Best Practices in Enterprise Search

Andy Moore. ................. 2 In Search of Search
I usually make it a rule for these interviews: No inside baseball. I’m normally much more interested in the so-called 50,000-foot view...how customers perceive their business needs and how these various KMWorld-y technologies can address them...Or how buyers organize themselves.

Bjørn Olstad and Silvija Seres . . . . 4 User Driven Information Discovery
The information chain is a pyramid. In past decades, information-intensive businesses profited from dropping storage costs. Computer systems offered simple.

Steve Papa, Endeca ............ 6 The Faceted Navigation and Search Revolution
You've likely heard the buzz by now: faceted navigation truly changes the ground rules for search and KM, and for the many related applications.

Mark Woollen, InQuira ......... 8 Accelerating Business Success with Intent-driven Search Interactions
In today’s hyper-competitive business environment, companies work hard to differentiate themselves and provide greater value to their customers.

Jeff Dirks, SchemaLogic ....... 10 Business Semantics Management: Empowered Enterprise Search Drives Competitive Advantage
For Global 5000 enterprises, the ability to optimize information management for competitive advantage will drive growth in market share.

Johannes C. Scholtes, ZyLAB ..... 12 Next-generation Deployment for Enterprise Search Tools
Consider the following: Throughout the course of the Enron investigation, the FBI has confiscated 12 million pages of paper documents plus terabytes of e-mail.

Chris Weitz, BearingPoint ..... 14 Making the Business Case for Search
The performance of any organization depends heavily on how well its people use information to make decisions and do their jobs. First, though.

An Engenium Success Story ... 16 Stalking the Smoking Gun: Top Law Firm Installs Concept Search Engine
In the legal community, search technology has become as critical to the preparation of major lawsuits as securing witnesses and researching case law.

Ian Davies, ISYS Search Software .... 17 Analytics: Measuring the Impact of Search
Quicker than you can say Sarbanes-Oxley, the enterprise search market has gone from serving a relatively quiet existence as a nice-to-have technology.

Mark Angel, Knova Software .... 18 Adaptive Search and Resolution for Service and Support
If lowering costs, closing more issues at first contact, or increasing customer satisfaction are on your contact center’s to-do list.

Steve Kusmer, WebSideStory .... 19 The “Other” Search
When online marketers and Internet technologists talk about “search,” the image of millions of shoppers, researchers and office workers typing queries.

Bill Fisher, MetaCarta ......... 20 Why Does Geography Matter in Enterprise Search?
While browsing through various historical maps featured in the “American Treasures” collection at the Library of Congress the other day, one thing became very evident.

Ken Bubeck, Open Text ......... 21 Distributed Search Technology
The need for fast, accurate search and retrieval continues to be a significant factor in driving software sales for Fortune 1,000 companies and their employees.

Bob Tennant, Recommind ...... 22 Beyond Search Results—Get the Whole Picture and Reap the Rewards
Search engines today are built with an innately document-centric view of the world. This is hardly surprising, given how the technology has evolved, but it is often not the paradigm best utilized for providing an answer to a user’s search request.

Laurent Simoneau, Coveo ..... 23 Managing Content for Added Value
There are many drivers that require companies to manage knowledge more effectively—the incredible rate at which electronic data is reproduced and stored.
In Search of Search

What Does the Marketplace See When It Looks in the Mirror?

By Andy Moore, Editorial Director, KMWorld Specialty Publishing Group

It’s a classic boom-bust cycle... except, like, on crack.

In the early days of search, we have a single platform. Or even how companies organize themselves into multi-disciplinary teams to spearhead enterprise technology purchases. Or even how CEOs feel about risk, like taking the perp walk on the 6 O’clock news.

But I almost never ask vendors what it’s like to be a vendor. Enterprise search, it turns out, is different. It’s been my observation that enterprise search has behaved like a micro-economic model of the typical technology adoption and maturity cycle—except, like, on crack.

So I tested my incredibly mature theory on a group of enterprise-search vendors in a series of conversations in February, 2006. Here’s what I learned:

Andy Moore: So, am I right? Is enterprise search on crack, or what? (I’m paraphrasing a little here, for the sake of brevity.)

Bob MacDonald: When a new technology—any technology—first hits, it has a singular definition, and people talk about it in a singular way. Search was like that, too. Then products start emerging, both expanding and breaking apart that original perception. Search started with just such a simple concept: You need to access the information you have. Then product vendors created solutions that focused on different aspects. Now—all of a sudden—you have those focused on e-commerce, those that are for legal or security requirements, or those for customer-facing applications in call center or service environments.

Jeff Dirks: It’s a classic boom-bust cycle, where the front-runners get out early and grab the IT budgets of the Global 5000, followed by consolidation and a narrowing of the leaders in the particular segment. I think that’s what’s happening in search.

But I also think it’s a much bigger addressable market than some of the analysts. We’re on the edge of a tectonic shift in the way big companies, especially, think about information management, and search is at the center of this.

Johannes Scholtes: Our customers have come to understand search much better than just five years ago. “Enterprise search” has changed, and has become a series of solutions optimized for specific customers’ needs. Some of us vendors are now doing more due diligence for each of the white papers in the series.

Chris Weitz: The adoption cycles among the customers and maturity cycles of the vendors are more accelerated due to the availability of Internet technology and rapid software development. Search will be assumed to be a platform going forward, not just a weird thing used by the IT department.

Silvija Seres: We have a single platform, and on top of that platform we create several, focused, search-derivative applications. At the heart of these specialized applications, there is a shared set of core capabilities—a scalable index; the flexible indexing of content and queries. Then you add the linguistics and media add-ons that allow you to create focused products.

Steve Papa: In the early days of search, there wasn’t much barrier to entry; a couple of people guys in a university could create a keyword search product. It wasn’t really any different than what you could do from a DOS prompt. Now the hard work and innovation are being done.

Silvija: We are now in this incredibly sweet spot in the market. It’s going so fast that we are all running to keep up. There’s an explosion in the volume of data, of course, plus there’s an explosion in the complexity of the data types and media. Users are becoming more demanding and sophisticated, they know what they want to do, and they want to do it on a huge scale.

So we’re going deeper into places where search is already in use. But we’re also going broader, into more markets. The technology analysts see this way, too. IDC is now adding structured data processes, such as business intelligence, into their analysis of this marketplace. So instead of

The Cast of Characters

<table>
<thead>
<tr>
<th>Jeff Dirks</th>
<th>President and CEO</th>
<th>SchemaLogic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andy Moore</td>
<td>Editorial Director</td>
<td>KMWorld Specialty Publishing Group</td>
</tr>
<tr>
<td>Bob MacDonald</td>
<td>Chief Marketing Officer</td>
<td>InQuira</td>
</tr>
<tr>
<td>Johannes Scholtes</td>
<td>President and CEO</td>
<td>ZyLAB</td>
</tr>
<tr>
<td>Silvija Seres</td>
<td>Vice President of Strategic Market Development</td>
<td>Fast Search &amp; Transfer</td>
</tr>
<tr>
<td>Chris Weitz</td>
<td>Managing Director</td>
<td>BearingPoint</td>
</tr>
</tbody>
</table>
predicting it as a $700 million market, with this broader view they now think of it as an aggregate $25 billion market.

Andy: They put a man on the moon 37 years ago. But it’s only the last few years that we’ve had decent search. What gives?

Johannes: People have been trying to optimize search for more than 50 years now, with mostly disappointing results until recently. The more technology that got thrown at it, the less efficiently it seemed to work.

Bob: I am amazed at how much of a blend of technology and intellectual theory there is in search. Back in the day of database development, everyone pretty much knew how to do it. There was an industry blueprint. You could pull together the people with computer skillsets and pretty much get into that business. But you cannot be effective in the search space without a high understanding of computational linguistics and semantic analysis, as well as computer science capabilities.

Johannes: On the Internet, you want to find the single answer that solves your problem...it’s what we call precision. But in auditing, discovery, disclosure or investigations, you want to find everything. The problem is, once you get to that 100% level of recall, your precision goes down the drain. So for those high-recall applications, you need to apply linguistic technology or visualization techniques to reach the level of precision AND recall you need.

Due to this, search is changing many markets. Beyond just the database market—where databases are becoming too large to be effectively managed by the database-providers’ tools—other markets such as advertising, auditing, internal discovery, law enforcement, legal...search is at the heart of major revolutions in all those markets.

Bob: Customers have come to realize that keyword search does not give them the specificity they need to respond to, say, a consumer’s question to the call center. They need the computational linguistics necessary to index their content for meaning, allowing them to pull an excerpt out of relevant documents, rather than just point to the document. That is a huge productivity jump.

Silvija: Search technology is built for huge volumes of complex data. Users who don’t necessarily know precisely how to “ask the right questions” are still being served. We’re just where we need to be.

Maybe it’s not an information overload: maybe it’s information underload. If you think about all the nature around us, we humans have learned how to handle that volume and complexity. That is where we are getting now with our computers. If we provide them with enough of the right information, with all the necessary context, and the right tools to filter and abstract, we will have a very intelligent tool. We’re just about there; it’s incredibly exciting.

Bob: The greater power of computing, and the lowering of those costs, have made it feasible for businesses to apply these tools. Our stuff came out of an 8½ year study at MIT. They weren’t trying to implement it for business; they were just trying to find a way—any way—to index and retrieve information with a completely automated system. The computing power needed to implement this stuff had to get to a price point where storing and retrieving it made sense. Now we’re there.

Andy: Let’s talk about the elephant in the corner—Google...

Chris: Search is badly defined and badly understood across the board. If you talk to business leaders and say “search,” they either draw a blank, or they say “Google” (It should be noted that this conversation took place the day following BearingPoint’s and Google’s announcement of a major alliance.)

"In the early days of search, a couple guys in a university could create a keyword search product."

If you’re eight or 80 years old, you now know what “Search,” with a capital S, means. And that’s because of Google. But enterprise search is currently the domain of the nerd in the IT department who is not making any business decisions. That’s changing, but it’s true.

Steve: It depends on the company, and who you’re talking to. If they’re using it to serve their customers, they’re taking it very seriously. The head of e-commerce at a large on-line retailer sees search as mission-critical.

Chris: Things migrate from features, to applications, to strategic levers, and we’re not anywhere close to the far right of that curve. Yet. We’re a huge systems integrator. We follow the money, and you’d have to use a magnifying glass to find the money in search. But we’re, to use an old adage, skating where the puck is going. I am clear about the fact that’s changing. That’s why we’ve invested in this search practice.

Jeff: It’s true that Google has captured the hearts and minds of the public. But I think of Google as a fairly straightforward search algorithm and a good way to monetize advertising in a Web setting. There couldn’t be a larger chasm of difference between the way the information stack is organized in companies, and the way Google works. We think of it as information-enabling key processes; search and discovery is a key part of that. But the use case might be global brand management across 1,000 content silos, or global customer self-help. So much is driven by the flow of information, where 15-20 years ago it would have been just the physical assets of the company.

Johannes: Everyone understands how the Google “popularity algorithm” works. If you search for the best hotel in New York, you’ll get an answer—the one that most people THINK is the best website for the best hotel. But in law enforcement, you don’t necessarily want the most popular website...you want, perhaps, all of them or the ones that are most relevant.

Chris: For most companies, pretty good is good enough. A lot of niche search vendors have overshot the market, and they’re doing more than people need.

Andy: Why is search such a big deal now?

Bob: Whether it’s an internal or external application, all the factors are in place to drive growth for the search market. For one, there’s not a single major corporation that doesn’t want to leverage Internet-type technologies for internal use to distribute information. And then on the external, customer-facing side, we’re going through a next-generation of corporate websites, through which the businesses are trying to drive interactions more effectively. And the most effective way to do both is through the search paradigm.

Steve: When people invest in new technology, they are increasing additional relationships among data. The irony is, the more you invest, the less accessible those relationships become. It’s not really about search, it’s about those relationships. The other challenge for search is open source stuff, such as Lucene. It’s even harder now for those two guys to compete with something that’s free!

Jeff: It’s not uncommon for us to hear from customers—the ultimate customers—who can’t find what they want on the website, and so they go to the call center and say “I didn’t want to call you, but I had no choice. Why don’t you save yourself some money, and me some time, and fix your site?”

Steve: The more people who are aware, the better. We find, though, that as people find the potential locked in the information assets, they then start to look for how to do a better job. With search high up on the agenda, higher people in the organization have gotten interested.

Silvija: It’s to the point where we are helping write business plans around search. We help companies understand how their whole strategy changes because of search. We help them operate in a new way.
Inverting the Corporate Information Pyramid:
User Driven Information Discovery

By Bjørn Olstad, CTO, and Silvija Seres, VP Strategic Market Development, FAST

The information chain is a pyramid

In past decades, information-intensive businesses profited from dropping storage costs. Computer systems offered simple and inexpensive data retention—in databases, document repositories and email archives. Within the scope of their design, these systems succeeded, creating massive amounts of digital information typically stored in application “silos.”

