

Research Report

Enterprise Search Markets and Applications

Capitalizing on Emerging Demand

June 2008

Lynda Moulton
Lead Analyst for Enterprise Search



THE GILBANE GROUP

Gilbane Group gratefully acknowledges the support of the sponsors of the research informing this report. This work would not have been possible without them. Please see the **Vendor Profiles** section of the report for detailed descriptions of these search suppliers and their offers.



Gilbane Group Inc.

763 Massachusetts Avenue
Cambridge, MA 02139 USA

Tel: 617.497.9443

Fax: 617.497.5256

info@gilbane.com

<http://gilbane.com>

Table of Contents

List of Figures	ii
List of Tables.....	iii
Preface	iv
Introduction	5
Market Landscape.....	6
Search Targets.....	6
Search Engine Models.....	9
Search Interface Options	10
Search Enhancement Tools	12
Market Demands/Market Realities	17
Verticals.....	18
Horizontals (Functional Groups).....	22
Planning for Enterprise Search: User Guidance	26
Enterprise Search Research.....	27
Field Experience with Search.....	30
Getting into the Customer's DNA: Vendor Guidance	41
Summary	46
Vendor Profile	49
Coveo	49
ISYS	53
Vivisimo	59
Appendix A: Enterprise Search Interview Lines of Questioning.....	64
Appendix B: Vendor Directory.....	68
Appendix C: Glossary	73
Appendix D: Bibliography	83

List of Figures

Figure 1: Four Major Search Product Attributes	6
Figure 2: Search Engine Target Options.....	7
Figure 3: Major Models for Licensing and Installing Search Software.....	9
Figure 4: Interface Styles for Searching.....	11
Figure 5: Search Related Products and Add-ons for Enhancing Content.....	13
Figure 6: Search Concepts with Detail on Content Stages	27
Figure 7: Survey Respondents' Role Distribution	28

List of Tables

Table 1: Companies with embedded search product	8
Table 2: Companies offering site search products.....	8
Table 3: Companies supporting principally Internet searching.....	8
Table 4. Survey Respondents' Product Experience (names of their choice)	30

Preface

This review of the search market for enterprises is framed by three points of view. Officially, I am an analyst for *enterprise search* for The Gilbane Group. Concurrently, I consult to companies working to align technology acquisition and deployment with focus on business propositions. Previously, I was a business owner responsible for developing and marketing a high-end enterprise content management system complete with end-user search and navigation interfaces (aka corporate integrated library system). This has been my work for nearly thirty years. It informs the kinds of research and information gathering I undertook to give the reader a slightly different angle. This is an analysis of a fledgling market and the adopters who struggle to make sense of all their options before and during deployment.

One goal of this report is to provide a summary of the marketplace that I hope will help buyers map their needs to search products that will deliver the most value in the shortest amount of time, and at a reasonable cost. Additionally, buyers will find practical guidance about product evaluation, selection, implementation, deployment and maintenance. A vendor directory and glossary round out the content directed to new buyers. Search users who are expanding or seeking improvements to their current experience will find advice on making incremental adjustments in deployment.

Vendors who are developing, selling, marketing, or supporting customers with enterprise search applications will find aids to competitive intelligence in the market landscape, directory and glossary. I hope these will give a different perspective on ways to “slice and dice” the playing field, and lead to a better understanding of clients and their needs. I am rounding out this study with some common sense commentary about what vendors need to be smarter about to build and sustain their own market presence. They are only too aware of how much competition is out there; hewing to market expectations with constant course corrections is a key to a winning strategy. Customers may not speak as directly as I will in this report.

This research would not have been possible without the foresight of Frank Gilbane in launching an official search practice area within The Gilbane Group in 2007, a year in which the number of vendors positioned in search eclipsed most other software industries, making it a strong growth area. I also owe thanks to Mary Laplante, Gilbane’s VP of Consulting, for believing that this was an area in which we have to make our presence known, especially in congruence with other content related application markets. Finally, to the un-named search technology pioneers, both buyers and sellers, who shared their stories with me, I hope your comments have found a worthy context in this report, and will bring awareness that result in product and market *wins* we all seek and expect as the search market matures. Thank you to all the anonymous contributors.

Introduction

Every year more and more jargon clutters our understanding of information technologies. Marketers' wordsmiths struggle to define new niches that differentiate them from companies in similar businesses. Popular culture elevates coined phrases to levels of buzz that become repeated *ad nauseam* with no clear definition. Is there a concrete definition for *enterprise search*? The answer is "no," but it is still laid out as its own marketplace, complete with directories, conference, seminars and a whole slew of vendors that claim to offer it. What this author set out to do is analyze and write about what Steve Arnold calls "beyond-the-firewall search;" we won't argue with that.

Just keep in mind that in many organizations it is possible to deploy practically the same technology for search outside the firewall as search for content within. The packaging and deployment will vary and search results will be vastly different (better or worse), but the core indexing and retrieval technologies may differ little. In this market review, the focus is on the buyer's perception, needs, and experience with search. The buyers we are considering are those seeking search applications for some defined domain of content that exists for indexing within the enterprise and for retrieval by the internal enterprise population, however the organization defines that. By this definition, searchers may be clients of a professional services firm, customers of a computer company, or students within a university. For some other thoughts on what constitutes *enterprise search* see [this commentary](#) by our President, Frank Gilbane.

The next two sections lay out the author's experiential view of how the buying and selling landscapes are self-organizing in early 2008. This year is the fifth anniversary of Information Today's Enterprise Search Summit in New York. The first few meetings were single track affairs held in one room with highly instructive presentations. These were designed to let future potential buyers know what products were possible, viable and appropriate, and what was needed to implement and deploy search.

Although search products have been available for enterprise procurement for over 30 years, the number of organizations (mostly government agencies and very large companies with significant R&D operations) that had actual experience with them was miniscule. In many ways, it still is. Major corporations and government agencies do not have installed and operational a single search application that covers all possible electronic content in the enterprise for retrieval in a single search interface. Some large professional services firms come close to that model, as do some small-medium businesses (SMB). Mostly, there are thousands of "instances" of search applications being deployed throughout enterprises of every type. Usage, at present, is in the phase that this analyst would call largely experimental or "hit or miss." As has been written elsewhere, there is a lot of pain and *seat-of-the pants* learning going on. Experts for support and experts for deployment and maintenance are few. These observations will bring the author to some strong guidance for vendors and advice to buyers to help each gain more traction as sellers and users, respectively.

Market Landscape

One way of categorizing search engines is to prioritize needs or search headaches. The following figure illustrates four different ways to approach investigating a search solution that will satisfy requirements in the enterprise. All categories deserve consideration, but buyers need to prioritize which will trump the others. This will undoubtedly hinge on who the principal (final) deciding entity is. End users will care most about the suitability of the interface to their style of searching and whether all the content they consider most valuable and relevant is indexed by the search engine. Product models and the architecture, as it suits current IT practices and infrastructure, will be a focus of the technology group. Business managers will usually focus on functions and features that support content analysis, reporting and visualization options or business intelligence. Asking each member of a selection team to prioritize their requirements in terms of must-haves and preferences will result in a first cut of products to seriously consider. Vendors need to know where their buyers place top priorities so they can help prospects to decide if what they have to offer is an appropriate fit.

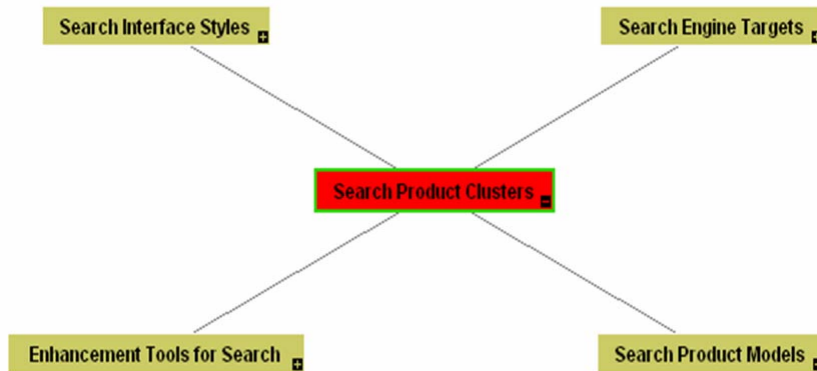


Figure 1: Four Major Search Product Attributes

Search Targets

At the top level, let's consider definitions and possibilities for each of these four categories of search targets.

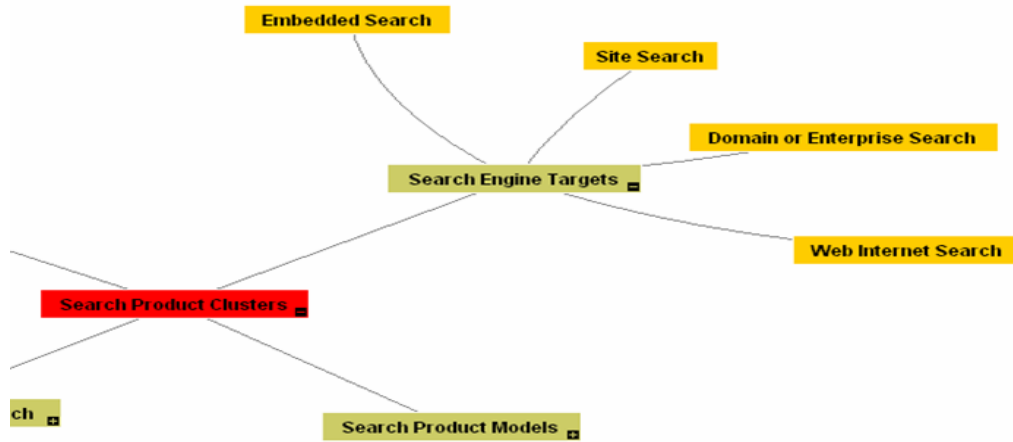


Figure 2: Search Engine Target Options

These categories should inform the first cut that buyers will make to effectively reduce the number of products to consider for selection. Keeping in mind that very few vendors have strength or offerings across all categories, focus is critical. It is important to remember that no enterprise will find a solution for all search challenges. The list below (which is not comprehensive) includes vendors that supply search for embedding or have their own search embedded in a suite of products for content management.

1. Embedded Search – Refers to search engines that are delivered as a component of another software application. There are several embedded search scenarios:
 - a. The search product is developed and maintained by the application vendor as part of the core product and not for use with any other application (e.g. search within MS Outlook or Adobe).
 - b. The vendor has its own search technology and embeds it as a service and continues to maintain and control distribution exclusively (e.g. Convera’s Excalibur for vertical search, Cuadra STAR for the STAR CMS).
 - c. The vendor OEMs another company’s search engine, which also continues to be available for other commercial use (e.g. Ontrack’s Engenium is available for OEMing).
 - d. The vendor adopts open source search, perhaps adding to it, depending on the larger user community’s enhancements and upgrades (e.g. Lucene with Siderean Seamark).

Access Innovations	Convera	Inmagic	IXIASOFT	nStein
Adobe	Cuadra Assoc.	Inquire	Liberty IMS	Progress Software
Attivio	Dieselpoint	Intranet	Lucene	SAIC
Clarabridge	dtSearch	IntelliSearch	Microsoft	Temis

Collanos	Index Engines	ISYS	Ontrack	Zylab
----------	---------------	------	---------	-------

Table 1: Companies with embedded search products

2. Site Search – Refers to products optimized for a single or bounded-use Web site, which may be internal, public facing or an extranet for a special purpose (e.g. customer support). There are many such products at reasonable costs suitable for non-profits, small businesses or sites limited in content, scope and usage; others scale to very large and sophisticated e-commerce domains for complex product offerings. PicoSearch at the low-end, Exalead for mid-range commercial use, and Endeca on the high-end for complex selling models would be among the products to consider plus any others on the following list. (note that Web CMS systems also offer embedded search for Web sites they manage.)

Baynote	Collarity	Mercado	SLI Systems
Blossom Soft	Endeca	PicoSearch	SurfRay
Convera	Exalead	Sinequa	

Table 2: Companies offering site search products

3. Domain or Enterprise Search – Refers to the use of a search product where the principal target is a collection of internal sites, collections of documents (file shares), applications containing data and/or documents, and may include a federating option that also searches external sites and presents the combined internal and external results organized in a prescribed format. Some new products in this category include special functions, such as focus on content conversion or management to an XML format, “e-Discovery,” email content, semantic search. See Appendix B, Vendor Directory, in which all the companies with products that are candidates for search across one or more domains of an enterprise are displayed in bold. Many of them also have products that target other content search models. Domain or enterprise search is the principle focus of this report.
4. Web Internet Search – Refers to engines that target content in across-the-Internet searches (e.g. Google and Yahoo), meta searches (e.g. searches results of sites searched by other search engines such as Clusty.com from Vivisimo) or searches across a collection of specific Internet domains (e.g. Wikia). This study will not comment on these Internet search engines, except regarding features that might be of interest within the enterprise. A very short list of Web Internet Search companies besides those mentioned is shown for its diversity.

Altavista	Eyealike	NorthernLight	Techrigy
Exalead	Grokker	Progress Software	VoiceTech Group

Table 3: Companies supporting principally Internet searching

Search Engine Models

Search engine models are grouped according to their generic installation types. Often business buyers are not aware of these possibilities when they begin investigating product types because they don't consider technical IT architecture implications. Buyers do need to have a discussion with their IT departments or those who will support the computing infrastructure and accessibility options to search.

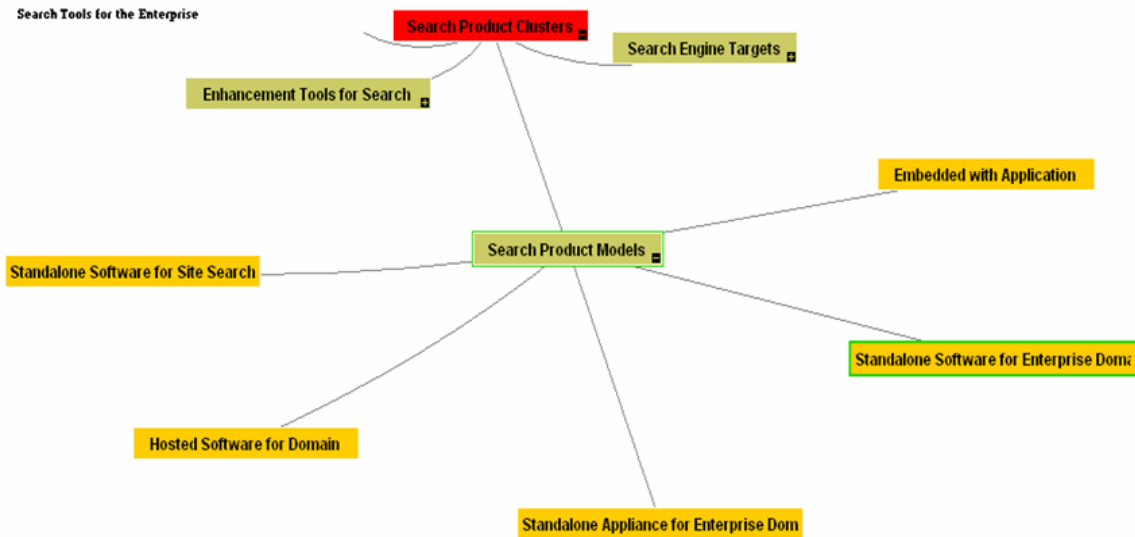


Figure 3: Major Models for Licensing and Installing Search Software

When an enterprise has a solid content management architecture, with strong governance of content flow managed by expert content specialists, search options can be narrowed quickly to match the architecture and applications being used for storing content. Search may be embedded in the document or content management system as with IBM's Notes or OpenText for large enterprise content environments, nStein, Inquire or Access Innovations for integrated special-purpose content management. In these cases, licensing the application includes a search function.

When standalone search is needed only for a single-bounded Web site, wiki or collaboration application, there are many low-cost options that snap into place rather easily, with minimal, but some, on-going support (e.g. Ontolica from SurfRay for Sharepoint). Licensing is usually confined in these situations to a single server, or defined repository of content, sometimes with limits on the number of documents to be indexed for a license category.

A third simple model is using a hosted search solution. This works well for non-profit enterprises wanting to easily expose content to its membership or the public without incurring capital costs. On a monthly fee basis, the vendor's own computer is managing the search software, content indexing and search activity. This is also a way for those supporting search in the enterprise to become acquainted with search tools in general without the overhead of managing the installation. A hosted solution is a good approach for any organization trying to become more familiar with how search engines function and what they have to offer in the way of administrative tools. Such solutions are also

appropriate for any organization that needs to ramp up quickly and has no capital budget in place. Confining what will be crawled and indexed to a sub-set of all possible content will prove a valuable learning experience. Dieselpoint, Funnelback and PicoSearch are just three that offer this option. Ask potential vendors about it; more companies are offering it all the time.

Finally, the model that receives the most attention is cross-domain enterprise search. It is the principal focus, for this study, of Gilbane discussions with several dozen enterprise search implementers. Their experiences reflect a number of surprising things about adoption, not the least of which is the limited scope of content that is actually being indexed by the applications described. While technological barriers are no longer a major issue to indexing millions of documents in dozens of formats (both structured and unstructured), it was difficult to find examples of this playing out in the field. Considering the companies offering products that index content across enterprises, regardless of the nature of content, we expected to find more case studies describing full deployment of this type. The reasons for why it is not happening more are pretty straightforward, and easily summarized. *The author believes that slow adoption is due to the learning curve needed to make search successful, lack of experts to do so, disappointments with legacy search that has not been well maintained, and the failure of enterprises to put a priority on quality content and governance.* More will be shared about this throughout this report, but the bottom line is that the barriers are neither search technology nor availability of options; there are plenty.

Search Interface Options

Usability will be the ultimate determinant of the success of any search product deployment. This extends not only to how easily a searcher will interact with the software through one or more interface options, but also to the value, accuracy, trustworthiness, and contextual relevance of search results. The searcher's first impression is how easily he can choose or dictate what he is searching for. Training, tutorials or "help" options are almost universally eschewed by users in preference to intuitiveness. Unfortunately, the design of interfaces that are sophisticated, intuitive and simple is an art that few implementation "engineers" have mastered. Experts like the designers at [User Interface Engineering](#) share copious insights into their development, testing and deployment experiences. The author wonders, however, if many development engineers actually seek this type of guidance, which has been built over decades.

