Preface

After the publication of *The Google Legacy: How Google's Internet Search Is Transforming Application Software* (2005), I spoke with dozens of people who dismissed Google's technology as "nothing special". This monograph presents information that provides additional insight into Google's inventiveness and its love of solving tough technical conundrums—what Google calls interesting problems—in elegant, clever, and mathematical ways. Google's catchphrase should be Eureka!

People have come to accept Google innovations as an everyday occurrence. By the time you read this, Google will have announced that it is mapping locations mentioned in books to Google Earth. So that a user can see them, although such capability may not be useful to the average Internet user, it is worth pondering that Google has the computational horsepower to perform a task that has never before in the history of man been attempted. Yet Yahoo's senior technologist told me, "Yahoo has better ways to deliver useful services." The comment may make the Yahoo executive get through the day, but Google delivers on a scale that is literally difficult or impossible for most people to grasp.

My main thesis is that Google's technology is special, and it makes the company a supra-national entity. Soon lawmakers will realize that Google will be challenging to regulate and control. Google's biggest vulnerabilities are its own management who must prove it is equal to the task of creating a billion dollar company without self-destructing meanwhile, competitors, unable to respond technically, are making an effort to crush Google with litigation.

This monograph is called *Google Version 2.0: The Calculating Predator* for a reason. Google is anchored in mathematics. Indeed, one can't appreciate what Google does or how its management team directs research, motivates engineers, or intimidates competitors without understanding mathematics.

Mathematicians and Strategists Calculate

Mathematicians calculate. I don't mean add or divide numbers. The sense in which I am using *calculate* embraces two meanings. First, calculation is a logical process or the use of a method that allows a mathematician to figure out that light goes 186,000 miles per second or that $E = m c^2$. Second, calculation means that one can determine by logic what action to take. Google is a calculating predator because it uses math as the foundation of much of its most sophisticated technology and because it plans its moves, actions, and tactics in a strategic way. Google doesn't do much by accident. Forget the popular media image of a bunch of crazy college kids. Google is managing its image in order to misdirect competitors, vendors, and even its users.

Google is an applied engineering company, but as you will learn in this monograph, Google blends several innovation techniques. Innovation at Google is the fuel needed to power the Googleplex and to satisfy Google's hunger for ever more powerful, capable systems and software.

The inventions described in this monograph, may remind you of other companies' innovations. For example, Google has invented "containers", a Google term for a variant of a virtual machine. The invention sweeps up personalization functions, portal technology, and repurposing of information in one basket.

Google can be viewed as an inspired copycat. Google is an information company, and the firms employees suck and filter information on an industrial scale. Google engineers are adept at combining insights of others and their own insights to create new constructs. Much work takes place in teams operating in a self-created pressure cooker to solve a problem. Not surprisingly, individual Googlers often have no broader conceptual picture of the overall company's direction.

Orchestrating the innovation, the flow of products and services, even the acquisitions like DoubleClick are the founders—Sergey Brin and Larry Page—and the duo's hand-selected CEO, Eric Schmidt. The Google public relations and marketing professionals work very diligently to make these three people "just folks" a tough job. Messrs. Brin, Page, and Schmidt are engineers most comfortable with other engineers. One of my researchers quipped, "Googlers are nerds of a feather". Clever and apt was this turn of phrase. Two of the triumvirate—Messrs. Brin and Page—are also mathematicians, most relaxed with eggheads at Stanford University or the Moscow State University.

How I visualize Google's management team as friendly, "calculating predators".

Competitors don't appreciate the fact that Google's top guns are mathematicians who calculate....for example, how to take a big bite out of traditional telecommunication companies' revenue.

In 1955, Martin Heidegger identified a type of thinking characteristic of modern life. Like Jacques Ellul and other thinkers, he said that technology poses practical and philosophical challenges to society.[1] In a little-known speech given to commemorate the composer Conradin Kreutzer, Heidegger said that thoughtlessness was part of modern life, noting, "Man today will eventually deny this flight from thinking." Heidegger's concern is that thoughtlessness allows calculative thinking to gain the upper hand. He adds:

> Calculative thinking computes. It computes ever new, ever more promising and at the same time more economical possibilities. Calculative thinking races from one prospect to the next. Calculative thinking never stops, never collects itself. Calculative thinking is not meditative thinking, not thinking which contemplates the meaning which reigns in everything that is."[2]

Google is perhaps today's best example of a company built on *calculative thinking*. Characteristics of calculative thinking include efficiency and logic, not emotional reactions. An elegant proof of a theorem bundles intelligence and beauty into a construct of great beauty. Google obviously needs revenue to expand; therefore, markets with inefficiencies that can be made more efficient with Google technology are logical targets. Analog telecommunications companies, for example, become prey to Google's more efficient technology. Like a grand master in chess, Google uses strategic feints to obtain its objective—winning the game.

