1

Connecting Information

> *The ultimate search engine would understand exactly what*
> *you mean and give back exactly what you want.*
> —Larry Page[1]

We live in the information age. As society has progressed into the postindustrial era, access to knowledge and information has become the cornerstone of modern living. With the advent of the World Wide Web, vast amounts of information have suddenly become available to people throughout the world. And searching the Web has become an essential capability whether you are sitting at your desktop PC or wandering the corporate halls with your wireless PDA. As a result, there is no better place to start our discussion of *connecting information* than with the world's greatest search engine—Google.

Google has become a global household name—millions use it daily in a hundred languages to conduct over half of all online searches. As a result, Google connects people to relevant information. By providing free access to information, Google offers a seductive gratification to whoever seeks it. To power its searches, Google uses patented, custom-designed programs and hundreds of thousands of computers to provide the greatest computing power of any enterprise.

Searching for information is now called *googling*, which men, women, and children can perform over computers and cell phones. And thanks to small targeted advertisements that searchers can click on for information, Google has become a financial success.

In this chapter, we follow the hero's journey of Google founders Larry Page and Sergey Brin as they invent their Googleware technology for efficient con-

---

[1]M. Prather, "Ga-Ga for Google," *Entrepreneur Magazine*, April 2002.

nection to information, then go on to become masters in pursuit of their holy grail—perfect search.

THE GOOGLE STORY

Google was founded by two Ph.D. computer science students at Stanford University in California—Larry Page and Sergey Brin. When Page and Brin began their hero's journey, they didn't know exactly where they were headed.

It is widely known that, at first, Page and Brin didn't hit it off. When they met in 1995, 24-year-old Page was a new graduate of the University of Michigan visiting Stanford University to consider entering graduate school; Brin, at age 23, was a Stanford graduate student who was assigned to host Page's visit. At first, the two seemed to differ on just about every subject they discussed. They each had strong opinions and divergent viewpoints, and their relationship seemed destined to be contentious.

Larry Page was born in 1973 in Lansing, Michigan. Both of his parents were computer scientists. His father was a university professor and a leader in the field of artificial intelligence, while his mother was a teacher of computer programming. As a result of his upbringing in this talented and technology-oriented family, Page seemed destined for success in the computer industry in one way or another.

After graduating from high school, Page studied computer engineering at the University of Michigan, where he earned his Bachelor of Science degree. Following his undergraduate studies, he decided to pursue graduate work in computer engineering at Stanford University. He intended to build a career in academia or the computer science profession, building on a Ph.D. degree.

Meanwhile, Sergey Brin was also born in 1973, in Moscow, Russia, the son of a Russian mathematician and an economist. His entire family fled the Soviet Union in 1979 under the threat of growing anti-Semitism and began their new lives as immigrants in the United States.

Brin displayed a great interest in computers from an early age. As a youth, he was influenced by the rapid popularization of personal computers and was very much a child of the microprocessor age. He too was brought up to be familiar with mathematics and computer technology, and as a young child, in the first grade he turned in a computer printout for a school project. Later, at the age of nine, he was given a Commodore 64 computer as a birthday gift from his father.

Brin entered the University of Maryland at College Park, where he studied mathematics and computer science. He completed his studies at the University of Maryland in 1993, and was awarded a Bachelor of Science degree. Following his undergraduate studies, he was given a National Science Foundation fellowship to pursue graduate studies in computer science at Stanford University. Not only did he exhibit early talent and interest in mathematics and computer

science, he also became acutely interested in data management and networking as the Internet was becoming an increasing force in American society. While at Stanford, he pursued research and prepared publications in the areas of data mining and pattern extraction. He also wrote software to convert scientific papers written in TeX, a cross-platform text processing language, into HyperText Markup Language (HTML), the multimedia language of the World Wide Web.

Brin successfully completed his master's degree at Stanford. Like Page, Brin's intent was to continue in his graduate studies to earn a Ph.D., which he also viewed as a great opportunity to establish an outstanding academic or professional career in computer science.

The hero's journey for Page and Brin began as they heard the call—to develop a unique approach for retrieving relevant information from the voluminous data on the World Wide Web.

Page remembered, "When we first met each other, we each thought the other was obnoxious. Then we hit it off and became really good friends. . . . I got this crazy idea that I was going to download the entire Web onto my computer. I told my advisor it would only take a week . . . So I started to download the Web, and Sergey started helping me because he was interested in data mining and making sense of the information."[2]

Although Page initially thought the downloading of the Web would be a short-term project, taking a week or so to accomplish, he quickly found that the scope of what he wanted to do was much greater than his original estimate. Once he started his downloading project, he enlisted Brin to join the effort. While working together the two became inspired and wrote the seminal paper entitled *The Anatomy of a Large-Scale Hypertextual Web Search Engine*.[3] It explained their efficient ranking algorithm, PageRank.