What is clear is that every organization has a cultural makeup to balance with its real business drivers for search. Successful design interface must reflect that cultural knowledge architecture. The ubiquitous search box and desire for the simplicity of the "Google experience" undermines efforts to give searchers without professional search training the full text experience they *want* with the search options they *need* to get professional results. They believe that it should be as easy to find the "test results on tensile strength of their proprietary nylon rope" as it is to find a "denim shirt with pearl encased snap closures in size medium" on the Internet. What they may not realize and understand is that this need for search to make their work more efficient is just as technologically possible within the enterprise, but that management is not convinced

that it will be cost effective to expend the human resources needed to design and build it. We are at a business standoff on this point.

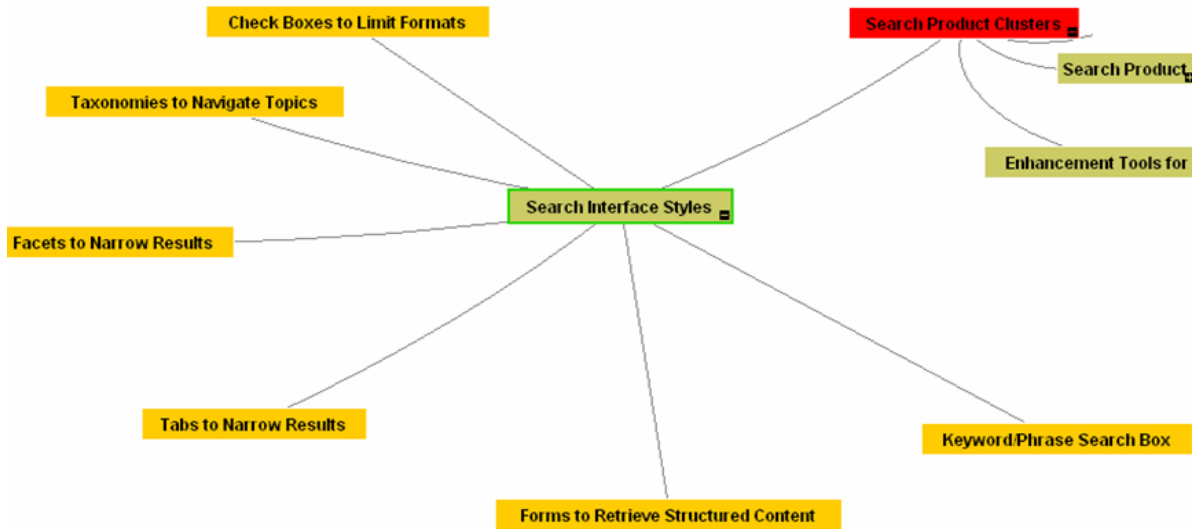


Figure 4: Interface Styles for Searching

In delineating the high-level interface options a buyer has to choose from, it is important to note that there is a human cost to design, develop, test, optimize, and maintain them. Not much will be gained from technology if it is treated as static or permanent. Consider the verb we use to express an item checked off on our “to-do list”: *done*. If a search installation is completed and treated as *done*, it will truly be finished delivering the value and efficiencies expected soon after deployment.

The following interface styles are not mutually exclusive. A good search portal design will offer options for different types of users, and with tailored layout and options for a specific audience. For each style, some considerations are:

Keywords/Phrase Search Box – Most searchers still don’t know when they need to use quotes to get an exact match on a phrase. Most won’t use them when they are needed. Then they do not understand why they retrieve content with the first word from the phrase in one paragraph and the third word four paragraphs later. Some search engines give preference to exact phrases, but others may push content to the top based on unseen metadata. Quotations may be helpful to explicitly bind a phrase or may not be needed. Decisions in tuning can control how and what the search actually looks for; decisions about design on the interface can clarify some huge misunderstandings. How about putting a little tip for help beside the search box like (e.g. type “cancerous cells” with quotes to find an exact match)?

Forms to Retrieve Structured Content – Many enterprise domain search products on the market can crawl and index database content, metadata associated with documents, as well as unstructured documents. Leveraging structure to retrieve more precise or relevant results depends on parsing the search options into a format that is really intuitive and optimized for a particular audience. While dozens of metadata or fields may be searchable, it is probable that only a few are going to be used regularly and

meaningfully by the target enterprise audience. For search form options, considerations regarding design and which data will be searchable require exploration, knowledge of the audience and experimentation to get it right. Ongoing feedback is critical for success.

Tabs (or buttons) to Narrow Results – Some search applications use a tabbed folder metaphor when there are a limited number of narrowing options (e.g. document types, broad categories, vertical search). These tend to fade into the background on a busy screen but can be helpful for simple applications. Google has used this visual option in the past but has changed from the tabbed approach to buttons and a “more” option with a drop down as the number of vertical search options increases. Targeted testing will determine if these search differentiators will actually be seen and used by the audience.

Facets to Narrow Results – A faceted approach is useful in situations where the selection of a small number of parameters will get the searcher quickly to a precise result in most cases. Product search is often parameterized for this reason (e.g. size, color, application, cost, and so on). Again, this requires expertise and a deep knowledge of both the content domain and target audience to get it right. A manufacturing company with a large number of components, products, or devices that need to be looked up on a regular basis will benefit from this style of interface and search tools that easily support it. Another business use is professional services where parameters might be client, case number, date ranges, and document type. In each case, metadata for categorizing content into the correct parameters is essential.

Taxonomies to Navigate Topics – Taxonomies are similar to facets but more extensible, deeper and richer in complexity. Enterprises with a highly diverse corpus of content with a great mixture of topics and target audiences (e.g. legal, financial, engineering, manufacturing) will find that metatagging with a controlled taxonomy provides better search experiences for a broad audience. Taxonomy also requires ongoing professional support to build and maintain, and content governance is essential. Metatagging is a discipline and requires a level of expertise in the nature of the business, information produced by the business, use of content and the users.

Check Boxes (toggles and other visual devices) to Limit Formats – One of the most common themes among users wanting to narrow search results is by type of content. Before they begin a search, they claim to know that what they want is in a PowerPoint presentation, an e-mail, or book. Giving users that option to restrict search up front often requires up-front programming for the search interface. Any search product that makes it easy to pre-define the type of material retrieved before executing a search has an advantage.

Search Enhancement Tools

An increasing number of products secondary to search engines have come to market in the past few years. Many more have existed for longer to support content management systems, library systems, business intelligence and automated report generation from database applications. Some of the latter have morphed into products that are offered by search company partners because of their value in improving the condition of

content to be crawled or the way search results are presented for further manipulation. The major categories of search enhancement tools are illustrated in Figure 5.

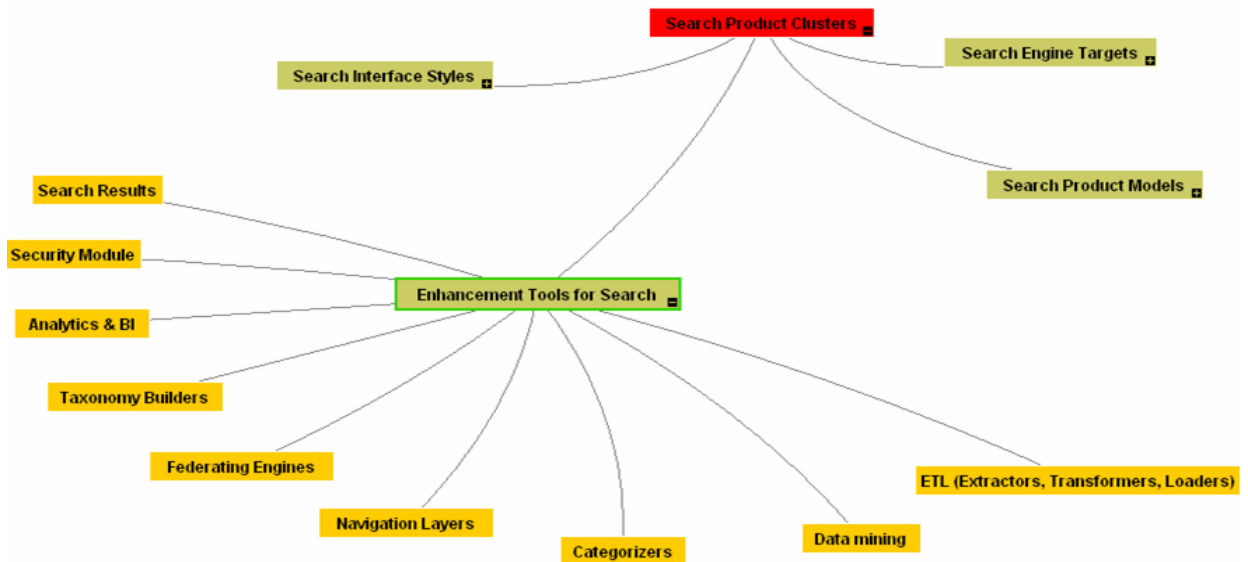


Figure 5: Search Related Products and Add-ons for Enhancing Content

Definitions of these types of software tools are contained in **Appendix C**, the **Glossary**. Still, some comments plus examples about when they are most appropriate may clarify why, when thinking about search procurement, buyers might also want or need to consider one of these applications.

Extractors, Transformers, Loaders (ETLs) – When an organization has a significant amount of content with data elements that could be used to create metadata for unstructured documents, ETLs are a perfect example of automation that saves significant labor. The processing might data mine (extract) the properties field for “author” names, usernames, dates last modified, titles, and so on. It is unlikely that the format of the data across multiple and large file servers will be “clean” enough to constitute good metadata. At its simplest, a transformer might apply pre-defined rules that will automate normalizing the data to a uniform standard. A loader module will be used to place the transformed content into an appropriate metadata form associated with the full-text document. Some commercial products focus on transforming data into XML or other standard formats while locally written scripts (e.g. PERL) can do the work on small numbers of documents fairly easily for loading into proprietary applications. Metadata that is cleaned and normalized will support significantly better metadata search or search navigation. Previously, the emphasis has been on ETLs for data warehousing applications but the author believes that enterprises seeking to improve search will also find them to be useful. (Companies that provide some of these solutions include: ISYS, IXIASOFT, MarkLogic, Olive Software, QL2, and Seaglex).

Data mining –To discover and evaluate before deciding what should be exposed to a search engine, a standalone data mining product can automate the knowledge asset discovery process. This can be useful to help an organization clearly define the nature, scope and amount of content that will be the target of a planned search application.

Because many search products are licensed based on number of documents, number of servers and/or number of “seats” for searchers, it is critical that companies have good data about these elements before finalizing the selection of any one product. It may mean the difference of tens or hundreds of thousands of dollars in initial licensing or costly upgrades at unscheduled intervals. Data mining over time will also give a sense of rates of growth and relative distribution of collections. Some data mining tools lend important assistance to the process of how content can or should be categorized, a step in the process of building meaningful taxonomies. Finally, a data mining process may be embedded with ETLs or provide valuable information about where ETLs can be applied. (Some of the companies in our directory that offer data mining functionality are: Attensity, Autonomy, Basis Technology, Business Objects, ISYS, and SAS.)

Categorizers – To expose the extent of, type and validity of existing metadata, standalone categorizers can lend support to data mining or data extraction processes. Text analytics products may perform this function or products complementary to search engines may be used. Categorization output is a significant indicator of how much and the nature of content that will be targeted by search. Used in conjunction with data mining, categorization output can form the basis for enhancing an existing taxonomy or building one from scratch. Many of the search engines that buyers consider will have built-in auto-categorizers. They will be used once the product is installed. (For an added layer of analysis to be applied to search, companies offering search enhancement tools with classifiers or categorizers are: Abreivity, Kroll Ontrack, Lexalytics and NorthernLight.)

Navigation Layers – To automate the process of search navigation into a multi-dimensional taxonomic structure and to create a simpler method of exploring content by facets, look for options for narrowing search. By exploiting a standard like RDF to find classes of metadata and organizing each class hierarchically, the navigation layer exposes the user to options for narrowing a search by any one of a number of facets at each point in the search. Navigation layers may be specially programmed or be an optional module for a particular search engine, or may be procured to work in a complementary way with a third-party search engine. (Candidate companies are: Dieselpoint, Endeca, Siderean, and Wand.)

Federating Engines – To merge the results of search performed with a variety of search engines, federating engines manipulate the content to present a normalized view of all the content, de-duplicated, and organized for optimal application within the enterprise. Federation sophistication ranges from simple mergers of results from a number of search engines to that which significantly enhances content before presenting the results to the searcher. Think of federation as post-processing of previously non-aggregated and non-associated content. The principal role for this software is when enterprises want to combine searching external and internal content in a single step where more than one set of indexes and indexing engines is involved in the process. (Some companies offering search federating include: EMC, Grokker, MuseGlobal, and SearchBlox.)

Taxonomy Builders – To provide content managers with controlled and standardized terminology lists, enterprises need tools to simplify the process of building and

maintaining lists of hundreds or thousands of terms. Some content software applications are tightly linked with an embedded taxonomy building and maintenance function. The advantage is that when terms change (e.g. spelling, term consolidation) a global change operation on content metadata is executed across managed content. Standalone taxonomy builders require application interfaces to detect additions, modifications and deletions, and then apply appropriate updates to target content. Since taxonomies should never be “finished” or static, the latter can present problems with synchronizing a large and rapidly changing topical landscape. For some applications like smaller web sites, these standalone products may be suitable. (Companies with standalone taxonomy builders or embedded tools are: Access Innovations, Cuadra Associates, MultiTees, Schemalogic, Wand, and WordMap.)

Analytics & BI – To truly leverage the total value in disparate enterprise content, software is needed not only to retrieve but also to analyze, classify, codify and re-constitute large amounts into new content that “tells a story.” With dozens of text analytics and BI software products on the market, it would seem that organizations are solidly embracing these tools. Sadly, the technological potential is vastly underutilized because it is little understood. Experts who understand the potential are not widely employed in IT departments, and only large enterprises seek out the consulting expertise needed to select and deploy the best options for any given vertical business. However, any organization not already using text analytics or BI tools should be aware, when seeking a search solution, that the value of content can be vastly extended by additional software that will help them sift through and usefully interpret huge amounts of data that could never be fully canvassed by human analysts. Many enterprise search products are adding these capabilities to their offerings, and there is trend toward convergence of BI or Text Analytics with what search companies offer. (Check out these companies to get a sense of the range of options: Basis Technology, Clarabridge, Endeca, Fast Search & Transfer, Information Builders, Lexalytics, SAS, and Temis.)

Security Modules – To ensure that only those who need select content see it, security software should be at the top of every enterprise search implementation. Ideally, all the tools that are needed to secure access will be embedded in whatever search engine is acquired and deployed to work with the local IT architecture. Before the final decision is made in the selection process, a clear picture of the IT infrastructure, architecture and components necessary for optimal security must be mapped and shared with installation, implementation and deployment personnel. This document should also be shared with and signed off by the selected search vendor. Security is something that must be understood by everyone. There should be no surprises.

If a current search solution does not provide the requisite security, the short-term solution is to ensure that no content be crawled and indexed with this product that cannot be viewed by the entire enterprise. Secure content needs to be confined through access control to search by other means. On the subject of security, it is worth noting that the author’s research revealed several instances where search software, once implemented to crawl all known enterprise domains and deployed, actually ended up revealing significant sensitive content that should have been restricted. It then had to be removed from the indexes promptly to safeguard confidentiality and other business

critical information. This must be tested for and corrected before opening search to the entire enterprise.

Search Sharing and Results Collaboration – Not to be confused with collaborative content management systems, this is a new area of improvement in the search environment, aka *social search*. The idea is that search, when it is a mission critical component to bring together relevant content for a project, can result in new content. It may take a searcher or analyst time to develop good search strategies to collect all relevant pieces into one results set, a set that could be used by others on a team. In addition to search products that support the option to share results with others and add tagging or social commentary through added functionality, there is a burgeoning population of so-called 2.0 products that support collaboration, including leveraging search results. Again, the technology may be way ahead of the imaginations of many in the enterprise, but you need to be prepared for demand for these tools. (Among the companies that are providing innovation in this area are: Attivio, Connotate, Endeca, and Vivisimo.)

Other Search “Hot-buttons” – While these have not been broken out as a separate class of product, there are other options arriving with the products already mentioned. A large one is *semantic search*, which is defined in numerous ways that overlap with text analytics, natural language processes, cognitive understanding, artificial intelligence and neural networks, among others. (Some of the products in our directory that leverage semantic-based computing to drive higher relevancy in search results are Brainware, ConceptSearching, Connotate, Knova, Kroll Ontrack, Nervana, Powerset, and Semantra.)

Finally, there are products that have specialties that may be of interest to enterprises: *meta search engines* or search engines that search results of other search engines (Clusty from Vivisimo, Grokker, and MindTouch are a few). Tools from integrators have other specialties. SAIC and Schemalogic handle *metadata management*, Vorsite and numerous others can *enhance Sharepoint search*, and LTU Technologies supports *image search*. There are these options and countless others for unique applications like mobile search, email search, and voice search. More keep coming.

Market Demands/Market Realities

The author believes that search market offerings are in a dichotomous relationship with search market buyers. This unfolds along a number of dimensions. The first is the issue of price, which for search tools is huge because the product list is so large and diverse. There is something for every buyer's price point. Price is the number one determinant of how the majority of potential buyers narrow the products they will consider with very few exceptions. The reasons are complicated, but among the problems for vendors are that a buyer's budget is rarely stated up-front. Partly this is because buyers want to learn as much as they can and want vendors to pay attention to them. They may know that a product is unlikely to be funded but there is always a next time. They want to know all they can about products to be able to offer up alternatives if what they do decide to procure fails at any point in the process of selection, procurement, implementation or deployment. This may seem like a cynical condition, but it is a reality. When pinning down search buyers on their selection process and asking the top reasons for their decisions, price is almost never mentioned up front. However, when the interviewer asks afterwards whether price was a factor, the answer is usually along the lines of "yes, probably the most important one."

The second dimension of the market is that the technologies being offered in products far outstrip the imaginations of those making selection and procurement decisions. With vendors defining features and functions that are little understood, or the implications of advanced options not easily grasped, buyers are overwhelmed with information that only confuses the differentiation process. In fact, people interviewed for this study complained about being overwhelmed with information, much of it not useful or appropriate to their investigation. There is so much more in many products than what most buyers can absorb and digest, especially if they are seeking a "first" search product.