Goals of This Monograph

One goal of this monograph is to make it easier for a reader with some knowledge of technology to understand that Google can do a great deal with its "as is" infrastructure. Its inventions in many cases do little more than load a new feature into the Googleplex. This new feature, like its telecommunications inventions, allow the Googleplex to function like a traditional telecommunications company. But Google can also be an Amazon or eBay, a bank, a video distribution company, and much more.

1. See Jacques Ellul, *The Technological Bluff* (William B. Eerdmans Publishing Co., 1990), translated by Geoffrey W. Bromiley.
2. Martin Heidegger, "Memorial Address," *Discourse on Thinking* (Harper Collins, 1959), translated by John M. Anderson and E. Hans Freund.

A second goal is to reveal that Google is different from other companies providing Web search and network services. Like Wal*Mart, Google operates at a scale beyond that of most organizations. Google addresses issues germane to petabytes of data and billions of queries per day; it is difficult to think about numbers this large. Understandably, Google's competitors tell me that Google is operating in the digital equivalent of an LSD hallucination. Microsoft, for example, struggles to understand Google's approach because Microsoft views the world in terms of licensed software, Office upgrades, and desktop or mobile devices running a version of Windows. Oracle sees Google as a partner, telling me, "Google's Bigtable is not a commercial product." IBM contacts tell me, "Google is an advertising company and has no credibility in the enterprise software market." Maybe they are right and I am wrong. I think I am right, but time will tell.

Many people dismiss Google because what the company has built is easy to duplicate.

A third goal of this monograph is to reiterate that Google, unlike Amazon or Yahoo, is built on mathematics, not engineering. The "guts" or inner core of Google is math, and the math that Google employs is extremely advanced. Forget Statistics 101 and Introductory Calculus; Google's use of mathematics pushes into the farthest reaches of number theory and computable relationships. Think in terms of hyperbolic geometry, the Lambda calculus, and even more abstruse techniques.

The title of this monograph is intended to communicate Google's mathematical foundation and its nature as a predator. It does eat markets and competitors. The turmoil at Yahoo is but one example of how Google copied the Overture online advertising model, tweaked it, and imbued it with predictive analytics. Google tossed in technology acquired from Applied Semantics (formerly Oingo), and the rest is financial history. Yahoo is, for all practical purposes, a distant second to Google suffering from Wall Street's disdain, and is for sale.

Fast Search & Transfer, once a leader in the enterprise search sector, finds itself in financial difficulty. Google's entry into enterprise search has disrupted that market. Fast Search did not diversify its revenues as did market leader Autonomy. As a result, the Google Appliance has sucked oxygen from Fast Search's sales efforts. Even Autonomy is fighting to survive. What will happen when Google gets serious about the enterprise software market? It has reshaped a once-dynamic sector, and Google was running a test, not a marketing campaign.

Heigegger's words, penned more than 50 years ago, could describe Google today. A short time ago, a senior executive from Sprint, one of the U.S. telecommunications companies partnering with Google told me: "We think Google is making up its telecommunications strategy as it goes along."

The statement is uncannily accurate. Heidegger makes explicit that calculative thinking leads to racing "from one prospect to the next". But it is also naive. Google is no newcomer to telecommunications. The company has more than a dozen patent applications related to telephony. These range from transceivers to a method of performing medical diagnostics via a wireless device. Another dozen patent applications relate to functions required for telecommunications services; for example, a system and method for efficient routing using a variation of the Hadamard function.

More telling, Google's interest in telecommunications reaches back to 1999, when the first telephony-related patent application was filed. Over the last eight years, one assumes that Google would do more than randomly wander from telco idea to telco idea. The march toward telecommunications is a core interest of Google's management, and it now pervades Gmail, individualized Google, Google's capturing digital images of storefronts, and its wireless initiatives with Apple Computer, the Spanish company FON, and the sharp-minded Sprint.

The data about market share from Nielsen NetRatings, Comshare, ZiffDavis, and other firms engaged in

reporting the scorecard for public search engines report widely varying data. But there's one thing on which the ratings services do agree. Google dominates online advertising. Google is so important to some businesses that established firms like BMW and Ricoh engage in Web page trickery to boost their Google ranking. A company that's not in the Google index may be unfindable.

Pragmatism and Cleverness: Obviously Rational Behavior

It is a fact that Google's core money machine is online advertising. Did Google invent online advertising? No, Google's engineers were inspired by the Yahoo-owned Overture system. Google then paid Yahoo about $1 billion prior to the Google IPO to end litigation related to Google's enthusiastic emulation of the Overture system.