Brin said about the experience, "The research behind Google began in 1995. The first prototype was actually called BackRub. A couple of years later, we had a search engine that worked considerably better than the others available did at the time."[4]

This prototype listed the results of a Web search according to a quantitative measure of the popularity of the pages. By January 1996, the system was able to analyze the "back links" pointing to a given website and from this quantify the popularity of the site. Within the next few years, the prototype system had been converted into progressively improved versions, and these were substantially more effective than any other search engine then available.

As the buzz about their project spread, more and more people began to use it. Soon they were reporting that there were 10,000 searches per day at

[2]D. A. Vise and M. Malseed, *The Google Story*, Delacourt Press, New York, 2005.
[3]S. Brin and L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Computer Science Department, Stanford University, Stanford, 1996.
[4]S. Brin and L. Page, "*The Future of the Internet*," speech to the Commonwealth Club, March 21, 2001.

Stanford on their system. With this growing use and popularity of their search system, they began to realize that they were maxing out their search ability due to the limited number of computers they had at their disposal. They would need more hardware to continue their remarkable expansion and enable more search activity. As Page said, "This is about how many searches we can do, and we need more computers. Our whole history has been like that. We always need more computers."[5]

In many ways, the research project at Stanford was a low budget operation. Because of a chronic shortage of cash, the pair are said to have monitored the Stanford Computer Science Department's loading docks for newly arrived computers to "borrow." In spite of this, within a short span of time, the reputation of the BackRub system had grown dramatically and their new search technology began to be broadly noticed.

They named their successor search engine *Google*, in a whimsical analogy to the mathematical term *googol*, which is the immensely large number 1 followed by 100 zeros. The transition from the earlier Backrub technology to the much more sophisticated Google was slow. But the Google system began with an index of 25 million pages and the capability to handle 10,000 search queries every day, even when it was in its initial stage of introduction. The Google search engine grew quickly as it was continuously improved. The effectiveness and relevance of the Google searches, its scope of coverage, its speed and reliability, and its clean user interface all contributed to a rapid increase in the popularity of the search engine.

At this time, Google was still a student research project, and both Page and Brin were still intent on completing their respective doctoral programs at Stanford. As a result, they initially refused to "answer the call" and continued to devote themselves to the academic pursuit of the technology of search.

Through all this, Brin maintained an eclectic collection of interests and activities. He continued with his graduate research interests at Stanford and he collaborated with his fellow Ph.D. students and professors on other projects such as automatic detection. At the same time, he also pursued a variety of outside interests, including sailing and trapeze. Brin's father had stressed the importance for him to complete his Ph.D. He said, "I expected him to get his Ph.D. and to become somebody, maybe a professor." In response to his father's question as to whether he was taking any advanced courses one semester, Brin replied, "Yes, advanced swimming."[6]

While Brin and Page continued on as graduate students, they began to realize the importance of what they had succeeded in developing. The two aspiring entrepreneurs decided to try and license the Google technology to existing Internet companies. But they found themselves unsuccessful in stimulating the interest of the major enterprises. They were forced to face the crucial decision of continuing on at Stanford or striking out on their own. With their

[5]D. A. Vise and M. Malseed, *The Google Story*, Delacourt Press, New York, 2005.
[6]Ibid.

realization that they were onto something that was important and perhaps even groundbreaking, they decided to make the move.

Thus our two heroes had reached their point of departure and they crossed over from the academic into the business world. As they committed to this new direction, they realized they would need to postpone their educational aspirations, prepare plans for their business concept, develop a working demo of their commercial search product, and seek funding sponsorship from outside investors.

Having made this decision, they managed to interest Sun Microsystems founder Andy Bechtolsheim in their idea. As Brin recalls, "We met him very early one morning on the porch of a Stanford faculty member's home in Palo Alto. We gave him a quick demo. He had to run off somewhere, so he said, 'Instead of us discussing all the details, why don't I just write you a check?' It was made out to Google Inc. and was for $100,000."[7]

The check remained in Page's desk uncashed for several weeks while he and Brin set up a corporation and sought additional money from family and friends—almost $1 million in total. Having started the new company, lined up investor funding, and possessing a superb product, they realized ultimate success would require a good balance of perspiration as well as inspiration. Nevertheless, at this point Google appeared to be well on the road to success.

Page and Brin have been on a roll every since, armed with the great confidence that they had both a superior product and an excellent vision for global information collection, storage, and retrieval. In addition, they believed that coordination and optimization of the entire hardware/software system was important, and so they developed their own Googleware technology by combining their custom software with appropriately integrated custom hardware, thereby fully leveraging their ingenious concept.

