



Next Generation Search

The hope, the hype and the chutzpah. Barney Pell, co-founder of Powerset, discusses why he believes that natural language search will be the next advancement in Internet search capability.

by Mark Duncan, askmar, www.askmar.com

Barney Pell, CEO and co-founder of Powerset was the lead speaker for the MIT/Stanford Venture Lab talk on March 20, 2007. Mike Sigal, CEO and co-founder of The Guidewire Group was the moderator. The three panelists were Christine Herron, director at Omidyar Network, Phyllis Reuther, CTO of Mobile Content Networks, and Munjal Shah, CEO and Co-founder of Riya.

Mike Sigal began the talk by noting that he did 15 minutes of Google searches to try and determine the 2006 search market size and its NASDAQ market capitalization and been unsuccessful in that time, citing this as an example of the need for better search. He noted that the Guidewire Group recently in reviewing the plans of a thousand startups, found that 25% of these companies had plans involving search; vertical search, horizontal search, interfaces optimized for mobile media, et cetera.

Powerset is a San Francisco based company with 40 employees. They are currently hiring 5 to 7 new people each month. It has raised over \$12M in capital from venture capitalists (The Founders Fund, CommerceNet, Amidzad) and angels. It recently acquired an lifetime, exclusive license to the natural language technology developed by PARC (Xerox Palo Alto Research Center) over the past 30 years. This technology includes a working code base and 60 core patents. They anticipate having a public beta at the end of 2007.

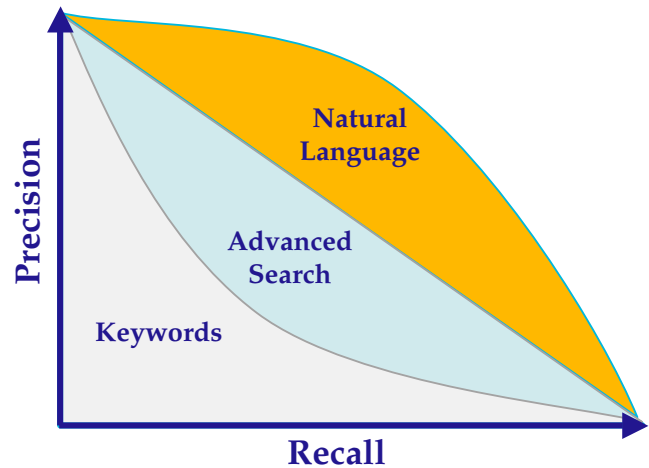
Natural language is computation intensive. Previously it took 5 to 30 minutes per processor for a given request. It is only in the past few years that the steady increase in processor improvement driven by Moore's Law has made natural language searches feasible, causing PARC researchers to say, "We're ready!"

Barney observed that the worldwide search market was \$4B in 2004 and is projected to grow to \$30B in 2010. Powerset plans to use an advertising business model since it is mature and proven. They see keyword based search engines as having reached a plateau, and that natural language processing is required to move beyond this.

The last 15 years have seen numerous search breakthroughs, but Google's use of keywords and page rank in 1999 was the most recent search innovation. Powerset believes its natural language represents the next major breakthrough for search in 2007.

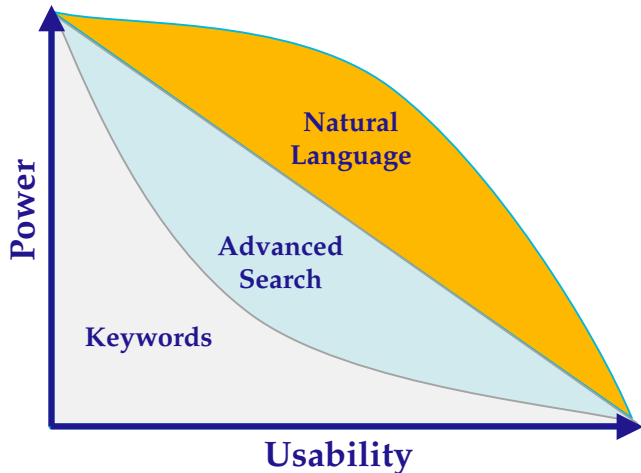
Year	Breakthrough	Company
1993	Text analysis	InfoSeek, Lycos
1995	Directory	Yahoo
1997	CPC Auction	Overture
1999	Keywords, Page Rank	Google
2007	Natural language	Powerset

With respect to precision and recall, keyword search provides good results, advanced search provides better results, and natural language provides the best results.