Google filters through most *interesting* innovations in computer science, information retrieval, and allied disciplines.[3] Then, with uncanny astuteness, Google seizes on the *interesting* part, leverages or reshapes it, and integrates the breakthrough with existing Googleplex functions.[4] Google innovates by building out functionality, using the search technology and advertising technology as its foundations.

It may turn out that only a country like China or India can "catch" Google. Individual companies are likely to have a tough time competing with Google. Google's engineers, motivated with Google stock units and a innate thrill of working on interesting problems, continue to leapfrog forward. Whether Google's engineers in Kirkland, Washington, solve a problem that befuddled Google's engineers in Moscow is irrelevant. There's another problem to solve, another company with smart people and interesting technology to acquire. Google uses a Lego block or Lincoln log approach, refined and fine-tuned over the last nine years. Unlike companies that have a push cart filled with technology junk, Google operates a race car with custom gear.

Google is not flawless. The complex systems can hiccup. Google's personalization service available at www.google.com/ig malfunctioned once, and thousands of users lost their custom settings and gadgets. Google Checkout can mysteriously "lose" an order. Gmail on some public terminals can leave a user's email session open and available to the next public terminal user. Nevertheless, Google's system operates reliably and remains comparatively bug free, secure, and blindingly fast, 24x7, regardless of the user's location.[5] More important is that Google's ability to innovate has continued to send a stream of new and modified products to users as "beta tests".

In my conversations with Google professionals and in my often-contentious encounters with Google wizards, I learned something new —some Google employees don't know much about Google's core technology. These Googley folks are smart, but Googlers work in silos, often unaware of other activities in the next four-person cube. The information is available on Google's Intranet, in popular books about Google by professional journalists like John Battelle and David Vine, but Googlers are too busy or too immersed in their own technology pool to worry much about the bigger picture.

3. The word *interesting* is Google-talk for a very important idea or issue. Google concerns itself with interesting problems, not with trivial ones. It approaches these problems in the way a math wizard tackles a tough problem—part intuition and part sheer effort. Google's engineers manifest the precept attributed to Thomas Edison: "99 percent perspiration and one percent inspiration".

4. The word Googleplex is used in this monograph to refer to the Google infrastructure, including its software, its hardware, its network, and its more than 30 data centers. Google, despite its growing range of desktop applications, operates from the "cloud"; that is, over a network connection. The Googleplex works like an ordinary desktop computer imbued with the fire-breathing capabilities of a supercomputer operating globally as if were under a user's desk or running inside a mobile phone.

5. One exception is accessing Google.com from China. Mysterious time-outs and inaccessible pages plague even casual Internet users. It's difficult to determine if this is a Google problem or a behavior imposed by "the Great Firewall of China".

Competitors are equally flummoxed. I was once asked by a client to prepare an executive memo on “How does Google innovate?”

I summarized what I had learned in my deep analysis of Google over the last five years and explained that Google was not quite as slap-dash as the company encourages outsiders to believe. The facts indicate that Google does a very good job of organizing its research, technology, and product activities. If anything, Google is a savvy, logical, calculating competitor. Google stalks a market, an acquisition candidate, a potential recruit. Google then strikes, quickly and in a cold-blooded way. My report emphasized the calculating and predatory aspects of Google’s innovation process. But this client rejected my analysis, preferring to believe that Google was fumbling its way forward, fueled by online advertising and Web users who didn’t know better.

Readers of this monograph will have an opportunity to decide for themselves. Lucky nerds or calculating predatory?

Flying with Open Source Instruments

Regrettably, Google provided me with zero information. I simultaneously shipped to my Google contacts and to my attorney each draft chapter in this monograph. I invited constructive criticism, additonal information, informal guidance. I do have a few Google emails reminding me that I have no official standing with regard to Google or its technology.

Then after the May 2007 Bear Stearns’ report *The Semantic Web*, based on information appearing in this monograph’s analysis of Google’s Programmable Search Engine, I learned from attorneys interacting with Google, “He is not incorrect.”

Where does the information in this monograph originate? The answer is, “Publicly-available information. In most cases, the information, is accessible via the Google index.” Google, I discovered, hides its information using code words or makes certain types of information hard to find. For example, if you don’t know the code word Sawzall, you will have a difficult time getting examples of code in Google’s database query language. Sawzall is the brand name of a popular power saw. So a casual Google search buries the programming information so deep in a Google result list as to be effectively hidden. Clever stuff?

Anyone can track down and read the source material I used in this, my second monograph devoted exclusively to the bits and pieces of Google ignored by the thousands who write about the company. Specifically I read Google’s technical papers conveniently listed on Google Labs’ Web page. In some cases, professional societies demand an outrageous sum for a document stuffed with equations, graphs, and references to hyperbolic geometry. But dozens of papers are available without charge. These are a treasure trove of information.

