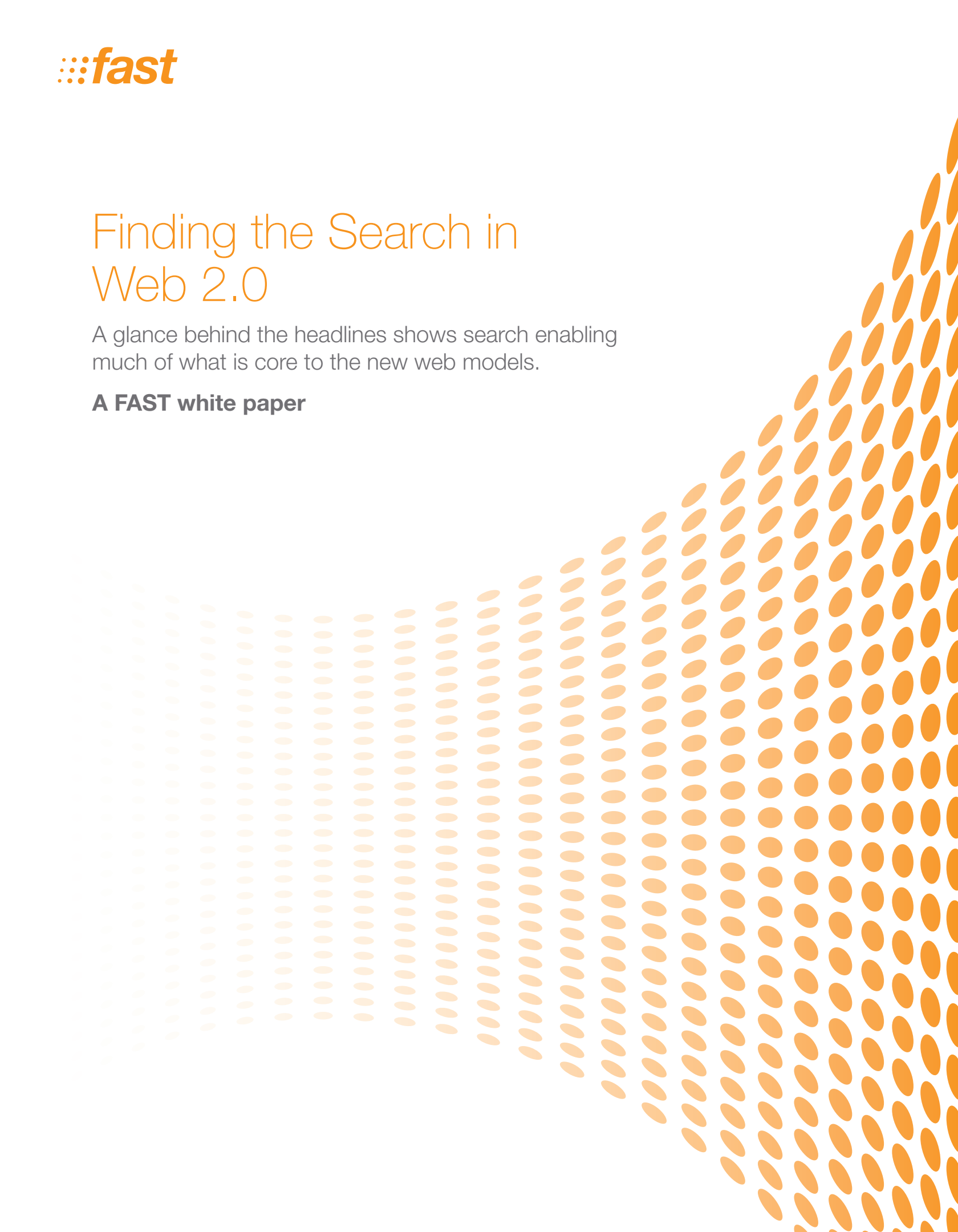
fast

Finding the Search in Web 2.0

A glance behind the headlines shows search enabling much of what is core to the new web models.