Similarly, with respect to power and usability, keywords provide okay usability, but bad power. Advanced search provides better power and worse usability (less than 5%

of users make use of it). Natural language search maximizes the benefits of both power and usability.



In talking to advertisers, they found that, “The longer

Consider the question, “Who did IBM acquire?” Compare the results provided by Powerset versus Google.

Powerset	Google
CADAM INC is sold to IBM	Lotus Software - Wikipedia, the free encyclopedia In 1995, Lotus had over 4000 employees worldwide and IBM's acquisition of Lotus ... that IBM feared did not materialize, many long-time Lotus employees did ... en.wikipedia.org/wiki/Lotus_Software = 37k - Cached - Similar pages
He then held the same title when Rational was acquired by IBM in 2003.	Oracle Corporation - Wikipedia, the free encyclopedia Ellison resigned in 2002, saying that he did not have the time to attend ... of Microsoft's SQL Server in the late 90's and IBM's acquisition of Informix ... en.wikipedia.org/wiki/Oracle_Corporation - 78k - Mar 10, 2007 - Cached - Similar pages
Rational was subsequently acquired by IBM in 2003.	Phoenix Technologies - Wikipedia, the free encyclopedia In 1992, Phoenix acquired Quatel, a leading BIOS supplier ... used on a joint development effort with IBM (called SurePath(tm)), but Phoenix did no further ... en.wikipedia.org/wiki/Phoenix_Technologies - 26k - Cached - Similar pages
Before founding Capco, he created Cimad Consultants and sold it to IBM.	IBM Lotus Notes - Wikipedia, the free encyclopedia Setting up archiving for the first time was complex, and often did not create an archive ... Since the IBM acquisition of Lotus, some industry analysts and ... en.wikipedia.org/wiki/Lotus_Notes - 66k - Cached - Similar pages

Natural language search provides search results that are more focused and directed than keyword searches.

the tail¹, the worse the problem.” For users, spam is killing search quality. If a better search mechanism is provided, users will migrate to it.