Google Inc. opened its doors as a business entity in September 1998, operating out of modest facilities in a Menlo Park, California garage.

As Page and Brin initiated their journey, they faced many challenges along the way. They matured in their understanding with the help of mentors they encountered such as Yahoo!'s Dave Filo. Filo not only encouraged the two in the development of their search technology but also made business suggestions for their project.

Following the company start-up, interest in Google grew rapidly. Red Hat, a Linux company, signed on as Google's first commercial customer. Red Hat was particularly interested in Google because the company realized the importance of search technology and its ability to run on open source systems such as Linux. In addition, the press began to take notice of this new commercial venture and articles began to appear in the media highlighting the Google product that offered relevant search results.

The late 1990s saw a spectacular growth in development of the technology industry, and Silicon Valley was awash with investor funding. The timing was

---

[7]"Search Us, Says Google," *Technology Review*, January 11, 2002.

right for Google, and in 1999, Page and Brin sought and received a second round of funding, obtaining $25 million from Silicon Valley venture capital firms.

The additional funding enabled them to expand their operations and move into new facilities they called the *Googleplex*, Google's current headquarters in Mountain View, California. Although at the time they occupied only a small portion of the new two-story building, they had clearly come a long way from a university research project to a full-fledged technology company with a rapid growth trajectory and a product that was in high demand.

Google was also in the process of developing a unique company culture. It operated in an informal atmosphere that facilitated both collegiality and an easy exchange of ideas. Google staffers enjoyed this rewarding atmosphere while they continued to make many incremental improvements to their search engine technology. For example, in an effort to expand the utility of their keyword-targeted advertising to small businesses, they rolled out the AdWords system, a software package that represents a self-service advertisement development capability.

Google took a major step forward when, in 2000, it was selected by Yahoo! to replace Inktomi as its provider of supplementary search results. Because of the superiority of Google over other search engine capabilities, licenses were obtained by many other companies, including the Internet services powerhouse America Online (AOL), Netscape, Freeserve, and eventually Microsoft Network (MSN). In fact, although Microsoft has pursued its own search technology, Bill Gates once commented on search engine technology development by saying that "Google kicked our butts."[8]

By the end of 2000, Google was handling more than 100 million searches each day. Shortly thereafter, Google began to deliver new innovations and establish new partnerships to enter the burgeoning field of mobile wireless computing. By expanding into this field, Google continued to pursue its strategy of putting search into the hands of as many users as possible.

As the global use of Google grew, the patterns contained within the records of search queries provided new information about what was on the minds of the global community of Internet users. Google was able to analyze the global traffic in Internet searching and identify patterns, trends, and surprises—a process they called *Google Zeitgeist*.

In 2004, Yahoo! decided to compete directly with Google and discontinued its reliance on the Google search technology. Nevertheless, Google continued to expand, increasing its market share and dominance of the Web search market through the deployment of regional versions of its software, incorporating language capabilities beyond English. As a result, Google continued to expand as a global Internet force.

Also in 2004, Google offered its stock to investors through an Initial Public Offering (IPO). This entrance into public trading of Google stock

---

[8]K. Kelleher, "Google vs. Gates," *Wired* 12(03): March 2004.

created not only a big stir in the financial markets but also great wealth for the two founding entrepreneurs. Page and Brin immediately joined the billionaire's club as they entered the exclusive ranks of the wealthiest people in the world.

Following the IPO, Google began to challenge Microsoft in its role as the leading provider of computer services. Google issued a series of new products, including the email service Gmail, the impressive map and satellite image product Google Earth, Google Talk to compete in the growing Voice over Internet Protocol (VoIP) market, and products aimed at leveraging the ambitious project to make the content of thousands of books searchable online, Google Base and Google Book Search. In addition to these new ventures, Google has continued to innovate in its core field of search by introducing new features for searching images, news articles, shopping services (Froogle), and other local search options.

It is clear that Google has become an essential tool for connecting people and information in support of the developing Information Revolution. Having established itself at the epicenter of the Web, Google is widely regarded as the "place to be" for the best and brightest programming talent in the industry. It is fair to say that, since the introduction of the printing press, no other entity or event has had more impact on public access to information than Google.

In fact, Google has endeavored to accumulate a good part of all human knowledge from the vast amount of information stored on the Web. The effective transformation of Google into an engine for what Page calls perfect search would basically give people everywhere the right answers to their questions and the ability to understand everything in the world.

