



The Birth of Google

Larry thought Sergey was arrogant. Sergey thought Larry was obnoxious. But their obsession with backlinks just might be the start of something big.

By John Battelle

It began with an argument. When he first met Larry Page in the summer of 1995, Sergey Brin was a second-year grad student in the computer science department at Stanford University. Gregarious by nature, Brin had volunteered as a guide of sorts for potential first-years - students who had been admitted, but were still deciding whether to attend. His duties included showing recruits the campus and leading a tour of nearby San Francisco. Page, an engineering major from the University of Michigan, ended up in Brin's group.

It was hardly love at first sight. Walking up and down the city's hills that day, the two clashed incessantly, debating, among other things, the value of various approaches to urban planning. "Sergey is pretty social; he likes meeting people," Page recalls, contrasting that quality with his own reticence. "I thought he was pretty obnoxious. He had really strong opinions about things, and I guess I did, too."

"We both found each other obnoxious," Brin counters when I tell him of Page's response. "But we say it a little bit jokingly. Obviously we spent a lot of time talking to each other, so there was something there. We had a kind of bantering thing going." Page and Brin may have clashed, but they were clearly drawn together - two swords sharpening one another.

When Page showed up at Stanford a few months later, he selected human-computer interaction pioneer Terry Winograd as his adviser. Soon thereafter he began searching for a topic for his doctoral thesis. It was an important decision. As Page had learned from his father, a computer science professor at Michigan State, a dissertation can frame one's entire academic career. He kicked around 10 or so intriguing ideas, but found himself attracted to the burgeoning World Wide Web.

Page didn't start out looking for a better way to search the Web. Despite the fact that Stanford alumni were getting rich founding Internet companies, Page found the Web interesting primarily

10 Years That Changed the World

[Intro](#)

[We Are the Web](#)

[The Birth of Google](#)

A Decade of Genius and Madness

[1995: Marc Andreessen](#)

[1996: Jerry Yang](#)

[1997: Jeff Bezos](#)

[1998: David Boies](#)

[1999: Pets.com sock puppet](#)

[2000: Shawn Fanning](#)

[2001: Mary Meeker](#)

[2002: Steve Jobs](#)

[2003: Howard Dean](#)

[2004-05: Ana Marie Cox](#)

for its mathematical characteristics. Each computer was a node, and each link on a Web page was a connection between nodes - a classic graph structure. "Computer scientists love graphs," Page tells me. The World Wide Web, Page theorized, may have been the largest graph ever created, and it was growing at a breakneck pace. Many useful insights lurked in its vertices, awaiting discovery by inquiring graduate students. Winograd agreed, and Page set about pondering the link structure of the Web.

Citations and Back Rubs

It proved a productive course of study. Page noticed that while it was trivial to follow links from one page to another, it was nontrivial to discover links back. In other words, when you looked at a Web page, you had no idea what pages were linking back to it. This bothered Page. He thought it would be very useful to know who was linking to whom.

Why? To fully understand the answer to that question, a minor detour into the world of academic publishing is in order. For professors - particularly those in the hard sciences like mathematics and chemistry - nothing is as important as getting published. Except, perhaps, being cited.

Academics build their papers on a carefully constructed foundation of citation: Each paper reaches a conclusion by citing previously published papers as proof points that advance the author's argument. Papers are judged not only on their original thinking, but also on the number of papers they cite, the number of papers that subsequently cite them back, and the perceived importance of each citation. Citations are so important that there's even a branch of science devoted to their study: bibliometrics.

Fair enough. So what's the point? Well, it was Tim Berners-Lee's desire to improve this system that led him to create the World Wide Web. And it was Larry Page and Sergey Brin's attempts to reverse engineer Berners-Lee's World Wide Web that led to Google. The needle that threads these efforts together is citation - the practice of pointing to other people's work in order to build up your own.

Which brings us back to the original research Page did on such backlinks, a project he came to call BackRub.