The Google technical papers include charts, graphs, and tables revealing some of the astounding performance metrics the company’s Googleplex delivers. In *The Google Legacy,* I did not use Google’s data because one reader said that it was unbelievable. I dutifully modified the file read speeds so Google’s system was only two or three times faster than the fastest systems then available for high-end systems. For this monograph, I did not include the most recent data that revealed Chubby (Google’s database file locking system) can outperform Oracle’s technology by a factor of 20. No one believes Google’s data, and I am tired of 25-year-old “experts” reminding me that Google can’t defy physics.

I also track of the United States Patent and Trademark Office’s patent applications and patents filed by Google. Google uses many different legal firms, consultants, and in-house lawyers, and some of these patent applications are not identified as belonging to Google. The give-away is the name of the inventor. In some cases, the application will include screenshots with a Google logo. Often a patent will be granted and only then will the assignee be revealed as Google. I believe that my collection of about 300 Google patent

applications and patents is about 70 percent complete. Rarely does a week pass without my researchers sending me one new Google patent application or patent.

The company maintains brief biographies of elite engineers. These provide some information. We chase down technical papers, Ph.D. dissertations, and similar information about certain Google professionals. For example, we have learned that when Jeff Dean's or Peter Norvig's name appears in a document, that document is worth examining. Amazing nuggets of information pop out. We first learned about the Bigtable technology in a technical paper written almost two years ago. The clue to this paper's existence was a blog posting by an engineer following a particular research area.

In this monograph, I fuse two incompatible types of information. First, I have the patent applications, patents, and technical papers. Then I have the "soft" discussion of how selected technologies disclosed in the public patent documents seems to relate to Google's capabilities, business tactics, and products and services.

Anyone reading the patent discussions may find that Google's technology is more sophisticated than counting inbound links to a Web page and figuring out how many times people click on content about Paris Hilton or a Honda CRV. We find that the patent applications are often accompanied by a "stub" or portion of the disclosed invention in one or more of the Google products or services. A good example is the "individualized Google" service. The "ig" service became available as a beta six months before the containers patent application was published by the USPTO. Many of Google's inventions, therefore, are visible or tantalizing glimpses of their functionality, available in the products and services you can use today.

We believe Google's patent activities and its formal technical articles in respected journals provide a window into Google's advanced engineering division. The beta products are demonstrations or tests of advanced technologies. If users click on the beta products or make use of the beta services, Google pays attention to guide subsequent activities for their technology. A good example is the use of Chubby and Google's mapping and satellite imaging products.The system and method for the technology making this possible is described in the Google technical documents. The writing is not intended for the casual reader, but effort invested in reading these Google materials provides useful information about Google.

The second element in the monograph is relating the advanced technology to Google's business tactics and strategies. There is very little information from Google in the Securities & Exchange Commission filings, the popular press, or in Google executives' public statements. Nevertheless, it is possible to look at what Google is doing with its more than 80 products and services, at its application programming interfaces, at its general business moves such as expressing an interest in participating in the Federal Communications Commission bandwidth auction, and at what competitors are doing in response to Google's actions. Each of the patent discussions is introduced by a business commentary and followed by a discussion of the possibilities the technologies described contain.

I believe that Google is inherently unpredictable. Therefore, some of my descriptions of technology will be general and in some cases incomplete.

Nevertheless, analysis of Google's inventions in certain key areas such as publishing and virtualization, coupled with consideration of business implications, is useful. To my knowledge, no one has attempted the type of invention deconstruction that I have undertaken. Also, no one has yet taken the Google technology and made an attempt to characterize what functions it can support. The nature of patent documents is to describe certain information in ways that can be quite difficult to understand.

I don't pay any attention to the legal validity of Google's patent applications or the patents themselves. I focus on the technology and relate that technology to Google's "as is" infrastructure and systems. The idea is to understand to some degree what Google can do with its more than 30 data centers, 465,000 servers,

and its purpose-built operating environment.

Google's marketing and PR people do a very good job of distracting analysts and journalists from the core activities of Google. Messrs. Brin and Page squabble about beds on Google's corporate jet. Eric Schmidt makes references to Google's interest in having wireless Internet access become a public good. A news story about Google is not complete unless a Googler is lounging on a bean bag chair or indulging in a gourmet, vegetarian meal with a chilled bottle of Odwalla juice in hand.

Readers who want in-depth business analyses of Google's advertising model or anecdotes that talk about "Sergey and Larry eating pizza in a Palo Alto restaurant" will be disappointed in this monograph. Google is a company founded by engineers, run by engineers, and built around the needs of engineers. It's great that Google makes customers happy and thrills thousands of advertisers with increased interest in their products. At the end of the day, Google wants its scientists, mathematicians, programmers, and computational linguists to be productive. A less forgiving soul than I might describe the majority of Googler as geeks, nerds, rocket scientists, or eggheads. I prefer to think of Googlers as computational and technical magicians. Several illustrations in this monograph show Google's founders against an image of Albert Einstein or as a magician holding fire in his palm. The idea is that Google is one of the most interesting manisfestations of technology I have encountered.