A FAST white paper



Introduction

Perhaps you first noticed it in 2000-01, at the lowest moments of the dot-com bubble crash: the sudden appearance on the web of a new encyclopedia created solely by volunteers, growing contributions at exponential rates. Wikipedia now has over 10 times the number of entries as the Encyclopedia Britannica and has become the de facto knowledge source for web users worldwide.

Or perhaps you noted the arrival of Web 2.0 when recently and notably Google caught the attention of everyone in the media industry and beyond by acquiring 14-month-old YouTube for an eye-popping $1.65 Billion.

Web 2.0 is an important phrase - often the phrase - on many lips today. Tim O'Reilly, CEO of O'Reilly Media, the iconic tech book publisher and event firm, was insightful and forward-thinking in creating this label - Web 2.0 - to describe a whole spectrum of innovative activity that has been taking place on the web since the bursting of the original "dot.com" bubble. In the 36 months since O'Reilly helped put together the original Web 2.0 conference, however, the promotional machinery of the technology press, the venture capital community, the media and advertising industries, and even the general business press has raised the decibel level and quantity of Web 2.0 hype nearly off the charts.

At FAST, we are looking beyond the hype to the pragmatic and the tangible in Web 2.0 developments. There what we see is the emergence of a whole new network of connections.

At FAST, we look beyond the hype to the pragmatic and the tangible in Web 2.0 developments. There, what we see is the emergence of a new network of connections.

MARKUS.ANGERMEIER@APERTO.DE RELEASED UNDER CC 2.0 DE ATTRIBUTION SHAPE ALIKE 11.11.05

Fundamental Transitions

In our view, the most important thrust of Web 2.0 is a marked shift in the engagement model between Web properties and their users or audiences. The Web from its early adoption in 1993-94 through roughly 2002 was primarily an access and download phenomenon (much like the traditional publishing model transferred intact to an electronic world). Since 2002, we see an accelerating trend toward a Web that is importantly driven by individuals' "uploads," that is, their active (and interactive)
participation.

Convergence Becomes Reality

We are looking at Web 2.0 and the development of search at a moment in time in which the global internet infrastructure is largely built out and connecting rapidly-expanding user communities and businesses with high speed broadband linkages. As these linkages serve an increasing proportion of worldwide businesses and populations, we are finally seeing the much-predicted convergence of media, broadcasting, entertainment, telephony, advertising, and data-based businesses.

The lines between internet, TV, and phone (and also entertainment devices like CD-players, game players, movie screens, PDAs, cameras, and radios) are now permanently blurred. This long-anticipated convergence is currently an area of constant and disruptive innovation for suppliers and consumers alike.

Where might search fit into these large trends? What role can it and will it play in Web 2.0? And what do we mean by "search" in this context anyway?

Is Search in the Picture?

In the depiction of the cloud of Web 2.0 presented above, the author has included trends, labels, site names, technologies, and new web dynamics (e.g. the "long tail") which have been associated with the Web 2.0 phenomenon—but search is not in the picture. Perhaps you would expect FAST to challenge the author on his lack of vision and his obvious oversight. But in fact we think there is real insight in this omission. If by search we mean the traditional keyword search box that was such a central characteristic of the Web 1.0 era, and was popularized by the original "search engines" Yahoo!, Excite, Lycos, and Infoseek, and then extended in terms of scalability by Alta Vista and FAST's own AllTheWeb, then we agree that that era of search is over, and that the usefulness of that Web 1.0 version of search has been redefined by more modern and ambitious approaches.

Search Beyond the Box

FAST's perspective is that both the capabilities and the business environments for search have left the commonly understood functions and behaviors of the traditional search box far behind.

For many, that traditional search box encapsulated both the first experience of the power of the web and a continuing experience of the failure of much of Web 1.0 technology to address the issues of volume/scale and relevance that confront us each time we fire off a question into the white box.

Is Google the Search for Web 2.0?

So if that early keyword box was inadequate, isn't Google's success in search an indication that the problem is solved, and we now just assume that Google is the peak of search evolution for Web 2.0? In our view, Google's page rank algorithm innovation, which permanently raised expectations for relevancy for most results lists, is an end only in that it represents the peak achievement for search in Web 1.0.