He reasoned that the entire Web was loosely based on the premise of citation - after all, what is a link but a citation? If he could divine a method to count and qualify each backlink on the Web, as Page puts it "the Web would become a more valuable place."

At the time Page conceived of BackRub, the Web comprised an estimated 10 million documents, with an untold number of links between them. The computing resources required to crawl such a beast were well beyond the usual bounds of a student project. Unaware of exactly what he was getting into, Page began building out his crawler.

The idea's complexity and scale lured Brin to the job. A polymath who had jumped from project to project without settling on a thesis topic, he found the premise behind BackRub fascinating. "I talked to lots of research groups" around the school, Brin recalls, "and this was the most exciting

project, both because it tackled the Web, which represents human knowledge, and because I liked Larry."

The Audacity of Rank

In March 1996, Page pointed his crawler at just one page - his homepage at Stanford - and let it loose. The crawler worked outward from there.

Crawling the entire Web to discover the sum of its links is a major undertaking, but simple crawling was not where BackRub's true innovation lay. Page was naturally aware of the concept of ranking in academic publishing, and he theorized that the structure of the Web's graph would reveal not just *who* was linking to whom, but more critically, the *importance* of who linked to whom, based on various attributes of the site that was doing the linking. Inspired by citation analysis, Page realized that a raw count of links to a page would be a useful guide to that page's rank. He also saw that each link needed its own ranking, based on the link count of its originating page. But such an approach creates a difficult and recursive mathematical challenge - you not only have to count a particular page's links, you also have to count the links attached to the links. The math gets complicated rather quickly.

Fortunately, Page was now working with Brin, whose prodigious gifts in mathematics could be applied to the problem. Brin, the Russian-born son of a NASA scientist and a University of Maryland math professor, emigrated to the US with his family at the age of 6. By the time he was a middle schooler, Brin was a recognized math prodigy. He left high school a year early to go to UM. When he graduated, he immediately enrolled at Stanford, where his talents allowed him to goof off. The weather was so good, he told me, that he loaded up on nonacademic classes - sailing, swimming, scuba diving. He focused his intellectual energies on interesting projects rather than actual course work.

Together, Page and Brin created a ranking system that rewarded links that came from sources that were important and penalized those that did not. For example, many sites link to IBM.com. Those links might range from a business partner in the technology industry to a teenage programmer in suburban Illinois who just got a ThinkPad for Christmas. To a human observer, the business partner is a more important link in terms of IBM's place in the world. But how might an algorithm understand that fact?

Page and Brin's breakthrough was to create an algorithm - dubbed PageRank after Page - that manages to take into account both the number of links into a particular site and the number of links into each of the linking sites. This mirrored the rough approach of academic citation-counting. It worked. In the example above, let's assume that only a few sites linked to the teenager's site. Let's further assume the sites that link to the teenager's are similarly bereft of

10 Years That Changed the World

[Intro](#)

[We Are the Web](#)

[The Birth of Google](#)

A Decade of Genius and Madness

[1995: Marc Andreessen](#)

[1996: Jerry Yang](#)

[1997: Jeff Bezos](#)

[1998: David Boies](#)

[1999: Pets.com sock puppet](#)

[2000: Shawn Fanning](#)

[2001: Mary Meeker](#)

[2002: Steve Jobs](#)

[2003: Howard Dean](#)

[2004-05: Ana Marie Cox](#)

links. By contrast, thousands of sites link to Intel, and those sites, on average, also have thousands of sites linking to them. PageRank would rank the teen's site as less important than Intel's - at least in relation to IBM.

This is a simplified view, to be sure, and Page and Brin had to correct for any number of mathematical culs-de-sac, but the long and the short of it was this: More popular sites rose to the top of their annotation list, and less popular sites fell toward the bottom.