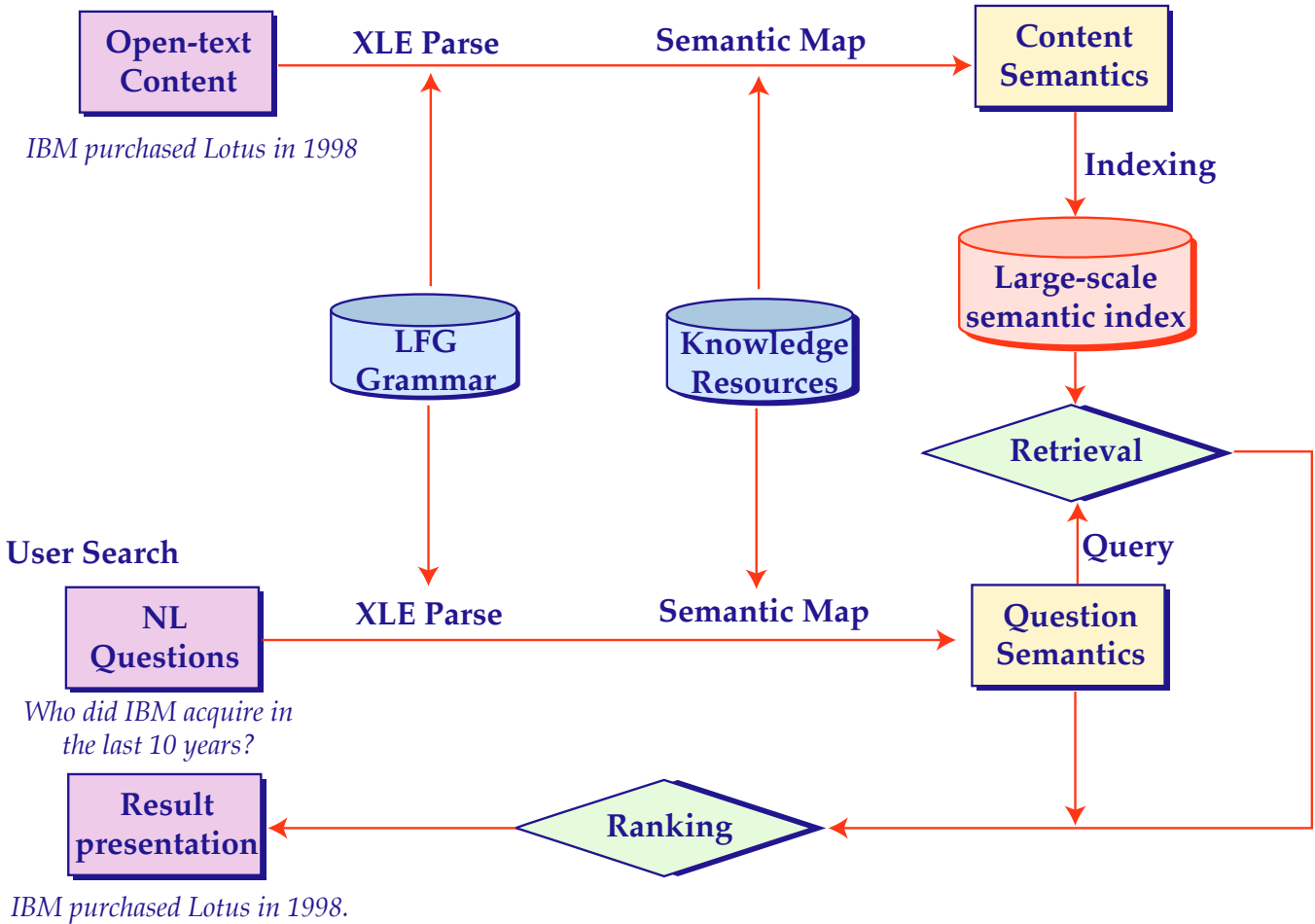
Being fluent in a language eliminates the need for translation, thus it would be ideal to be able to use natural language as a query language.

In summary, natural language is simple, it decreases costs, and it enables better targeting.

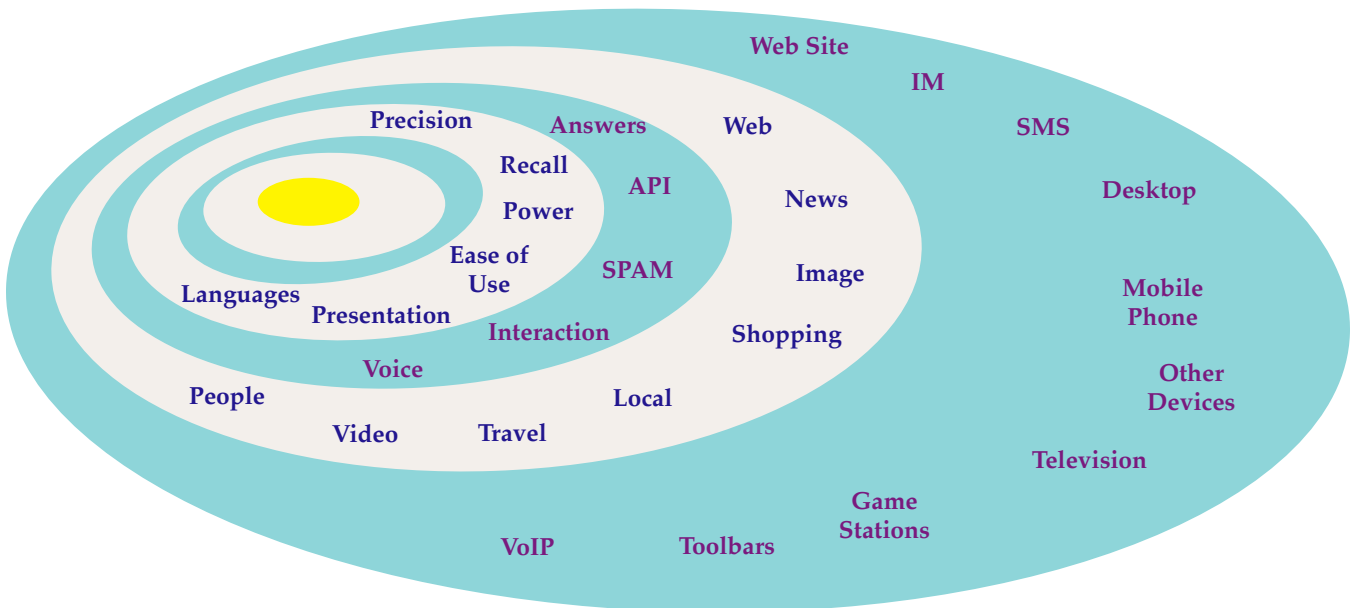
¹ *The Long Tail: Why the Future of Business is Selling Less of More*, by Chris Anderson, Hyperion, 2006