Page and Brin could not have achieved their technological success without having a clear vision of the future of the Internet. Page recently commented in an interview that he believes that in the future "information access and communications will become truly ubiquitous," meaning that "anyone in the world will have access to any kind of information they want or be able to communicate with anyone else instantly and for very little cost." In fact, this vision of the future is not far from where we are now.[9]

Page also noted that the real power of the Internet is the ability to serve people all over the globe with access to information, which represents empowerment of individuals. The ability to facilitate the improved lives and productivity of billions of human beings throughout the world is an awesome potential outcome.

And the ability to support the information needs of people from different cultures and languages is an unusual challenge. Page stated in an interview that "even language is becoming less of a barrier. There's pretty good automatic translation out there. I've been using it quite a bit as Google becomes

---

[9]S. Brin and L. Page, *The Future of the Internet*, speech to the Commonwealth Club, March 21, 2001.

more globalized. It doesn't translate documents exactly, but it does a pretty good job and it's getting better every day."[10]

Even with translation and global reach, however, there remain significant challenges to connecting the people of the world through advanced information technology. One of the challenges is the potential for governmental restrictions on the access to information. Encryption technology, for example, inhibits the power of governments to monitor or control such information access. However, a 1998 survey of encryption policy found that several countries, including Belarus, China, Israel, Pakistan, Russia, and Singapore, maintained strong domestic controls while several other countries were considering the adoption of such controls.[11]

The phrase "Don't be evil" has been attributed to Google as its catch phrase or motto. Google's present CEO, Eric Schmidt, commented, in response to questions about the meaning of this motto, that "evil is whatever Sergey says is evil."

Brin, on the other hand, said in an interview with *Playboy Magazine*: "As for 'Don't be evil,' we have tried to define precisely what it means to be a force for good—always do the right, ethical thing. Ultimately 'Don't be evil' seems the easiest way to express it." And Page also commented on the phrase, saying "Apparently people like it better than 'Be good'."[12]

Page and Brin maintain lofty ambitions for the future of information technology, and they communicated those ambitions in an unprecedented seven-page letter to Wall Street entitled "An Owner's Manual for Google's Shareholders," written to detail Google's intentions as a public company. They explained their vision that "searching and organizing all the world's information is an unusually important task that should be carried out by a company that is trustworthy and interested in the public good."[13]

In response to questions about how Google will be used in the future, Brin said, "Your mind is tremendously efficient at weighing an enormous amount of information. We want to make smarter search engines that do a lot of the work for us. The smarter we can make the search engine, the better. Where will it lead? Who knows? But it's credible to imagine a leap as great as that from hunting through library stacks to a Google session, when we leap from today's search engines to having the entirety of the world's information as just one of our thoughts."[14]

At this junction, Page and Brin find themselves in a state of great personal wealth and great accomplishment, having created a technology and company that is profoundly affecting human culture and society. The two computer

---

[10]Ibid.

[11]"Cryptography and Liberty 1998, An International Survey of Encryption Policy," February 1998. `http://www.gilc.org/crypto/crypto-survey.html`.

[12]"Google Guys," *Playboy Magazine*, September 2004.

[13]From Google's letter to prospective shareholders. `http://www.thestreet.com/_yahoo/markets/marketfeatures/10157519_6.html`.

[14]"Google Guys," *Playboy Magazine*, September 2004.

scientists have traveled far in their hero's journey to carry out their vision of global search, having developed skills and capabilities for themselves as well as for Google and the Googleware technology. As they succeeded, their search technology became a key milestone in the development of the Information Revolution. Their journey is not over, however. Before continuing their story, let's digress into the historical context.

THE INFORMATION REVOLUTION

Over past millennia, the world has witnessed two global revolutions: the Agricultural Revolution and the Industrial Revolution.

During the Agricultural Revolution, a hunter-gatherer could acquire the resources from an area of 100 acres to produce an adequate food supply, whereas a single farmer needed only one acre of land to produce the equivalent amount of food. It was this 100-fold improvement in land management that fueled the Agricultural Revolution. It not only enabled far more efficient food production, but also provided food resources well above the needs of subsistence, resulting in a new era built on trade.

Where a single farmer and his horse had worked a farm, during the Industrial Revolution workers were able to use a single steam engine that produced 100 times the horsepower of this farmer–horse team. As a result, the Industrial Revolution placed a 100-fold increase of mechanical power into the hands of the laborer. It resulted in the falling cost of labor and this fueled the unprecedented acceleration in economic growth that ensued.

Over the millennia, humans have accumulated great knowledge, produced a treasury of cultural literature, and developed a wealth of technology advances, much of which has been recorded in written form. By the mid-20th century, the quantity of accessible useful information had grown explosively, requiring new methods of information management; and this can be said to have triggered the Information Revolution. As computer technology offered great improvements in information management technology, it also provided substantial reductions in the cost of information access. It did more than allow people to receive information. Individuals could buy, sell, and even create their own information. Cheap, plentiful, easily accessible information has become as powerful an economic dynamic as land and energy had for the two prior revolutions.