As they fiddled with the results, Brin and Page realized their data might have implications for Internet search. In fact, the idea of applying BackRub's ranked page results to search was so natural that it didn't even occur to them that they had made the leap. As it was, BackRub already worked like a search engine - you gave it a URL, and it gave you a list of backlinks ranked by importance. "We realized that we had a querying tool," Page recalls. "It gave you a good overall ranking of pages and ordering of follow-up pages."

Page and Brin noticed that BackRub's results were superior to those from existing search engines like AltaVista and Excite, which often returned irrelevant listings. "They were looking only at text and not considering this other signal," Page recalls. That signal is now better known as PageRank. To test whether it worked well in a search application, Brin and Page hacked together a BackRub search tool. It searched only the words in page titles and applied PageRank to sort the results by relevance, but its results were so far superior to the usual search engines - which ranked mostly on keywords - that Page and Brin knew they were onto something big.

Not only was the engine good, but Page and Brin realized it would scale as the Web scaled. Because PageRank worked by analyzing links, the bigger the Web, the better the engine. That fact inspired the founders to name their new engine Google, after googol, the term for the numeral 1 followed by 100 zeroes. They released the first version of Google on the Stanford Web site in August 1996 - one year after they met.

Among a small set of Stanford insiders, Google was a hit. Energized, Brin and Page began improving the service, adding full-text search and more and more pages to the index. They quickly discovered that search engines require an extraordinary amount of computing resources. They didn't have the money to buy new computers, so they begged and borrowed Google into existence - a hard drive from the network lab, an idle CPU from the computer science loading docks. Using Page's dorm room as a machine lab, they fashioned a computational Frankenstein from spare parts, then jacked the whole thing into Stanford's broadband campus network. After filling Page's room with equipment, they converted Brin's dorm room into an office and programming center.

The project grew into something of a legend within the computer science department and campus network administration offices. At one point, the BackRub crawler consumed nearly half of Stanford's entire network bandwidth, an extraordinary fact considering that Stanford was one of the best-networked institutions on the planet. And in the fall of 1996 the project would regularly bring down Stanford's Internet connection.

"We're lucky there were a lot of forward-looking people at Stanford," Page recalls. "They didn't hassle us too much about the resources we were using."

A Company Emerges

As Brin and Page continued experimenting, BackRub and its Google implementation were generating buzz, both on the Stanford campus and within the cloistered world of academic Web research.

One person who had heard of Page and Brin's work was Cornell professor Jon Kleinberg, then researching bibliometrics and search technologies at IBM's Almaden center in San Jose. Kleinberg's hubs-and-authorities approach to ranking the Web is perhaps the second-most-famous approach to search after PageRank. In the summer of 1997, Kleinberg visited Page at Stanford to compare notes. Kleinberg had completed an early draft of his seminal paper, "Authoritative Sources," and Page showed him an early working version of Google. Kleinberg encouraged Page to publish an academic paper on PageRank.

Page told Kleinberg that he was wary of publishing. The reason? "He was concerned that someone might steal his ideas, and with PageRank, Page felt like he had the secret formula," Kleinberg told me. (Page and Brin eventually did publish.)

On the other hand, Page and Brin weren't sure they wanted to go through the travails of starting and running a company. During Page's first year at Stanford, his father died, and friends recall that Page viewed finishing his PhD as something of a tribute to him. Given his own academic upbringing, Brin, too, was reluctant to leave the program.

Brin remembers speaking with his adviser, who told him, "Look, if this Google thing pans out, then great. If not, you can return to graduate school and finish your thesis." He chuckles, then adds: "I said, 'Yeah, OK, why not? I'll just give it a try.'"

From The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture, copyright © by John Battelle, to be published in September by Portfolio, a member of Penguin Group (USA), Inc. Battelle (battellemedia.com) was one of the founders of Wired. © 2007 CondéNet Inc. All rights reserved.

Use of this constitutes acceptance of our [User Agreement](#) and [Privacy Policy](#)