Google Version 1.0 was the Web search and advertising company. The Version 1.0 image continues to shape thinking about Google. Google controls more than 60 percent of search traffic so an Internet user is more likely to use Google than Live.com, Yahoo.com, or Ask.com.

Google Version 2.0 is another creature entirely. I use the term *Googzilla* to describe the current incarnation of Google. The idea is that Googzilla is big, powerful, and indifferent to the insects and ants crushed by its massive paws. Google Version 2.0 shares the search and ad capabilities of Google Version 1.0, but it is much, much more. In this monograph you will learn that Google could, with little effort and time, become a full-fledged scientific and technical publisher. With a little more effort, Google could become the largest information publishing, distributing, archiving, and retailing operation in the world.

The Google Stack: A New Type of Operating System Arrives

The company's infrastructure makes it trivial for Google to load different instructions into the Googleplex. These instantly imbue Google with the ability to operate as a publisher, financial institution, or media company. Just as a new software program allows you laptop to perform new functions, Google's infrastructure can deliver functions other than search and retrieval.

The illustration below provide's an overview of what I call the Google Operating Environment or GOE. This is a representation of how Google's inventions fit together. Google has not made available this type of representation. I have simplified many details revealed in Google's public documents to highlight the main features of what I call Google's "as is" operation.

The inventions and my attempt to explain how they work and what functions each supports fit into this stack. Some like Dr. Guha's Programmable Search Engine provide operating system functions as well as functions that can be accessed by the APIs and variants surrounded by a broken line. Others operate behind the scenes like Dr. Patterson's "Xift" variants. And others like the extended database inventions allow Google to make another device function as part of a Google grid or become an extension of the Googleplex itself. Even something as trivial as a personalized interface displayed to a single user taps into services and functions throughout the stack. Google has developed a next-generation applications platform, development infrastructure, and "wrapper" architecture that gives the Google engineers greater degrees of freedom than engineers at Amazon, eBay, Microsoft, or Yahoo enjoy.

Google has a big vision. "All" the world's information remains the goal. That means the Internet,

enterprise services, telephony running on Internet Protocol, and the other functions that can be made more efficient by shifting to a global, digital computational system. Keep in mind that Google has been working on these technologies in a calculated—that is, mathematical—way. Google moves forward a step at a time until the problem is solved.

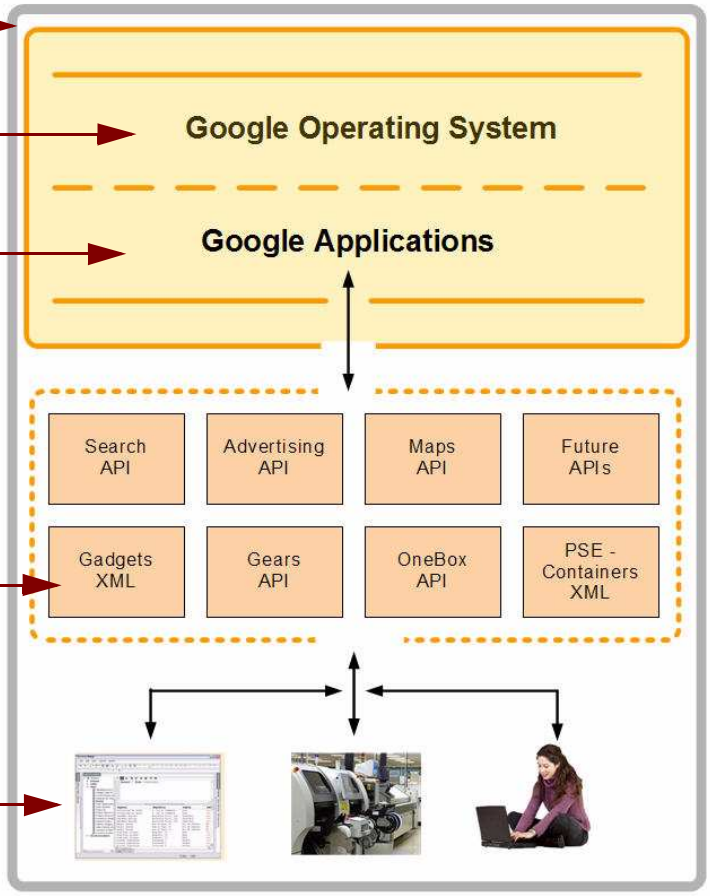
The gray border represents a user accessing a Google service or running a Google desktop application “stub”.

The operating system is Linux, Google-coded extensions, and core functions such as the Google File System or Bigtable.