The competitive landscape has changed. Return on intellectual capital drives success at least as much as transactional growth does, and infrastructures that served information storage well enough in the past fail miserably at the task of exploiting the enterprise’s information. Companies also face new risks, such as compliance laws demanding higher levels of visibility. Information access has become strategic: In the knowledge-based economy, companies win by providing effective access to information.

Consider the traditional enterprise information infrastructure: an information value chain with production at the low end, consumption at the high end. The levels in the chain create a pyramid that widens from top to bottom, its width at any point corresponding to investment and use at that level. In most enterprises, expensive RDBMS storage and legacy architectures make this pyramid bottom-heavy. We believe there is an innovative way to turn the cost structure of this pyramid on its head.

Inverting the Pyramid

To quote the CIO of a large, multinational enterprise: “We are world champions at producing information. Finding what we need is an entirely different matter.”

At the bottom of the pyramid are data sources, which in this enterprise include corporate, regional and local documents, in five languages and 220 formats. To manage this information and its associated applications, the enterprise has large teams of database architects, programmers and managers, plenty of software licenses and numerous high-end servers.

Further up are tools and people who organize and refine, people who provide access for knowledge workers. Librarians, application managers and user support spend man-year upon man-year building taxonomies, applying metadata, controlling access and developing new information services.

At the top of the chain are the information consumers: employees, analysts, management and customers looking for product information.

Information access in RDBMS and legacy systems is inflexible, costly and slow. Metadata maintenance is complex, expensive and imprecise; new information services are major undertakings; the system architecture is scattered and expensive to maintain; coordination of work is non-trivial. There is duplication of work; there is duplication of content; and still, important information gets lost. The information is often untimely.

If we introduce an information architecture based on a modern search platform, our bottom-heavy pyramid changes dramatically. A good search platform turns it on its head, and thereby improves information access, reduces total cost of ownership and increases overall enterprise performance.

Immediate Information Discovery

The traditional pyramid grew out of a focus on transactions. There was less data, and what there was of it was mostly structured. Times have changed: Information consumers understand the value of accurate, timely information and know how to use it. Better service affects their performance directly. Better information access is now a strategic requirement. This new “immediate information discovery” model, with intuitive search front ends, relevant results, sub-second performance and advanced tools to filter and refine the answers, drives many new competitive business models.

The immediate information discovery model addresses needs throughout the value chain. In our example enterprise, the management wants to grow the business, so new revenue streams and low TCO are important. IT managers like the innovation potential it affords their services; product managers love the increased user traffic. Existing users relish new ways to access and use information: Contextually sensitive search and navigation, user-specific relevancy ranking, spell checking, search for similar documents...all combine to create new ways to tap company information. New customers are attracted by new search-based services: vertically focused subject search, topic maps, update alerts and search for experts. These capabilities open new revenue channels, increase the...
number of visitors, enhance user satisfaction and improve business performance.

New services include digital reports and books, video and audio content search, enhanced topic search and reference research, all combined with the traditional strengths of the enterprise: subject matter expertise and proprietary content management. And the cost aspect: Through this search-based model, the example enterprise has seen 60%-80% savings in HW/SW acquisition costs through fewer licenses and the use of commodity servers, and 10% savings in operation and scaling costs through faster installations, easier scalability, simplified feature extension and fewer people.

**Semantic Index**

Good enterprise search engines turn the information pyramid around. They enhance the whole information chain, placing the bulk of functionality where it is most needed: at the information consumption point. Some search engines, however, go beyond information access enablement. They transform the whole information storage, processing and consumption model by enabling true semantic—meaning-based—information handling.

This model extends the “semantic web” idea to all content, structured or unstructured. It automates the recognition of information context and meaning, at several levels, and replaces costly and slow specialist data mining tools with a single user-friendly system that enables end-users to create sophisticated queries and get accurate results at lightning speed. The semantic model maintains enterprise control over business and security paradigms, but moves the power balance in information access from producers to consumers, where it belongs.

This contextually aware information model is based on an embedded part of the search engine called the **semantic index**. The heart of a contextual search engine, this index stores files, database content and multimedia with no loss of structural information. It retains HTML structures and database records; it recognizes a practically infinite number of entity names; it knows the granularity of sentences and paragraphs. It translates the original context into searchable elements, and thereby brings an unprecedented level of “understanding” to information providers and consumers. It is grounded in years of research and development, and is the core of next-generation search engines.

This index can be explored on two levels: the mechanistic representation level, which affects the richness with which information can be encoded in the search engine, and the intelligent information analysis level, which affects the user’s power to explore context-rich information. Full contextual representation of data in the semantic index has two main consequences for end users. One is that the index contains enough information to support any XML-based query, and thus opens the way for full integration of structured and unstructured data. It accepts complex queries based on XPath or X-Query and solves them in sub-second time. Any native content can be made searchable with its structure preserved, without a priori knowledge of data structures or schemas. This is true schema flexibility: schema independence, but with full schema exploitation.

Contextual analysis lets users change the granularity of search at will: they can choose the atomic unit of search dynamically. This is a fundamental change in the way search engines work. Traditionally, the unit of indexing (e.g., a Web document or a PDF file) was also the unit of retrieval. Now the unit of recognition and retrieval can be moved to a sub-document level. This means improved search precision and less work for users—no more wading though long documents in search of the relevant sentence, or listening through an entire newscast in search of the

“There is an innovative way to turn the cost structure of this pyramid on its head.”
The Faceted Navigation and Search Revolution

By Steve Papa, Founder and Chairman, Endeca

“The future is faceted.”
—Peter Morville, President and Founder, Semantic Studios

You’ve likely heard the buzz by now: faceted navigation truly changes the ground rules for search and KM, and for the many related applications where users need to find information, ranging from subscription content services to directories. It complements search and relevance ranking to fill a big hole in the process of making knowledge reusable. And not only does faceted navigation help end users far more easily find what they’re looking for, it also helps content owners efficiently manage content...or, in reality, to make it even feasible to manage content at all.

Below, I’ll talk about the distinct properties of Faceted Navigation that create those benefits, and later, explain how classic information science combined with breakthroughs in computer science bring it to life in a form called “guided navigation.”

There is no universal way to categorize content; relevance is in the eye of the beholder.

What is faceted navigation? It’s a way to browse information, or to refine long lists of search results, along multiple dimensions, aka facets. These are orthogonal lenses through which to view the world. For example, I might search for an expert by facets like name, project, company or date—and more likely, by some combination of those facets, selected in any sequence.

Facets feel instantly comfortable because we naturally think of the world from multiple perspectives. It’s like the fable of the five blind men with the elephant, or the tale of Rashomon, where each witness to the same event recounts it from a different angle. There is no universal way to categorize content. Relevance is in the eye of the beholder; the same piece of content might be useful to different people for different reasons, or even to the same person for different reasons depending on their goal.

Busch’s Law

So how do facets work their powers? First, we need to state what I’ll call Busch’s golden law of facets, named for Joseph Busch of Taxonomy Strategies, a past president of the American Society for Information Science and Technology:

Four facets of 10 nodes each have the same discriminatory power as one taxonomy of 10,000 nodes.

That’s stunning. That means that with facets, I can describe a collection with 40 nodes (aka subject categories) that would take a taxonomy 10,000 nodes to describe. That’s for an idealized case, of course, but the gist of it holds true in the real world. The bottom line is that with facets, we can make do with orders of magnitude fewer nodes than we needed in a taxonomy.

That’s because taxonomies are a type of pre-coordinate indexing, meaning that its builder anticipates the compound subjects people can browse along later, like “18th Century French History.” In contrast, faceted navigation is based on post-coordinate indexing, meaning that end-users string together their own compound subjects at search time. They do this by combining simple elements from multiple facets, in this example, (Time: 18th Century) + (Country: France) + (Topical Subject: History).

Reducing the number of categories we need by orders of magnitude leads directly to two primary benefits:

1. Faceted navigation helps users more easily find what they’re looking for. First, and most simply, it’s much easier for users to grok 40 nodes than 10,000, so in practice, they start browsing more. More importantly, faceted navigation offers users many different paths to each item of content—often dozens or hundreds of paths.

For example, before Barnes & Noble.com deployed Endeca, users could browse books through a taxonomy that offered roughly 250,000 paths. Today, with their BookBrowser, Barnes & Noble offers their customers literally billions of paths to browse for books. That difference—between less than a million paths and billions of paths—is the incremental value facets bring.

2. Faceted navigation lets content owners streamline information management processes. Most obviously, if you need orders of magnitude fewer categories or nodes than before, it becomes exponentially simpler to manage them. This leads to cascading benefits. It makes the work of auto-classifiers simpler, since they have fewer buckets to pick from. Also, facets operate independently of each other, which leads to a “schemless” data model. This makes it simple to combine heterogeneous collections, because you don’t need to mesh facets into a single taxonomy, and it makes it simple to add facets incrementally over time.

Facets Pre-date Computers

Librarians have been organizing the world’s information since before Tim Berners-Lee was just a gleam in his parents’ eyes. Of course, the invention of the Web and intranets has created critical new

Steve Papa

Steve Papa, chairman and founder of Endeca, started the company in 1999 to help all people overcome information overload. Under Papa’s leadership, the company has rapidly become a leading provider of innovative information access and analysis software solutions to the Global 2000. Today, Endeca’s marquee customer base includes organizations like Bank of America, Barnes & Noble, Boston Scientific, Classmates.com, NASA, IBM, The Home Depot, Harvard Business School, The Library of Congress, Putnam Investments, Texas Instruments, Wal-Mart and Zagat Survey. Before starting Endeca, Papa was an early employee at the search engine company Inktomi, where he worked with unstructured data on a massive scale, and before that, at the data warehouse company Teradata, where he worked on structured data also on a massive scale.
problems for information scientists to solve, but some universals of knowledge organization persist from the age of bookshelves. Faceted navigation is based on one such pillar of library science, known as faceted classification. This was codified rigorously in the 1930s by S.R. Ranganathan, who is now revered as a founding father of library science, in a book called *Colon Classification*. As David Weinberger wrote of the master in “Rediscovering Ranganathan” in *Forrester Magazine*, “Yes, that is the world’s worst-ever title for a book, but the system it outlines for classifying books—using categories separated by colons—was revolutionary.”

Useful as faceted categories were, in the physical library you still had to ultimately place each book on a single shelf, making it impossible to roam the stacks by the facet of your choice. But mid-century, a librarian named Mortimer Taube took another run at it. Taube’s innovation was the Uniterm system, which marries facets and post-coordinate classification to a mechanical index that made the ideal of faceted browsing possible. Finally, the value of faceted classification was unlocked byfaceted navigation. In Taube’s system, books were cataloged on edge-notched punchcards, where a punched notch indicated the assignment of a facet to that book. To browse for a book, a librarian inserted a rod into a notch in the stack of index cards and lifted, leaving behind just the cards classified by that facet/notch (a variation on the system used light boxes instead of rods to line up notches). Then, he repeated this step with another facet, and so winnowed down to a small set of highly relevant books. Unfortunately, it took considerable training and dexterity to use the system, which means it never made it out from behind the reference desk.

There are some instructive lessons to take away from this brief library history:

1. Faceted *classification* and faceted *navigation* are complementary—but not the same.

Obvious though it might sound, it took both Ranganathan and Taube to produce a great system. You’ll often hear people today use faceted classification interchangeably with faceted navigation, but they don’t always come paired. Faceted classification is widespread today; faceted navigation is still too rare. And this isn’t just a nit to pick because it leads to our next observation:

2. Navigation unlocks the latent value in classification. Almost all content already has some faceted classification or structure that can readily be used for navigation. In fact, rarely have we met a client who didn’t think they had unstructured content, yet almost always do we find it to be full of valuable structure. So why don’t they recognize its latent value? Until you see it with navigation for the first time, it’s not obvious—and once you do, it’s indispensable.

3. Iteration is the key to browsing. Taabe had it right. Contrast these ways of browsing:

   - “I want books on Austrian music history;” or
   - “I want books on music; of those, I want the history ones; of those, I want the ones about Austria.”

The former is a search in the ideal world; it assumes that what we’re looking for exists, and that we’ve accurately predicted how the content will be described. It’s a pre-coordinate search. The latter corresponds more closely to a real-world search, where we constantly deal with “empty shelves.” By taking successive steps along multiple facets, we can winnow down, back up, and avoid dead ends.

4. If the user interface isn’t simple, the system won’t survive.

Taube had the UI wrong, which is why he never made it out from behind the reference desk. His system was good in theory, but everyone must be able to use it, without any training.

**Case in Point**

While we’ve been discussing examples of browsing for books in libraries, chances are you’re reading this because you want to help your business by helping people reuse documents on intranets, engineering designs in PLM systems, intelligence for Homeland Security or some other kind of electronic enterprise content. In fact, the early adopters of faceted navigation have been enterprises, not libraries. In the past five years, Endeca has launched faceted navigation solutions in 20% of the Fortune 100, and among more of the top 100 online retailers than solutions from all competitors.

So what happens when faceted navigation finally makes it back to Ranganathan and Taube’s home turf? In January of 2006, North Carolina State University pioneered the use of faceted navigation on an OPAC, or electronic card catalog, which you can try at www.lib.ncsu.edu/catalog/.