The Google accomplishment is in effect a proto-Web 2.0 phenomenon, in that it gains its incremental improvements in relevance explicitly by taking advantage of the implicit wisdom of social interaction represented by the multitudes of hyperlinking acts of individual web authors and publishers around the globe. But there are two fundamental limits to the way that Google is handling this innovation that locate its version of search squarely

FAST's perspective is that both the capabilities and the business environments for search have left the commonly understood functions and behaviors of the traditional search box far behind.

in the Web 1.0 world.
The first core limitation of Google's approach is that the user experience model – the practice of responding to each query with a (long) list of links is fundamentally flawed. This Web 1.0 model confronts the user with a 5-stage process that is even now being relegated to the dust bin of search experience history by new innovations from FAST and others that we will touch on below. The 5 stages, of course, are:

1) Ask (type in search term or terms)
2) Review (scan lists of pages)
3) Make selections from list & navigate to the chosen page
4) Read, weigh relevance of the selected page
5) Repeat

In the current stage of development of Web 2.0, pioneering practitioners of search are already eager and, with FAST search technologies, able to abandon this familiar user experience model. The replacement experience that these pioneers are innovating with FAST software leaves the 5-stage sequence behind and replaces it with an architecture for interaction designed to connect users directly with answers. The next generation of search will deliver decision support environments that time-shift users' searches into parallel channels and connect them to actions and processes that set a new bar for interaction quality.

The second core limitation of the Google approach is that the search leaves the insights it picks up from analyzing structures of links across the web at a web-wide level. This "global" insight limits the ability of the search to capture local nuances that can make the difference between a successful

and a failed search for specific audiences or community networks. The approach does not consider the patterns of interaction or knowledge conventions that emerging Web 2.0 communities take for granted.

Sentiment analysis is the raw material for both predictive market intelligence and proactive Web counter-intelligence programs, and sentiment analytics are becoming core to the Web 2.0 search platform.

However, this limited, "generalized" results approach is highly supportive for the keyword-based advertising business model that is fueling Google's business success. For now, Google's Web 1.0 search success appears to be restricting the firm's ability to continue to innovate into the Web 2.0 search experience.

Where to Find Search In Web 2.0

The new search for Web 2.0 leaves the traditional search box behind and replaces it with a next-generation search that takes the leverage offered by Web 2.0 participation – user-generated content, knowledge-sharing communities and social networks, local rating and tagging, and highly interactive interaction and programming models – to create extreme precision in relevance models delivered through emergent consumption-centered architectures. In the segments that follow, we review what we believe to be the most important areas of current development where search impacts Web 2.0, and summarize what FAST is contributing to innovate in these environments.

Consumer Generated Media

One of the earliest take-up areas in Web 2.0 centers on publishing, and the traditional publishing industry is being rocked by the rapid emergence of consumer-generated, open source media. As of July 31, 2006, Technorati, the blog aggregation web site, reported that there were 50 million blogs being published worldwide, and the "blogosphere" is doubling every 200 days.

We noted above the phenomenal growth and success of Wikipedia, the consumer-generated encyclopedia. This open source of accumulating knowledge from around the world not only challenges the traditional models of the encyclopedia business, which had already been permanently disrupted by Microsoft with its Encarta gambit for local PCs in the '80s, but also now offers effectively "real time" knowledge encapsulation on many developing events across the globe.

FAST is responding to the blog and wiki phenomenon not just by providing tools to search and analyze both the text content and the metadata content of the blog pages, but also by working closely with customers to adopt new practices that adapt to the business impacts of these micro-publications.
Because they are by nature "opinionated," blogs serve as a major new source of market intelligence, and so in one new Web 2.0-enabled innovation, savvy companies' marketing groups are mining blog content for sentiment indicators about products, brands, corporate political activity or executive behavior. Sentiment analysis is the raw material for both predictive market intelligence and proactive Web counter-intelligence programs, and sentiment analytics are becoming core to the Web 2.0 search platform.

FAST's Contextual Insight® software also helps blog searchers address the issue of lack of structural, taxonomic, or professionally marked-up metadata in "informal" media. This software enables automatic analysis and metadata generation at the concept, entity, and sub-structural levels of the text, allowing casual documents to offer many of the accessibility characteristics of professionally published media.