The Google applications like Google Docs run on the Googleplex on the GOS.

The APIs and their variants allow developers to customize interaction with the Googleplex.

The “user” may be a process, a legacy system, or a person connected to a network. The user can use XML to instruct the Googleplex to take certain permitted actions on behalf of the user.



There is no Google operating system. There is a stack, and Google has virtualization technology that makes it possible for any device connected to the network to work as a virtual machine. Google’s vision, if carried to its logical end, is that Google becomes the Internet.

Google has most of the pieces in place and operating. Engineers call this an “as is” system. *As is* means that Google has the environment online, operational, and in daily use. Most of Google’s competitors are in the “to be” phase of their development. Based on my analysis of Google’s technology and its principal competitors, Google enjoys a technical lead time of nine to 24 months as of August 2007. Google’s lead time may be greater because some competitors like Amazon make innovations by “hacking” a relatively narrow fix to a known problem. Google, on the other hand, solves major problems such as the stability and performance of traditional relational database design. So what a competitor touts as a solution may be a stopgap. So, in a sense, the larger engineering gap between Amazon and Google is not being closed in a significant way. Startups say they are “like Google” or will be “Google killers”. Venture firms may buy this line, but I don’t. Google’s advantage is its engineering foundation, rooted in math, hardware engineering, and software engineering. Google is not an integration company. Google is an applied engineering company, maybe the most skilled applied engineering company since Thomas Edison cranked out inventions more than a century ago.

Upside and Downside of Google's Dominance

Google has an upside for users. The petabytes of information reachable by a browser can be accessed easily and quickly. The massive infrastructure required to index more than 20 billion Web pages costs most users nothing. The bill is paid by advertisers eager and willing to pay as much as $150 for a single "click" on an advertisement.

Google has used its billions to reinvest in its infrastructure. The patent applications and patents described in this monograph are fungible evidence that:

- Google tackles difficult, fundamental computing problems, often from several different technical starting points. For example, personalization of information services can be implemented using sophisticated tracking methods and / or with predictive statistical systems. Some operate at command-central, the Googleplex itself. Others are triggered by a user's device, reporting the user's location and present information actions.

- Google focuses on economies. It is the one online company that embraces the efficiency of the "calculative thinker". Efficiency at Google does not mean reducing costs by eliminating certain expensive procedures; for example, improving the efficiency of a programmer by shifting certain functions to the Google operating environment so the programmer doesn't have to parallelize certain functions manually. Google focuses on making the overall system perform more calculations in a specific time interval or reading data at a mind-boggling terabyte per-second gulp.

- Google embraces scale. Wal*Mart and Exxon are similar to Google because these two companies understand and nurture scale. Wal*Mart's dominance of retailing is a consequence of doing certain functions very well across a large range of tasks. Exxon handles energy at a comparable scale. Competitors find that companies operating at scale are difficult to understand, dislodge, and explain. Scale distances competitors technically, financially, and conceptually. In online, Google scales. Yahoo cannot scale, and the turmoil at the company, indeed within its ad business that Google copied and then improved upon, makes it a distant second to Google for the foreseeable future.

Of course, there's a downside to the emergence of Google and its scalable network potential. This monograph will attempt to explain three key issues.

First, most observers and users of Google have absolutely no notion of what Google can do as a company. Google has woven search and advertising into what I call the DNA of the company. As a result, Google has the luxury of developing business models that can rely on advertisers to foot the bill, charge users for services just as conventional Google competitors generate revenue, or blend the free and the fee in different ways. Competitors lack this flexibility. Microsoft may talk about delivering ad-supported services, but the reality is that Microsoft's $50 billion in revenue depends on license fees. Wall Street won't stand for Microsoft cannibalizing its existing revenue streams in a significant way. Result? Google is operating with more freedom than Microsoft, and it does not have to spend money to support its legacy applications. So, on both the revenue and cost side of the balance sheet, Google has an advantage.

Second, Google's infrastructure costs money. Upwards of $2 billion a year is invested in data centers, research, and engineering at Google. Amazon, by way of contrast, invests in the range of $350 to $450 million range. Google is, as you will learn, running an innovation process that combines elements of the legendary Bell Laboratories' "blue sky" invention with the applied innovation honed to razor sharpness by the 3M Corporation. A Google invention is characterized by its combination of insight (the Hadamard

