washingtonpost.com

Search For Tomorrow
We Wanted Answers, And Google Really Clicked. What's Next?

By Joel Achenbach
Washington Post Staff Writer
Sunday, February 15, 2004; Page D01

In the beginning -- before Google -- a darkness was upon the land.

We stumbled around in libraries. We lifted from the World Book Encyclopedia. We paged through the nearly microscopic listings in the heavy green volumes of the Readers' Guide to Periodical Literature. We latched onto hearsay and rumor and the thinly sourced mutterings of people alleged to be experts. We guessed. We conjectured. And then we gave up, consigning ourselves to ignorance.

Only now in the bright light of the Google Era do we see how dim and gloomy was our pregooglian world. In the distant future, historians will have a common term for the period prior to the appearance of Google: the Dark Ages.

There have been many fine Internet search engines over the years -- Yahoo!, AltaVista, Lycos, Infoseek, Ask Jeeves and so on -- but Google is the first to become a utility, a basic piece of societal infrastructure like the power grid, sewer lines and the Internet itself.

People keep finding new ways to use Google. It is now routine for the romantically savvy to Google a prospective date. "Google hackers" use the infiltrative powers of Google to pilfer bank records and Social Security numbers. The vain Google themselves.

It was about three years ago that the transitive verb "to Google" entered the lexicon, but it was only last year that Google passed all rival search engines in the number of queries handled -- now upwards of 200 million a day. So phenomenal is its success that some industry watchers think an initial public offering of Google stock could raise $20 billion and trigger a second dot-com boom.

"You build a better mousetrap and the world will beat a path to your door," Stewart Brand, computer guru and president of the Long Now Foundation, says of Google. "A wider path, I think, has never been beaten in the history of the world. It's an astonishing mousetrap story."

In the dot-com world, nothing stays the same for long, and it's not clear that Google will forever maintain its dominance over such ferocious rivals as Yahoo! and Microsoft. But the business story of Google is less interesting than the technological one: If information is power, then Google has helped change the world. Knowledge is measurably easier to obtain. Google works. Google knows.

The world used to be transformed by voyages of discovery, religious movements, epidemic globe-circling diseases, the whims of kings and the depredations of armies. But over the centuries, technology has emerged as the primary change agent, the thing that can shrink a planet, undermine dictators and turn 14-year-olds into publishers.

The question is, who's going to build the next mousetrap? What will it do? The laboratories of Internet companies are furiously trying to come up with the next generation of search engine. Whatever it is and whatever it's called, it will likely make the current Google searches seem as antiquated as cranking car engines by hand.

Mom, What's a Library?

The transition into the Google Era has not occurred without some anguish. The stacks of a university library can be a rather lonely place these days. Library circulation dropped about 20 percent at major universities in the first five years after Internet search engines became popular. For most students, Google is where all research begins (and, for the frat boys, ends).

A generation ago, reference librarians -- flesh-and-blood creatures -- were the most powerful search engines on the planet. But the rise of robotic search engines in the mid-1990s has removed the human mediators between researchers and information. Librarians are not so sure they approve. Much of the material on the World Wide Web is wrong, or crazy, or of questionable provenance, or simply out of date (odd to say this about a new technology, but the Web is full of stale information).

"How do you authenticate what you're looking at? How do you know this isn't some kind of fly-by-night operation that's put up this Web site?" asks librarian Patricia Wand of American University.

Students typically search only the most obvious parts of the Web, and rarely venture into what is sometimes called the "Dark Web," the walled gardens of information accessible only through specific databases, such as Lexis-Nexis or the Oxford English Dictionary. And most old books remain undigitized. The Library of Congress has about 19 million books with unique call numbers, plus another 9 million or so in unusual formats, but most have not made it onto the Web. That may change, but for the moment, a tremendous amount of human wisdom is invisible to researchers who just use the Internet.

"For a lot of kids today, the world started in 1996," says librarian and author Gary Price.

And yet Berkeley professor Peter Lyman points out that traditional sources of information, such as textbooks, are heavily filtered by committees, and are full of "compromised information." He's not so sure that the robotic Web crawlers give results any worse than those from more traditional sources.

"There's been a culture war between librarians and computer scientists," Lyman says.

And the war is over, he adds.

"Google won."

Advanced Search

In the early days of search engines, finding information was like fishing in a canal: You might hook something good, but you were just as likely to reel in an old tin can or a rubber boot. Now you often find exactly what you want.

One reason Google works so well today is that there's so much for its robotic crawlers to explore. Google initially searched about 20 million Web pages; the company's home page now boasts that it searches 3,307,998,701 pages.