It immediately drew rave reviews from the toughest KM critics out there—the librarians. NCSU themselves called it “a groundbreaking approach...the first of its kind in a library...empowers users to quickly locate the items they’re looking for or to explore the multifaceted research collection in depth, exploiting both the software’s cutting-edge capabilities and the library’s many decades of investment in detailed cataloging and classification.” And just a handful of librarian blog excerpts:

- “The best library catalog interface I have come across to date;”
- “It’s brilliant—faceted browsing and relevance ranking really improve access. It makes the search experience closer to what users expect when searching on the Web;”
- “NCSU libraries butcher the pig. Andrew Pace of the NCSU Libraries, who has said that making minor changes to library catalog systems is like putting lipstick on a pig, has unveiled a revolutionary new interface to their catalog;” and
- “The bar has been raised.”

Ranganathan and Taube had it right all along, and now without the restrictions of 3-D bookshelves, the power of their principles has come to life.

**Guiding Light**

So what is the difference between faceted navigation and Guided Navigation? Guided Navigation is the name for Endeca’s version of faceted navigation. It’s a meeting of classic information science, breakthrough computer science and gritty, real-world software tools and business processes. Guided Navigation wraps search results in a context that shows users how to refine and explore their results, while constantly removing dead ends. With hundreds of commercial deployments now, Endeca has continually built its expertise with leaders like IBM, Home Depot and the DIA back into the best practices reflected in the tools, UIs and APIs that give content owners editorial flexibility and managerial control over their sites, turning faceted navigation into a solution to a business problem.

To learn more about facets and guided navigation, I recommend the first issue of Philip C. Murray’s “Barrington Report on Advanced Knowledge Organization and Retrieval,” and David Weinberger’s “Rediscovering Ranganathan” in *Forrester Magazine*.

Endeca’s Guided Navigation®, Search and Analysis solutions help people find, discover and analyze information in ways never before possible. For technology leaders like Walmart.com, NASA, IBM and John Deere, Endeca is creating value by helping their users more effectively find the information they need. And reflecting the broad applicability of these new solutions across the enterprise, Endeca’s hundreds of commercial deployments span a wide variety of industries—financial services, media and publishing; manufacturing and distribution; government and retail.
Accelerating Business Success with Intent-driven Search Interactions

By Mark Woollen, Vice President, Products, InQuira

In today’s hyper-competitive business environment, companies work hard to differentiate themselves and provide greater value to their customers. Despite numerous investments in Web content and customer relationship management (CRM) technologies, companies continue to struggle with transitioning their customers to lower-cost interaction channels. Businesses often fail to maximize their marketing and sales conversion opportunities due to complex, customer-facing websites that overwhelm users with too much information that bears little relevancy to customers’ specific needs. Across industries, companies struggle to deliver Web self-service options to consumers who would rather not visit the store or dial into the toll-free support line. Marketing and service organizations are beginning to realize that they cannot deliver a successful one-to-one real-time customer experience on the Web without an inherent ability to first understand the customer’s purpose or intent for that specific visit. Nothing accelerates the path to value more than capturing and understanding what your customers want to accomplish in the next five minutes, and delivering a relevant and personalized experience that makes it possible for them to achieve their objectives. Intelligent search is the enabler for exactly that type of experience.

Companies lack neither the knowledge of their customers nor the content required to address their needs. What companies lack is the ability to put their wealth of information contained therein will be a close enough approximation to what the visitor actually needs. The operating assumption is that you design the website experience around the largest segment of visitors—in our examples, the shoppers who already know which vehicle they want to research, or the customers experiencing a common problem. Neither approach addresses the needs of the few—those individuals with unique needs that fall outside the boundaries of the business’ targeted segments. Taken in their aggregate, the “few” will generally represent a larger visitor population than a company’s targeted segments. Amazon and eBay have created new business models leveraging the interconnected nature of the Internet to create markets for “the few.” Yet most companies have not applied those same lessons to their own sites. The underserved visitors represent lost revenue opportunities, and arguably—because they are unable to resolve their specific support needs online—the greatest consumers of expensive CSR expertise.

Find information that address their needs in real time.

Delivering a personalized and effective real-time customer experience requires that companies adopt a new set of best practices for successfully engaging their customers online. These best practices include:

- Understanding the customer’s intent for each online interaction;
- Matching the right enterprise resources to the customer’s objectives for each specific interaction; and
- Delivering an intent-driven personalized experience for each customer interaction.

By implementing these best practices, companies can successfully accelerate their time to value from critical online customer interactions—accelerating, for instance, conversion paths that generate revenue and customer self-service processes that resolve problems.

Understand the Customer’s Intent for Each Online Interaction

People go online to satisfy a wide range of needs, and behavior may vary depending on what it is the visitor is trying to accomplish. This is one of the key reasons it is so challenging to create a navigation-based Web experience that meets the needs of all customers. When customers attempt to use the search facilities of a company’s website to overcome the inherent limitations of a website’s standard navigation, they typically try one of two approaches. Either they will guess at how to articulate their needs in two or three words, hoping the search returns the information they need, but not really believing it will; or they explicitly express their needs much as they would in conversation.

With traditional keyword-based search engines, both approaches carry risks for the searcher—the first will likely return a long list of results that may or may not contain the needed information. The searcher often does not know until he clicks through one or more of the results. The latter approach will return even more results, and it is unlikely they will be relevant to the searcher’s natural language query. In reality, few people characterize their problems, needs or objectives in exactly the same way, so businesses respond by designing customer experiences around the information or questions they deem to be most appropriate for their largest segments of site visitors. Automobile manufacturers structure their websites by model, which, by design, assumes visitors already know which vehicles they want to research.

Many customer support sites direct visitors to a list of top FAQs, hoping that the information contained therein will be a close enough approximation to what the visitor actually needs. The operating assumption is that you design the website experience around the largest segment of visitors—in our examples, the shoppers who already know which vehicle they want to research, or the customers experiencing a common problem. Neither approach addresses the needs of the few—those individuals with unique needs that fall outside the boundaries of the business’ targeted segments. Taken in their aggregate, the “few” will generally represent a larger visitor population than a company’s targeted segments. Amazon and eBay have created new business models leveraging the interconnected nature of the Internet to create markets for “the few.” Yet most companies have not applied those same lessons to their own sites. The underserved visitors represent lost revenue opportunities, and arguably—because they are unable to resolve their specific support needs online—the greatest consumers of expensive CSR expertise.
The diverse way in which people express their needs requires that companies implement intelligent search and dynamic navigation capabilities that can capture and comprehend the customer’s intent for his or her visit to a company’s site. Understanding customer intent is multi-dimensional in nature and requires the ability to understand both a customer’s immediate need, and infer in real time related secondary or tertiary needs. For example, a searcher who wants to know “which vehicles get the best mileage” might next logically inquire about potential fuel cost savings from buying a more efficient car. The ability to comprehend the complete depth of customer need—the customer’s intent—is essential for addressing their most urgent requests immediately and intelligently anticipating related information needs. By implementing an ability to understand customer intent, companies can significantly improve their customers’ online experience and accelerate customers’ paths to success.

**Match Enterprise Resources to Customer Intent for Each Interaction**

The ability to comprehend customer intent provides companies with powerful capabilities to determine which enterprise resources are best equipped to address customer needs. All too often, companies overwhelm their customers with too much content that contains too little customer relevance. Understanding the intent behind each customer interaction enables companies to more efficiently focus their enterprise resources on customer-specific needs. Content, in this context, need not be limited to static information. Companies can also leverage an understanding of intent to determine if and when transactional systems and their related business processes should be exposed to customers. For example, if customer inquiries related to billing (e.g., disputes, balance inquiries, etc.) are a common customer need, then integrating search with the enterprise billing system is of critical importance in being able to adequately address customer needs. Matching enterprise resources to customer intent provides powerful dividends to all parties. Customers obtain easier, more efficient access to the content and data that address their specific needs, leading to greater satisfaction and desire to use the Web as their primary channel of interaction. Better alignment between content management and customer intent enables more effective use of existing content, eliminates the creation of redundant content, and extends the reach of transaction-based systems-of-record to provide customers with real-time access to critical information.

Businesses that succeed in matching enterprise resources to the real-time intent behind each unique customer interaction will create a distinct competitive advantage. Deployed properly, this ability to match enterprise resources with the real-time needs of customers can accelerate marketing and sales conversion paths, or problem resolution processes, creating value for both the visitor and the company whose site the customer is visiting.

**Deliver an Intent-driven Personalized Experience for Each Customer Interaction**

The ability to deliver a real-time, personalized experience that successfully addresses the needs of the visitor is the only true barometer for successful online customer interactions. Determining the customer’s intent for a particular visit allows companies to understand the full extent of their customers’ needs. Matching customer intent to enterprise content and data is critical for identifying the right information to address the customer’s complete set of needs. The intent-driven experience enables companies to present a unified face to the customer—one where service, sales and business partners can present relevant solutions and offers specific to the context of a customer’s need. By providing customers with real-time, personalized responses to their inquiries, companies are initiating a Web-based dialogue with their customers. Should customers need to perform additional research to achieve their objective for that visit, the personalized experience provides them with personalized navigation that can be used to guide customers to their desired information.

Customer intent-driven search provides companies with a powerful means of differentiating themselves in a crowded and confusing online market. By focusing on the needs of customers and delivering appropriate content and processes specific to those needs, companies can accelerate their time to value from every customer interaction.

InQuira improves the quality of customer interactions through websites and contact centers with integrated intelligent search, knowledge management and analytics applications. InQuira’s solutions leverage patented intelligent search technology to understand a customer’s intent in real time, and dynamically create a personalized response that incorporates the right information, data and links to applications from across an enterprise to meet that customer’s need. InQuira’s blue-chip customers include AT&T, Bank of America, Chrysler, BEA Systems, Fidelity Investments, Honda, Sovereign Bank and SunTrust.

The InQuira 7 suite of applications enables companies to provide customer-centric service, online and through assisted service channels, across marketing, sales, service and support business functions. More information is available at www.inquira.com.

**“The ability to deliver a real-time, personalized experience that addresses the needs of the visitor is the only true barometer for successful online customer interactions.”**

**“All too often, companies overwhelm their visitors with too much content that contains too little customer relevance.”**
Business Semantics Management

Empowered Enterprise Search Drives Competitive Advantage

By Jeff Dirks, President and Chief Executive Officer, SchemaLogic

For Global 5000 enterprises, the ability to optimize information management for competitive advantage will drive growth in market share while increasing margins through greater business process efficiencies. Your employees and partners need the ability to access the right information at the right time to make the day-to-day business decisions that support your enterprise. While optimized access to information is crucial, it won’t be achieved until your enterprise systems can efficiently sift through your large and growing body of content to deliver relevant, comprehensive results. Business semantics management (BSM) enables greater findability for key assets across the hundreds of content and data repositories that exist in the typical global enterprise of today. BSM manages the ever-changing semantics of the business—throughout its lifecycle.

For most large companies today, one thing stands in the way of optimized information access: semantic inconsistency. IT organizations in large companies have typically deployed numerous information management systems, which have never learned to speak the same language. For example, the application that manages content generated by the R&D group for a major pharmaceutical company knows the newest product offering as CardioFDATrial3, while marketing, purchasing and shipping know it as Heart Health. In the event of a recall or proposed product name change, how will the pharmaceutical company find all of the information related to that product necessary to support an FDA recall or global re-branding effort?

BSM technology allows organizations to specifically address this universal problem. And, as we’ll discuss in detail in this article, the business benefits extend further than most companies realize.

The Evolution of Enterprise Information

Nearly two decades ago, companies began to introduce databases and enterprise applications to manage financial, manufacturing and supply chain transactions for lines of business. The data these programs generated was typically the result of a transaction (Invoice posted, inventory depleted, sales order shipped) resulting in data being stored in rows and columns in a database table. Access to this data, however, was limited to a handful of technical and knowledge workers. The result was rigid systems unable to respond to changes within the business and not representative of how over 80% of the information assets in today’s typical company are created, stored and accessed.

Since that time, the amount of unstructured information generated by companies—information contained in documents, spreadsheets, email and Web sites—has exploded. This type of content represents the majority of a typical company’s information assets. Because unstructured content is incredibly complex, the different applications used to manage this information have each developed their own languages/semantics for categorizing and classifying it. At the same time this information evolution was occurring, companies were becoming increasingly more virtual and global. The resulting cultural and language differences introduced even more inconsistency.

The systems that manage unstructured information do so within a certain context or perspective—one that is shared exclusively by the employees or groups that use the systems. For a large enterprise, employees need to access all relevant information that will help them do their jobs and make day-to-day business decisions, regardless of the original
context of that information. Unfortunately, this is not possible in today’s enterprise information environment.

The Market Demand for Business Semantics Management

In addition to the strong business case for optimizing information access, employees, due to what has been called the “Google effect,” now have very high expectations when it comes to information accessibility. Because enterprise information is stored in silos, in disparate applications, and is categorized and classified using inconsistent semantics, information access today falls far short of what is needed by the business and expected by employees.

