 million in the first half of 2006, most of which came from its acquisition of Verity. Open Text projected around \$200 million in revenues for the same period. While far from insignificant, these figures are dwarfed by Google's revenues of close to

¹¹¹ Tom Eid, Market Share: Enterprise Content Management Software, Worldwide, 2003-2005 (Stamford, CT: Gartner, 2006).

¹¹² Paula J Hane, Autonomy and Verity Join Forces in Enterprise Search Market (InfoToday, November 14 2005 [cited September 14 2006]); available from <http://www.infotoday.com/newsbreaks/nb051114-3.shtml>.

\$5 billion. Interestingly, Google has come up with a different model to sell its technology for enterprise use: the “search appliance,” a bright yellow server loaded with Google indexing and search routines ready to index intranet data and (with suitable bridges) data from other sources such as Salesforce.com records.¹¹³

CONCLUSIONS

It is impossible to imagine the story of Internet navigation services having developed the way it did without the enormous flood of money into the field during the dot com boom. Search and portal firms were funded by venture capitalists in large numbers, encouraged to grow rapidly and became publicly traded companies long before they were profitable or indeed before anyone had a clear idea of how the market would develop. The portal concept was attractive to an immature industry composed of firms desperate to show the rapid revenue growth their investors demanded, but this ill-considered diversification led AltaVista, Lycos, Infoseek and Excite to ruin. Doing one thing well proved a better strategy than doing many things indifferently. Why would users favor a portal offering second- or third-rate implementations of email services or music sales when the best sites in each category were just a click away? In a different business environment it would not have been necessary for the portals to waste quite so many billions of dollars to illustrate this. Only Google, fortunate enough to be able to learn from the mistakes of its predecessors and led by headstrong founders skeptical of conventional business wisdom, stayed focused on providing users with the best possible search results.

¹¹³ Anonymous, [Google Rolls Out Corporate Search](http://news.bbc.co.uk/1/hi/business/3759878.stm) (BBC News, October 20 2004 [cited September 16 2006]); available from <http://news.bbc.co.uk/1/hi/business/3759878.stm>.

In the end, the story of search functions as a kind of parable to explain the strengths and weaknesses of the Web itself. Search is an essential feature of any distributed information system, but one neglected entirely in the Web's original design in favor of simplicity and decentralization. The Web effectively relied on free market innovation, rather than central planning, to provide its navigation capabilities. Dozens of rival search services battled to work around this fundamental deficiency, with considerable success. When the early leaders adopted suicidal business strategies, this free competition between technologies and business models allowed Google to step in with a clearly superior product. But even Google's search results are still full of broken links and out-of-date results. And it remains impossible to see a complete list of pages linking to a particular site, a key part of Nelson's original vision for hypertext. Yet it seems unlikely that any system encumbered with a central database or registry of links could have grown as quickly and evolved as fast as the Web did.

Given that Web pages supply almost none of the metadata (information snippets such as the name of the author, date, publisher, or keywords) relied on by traditional information retrieval systems the effectiveness of current search engines is most impressive. But the lack of metadata makes search engines work hard to achieve sometimes mediocre results. For all Google's cleverness it can only go so far in mitigating the fundamental limitations of the Web. Various efforts are under way to address this issue, most ambitiously a project known as the Semantic Web proceeding under the direction of Berners-Lee.¹¹⁴ It is not clear whether this complex model will ever be widely adopted by Web publishers, but more and more areas of the Web are adopting simpler mechanisms to support metadata. One of the most discussed Web

¹¹⁴ Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer. (Cambridge, MA: MIT Press, 2004).

trends of recent years, the folksonomy, describes systems that allow users to assign their own arbitrary tags to things.¹¹⁵ These tags can then be used by others to search or browse through the material. Popular new sites such as Flickr and YouTube are full of pictures and videos tagged with descriptive terms by visitors to the site. Other services, such as furl.net and digg.com, allow users to assign tags to websites.

Meanwhile, the Web currently supports just one hugely successful portal (Yahoo) and just one hugely successful search engine (Google). As Yahoo has added strong search capabilities and Google has rounded out its range of services their capabilities have begun to converge, even as their personalities remain quite different. Between them they dominate the markets for Internet search advertising and for syndicated advertising displayed on other websites. Whereas conventional wisdom in the late 1990s favored the full-service portal concept and held that Web publishing would be dominated by a few major companies, today people celebrate the Internet's ability to make niche markets profitable by reducing the distribution and production costs necessary to supply books, films, and other entertainments to small audiences (an idea known as the "long tail"). Amazon and eBay make it easy for buyers to find formerly obscure works or goods that fit their taste. Likewise, in the world of Web publishing Yahoo and Google have established themselves as the vital central points where searchers can find what they are looking for and advertisers can find buyers for their products. The profound centralization of search traffic and syndicated advertising revenue in the hands of these firms has supported an equally profound decentralization of Web publishing.

¹¹⁵ John Markoff, "Technology: By and for the Masses" New York Times, June 29 2006, C.1.