So, the dichotomy is this: there is a vast oversupply of technology with limited buying power in the market spread among a lot of vendors.

It isn't the number of systems that will be procured (total market capacity), but the lack of purchasing power across the market for high ticket systems.

However, there are a couple of mitigating factors. One is that many vendors do not depend entirely on search product revenue and are only gradually ramping up in this marketplace. Thus their visibility in search is low; for them, failure to sell search is not a potential "show-stopper" over the long term. Also, most software companies that are exclusively in the search market are focused on delivering value and building brand recognition. They are also running lean and reinvesting earnings in technology. Finally, they are packaging search for specific vertical markets, which will give them strength over highly diversified companies.

All of this is to the benefit of buyers who will grow in sophistication as they work through the process of implementing and deploying several products, moving from

inexpensive and simple to more complex offerings. As the market matures, so will the early product leaders. The ones who are going to survive will have more advanced offerings that keep up with client demands. As well, they will be building solid service and support operations. The latter will improve the revenue stream from a slug of capital licensing revenue late each year, to more even income recognition on maintenance revenue, month-by-month.

Finally, with more experienced users in the market for any software package, a supply of experts will be available to provide consulting implementation support, either as third-party consultants or working for the vendor. Widely deployed products gain immeasurable strength when enterprises know that they can employ expertise to help with implementing and maintaining a product. Buying decisions made by upwardly mobile professionals between product X and product Y professionals may have as much to do with how experience with a product will enhance their own expertise as anything else. They want marketable skills. In the author's opinion, this is one reason that technically excellent products may sometimes fail in the market, something buyers and sellers need to think about.

Verticals

Great emphasis is placed on product positioning in vertical markets. Hundreds of product directories are organized around vertical markets, conditioning buyers to look first at products used in organizations that share their vertical. This is not necessarily relevant, especially in the area of search products, but it is the nature of the business. It sometimes defies logic that buyers would seek the same tools as a competitor when it is entirely possible that another product with a unique strength might enable their enterprise to differentiate itself by innovative use of search to its competitive advantage.

- Enabling employees to search more efficiently or to share search results easily on major projects would both be a cost savings and speed products to market more rapidly.
- Perhaps plugging a hole in managing email compliance or needing to enhance search output by providing financial and business analysis will give the enterprise a more solid operational footing than a competitor.

Understanding the nuances of search and its potential is a serious professional undertaking. The selection should not be left to minors who can do no better than look for products established in a particular vertical.

There are product differentiators that play better in select industries. Vendors positioning their offerings for their strengths should consider where their product marketing efforts have the best opportunities. Some fundamentally unique needs in various industries are noted in the following. This is not intended as an exhaustive list but a reminder that some differentiators are significant when it comes to value in search tools.

Professional Services – Consulting, law, and accounting firms generate huge amounts of intellectual output, much of it in similar formats and topics. These enterprises depend on search that leverages particular metadata, case numbers, client names, dates

and date ranges, and type of content. Because of the large volume, performance and scalability are particular issues. Products that specialize in searching text and returning it in rich contextual formats (clustering) to help differentiate similar documents quickly are favored in this industry. Finally, professional services firms have been early adopters of document management technologies. Therefore, any search product selection process must come to grips with legacy document or content management systems that are well established in the enterprise. Buyers will want to confirm that appropriate connectors for applications already in use are supported.

Engineering and Manufacturing – Electronics, metallurgical, process and heavy equipment are industries closely related due to their focus on invention and on device and process design. Unlike the physical sciences where rigorous scholarly research is the norm, engineering findings are accrued in test and measurement results. Data accumulates in formats of their own unique structures, in computer aided design (CAD) systems, drawings and specifications documents. Because much of the content is graphical in nature, good automated retrieval depends on metadata, which may or may not be present or normalized. ETL products for “discovering” and normalizing metadata, and search products that been optimized through APIs or other modules to work well with CAD and other imaging systems will top the list of products most favorable for this vertical.

Biotech and Pharmaceutical – Drugs, medical devices, and genomics are the top product types in these industries. More than any other vertical, there is substantial uniformity in how they operate because their fundamental work is based on biological and chemical sciences with centuries-old methodologies, backed up with scholarly research by both other scientists and original bench science content. Furthermore, the path from invention or discovery to market is tightly regulated and requires rigorous documentation of a particular format and type. Companies that bypass “the rules of engagement” in this industry do so at their own peril. Short circuiting the process eventually catches up with them. Search products must meet a higher standard for relevancy because this vertical expects more from search. It has been using search technology for decades and knows how to use it. Additionally, published content acquired through various fee-based services has long been a core electronic resource. Search systems that are acquired for finding and presenting results from internal content must provide superior federating technology that filters, merges, enhances and presents internal and external content together. Users in this vertical are extremely sophisticated and demanding about how search results are presented.

Chemical, Materials, Energy – Chemical development and processing, materials science, petroleum and fossil fuels, and alternative energy all share characteristics of engineering and pharmaceutical industries. The obvious reason is that they are populated with both engineers and chemists of various disciplines. Materials science and alternative energy oriented firms are growth industries with significant innovation. However, other established energy industries and chemical companies are also venturing into new technologies and research. Search in mature organizations must deal with vast amounts of legacy content, much of it previously indexed or cataloged manually or in legacy “home-grown” systems. A complete audit of this older material

and a plan for how to get it into formats for crawling and indexing can be aided by data mining and ETL products.

Having direct experience in these industries, the author makes a strong case for investing in legacy content conversion. Companies in this vertical have huge amounts of content worth mining as the world struggles to quickly build solutions to solve energy and environmental problems. Historically, there have been thousands of similar initiatives that never came to maturity but may have great value now because enabling technologies have been developed that will improve economic viability. This is an issue peripheral to search but is a perfect example of where search and search related technologies can be leveraged to advance mission-critical programs by building on the past.

Transportation – Commercial carriers, shipping, and public transport are industries directly linked to but very different than energy companies. These businesses depend on data, more than any other content, to manage equipment, supplies, fleets, personnel, scheduling and traffic control. Operations, research and financial analysts work small margins in these industries to make efficiency gains and control costs by optimizing operations. Search products with embedded analytics, BI, and modeling capabilities or products that can be easily integrated with these functions to leverage piles of orthogonal raw data are the best bets for consideration.

Consumer Products – Clothing, appliances, and household goods businesses are now highly diversified or owned by highly diversified organizations. They will find benefits in search products similar to those suitable for chemical companies for their R & D content or those for transportation companies for their raw materials, manufacturing and supply chain operations. Because this vertical depends on extensive advertising and marketing, departments that are heavily invested in these operations will find benefit in the search products that support social and collaborative functions because creative team work is such an essential aspect of their functions. Tools for rapid sharing of content containing ideas and market research should be at the top of this enterprise group's focus.

Publishing and Media – Entertainment, news, and print media are experiencing intense competition, market upheavals and extensive experimentation with new business and distribution models. More than any other industry they depend heavily on search to do business as well as in manage content, their biggest business assets. Little can be said about what they need in search products in a generic statement, for “one-size” does not fit all. They will be pushing the boundaries of what search can do for the bottom line to reach the public arena and to push the creative envelope internally. For vendors, it is valuable to recognize the opportunities, volatility and churning in this business; vendors need to use imagination and inventiveness to position products successfully in this market. While traditional publishing has historically been risk averse, media companies are being forced to change, too. It is critical to make search applications and technologies easy to understand, adapt to, and adopt.

Telecommunications – Devices, networks, and carrier companies have two sides to what search means. Internally, they will seek search solutions similar to what

engineering and manufacturing use, particularly for their own invention and engineering. They behave like other electronics related companies. However, increasingly the author sees a rising opportunity for them to leverage their own hardware and signals technologies to deliver search externally, for both Internet search and enterprise search (which is beginning to be an offering, starting with remote access to email). While search product vendors are pushing search options to device users, it may well be that the evolving business model will be partnerships in which device manufacturers will optimize for particular types of search, or carriers will bundle search brands with service options. The possibilities for search integrated with this vertical's products and services are vast and waiting to be exploited.

Aerospace and Defense – Aircraft, weapons, and signals intelligence/radar can be stealthy businesses, to make a pun. What they invent, design and build is dependent largely on government procurement, which in turn is driven by politics and world events. This industry has core engineering and manufacturing operations, but winning contracts is often achieved more by the promise of being able to deliver in reaction to a request for proposal (RFP) than innovative ideas pushed through a standard R & D process. In this business, the ability to leverage vast amounts of legacy content for the purpose of efficiently preparing proposals, keeping track of hundreds or thousands of contractual components and milestones, and preparing technical documentation for delivery with unique products is a challenge of scale and complexity. Products delivered to specification are usually built from existing components plus newly invented components. They may be assembled in new and inventive ways. All of this requires that contractors are able to manage the content associated with each variable by itself, as well as integrated in new designs. Search products that integrate with popular document management systems (e.g. Documentum, Astoria) and CAD systems will be favored in this industry.

Finance – Investment firms, banks, and regulatory agencies operate on structural data and much of it in real time. Search coupled with analytical and BI tools is critical, and it must scale to handle ever increasing amounts of content in very specific categories and time frames. Like law and scientific research, federating internal and external content is vital, as well as being able to provide visual graphic representations from results (e.g. charts, graphs, spreadsheet images).

Non-profits – Museums, foundations and philanthropic groups, and institutions sponsoring social programs are among those looking for solutions that appreciate the visibility search will have in a public setting. Many of these organizations have valuable institutional content that they want to share with their members or the public at large. Because so much enterprise-centric search is behind-the-firewall in other verticals, it is difficult for vendors to point to demonstrations of their products in action. Giving special attention to non-profits, aiding them with making visually attractive and elegant but simple search interfaces, is one way for search vendors to gain market exposure. Finally, the diversity of content, objects, images, video, audio and text make for a rich target for those search products that can index and provide Web access to these types of information.

Horizontals (Functional Groups)

It has long been assumed that “enterprise search” targets opportunities that bring entire organizations under a single umbrella of search. Only large search vendors FAST, Endeca, and Autonomy were positioned that way at various stages of market evolution. Enterprise platform vendors like IBM and Oracle have been positioned to handle that promise as well, providing tools to integrate a multitude of software applications with search for companies well-heeled enough to afford IBM’s consultants or other system integrator partners.

However, numerous more recent search vendors have staked their reputations on products targeted at particular professions and verticals to solve unique search challenges, many noted in the previous section. One difficulty of doing business is to identify markets that are large enough to create a good value proposition for inventors and their investors. Even after three-plus decades of search in the enterprise, search is still poorly understood by most professional groups within organizations. This is now changing but historically, with search only being taught by search service vendors or in graduate library schools, most professionals came to their workplace with no particular expectations about search as a workforce tool. As a result, there is a significant variety of knowledge, behavior and skills related to search within the enterprise and what it may or may not do to help with work. More obvious is the fact that everyone’s experience is colored by Internet search engines, Yahoo and Google being the most used.

The certainty of cultural diversity even within a vertical market, coupled with specific professional expertise about content in a particular discipline, presents any product vendor with a marketing and selling situation of importance. Selling to a particular group of professional uniformity only requires focus on what their understood needs are. That is one reason why vendors often take their products to professional meetings where a unified message within one vertical works well. When a vendor is called into an organization to make a presentation to a diverse group of professionals, all with different levels of search expertise and all with biases about how search should work for them, it is just about impossible to deliver a message that resonates positively to all present.

This is obvious to search solution providers and experienced marketing and sales professionals. As a former vendor, the author is encouraging buyers to be sensitive to this because it will help them be respectful of a process that needs to be followed to get the best and most appropriate product for the organization. It will also stand them in good stead over the years to begin to have a meaningful and transparent dialogue with companies with whom they expect to do business, now or in the future. There is nothing to gain by building an adversarial relationship with suppliers, and everything to gain by building an atmosphere of trust and clear expectations. One simple example: people don’t like to discuss budgets. This often seems to be the elephant in the room because buyers don’t like to admit that they don’t have control over money, or as much as they would like. They are defensive and fearful that they will not get attention from a vendor if they lack purchasing power. If buyers receive this reaction before the sale is closed, imagine how poorly the vendor will support them after. Look elsewhere; there are

enough options in the market now. Be upfront about a non-negotiable budget and look at products that are in the target price range. Another alternative, given some flexibility, is to work with the favored vendor(s) to help justify spending more than planned.

Here are a few comments about various professional functional groups as they relate to search within the enterprise. These are generalizations and are intended for a couple of types of readers. For leaders of a search selection or implementation team, particularly in a small or mid-sized (under 2,000 employees) organization, communicating with people from each group will help the long-term success of implementing search across the enterprise. They are potential stakeholders. Not recognizing their presence in the organization and their needs means that the chosen solution may need to be replaced sooner than planned or supplemented with others. It is also important to understand the range of diversity of needs for search; buyers may, in fact, decide that two products would be more suitable than one because there is no overlap of need for specific content domains. For vendors, if representatives of any of the following groups are involved in the selection process, it certainly helps to be able to identify with their professional content search interests.

Ideally, Human Resources personnel should have access to most employee generated content, at least to the extent that they know and understand the generalities of an employee's work product. Their most critical need, however, is access to all business organizational documents relating to how the company is organized and run, regulatory and personnel related legal documents, and all instruments of personnel management, program and benefit plans, hiring and contracting materials. If the public web site is not sufficiently complete about the organization's mission, products, and services, they will certainly need access to internal documents that supplement what is not public facing.

Finance and administration needs quick look-up access related to cross-enterprise procurement, manufacturing supply chains, payrolls, product and service vendors, customers and deliveries. Most of this information is in databases so access to that structured data in ways that are easily understood and presented will be highest on their list of priorities.

Text and unstructured documents and reports) and data from other applications (e.g. test results, design drawings) are primary resources for R & D. If the organization uses document management and/or content management systems for structural management of unstructured documents, the search engine must access all those relevant applications and databases, as well as any file shares. Finally, it is increasingly common for R&D to work in communities of practice different than their formal organizational groups. Work products are shared and collaborated on through systems like Sharepoint. Work in progress is usually just as important to search as internally published final reports. Finally, being able to link experts with the content they have produced is a very high priority for R & D.

Manufacturing needs process and equipment manuals and specification drawings, plus access to standards. Supplier catalogs, customer requisitions and delivery information are essential for them or any other group charged with getting product to the customer.

Business Development and Competitive Intelligence are two areas for which external search are probably more important than enterprise focused search. Their external access requires substantially more content than what is found through generalized Internet search. Subscription based content or content for a fee is also vital to understanding markets and competing products. The large numbers of these services means having IT resources available to make access easy and to help with federating content from outside with that inside the organization. Internal access needs will usually focus on information generated by the sales and marketing organization with occasional need to view customer documents that result from earlier sales and support services.

Customer support is often neglected in search initiatives because the organization fails to realize how much efficiency and positive reaction from clients will be gained by having snappy, complete and accurate responses from the support team. Good search for support personnel requires a comprehensive knowledgebase into which all products and process-related documentation is fed, or through which it is searchable and accessible. Any data capturing mechanism that is used by support to record customer inquiries and responses provided must also be part of the knowledgebase. Of course, these content input resources must be populated and maintained to be of most value. When search works well for customer support, the benefit to the organization will be apparent. Search logs that can be easily tracked for activity from customer support will prove highly valuable for seeing where the knowledgebase has holes that need plugging. The needs in this area are really low-hanging fruit.

Marketing needs a pipeline to anything product or service related. Technical staff are often assigned to marketing departments to brief writers and aid marketing campaigns and collateral development. They represent the voice of the potential customer and are a good resource. In addition, marketing needs access to technical documentation and project business documents that will enable their own internal research when they are working to develop product descriptions. Having access to customer support records, or other material that will help them “tell a story” about how products are being used in the marketplace, will be an asset.

Sales people are often tough to pin down about what they need to be able to search. At a minimum, having access to lead generation data, marketing collateral, and legacy information about current customers who might be prospects for new products will be at the top of the list. To this add email search, since the sales process is most often managed through contact lists and exchanges with prospects, and the lead tracking databases that they themselves need to update. Experience in many organizations shows that too much information stays in email folders; it never makes it to more structured applications for managing prospects. A realistic approach is that sales people may be the neediest of search across highly unstructured content.

Legal departments in large enterprises operate more like professional services firms, described in the vertical market section. However, in small and medium organizations, there are often only one or two attorneys who will rely heavily on search to keep them abreast of issues related to regulatory, finance, intellectual property, intellectual assets, idea records, contracts, and licensing as it directly relates to the organization. Again,

Enterprise Search Markets and Applications

like professional service firms, search that federates internal and subscribed legal content in unified search is the gold standard content architecture.

Planning for Enterprise Search: User Guidance

Thinking about search in the context of its application (e.g., what it is for) and how it relates to the content environment is at the highest level model for beginning to plan for search. Figure 6 below illustrates four major search concepts covered earlier in this report on the left. But before thinking about products to review, seeing where content comes from, how it is built, staged and stored will provide a roadmap to narrow the search for search. To pursue this understanding, review the breakdown on the right side of Figure 6.

The bottom tail indicates a general sequence from content creation to ultimate use, content reconstitution as it is applied in some form. The second branch in the middle reflects three stages of content evolution from its original creation and editing processes, metadata creation (index), extraction and insertion from other content sources, and use of various look-up and validation tools in the creation process. Validation tools are relevant if they include taxonomies to exploit in search navigation. If metadata is created for content, selecting a search product that will use it effectively is important. Content creation may also be the result of an actual search process from which it is reconstituted as new content (e.g., mashups and analytics). Finally, the top right reflects a series of possible end purposes for content—post-creation processes that may be the ultimate roles of the search product to be implemented, or one complementary to search.

Why is this framework important to recognize?

- First, because when buyers engage in search product selection and contemplate the content to which the search engine will be exposed, they need to know where that content is and what attributes of it users want to index. Knowing in advance all the content creation processes that exist in the enterprise or target area will be invaluable in explaining to vendors what will be searched, and how. How many applications or databases contain target content, what servers these reside on, how they are networked in the enterprise, etc?
- Second, establishing all the expectations about how, what, for whom, and why users want to be able to search will lend clarity to defining which products will actually meet business goals.