The falling cost of information has, in part, reflected the dramatic improvement in price–performance of microprocessors, which appears to be on a pattern of doubling every eighteen months. While the computer has been contributing to information productivity since the 1950s, the resulting global economic productivity gains were initially slow to be realized.

Until the late 1990s, networks were rigid and closed, and time to implement changes in the telecommunication industry was measured in decades. Since then, the Web has become the grim reaper of information inefficiency.

For the first time, ordinary people had real power over information production and dissemination. As the cost of information dropped, the microprocessor in effect gave ordinary people control over information about consumer products.

Today, we are beginning to see dramatic change as service workers experience the productivity gains from rapid communications and automated business and knowledge transactions. A service worker can now complete knowledge transactions 100 times faster using intelligent software and near ubiquitous computing in comparison to a clerk using written records. As a result, the Information Revolution is placing a 100-fold increase in transaction speed into the hands of the service worker. Therefore, the Information Revolution is based on the falling cost of information-based transactions, which in turn fuels economic growth.

In considering these three major revolutions in human society, a defining feature of each has been the requirement for more knowledgeable and more highly skilled workers. The Information Revolution signals that this will be a major priority for its continued growth. Clearly, the Web will play a central role in the efficient performance of the Information Revolution because it offers a powerful communication medium that is itself becoming ever more useful through intelligent applications.

Over the past fifty years, the Internet/World Wide Web has grown into the global Information Superhighway. And just as roads connected the traders of the Agricultural Revolution and railroads connected the producers and consumers of the Industrial Revolution, the Web is now connecting people to information in the Information Revolution.

The Information Revolution enables service workers today to complete knowledge transactions many times faster through intelligent software using photons over the Internet, in comparison to clerks using electrons over wired circuits just a few decades ago.

But perhaps the most essential ingredient in the Web's continued success has been search technology such as Google, which has provided real efficiency in connecting to relevant information and completing vital transactions. Now Google transforms data and information into useful knowledge, energizing the Information Revolution.

DEFINING INFORMATION

Google started with Page and Brin's quest to mine data and make sense of the voluminous information on the Web. But what differentiates information from knowledge and how do companies like Google manipulate it on the Web to nourish the Information Revolution?

First, let's be clear about what we mean by the fundamental terms *data, information, knowledge*, and *understanding*.

An item of data is a fundamental element of information, the processed data that has some independent usefulness. And right now data is the main thing you can find directly on the Web in its current state. Data can be considered the raw material of information. Symbols and numbers are forms of data.

Data can be organized within a database to form structured information. While spreadsheets are "number crunchers," databases are the "information crunchers." Databases are highly effective in managing and manipulating structured data.[15]

Consider, for example, a directory or phone book that contains elements of information (i.e., names, addresses, and phone numbers) about telephone customers in a particular area. In such a directory, each customer's information is laid out in the same pattern. The phone book is basically a table that contains a record for each customer. Each customer's record includes his/her name, address, and phone number. But you can't directly search such a database on the Web. This is because there is no schema defining the structure of data on the Web. Thus what looks like information to the human being who is looking at the directory (taking with him his background knowledge and experience as a context) in reality is data because it lacks this schema.

On the other hand, information explicitly associates one set of things to another. A telephone book full of data becomes information when we associate the data to persons we know or wish to communicate with.

For example, suppose we found data entries in a telephone book for four different persons named Jones, but all of them were living within one block of each other. The fact that there are four bits of data about persons with the same name in approximately the same location is interesting information.

Knowledge, on the other hand, can be considered to be a meaningful collection of useful information. We can construct information from data. And we can construct knowledge from information. Finally, we can achieve understanding from the knowledge we have gathered.

Understanding lies at the highest level. It is the process by which we can take existing knowledge and synthesize new knowledge. Once we have understanding, we can pursue useful actions because we can synthesize new knowledge or information from what is previously known.

Again, knowledge and understanding are currently elusive on the Web. Future Semantic Web architectures seek to redress this limit.

To continue our telephone example, suppose we developed a genealogy tree for the Joneses and found the four Joneses who lived near each other were actually brothers. This would give us additional knowledge about the Joneses in addition to information about their addresses. If we then interviewed the brothers and found that their father had bought each brother a house in his neighborhood when they married, we would finally understand

---

[15]"Databases from Scratch I: Introduction." `http://brebru.com/databases_from_scratch_1.html`.

quite a bit about them. We could continue the interviews to find out about their future plans for their offspring—thus producing more new knowledge.