It’s estimated that 35% to 60% of the $300 billion spent annually on systems integration is due to semantic inconsistency—a reality that represents a gaping hole in the information management fabric of the enterprise. This does not imply shortcomings in the applications commonly used to manage information, rather it demonstrates a market need for an overarching solution that manages and reconciles semantics across enterprise applications.

Business semantics management solutions enrich and extend existing applications and help them reach their full potential. And they enable companies to deliver comprehensive search results that are more relevant, more precise and more appropriate for the enterprise than Google ever dreamed of.

For most large companies, corporate terminology and semantics are under the control of a select few technical and knowledge workers, like database administrators and taxonomists. These workers manage a very structured data world using databases, spreadsheets and documents. Unfortunately, more than 80% of an organization’s information assets exist outside the structured world and knowledge workers cannot access the information due to the complex and diverse storage methods. BSM solutions provide a framework for managing, reconciling and evolving enterprise semantics across both unstructured and structured information assets. When all content is classified and categorized based on a system that understands the links between inconsistent terminologies, the optimized information access will lead to improved communications and decision making within the organization and among customers, suppliers and partners.

The Role of Business Semantics Management as a Service

The role of BSM solutions, such as those pioneered by SchemaLogic, is to create an overarching semantic model to extend enterprise taxonomy deeper within the organization, as well as outside the firewall. As changes occur in the business, these changes will be expressed via the semantics that describe the business—be it a formula, a product name, a marketing brochure or a job description. BSM provides a way to manage these changes, notify the right parties and keep critical systems in sync.

When the solution is deployed across the enterprise, employees, departments, customers and partners are given real-time access to information trapped in silos across enterprise applications, as well as to “orphaned” unstructured data. Real-time access to this information assets enables companies to make better, more insightful business decisions, as well as optimize resources and operations. In addition, BSM helps companies bring governance to an area that was formerly too complex to monitor or manage.

From a practical point of view, managing, reconciling and evolving enterprise information assets through business semantics management solutions enables:

- **B2B and B2C applications**: Better match of goods and services to prospective buyers and more effective participation in trading communities spanning supply and demand chains.
- **Enterprise search and vertical search**: Optimize enterprise search systems by managing, reconciling and evolving enterprise taxonomies and metadata.
- **Enterprise portals**: Ensure employees, partners and customers have access to the information they need by optimizing access to information assets.

Benefits Speak for Themselves

Companies are gradually beginning to recognize that information management, while essential, cannot alone optimize the “findability” and timely delivery of a company’s information assets. The current disconnect between structured and unstructured information, coupled with the organic evolution of the semantics that describe that information, requires enterprises to have an overarching semantic framework with the capability to reconcile inconsistencies across information management systems.

In addition, as information continues to change and grow, a critical component of BSM is a platform and process to engage stakeholders across the enterprise—as opposed to only technical and knowledge workers—in the continued evolution of the semantic framework.

Business semantics management makes companies more competitive by:

- Enriching information management and search applications;
- Supporting and facilitating SOA and Web service oriented architectures;
- Optimizing global brand and product management;
- Enabling governance across structured and unstructured information silos;
- Improving communication/participation among employees, and with customers, suppliers and partners; and
- Fueling development of the business insights that enable your enterprise to make critical decisions.

Use Case: IBM Self-Service Portal

IBM turned to SchemaLogic for help with a key component of their W3 Portal (an HR initiative). The company needed a solution to power a tool called “Blue Pages,” which would allow managers at IBM to locate resources within the company that have very specialized skills. As with any large organization, IBM had many ways of describing the same skills and expertise across the company. Without a way to reconcile these inconsistencies across enterprise systems, managers could not identify internal resources with the necessary skills, and instead, needlessly hired outside consultants to support their projects.

With BSM from SchemaLogic, IBM better utilizes its existing resources and reduces reliance on outside contractors. IBM estimates the solution has saved the company $680 million. IBM is currently expanding the technology to the Global Services and Business Consulting arms of the business.

SchemaLogic is a global leader in reconciliation and change management for Global 5000 companies, providing a business semantics management (BSM) solution to manage the ever-changing language of the enterprise.

SchemaLogic’s solution suite helps organizations establish “communities of collaboration” through living corporate semantics (terminology, definitions, knowledge and know-how) that enable enterprises to increase their competitive advantage and reduce operational costs.

For more information about SchemaLogic, call: 425.885.9695 or visit www.schemalogic.com.
Comprehensive E-discovery and E-disclosure Technologies

Next-generation Deployment for Enterprise Search Tools

By Johannes C. Scholtes, President & CEO, ZyLAB North America

Consider the following: Throughout the course of the Enron investigation, the FBI has confiscated 12 million pages of paper documents plus terabytes of e-mail and other electronic files. Investigators faced a huge challenge in being able to thoroughly, accurately and promptly archive all of this information, make it completely searchable, conduct sophisticated queries and create evidence annotations. Aside from having to achieve all of these goals, investigators also had to organize and disclose evidence in a logical, complete and auditable manner.

A vital element within organizations has risen to the point that “more than half of Global 2000 enterprises will have adopted an e-discovery software solution” by 2007. With this growth, the number and types of users involved with e-discovery activities—as a necessary component of both increased operational efficiency and defense against negative audits—will increase. Regardless of the individual roles and responsibilities various users may have, at the core of their e-discovery activities will be their desired capability to quickly and accurately search and find, on an enterprise level, the data for which they are looking.

“Enterprise search is not, in and of itself, enough to address the requirements of large-scale investigations.”

For another challenging example, consider the information discovery logistics faced by the investigators involved with the International Criminal Tribunal for the former Yugoslavia (ICTY). This ongoing case involves the analysis of tens of millions of pages of case documentation, in more than 10 languages, as well as vast amounts of electronic files, e-mails and attachments, third-party databases and the testimonies of hundreds of suspects, witnesses and victims. Adding to the complexity, hundreds of attorneys and prosecutors involved in the case can request different sets of data. Needless to say, the question of how best to manage and distribute all the case information was one of the first issues that had to be addressed.

These stories provide just a couple of examples of why, according to the recent Gartner report, Understanding the E-Discovery Vendor Landscape, the consideration of e-discovery as a key consideration that must accompany enterprise search capabilities in order to effectively conduct large-scale e-discovery and e-disclosure activities.

Even though principles of enterprise search may be at the heart of any good e-discovery (and, by extension, e-disclosure) solution, traditional enterprise search is not, in and of itself, enough to address the full range of requirements users face when conducting large-scale investigations. Times have changed, and many vendors currently claiming to be e-discovery providers still are not completely tuned in to this new reality. Subsequently, these vendors fail to realize and properly address the new sets of rules and expectations associated with not only enterprise search technology in an e-discovery context but also the full range of corresponding tools needed to create the comprehensive e-discovery and e-disclosure solutions demanded by an increasingly sophisticated user base.

In addition to requiring redeployed enterprise-search technology, large-scale investigations like those presented here also need extensive capabilities for long-term data storage and comprehensive knowledge management. The following sections profile the key considerations that must accompany enterprise search capabilities in order to effectively conduct large-scale e-discovery and e-disclosure activities.

Keys to Success

1. Secure and sustainable long-term data storage. Taking into consideration the long-term, high-volume archiving needs profiled in the stories above, in addition to the need to handle a variety of data types in a very secure environment, only XML-based storage framework can address these needs. For criminal tribunals, for example, it may take decades before enough evidence is found, processed and presented to warrant the arrest of a war-crime suspect.

In an instance such as this, having ready access to all of the original investigation data is imperative, as is the need to be able to work with it in its original form, without expensive conversions. Only an open, XML-based framework provides a truly viable solution.

2. Ability to handle all discovery factors involved in large-scale investigations. During the development of a large-scale e-discovery project, all required tasks need to be acknowledged and taken into account when evaluating the feasibility of any proposed technology solution. Many available tool suites can do a good job on a certain aspect of e-discovery (such as large-scale storage), but few can actually address on a suitable level all of the following tasks that combine to make a comprehensive e-discovery solution:

   Full-scale data capturing capabilities: An investigation can only start if all targeted data has been captured. Although many current tool suites default to electronic data as the
model on which to focus their capturing capabilities, one of the biggest capturing issues to address in large-scale investigations is still paper focused. To make sure that all information is accounted for, capabilities must be in place to handle low-quality, odd-sized, double-sided and/or corrupt documents as well as recover lost clusters and deleted files. Ideally, any e-discovery solution should have appropriate capturing tools for all types of media: paper, email, multimedia or other types of electronic files.

In terms of the OCR component of the capturing process, other capabilities need to be taken into consideration, such as automatic language recognition, bitmap identification and the OCR-ing of bitmaps in attachments. In addition, a full range of auditing and analysis tools, plus tools for workflow, digital signatures, and the recovery of deleted files and lost clusters, are all very important aspects of any e-discovery software solution. These tools help support the duties of various responsible parties and can help expedite the discovery process (given that an actual discovery strategy is in place to support).

In addition, a very clear link must exist between physical data (paper, hard disks, servers) and electronic copies. Inventories that were archived before conversion to electronic archives need to be accounted for. Quality control mechanisms must be part of the equation because the existence of these types of procedural components will be a prominent aspect of any audit.

Advanced searching capabilities, particularly fuzzy and multilingual. The key motivation for users engaged in e-discovery is, simply, to not miss any archived e-mail, document or electronic file that may have relevance to their particular investigative activity. Therefore, e-discovery searching should concentrate on 100% recall rather than on precision (like with Google). However, attaining full recall can be compromised not only by the existence of multiple languages in a document repository and/or different document sizes, structures, types and domains, but also because many times users simply do not know exactly what it is they are looking for. Fuzzy search and multilingual search tools are essential in this context. Good investigators will recognize a smoking gun, but they need the tools that enable them to find the words to retrieve it.

Hit-highlighting, hit-navigation, customizable relevance ranking, optimized use interaction, text-mining and visualization tools: Sophisticated searching capabilities in an e-discovery context must carry with them accompanying tools to help users review large result sets: hit-highlighting, hit-navigation, customizable relevance ranking, optimized user interaction, text-mining and visualization tools are among the essential tools. Being able to search large amounts of information is one thing, but to be able to actually find specific bits of information is quite another.

ZyLAB builds solutions that meet companies’ e-discovery and e-disclosure needs. ZyLAB’s signature tool suite, ZyIMAGE, was originally designed alongside legal, security and law enforcement professionals to meet the exact e-discovery and e-disclosure capabilities profiled in this paper. Investigators worldwide rely on ZyIMAGE’s open XML-based information access platform as a framework with which to manage huge data sets. The advanced enterprise-search capabilities and robust text-mining tools available in ZyIMAGE—coupled with support for over 200 languages and 370 file formats, including paper, e-mail and multimedia—continue to make ZyLAB a leading solution provider for the most complex and far-reaching investigations. In fact, ZyLAB is the preferred software provider for the forensic units of KPMG and PricewaterhouseCoopers, as well the FBI (Enron), the International Criminal Tribunals for Rwanda and the former Yugoslavia, and the Italian Justice Department (Pramat fraud investigation, the largest ever in Europe), just to name a few.

“Being able to search large amounts of information is one thing, but to be able to actually find specific bits of information is quite another.”

Of course, format and bundle information in a logical way, create and manage annotations or review timelines for specified users, alert users when new information has been added to a folder, develop newsgroups for discussing information and so on.

3. Ability to handle all disclosure factors involved in large-scale investigations.

Finding and implementing tools that support established disclosure procedures is tricky, as the quality of any type of software solution is directly related to the appropriateness and thoroughness of the internal procedures it supports. Obviously, disclosure procedures must be set up in a way that they meet the needs of the company and the mandated disclosure rules to which they must adhere. Moreover, procedures must be “do-able” by the staff tasked with fulfilling them, particularly in terms of staff being properly trained and understanding the rationale behind their document and records management activities.

On a structural level, when information is ready for disclosure, the solution that supports this activity must be able to publish the information in a variety of formats, depending on procedural or recipient requirements. An innovation solution should be able to easily disclose information in PDF, TIFF, DII, XML and so on, as well as publish to third-party solutions (such as legal software like CaseMap).

ZyLAB is an innovative developer of affordable content management and compliance solutions for paper-intensive organizations. ZyIMAGE, ZyLAB’s flagship solution, helps small and medium-sized businesses (SMBs) and government organizations digitally file and manage millions of pages of paper, electronic documents and email. High-quality search and retrieval features (which support over 200 languages) give users the ability to easily organize, investigate and distribute information.

With more than 7,000 installations worldwide and more than 500,000 users, ZyLAB has a wide breadth of experience and knowledge across a variety of different industries and business applications. For more information visit: www.zylab.com
Making the Business Case for Search

By Chris Weitz, Managing Director, BearingPoint

The performance of any organization depends heavily on how well its people use information to make decisions and do their jobs. First, though, they have to find the information they need.

And that’s not so easy. It’s estimated that knowledge workers spend up to half their time looking for actionable information—information they can use to bring value to the organization. This is hugely frustrating, especially to frequent users of Internet search engines who have become accustomed to the speed, power and convenience of widely available, powerful search technology. Increasingly, enterprise users also are demanding faster and easier ways to locate the information they need for decision making and task completion.