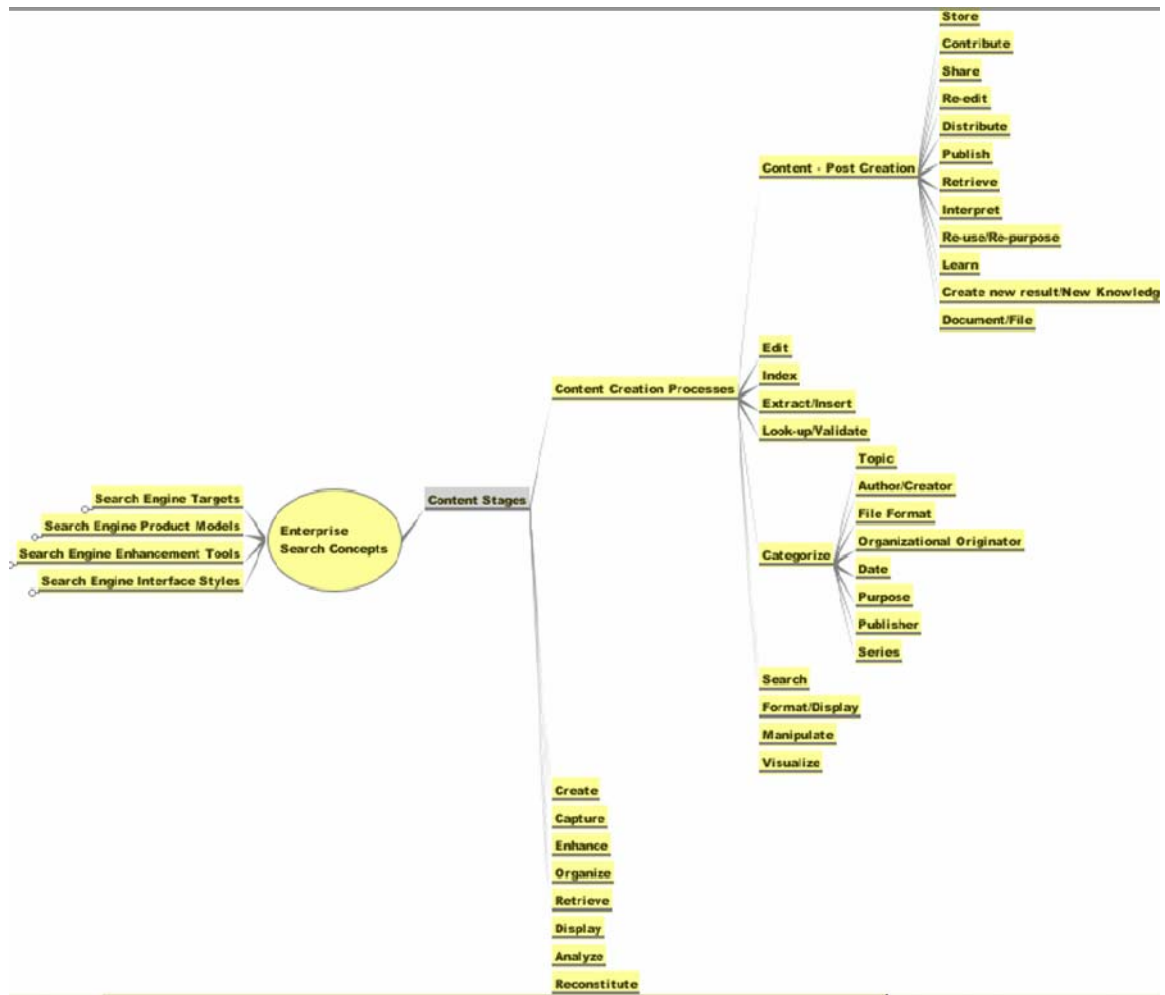


Figure 6: Search Concepts with Detail on Content Stages

Enterprise Search Research

The next sections reflect information about a couple dozen enterprises the author has become familiar with over the past few months, recording their experiences on selecting and leveraging search products. The summaries reflect both common experiences and behaviors, and some commentary about unique variations in experience. While all those interviewed have had a net positive experience with search, and most expectations have been met, when they were pressed, all found something to reflect on as needing improvement. All of those interviewed in depth planned search deployment in a phased approach, expecting to build on the knowledge and experience they have gained already to continually improve outcomes for users.

Many organizations are adopting and adapting search as the center of their portal or intranet design, making it one of the gateways to enterprise content. In addition to search through an intranet for access by employees, some of the content may still reside in applications for a program or project, or be otherwise secured and searched through embedded, application-specific search.

These summaries are provided to encourage readers who are just beginning to make inroads with enterprise search; they are in good company, with many others “just getting their feet wet.” For users looking for a change, for big improvement or replacement of the search products already installed, there are some ideas for how to do it better the second or third time around.

About the Survey: General Findings

There was a Y/N question for participants: “Are you a consultant?” Fifteen of 36 responded *No*, eleven *Yes*, and ten did not answer the question. Across all respondents, distribution of role involvement with enterprise search was interesting. There were five options that any respondent could check: Selection, Implementation, Tuning, User and Other. Eight checked off having a single role, and five checked “Other” or did not respond. Everyone else has had multiple levels of involvement.¹

Participation in Enterprise Search Product Activities

Activity	Role			Totals
	Consultant	Employee	No-Response	
Selection	7	11	6	24
Implementation	9	14	6	29
Tuning	4	9	4	17
User	4	11	4	19
Other	3	3	3	9

Figure 7: Survey Respondents’ Role Distribution

As illustrated in Figure 7, most respondents were involved in product selection and implementation. There is some indication that consultants tend to move out of the picture once a product is selected and installed. While not fully reflected in Figure 7, it is noteworthy that while most consultants checked a box that indicated they were also involved in search in their own organizations, they rarely indicated that they are users of search. However, non-consultants involved in search selection and implementation checked the box indicating that they were also users. One-on-one interviews tend to bear out the fact that employees who are involved in responsibility for search identify closely with the user community. Because there was so much rich information from non-consultants, the interviews did not include consultants. Previous experience with consultants, of which the author is one, would indicate a wide disparity between

¹ The sampling is not large enough to be statistically significant and respondents were self-selecting so the data itself can only indicate some trends in experiences.

consultants who have significant direct experience with search as users and those who only use search in short stints. Where possible, Gilbane interviewed people who were also search users.

Product Experiences Tallied

Of 36 respondents, 26 had experience with two or more search products with the average being three. Six indicated experience with only one and four did not name any.

Of the respondents who were later interviewed, few had enough experience with more than one product to feel comfortable extensively comparing the product selected for the interview with others they had experienced. However, some made comments about second or third products in passing that are reflected in the following experience summaries.

Table 4 lists the company or product names as labeled by respondents on the survey. Rows for five names were presented on the survey form, but clearly some people had more experience than with five products. Looking across the data, variants of search for Microsoft environments (including MOSS) had a high count, as did the family of Autonomy products (including Verity and UltraSeek) and those for Google.

Autonomy (8)	Lucene/Solar/Nutch4 (3)	West KM (1)
Autonomy IDOL (3)	Mac Spotlight (1)	WorldCat (depends on ES view) (1)
Ultraseek (2)	Microsoft (1)	X1 (1)
Coveo (CES) (1)	MS Livesearch (1)	
Dd (1) – may not be valid	Microstrategy (1)	
DTSearch (1)	Mondosearch (1)	
Endeca (3)	Muse's content federation (1)	4 others (1)
Fast ESP (4)	Novell Quick Finder (1)	All the Web search engines (2)
Google (9)	Ontolica (1)	various bespoke index-based search (1)

Google Appliance (5)	OpenText/Livelink (3)	Several OEM searches (1)
Google Apps (1)	Recommind (3)	Univ. library catalogs (1)
Google Custom Search (1)	MOSS/Sharepoint (8)	West KM (1)
Google Desktop (2)	SLI Systems (1)	WorldCat (depends on ES view) (1)
Homegrown based on swish++ (1)	Verity (3)	X1 (1)
IBM OmniFind (2)	Verity K2 (1)	
Intellisearch (2)	Vivisimo (4)	
Lotus Notes indexed search (1)	Express Search (Vivisimo's Velocity, targeted at Interwoven's WorkSite DMS) (1)	

Table 4. Survey Respondents' Product Experience (names of their choice)

Field Experience with Search

The following information is presented in an anecdotal, non-quantitative format for one important reason. After thirty years, search in the enterprise market is still highly fragmented. Data that have been widely published about the installed base is from vendor records of number of sales. Having been on the vendor side, the author can vouch for poor accountability of these statistical summaries. They rely largely on how the vendor counts and are not often verified. The industry knows this but most users do not; as a result, this data is hard to codify and quantify. The majority of companies in this business are privately held. They may also have multiple products, or be part of larger corporations in totally different businesses. Therefore, counting customers or installations is left to the vendor's judgment. Whether customers are still on maintenance or actually using the product is another unknown when perusing customer lists. Deployment and licensing models differ radically, so that comparing five seat licenses for a 100-person firm with a 2,000 seat license for all employees in a firm is not really meaningful. Licenses to non-profits or for educational purposes are often free or low cost, which may also mean that the vendor is not providing much support. The impact on the vendor's resources in these cases is minimal, so one cannot judge

their customer support operation as servicing 200 accounts if only 20 are paying for full service.

As pointed out earlier in this report, search technology underlying dozens or even a couple hundred commercially-deployed products is quite advanced. Therefore, these sections of the report with user experiences are meant for the reader to learn how others have gone about selecting and procuring their products; what happens in the buying, installing, implementing and deployment phases; and, most important, what the end-user experience is like. Finally, after working with a vendor and its product for a while, how does an enterprise feel about the business experience and what would they do differently the next time?

Selecting Search Products

Among the enterprises studied, the process of selecting a search product for widespread enterprise use generally stretches over a year to 18 months. However, the process varies widely from one organization to another, ranging from selecting and testing a product recommended by someone in another company, to adoption of a product supported by the parent company, to engaging in a lengthy and time-consuming winnowing of options from a long list, reading volumes of documentation and “test-driving” one or more possibilities. Among the parties interviewed, one individual, whom the author would call the “search champion” or “search evangelist,” usually controlled and guided the selection process. It appears to be the norm that this individual will continue to be actively involved in implementing and have some role in on-going support and maintenance. These people are also keeping an eye on the marketplace to learn what is coming along that might be better for enhanced functionality and features. In other words, they are actively engaged in search even though most had other significant job responsibilities. Their engagement and interest in search usability and technology was probably a strong factor in their response to the survey and willingness to be interviewed.

Interviewees were asked to name two or three top criteria that had to be met before considering a product. Here is the list as supplied by ten organizations, all heavily vested in scientific and technical research activities or legal research.²

Highest Priority Criteria for Selecting a Product (number with similar comment)

- CIO had to be comfortable with what he saw (1)
- Clustering (automated), Facetted search (3)
- Correlate people (expertise) information with the content they had worked on (1)

² They are presented alphabetically to collocate similar topics.

Enterprise Search Markets and Applications

- Cost must be five digits or less (5)
- Cost structure needs to not penalize high volume of content when the firm is small (1)
- Cost to fit a budget under \$500K including overhead (1)
- Ease of deployment; short cycle (weeks or couple of months) (2)
- Ease of support and maintenance (less than one FTE) (1)
- Integrate search with .NET-based customer software (1)
- Integrates external search results from subscribed content with internal work driving the research (1)
- Integration with application that had embedded search PLUS other internal content on file shares (2)
- Metadata indexing stored in a variety of formats and normalized; needs to leverage property data (3)
- Operates in an MS-SQL environment for the database back-end (1)
- Proven connection with Interwoven (1)
- Proven connection with LiveLink (1)
- Proven connection with Sharepoint (3)
- Proximity searching (1)
- Relevancy and quality of results (2)
- Scalability (1)
- Security; maintenance of access control lists (3)
- Speed of indexing and search speed (2)
- Structured and unstructured content; DB content; PDF searching (3)
- User interface that people would actually use; need to be able to get answers quickly when speaking with clients (2)
- Vendor experience in their industry (2)

Selection Activities that Impacted Decision-making (number with similar comment)

- Did a huge amount of research to ID products that would match requirements, narrowed it to one, downloaded, converted metadata, indexed and demoed it. (1)
- Not a lot of comparison recently; in the past (2 years) enterprise search appeared to be too costly to consider. (1)
- Considered Lucene open source but overhead to maintain was too great (1)
- Had research experts in house do the searching for possible products (2)
- Got a trial version, set it up and let others test drive it (POC) (2)
- Always testing products keeping in mind the amount of maintenance and overhead needed to make it work the way they want (1)
- Inherited the search indexing system when the company that built it was acquired (1)
- Recognition that different search products will be required for a variety of content uses and sources (expect to deploy two or three search products) (1)
- Team of five narrowed the choice to one from eight systems with significant internal technical support (1)
- Experience with other systems in-house influenced key selection criteria (1)
- Found some products to be too Web search oriented (1)
- Definition among products of what constituted support for thesaurus was very uneven, not well-defined (1)
- License came through the parent company easily and that was more attractive than procuring other products on the market (1)
- Proof of Concept (POC) might be considered but seems like a waste of time and resources (1)
- Knew the selected system **was it** as soon as he started testing it; up to then had looked a dozen or more that could not make the grade. (1)

Purchasing

Only a couple of the interviewees had significant negative comments about the business relationships with vendors. It is not surprising that those were about vendors not selected. Most notable were comments about the complexity of licensing models and options. With increased virtualization, it is not an easy calculation to define

1. how many servers are involved in a license that is priced based on which servers are being crawled,

2. where searchers sit in the network relative to the installation, or
3. how the software will be deployed in the network.

Add to that mix what defines a server and its characteristics. Finally, numbers of documents indexed often does not correlate to number of users being served, especially in professional services firms. A firm with 60 professionals but several million documents, many in the form of emails or short memos, will have a hard time justifying the same price tag as a ten-thousand employee firm with fewer documents.

Most people interviewed had engaged in serious, hands-on evaluation before purchasing, either with a downloadable demo version, or a full-blown “proof of concept” (POC) with their own content being crawled and indexed. This is highly recommended to ensure that expectations will be met and to give stakeholders an opportunity to get their hands on the product. It also helps to get the best deal when buyers can point to a problem or deficiency for their organization, and can then work out with the vendor how it will be resolved now or in the future, a trade-off on price, or customization at the vendor’s expense to close the business arrangement. The author highly recommends a POC. Often the POC can be converted to a license so that work done for test purposes is not lost.

The author has made the observation that the higher-priced systems seem to have the most requirements for customization and heavy overhead, while the lower-priced systems offer significant benefits with minimal tuning. This is a difficult value proposition to understand, especially for large enterprise buyers. They tend to equate cost with value and are suspicious that a lot of technology can come “out-of-the-box” and still be appropriate, scalable and reliable. The cases examined reinforce the author’s belief that enterprises should be looking for good value at a low cost, especially for search within divisions of large enterprises or in small and medium businesses.

Having periodic Q & A sessions with a vendor throughout the decision process is important. This will reveal a number of things. After speaking with them a second or third time, are buyers left with the sense that they know the company and that they hear what they say about requirements? If the representative can’t answer a question, does he or she get back with a response that is timely and is the question actually answered? This is also an opportunity to ask them questions about their company. One of the author’s frequent questions is, “How are you using this product yourself? What are some of the benefits that you find using it for your sales job?”

An often-repeated comment by those looking to replace search products or planning to procure for the first time is that Google Application Server (GSA) is the product to beat. The reason is partly due to price but also simplicity of the business model. The pricing is easy to understand, the procurement is straightforward, the application is quick to install and deploy out-of-the box, and buyers only have a few models to choose from. Buyers do not like to hear that they can select from an endless list of pricing structures and a long menu of optional add-ons. Having all those choices also strings out the procurement process as buyers agonize over whether they are making the right choice.

Buyers want to know a fixed price range for a vendor's products very early in their selection and narrowing process. Requiring a lengthy engagement period for presentations, demos, and talk before a vendor is willing to even give a price range is a waste of everyone's time and effort. Organizations have budgets, and will purchase within a range. It is foolish for vendors to try to determine how much enterprises are willing to spend before giving out any pricing information. There are now enough products on the market that are relatively low budget and packaged appropriately for most enterprises that buyers should not have to get caught up in the "request for quotation" (RFQ) process until very late in their acquisition cycle. Only if they have a general idea of what a vendor's pricing looks like, and if know that they have the budget, should they engage in in-depth discussions. If a buyer gets to the RFQ and the quotation is a big surprise, way out of range from earlier discussions, a red flag should go up immediately about what the long term business relationship might be like.

Finally, any capital procurement of software should be vetted for the amount of human overhead that will be required for initial installation, implementation, and ongoing support, whether the human resource is internal or consultants. *More disappointment with search products is caused due to inadequate human resources than any other cause. More failures are due to human factors, rarely the technologies.* Support is needed for:

- Adequate implementation and tuning
- Intelligent and appropriate search interface design
- Ongoing refinements that continue to add value and leverage content in ever more sophisticated ways
- Continuous tracking of search logs to analyze and understand what is being searched and how successfully
- Routine updates/upgrades and adoption of new product features

The bottom line is that as with any other business applications, search needs care and feeding for which budget must be available. Without that, search will fail, and the organization will not have another chance to do it better or replace the product for a long time. Talk to other users of the product to test what the vendor say about support.

Implementing

The implementation stories are far more complex and varied as to how enterprises came to a final deployment. Time ranged from weeks to years, and some were still working out the kinks of POCs that were evolving into the final implementation. What was remarkable was the number of quick (a couple of weeks to a few months) implementations that were deployed and got immediate adoption by a large population with continuing escalation of use over several months. Taking some liberty to disguise specific cases, here are some stories.

Enterprise A chose to install the search product shortly after downloading it to make sure it performed as they wanted it to. Within six months, after writing a few macros to

reformat legacy content and add metadata, it was rolled out with default options to see how users, about 50 internal professionals and a few hundred subscribers, would react. They already had a simple taxonomy they could leverage. Over the past year, they have developed some custom settings and minor improvements; the developer is not a computer science person and has other responsibilities. Users are very happy, and he is pleased with the ease of use.

A search product that came with an acquired company at **Enterprise B** was in use with a limited presence, but lacked comprehensive coverage of the organization's content. A person who joined the company and worked in knowledge management discovered another indexing tool embedded in their own products. Because that product is developed for commercial use and needs a test bed before releases, a synergy has developed that benefits the company ("eating its own dog food") and clients because they can influence product development. The indexing engine has now been accepted as the tool for ensuring content consolidation. They recognize that ongoing support is needed and are proactively adding content. It has become the corporate intranet and gains mindshare all the time. The search resource is becoming a trusted asset as coverage becomes more complete, as they take care to make sure items are adequately identified, and as the engine is tuned to ensure relevance of search results.