Currently Powerset is using Wikimedia with its 30 million sentences to provide the content acquisition used to train the system. Once the large-scale semantic index and knowledge resources have been formed, natural language user search queries can be made.

Content Acquisition



As the following illustration shows, natural language processing changes the core elements of many products and distribution platforms:



Powerset could have focused on various verticals such as travel, news, business, people, health, shopping, entertainment and local. Instead, it decided to go broad, for everything, the entire \$30B search market.

Questions

At this point, the other panelists spoke briefly. Phyllis Reuther of Mobile Content Networks, www.mcn-inc.com, noted that they provide real-time vertical mobile search solutions, enabling users to quickly find ring tones, images, games, or local information.

Munjal Shah of Riya, www.riya.com, provides visual search. It allows you to use a photo for a query, enabling you to find a fabric that matches a given chair. They have raised \$19M, have 45 people, and have focused initially on shopping, soft goods, carpets and fabrics.

What is your revenue model? Market size?

There is currently \$4B being spent on search advertising, and this is forecasted to increase to \$22B. There is nothing new about our business model, we make our money connecting advertisers to searchers.

When is it more appropriate to do keyword searches?

A short two or three keyword search that results in a website is something that is very difficult for natural language.

Searches usually start with two or three words, but quickly, you have to censor yourself. Natural language search wins as the number of search terms increases.

How do you support multiple languages?

All languages are different, at Powerset we are language independent, we have unique grammars for each language.

It takes about a year to develop the grammar and lexicon for a language. PARC about 15 years ago, distributed its system to linguists. This has resulted in a community that gets together twice a year. We expect to be able to leverage and harvest this effort.

Indexing, how do you catch up to Google?

The cost of natural language indexing is much greater than with keywords. The margins in search are 100x. Google has a margin of 33%, but this is the result of auctions bidding them up.

There are limits to how many pages you have to index, we have several people working on this problem. In any event, in five years, computation will not be an issue.

Powerset has partnered with Amazon, we are paying 10 cents per cpu/hour; thus additional computation is all variable costs; we don't have to buy hardware.

We start out with a smaller index that is deeply processed, and supplement it with a keyword index.

Why are you talking about this now, if you're not going to be ready until the end of 2007?

It took 15 months to negotiate a license with PARC, we just signed the deal a few weeks ago. In any event, we don't believe in being fully stealth. As well, various blogs commented on us, and gave us the incentive to go public sooner, as well as an article in the New York Times in January 2007.

The key battle is for talent, ideally you want people to come to you. One article resulted in 400 people applying to come work for us.

Natural language is very hard to do. What are the alternatives to keyword search that shrink the opportunity for natural language?