A Changing Landscape Driven by Users

Due to the increased availability of broadband communications, the general public and workers alike increasingly rely on the Internet for everyday tasks and activities. In fact, many users take for granted that the Web will offer services that are easy to understand, easy to use and accessible anytime, anywhere. At the same time, many public- and private-sector organizations are now using Internet software services as a way to reduce costs and deliver services to a demanding and physically diverse end-user population.

These influences, especially the rapidly changing expectations of enterprise users, are driving a major shift in the software landscape. Enterprise users now expect the same level of functionality, transportability and ease-of-use in the enterprise as they experience as consumers on the Internet. The hard line between consumer software and enterprise software is steadily disappearing.

As a result, software and solutions vendors must de-emphasize the old model of large, highly complex, monolithic systems (and integration projects) and create solutions that are rapidly innovated, Internet-based, modular and easy to deploy and deploy. For companies to sustain this rapid innovation, speed of development and simplicity are critical factors to their success.

With the increasing demand for search capabilities in the enterprise, it is important to understand what type of information enterprise users need and want. Today, most users conduct two types of search in the enterprise, as described below.

Some knowledge workers spend most of their time engaged in specialized search. They log into a specific system either to update or use data. For example, an analyst may search within the finance system to recover all accounts-receivable records associated with a specific customer. An auditor looking for evidence may search within a messaging archive for all the e-mails sent between two people over the last six years. A government intelligence agency may search all phone records for a specific individual or organization.

In contrast, searching the enterprise is the generalized search for information in any location, within a department, across the entire organization and beyond. Searching the enterprise is like Internet search, which aggregates results from all available information sources into one unified result. It is different from specialized search, which has complex search strings and filtering of data sources. The source of information in a search of the enterprise can range from the general to the specific, from arbitrarily found Web pages to structured documents and databases.

Search solutions for the enterprise can be used for recovery, discovery and exploration, for example:

**Recovery**
- Where is the schematic for this part number?
- Where is the resume of that job candidate?
- What documents satisfy this audit request?

**Discovery**
- What are the current organizational policies?
- Who else in the organization or elsewhere has done something similar?
- Who has done something I can reuse?

**Exploration**
- What does the press think?
- What are our competitors doing in this area?
- What does/should it cost?

By improving individual employee productivity, search solutions for the enterprise offer a number of significant benefits. They can improve customer service by helping service representatives answer customer questions and solve problems more quickly, efficiently and completely. They can contribute to higher sales by making crucial information more readily available to the sales force. They can improve the quality of decisions by making the relevant information available.

Search solutions for the enterprise also can help improve an organization’s agility and the ability to respond to market threats. They are a way to leverage intellectual capital by finding, rather than recreating, existing information. They foster innovation through improved information discovery and can contribute to shortening response and development cycle times.

Why It’s Hard to Find What You Need

Why do people spend so much time searching for information? To start with, the number and complexity of sources of useful information are growing constantly. That means more data is stored in more places and is harder to find—in enterprise transactional databases and knowledge repositories, as well as in unstructured document collections both within and outside the enterprise.

As a result, searching for information today can be inconvenient, ineffective and time consuming. Users may have to log into different systems and use different techniques to drill down for what they need. Disparate user interfaces present data differently, creating inconsistencies and confusion. For the user, this means lost productivity.

The lack of effective search solutions can represent a huge hidden burden. Significant staff time is lost searching for information, time that could be reclaimed by enabling good search solutions.

Existing information portals provide some relief. But as the number of underlying systems grows, portals must be constantly expanded to keep up. Such expansion can lag the availability of data ready to be put to use, as well as be expensive and hard to scale.

On the other hand, general-purpose search solutions are not primarily focused on compliance or deep research problems, where every data record has significance. Although they can return all items matched by narrow and focused search requests, search solutions for the enterprise typically strive to satisfy efficiently the most widely needed types of requests for the broadest community of users.

To be effective, search solutions for the enterprise should be:

**Universal**—The home page of the search solution should serve as the general user interface to the organization’s information. It will only become such a tool if end users perceive it as the best starting point. Even if, as is the case at the beginning of deployment, it does not encompass...
all available information, it should include enough to be immediately useful.

**Intuitive**—End user experience is paramount. A common, simple and familiar interface is key to user adoption, allowing the delivery of the timely, relevant content required by employees to do their jobs. While the search engine will scour many repositories with different search tools, details of each repository’s personality should not show in the returned results.

**Responsive**—The search solution must perform well if users are to accept it. This means that the solution should pre-index and pre-store results and provide rapid return of search results.

**Secure**—The search solution should search all accessible content and only display results appropriate to the role of the viewer with the same level of controls in place.

**Expandable**—The search solution must provide for the seamless addition of capabilities to access and index new repositories and new data better.

**Cost-effective**—The search solution should not require complex administration, be hard to use or be expensive to deploy. A turnkey solution requiring minimal maintenance is often preferable. While the amount of enterprise storage now grows exponentially, the cost of search should not do the same.

**Scalable**—Scalability goes hand in hand with cost-effectiveness. Increasing search volume should not result in higher cost. Boosting capacity should require minimal hardware, and the process of adding that hardware should be undetected by the user community. Systems that require intermediation—tagging, classifying and organizing information—tend to be harder to maintain and not scalable.

**Robust**—As an enterprise utility, the search solution should be highly available and tolerate disruption of the surrounding infrastructure.

### Integrating Search Into the Enterprise

Deploying a basic search solution for the enterprise can be as simple as installing it on the organization’s network, pointing it at an intranet repository and turning it on. The target user community will realize immediate benefit with minimal configuration and customization. However, large-scale deployment across complex organizations with database retrieval requires consideration of other issues to truly unlock the value that can be provided by search solutions, specifically:

- **The search solution should interact with existing knowledge management systems.** A search solution can deliver the greatest value if it complements existing knowledge management tools. Interfaces should leverage existing investments in content management. Regardless of how much work may have already been completed within an enterprise, search solutions will provide incremental benefits to the existing solution.

- **Search results returned to an individual knowledge worker should be highly relevant to the task they are trying to accomplish.** A critical component for broad adoption will be how effective users find the actual results of their search. Providing as much relevancy as possible through mass customization will be a key component.

- **The search solution should integrate with heterogeneous systems for retrieval of key information and make data available to the search tool as it is dynamically updated.** A good search solution will recognize various file types found within the organization. This offers the potential to obtain information from a wide range of structured and unstructured data sources, creating a comprehensive search environment.

\[
\text{“The hard line between consumer software and enterprise software is steadily disappearing.”}
\]

Obtaining this information frequently requires:

- **Custom adaptors**—Search engines can find and index many types of information, but extracting information from enterprise resource planning, customer relationship management and other systems requires additional software. Information must be extracted in a way that preserves the integrity and security of underlying data, while providing access to people with the need to obtain it.

- **Database feeds**—Enterprises need to make selected content of key databases readily accessible through search solutions. This data, which is typically dynamic, is made accessible through standard interfaces, such as extensible markup language (XML) feeds, without affecting the source system and network response time.

- **Search tools should fit into the organization’s overall information architecture.** A good search solution should be deployed so that it effectively and efficiently integrates with the overall enterprise IT environment. Two key areas of concern are security and disaster recovery:

- **Security**—Searching the enterprise highlights a range of security challenges related to consolidating information and presenting that consolidated information to users. The nature of these challenges depends on the compliance and regulatory issues of specific industries and the current security practices of the organization. Enterprise security must be evaluated in the context of today’s rapidly changing environment.

- **Disaster recovery**—Because it does not store user data, a search solution may not fit the profile of a typical database or application system in disaster recovery strategies. However, a catastrophic system failure or data loss in the enterprise would compromise the effectiveness of search. A suitable disaster recovery strategy, encompassing the search solution for the enterprise, should be implemented.

### Realizing the Benefits of a Search Solution

An effective search solution for the enterprise can deliver many benefits. For end users, it can increase productivity by making relevant information easily accessible and retrievable. Workers can focus on using information, not on finding, storing and organizing it, gaining the full value of the enterprise’s new search capability.

The use of a familiar interface also can reduce the need for training and dramatically increase the level of adoption by the user community. A search solution also can support significant operational improvement and lower IT expenses by reducing dependency on numerous systems and integration points.

Once enterprises fully understand the benefits of leveraging search solutions to improve the productivity and effectiveness of their workforce, they can then begin to explore the opportunities to transform the ways they use their information as never before possible.

That sounds like a real winning formula!  

---

*Chris Weitz, BearingPoint managing director, leads the search solutions practice. He has more than 25 years of experience with information systems spanning a wide range of organizations, from Fortune 500 companies to technology startups. At BearingPoint, he manages the planning, development, integration and implementation of search solutions for the enterprise and other major IT projects worldwide. He can be reached at 1.866.661.FIND.*
In the legal community, search technology has become as critical to the preparation of major lawsuits as securing witnesses and researching case law. The electronic file in a given case can easily encompass hundreds of thousands of pages. Using keyword filters to mine that mountain of information can not only be inefficient but also leave important evidence buried simply because it fails to precisely match the language of the query.

To solve the problem, Foley & Lardner LLP recently became the first law firm to equip attorneys and staff with in-house concept searching capabilities when reviewing electronic evidence. The concept search engine forms the cornerstone of a custom litigation-support platform that Foley built for its large national litigation department.

The system promises to save hours of work while also yielding more thorough results, including information that cannot be found through keyword searches. This stems from the system’s ability to infer meaning and perform human-like associations through sophisticated mathematical algorithms modeled on human information processing.

As a result, the system can find relevant documents even if they do not share words or phrases with a query. A search on asbestos, for example, will automatically return documents concerning associated topics such as mesothelioma and Johns-Manville because the system “understands” the relationships among them.

In addition, users can search by sentence, paragraph or the entire document as well as word and phrase; perform thorough document reviews even if they have no knowledge of specific industry terminology; use natural language queries without complex syntax; and conduct conceptual, fielded, Boolean and metadata searches simultaneously.

Digging Deeper

Foley’s commitment to concept searching reflects the ongoing explosion in electronic data associated with litigation and the need to find better ways to explore the data pool. Precision searching is imperative because any given document may contain a smoking gun or exculpatory evidence that may influence the outcome of a case.

In 2004, for example, former star investment banker Frank Quattrone was convicted of obstructing a federal investigation into stock offerings at Credit Suisse First Boston in part because of an email telling employees it was “time to clean up those files” after he learned of the investigation. Similarly, a Massachusetts wrongful death suit involving the diet drug combination Phen-Fen was quickly settled after the discovery of an email from an employee in the pharmaceutical company’s accounting department who complained about “writing checks to fat people worried about a silly lung problem.”

Concept searching reduces the risk that this kind of vital evidence will be overlooked. It also reduces the number of queries required to complete the electronic discovery process. This, in turn, shortens the review cycle and helps control costs to the client.

Foley expects not only to shorten and sharpen its electronic discovery efforts, but also to allocate its human resources more intelligently. This is because the firm is using the relevance rankings returned by the system to determine which issues will be handled by senior attorneys, associates and paralegals, respectively.

Significance for Enterprises

While Foley’s embrace of concept-based searching involves the ability to hunt for information across disparate internal data stores. This saves time by eliminating the need for users to search each business application and database individually. It also allows results from all sources to be returned in a single results set for faster processing.

With the availability of concept search engines like Semetric that can be embedded in custom applications with virtually any combination of components, organizations today have new opportunities to harness the power of conceptual search for uses ranging from electronic discovery to corporate knowledge management systems, intranets, customer service portals and beyond. For law firms, Foley & Lardner is leading the way.

Leveraging Semetric’s Advantages

By choosing Engenium’s Semetric technology as its concept search platform, Foley also gained a number of additional advantages related to Semetric’s patented architecture. These include:

◆ Complete automation—Unlike other products, Semetric does not require manual assembly, input and updating of synonym lists, lexicons or thesauri in order to build the knowledge set required for document searches. Instead, Semetric indexes and adjusts to changes in document sets and new vocabulary automatically. This dramatically reduces setup and maintenance time, lowering overall costs.

◆ Progressively smarter searches—Semetric’s patented “portability” from one information set to another is enabling Foley to share the knowledge set built for a given case with each successive case in the same litigation category or practice area. This feature—unique in the industry—provides an incremental improvement in search results that, in turn, will give Foley’s litigators a competitive advantage.

◆ Preliminary case overview—In addition to Engenium’s Semetric search engine, Foley is using the company’s Autometric clustering engine to make it possible to intelligently group documents according to theme even before users begin the search process. These thematic “clusters” provide a snapshot of key issues that can help users structure and prioritize their research. Any document in any “cluster” can then be searched through Semetric, yielding maximum results in minimum time.

Engenium is a leading provider of intelligent search and search-related products designed to improve information access for business users. Founded to overcome the shortcomings of existing information retrieval solutions, the company has incorporated a full range of search capabilities from keyword to conceptual in its patented and award-winning Semetric® platform in order to streamline interaction with business applications and services.
Analytics: Measuring the Impact of Search

By Ian Davies, Founder and Managing Director, ISYS Search Software

Quicker than you can say Sarbanes-Oxley, the enterprise search market has gone from serving a relatively quiet existence as a nice-to-have technology to playing a critical role in today’s emerging information management solutions. The good news is this evolution has spawned rapid innovation, greater extensibility and a broader range of available solutions. The bad news is the landscape is muddier than ever, leaving customers with the difficult task of making sense of the market and its players.