Over the past year, the author has talked to a number of enterprises that have Google Search Appliance. The general consensus is that GSA does a good deal out of the box for its price, but with good internal technical support, it can do a lot more. Time saved on installation and maintenance of the appliance is better used for customizing options, tuning and writing interfaces, and integrating target content. Its strengths are its ease of purchase, installation and maintenance. On the down side, all users have issues with trying to influence Google or gaining the ear of Google's engineers about the importance of features unique to the enterprise. More than one company commented on Google's lack of understanding of search needs for enterprise users. They seem to respond to all search as they would to an Internet search experience. Lack of support for clustering results and poor security options topped the list of complaints. Respect for the criticality of search when it is being used to find business content as part of the business process also seems to be lacking. When a server crashed at one company they were requested to re-boot repeatedly until they were told a replacement server would be shipped. If they had not been running a redundant server, they would have been without search for a number of days, not an acceptable response. Most GSA users have adopted it for expedience in "getting going" with search, or as a stopgap until something better comes along. They are constantly surveying the marketplace, but in the meantime are gaining critical experience that will benefit them when they migrate to a successor product. In the meantime, Google is not standing still either so they may hold on to these users over the long term.

Enterprise C selected and deployed a search product in a few weeks after investigating the search market over a couple of years and studying five very diverse products. That study effort rendered significant understanding about the nature of search technology. Knowing that the user interface would be the primary key to adoption, and faceted or parametric search would deliver great benefit, it was the interface that sealed the selection, and it has paid off. Incidentally, this customer

indexed across all file shares, which immediately revealed in testing content that they needed a layer of security before it could be rolled out. Rather than agonize over every corpus of content before adding it, they added everything and then subtracted or secured what needed to be secure. This approach is one that seems to be growing in popularity.

Enterprise D is a good example of a key trend in search: serving a special business population within an enterprise by solving a particular problem. In the case of this organization, a legacy system was in place for indexing content coupled with custom code to route documents to create an alerting service. A turnkey enterprise search product was acquired to replicate the model, enhance relevance of retrieval, improve indexing speed, include more types of content in the index, and federate internal and external content. After a few years of dissatisfaction with the previous solution and rejection of a number of search options because of document-based pricing, the organization found the current system. It is being heavily developed by their own engineers to meet very controlled and specific business requirements. This is a situation in which a measurable improvement during the POC over the previous system will determine the success of the new deployment.

A second case of leveraging search bundled with a special business application is being rolled out for **Enterprise E**. The product being deployed is targeted to subscribers of published specialized content with the purpose of correlating that content directly with internal content for work on specific projects. It was deployed originally with defaults and had a very low adoption rate, which proved to be the result of lack of training and professional scrutiny of what was included in the internal crawled content. Now that the enterprise has engaged a full-time expert, and makes use of subject matter experts to govern content additions, remove junk, and build a mentoring and training program, usage has jumped to 50% of eligible professionals. Because they are being coached and shown how to integrate it with their work, the success rate of use and understanding has improved dramatically.

One comment made in several discussions was the importance of context in search results. Implementers stressed the need to make sure that search results have good descriptions so people understand what they have found.

Finally, a few of the implementers underscored their increased commitment to incrementally improve the success and adoption of whatever search product is deployed. Their organizations are doing this by assigning an internal administrator or consultant to the implementation and deployment effort. One state of evolution that seems to be the norm is adding faceted or parametric search, and/or clustering based on taxonomies or controlled vocabularies. Some commentary in the technical press about search has expressed the view that the labor involved in taxonomies is not a sustainable model and that semantic search will replace automatic indexing coupled with human maintained taxonomies. The author's view is that the learning curve on what belongs in a taxonomy and how to develop and maintain it has been a slow process, but has caught on. When an enterprise understands how and why search results categorized appropriately for their unique enterprise needs brings them benefits, they embrace and commit to that model. When administrators learn that a

topical navigation model with drill-down options helps special audiences understand the scope of content, they move to build out that feature. On the other hand, semantic search is only rooting itself slowly in specialized domains, and the evidence is slim that it has widespread deployment in enterprise search yet.

Deploying and Adoption

With an increasing number of search products on the market that install easily and provide significant, out-of-the-box features and functionality, buyers are trending toward getting started with the default set-up and then tuning to user reactions. Often these users are in a test mode with a representative group of early adopters for in-house application. Administrators and search “evangelists” within enterprises try to select individuals or groups that have a pressing business need for search, a well-defined search problem or a special interest in search applications in general. Information professionals, such as librarians, may have a longer term experience with search technology and can be valuable for product selection and evaluation, and for on-going tuning and maintenance or training. However, it was noted that their advanced level of understanding of search does not necessarily make them the best representatives to judge user interfaces or how special features might be viewed by the larger general population of an enterprise.

One administrator pointed out that new employees make the best testers because they are in need of a lot of information to get started quickly. The author concurs that “newbies” are a great source of concrete evidence if search is working or not. They can also bring insights into other ways to deploy from their previous experiences. They are worth proactively reaching out to for commentary.

It is long overdue that organizations planning to deploy search products understand the need for perpetual oversight and management. Software applications evolve and require maintenance and upgrading. The richness of the technology requires skilled tuning and adjustments to ensure continued value. Our interviewees also strongly encouraged a program for training, some type of regular surveys, and generally a level of regular contact with users to get a sense of “how it’s going.” Responsiveness to what is learned in these communications will result in continual improvements and enhancements to the interface, scope and quality of the content being served, and trust in the search tool’s value.

Deployment of search often brings about innovation in content and knowledge asset management. For example, the idea that search can be used as a tool to learn about what is available to leverage within the enterprise reflects its close kinship with data mining. Because search tends to be used more heavily with unstructured plus structured content, it is a natural application for bringing siloed or disparate content into a unified view that can itself be instructional.

One surprise emerging from the research is the sense of afterthought among the people interviewed about their use of log files to get a better understanding of how and for what search is being used, and how much. Not routinely examining search logs is a

missed opportunity for understanding weakness is the scope of content, discovering terminology that belongs in taxonomies or ontologies, and learning more about the user community. When this was brought to the attention of administrators, several made a note to add this to their routine administration activities.

Maintaining

It is clear from all responders that search products at every level, even those quick and easy to install, are not technologies that implementers can install, then walk away. Every interviewee, even those who would not have ongoing responsibility for administration, expects to be involved for the long term in some capacity to ensure positive outcomes for users.

Ease of upgrades seemed to be better for the lower-budget products. Those vendors appear to be providing excellent value in terms of understanding that low license cost needs to also reflect low support costs. That is encouraging and good news for those looking to begin with a modest budget.

Their planned use of every feature included in upgrades was commented on by several administrators. This seems to be a sign that none of the products are mature to the point where they fully satisfy all requirements. Some administrators had gone ahead using APIs to develop custom features for specialized content or to manage search results in a more elegant fashion, and were glad when the vendor included similar features in later releases. However, this usually required some adapting of a new release for the installation because the custom work either could be scrapped or had to be re-deployed in some fashion. This relates to another point brought up several times about enjoying working with specific vendors because they were responsive and did take suggestions enthusiastically. Finding vendors early in their development who want to hear new ideas from clients is an excellent sign of a business relationship that will last because these are the vendors who are more likely to survive.

All GSA users seemed a little less pleased with the business relationship and responsiveness to problems. Several noted that bugs fixed once tended to reappear in later releases. The company size, newness to building a close relationship to enterprise clients and number of enterprise GSA systems (a couple of thousand) they have deployed in a short period of time reflect that they still have a learning curve to sustain this business. However, they do not lack for resources and will undoubtedly figure out things like code control, customer support, and other “must haves” for enterprises, if they remain committed to this market.

Human Resources/Staffing

Wrapping up guidance for buyers and user communities, it is clear that enterprise search deployed for a company division, or small or medium-sized organization needs to have at least one full-time equivalent (FTE) for ongoing product support. Dividing the role between two people works well because there is human redundancy and back-up. This also provides continuity if one person leaves. Most organizations with a high ratio of research oriented professionals to total employees make use of departmental or

Enterprise Search Markets and Applications

project subject matter experts for some aspect of content, metadata, and taxonomy support to ensure the quality of content contributed for indexing, and the relevancy of search results and how they are presented. The smallest enterprises sometime struggle with a person having other duties but soon come to realize that search is a full-time commitment for someone. Search deployment, tuning and maintenance are never done.

Search administrators and champions come from diverse backgrounds because search is “owned” by such diverse business units in organizations. Search technology implementation and usage are not typically taught as part of a computer science curriculum. Evangelists and those who end up actually responsible for its “care and feeding” after the IT group sets up an environment for its installation come from various disciplines. Among interviewees, besides those with a computer science background, were business managers, subject matter experts and library and information science people with an orientation or passion for technology. Several had a knowledge management focus that they applied to their work, and search seems to fit naturally with that.

Subject matter expertise is highly valued for development of taxonomies, content selection and developing metadata guidelines. The role of librarians is shifting from being searchers of external search to coaching users on the internal search tools and how to leverage them for finding the best content.

Getting into the Customer's DNA: Vendor Guidance

This section is for vendors. It is offered as advice so that they can become better at what they do. Buyers can learn a lot by reading about some of the challenges vendors face, too.

How can those who develop, market, sell and support search products build healthy and successful business relationships with buyers and users? Really listening to the voice of the user and working to understand where their beliefs, frustrations, and positive attitudes originate will help get closer to those individuals. As hard as it is to listen to negative feedback, often as a result of transference of a previous bad experience with another vendor, it can be turned into an opportunity for moving the conversation in a positive direction.

The Communicating Experience Must be Two-way – The author was in the vendor's shoes selling products for over twenty years, and so is empathetic to those whose ornery customer experiences put them in a bad frame of mind. When the tone of conversation with a customer or prospect is negative, it can be difficult and some people just can't respond appropriately to negative feedback. These employees do not belong in a conversation with prospects or clients. If they are developers and good at it, try to keep them away from the buyer, no matter how smart and gifted their technology skills are. If a prospect or client insists on talking to the "brains" or "guru" behind a product, vendors can brief that employee about how and what is expected in the way of behavior. Who is you? Do it in a way to clarify what is on the line for them and the company. It also does not hurt to set some expectation with the customer as to what this person's strengths are, leaving out communication skills.

Why the Trade Show Experience Matters – This may seem like obvious commentary and strange for a report of this nature. However, there have been too many experiences with vendors in the author's recent past where a vendor representative on the floor at trade show demonstrated particularly poor judgment in talking with me, even after introducing myself either as an analyst or consultant seeking solutions for a client. Being "lectured" about technology, why questions were not valid, or what was "really needed" does not sit well. The author would not recommend doing business with a company that showcases products in this manner. If this is a public, pre-sales performance, imagine what the post-sales experience would be like. As a vendor, people who interact with prospects or clients need to be groomed with some sales training or a coach to work on this critical skill. *As has been stated several times in this report, it is*

*not about the technology. It is about what works for the buyer or buyer's users and what the working relationship is going to be over time.*³

Setting Expectations in the Selling Process – It is very difficult when a prospect has become enamored of a product or company through some external force, and over time the vendor discovers that the offer is really not the best solution, or that the prospect can't afford it, or that they do not have adequate resources to support it. Sometimes buyers want to believe it *is* the solution because they need one and have run out of time, have a poor understanding of search, or a respected colleague in another company is using it and they want the comfort of a "best buddy" to be there with guidance and advice. No vendor wants to turn away a sale but the consequences of going ahead in the face of certain failure, disappointment or no chance for success is a situation worth confronting honestly. Again, communication is everything, so explain the dilemma as empathetically as possible and offer concrete guidance on what can and cannot be delivered. Look for openings in the conversation to suggest alternative solutions while leaving the door open to the fact that they may decide that they really do need what the vendor has, or they may find the product is appropriate later for its true value.

Also, when a prospect explains his top three priorities for search in very specific language and has reasons laid out for it, don't argue if the requirement can't be satisfied, and don't try to "paper-over" the lack of the feature. The author coaches clients to be good buyers by stating up front every "drop-dead" issue they have (issues that if not addressed by the product are "show stoppers".) The vendor must reciprocate by acknowledging non-compliance with the request. It can come back to benefit the vendor because, if the desire is for something no vendor can supply for the prospect's budget, chances are that trust will have been built, trust that will encourage them to return for another look.

One other expectation must be spelled out when selling. *Search products do not deliver value without some human involvement on an ongoing basis.* Develop a description for the type of internal support that will be required; keep it simple but honest.

Story-Telling – Having a relevant case study in the vendor's marketing arsenal is so important to being able to convince buyers that sales people are listening to them and have a solution that will work. Selling solutions, which is what search products need to be positioned as, is about building a story line with which a buyer identifies. The author once worked with a fine sales person who had come from his first selling job as an order-taker for standard, off-the shelf and low cost items. Selling products in the high five-figure range was a new experience. We would go out to prospects together, he to lay the groundwork for a presentation that I would do until he was comfortable doing these trips alone. After a few trips together he made the observation that every presentation I did was different and I had new examples of product deployments every time. This was

³ The author commented earlier to users that having an adversarial relationship with a vendor is of no benefit. So, it cuts both ways.

something that I did rather naturally because I was focused on what the prospect was asking for and what they said they needed for their enterprise. My stories were real examples of others with similar needs and how they used the product to address that problem. Talking to customers all the time is the only way to accrue this knowledge and expertise. Just as customers should not expect to just install the product and have it be self-tending, a vendor must keep going back for more conversations to learn how the product is working and being adopted in the enterprise.

Getting One's House in Order – Talking to vendors and other customers of the vendor about their support services is something the author suggests to clients. Be prepared for this conversation with prospects by deploying an internal database for managing support operations, all client information including contacts and their roles, and for tracking every call, no matter how trivial. When a vendor is successful and growing rapidly, this is the most difficult retrofit they will undertake if it's not done in the first place. Early days of deploying and support often depend on developers to engage heavily with clients. This is not bad and each learns great stuff from the other. Bonds are formed but those early communications are also an opportunity to begin building a knowledgebase of:

- Tips and tricks
- Client needs that are candidates for enhancement
- Bugs or design flaws that need attention
- Selling stories
- Logging and monitoring customer experiences for future reference
- Fine tuning the operation and understanding gaps in the vendor/client relationship
- Recommendations or advice given about how to solve a problem
- Error messages
- Etc.

This knowledgebase will form an essential backbone to evolution as a vendor because it will contain data that can be mined and analyzed for decisions about product development, need for better product documentation or help, improvements to installation and tuning, which customers are struggling or not happy, and so on. A knowledgebase can also greatly diminish the need to add significantly to the support team, even in periods of rapid growth. Make it a performance measure for every employee who answers calls that all calls get recorded. Without it, the wheel will be re-invented endlessly and valuable data about operations, growth and performance will never be recovered. Also, customers are very impressed when they call and a support person can state, "The last time you called you wanted to know ...How did our recommendation to ... work out for you?"

Communications – A number of discussions with customers revealed a situation that gets little attention but obviously "bugs" administrators, "surprises." They like to be communicated with by vendors and they like full disclosure. As an analyst, the author receives a lot of communications from vendors for the purpose of maintaining visibility

in the marketplace. Customers want to receive specific information that helps them with their product usage, “how-to” information, sharing of a knowledgebase, known information about problems, who other customers are so they can form communities of practice for mutual sharing of experiences, and so on. We know from the vast amount of marketing and product information that is churned out that vendors have the resources to generate a lot of communications but more information directed at customers for their support would be an improvement welcomed by many.

Licensing and Pricing – There is very little pricing information available to buyers on vendor web sites or in product directories anywhere. A few buyers that author has spoken with recently commented on the awkwardness of having been given early “ball-park” cost figures on which an enterprise budget was prepared, only to learn during final contracting that the numbers were very different. Part of the reason is the complexity of licensing options and models. This gets also to a packaging issue for various platform configurations. The scope of this report is not designed to make specific pricing recommendations although there are quite a number of possibilities for creating simple and straightforward pricing models that work. They may require some technology to enforce and deploy product in a way that is compatible for use. Among our twelve intensive discussions, three users had serious problems because they procured the wrong configuration.

Virtualization and Web-based installations do present new problems not present with client-based software. However, it may be time to consider some new options for pricing based on usage and scale that work fairly across all shapes and sizes of enterprises. This is one area where innovation could create some healthy new business opportunities for some vendors. The bottom line is that few enterprises will get into search products for the first time without a modest price point and a simple option for deployment that does not require months of negotiation between vendor and purchasing. The market will grow much faster with simpler low cost products.

Preparing the User – This goes hand-in-hand with setting expectations in the selling process but needs reinforcing when a new customer is getting ready for installation, and implementation. Another type of communication piece that would be welcomed is a “what to expect” document that gives simple steps for installing, setting up, tuning and FAQs built from other customer experiences. The latter can easily be harvested if a customer support database has been established. The author learned that some vendors begin bulletin boards, but don’t seed or maintain them. They may start customer self-service sites but fail to promote them or guide users to them. Vendor support and access to human beings is very important for new users; make sure customers can get to human help easily and quickly in the early stages. Once they are comfortable with the product and confident that other paths to information will work, extra effort to make self-service at least as well designed as the product will go a long way.

Ideas for Product Improvements – The first recommendation builds on something mentioned earlier; administrators acknowledged that they should be looking at and analyzing log files. This seems like a great opportunity for administrative reports along the lines of text analytics to reveal information about product usage. A second idea is that the fundamentals of old-fashioned command language search (proximity

operators, ranges of dates and integers, and Boolean operators) are there in many products but require the use of drop-down menus that are neither obvious nor intuitive. There is a lot that could be offered in the way of better design for interfaces to these advanced search options. Finally, during internal testing of new releases of your product, you might want to set up some times for testing events when your clients can visit your offices or an off-site location to explore new product developments with your development team present to receive feedback. These times can be structured to give customers the opportunity to network and share ideas with each other. By inviting them to share their findings as a group, you will discover that, in a community, really poor ideas get filtered out but good and creative ideas get exposed and pre-vetted before you invest in an effort that ends up with little buyer interest.