Incremental improvements to keyword searches are going to happen. You can do personalization, where you use your knowledge of the person doing the search. However, this doesn't solve the intent problem.

Clustering results also allows slicing and dicing better. Conceptual search allows semantic relationships to be formed. Having specialized knowledge of terms like "strappy shoes," or "Mary Jane" as it applies to women's fashion.

What about approaches that add metadata to web pages to improve search results?

Humans created information on the web, but it gets processed by machine. If metadata was added when it was created, it would reduce search ambiguity.

This is a classic chicken and egg problem. Somehow, you have to get the effort started. At the very least, it needs a lot of volunteers on the back-end. In any event, Powerset is able to take advantage of any preliminary work done.

Metaweb Technologies, www.metaweb.com, provides a rich semantic structure that may be useful in this context.

What about conversation search?

Search is already moving in this direction. and is already happening at the visual search level.

How do you see advertising being done in natural language?

You can buy a sentence, keywords, or concept. For example, you could buy the concept, "college clothing."

In Search of Search

A Market Analysis by The Guidewire Report, February 5, 2007, www.guidewiregroup.com

In the current market, there is perhaps no more heavily invested technology category than search, especially in a market with a dominant player, a dominant business model, and an absolutely unbeatable capital advantage. Dozens and dozens of companies are chasing Google for one reason: tens of millions of people search the Internet and still have a hard time finding what they are looking for.

Indeed, the Top 5 sites in the world are search sites (according to Alexa rankings), if you'll grant that YouTube, ranked No. 5 at the time of this writing, is effectively video search. Each of these businesses is ad supported and each would claim that they target ads based on the interests-as expressed in search terms-of their audience.

It's the same story for nearly every search start-up in the market today, and that may well be their Achilles heel. But more on that later.

In defining search as a "market," one needs to separate the technology platform from the expression of that technology to recognize that many "markets" are being affected by new search methodologies. We see a concentration of activity in three areas of the search market: infrastructure, vertical, and social.

We define "infrastructure" to mean fundamental search technology typically commercialized as middleware, a software appliance, or delivered as a specific application. Companies playing in the infrastructure sector might best be thought of as "applied search" technologies, where their search methodologies are applied to specific data sets, typically in large enterprise data centers.

That definition casts a wide net over a set of start-ups developing core search technologies. The French start-up **Exalead**, for example, provides a uniform search metaphor across data sets from desktop to the Web. **Reveal Technology, Inc.** has just launched a peer-to-peer search application to find documents across workgroup computers. **Xi Technologies**, an Idealab company, utilizes fast indexing technology to enable find-as-you-type search of desktop content.

No matter the technical approach, the vast majority of the search infrastructure companies are applying their technology to limited domains.

Seed-funded **System One** blends collaboration with search to deliver an enterprise information discovery engine. No doubt that is why we are seeing a resurgence of natural language technologies applied to information

search. Despite decades of fundamental research, the ability for algorithms to handle intention and nuance of meaning remains elusive on a mass scale. The companies playing in the natural language space are doing so in all manner of ways and have one core belief in common: computers should understand and interact with people at a higher level. Some companies to watch include **Powerset**, **Hakia**, **HeaciCase**, **Radar Networks**, **TextDigger**, and **Snap**.

No matter the technical approach, the vast majority of the search infrastructure companies are applying their technology to limited domains (a desktop hard drive or an enterprise server farm). In fact, we have only recently discovered an as-yet-unannounced company working on a very early stage project to index the Internet in a more scalable manner than that of Google, Yahoo!, MSN and other leaders. It is too soon to tell if this company will be the next Google, but as the "informationsphere" expands, the sheer volume of content will demand that search technologies will have to scale by using more efficient algorithms, not more commodity servers.

These examples demonstrate the diversity of applications for search technology. Indeed, as businesses and consumers generate-and save-increasing amounts of content, be it structured data or rich media, existing search methodologies are challenged to keep up.