"In 1996, if you tried to Google someone, if Google existed, it wouldn't have been a very satisfying experience," says Seth Godin, author of a number of best-selling e-books. "We hit a critical mass of really valuable stuff that was online, I think, about 2000."

The expansion of the information universe makes the navigational tool all the more valuable. And yet the search function at first seemed to be an unglamorous computer application. The pioneering search engine companies, including Yahoo!, Excite, AltaVista and Lycos, wanted to transform themselves into something snazzier, a "portal," the full gee-whiz Internet Century home page that would offer the user a link to everything between here and Neptune, plus plane tickets.

But the history of computer technology is full of companies that failed to see the potential glory right in front of them. In the early 1980s, IBM thought that the "operating system" within the computer wasn't nearly as important as the hardware, the box itself. And then Microsoft, which benefited from that oversight, became so focused on software programs that it was slow to capitalize on the Internet revolution, leaving Netscape to create the first commercial Web browser. And then almost everyone underestimated Search.

Not Google. When the company debuted in September 1998, it looked like a throwback. This wasn't a portal. The home page showed mostly white space, anchored by a little rectangle, a box, perfectly blank. Fill in blank and get results. This was plain ol' boring Search, without news headlines, plane tickets, e-mail or any other bells and whistles.

But what results! Google has farms of computers working in parallel. You can put in a couple of words and -- *gzzzzt!* -- get 600,000-plus results within some preposterously brief amount of time. (Google brags about it: "Search took 0.17 seconds." Showoffs!)

Google, the creation of Stanford graduate students Sergey Brin and Larry Page, is like many other search engines in its basic operation. It has powerful software programs that automatically "crawl" the Web, clicking on every possible link, scouting the terrain. What has made Google special is that, in assessing the quality of sites, it takes note of how many other pages link to any given page. This is an old idea from academia, called citation analysis. If many Web sites link to a particular page, the page rises in Google's vaunted "page rank" and is more likely to be on the first page of the search results.

"You're getting the advantage of the group mind," says Paul Saffo, a research director at the Institute for the Future.

This is a key concept: As the Web has grown, it has developed a kind of embedded wisdom. Obviously the Web isn't a conscious entity, but neither is it a completely random pile of stuff. The way one part links to another reflects the preferences of Web users -- and Google tapped into that. Google, in detecting patterns on the Web, harvested meaning from all that madness.

This points the way to one of the next big leaps for search engines: finding meaning in the way a single person searches the Web. In other words, the search engines will study the user's queries and Web habits and, over time, personalize all future searches. Right now, Google and the other search engines don't really know their users.

For example, Saffo isn't really interested in the stuff that most people look for when they do a Web search. He's one of the premier futurists of Silicon Valley and fondly recalls the days, back in the 1980s and early 1990s, the pre-Web era, when the Internet was the reserve of the technological elite who posted their brilliant thoughts on electronic bulletin boards. Now, everyone from about third grade up has an e-mail address and loiters around the Web as though it's the corner 7-Eleven. The results of a Web search reflect the tastes of a broad swath of ordinary Americans who in some cases are still wearing short pants.

"The more people get on the Web, the more the Web becomes the vaster wasteland that is the successor to the vast wasteland of television. I don't care what the majority of people are looking at, because the majority of people are really boring," Saffo says.

He needs a better search engine. He needs one that knows that he's a big-brain tech guru and not an eighth-grader with a paper due.

"The field is called user modeling," says Dan Gruhl of IBM. "It's all about computers watching interactions with people to try to understand their interests and something about them."

Imagine a version of Google that's got a bit of TiVo in it: It doesn't require you to pose a query. It already knows! It's one step ahead of you. It has learned your habits and thought processes and interests. It's your secretary, your colleague, your counselor, your own graduate student doing research for which you'll get all the credit.

To put it in computer terminology, it is your *intelligent agent*.

Calling Agent 001101

No one knows how the intelligent agents of the future might really work, and once you venture more than a few months out you're already into some seriously fuzzy territory. But you might imagine that this intelligent agent could gradually take on so many characteristics of your mind that it becomes something of a digital doppelganger, your shadow self.

To borrow and slightly distort something from "Star Trek," it's like your personal digital Borg, having absorbed your thoughts and melded them with an existing software program.

Perhaps this digital self could become a commodity, something marketable. Imagine that you have to write a paper for a class about the future of search engines. You don't want to use your own lame, broken-down, distracted, gummed-up-with-stupid-stuff virtual secretary to do your research. You want to download Bill Gates's intelligent agent, or Paul Saffo's, or Sergey Brin's, to help you ask smarter questions and find the best answers.