Summary

The enterprise search market is bursting at the seams with innovation and new products. The amount of information gathered from just a few search adopters in a five month period is quite revealing. The author noted similarities in their experiences across a great variety of organizations, but all the people interviewed were learning on their own for the most part. There is no play-book for deploying search products and getting them to work the way any unique enterprise culture wants them to behave. The variables for customizing are endless and the array of products is actually so feature-rich that deployment can be quite overwhelming for first time implementers. There is an enormous amount of experimentation going on, with most customers thinking that everyone else is more advanced than they are.

The marketplace is training experimenters how to procure, install, and implement search. There are studies that point to dissatisfaction with search within the enterprise, which is understandable if those polled are users not actively using a search solution tailored to their enterprise or their business needs. There are still thousands of deployments of search in organizations that were part of a legacy system, installed and left to index some legacy content and never touched again. Somehow, these old search systems never get “turned off” or “turned out.” When the newer systems replace them, lack of education or visibility and hand-holding results in whole populations left with the archives and dusty silos of now useless information. The author believes that this is where many complaints originate.

The committed search champions and administrators the author interviewed reflect a much more positive state of affairs. There are huge opportunities for market growth and market innovation.

If we can move from thinking that enterprise search means one search product for all content, across all repositories, for all employees with a simple search box, and instead get creative in thinking about why we search and all the possibilities for how to get to answers, we will be arriving at a very different vision in the next couple of years.

Vendors have to make it simple and communicate a simple vision; users need to catch up with the technology. There is much to digest and embrace for everyone. Here are two lists, one for buyers and one for sellers to sum up the key points about search for you to know.

Sellers

Market consolidation has been largely among high-end leaders and platform companies acquiring niche products that build on their architecture. The diversity of requirements in enterprises has placed a premium on niche players, so far. Jeff Cutler of Answer.com made a comment in 2007 that in the Web search market a 1% market share is worth \$1

billion. The enterprise market is years away from that type of valuation, but as it evolves there will be much value to spread around.

These are some of the basics of what buyers are seeking:

- Simple cost and licensing model
- Security – authentication through single sign-on
- Ability to search unique content with each niche having special requirements
- Better search interface design, more intuitive
- Good support and expertise that helps them meet their unique mission goals
- Good relevance in results through effective categorization and tuning tools
- Facetted and parametric search
- Clustering
- Scalability and performance
- Easy installation and maintenance
- Interface design options easy to deploy for non-programmers
- Add-ons for ETL, reporting and collaboration, text analytics
- Messaging and guidance to help buyers in their pursuit of enterprise success

The financial thresholds are in these rough categories (what will be spent):

- \$5 – 15K plus services in the \$1 – 2K range
- \$16 – 30K plus services in the \$3 – 25K range
- \$20 – 50K plus services in the \$5 – 50K range

Note that as budget for capital purchase increases, buyers are also willing to spend more for services as a percentage of the license cost. They are probably aware of the need for human resources to maximize product success.

Buyers

Search champions within enterprises are all too aware that ability to move forward with new technology is contingent on many variables. New management, mergers and acquisitions in their own organization or in the marketplace of vendor they are targeting, downturns in business, or sudden elevation of a more critical project that requires IT attention are all common events that derail capital purchasing plans.

Keeping an eye focused on search as a solution to gain efficiency for all employees through self-service is the attitude to adopt for as long as it takes. Being prepared with information about vendors, costs, and new market developments that might be a quick-sell are the ways that a search evangelist has to arm himself for any opening to move toward a purchase.

The previous section highlighted the must-have features and options for buyers in the current market. In addition, buyers will want to approach vendors with the following in their specification:

Enterprise Search Markets and Applications

- List of other applications and tools that the search product must have connectors for and experience in deploying them in the past
- Expectations regarding the installation, what computing and network environment is already in place and available
- A list of all the content expected to be incorporated in search and rough estimates of size and the sequence in which the search index will be built
- Number of potential users, estimated concurrent and actual totals
- A description of people available to work on implementation and deployment to get the vendor's assessment of gaps and to assess where additional consulting support might be necessary
- Any special content retrieval requirements (e.g. proximity search with-in sentence, ability to search mathematical formulas, voice search, metadata management tools)

Keeping these major points in mind, sellers and buyers will continue to capitalize on this growing software market's potential.

Vendor Profile

The three vendors profiled in this section are the companies who sponsored the research underlying this report. They appear in alphabetical order. *Appendix B: Vendor Directory* presents a larger listing of suppliers of search and search-related technologies and services.

Coveo



<http://coveo.com>

Representative Customer Insights

As soon as I saw the system, I knew it was the one that would work for us.

On the Coveo business relationship: Working with them was a refreshing experience after working with hundreds of companies...know how to relate to customers...seem to be focused on our type of company.

History and Competitive Positioning

In the fall of 2004 Copernic spun off its enterprise search product to form a new information access firm, Coveo. Coveo was able to capitalize on its core knowledge of search to launch its total enterprise search offering. The announcement was heralded as a potentially disruptive force in the enterprise market, as noted in this [Traffick.com, Search Engine Enlightenment blog entry](#). Laurent Simoneau, formerly Copernic's COO, became the President and CEO of the new firm, a position he still holds as originally announced in this [Press Release](#). Since their launch, they have released the Coveo Enterprise Search (CES) product up to version 5.1, which was released in early 2008, and have grown to over 600 deployments.

In fact, the successes have brought the company a surge in investments with the interest of a new evangelist investor in 2008, the highly successful businessman Louis Tetu, founder and former leader of Taleo Corporation; he is now Coveo's Executive Chairman. This [article](#) in [Forbes](#) describes the investment and Mr. Tetu's new involvement in the firm. What it does not express is his tremendous energy and the potential he sees for the company, expressed in a phone conversation with Gilbane in March 2008. It was clear that his drive to bring new investment into Coveo reflects his belief that this market will grow enormously, but more importantly a platform with Coveo's capability will scale to handle all the vast and complex content assets of the enterprise. More recently, Coveo has added two members to its board with extensive

entrepreneurial experience, including J. Alberto Yépez, who has held senior roles at Apple and Oracle; and Howard Gwin, who has held senior roles at IBM, Xerox and Peoplesoft.

In March 2008 Coveo announced its comprehensive G2B Information Access Suite, built on the original Coveo Enterprise Search platform; it is designed to leverage intelligent searching across the enterprise, described in the following section.

Descriptions of the Offerings

The Coveo G2B Information Access Suite offers specific packaged solutions that can be linked together for a total view of company information from typically disparate silos and locations. Taken as a whole, the G2B Information Access Suite reflects the company's commitment to and investment in ensuring that existing and new customers have 360° coverage of enterprise content. Built on top of the CES platform, the G2B Information Access Suite has been fleshed out with new optional applications, including Coveo G2B for Intranets, Coveo G2B for Custom Applications, Coveo G2B for Email, Coveo G2B for CRM and Coveo G2B for Multimedia.

Coveo G2B for Intranets demonstrates the value of a collaboration platform by giving workers fast access to what they need from large stores of documents sitting across platforms such as Documentum, FileNet and SharePoint.

Coveo G2B for Email is best understood by visiting the [Coveo Labs page](#) on their Web site and viewing one or both of the demo offerings. This provides both guided scenarios for understanding the full range of applications and the power of Coveo search. It also illustrates the level of complexity that can be achieved by using the faceted navigation for drilling down into search results. Finally, users can construct their own searches and discover for themselves how this add-on will work within their enterprise.

Coveo G2B for CRM delivers additional power and value with connectors to all major CRM applications, including *salesforce.com*. This module simplifies the aggregation of data from *salesforce.com* activity without the need for special report formats.

Finally, **Coveo G2B for Multimedia** brings Coveo's core search power to intelligently discover relevant audio and video content. It uses Coveo's patented audio-to-text converter to render text content for indexing. Any industry with substantive internal multimedia content assets with audio attributes should be considering this option.

For custom deployments of search in the enterprise, Coveo offers the **G2B Custom Applications Toolkit**. Enterprises of all sizes with modest internal technical support have the ability to present a professional search experience.

Strengths

Among the strengths that impress Gilbane beyond Coveo's reputation for quick and easy installation and deployment are the following:

Enterprise Search Markets and Applications

- Pricing is reasonable and easy to understand. It is based on the number of units to be indexed and scales as needed as indexing capacity grows. Customers can begin with a modest license size at a departmental-level and add more capacity as they are ready to extend the reach of search into bigger groups or repositories newly discovered or to content in legacy systems with latent value that are ready for integration with other enterprise content.
- Ease of incorporating legacy database content for indexing by writing macros to extract metadata from locally built databases
- Ease of installation and set-up with minimal technical expertise. Speed of indexing and deploying
- Intuitive interface for users, out-of-the-box
- Availability of faceted navigation without additional programming
- Availability of proximity searching
- Document access permissions are handled well by the security functions
- Scales well without performance impact
- A [Knowledgebase](#), plenty of [Help](#), and a reputation for great customer service including IM and email support
- Nearly 30 Partners to support customization, system integration with other applications and deployment
- An OCR add-on is available

Strategic Advantages

Coveo has focused on where the most unstructured content and transactional content resides in enterprises (e.g. PDFs; Office documents; intranet repositories, including SharePoint; email with and without attachments). They make a point of indexing it well, presenting excellent instantly usable content in results that users recognize. Their navigation model is intuitive and clear. They have understood that search has to be simple from installation to deployment, from both the office and mobile devices.

Coveo is moving quickly to gain market share among the least served enterprise user groups, those who are outside the firewall on a regular basis and those heavily engaged in Web 2.0 tools for collaboration. With the addition of G2B for Intranets, G2B for Email, G2B for CRM and G2B for Multimedia, Coveo is making it very easy for enterprises to see a strong value proposition for search, which might not have been obvious earlier. The email play and linkage with *salesforce.com* will excite road warriors whose access to enterprise content has been so poor that they fail to use search and go without having critical information in a timely way.

Finally, Coveo has stayed close to its customers and works with them to achieve success; that reputation has not diminished, even as they have had rapid growth. Their new investors understand the importance of these relationships and bring with them experience in successfully growing start-ups. The infusion of capital and recognition that they need to keep the feel of a “small company” positions them well.

Futures

Coveo is well capitalized and is putting the capital to work with new technology and marketing. Also, their sensitivity to and understanding of pain points in the marketplace should continue to give them healthy sales. This will be sustained as long as they continue to listen to the marketplace. As Gilbane sees the pressures grow in enterprises for niche and compartmentalized search within divisions, groups and special programs, products at Coveo's price point will be recognized as offering excellent value for solid technology. We expect them to maintain that focus.

Customer Testimonials

In keeping with the number of case studies that already appear on Coveo's [Web site](#) we found no surprises when we reached out independently to a number of their clients. Among the types of comments that came out of lengthy interviews about all aspects of their use of the product:

We had to have a user interface that people would actually use and knew in advance that adoption would be a big deal...it had to be usable when we were in the midst of talking to a client to find answers quickly....when first demonstrated it was a no-brainer for top management.

We put effort into linking financial information with people and their projects to create metadata that would reveal expertise...from this CES was able to deliver facets that would let users narrow searches by client or by the person working on a project or connected to that client.

Permissions were important and Coveo reads active directory permissions, which we just adopted.

On what you should have known before you purchased CES: Nothing I have looked for isn't there...wish I could have gotten it sooner...looked at six other products but immediately settled on Coveo once we saw it.

Needed to leverage the data in the properties fields in Office documents, plus full-text...Coveo did that.

Had worked with another system that people did not use because it just confused them; the Coveo interface was much easier, especially for people who do not search often.

We are getting good feedback from our users...and good value from the product.

In a bake-off between GSA and Coveo, the latter had two things out of the box that we needed...ability to preview large documents with keywords highlighted and the "QuickViews" default with thumbnails.

Corporate Facts

Coveo Canadian Headquarters: 2800 St-Jean-Baptiste, Suite 212 Québec, QC G2E 6J5

Phone: 418-263-1111, Fax: 418-263-1221

Coveo US Headquarters: Riverside Center, 275 Grove Street, Suite 2-400, Newton, MA 02466

Phone: 781-371-0511, Fax: 617-663-4801

Coveo California Office: 120 Hawthorne Ave., Suite 100, Palo Alto, CA 94301

Phone: 650-475-8021, Fax: 650-475-8024

Officers: **Louis Tetu**, Executive Chairman; **Laurent Simoneau**, President and CEO; **Jean Lavigreur**, Chief Financial Officer; **Marc Sanfaçon**, Chief Technology Officer; **Richard Tessier**, Executive Vice President, Products; **Benoit Leclerc**, Executive Vice President, Sales; **James Waters**, Vice President, Global Marketing.

Sales: 1-800-635-5476 (US/Canada toll-free); 00-800-2673-7642 (International toll-free)

Support: For customers using a valid support plan 1-866-266-1583 (US/Canada toll-free), 1-418-266-1583 (Europe)

Employees: ~ 70 worldwide

Pricing: Coveo's pricing is based on indexing 'units' specific to the repository (ie, Salesforce.com, MS Exchange, etc.) from which businesses are looking to access information.

Status: Privately Held

ISYS



<http://isys-search.com>

Representative Customer Insights

When search goes beyond its input/output function, to offer post-query searches and other advanced features, the value it provides is immense...150 hour project reduced to two hours.

ISYS' flexibility and knowledge go far beyond any expectations I had of the company.

History and Competitive Positioning

After 20 years and with its roots clearly established in the business and government sectors, ISYS is a major presence in the enterprise market. The company has more than 14,000 installed customers, the highest number of any company in this market, with platforms from desktops to enterprises on servers supporting as many as 10,000 users.

ISYS was founded and is led by Ian Davies in Sydney, Australia, where he still serves as the chief technology expert. His vision of standalone software for text retrieval was launched on a DOS platform in 1988. ISYS has continued to evolve its technology in response to customer demands and is now operating on Windows servers including Vista, and more recently Linux. ISYS also supports UNIX installations. Product functionality has kept pace with the market, making ISYS a solid choice for enterprise search and an award winning one, as well. Ian's philosophy of search is captured in this [recent interview](#) with Steve Arnold.

ISYS has been a global company since its beginning; with headquarters in Australia, it also maintains offices in Colorado and the UK for Europe. Long positioned for the mid-market, customers including major corporations, government agencies, financial institutions, and law and insurance firms have found that ISYS offers the flexibility to provide excellent retrieval capabilities as well as advanced tools for manipulating retrieved content to derive extra value for their operations.

ISYS has extended the reach of enterprise search to include the same capabilities in products for the desktop, Web intranet portals or public Web sites, and embedded with other applications. Add-on modules leverage search on any of these platforms for email retrieval, integrating internal and external search results, and publishing content with ISYS search for distribution.

Descriptions of the Offerings

With over 14,000 customers representing just about every industry, ISYS Search has a clean line-up of products that won't confuse the buyer.

The first enterprise-class offering is **ISYS: desktop** used to search across individual PCs or entire networks with all the ISYS Search functionality through its own search window for an individual to administer indexing, and to search.

The **ISYS: web** product provides the same ISYS Search functionality in a web-based environment, whether it is an intranet, or internal or external website. It is ideal for public facing websites or portal intranet applications and requires a server environment.

For OEMs **ISYS: sdk 8** for developers is the ideal solution for embedding a small footprint search engine within applications. A list of partners delivering ISYS for

embedded search tells the story: EMC Corporation, TOWER Software, Autodesk, World Software Corporation, Microdec and Konica Minolta.

In addition to these three principal products, three other modules are available to enhance ISYS Search: [ISYS:spider](#) to add Internet sites or internal web sites to ISYS indexes including a Lotus Notes server; [ISYS: email.search](#), a low cost option for standalone searching within mailboxes; and [ISYS: publisher](#), a search tool to deliver with content to be published on a CD.

Follow the product links for more details and take advantage of a trial download to run a proof-of-concept (POC). This link to [technical tips](#) outlines tuning and administrative options for the products.

ISYS has a steady flow of product enhancements and new options. Industry recognitions and their presence at meetings are important to staying informed about their products. To maintain awareness of [new developments](#) or visit with ISYS [at meetings](#), check these pages regularly. The business press has also been keeping an eye on ISYS as evident in articles in [Forbes](#) and [Information Week](#).

Strengths

ISYS' installed base indicates that it delivers capabilities recognized as necessary by enterprises of every size and industry. Strengths are comprehensive, and include everything buyers would expect as a default in enterprise search; here are the more interesting features:

- Indexing supports 200+ file formats [including numerous application specific](#) formats and content in all popular databases
- Indexing supports 60 languages
- Multiple query options: Web-style, ISYS style with Menu-assisted search (command-based with operators for several variations on proximity, field specific for most metadata, exclude terms, Boolean, etc.), and Natural Language query
- Indexing content *in situ* (no requirement to move content to designated servers, file shares, out of applications or databases). Simple setup for pointing to “what” needs to be indexed
- Content for indexing “feeds” option
- Rapid indexing performance with low overhead; extreme scalability
- Highly sophisticated search analytics reporting, *Search Trends*, enables organizations evaluate how ISYS is used and where implementation can be tuned for improvements
- [Forums](#) for customer sharing and [FAQs](#) with good online support interaction with clients
- Ability to establish relevancy based on “source biasing”
- Range searching for numbers, dates, and patterns

- Entity extraction (e.g. people, organizations) to supported faceted search refinements
- Sorting by any metadata field
- Simple adoption model that encourages migration from *ISYS: desktop* to *:web*

Strategic Advantages

ISYS' chief advantage in the marketplace is the simple product models it offers and how well packaged they are for adoption and deployment, both within an enterprise and within applications for deployment by OEMs. OEM partnerships already mentioned with companies like Tower Software (soon to be acquired by HP) and EMC reflect the solid technology that ISYS has to offer. When a non-profit organization as prominent as Columbia University Medical Center (CUMC) makes a strategic decision to use *ISYS: desktop*, there is a good reason. There is a cost benefit to ensuring quick, reliable access to a large collection of regulatory documents and medical studies with rapid implementation. Working with an organization like Columbia at the entry level is a winning strategy when the vendor has scalability and products that will also help users at the next level.

With the largest growth in content management occurring with SharePoint implementations, plenty of legacy Lotus Notes applications, and thousands of professional services firms using Interwoven Worksite, having indexing connectors and simple deployment options for just these three products is good positioning for ISYS. Active directory support secures permission-based access to content. Users only need a single sign-on.