Aside from standard time-to-market and pricing variables, organizations are now entertaining additional key considerations when evaluating enterprise search. It isn’t uncommon to hear, “Okay, I found my documents. What else can you do for me?” In other words, customers are expecting more from their search investments, looking to vendors to deliver greater value and a higher level of proactive knowledge management capabilities.

Beyond Search

Although not as pervasive among enterprise search technology as one might think, search analytics serves as one way the industry is helping deliver greater value to its customers. Whereas a Web analytics software package tells you how your visitors navigate your site, search analytics offers clear reporting on what these users actually search for, what they find and don’t find and the trends in their search behavior.

Consider some sample reports that search analytics delivers:

◆ Most popular searches over a given period of time;
◆ Top searches with no results found;
◆ Searches with no results viewed;
◆ Top ascending and descending searches; and
◆ Search volume (peak and quiet times).

The benefits of search analytics are broad in appeal, yet few businesses are fully leveraging this valuable insight. Consider the potential ramifications of tapping such a tool across several professions in an organization:

Content managers. A primary goal of content management is to optimize the structure of information to ensure higher quality classification and retrieval. Analytics plays a vital role in directing content managers to deficiencies in their content. It does this by highlighting searches that yield no result views, thus prompting these professionals to either tweak metadata or create new content to better satisfy a user’s expressed interest.

Administrators and Webmasters. Standard reporting features also illustrate top searches that produce no results. Often, these stem from incorrect spelling or variations in word choice. By analyzing these occurrences, administrators can opt to create custom synonyms to establish an association between what users enter and what the company presents as standard terminology.

Marketers and merchandisers. One of the challenges in both marketing and merchandising is anticipating what customers want and addressing those needs at the most critical time. By monitoring ascending and descending trends in search terms, marketers and merchandisers can construct campaigns and promotions designed to capitalize on hot trends or resurrect interest in a waning topic. Additionally, by simply viewing a list of top searches, these individuals can use this insight as keyword fodder for pay-per-click advertising programs like Google AdWords.

Executive management. Strategic decisions require considerable quantifiable data. Search analytics supplies this information to management, giving them valuable insight into customer wants and needs. More importantly, these wishes are expressed in a customer’s own words, leaving little question as to the demands and direction of a target market.

It goes without saying that the value inherent in search analytics isn’t confined to public Websites. An organization can apply these tools to all of their online properties, from corporate intranets to partner portals and extranets. Without this insight, the impact and effectiveness of an enterprise search solution can never be fully measured or optimized.

Established in 1988, ISYS Search Software is a global supplier of enterprise search solutions for business and government. The company’s award-winning software suite offers a broad range of products designed for searching desktops, networks, Websites, intranets and custom applications. ISYS has been deployed by thousands of organizations operating in a variety of industries, including government, legal, law enforcement, financial services, healthcare and recruitment.

ISYS SearchTrends serves as a core component of the ISYS:web 7 search solution for Websites, intranets, portals and custom Web applications. Providing out-of-the-box, browser-based reporting capabilities, ISYS SearchTrends delivers detailed insight into user behavior, search trends and patterns and overall search activity. Through this functionality, organizations and their employees can extend the value of their enterprise search solution by applying the insight gained to their everyday work lives.
Adaptive Search and Resolution for Service and Support

By Mark Angel, CTO, Knova Software

If lowering costs, closing more issues at first contact, or increasing customer satisfaction are on your contact center’s to-do list, chances are you have been considering a knowledge management suite—probably including business process support that can integrate with your CRM or incident management system.

Great search is key to achieving the benefits of such a solution—there’s no point in managing knowledge if it’s too hard to find—but most search solutions for service and support fall short for three reasons:

Search for service and support is different—In the standard search model, people know what they’re looking for even if they’re not sure how to ask for it. In service and support, customers have an issue, but they don’t know how to describe the content that will resolve it. In a customer service environment, a natural language processing (NLP) search needs to be geared toward guiding people to a resolution even though they don’t know what it will be.

Search must adapt to user queries—Search companies are often founded on a technology vision: once users have associative search, case-based reasoning, query intent, and support fall short for three reasons:

Search IS a critical component of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Search IS a critical component of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.

Mark Angel has worked in the fields of knowledge management, search, statistics and software development for more than two decades. As CTO of Knova Software, Angel is responsible for the product and technology vision of the company. Angel founded Kanisa in 1997 and joined Knova through a merger with ServiceWare Technologies in 2005.

Knova Software is a leading provider of service resolution management applications that reduce service costs, increase revenues and improve customer satisfaction. Built on a next-generation search and knowledge management platform, Knova’s suite of knowledge-empowered customer service applications automate the resolution process across multiple channels including contact centers, help desks, email, forums and self-service sites. For more information, visit www.knova.com.
The “Other” Search

By Steve Kusmer, Senior VP, Search and Content Solutions, WebSideStory, Inc.

When online marketers and Internet technologists talk about “search,” usually the image of millions of shoppers, researchers and office workers typing queries into Google or Yahoo or MSN comes to mind. However, there is another search that deserves the attention of site administrators and e-business managers. What is this “other” search? It is the search function on your Web site, also known as “site search.” Studies have shown that the conversion rate of visitors using site search on e-commerce sites is about three times that of visitors not using site search. Further, up to 90% of visitors will use site search when looking for your products, services or content.

Delivering Results

Consider the following customers using site search solutions from WebSideStory:

◆ Bloomingdale’s reported that the ROI on their search investment was 10X—or 1,000%.
◆ Palm reported that the conversion rate of their site search increased 60% using a best practice described below.
◆ A Fortune 100 company’s Web site recently saw a 4%-5% decrease in customer support calls after a recent search upgrade.

One in three visitors use site search first—before doing anything else on a site. If they are disappointed in their experience, the chances of them returning are slim. Visitors are telling marketers, in their own words, why they came to the site, and the site is not responding to those requests.

The smartest Web teams and the most savvy Web marketers are investing in and attending to site search with the same high priority that they associate with optimizing for, or buying keywords on, the major search engines, investing in email marketing software solutions, or attracting and converting prospects to customers. Furthermore, by combining powerful site search with informative Web analytics tools, these same website managers are leveraging site search to drive up sales, influence site visitors, better serve customers and reduce costs associated with customer service.

Best Practices

To get the best return on your site search investment, consider the following best practices:

Best Practice #1: Place your search box within your navigation.
Site search should be something that visitors can access on every page of your site, whether they are in the shopping cart, the gift registry, the customer service section or in the product information pages. It should remain in the same place, just like any piece of global navigation. It should be at the top and should be prominent. Your visitors will be nearly twice as likely to use search if you follow this best practice. After you invest and optimize site search, you want visitors to use it.

Best Practice #2: Actively tune the “Valuable Top.” For typical Web sites, the top 100 or 200 search phrases will comprise 40% of all searches done on the site. We call this the “valuable top” of all search phrases, and it warrants your active attention. As an example, Bloomingdale’s discovered through their site search that the popular brand Juicy Couture was near the top of their queries, and tuned their site to deliver branded results.

Best Practice #3: Automatically handle the “Long Tail.” After tuning the valuable top of search phrases, look for patterns among the “long tail”—the rest of your search terms—because there might be clusters that warrant attention. For example, you might see that customers are using product numbers. These can be quickly handled by enabling your site search to search for this data. In addition, make sure your search engine handles misspellings automatically; it’s one of the simplest ways to handle a broad set of queries that lead to results not being found.

Best Practice #4: Provide rich results. Go beyond vanilla text-only search results. For example, the Web team at Palm recognized that the top search terms include product names, such as “zire.” Now, the system injects rich promotions directly into the search results, the site search conversion rate increased 60%.

Best Practice #5: Guide your visitors. Because visitors tend to enter general search terms of only one or two words, provide a facetting and filtering interface on the search results pages so that visitors can narrow their request/query down as they go. Create an experience that’s more like an offline store (browsing, asking questions, seeking help, building relationships between items). If you do not carry something online, but offer it in an offline location, use that opportunity to point the searcher to the answer to their question.

Best Practice #6: Measure, Manage, Optimize. Read your search reports for top searches and searches that result in zero results, and analyze trend data. Analyze how search contributes to various conversion events on your site—registrations, purchases, customer service inquiries. Follow searchers through their entire visit and do path analysis. Incorporate this data in making decisions about content changes, search tuning initiatives and other online activities. Search is a window into the visitor’s mind. Compare that information to the actions they actually take after searching and you will gain valuable information about your potential customers.

Driving Business Performance

Investments in site search at leading Web sites have led to the following representative results:

◆ Site search conversion rate increased by 50%.
◆ Online sales grew by 35%.
◆ Search-to-conversion rate improved by 49%; and
◆ Sales increased by 20% immediately.

If the “other” search is just a box on your site, you are missing a valuable opportunity. Take action on the best practices in this article and you will see immediate benefits to your online business.
Location, Location, Location—With Information, Information, Information

Why Geography Matters in Enterprise Search

By Bill Fisher, Business Development Executive, MetaCarta, Inc.

While browsing through various historical maps featured in the “American Treasures” collection at the Library of Congress the other day, one thing became very evident: For hundreds of years, people have been making maps to pass along wide ranges of information critical to businesses, such as the location of international telegraph communication lines or representations of oceanic currents for international trade. Many of the world’s greatest thinkers and innovators used maps to not only visualize their ideas and inventions, but to also provide information about key business concepts.

Considering the importance of information flow in today’s business environment, I started thinking about the recent increase in the number of enterprise solutions that leverage digital maps. There are so many applications: identifying a disease outbreak, researching a real estate development, looking for the next great oil reserve—the list is enormous.

So, why does geography matter in the enterprise? For public and private entities, enterprise search represents an important tool for knowledge workers to make better, more efficient decisions. Knowledge workers need to know who, what, why, when, how, and where. They need to visualize pertinent information on a map in seconds, with great accuracy, from any work platform.

There’s a problem: approximately 30% of most organizations’ data and documents are not found in a typical search. Finding important information in the enterprise encounters the same problem: unstructured content such as notes, articles and reports may not be accessible. Content stored in structured databases may be, but it represents a smaller fraction of what information exists in the enterprise.

According to the National Academy of Sciences, as much as 85% of content is unstructured. Furthermore, 80% of decisions made by knowledge workers in an enterprise are derived from information stored in unstructured format.

Many of the “silver bullets” of business intelligence reside in unstructured content. For example:

- A string of reports about walk-in patients to a free clinic in Staten Island that have all been diagnosed with a rare strand of influenza;
- An investigative reporter’s story for a small Arizona newspaper that cites specific areas where migrants have crossed the border trying to enter the United States; and
- A thick digital file from a retired geoscientist which contains notes and reports on Alaskan North Slope crude and other key corporate knowledge that spanned his 30-year career.

These examples show that unstructured information is pertinent, but not always widely available and accessible with a search tool. They also demonstrate how location paints a better picture of a problem.

Location + Intelligence = Comprehensive View

Location, location, location—with information, information, information. 70% of all unstructured content contains a geographic reference. Visually representing unstructured content on a map further tells a story. In a knowledge-worker economy, location and information must fuse, but linking them requires enterprise technology that can accurately and logically search structured and unstructured data. This requires a way to geospatially link unstructured content such as documents to a map. This drives the emergence of a new focus in enterprise search: location-based search technology.

With geography becoming more and more important to mission-critical decisions—from national security and intelligence to finding where to drill for oil—this fusion of geographic search for unstructured content provides unmatched context and intelligence about any location.

Geographic text search (GTS) enables users to rapidly locate high-relevance unstructured and structured documents using both keywords and a map as a filter. Documents in various types and formats can be located based on their geographic references contained within the text. Knowledge workers can now answer the question “what do I have in my document collection that talks about this place?”

First, this requires a process called “geoparsing.” Technically speaking, geoparsing is the process of applying natural language processing (NLP) on an unstructured text document to identify geographic references—explicit or implied. This goes beyond simple string matching. By considering all text in a document, a geographic text search system can consider contextual clues to more accurately determine the exact geographic reference and its location mentioned in a document. For example, the word “London” in a document might imply London, England. But it could also refer to “London broil.”

For each geographic reference identified in a document, a set of latitude/longitude coordinates is assigned to the document. Confidence values are used to rank potential locations as defined by a gazetteer. The result is a comprehensive index of documents marked with one or more coordinates. Geographic text search relies on this index to supply accurate search results to users.

GTS systems, such as MetaCarta’s, feature a full-text geographic metadata index, specially optimized for geographic queries. Solutions that supply geographic text search through integration with a dual index strategy produce very low performance query speeds. The solution is a highly specialized, efficient search index for extremely fast searches on simultaneous keyword and geographic queries.

Geographic text search solutions have emerged that enable knowledge workers to find accurate, relevant content in seconds using a map. The ability for any organization to leverage geography to better manage unstructured content creates intelligence that expedites work processes and organizational success. Location combined with intelligent geographic search technology is a harbinger to the knowledge-worker economy.