function) with the other components of the Googleplex (the Google File System, the Chubby locking system, Google's MapReduce libraries, etc.). Google is building out its infrastructure, using its massively parallel distributed computing environment to deliver a wide range of applications. In contrast, Amazon looks to university graduate students to find individuals who have a clever work-around to a problem. To illustrate: consider Amazon's approach to databases. Amazon uses the Oracle database as its information management engine. Google uses Bigtable (a multi-dimensional database that consists of infinite rows) and various traditional SQL databases like MySQL. Google engineered its file system, its master-slave architecture, and its storage strategy to deliver scaling, speed, and flexibility. Amazon's approach relies on commercial-off-the-shelf software and is subject to the same bottlenecks that plague anyone using a Codd database to manipulate many terabytes of data. Amazon's solution was to use a variation of the content based caching technology to create a wrapper—a layer of middleware—to get more mileage out of centralized datastore design. Amazon's solution was cheaper, and it delivered a needed performance boost with an increase in reliability. The downside is that Amazon's solution lacks the flexibility of the Google solution. Going forward, Google's willingness to tackle and solve difficult computing problems gives it an edge over Amazon. Amazon describes itself as a technology company, but the description is similar to a high school football team saying that it is as good a professional team as Manchester United.

Third, Google solved the problem of search and then grafted a variation of the Overture ad system on the original Googleplex plumbing. Google discovered that by solving the problems associated with search and advertising, its engineers had stumbled on one of those happy, unexpected circumstances in science. The original solutions proved to be more widely applicable than the original inventors ever intended. Google, in effect, found a solution that could support many other, unrelated problems. To the outside observer, Google seems to be a company able to innovate in hours or days. When the rumor of eBay's acquisition of StumbleUpon.com surfaced, Google had its own version called Recommendations deployed and publicly available within four hours. In the world of online innovation, Google is not just fast, Google is operating with an efficiency that is not often encountered in any business. For a company of its size, Google remains remarkably nimble and surprising. Google can deliver a great many products and services from its malleable framework. Telecommunications, motion pictures and entertainment, financial services, and publishing are just a few of traditional market sectors that Google can enter and disrupt without too many technical gyrations. Most of these market sectors are blissfully ignorant of Google's capabilities.

I had a wealth of information about Google's technical inventions. The table below provides a summary of content we analyzed but did not include in this monograph. Perhaps in a few months I will convert our analyses and PowerPoint briefing slides to narrative form.

As you might imagine, after my bout with cardiologists in February 2007 and then the push to create the chapters in this monograph, I'm not bursting with enthusiasm for another long slog through Google's technology, running gun battles with Google's attorneys, and the Type A fretting I hear from investment analysts convinced that Google sells ads and fuels the search engine optimization industry.

In this monograph, you will not find detailed information about a number of very important Google inventions. Perhaps later?

Table 1: What Is Not Included in This Monograph

Subject	Comment
AdWords and AdSense	Like search engine optimization, the subject of spending money and earning money via advertising is of little interest to me. We do reference certain inventions that allow Google to do a better job for its advertisers. For information on these topics, contact an ad agency.
Descriptions of Google's products and services	Popular books like *Googlepedia: The Ultimate Google Resource* by Michael Miller now in its second edition ISBN 0-7897-3639-X exhausts this topic
Getting a job at Google	This is easy. Be really smart and do something that Googlers find useful. Otherwise, forget it. No need for elaboration on this topic.
Google and telecommunications	I have a separate 50-page monograph in preparation on this topic. The complexity of the 14 Google inventions in this field requires a different editorial approach and more complex diagrams than possible in this monograph.
Google Applications	Google offers a wide range of applications and is edging closer and closer to major actions in this area. The technology of the applications is in this monograph. I chose not to elaborate on Google's inevitable push into dataspace services, customer relationship management, and desktop applications like word processing and mail.
How Google search works	Google does not have a single approach to search.for general information popular books by John Battelle, David Vine, and others do a good job providing this information. In a pinch, consult Wikipedia.
How to make money via Google	The focus of this book is on Google's technology. Although there are discussions of the business implications of certain Google actions, I excluded the type of information in such books as *The Five Dysfunctions of a Team* or *Leadership Secrets of Attila the Hun*
Privacy and Google	Our briefing on Google's ability to track individual users triggers strong reactions. Our focus is on the technology, not its social and legal consequences. We decided to omit these "softer" topics. If you are concerned about privacy, don't use the Internet without anonymizing tools.
Programming for Google	Other publishers cover the basics of programming languages that Google supports. I have included some code snippets to show what Google is encouraging. I recommend titles from O'Reilly Media or comparable high-quality technical publishers. Google provides a wealth of information at http://code.google.com
Search engine optimization (SEO)	This is today's jargon for creating Web pages with information and useful index terms. As a matter of principle, I find this topic distasteful and suggest that you attend a Search Engine Strategies conference or hire one of the companies advertising on the Web logs SearchEngineLand.com or SearchEngineWatch.com
Security at Google	Google runs a tight ship. Security breaches do occur, but in most cases, they are quickly patched. We do not include information about what Google does and how the company does it. We do have these data, but the information is not germane to the central argument about Google's core technology.