Emergent Communities

Another important trend within the set of Web 2.0 phenomena is the increasing popularity of social networks and "community" sites. Younger people have taken to the idea of publishing themselves on Web media, assisted by innovative entrepreneurs who have developed sites like Friendster, MySpace, and Facebook that have proved both attractive and "sticky" for large audiences to share and connect around all kinds of personal "stuff."

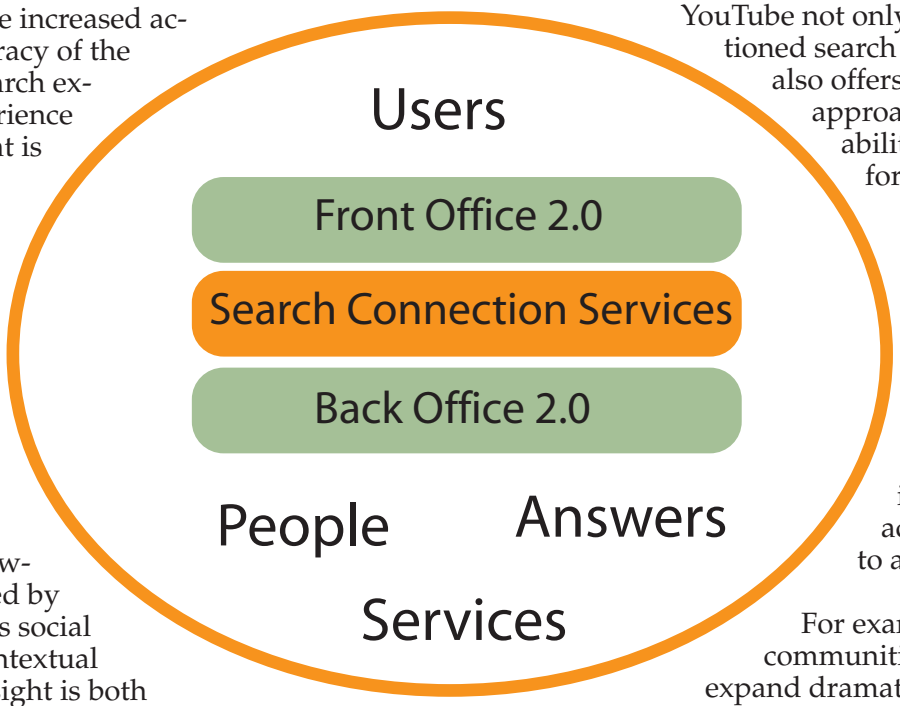
Social Search

The development of these online social networks makes possible some interesting and suggestive developments for search technologies that FAST is incorporating into the software it delivers today to

online community businesses.

Unlike users of conventional search, the members of communities share a large body of information about themselves with other members of the community, particularly with their closest associates – their buddies or friends. The new search technologies can take these implicit community member profiles and relationships into account when interpreting search questions of community members. This approach can tighten interpretations of the relevance of specific content items. It can also leverage the opinions of a group of friends to provide guidance for recommendations or suggestions of what a particular community member would find most useful or fun.

The increased accuracy of the search experience that is powered by this social contextual insight is both refreshing and directly useful to users and at the same time provides ways for community business sponsors to monetize the attention and transactions of community members.



Video Communities

Another community sharing activity associated with Web 2.0 is rich media uploading, initially photos and music, but now increasingly full-motion video, for general consumption and entertainment across the Web. Early sites like Napster were doomed by their own success at transforming access models for the music industry. But that success led to a new model for online music that has re-invigorated a whole company, Apple, around the iPod and the iTunes music store destination site.

Social photo-sharing and tagging site Flickr built a franchise big enough to attract Yahoo!. And in 2006, after less than a year of operation, video-sharing site YouTube was delivering 100 million "showings" per day, and making business arrangements with mainstream media players like Fox and Warner that have the potential to impact the online strategies of the media industry. This impact was clearly in the sights when Google made its $1.65B acquisition of YouTube in October, 2006.