If we could manipulate data, information, knowledge, and understanding by combining a search engine, such as Google, with a reasoning engine, we could create a logic machine. Such an effort would be central to the development of artificial intelligence (AI) on the Web.

AI systems seek to create understanding through their ability to integrate information and synthesize new knowledge from previously stored information and knowledge. An important element of AI is the principle that intelligent behavior can be achieved through processing of symbolic structures representing increments of knowledge. This has produced knowledge-representation languages that allow the representation and manipulation of knowledge to deduce new facts from the existing knowledge.

The World Wide Web has become the greatest repository of information on virtually every topic. Its biggest problem, however, is the classic problem of finding a needle in a haystack. Given the vast stores of information on the Web, finding exactly what you're looking for can be a major challenge. This is where search engines, like Google, come in—and where we can look for the greatest future innovations to come when we combine AI and search.

Larry Page and Sergey Brin found that the existing search technology looked at information on the Web in simple ways. They decided that to deliver better results, they would have to go beyond simply looking, to looking good.

LOOKING GOOD

Commercial search engines are based on one of two forms of Web search technologies: human-directed search and automated search. Human-directed search is search in which the human performs an integral part of the process. In this form of search engine technology, a database is prepared of keywords, concepts, and references that can be useful to the human operator. Searches that are keyword based are easy to conduct but they have the disadvantage of providing large volumes of irrelevant or meaningless results. The basic idea in its simplest form is to count the number of words in the search query that match words in the keyword index and rank the Web page accordingly. Although more sophisticated approaches also take into account the location of the keywords, the improved performance may not be substantial. As an example, it is known that keywords used in the title tags of Web pages tend to be more significant than words that occur in the Web page, but not in the title tag; however, the level of improvement may be modest.

Another approach is to use hierarchies of topics to assist in human-directed search. The disadvantage of this approach is that the topic hierarchies must be independently created and are therefore expensive to create and maintain.

The alternative approach is automated search; this approach is the path taken by Google. It uses software agents, called Web crawlers (also called spiders, robots, bots, or agents), to automatically follow hypertext links from one site to another on the Web until they accumulate vast amounts of information about the Web pages and their interconnections. From this, a complex index can be prepared to store the relevant information. Such automated search methods accumulate information automatically and allow for continuing updates.

However, even though these processes may be highly sophisticated and automatic, the information they produce is represented as links to words, and not as meaningful concepts.

Current automated search engines must maintain huge databases of Web page references. There are two implementations of such search engines: individual search engines and meta-searchers. Individual search engines (such as Google) accumulate their own databases of information about Web pages and their interconnections and store them in such a way as to be searchable. Meta-searchers, on the other hand, access multiple individual engines simultaneously, searching their databases.

In the use of keywords in search engines, there are two language-based phenomena that can significantly impact effectiveness and therefore must be taken into account. The first of these is polysemy, the fact that single words frequently have multiple meanings; and the second is synonymy, the fact that multiple words can have the same meaning or refer to the same concept.

In addition, there are several characteristics required to improve a search engine's performance. It is important to consider useful searches as distinct from fruitless ones. To be useful, there are three necessary criteria: (1) maximize the relevant information, (2) minimize irrelevant information, and (3) make the ranking meaningful, with the most highly relevant results first.

The first criterion is called recall. The desire to obtain relevant results is very important, and the fact is that, without effective recall, we may be swamped with less relevant information and may, in fact, not receive the most important and relevant results. It is essential to reduce the rate of false negatives—important and relevant results that are not displayed—to a level that is as low as possible.

The second criterion, minimizing irrelevant information, is also very important to ensure that relevant results are not swamped; this criterion is called precision. If the level of precision is too low, the useful results will be highly diluted by the uninteresting results, and the user will be burdened by the task of sifting through all of the results to find the needle in the haystack. High precision means a very low rate of false positives, irrelevant results that are highly ranked and displayed at the top of our search result.

Since there is always a trade-off between reducing the risk of missing relevant results and reducing the level of irrelevant results, the third criterion, ranking, is very important. Ranking is most effective when it matches our information needs in terms of our perception of what is most relevant in our

results. The challenge for a software system is to be able to accurately match the expectations of a human user since the degree of relevance of a search contains several subjective factors such as the immediate needs of the user and the context of the search. Many of the desired characteristics for advanced search, therefore, match well with the research directions in artificial intelligence and pattern recognition. By obtaining an awareness of individual preferences, for example, a search engine could more effectively take them into account in improving the effectiveness of search.

Recognizing ranking algorithms were the weak point in competing search technology, Page and Brin introduced their own new ranking algorithm—PageRank.