Bill Fisher brings 10 years of experience in technology marketing to his role as business development executive at MetaCarta, Inc. He drives marketing communications and programs, product marketing, interactive/Web marketing and plays a significant role in building MetaCarta’s brand as the leader in geographic intelligence solutions.

MetaCarta provides users with map-driven geographic search, geographic referencing and data visualization capabilities. MetaCarta products make data and unstructured content “location-aware,” making that information geographically relevant. Founded by a team of MIT researchers in 1999, MetaCarta is privately held, with US headquarters in Cambridge, MA, and offices in Vienna, VA, and Houston, TX. For more information, please visit www.metacarta.com.
Distributed Search Technology

By Ken Bubeck, Manager, Business Development, Discovery Products Group, Open Text, Inc.

The need for fast, accurate search and retrieval continues to be a significant factor in driving software sales for Fortune 1,000 companies and their employees, as well as for the general public. A recent survey found that search is now the second most popular online task behind email. This is due in large part to the commercial and technical success of Google. More and more senior managers are saying “Why can’t I find information in my organization like I can find information on the Web using Google?” Many customers and prospects over the last year have expressed a strong desire in their organizations (often coming from senior management) to consolidate and simplify the searching experience.

Federated search and access to all of an organization’s repositories is also very important. Users would like to have a single search facility that searches all of their available information and repositories, including their own desktops; aggregates and de-duplicates the results; and builds clusters of results. This can be handled by “search federation” technology that communicates with the search engine of each individual repository or by “content extraction” technology that exposes the information in the repository to a single master search and indexing system. Either or both of these methods are being required, depending on the functionality needed by users.

In the specialty search marketplace, there is an increasing interest in domain-specific search technology for electronic discovery and litigation support. In the intellectual property and commercial information markets, interest is on the rise for scalable, stateful, 24x7 search technologies that incorporate all of the new capabilities described above as well as visualization capabilities to help users quickly navigate to the most useful information. In the email management and archiving arena, users are interested in being able to search metadata seamlessly along with the text of messages and even the text inside attachments.

The components of a distributed architecture system are:

◆ Search director. Manages the entire distributed system across all servers. It forwards the requests to the appropriate server and combines search results from the distributed servers.

◆ Shared search engines. Search engines that are user-independent and can be shared and configured in engine pools for optimal performance and utilization of server resources. This eliminates the requirement of a dedicated engine per user.

◆ Search broker. Manages pools of shared engines to most efficiently handle simple and very complex queries, splitting queries among multiple engines as needed for optimal performance. The broker balances incoming requests between engines, selecting idle engines whenever possible to handle new requests.

◆ Load balancing. Manages distribution of workload across servers of varying capacity for optimal utilization and performance. It permits different balancing schemes to be enabled by user-configuration criteria; workload can be directed to specific servers to permit administration on others.

◆ Static splitting. Manages distribution of workload by splitting searches across servers when needed.

◆ Configuration server. Serves as a repository and management tool for centrally managing all configuration and initialization files.

◆ Applications manager. Monitors critical processes, starts and stops processes as needed, identifies failed services and auto-restarts services to ensure a more fail-safe operation.

Discovery Server

Discovery Server is the Open Text re-branded name for the BRS/Search product line. BRS/Search is being used as a general purpose search engine in over 200 installations around the world. In 2004, Version 9.0 was released as a completely redeveloped information discovery tool with stateful capabilities, taxonomy navigation, auto-classification, concept and fact extraction and a Java-based standard UI. It is being actively developed and evolving into a true distributed search product able to handle a large amount of legal-related data. It offers several APIs and can be easily integrated into OEM products. Open Text is working with large organizations to make Discovery Server a fully scalable, fault tolerant, 24x7 system with instantaneous failover.

The Open Text COTS product that addresses some of these needs is Distributed Search. It has been developed over the last four years to address large amounts of patent-related data, where complex inquiries are being applied by more than 4,000 simultaneous users hitting large volumes of data. The architecture developed for the Distributed Commands version of BRS/Search takes advantage of shared storage areas of multiple machines.

The Distributed Search product is a suite of servers working together to provide consistent access to BRS Engines. It uses an administration tier containing a search director component. The director supports splitting searches across multiple machines with duplicate database sets and distributing commands to the machine that is least utilized. End-user requests can be piece-qualified to a single database set or to pieces of all database sets.

The Challenges

The issues and challenges around search today are significant. We are seeing increasing interest in the additional search capabilities that were described above. As the business applications being built begin to rely more and more on robust, full featured, scalable search technologies used in ways that provide competitive advantage, the suppliers of search technologies are going to see increased pressure to keep up and even surpass the state-of-the-art in the marketplace. Distributed search technology is only one approach to these challenging problems.
Beyond Search Results—Get the Whole Picture and Reap the Rewards

By Bob Tennant, CEO, Recommind

Search engines today are built with an infinitely document-centric view of the world. This is hardly surprising, given how the technology has evolved, but it is often not the paradigm best utilized for providing an answer to a user’s search request. When users search, they are searching for lots of different kinds of things.

In many situations, a search is performed not to identify a document, but to identify another kind of object: a product, a person, a project, etc. A superior search system enables users to quickly identify documents, but also to identify these other kinds of entities. This is accomplished by incorporating a view of the world that contains more objects than just documents.

How do search systems that contain representations of different kinds of objects differ from the traditional search systems on the market today?

Consider a commerce website that sells a particular kind of product—in this case, let’s say fasteners, or ways of joining two pieces of material together. Each product (fastener) has a title and a short description. These represent a kind of document that traditional search systems can index, search and, hopefully, find when keywords contained in the query are also contained in the description.

But suppose that you, as the user, are looking for a particular kind of fastener—a nut and bolt—to join two pieces of metal together. You’re looking for that particular kind of fastener not evolved, but because that is what came to mind. In a traditional search system, when you search for “nut and bolt” and variations thereof, nuts and bolts will be returned, but not screws, glues or rivets. If you can’t find an appropriate nut and bolt, you’re at a loss, and the site loses your business.

Now consider what would happen if the site deployed a more advanced search system. The site can avoid losing your business by returning a suitable substitute, in this case perhaps metal screws or epoxy. To do so, the search system needs to be able to have an idea of (a.) what products belong to the fastener class; (b.) what fasteners are useful for joining metal; and (c.) what the prices of the items are. One way to solve this problem would be to add this information to every document, a time-consuming process that results in high maintenance costs as inventory and prices change over time. Another way is for the system to contain an understanding of a fastener object, building a virtual representation of such an object from the various systems (product catalog, joining FAQ, product inventory and pricing) involved. In this way, as pricing and recommended joining methods change over time, the information the search system returns remains current—at a low maintenance price tag.

In the example above, more than just the document is being searched on. Information is being pulled from other systems to represent the “fastener” object and substitutes can be provided whose own descriptions might not contain any of the keywords contained in the first product’s description. A substitute product might be suggested because it is also a fastener, can be used for metal, and is about the same price, for example.

The benefit of the more advanced search system building a virtual fastener object is readily apparent. This is the most basic case, however. In a customer-facing environment, like an online store or a customer service center, many of the relationships described above will be outlined explicitly—a fastener class will be explicitly created, for example. The benefits of reaching beyond simple document search become even greater in many internal search applications, where these object classes have grown up organically and are implicit in the way the organization works.

Implicit objects are all around us in organizations. One common search we’re all familiar with is to identify someone in your organization who can help with a particular issue. Your colleagues have written documents, written e-mails, billed time, worked on particular accounts and done myriad other things that provide a virtual representation of who they are. All of this information can be exploited in the manner described above to identify a person to help with a particular task, by the search system pulling all of these disparate documents from different systems into a virtual person object. As a result, when trying to find who was involved in that Enron gas deal, searching for “Enron gas deal” can help you identify all those involved, in an explicit manner. The same information can then be used to identify other deals that were similar, perhaps one in Europe done by Gazprom. By knowing that the European deal was similar, you can identify people in the London office who can help with a new gas deal you’re working on.

Whether finding the right product, finding the right person, or finding a similar deal is your aim, when considering your information infrastructure, make sure your search system can get the whole picture and you will reap the rewards.

Bob Tennant is the chief executive officer for Recommind. Prior to joining Recommind he served as the managing director of Swiftsure Capital and in senior positions at Sun Microsystems, where he led the marketing of the Java software platform into consumer electronics. Tennant is a board member of several successful young companies and a graduate of the University of Victoria, the University of California at Berkeley and Columbia University, from which he has an MBA.

Recommind’s MindServer product suite delivers search results in full context, enabling users to retrieve information on a product or project basis or to identify colleagues with particular expertise. With faster access to the right information, organizations such as Bertelsmann, Dupont, NIH, and leading law firms Morrison & Foerster and Cravath, Swaine & Moore are using MindServer to save time, increase sales, and increase the value of information assets.
Managing Content for Added Value

By Laurent Simoneau, CEO, Coveo Solutions Inc.

“Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information on it.”—Samuel Johnson

There are many drivers that require companies to manage knowledge more effectively—the incredible rate at which electronic data is reproduced and stored, the increasing role of intranets for collaborative workgroups, and organizations working to meet regulatory compliance. Enterprise search has been thrust to the forefront as a strategic business application by all of these factors and many more. Organizations now understand that enterprise search is more than a pedestrian business application—it provides a strategic business function by supporting knowledge workers and managing intellectual property. When viewed in this light, organizations tend to move away from viewing search applications as a set of features and cast an eye toward the overall value of unlocking stored knowledge and protecting intellectual property.

For example, can the application successfully negotiate access to content stores without professional services, keeping investment costs down while driving access to information up? Will the application apply techniques that classify and organize results and embed the search function behind the scenes in operational applications? Does the application deploy quickly, with security, and make users more productive? Does the application add value to the IT ecosystem by using existing assets that leverage identity, authentication and authorization to provide secure information access? For those organizations that view search as a critical function, the answers to these questions quickly identify the appropriate application required to support knowledge management as a strategic asset.

PRTM: Search is a Critical Business Asset

PRTM (www.prtm.com) is an example of an organization that views search as the most effective way to manage intellectual property. Founded in 1976, PRTM is recognized internationally as a leading management consultancy, and ranked as one of the “50 Largest Consulting Companies in the World” by Consulting News. PRTM is noted for its leadership in product development and supply chain management and is highly valued for a focus on results, developing and executing strategies through improvements in best practices, IT-enablement and core business processes.

In 2004, PRTM launched a knowledge management initiative to better control its intellectual property. Steve Jackson, PRTM’s CIO, recognized that both intellectual property and effective search are key business assets, and both are equally essential for effective knowledge management—“when intellectual property is important, having an effective search engine is absolutely essential,” says Jackson. PRTM standardized on Microsoft SharePoint as an information management platform, and selected Coveo Enterprise Search for SharePoint (CESS) as the added-value complementary search application. PRTM found that other search applications did not meet their key requirements.

PRTM downloaded the search application and had it up and running in a matter of hours. This was a key benefit to PRTM because, according to Jackson, “one of our guiding principles is that if it is difficult to implement it will be difficult to maintain.” During the initial pilot, the search application indexed more than 300,000 documents on file servers and websites in PRTM’s east and west coast offices. Today, PRTM is managing almost 400,000 documents and 140 SharePoint sites. The system is available to 400 users across the globe, and handles approximately 3,500 user queries per month.

Benefits and Lessons Learned

The ability of the search tool to seamlessly work over a WAN and easily index file shares are key requirements for PRTM. Support and minimal maintenance requirements are also important to PRTM. During the trial, the system uncovered misfiled and unsecured documents, enabling PRTM to improve document security and adhere to their corporate compliance policies. Finally, as far as user acceptance, Jackson reports, “No news is good news as far as our users are concerned. Our users let us know when something doesn’t work.”

Laurent Simoneau’s expertise in the search industry stems from more than 10 years of experience in developing and bringing to market innovative and award-winning search products. He comes to Coveo from Copernic, where he was the Chief Operating Officer and responsible for orchestrating the company’s enterprise products division spin-off into Coveo.

Coveo Enterprise Search delivers secure, unified search across all documents and multi-media files located in file systems, databases, enterprise applications, email servers, intranets, and web sites. An award-winning product, and based on industry standard .Net and ASP technologies, Coveo Enterprise Search delivers value with out-of-the-box security for regulatory compliance, unparalleled accuracy based on file monitoring, de-duplication, concept extraction and summarization, and a deployment cycle of less than 24 hours. Coveo’s enterprise clients include AAA, AMN Healthcare, Eli Lilly, Fannie Mae, International Paper, Hewlett Packard, Lockheed Martin, NATO, NASA, AC Nielsen, and Verizon.

For additional information, please visit www.coveo.com or contact us at 800.635.5476.
SEARCH REVOLUTION continues from page 5

relevant snippet. Plus, contextually aware analysis allows on-the-fly text mining, where users can pose queries previously reserved for expert data miners with too much time on their hands. Such queries combine structural data context and schemas with recognized entities, categories and other implicit information, then use sophisticated statistical matching to answer questions like: “Which persons appear in sentences that contain any company names and the word ‘scandal’?”