FAST search offers video sharing sites like YouTube not only highly accurate, socially conditioned search around its clip collections, but it also offers direct video and audio mining approaches that can enhance the findability and monetization possibilities for this media.

In the advanced scenarios of Web 2.0 content sharing, community-generated tags and tag clouds provide an initial level of access that is helpful for navigation and for cueing high quality search, but the techniques for directly mining the rich media itself can take the intimacy and accuracy of the search experience to a new level.

For example, rich media search enables communities, bloggers, and publishers to expand dramatically the usability of innovations like the podcast and the video podcast by "cracking the DNA" of the presentations and offering the resulting "handles" to the content to searching audiences, whether local or global.

Search Connection Services

As the diagram above indicates, perhaps the most powerful capability of search software is its position as a service broker between the intention of a content- or information consumer and the content or product that would satisfy that consumer's current quest.

Staging the Conversation

The new Web 2.0 search connection services en-

The core contribution of search-delivered filters is to separate the "signal" of interest from the "noise" of overwhelming quantities of content in any Web environment. The commercial impact of knowing how to build filtered environments and user experiences could not be greater.

able online businesses to develop architectures of participation which allow users to discover each other and explore shared interests in the context of the online interaction space.

The search services help harness the collective insights and perspectives of these users to keep content and product offerings relevant, often in real time.

The Web 2.0 search connection services also should provide a deeply contextual advertising infrastructure that pushes past the Web 1.0 models to offer integrative, suggestive, community-sensitive, and proactive services.

Exposing the Tail

In his new book The Long Tail, Chris Anderson refers to this function of the new technologies as "filters," by which he means the kind of spectrum of extended services capabilities which FAST incorporates in its solution portfolio and FAST's customers are implementing in Web 2.0 commercial environments.

These "filters" include advanced search technologies, recommendation engines, social networks, taxonomies and navigation aides, collaborative filtering, entity extraction, intelligent usability practices, and all the other infrastructure facilities that promote discovery and context online.

The core contribution of these filters is to separate the "signal" of interest from the "noise" of overwhelming quantities of content in any Web environment. The commercial impact of knowing how to build filtered environments and user experiences could not be greater.
As Anderson states: "In today's Long Tail markets, the main effect of filters is to help people move from the world they know ("hits") to the world they don't ("niches") via a route that is both comfortable and tailored to their tastes. In a sense, good filters have the effect of driving demand down the tail by revealing goods and services that appeal more than the lowest-common-denominator fare that crowds the narrow channels of traditional mass market distribution."

What Anderson has put his finger on is the fact that the economic benefits of the Long Tail can not be realized without the portfolio of technologies and practices that makes the discovery, evaluation, and interaction with these digital products possible. It is this portfolio that makes up the new search for Web 2.0, and it is the nuanced deployment of these technologies across more and more online
businesses and communities that will bring the ideas and trends behind Web 2.0 out of the realm of speculation about futures and into the machinery of day-to-day business online.

To Do's

Will Web 2.0 totally transform the "competitive rules" by which business operates today? No. What it will do is provide a whole range of opportunities for nimble firms in many industries to take advantage of the newly "social" and participatory Web to deliver value to online customers. The speed of this change will threaten a number of firms and a number of traditional business models, but the innovative among them will help pioneer the new Web economy to their lasting benefit.

About FAST

FAST is the leading developer of enterprise search technologies and solutions that are behind the scenes at the world's best known companies with the most demanding search problems. FAST's solutions are installed in more than 3500 locations.

FAST is headquartered in Oslo, Norway and Needham, Massachusetts and is publicly traded under the ticker symbol 'FAST' on the Oslo Stock Exchange. The FAST Group operates globally with presence in Europe, North America, the Asia/Pacific region, South America, the Middle East and Africa. For further information about FAST, please visit www.fastsearch.com.

For any feedback or questions related to this paper, please contact us at feedback@fastsearch.com.

FAST™
www.fastsearch.com
info@fastsearch.com

Regional Headquarters

The Americas
+1 781 304 2400

Europe, Middle East & Africa (EMEA)
+47 23 01 12 00

Japan
+81 3 5511 4343

Asia Pacific
+612 9929 7725

FST-1000256-01