There are primitive intelligent agents already. Amazon.com makes book recommendations based on your previous purchases and the judgments of others who have liked the same books you've liked. But this form of collaborative filtering is still fairly crude.

Microsoft senior researcher Eric Horvitz describes a variety of new and future technologies in which software is more active, more of an entity, no longer just some inert codes waiting for the user to issue a command. For example, there's a program he already uses called IQ, for "implicit query."

"As you're working, we continue to formulate queries in the background, that the user doesn't even know about. They're happening very quietly," Horvitz says.

But Horvitz is keenly aware that people don't want a program that's too pushy, that's constantly interrupting. Humans have limited powers of attention. Software, says Horvitz, "needs to be endowed with the kind of common courtesies we'd expect from a well-mannered colleague."

And lurking over the future of such programs is the dilemma of privacy. There's valuable information in the way people use the Web, but they may not want others, or even a machine, to pay close attention to every place they venture. How do you create an intelligent agent that knows when to look away? How do you avoid what Horvitz calls the "monster possibilities"?

What everyone wants is a reasonable, discreet intelligent agent, like an English butler. It should be one that can get things accomplished, to take the extra steps even without being prompted.

"I don't think anyone wants a search engine," says Seth Godin. "I think people want a find engine."

Find, and do. Solve problems. *Make it so*.

"I often use the analogy of Web agents being like travel agents," says James Hendler, a computer science professor at the University of Maryland. "When I go to my travel agent and say where I want to go, they don't usually just say, 'Yes, you can get there.' They give me some options of different ways to get there. They think about some things I might have forgotten. Do I need a car, do I need a hotel reservation? And then they go do it for me."

Computers as a general rule do only what they're told to do. They don't have artificial intelligence in the classic sense. They have no common sense. IBM's Gruhl, the chief architect of a new product called WebFountain, points out that no computer has ever learned what any 2-year-old human knows.

A computer, he says, can become easily confused by the sentence "Tommy hit a boy with a broken leg." The computer doesn't understand that a broken leg is not going to be an instrument used in an attack. "Common sense, how the world works, even something like irony, are very difficult for computers to understand," says Gruhl.

Semantic Discussions

To achieve common sense, the Web needs to go through the infantile process of self-discovery. The Web doesn't really understand itself. There's lots of information on the Web, but not much "information about information," also known as "metadata."

If you're a robotic search engine, you look for words in the text of a page, but ideally the page would have all manner of encoded labels that describe who wrote the material, and why, and when, and for what purpose, and in what context.

Hendler explains the problem this way: If you type into Google the words "how many cows in Texas," Google will rummage through sites with the words "cow" and "many" and "Texas," and so forth, but you may have trouble finding out how many cows there are in Texas. The typical Web page involving cows and Texas doesn't have anything to do with the larger concept of bovine demographics. (The first Google result that comes up is an article titled "Mineral Supplementation of Beef Cows in Texas" by the unbelievably named Dennis Herd.)

Hendler, along with World Wide Web inventor Tim Berners-Lee, is working on the Semantic Web , a project to implant the background tags, the metadata, on Web sites. The dream is to make it easier not only for humans, but also machines, to search the Web. Moreover, searches will go beyond text and look at music, films, and anything else that's digitized. "We're trying to make the Web a little smarter," Hendler says.

But Peter Norvig, director of search quality at Google, points out that the current keyword-driven searching system, clumsy though it may be and so heavily reliant on serendipity, still works well for most situations.

"Part of the problem is that keywords are so good," he says. "Most of the time the words do what you want them to do."

Billions of dollars are at stake in this race to invent the next mousetrap, and Google faces serious challenges. Yahoo! has long had a partnership with Google, using it to power many of its searches, but Yahoo! has since acquired two other search engine companies, and plans to drop Google in favor of its own Web crawlers. Microsoft, meanwhile, is sure to make search a fundamental element of the next version of its operating system , due in 2006 and code-named Longhorn.

Will Google get steamrolled like Netscape?

"We spend most of our time worrying about ourselves and not our competition," says Google's Norvig.

Technology creates a horizon beyond which human destiny is unknowable, because we can't anticipate all the crazy stuff that brilliant people will invent. The author Michael Crichton has pointed out that a person in the year 1900 might have contemplated all the human beings who would be on the planet in the year 2000, and wondered how it would be possible to obtain enough horses for everyone.

And where would they put all the horse droppings?

Specific predictions are usually wrong. But a general trend has emerged over the course of centuries: Information escapes confinement. Information has been able to break free from monasteries, libraries, school-board-sanctioned textbooks, and corporate publishers. In the Middle Ages, books were kept chained to desks. Information is now completely unchained.

It has a life of its own -- and someday perhaps that won't be just a metaphor.