GOOGLE CONNECTS INFORMATION

Just as the name Google is derived from the esoteric mathematical term googol, in the future, the direction of Google will focus on developing the esoteric "perfect search engine," defined by Page as something that "understands exactly what you mean and gives you back exactly what you want." In the past, Google has applied great innovation to try and overcome the limitations of prior search approaches; PageRank was conceived by Google to overcome some of the key limitations.[16]

Page and Brin recognized that providing the fastest, most accurate search results would require a new approach to server systems. While most search engines used a small number of large servers that often slowed down under peak use, Google went the other direction by using large numbers of linked PCs to find search results in response to queries. The approach turned out to be effective in that it produced much faster response times and greater scalability while minimizing costs. Others have followed Google's lead in this innovation while Google has continued its efforts to make its systems more efficient.

Google takes a parallel processing approach to its search technology by conducting a series of calculations on multiple processors. This has provided Google with critical timing advantage, permitting its search algorithms to be very fast. While other search engines rely heavily on the simple approach of counting the occurrences of keywords, Google's PageRank approach considers the entire link structure of the Web to help in the determination of Web page importance. By then performing a hypertext matching assessment to narrow the search results for the particular search being conducted, Google achieves superior performance. In a sense, Google combines insight into Web page importance with query-specific attributes to rank pages and deliver the most relevant results at the top of the search results.

The PageRank algorithm analyzes the importance of the Web pages it considers by solving an exceptionally complex set of equations with a huge

---

[16]Quotes from `http://www.google.com/corporate/tech.html`.

number of variables and terms. By considering links between Web pages as "votes" from one page to another, PageRank can assign a measure of a page's importance by counting its votes.

It also takes into account the importance of each page that supplies a vote and, by appropriately weighting these votes, further improves the quality of the search. In addition, PageRank considers the Web page content, but unlike other search engines that restrict such consideration to the text content, Google consider the full contents of the page.

In a sense, Google attempts to use the collective intelligence of the Web, a topic for further discussion later in this book, in its effort to improve the relevance of its search results. Finally, because the search algorithms used by Google are automated, Google has earned a reputation for objectivity and lack of bias in its results.

Throughout their exciting years establishing and growing Google as a company, Page and Brin realized that continued innovation was essential. They undertook to find innovative services that would enhance access to Web information with added thought and not a little perspiration. Page said that he respected the idea of having "a healthy disregard for the impossible."[17]

In February 2002, the Google Search Appliance, a plug-and-play application for search, was introduced. In short order, this product was dispersed throughout the world populating company networks, university systems, and the entire Web. The popular Google Search Appliance is referred to as "Google in a box."

In another initiative, Google News was introduced in September 2002. This free news service, which allows automatic selection and arrangement of news headlines and pictures, features real-time updating and tailoring, allowing users to browse the news with scan and search capabilities.

Continuing Google's emphasis on innovation, the Google search service for products, Froogle, was launched in December 2002. Froogle allows users to search millions of commercial websites to find product and pricing information. It enables users to identify and link to a variety of sources for specific products, providing images, specifications, and pricing information for the items being sought.

Google's innovations have also impacted the publishing business with both search and advertising features. Google purchased Pyra Labs in 2003 and thus became the host of Blogger, a leading service for the sharing of thoughts and opinions through online journals, or blogs (weblogs).

Finally, Google Maps became a dynamic online mapping feature, and Google Earth a highly popular mapping and satellite imagery resource. Using these innovative applications, users can find information about particular locations, get directions, and display both maps as well as satellite images of a desired address.

[17]D. A. Vise and M. Malseed, *The Google Story*, Delacourt Press, New York, 2005.

With each new capability, Google expands our access to more information[18] and moves us closer to Page's holy grail—perfect search.

At this junction, Page and Brin have finally completed their hero's journey. They have become the *masters of search*; committed to improving access to information and lifting the bonds of ignorance from millions around the world.

PATTERNS OF DISCOVERY

Larry Page and Sergey Brin were trying to solve the problem of easy, quick access to all Web information, and ultimately to all human knowledge. In order to index existing Web information and provide rapid relevant search results, their challenge was to sort through billions of pages of material efficiently and explicitly find the right responses.

They were confident that their vision for developing a global information collection, storage, and retrieval system would succeed if they could base it on a unique and efficient ranking algorithm.

The process of inspiration for Page and Brin became fulfilled when they completed their seminal paper entitled *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, which explained their efficient ranking algorithm, PageRank. In developing a breakthrough ranking algorithm based on the ideas of publication ranking, Page and Brin experienced a moment of inspiration.