ISYS seems to attract customers with the need for strong content analysis capabilities. Gilbane interviewed several users who selected ISYS because they had very specific content “slicing and dicing” requirements. Even Steve Arnold, who has worked with and evaluated hundreds of search engines, has put ISYS to good use for patent analysis as described in his recent [blog posting](#). Also, take a look at this summary of [analytics reporting](#), a support feature that comes with ISYS.

Finally, by offering an embedded search option, ISYS has good penetration where search needs to be included to leverage the value of content in various distribution models. It also gives ISYS added exposure; although it might not be obvious, users who like the way it works will track down the technology. Between embedded search and the number of [ISYS powered public sites](#) Web sites, there are plenty of ways to become familiar with the great variety of deployment options.

Futures

ISYS has a solid track record that few enterprise search vendors can match; given the rise in visibility of search, largely driven by Internet search, the company has even more potential going forward. With a solid support operation already in place for a product that seems not to require a huge amount of support, they can stand to grow at a faster pace. Of course, this can be a mixed blessing for the marketplace because great

successes don't often come without risk. This is a company that has clearly benefited from excellent algorithmic design and vision by a founder who is still in charge.

Gilbane is encouraged to read that others in management at ISYS are sustaining that core vision for how search should work and needs to be implemented. They come across as pretty pragmatic and solid people, as reflected in this title for a talk by Terry Clift, General Manager of Global Operations: *Forget "One Size Fits All," Search is an Iterative Process*. That's typical of the kind of messaging ISYS uses to demonstrate that they know how to sell it straight up. Iteration and on-going "care and feeding" are a theme Gilbane stands behind, as well.

Customer Testimonials

Commentary from Gilbane is not necessary to be convinced that ISYS has good stories to back up the reasons for their very broad deployment. Numerous [case studies](#) are shared on their Web site. By comparison, it is rare to find so many customers willing to share their experiences in a public forum, so this can be viewed as a net positive.

Gilbane first experienced a very enthusiastic ISYS client at its 2007 Boston Conference, where he was presenting. His willingness to fly in for the day from halfway across the country to talk about his unique application was enough to get our attention. Further investigation and talking with clients brought the following comments. Some of the applications are rooted in ISYS strength for exposing content for extraction and deeper analysis.

Needed to be able to define data for a pre-determined set of companies, then sift through that data to garner specific bits of information, extract it from over a million documents manually...had to be a better way. This was the basis for selecting ISYS.

None of the competitors were able to offer sophisticated search combined with the ability to process information on a post-query basis...In seconds ISYS found more than 7,500 potentially relevant documents out of more than 6,000,000 in the set...manually processing these search returns is the most tedious and mistake-prone part of any research project. "the magic of ISYS really makes a difference"

Selected product for strength of proximity search, speed of search and indexing and ability to integrate it into our own custom software...will be transparent to users that they are using ISYS.

Working with ISYS was a pleasure...The helpful support and visionary guidance ISYS provided was an essential ingredient in creating a search solution that worked with my particular situation.

Have called on support and they have been very responsive.

Tuning and implementation have been good...built-in utilities work well.

We chose ISYS after seeing it met our needs in a very cost-effective manner...We were able to fine-tune the product to make it work just as needed. We are now considering

Enterprise Search Markets and Applications

further uses of ISYS technology to help us streamline our archival (needs) and retrieval of documents.

Corporate Facts

ISYS Search Software is headquartered in Sydney, Australia, and maintains offices in Denver, Colorado, and Chester, England. Contact the nearest office for any sales inquiries, technical support needs or general questions.

The Americas: ISYS Search Software, 8765 E Orchard Rd #702, Englewood CO 80111, USA

Phone: +1 303 689 9998, Fax: +1 303 689 9997, Toll Free: +1 800 992 4797

Sales: info-us@isys-search.com

Australia & Asia Pacific: ISYS Search Software, PO Box 1312, Crows Nest NSW 1585 Australia

Phone: +61 2 9439 5800, Fax: +61 2 9439 8569

Sales: info-au@isys-search.com

EMEA: ISYS Search Software, The Steam Mill, Steam Mill Street, Chester CH3 5AN UK

Phone: +44 (0) 1244 313216, Fax: +44 (0) 1244 313003

Sales: info-uk@isys-search.com

Officers: **Ian Davies**, Founder and Managing Director; **Terry Clift**, General Manager, Global Operations; **Derek Murphy**, President, The Americas; **Michael Blaszak**, Vice President, Sales; **Dave Haucke**, Vice President, Global Marketing; **Nigel Wheeldon**, Group Sales Manager, EMEA.

Employees: ~ 50 Pricing: Desktop \$1,000 and up; Web \$15K and up (not document-based)

Status: Privately Held

Vivisimo



<http://vivisimo.com>

Representative Customer Insights

Once we deployed Vivisimo we went from 50 to 1,500 searches per day (on our intranet).

[We] selected Velocity because it could display search results the way we wanted to see them. We were seeking a single application for touching all content that researchers need to do their work with the requisite authority.

History and Competitive Positioning

Vivisimo comes by its search innovations quite purely, straight from its academic roots at Carnegie Mellon where its co-founders and leaders established their vision. While collaborating in 1998 on research to make large lists of search results more understandable to users, Raul Valdes-Perez, a computer science professor with a background in artificial intelligence, and Jerome Pesenti, also a computer scientist, began their professional relationship. In 2000 Vivisimo was launched with just a handful of people, many of them Carnegie Mellon students and graduates.

In reviewing a history of when various companies began to appear in lists of enterprise search products, it was interesting to see that Vivisimo did not make a name for itself until 2005. But when it did, it was viewed as visionary. This reflected awareness that its high-performance indexing and scaling attracted a significant customer, the federal government's Government Services Administration. Vivisimo's search engine was selected to replace the original system for FirstGov.Gov, Fast Search & Transfer, which could not keep up with the growing number of documents online. As well, the old system cost \$3.2 million a year to support only eight million documents. Vivisimo's Velocity, paired with MSN Search, supports 40 million documents at an annual cost of \$1.8 million. The system has since been renamed USA.gov.

It didn't take Vivisimo long to be cited as "Best Enterprise Search" in 2006 by InfoWorld testers. Since then the National Library of Medicine (NLM) selected Vivisimo for MedlinePlus.gov. As a top government news and information site for access to over 700 topics in English and Spanish, it provides links to information from 16 million articles from over 5,000 journals. It has garnered the top rating by both the English and Spanish speaking public among government sites. For decades NLM has been an innovator in search technology development and adoption, so it is a positive statement that they selected Vivisimo.

But it is Vivisimo's Velocity is driving its current success, delivering *search done right* (their tag line) to the enterprise with high performance, scalability, a security authentication model that respects existing enterprise security framework, and attention to adaptive usability. To their list of numerous government clients in the U.S. and globally, they have added significant corporate and professional services firms. Their customer base grew to over 150 by the end of 2007, and they are well positioned to beat the leaders on technology for some time to come. Vivisimo has technology that continues to evolve rapidly. The company has raised its first round of major funding (\$4 million) to ramp marketing and sales.

Descriptions of the Offerings

Vivisimo has two product lines, the Velocity Search Platform for the enterprise and Clusty for metasearch used by educational institutions and non-profits. This review focuses on Velocity, although much of the underlying technology is the same. Since its inception, Vivisimo has been noted for its advanced federated and clustering capabilities that act as a transformational layer to federate search results and present them in semantically grouped topics to help users easily narrow the focus of a search. This engages users in a simple mechanism of text mining in which they can refine their view of the results. Because the search engine in Velocity has been optimized to index very large domains of various levels of complexity, it is ideal for managing enterprises in which the content is highly diverse in topical scope and type. To this, Vivisimo has added a security model that does not force a particular authentication architecture on an enterprise but allows the IT department to work with Vivisimo within its own model.

This is all explained in a self-guided [product overview](#); it is highly recommended. Throughout the Vivisimo site there are numerous demonstration options to "test-drive" public implementations. These illustrate the flexibility of the product and diversity of deployments and will give the reader a good feel for the product.

Unlike any other vendor deploying large-scale, high-end search, the cost effectiveness of Velocity is obvious when considering the ease of deployment, powerful but intuitive administrative tools requiring low human overhead, and rapid indexing. All of this ensures implementations in a [matter of weeks](#), rather than months or years. The product is priced and designed to bring value to the enterprise quickly.

To better understand [Velocity Enterprise Search](#), review the options for various deployments: Intranet, Corporate Web site, or Special Applications.

Vivisimo has three areas of focus for the **government**: Citizen Web (e.g. USA.gov, MedlinePlus.gov), Government Intranet Search for use within an agency, and search for the Intelligence sector.

[Velocity for Mobile](#) is really an extension of Velocity for the Enterprise because it provides simplicity of access to the Vivisimo experience within a mobile device as easily as through the desktop or a laptop. Security to enterprise content is provided through a VPN connection just as a remote user on a laptop might operate.

Finally, Vivisimo's OEM partners are delivering Vivisimo search technology in their applications for special audiences. Of particular note is the recent announcement that Interwoven, with its large presence in the professional services market, will resell Velocity Search Platform and has also embedded that capability into the Interwoven Worksite product, replacing Autonomy's Verity.

Strengths

Velocity is efficient in every sense, reflecting that Vivisimo has learned about what users do not embrace in enterprise search. Instead of replicating old technologies in new packaging, they have taken a fresh approach that shows in their product usability, a look and feel in keeping with Web 2.0 tools, and attitude about what drives buyers to enterprise search. They consistently work for understanding. Here are some non-standard features and strengths:

- Clustering technology does not require customer pre-defined taxonomies but supports enhancing terminology and language through simple administrative tools to add synonyms, tune relevance, etc.
- A "remix" feature enhances clustering. When the clusters are not sufficiently granular or meaningful for a user to drill down in a pile of results, the *remix* button effectively re-clusters by exploiting the content to discover more terminology that might put content into other folders. This [article](#) summarizes the feature, which works in Clusty.com and Velocity.
- Personalization at multiple levels: by search administrator who can establish role-based result sets for individuals or groups, relevancy settings and navigation options at user or group level, and by users who need to be alerted when new content is indexed that matches a pre-defined query
- Collaboration options that leverage query strategies and query results among members of a practice area or project group so that members or leaders can save, share or export queries or the results. This option is designed to support social tools, such as tagging and annotation.
- Social networking by clustering individuals with their content. Together with the collaboration tools, this enhances the Web 2.0 look and feel.
- Support for over 80 languages
- Pre-built connectors for: EMC Documentum, MS Exchange, Lotus Notes, SharePoint, Oracle DB, Salesforce CRM, EMC Legato Email Extender, Symantec Enterprise Vault and numerous other proprietary formats
- Support for federation of internal content search results with external search results from subscribed services, third-party feeds or Z39.50
- High volume search performance in the order of 1,000 queries per second on a 62-bit processor, with a small index footprint 8-20% of original full-text data size
- Rapid deployment in very large enterprises (e.g. three months for USA.gov)
- Intuitive administrator interface and tools with no need for APIs

Strategic Advantages

Vivisimo's interface in the enterprise has received this reaction among searchers: "It's easy, intuitive and powerful." Adopters seem to quickly "get it" when it is deployed; whether to the public or among a more sophisticated corporate population, search usage ramps dramatically. With Google Search Appliance (GSA) often procured as a stop-gap measure to fill a need for enterprise search, users soon discover that what they liked on the Web does not automatically translate to a good experience internally. With Fast and Autonomy looking stale and Endeca's cost out of reach for many, Vivisimo could be the solution to beat.

Vivisimo has already demonstrated ability to gain a foothold in enterprises with a need for practicality and expedience—government agencies. Content in the government tends to be voluminous and complex, and IT departments are bare-bones and over worked. Since Vivisimo works well and economically for the largest search operations of the government, it has already passed a critical test. Major corporate clients have also had successful deployments that they are happy to talk about: Procter & Gamble, Tyco Electronics and Organon, a division of Schering Plough, are just a few.

Taking a view of the multi-faceted human aspects of search is novel and welcomed in the 2.0 era. Gilbane was impressed in several presentations to see the extent to which Vivisimo uses its own tools to work collaboratively across all the content in the company, sharing ideas, content, search strategies and easily locating expertise within the company using Velocity. This was true of all employees we met in several situations. When a company enthusiastically embraces its own technology and readily demonstrates work benefits "at the drop of a hat," it gives us confidence that good stuff will continue to appear in the product, and when it appears it will be well tested.

The [lineup of OEM re-sellers](#) is solid. These gateways into new accounts will only give Velocity more corporate exposure. It is a tribute to the technology that these vendors have adopted it because they definitely will not put up with technology that breaks.

Futures

Test drives of the Velocity product inevitably resulted in some questions and suggestions about how to make it better; we were encouraged at how receptive and interested our demonstration coaches were. Already, we are seeing how seriously Vivisimo listens to its audience with the *remix* feature and mobile search. We expect to see this innovation continue as Vivisimo stays in step with demanding clients.

In early 2008, Vivisimo hired a VP for Strategic Alliances, a new position. This demonstrates a commitment to how they are expanding the business. With the investment funding and new partners, marketing and sales are going to be priorities.

Fast and Microsoft have just launched the enterprise search division to build a new platform beginning with older architectures. Vivisimo began with new technology and new customers whose expectations they met quickly. They are not saddled with legacy clients and old search models that may be difficult to migrate into a new platform. As Vivisimo's more established competition works to bring its development operations,

Enterprise Search Markets and Applications

vision and customers into focus, Vivisimo has a head start at the high-end with products that truly look like *search done right*.

Customer Testimonials

Customer comments are widely published in the press from government users, and from public presentations. This [article](#) affirms the basis for adopting Vivisimo and MSN search to replace Fast. Other comments reinforce customer confidence in the product and company:

In head to head proof of concept between Fast, Autonomy and Vivisimo, Vivisimo outperformed on every metric. Autonomy could not handle single-sign-on at all.

A test of indexing 700,000 documents took six weeks with Autonomy and two days with Vivisimo. . . the semantic clustering was a bonus feature.

We were already considering Vivisimo but when Express Search with Interwoven was announced, we fast-tracked our investigation.

Corporate Facts

Vivisimo, Inc. Headquarters 1710 Murray Avenue, Pittsburgh, PA 15217

Phone: +1-412-422-2499, Fax: +1-412-422-2495

Vivisimo European Offices

Paris office: Phone: +33-066-324-4792, Fax: +33-014-535-3061

London office: Phone: +44-208-940-8773

Employees: 100 Privately held

Officers: **Raul Valdes-Perez**, Chief Executive Officer; **Jerome Pesenti**, and Chief Scientist; **Christopher Palmer**, Chief Technology Officer; **Dennis Brestensky**, Vice President of Operations; **John Dean**, Chief Financial Officer; **Jeff Hornung**, Vice President of Strategic Alliances; **Joe LeBas**, Vice President of Sales, **Rebecca Thompson**, Vice President of Marketing.

Appendix A: Enterprise Search Interview Lines of Questioning

The user experiences described earlier in this report were discovered through direct lines of questioning by the author. In some cases information was also gathered from presentations made by individuals involved in search technology for their organizations in webinars or at meetings, and the Q&A sessions that followed. Information in white papers, blogs, and articles all informed the author's knowledge of behaviors and experiences in the marketplace.

This appendix summarizes the types of information sought during the research and the sequence for questioning. It is intended to stimulate buyer thinking about how to frame needs, intent, and process of selecting a search product. Search vendors may gain a new appreciation for how much a buyer needs to know before making a procurement, and all the work that lies ahead to gain value from their ultimate procurement. Building healthy relationships and mutual respect between buyers and sellers is one goal of this study.

Demographics/Organization Being Described

Industries served (vertical)

Size of Organization where search product was deployed

Scope of the collections being indexed and searched

Size of collections

Entire corporation, division, business unit

Demographics/Interviewee

Formal Background or training

Work experience

Previous work with system

Contact information

Demographics/Product (discussed in the interview)

Name of Product

License Size and configuration

Modules

Versions

Selection Criteria

Problem definition (or opportunity creation)

Cost

Technical constraints/Requirements

Selection Process

How many products

Basis for first cut

Narrowing

Vendor business relationship

Implementation/Deployment

Platform requirements

Ease of installation

Architecture requirements for user access

Tuning/Administration

Expertise required to perform tuning/administration

Length of time to become familiar with product administration options

What options do you like best/find most useful

What problems have you encountered with administration operations?

Have you been able to achieve the tuning and feature enablement that you needed/expected?

What should the vendor consider adding to make the tuning/admin function better/easier?

Usability/Adoption

How long to install and activate

How did you go about introducing the search engine?

What were the early experiences?

Where did you need to make changes: deployment or training?

Enterprise Search Markets and Applications

Was there a plan for “rolling it out”?

How well did the plan work out

What were the most positive outcomes (reports of successful use)?

How well was the product received?

What would you do differently?

What are your users reporting about their experiences with the product?

Other Roles

What skills other than your own are needed for:

- Selection
- Implementation/Deployment
- Tuning/Administration

Business Relationship with Search Vendor

How would you describe the sales/selection process with the vendor of the product you selected?

How would you describe the support you have received from the vendor?

What would you have valued in terms of support or business relationship that you didn't experience?

What were the most positive aspects of your vendor experience?

What Should you Have known?

Before selecting a product

Before implementing a product

Before rolling the product out to users

Disappointments

What has been your biggest disappointment?

Where does the problem lie?

Is there any hope for remediation?

Labor intensity of support

For your size of organization, is the product selected appropriate in terms of support required

Accuracy and Scope of retrieval

Describe any issues with the way the product retrieves content and returns results

Is there content that should be retrieved and isn't

Describe any security related issues that have been or will be a concern

Scalability/Performance

What issues have you encountered concerning disk storage requirements?

What issues have you encountered concerning other hardware requirements?

How well is the product performing in the area of speed on retrieval and indexing updates?