Working on known data sets makes the search problem more manageable, and that's why many new search companies focus on vertical markets. At Guidewire Group, we think "vertical search" has-oddly enough-both a vertical and a horizontal axis.

Along the vertical axis are subject areas - healthcare, consumer electronics, shopping, etc. These vertical silos enable tighter taxonomies that act on specific data sets to deliver better results. In the last year, we've seen search and categorization technologies applied to consumer electronics (**Retrevo**), healthcare (**Kosmix**), and company information (**Zoominfo**), among other vertical content areas.

Along the horizontal axis are data types - photos, video, music, 3D objects, and the like. These horizontal slices enable engineers to target specific attributes of the data type to hone search results. Sites such as **YouTube** and **Flickr** merge rich media sharing with searching, a necessity as their user-generated content databases become massive. We expect to see significant activity in vertical search that cuts across the Websphere. **Big-gerBoat** provides aggregated search for a range of media types. **Blinkx** and **Dabble** index video, **Krugle** programming code, **Qloud** music discovery, **Boorah** restaurant reviews, and Europe-based **Tablefinder** restaurants and restaurant reservations.

Smart algorithms, focused on clear taxonomies and specific data types, are improving search results. Yet smart

people can offer more intelligent processing than a thousand servers when the collective knowledge of the community is leveraged against an information problem. That's the basis of "social search," which might be more aptly labeled social information sharing. Social search approaches layer human insight over basic search methodologies to provide context, texture, and ranking to information.

Social search addresses the challenge of finding rather than searching. It is, effectively, about narrowing search results from the thousands or even millions of possibilities that a Google search might offer, for example, to the few really useful and vetted results recommended by fellow travelers. These social approaches also deliver serendipitous results that cannot be attained from specific search terms applied to literal algorithms.

As more start-ups chase the search market, business models and market assumptions will be severely tested.

Some players are approaching social or contextual search from a vertical perspective, like **Tourist Republic**. This Europe-focused site simplifies the search for travel resources by aggregating specific content on this topic and then allowing users to share their own information with each other and rate the content. This adjusts how the content appears in future searches. **Vozavi** uses crawling and semantic technology along with input from shopping "experts" to provide search for shoppers seeking product reviews. **Yoono's** social and contextual search engine enables users to find relevant sites that are based on their own previous online history as well as the collective set of preferred sites from others in the system who have also indexed their bookmarks. **Aggregate Knowledge** is a service that uses powerful analytics to crunch large data sets and provide personalized content recommendations on each page viewed by site visitors, based on their preferences. Netherlands-based **bliin** takes GPS technology and integrates it with social networks and localized search, creating a system that gives people recommendations and information based not only on where they are but also on who else in their social network has been there.

As more start-ups chase the search market, business models and market assumptions will be severely tested. Most search-focused start-ups are chasing the advertising market, an approach that seems logical enough. If one assumes that an individual's search terms are an explicit declaration of interest and intention, then ads can be surgically targeted to address those declarations. More highly targeted ads are more valuable, the theory goes, and that makes the search site that much more valuable to advertisers.

That theory, however, may find itself at odds with Web advertisers' continuing attraction to mass-market traffic.

On one hand, Web sites herald big traffic numbers to support handsome ad rates. Mass audience equals advertising reach equals efficient advertising spends. This is the old thinking that drives mass media advertising in print, broadcast, and online.

On the other hand, Web advertisers also talk about traffic profiles and champion the "long-tail" theory to place a high value on a tightly targeted audience. Indeed, advances in search technology, local content filtering, and algorithmic and collaborative discovery mechanisms are all about fine-tuning the user experience and delivering to a specific individual exactly the right information at exactly the right moment. Most marketers would agree that this precise, motivated Market of one, where conversion is not just probable, but likely, is the most valuable ad spend. Today's technologies identify and deliver that Market of One, yet few sites have figured out how to calculate and collect on that value efficiently.