But they didn't stop there. They also believed that optimization was vitally important and so they developed their own Googleware technology consisting of combining custom software with custom hardware, thereby reflecting the founders' genius. They built the world's most powerful computational enterprise, and they have been on a roll ever since. Page stressed that inspiration still required perspiration and that Google appeared destined for rapid growth and expansion. In building the customized computer Googleware infrastructure for PageRank, they were demonstrating the 1% Inspiration and 99% Perspiration Pattern.

The result was Google, the dominant search engine connecting people to all of the World Wide Web's information.

FORECAST FOR CONNECTING INFORMATION

For many of us it seems that an uncertain future looms ahead like a massive opaque block of granite. But just as Michelangelo suggested that he took a block of stone and chipped away the nonessential pieces to produce *David*, we can chip away the improbable to uncover the possible. By examining inventors and their process of discovery, we are able to visualize the tapestry of our past to help unveil patterns that can serve as our guideposts on our path forward.

---

[18]S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertexual wab Search Engine," *Computer Networks*, 30(1–7), 107–117, 1998.

Page and Brin invented an essential search technology, but their contributions to information processing were evolutionary in nature—built on inspiration and perspiration. One forecast for connecting information is that we can expect a continued pattern of inspired innovation as we go forward in the expansion of search and related technology.

Discoveries Requiring Inspiration and Perspiration

In considering the future for connecting information, we expect that improved ranking algorithms will ensure Google's continued dominance for some time to come. Extrapolating from Google's success, we can expect a series of inspired innovations building on its enterprise computer system, such as offering additional knowledge-related services.

Future Google services could include expanding into multimedia areas such as television, movies, and music using Google TV and Google Mobile. Viewers would have all the history of TV from which to choose. And Google would offer advertisers targeted search. Google Mobile could deliver the same service and products to cell phone technology. By 2020, Google could digitize and index every book, movie, TV show, and song ever produced, making it available conveniently.

In addition, Google could dominate the Internet as a hub site. The ubiquitous GoogleNet would dominate wireless access and cell phones. As for the Google browser, Gbrowser, it could replace operating systems.

However, our vision also includes connecting information through developing more intelligent search capabilities. A new Web architecture, such as Tim Berners-Lee's Semantic Web, would add knowledge representation and logic to the markup languages of the Web. Semantics on the Web would offer extraordinary leaps in Web search capabilities.

Since Google has cornered online advertising, it has become progressively more precision targeted and inexpensive. But Google also has 150,000 servers with nearly unlimited storage space and massive processing power.

Beyond simply inspired discoveries, Google or other search engine powers could find innovations based on new principles yet to be proved, as suggested in the following.

Discoveries Requiring New Proof of Principle

Technology futurists such as Ray Kurzweil have suggested that Strong AI (software programs that exhibit true intelligence) could emerge from developing Web-based systems such as that of Google. Strong AI could perform data mining at a whole new level. This type of innovation would require a proof of principle.

Some have suggested that Google's purpose in converting books into electronic form is not to provide for humans to read them, but rather to provide a form that could be accessible by software, with AI as the consumer.

One of the great areas of innovation resulting from Google's initiatives is its ability to search the human genome. Such technology could lead to a personal DNA search capability within the next decade. This could result in the identification of medical prescriptions that are specific to you; and you would know exactly what kinds of side effects to expect from a given drug.

And consider what might happen if we had the perfect search. Think about the capability to ask any question and get the perfect answer—an answer with real context. The answer could incorporate all of the world's knowledge using text, video, or audio. And it would reflect every nuance of meaning. Most importantly, it would be tailored to your own particular context. That's the stated goal of IBM, Microsoft, Google, and others. Such a capability would offer its greatest benefits when knowledge is easily gathered.

Soon search will move away from the PC-centric operations to the Web connected to many small devices such as mobile phones and PDAs. The most insignificant object with a chip and the ability to connect will be network aware and searchable. And search needs to solve access to deep databases of knowledge, such as the University of California's library system. While there are several hundred thousand books online, there are 100 million more that are not.

The perfect search will find all this information and connect us to the world's knowledge, but this is the beginning of decision making, not the end. Search and artificial intelligence seem destined to get together.

In the coming chapters, we will be exploring all the different technologies involved in connecting information and we will be exploring how the prospects for the *perfect search* could turn into *ubiquitous intelligence*.

First, ubiquitous computing populates the world with devices using microchips everywhere. Then the ubiquitous Web connects and controls these devices on a global scale. The ubiquitous Web is a pervasive Web infrastructure that allows all physical objects access by URIs, providing information and services that enrich users' experiences in their physical context just as the Web does in cyberspace. The final step comes when artificial intelligence reaches the capability of managing and regulating devices seamlessly and invisibly within the environment—achieving ubiquitous intelligence.

Ubiquitous intelligence is the final step of Larry Page's perfect search and the future of the Information Revolution.