The semantic index cannot be implemented as an incremental change to existing search platforms. It is not simply a new feature, a new capability or grafted-on handling of a new content type. It is a search revolution.

Budding Personal Pyramids

We have explored how an enterprise can invert its information pyramid to provide its users with the information they need—simply, quickly and cheaply. We have explored how an enterprise can use the contextual power of the semantic index to make the pyramid intelligent. What more can the enterprise do for its employees and customers?

Well, people have personal data. Laptops and PDAs contain a wealth of information in files and emails, chats and blogs. People want access to this personal information in the same way that they have access to corporate information or information from the Web. So they install desktop search tools, from random providers. These tools have known security issues. What they don’t have is finely tuned access to corporate information. What they don’t have is all the advanced linguistic and analysis tools described above. It is like giving people a Ferrari to explore corporate information, but a Matchbox toy car to play with their personal information.

This toy car can do damage. Security leaks are one thing. Traffic leaks, leading users away from the portal and to other search engines, is a strategic loss for the enterprise.

The agile enterprise can provide protection of its content and its users in the face of demands for personal searching. It can provide its user base with its own personal search platform (PSP), based on the same principles as the enterprise search platform, and compatible with it. This has many advantages. First, there is reuse of information between the platforms, with the same capabilities of contextual insight, semantic indexing, and advanced processing and mining tools. Second, there is significant reuse of skills, processes and resources for deployment and maintenance of the two systems. They share functionality in the control systems. Third, the feature set is greatly expanded, compared with freely available desktop search products. It includes navigation, content preview, highlighting, similarity searching, taxonomy usage and native content editing. Finally, where desktop search tools tend to be monolithic, PSP is modular. The main modules of PSP—the indexing engine, the content source APIs, the federation engine, the front-end APIs and the control engine—are all flexible and separately programmable, allowing the fine-grained granularity in control and deployment enterprises need to make this PSP their own—to simultaneously support their users’ needs and their business models.

Our example enterprise uses PSP to connect its employees’ personal content with the company’s intranet content, and to enable offline searching of some of its central resource management tools. Another example, a large national portal, allows its customers to download their own copies of PSP. Since customers can search both the Web and their own content from their personal PSP installations, this portal has effectively locked in its traffic in a user-friendly way, even when people “only” want to search their own files.

PSP provides enterprises with tools that enable safe access to personal information and increase functionality far beyond that of standard desktop platforms. PSP protects company brands by avoiding traffic leaks to other search engines.

Pyramid To Go?

Our example enterprise now has a consumer-focused, intelligent enterprise search that allows its users safe access to personal content. But there is more—mobile intranet, for example. Users like their mobile phones and their PDAs, and the enterprise is considering the third pillar of search in an enterprise setting: the mobile search platform (MSP). MSP allows users to search corporate or personal content from handheld devices—to find a customer address while on the road, or browse digital products from preferred content partners. The enterprise can also use MSP to create its own crawl of the mobile Web, gathering only high-quality documents relevant to its user base. MSP provides the advantages of the other two platforms—modularity, contextual insight, high-speed searching in a user-friendly context—and exploits still more synergies between personal search, mobile search and enterprise search. Shared content, shared features, economies of scale. MSP is secure and tunable. Our example enterprise is now fully searchable, and the true value of its information is seen in its users’ ability to use it, not just their ability to produce it.

Traditional information systems optimize information production and storage, not information consumption, and too often leave would-be information consumers unsupported and uninspired. To stay competitive, the enterprise needs to enable and entice its employees and customers to make the best possible use of what may be its most valuable asset: information. Companies that apply information, instead of just creating and storing it, have a huge and sustainable strategic advantage.

Bjørn Olstad, Ph.D., serves as the chief technology officer at Fast Search & Transfer (FAST). Before joining the company, Dr. Olstad held key positions within General Electric Medical Systems, including director of research and development for Cardiac Ultrasound. He has served as a professor in computer science at the Norwegian University of Science and Technology (NTNU), where he was awarded the youngest professorship ever.

Silvija Seres works in Norway as a vice president of strategic market development with FAST. She holds an MBA from INSEAD in France and has extensive scientific background in algorithm design and optimization, with a Ph.D. and Prize Fellowship from Oxford University. She has held several visiting posts at leading research institutions and has developed international educational programs.
**The 2006 Best Practices Series...**

<table>
<thead>
<tr>
<th>Month</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>Business Process Management</td>
</tr>
<tr>
<td>February</td>
<td>E-mail &amp; IM Management</td>
</tr>
<tr>
<td>March</td>
<td>Business &amp; Competitive Intelligence</td>
</tr>
<tr>
<td>April</td>
<td>Enterprise Search</td>
</tr>
<tr>
<td>May</td>
<td>Enterprise Content Management</td>
</tr>
<tr>
<td>June</td>
<td>Government</td>
</tr>
<tr>
<td>July/August</td>
<td>Imaging &amp; Forms Processing</td>
</tr>
<tr>
<td>September</td>
<td>Financial Services</td>
</tr>
<tr>
<td>October</td>
<td>Records Management &amp; Regulatory Compliance</td>
</tr>
<tr>
<td>November</td>
<td>Knowledge Management</td>
</tr>
<tr>
<td>December</td>
<td>Digital Asset Management</td>
</tr>
</tbody>
</table>
Solving Your Findability Dilemma

- Tuning Search
- Search Logs & Metrics
- Entity Extraction & Clustering
- Federated Search

- Taxonomies, Metadata & Search
- Optimizing Performance
- User-Generated Metadata
- Increasing Relevancy

Full Conference
Gold Pass, May 22–24 (includes 2-day pass plus 2 half-day workshops) ........................................By April 28 $1095
(After April 28 $1195)

AM W1—Taxonomy Fundamentals
AM W2—Information Architecture
AM W3—Integrating Search & Taxonomy

PM W4—Building Taxonomies
PM W5—Enterprise Search 101
PM W6—Content Management & Search

Full 2-day conference, May 23–24 ................................................................. By April 28 $895
(After April 28 $995)

Pre-conference Workshops
Monday, May 22 (price per workshop) ......................................................... By April 28 $195
(After April 28 $215)

AM W1—Taxonomy Fundamentals
AM W2—Information Architecture
AM W3—Integrating Search & Taxonomy

PM W4—Building Taxonomies
PM W5—Enterprise Search 101
PM W6—Content Management & Search

Note: All prices are per person. Faculty and schedules are subject to change; registrants should visit the Web site at www.EnterpriseSearchSummit.com to view specific schedules.
### Conference at a Glance

#### Monday, May 22, 2006
- **9:00 a.m. – 12:00 p.m.**
  - **W1:** Taxonomy Fundamentals
    - Busch & Daniel
  - **W2:** Information Architecture
    - Morville
  - **W3:** Integrating Search & Taxonomy
    - Pohs & Kamm
- **1:30 p.m. – 4:30 p.m.**
  - **W4:** Building Taxonomies
    - Hlava
  - **W5:** Enterprise Search 101
    - Rappoport
  - **W6:** Content Management
    - Earley & Regli

#### Tuesday, May 23, 2006
- **8:00 a.m. – 9:00 a.m.**
  - CONTINENTAL BREAKFAST
    - Sponsored by Coveo
- **9:00 a.m. – 10:00 a.m.**
  - WELCOME & KEYNOTE
    - Ambient Findability
    - Peter Morville, President, Semantic Studios
- **10:00 a.m. – 10:30 a.m.**
  - Ten Things No One Tells You About Enterprise Search
- **10:30 a.m. – 11:00 a.m.**
  - COFFEE BREAK
  - Visit the Enterprise Search Showcase
- **11:00 a.m. – 11:45 a.m.**
  - How Does Everyone Else Do This?
- **11:45 a.m. – 12:30 p.m.**
  - Boosting Findability
- **12:30 p.m. – 1:45 p.m.**
  - ATTENDEE LUNCH
- **1:45 p.m. – 2:15 p.m.**
  - BREAKOUT A
    - A-1: Beyond Search: Intelligent Use of Intelligence
  - BREAKOUT B
    - B-1: User-Centered Results Page Design
  - BREAKOUT C
    - C-1: E-Commerce-Centered Site Search
  - BREAKOUT D
    - D-1: Faceted Navigation of User-Generated Metadata
  - BREAKOUT E
    - E-1: User Experience & Human Behavior
  - ATTENDEE LUNCH
- **2:30 p.m. – 3:00 p.m.**
  - CONTINENTAL BREAKFAST
    - Sponsored by Coveo
- **3:15 p.m. – 3:45 p.m.**
  - COFFEE BREAK
  - Visit the Enterprise Search Showcase

#### Wednesday, May 24, 2006
- **8:00 a.m. – 9:00 a.m.**
  - CONTINENTAL BREAKFAST
  - Enterprise Search in an ECM World
- **9:00 a.m. – 9:45 a.m.**
  - Enterprise Search in a Powerhouse: Integrating Search
- **9:45 a.m. – 10:15 a.m.**
  - COFFEE BREAK
  - Visit the Enterprise Search Showcase
- **10:15 a.m. – 10:45 a.m.**
  - Web Services Solutions for Metatagging Challenges
- **10:45 a.m. – 11:30 a.m.**
  - Contextual Analysis, Entity Extraction, & Innovative Search Solutions
- **11:30 a.m. – 12:15 p.m.**
  - ATTENDEE LUNCH
- **12:15 p.m. – 1:30 p.m.**
  - CONTINENTAL BREAKFAST
  - Tuning Search: Analytics, Search Logs and Best Bets
- **1:30 p.m. – 2:00 p.m.**
  - Tuning Search: Best Practices
    - Tips for Using Search Logs & Reports
    - Creating Best Bets & Synonyms
    - Tuning Search to the Structure of the Content
    - Using Metrics & Log Analysis to Identify Customer Needs
- **2:00 p.m. – 3:15 p.m.**
  - CONTINENTAL BREAKFAST
    - Sponsored by Coveo
  - Steve’s Scoop: Who’s Hot in Enterprise Search & What’s Coming

---

#### Easy Ways to Register

   When registering on the Web site, please refer to the priority code above your name on your mailing address label.

2. Phone:
   - (800) 300-9868 or (609) 654-6266

   Confirmation: All registrants will receive confirmation of their registration by mail. Registration packets may be picked up at the Registration Desk beginning at 8:00 a.m. on the Pre-conference Day.
   Cancellations received prior to April 28 will receive a full refund minus a $50 processing fee for that event. There will be no refunds for cancellations made after April 28.
For more information on the companies who contributed to this white paper, visit their websites or contact them directly:

BearingPoint, Inc.
1676 International Dr.
McLean VA 22102
Ph: 866.BRNGPNT or 703.747.3000
FAX: 703.747.8602
Contact: sales@bearingpoint.com
Web: www.bearingpoint.com

InQuira Inc.
851 Traeger Avenue, Suite 125
San Bruno CA 94066
Ph: 650.246.5000
FAX: 650.246.5036
Contact: mrlng@inquira.com
Web: www.inquira.com

Recommind, Inc.
170 Columbus Ave., Suite 310
San Francisco CA 94133
Ph: 415.394.7899
FAX: 415.397.3811
Contact: info@recommind.com
Web: www.recommind.com

Endeca Technologies, Inc.
55 Cambridge Parkway
Cambridge MA 02142
Ph: 617.577.7999
FAX: 617.577.7766
Contact: sales@endeca.com
Web: www.endeca.com

ISYS Search Software, Inc.
8775 E. Orchard Road, #811
Englewood CO 80111
Ph: 800.992.4797
FAX: 303.689.9997
Contact: info@isys-search.com
Web: www.isys-search.com

SchemaLogic Inc.
620 Kirkland Way, Suite 100
Kirkland WA 98033
Ph: 425.885.9695
FAX: 425.883.0117
Contact: sales@schemalogic.com
Web: www.schemalogic.com

Knova Software
10201 Torre Avenue, Suite 350
Cupertino CA 95014
Ph: 800.572.5748
Contact: info@knova.com
Web: www.knova.com

WebSideStory
1111 Bayhill Drive, Suite 285
San Bruno CA 94066
Ph: 650.244.1400
FAX: 650.244.1401
Contact: sales@atomz.com
Web: www.websidestory.com

MetaCarta, Inc.
350 Massachusetts Avenue
Cambridge MA 02139
Ph: 617.661.6382
FAX: 617.661.6384
Contact: info@metacarta.com
Web: www.metacarta.com

ZyLAB North America LLC
1577 Spring Hill Road, Suite 420
Vienna VA 22182
Ph: 866.995.2262
FAX: 703.991.2508
Contact: info@zylab.com
Web: www.zylab.com

Produced by:

Kathryn Rogals
207-338-9870
kathy_rogals@kmworld.com

Paul Rosenlund
207-338-9870
paul_rosenlund@kmworld.com

Andy Moore
207-236-0331
andy_moore@verizon.net

For information on participating in the next white paper in the “Best Practices” series. contact:
paul_rosenlund@kmworld.com or kathy_rogals@kmworld.com • 207.338.9870