Appendix B: Vendor Directory

Companies principally focused on enterprise search are in bold

Company Name	URL
Abrevity	http://www.abrevity.com/
Access Innovations	http://www.accessinn.com/
Adobe	http://www.adobe.com/products/lifecycle/?ogn=EN_US-gntray_prod_livecycle_home
AltaVista	http://www.altavista.com/ see also: Fast
ATG	http://www.atg.com/en/products/commerce_search.jhtml
Attensity	http://www.attensity.com/
Attivio	http://www.attivio.com/
Autonomy	http://www.autonomy.com/content/home/index.en.html
BA-Insight	http://www.ba-insight.net/products.html
Basis Technology	http://www.basistech.com/
Baynote	http://www.baynote.com
Blossom Software	http://www.blossom.com/index.html
Brainware	http://www.brainware.com/
BRS/Search	SEE: OpenText
Business Objects	http://www.businessobjects.com/product/information_discovery/ , See also: SAP
Clarabridge	http://www.clarabridge.com/Products/BISearch/tabid/106/Default.aspx
Clusty	SEE: Vivisimo
COGITO	SEE: Expert System
Collanos	http://www.collanos.com/
Collarity	http://www.collarity.com/
Concept Searching	http://www.conceptsearching.com/web/
Connotate	http://www.connotate.com/
Convera	http://www.convera.com/
Coveo	http://www.coveo.com/en/default.aspx
Cuadra Associates	http://www.cuadra.com/

Enterprise Search Markets and Applications

Company Name	URL
Data Harmony	SEE: Access Innovations
Deki Wiki	SEE: MindTouch
Dieselpoint	http://www.dieselpoint.com/featurematrix.html
Documentum	SEE: EMC
dtSearch	http://www.dtsearch.com/
EasyAsk	SEE: Progress Software
EMC	http://www.emc.com/products/detail/software/eci-services.htm
Endeca	http://endeca.com/
Engenium	SEE: Kroll Ontrack
Exalead	http://corporate.exalead.com/enterprise/l=en
Excalibur	SEE: Convera
Expert System	http://www.expertsystem.net/?lang=1
Eyealike	http://www.eyealike.com/index.php
Fast ESP	SEE: Fast Search & Transfer
Fast Search & Transfer	http://www.fastsearch.com/
Funnelback	http://funnelback.com/
Google	http://www.google.com/enterprise/intranet_search.html#utm_medium=et&utm_source=us-en-et-bizsol-0-finderB-all&utm_campaign=en
Grokker	http://www.grokker.com/
Hummingbird Search Server	SEE: OpenText
IBM	http://www-306.ibm.com/software/data/enterprise-search/
IDOL	SEE: Autonomy
Image-Seeker	SEE: LTU Technologies
Index Engines	http://www.indexengines.com/
Information Builders	http://www.informationbuilders.com/products/webfocus/index.html
Inmagic	http://www.inmagic.com/index.html
InQuira	http://www.inquiria.com/
Instranet	http://www.instranet.com/index.asp
IntelliSearch	http://www.intellisearch.no/Solutions/

Enterprise Search Markets and Applications

Company Name	URL
InXight	SEE: Business Objects
iPhrase	SEE: IBM
ISYS	http://www.isys-search.com/
IXIASOFT	http://www.ixiasoft.com/default.asp?xml=/xmldocs/webpages/webpage-profile.xml&section=1#
K2 Enterprise	SEE: Autonomy
Kaidara	http://www.kaidara.com/
Knova	http://www.knova.com/
Kroll Ontrack	http://www.engeniumsearch.com/
Legato	SEE: EMC
Lexalytics	http://www.lexalytics.com/index.php
Liberty IMS	http://www.libertyims.com/index.html
LiveLink	SEE: OpenText
Longitude	SEE: BA-Insight
LTU Technologies	http://www.ltutech.com/en/
Lucene	http://lucene.apache.org/java/docs/
Luxid	SEE: Temis
MarkLogic	http://www.marklogic.com/
Mercado	http://www.mercado.com/
Microsoft	http://www.microsoft.com/enterprisesearch/
MindServer	SEE: Recommind
MindTouch	http://wiki.mindtouch.com/
MondoSoft	SEE: SurfRay
MultiTees	http://www.multitees.com/
MuseGlobal	http://www.museglobal.com/
Nervana	http://www.nervana.com/
NetWeaver	SEE: SAP
NorthernLight	http://www.northernlight.com/
nStein	http://www.nstein.com/
Olive Software	http://www.olivesoftware.com/
OmniFind	SEE: IBM
Ontolica	SEE: SurfRay

Enterprise Search Markets and Applications

Company Name	URL
Ontrack Engenium	SEE: Kroll Ontrack
OpenText	http://www.opentext.com/2/sol-products/sol-pro-knowledge-management/pro-ll-federated-query-server.htm
Oracle	http://www.oracle.com/database/secure-enterprise-search.html
Paglo	http://paglo.com/
PicoSearch	http://www.picosearch.com/
Polyspot	http://www.polyspot.com/Home.aspx
Powerset	http://www.powerset.com/about
Presto	SEE: Inmagic
Progress Software	http://www.progress.com/index.ssp
QL2	http://www.ql2.com/
Recommind	http://www.recommind.com/
Rosette Linguistics Platform	SEE: Basis Technology
SAIC	http://www.saic.com/products/software/teratext/products/
SAP	http://www.sap.com/usa/solutions/informationworkers/enterprisearch/index.epx
SAS	http://www.sas.com/technologies/analytics/datamining/
Schemalogic	http://www.schemalogic.com/
Seaglex	http://www.seaglex.com/
SearchBlox	http://www.searchblox.com/
Semantra	http://www.semantra.com/
Siderean	http://www.siderean.com/
Silobreaker	http://www.silobreaker.com/
Sinequa	http://www.sinequa.com/index.html
SLI Systems	http://www.sli-systems.com/
STAR	SEE: Cuadra Associates
SurfRay	http://www.surfray.com
tazti	SEE: VoiceTech Group
Techrigy	http://www.techrigy.com/
Temis	www.temis.com/
Terabase	http://www.terabase.com/

Enterprise Search Markets and Applications

Company Name	URL
Teragram	SEE: SAS
TeraText	SEE: SAIC
Texis	SEE: Thunderstone
TEXTML	SEE: IXIASOFT
TextWorks	SEE: Inmagic
Thunderstone	http://www.thunderstone.com/texis/site/pages
TREX	SEE: SAP
UltraSeek	SEE: Autonomy
Velocity	SEE: Vivisimo
Verity	SEE: Autonomy
Vivisimo	http://vivisimo.com/
VoiceTech Group	http://www.voicetechgroup.com/
Vorsite	http://www.vorsite.com/Default.aspx
WAND	http://www.wand.com/core/AboutUs.aspx
WebFOCUS	SEE: Information Builders
WebSphere	SEE: IBM
Wikia	http://search.wikia.com/wiki/Search_Wikia
WordMap	http://www.wordmap.com/
X1	http://www.wordmap.com/
Xerox PARC	SEE: Powerset
ZyLAB	http://www.zylab.com/

Appendix C: Glossary

Term	Definition	Synonym
Aggregation	Activity for forming distinct sets of content	
Analytics	Data that helps track business trends Records that describe part of a larger domain Sophisticated version of data	Text analytics
Application programming interface	Vendor supplied add-on software tools to facilitate programming new features or functional enhancements to integrate a software product with other applications	API
Associative structures	Data models showing linkages among different types of data records (e.g. customers to transactions)	
Authority	Validating entity	
Authority control	Methods and lists employed for validating terminology and other content normalizing values in data maintenance	
Auto-categorization		See Categorization
Boolean searching	Use of explicit commands to limit or narrow the scope of a search (AND), expand its scope (OR), or exclude explicit content (NOT). e.g. search for content limited to containing both "energy" AND "solar", where AND is the command.	
Business analytics	Technique to visualize and analyze business data to support decision making	

Term	Definition	Synonym
Business intelligence	Technologies that gather, store, analyze and make accessible data to help enterprise users make better business decisions. It includes decision support, query and reporting, online analytical processing, statistical analysis, forecasting and data mining. Enhancing data into information and then into knowledge. Traditionally focused on extracting and manipulating data from structured databases including numeric data. Viewed by some as the umbrella for other technologies including text mining and analytics.	BI
Categorization	A computational or human activity assigning labels to sets of content to explicitly aggregate by label	
Citation	Information that accurately defines and describes a publication or data file	RT Results
Clustering	Process for gathering unstructured content into a common space for the purpose of grouping it with content on the same topic	RT: Aggregation Categorization
Collaboration	Describing shareable processes and/or content within an application.	See also: Social search
Concept search	Retrieval of content through automated means that take contextual information, not just key words, into account when determining the relevancy of the content.	See also: Semantic search
Connectors	Software tools supplied by search vendors or built internally to support data exploitation by a search engine.	Adapters
Content	The target of search regardless of format or medium. Everything included in a collection of files	

Enterprise Search Markets and Applications

Term	Definition	Synonym
Content intelligence	Generic term for a collection of technologies that automate the process of interpreting content analytically and present the results in a structured format.	See also Business intelligence
Context	Surrounding content that elucidates and clarified a set of data	
Controlled vocabulary	Terminology from approved lists used for tagging content	
Crawling/Spidering	Computer programs, usually part of a search engine, that traverse a specified set of domains for the purpose of indexing all content encountered	
Cross reference	Information that guides to another piece of content. In a controlled vocabulary a term pointing to another term for required or alternative usage in indexing and for prompting during a search dialogue.	
Data aggregation	Inclusion or clustering models for heterogeneous data sets	
Data federation	Organized data state formed by merging and normalizing a collection of similar data objects	
Data mining	Computerized process for extracting content from structured repositories	Data mining . See also Text Mining
Data normalization	Standardization of identical data elements (reducing fields to the simplest meaningful or workable structure). Applied consistency.	
Data warehouse	A central repository or information infrastructure that stores or logically connects a collection of databases and associated content with characteristics and controls that enable sharing and federated retrieval.	
Database	Repository of data organized by explicit records and fields, or tables, rows and attributes	

Enterprise Search Markets and Applications

Term	Definition	Synonym
Digital asset management	A type of content management that automates the application of rigorous governance rules for how the content is created, modified, and maintained with access controls.	
Domain	A corpus of content bounded by system architecture definitions.	
Dublin core	A standard 15-element metadata element set maintained at http://dublincore.org/ as a baseline for content.	See also Metadata
Embedded search	Retrieval algorithms delivered as a part of a software application for searching the content within the application.	
Enterprise search	Software used to index and retrieve content that exists within an organization, ideally optimized for specific enterprise business requirements.	
Entity extraction	A process of content analysis by which the software identifies and classifies data by type or attribute for the purpose of creating metadata from unstructured content.	
ETL	Extract, load and transform suite of algorithms or programs	
Extractors	Software programs that harvest data content from databases, files or other applications, usually for the purpose of then manipulating the data for eventual exposure to other applications or search engines.	
Faceted navigation	In a search interface, the exposure of a controlled terminology list with facets (classes of concepts) with drill-down (broader to narrower) capabilities to facilitate moving through the facets to obtain different groups of content results.	

Term	Definition	Synonym
Federated search	Process of retrieving content either serially or concurrently from multiple targeted sources that are indexed separately and presenting results in a unified display.	See also Federation
Federation	Expansion of the concept of aggregation. It has play in a multi-domain environment (internal sites or a mix of internal and external). Across domains it supports at least four distinct functions: Integration of the results from a number of targeted searchable domains, each with its own search engine Disambiguation of content results when similar but non-identical pieces of content might be included Normalization of search results so that content from different domains is presented similarly Consolidation of the search operation (standardizing a query to each of the target search engines) and standardizing the results so they appear to be coming from a single search operation	
Filtering	Applying other search criteria to narrow or alter the results of an existing search or stored search strategy.	
Full text search OR “free” text	Retrieval of strings found within the full content of a collection of files	Full text retrieval; free text search; unstructured search
Fuzzy search	Content retrieval algorithms that have rules for what content is relevant to match a query. (e.g. finding all words that are alternative grammatical forms of elevate or mean the same thing as elevate.)	
Histogram	Frequency distribution display or model – visualization presentation	

Enterprise Search Markets and Applications

Term	Definition	Synonym
Hosted search	Retrieval software is installed and supported on computing infrastructure that is maintained by the vendor; search algorithms operate from that host on user-controlled content, which may or may not reside permanently on the host.	
Index	Systematically arranged list; in computerized systems it is a representation of content to speed retrieval by the governing algorithms.	
Indexing	A human intellectual process for organizing content to optimize retrieval. A computerized process for organizing content to optimize retrieval	
Integrated information system	Connected data structures and workflow procedures with common features supporting a unified architecture and operational method.	
Interface (Search)	The architecture controlling the methods and design through which a user executes a search.	
Keyword	Non-controlled terminology; language extracted from the content literally	
Keyword search	Query request for literal text as crawled and indexed by a search engine	
Knowledgebase	A domain specific data repository of facts or rules accessible in machine readable format to support software applications	
Link	URL address explicitly connecting content in one location to content in another (my be within a document, site, or remote)	Hyperlink
Loaders	Software applications designed to transfer data from one database to another often couple with transformers	

Term	Definition	Synonym
Metadata	Explicitly defined labels for structuring content that describes any document or file regardless of the native format.	Citation, Properties, Bibliography
Natural language query	Search expressed as a question by a native speaker who asks for information	
Navigation	Method of traversing content with a device (e.g. mouse), or accelerator keys through a structured layer of content to reach other content (e.g. drilling down through a taxonomic structure)	
OEM	Original equipment manufacturer; used to explain relationship of a supplier to another organization whose product is embedded in the delivered application.	
Ontology	An assembly of concepts in which all possible relationship that might exist between and among concepts is explicitly mapped	
Open source search engine	Retrieval software available without licensing costs and customizable by the acquiring organization. E.g. Lucene	
Parametric search	Interface architecture supporting the selection of multiple variable criteria in a single search pass. e.g. to find all products within a class, with specific properties, and applied to select industries.	
Personalization	Self management of the software application's interface	
Phrase search	Retrieval query specifying explicit adjacency of two or more terms in the order expressed in the query.	
Portal	Web-based page of links serving as points of entry to specific content, other web sites, and applications.	
Prompt	Interface symbol or text indicating that a user response is required to proceed with the transaction	

Enterprise Search Markets and Applications

Term	Definition	Synonym
Repository	A database or file structure for electronic content	
Results	The citations or partial content of data retrieved in a search	
Retrieval	Process of accessing content through the act of searching	
Search	Process classification for all software designed to retrieve content.	
Search appliance	Hardware bundled with search software designed to be plugged into an existing computer infrastructure (e.g. network) to begin the process of crawling and indexing target content within the network.	
Search engine	Software with algorithms specifying how data is to be retrieved from one or more indices.	
Search intermediary	Individual who interprets what a user wants to find and performs retrieval operations on behalf of the user.	
Search platform	Suite of software products that together enhance simple index searching with additional functions related to content (e.g. transformation, analysis, reporting)	
Searching	Using retrieval software or a non-automated process for finding content	
Security	In a search environment, the search engine functions that support access controls to content through authorization validation.	
Semantic search	Use of natural language or meaningful queries to find content through retrieval software designed to understand linguistically meaningful questions and the target content.	
Site search	Option using navigation or a search box to retrieve only content from a specific Web site domain.	

Enterprise Search Markets and Applications

Term	Definition	Synonym
Social search	Option within a search interface environment to share and annotate search results using collaborative features.	
Sort	Arrange or order data in a defined sequence.	
Stemming	A form of fuzzy search logic that reduces a word to its fundamental root and looks for any word with that root. (e.g. a search for stemming would also retrieve stem, stems, and stemmed)	
Structured content	Data stored in a database or explicit metadata stored in a software application	
Structured search	Use of pre-defined forms or explicit commands to give bounds to query criteria and parameters. (e.g. restricting the search for a word to the title field)	
Tag and tagging	Use for semantic labels or functional tagging that indicates the purpose of a topic or conceptual string. Different that cataloging in which metadata values are being assembled congruently to the content. Tags usually reside embedded in the content.	
Taxonomy	Hierarchically ordered list of terminology approved for tagging or categorizing a corpus of content. Also, often exposed in the search interface to form the framework for navigated search.	
Text mining	Extracting interesting and non-trivial information and knowledge from unstructured text. Interdisciplinary field that draws upon: Information retrieval Data mining Machine learning Statistics Fact extraction Computational linguistics	Information extraction, Intelligent text analysis, Text data mining, Knowledge discovery in text

Term	Definition	Synonym
Thesaurus	A list of terms that are assigned simple relationships, cross references, scope notes, usage notes and other directives. A thesaurus is often more comprehensive than a taxonomy but less complex than an ontology.	
Trackback	A URL from one piece of content to the URL of another.	
Transformers	In data and content management, tools to normalize or otherwise systematically change data.	
Unstructured content	Content not organized in a formal structure; files not in a database (e.g. a Word document)	
Visualization	Graphical or image representation of data to reflect some understood relationships that reflect information or reveal knowledge about the data.	
Web search	Retrieval from a domain of content exposed to a single or multiple Web sites.	
XML	Acronym for eXtensible Markup Language. An infinity customizable markup language for defining the metatags, descriptions of kinds of content within or applied to a domain of content.	

Appendix D: Bibliography

2007 was such an active year for the topic of enterprise search that hundreds of articles could have been included in this brief bibliography. Almost half way through 2008, as noted earlier in this report, debate about whether or not we are in a post enterprise search era has surfaced. Needless to say, much is still being written about the subject and will for as long as it takes to resolve debate about all classes of software designed for the “enterprise,” maybe decades. The following articles are a good starting place to round out an understanding of the search marketplace, and five blogs are presented. Readers will find links to dozens of other resources if they consistently peruse the blogs. The articles will send them in other directions.

Angel, Gary. [***Web Measurement and Analysis for Internal Search***](#). DM Direct, September 2007

DuPont, Ben. [***Enterprise Search: Seek and Maybe You'll Find***](#). Intelligent Enterprise, June, 2007

Kho, Nancy Davis. [***Smart search: business intelligence and search converge***](#). eContent Magazine, Nov. 1, 2007. pp. 50 – 57.

Lamont, Judith. [***Search: sophisticated yet simple***](#). KMWorld, April 1, 2008.

Malik, Shadon. [***The Silent March of Data Visualization***](#). DM Review Special Report, September 2007

Links to Other Resources and Blogs

Enterprise Search Sourcebook: <http://www.enterprisearchcenter.com/SourceBook/>

Lee Romero of Novell on Enterprise Search: <http://blog.leeromero.org/>

Steve Arnold on Beyond Search: <http://arnoldit.com/wordpress/>

Curt Monash on Enterprise Search:
<http://www.texttechnologies.com/2008/01/14/enterprise-search-versus-web-search/>

Lynda Moulton on Enterprise Search: http://gilbane.com/search_blog/