In the meantime, we see a troubling reliance by early start-ups on a few ad networks, most notably Google's AdSense. While these ad networks provide a readily accessible revenue stream for search start-ups, they can also create a dependence on another company's business model that we believe builds uncontrollable risk into a startup's business plan. Call us old fashioned, but we like businesses that ask real customers to pay real money for the value they receive. For this reason, the search infrastructure companies may have an advantage over Web-based search companies, particularly if the latter cannot monetize their value independently.

In Silicon Valley, the Race Is On to Trump Google



James C. Best Jr.

By MIGUEL HELFT, New York Times, January 1, 2007

SAN FRANCISCO, Dec. 29 — In brand-new offices with a still-empty game room and enough space to triple their staff of nearly 30, a trio of entrepreneurs is leading an Internet start-up with an improbable mission: to out-Google Google.



Jessica Brandi Lifland for The New York Times, From left, Steve Newcomb, Lorenzo Thione and Barney Pell are the founders of Powerset, a search engine.

The three started Powerset, a company whose aim is to deliver better answers than any other search engine — including Google — by letting users type questions in plain English. And they have made believers of Silicon Valley investors whose fortunes turn on identifying the next big thing.

“There’s definitely a segment of the market that thinks we are crazy,” said Charles Moldow, a partner at Foundation Capital, a venture capital firm that is Powerset’s principal financial backer. “In 2000, some people thought Google was crazy.”

Powerset is hardly alone. Even as Google continues to outmaneuver its main search rivals, Yahoo and Microsoft, plenty of newcomers — with names like hakia, ChaCha and Snap — are trying to beat the company at its own game. And Wikia Inc., a company started by a founder of Wikipedia, plans to develop a search engine that, like the popular Web-based encyclopedia, would be built by a community of programmers and users.

These ambitious quests reflect the renewed optimism sweeping technology centers like Silicon Valley and fueling a nascent Internet boom. It also shows how much the new Internet economy resembles a planetary system where everything and everyone orbits around search in general, and around Google in particular.

Silicon Valley is filled with start-ups whose main business proposition is to be bought by Google, or for that matter by Yahoo or Microsoft. Countless other start-ups rely on Google as their primary driver of traffic or on Google’s powerful advertising system as their primary source of income. Virtually all new companies compete with Google for scarce engineering talent. And divining Google’s next move has become a fixation for scores of technology blogs and a favorite parlor game among technology investors.

“There is way too much obsession with search, as if it were the end of the world,” said Esther Dyson, a well-known technology investor and forecaster. “Google equals money equals search equals search advertising; it all gets combined as if this is the last great business model.”

It may not be the last great business model, but Google has proved that search linked to advertising is a very large and lucrative business, and everyone — including Ms. Dyson, who invested a small sum in Powerset — seems to want a piece of it.

Since the beginning of 2004, venture capitalists have put nearly \$350 million into no fewer than 79 start-ups that had something to do with Internet search, according to the National Venture Capital Association, an industry group.

An overwhelming majority are not trying to take Google head on, but rather are focusing on specialized slices of the search world, like searching for videos, blog postings or medical information. Since Google’s stated mission is to organize all of the world’s information, they may still find themselves in the search giant’s cross hairs. That is not necessarily bad, as being acquired by Google could

be a financial bonanza for some of these entrepreneurs and investors.

But in the current boom, there is money even for those with the audacious goal of becoming a better Google.

Powerset recently received \$12.5 million in financing. Hakia, which like Powerset is trying to create a “natural language” search engine, got \$16 million. Another \$16 million went to Snap, which has focused on presenting search results in a more compelling way and is experimenting with a new advertising model. And ChaCha, which uses paid researchers that act as virtual reference librarians to provide answers to users’ queries, got \$6.1 million.

Still, recent history suggests that gaining traction is going to be difficult. Of dozens of search start-ups that were introduced in recent years, none had more than a 1 percent share of the United States search market in November, according to Nielsen NetRatings, a research firm that measures Internet traffic.

Amassing a large audience has proved to be a challenge even for those with a track record and resources. Consider A9, a search engine owned by Amazon.com that received positive reviews when it began in 2004 and was run by Udi Manber, a widely recognized search specialist. Despite some innovative features and early successes, A9 has captured only a tiny share of the market. Mr. Manber now works for Google, where he is vice president of engineering.