Let me add that I do have briefings on these topics. If you are interested in learning about these topics, let me know. However, I don't plan on writing a monograph on these topics. Contact me by sending an email to sa [at] arnoldit [dot] com.

Google beyond 2007

A competitor with enough courage and money could arguably wound, even kill Google. But Google's greatest weakness is itself. The company must demonstrate that it can manage the large number of brilliant people it employs. Google must recognize that lawyers can be as deadly as a fracture in a management team. Google, therefore, is not invincible, nor will its upward trajectory continue at the blistering pace set in the post-IPO to 2006 months. Google also won't go quietly into that good night. Companies choosing to underestimate, ignore, or assume that Google is a one-trick pony are likely to find themselves surprised by Google's "calculative" and predatory actions.

Readers of this monograph may wonder why certain topics such as telecommunications are given short shrift. Others may puzzle over the omission of detailed analyses of Google Docs, Google's application programming interfaces, or explanations about how to use Google's command language to perform better searches. The answer is that I have had to select from an overwhelming amount of information.

In consultation with my colleagues and publisher, I decided to focus on technologies and issues not covered in other books, popular and trade press articles, and the many Web logs that track every twitch and fidget of Google.

I apologize for the complexity of some of the topics explored. I have tried to catch errors and asked those who have read drafts of the monograph to identify mistakes and problems. Any errors are mine. I welcome communications from readers, in the interest of accuracy.

My hope is that others will build on this work, using Google's technical information as a viewport into the inner workings of a company that has become a world leader in online technology. I have been able to cover only a small portion of Google's mosaic of technical invention. I have a deep respect for the computational basis of much of the company's success.

I also have learned that Google was the right company at the right time. Few appreciate that Sergey Brin and Larry Page were aware of the innovations in search that powered AltaVista.com. Hewlett-Packard, a great company in its own sector, had no clue about the importance of the AltaVista.com technical innovations. Messrs. Brin and Page did. When capital became available to Google's founders, Messrs. Brin and Page were quick to hire Jeff Dean, Sanjay Ghemawat, and others from the AltaVista.com team. These people brought more than five years' hands-on experience with the problems of search. Furthermore, it is not widely known that much of the Googleplex flourished in the fertile environment of the AltaVista.com team.

At the moment of Google's founding, Excite, Lycos, and Yahoo—once leaders in Internet search—decided to become portals. In one of the great ironies of online, Google has circled back to the portal idea, blending it with virtualization technology. Now, Google's competitors are rushing to close the gap in search as Google reinvents the portal without the engineering weaknesses of those early 1998 efforts. Google's success was given a boost because search competitors took their eye off the ball. Google hit the ball out of the park.

Messrs. Brin and Page, whether by luck or good judgment, determined that algorithms and mathematics provided the intellectual foundation for processing large amounts of content. The original PageRank invention was a blend of bits and pieces from different disciplines. The mix was unique, and Google demonstrated that most users were able to find useful results from the Google search engine with little effort on their part. Type a word, hit enter, and pick a relevant result. Simple. But under the hood, the

mathematical engines of Google were forcing its engineers to find ways to deliver sub-second response time regardless of the demand on the system. Software and hardware engineering fused to deliver the computational engine needed to deliver on Messrs. Brin's and Page's insight regarding information retrieval.

Finally, after Google shot to the number one spot in search engine usage, competitors and competitors-to-be swallow Google's PR KoolAid® in thirsty gulps. Google has great food. Google people ride Segways. Google hold great parties. Google is a college toga party decorated with cheerful primary colors. Google is a big, disorganized nerd fest. Readers of this monograph will learn that Google is calculating. It uses math; it embraces math; and it values the objective outputs of math. It is also calculating because Google knows that it wants to be a $100 billion company. Google wants to be the next supra-entity, wresting revenue from different Internet protocol-based operations. Google has evolved into a teenage Googzilla, and it is a calculating predator. Math does not understand such sophisticated notions as good and evil.

I want to thank the many people who provided information on and off the record. Several individuals warrant a special thanks. These include Erik S. Arnold, my son, who talked with me about the ideas in this monograph and then read each of the sections in various forms. My long-time colleagues—Don Anderson, Tony Safina, and Stuart M. Schram IV. Each of these people worked many hours to assist me in finding, cataloging, analyzing, and describing Google's inventions. Kay Arnold and our wonderful boxer, Tyson Arnold, for sticking with me through a 24-month process. Late nights and missed dog walks roiled their days, and I apologize for my forgetting the normal world when digging into Googzilla's innards. Finally, Ulla de Stricker, to whom this study is dedicated. For more than 25 years, Ulla has discussed, critiqued, and revised my drafts for this monograph. She is Canada's most accomplished information expert, and she is an editor without parallel.

Stephen E. Arnold
August 2007
Harrod's Creek, Kentucky