The setback apparently has not stopped Amazon or its chief executive, Jeffrey P. Bezos, from pursuing profits in search. ChaCha said it counts an investment company owned by Mr. Bezos among its backers, and Amazon is an investor in Wikia. An Amazon spokeswoman said Mr. Bezos does not comment about his personal investments.

Some start-ups are similarly bullish. “We expect to be one of the top three search engines,” said Riza C. Berkan, the chief executive of hakia. It is a bold claim, given that hakia’s technology is not yet ready for prime time, and Mr. Berkan readily concedes it will take time to perfect it.

The dream, however, is quintessential Silicon Valley.

“It is hard for me to believe that anybody thinks they can take Google’s business from Google,” said Randy Komisar, a venture capitalist who was once known as Silicon Valley’s “virtual C.E.O.” for his role as a mentor to scores of technology firms. “But to call the game over because Google has been such a success would be to deny history.”

In some ways, the willingness of so many to make multimillion-dollar investments to take on Google and other search companies represents a startling change. In

the late 1990s, when Microsoft dominated the technology world, inventors and investors did everything they could to avoid competing with the software company.

Yet many of today’s search start-ups are putting themselves squarely in the path of the Google steamroller. Most explain that decision in similar ways.

They say that Google’s dominance today is different from Microsoft’s in the late 90s when its operating system was a virtual monopoly and nearly impossible to break. In the Internet search industry, “you earn your right to be in business every day, page view after page view, click after click,” said Barney Pell, a founder and the chief executive of Powerset, whose search service is not yet available.

They also say that the market for search simply is too large to resist. Google, which, according to Nielsen, handles about half of all Internet searches in the United States, is valued at an astonishing \$141 billion. So, the reasoning goes, anyone who can grab even a small slice of the search market could be well rewarded.

“You don’t need to be No. 1 to be worth billions of dollars,” said Allen Morgan, a partner at Mayfield Fund, a venture capital firm that invested \$10 million in Snap. The company is also backed by Bill Gross, an Internet financier who pioneered the idea of linking ads and search results, only to see Google seize the powerful business model and improve on it.

Almost all of today’s search entrepreneurs also say that Google’s success lends credibility to their own long-shot quest.

When Lawrence Page and Sergey Brin first started tinkering with what would become Google, other search engines like AltaVista and Lycos and Excite were dominant. But the companies that owned them were distracted by efforts to diversify their businesses, and they took their eye off the ball of Internet search and stopped innovating.

Some now say that search has not evolved much in years, and that Google is similarly distracted as it introduces new products like word processors, spreadsheets and online payment systems and expands into online video, social networking and other businesses.

“The more Google starts to think about taking on Microsoft, the less it is a pure search play, and the more it opens the door for new innovations,” said Mr. Moldow, the Foundation Capital partner. “That’s great for us.”

But at the same time, Google, Yahoo and Microsoft have thousands of engineers, including some of the world’s top search specialists, working on improving their search results. And they have spent billions building vast computer networks so they can respond in-

stantly to the endless stream of queries from around the world.

Search “is becoming an increasingly capital-intensive business,” said Marissa Mayer, Google’s vice president for search. That makes it harder for start-ups to catch up to the giants, she said.

That is not stopping entrepreneurs from betting that they can. Powerset has search and natural-language experts among its two dozen employees, including former top engineers from Yahoo and a former chief linguist from Ask Jeeves, Ask.com’s predecessor. They are the kind of people who could easily land jobs at Google or Microsoft or Yahoo.

Steve Newcomb, a Powerset founder and veteran of several successful start-ups, said his company could become the next Google. Or, he said, if Google or someone else perfected natural-language search before Powerset, then his company would make a great acquisition for one of the other search companies. “We are a huge story no matter what,” he said.

Ms. Dyson, the technology commentator and Powerset investor, captured the optimism more concisely and with less swagger. “I love Google,” she said, “but I love the march of history.”