
Choice: A Growing Problem in Enterprise Search

The *Wall Street Journal's* Whitens, Brightens, and Confuses (February 23, 2011, D-1) raised an issue that plagues enterprise search and content processing. The story documented that in a typical American supermarket, the consumer is faced with more than 350 different types of toothpaste. One interesting fact the *Wall Street Journal* offered was that in the US 69 new toothpaste varieties entered the market in the last 12 months. How different can toothpaste be?

The point of the article was that consumers face a bewildering number of choices. The impact of toothpaste proliferation is that “consumers are annoyed.” More significantly, the article asked a question that strikes to the heart of product proliferation, “Does brand matter?”

Enterprise search and content processing may be headed down the same path. I recently reviewed the enterprise search products from 32 vendors. Some of the vendors have high profiles. Examples in my analysis were Autonomy, Endeca, Exalead, and Microsoft. Other vendors were important but had more focused brand recognition, often within niche markets in the enterprise. Examples in my analysis were FLAX, Intelligenx, and SurfRay.

I extracted the pool of 32 from my much larger list of search and content processing vendors. The master list includes more than 200 vendors. Many of the companies on my list are surviving but lack the interest or the resources to illuminate the radar of the Fortune 500. Firms I excluded from my list of 32 include Sowsoft LLC, Sprylogic Technology Ltd , and X 1 Technologies, Inc.

The consulting firms with an enterprise search practice narrow the scope of their coverage to a handful of companies. I think the search consultants are responding to what their clients ask and, to some degree, what search vendors are covered in their competitors' reports.

When I wrote the first three editions of the Enterprise Search Report (2003 to 2006), I departed from the convention of covering six or eight vendors. I worked through 20 or 30 companies' search systems and learned that the job was simply too big and impractical to be sustainable.

I had bumped into the toothpaste problem. Does the enterprise market need hundreds of choices for information retrieval? The answer is, “Some people think so.” I am not so sure.

Most of the new comers to the enterprise search market focus on a specific problem in the market. The angle for some new vendors their search system gains fancy visualization features. Sophia Search emphasizes that its semiotics approach understand the context of enterprise content. For open source vendors, the differentiator is a willingness to allow licensees to make changes to the search system. FLAX, Funnelback, and Lucid Imagination, among others are more open. That flexibility along with eliminating software licensing fees may be one reason why Cisco Systems and LinkedIn have embraced open source search. When new entrants gain traction, news of those deals spurs other start ups to consider the eEnterprise market.

Many search vendor pitch venture capital firms or other funding sources. Search is expensive, and even established companies need infusions of capital to cope with support, research and development, and marketing demands. One frequent assertion a search vendor makes is that the market problem is sufficiently large to warrant developing and marketing a new search solution. Investors seem to be open to these arguments. BA Insight, Lucid Imagination, and Palantir Technologies received funding. Palantir obtained a staggering \$90 million investment last year, dwarfing Lucid Imagination's \$16 million in 2010 funding. In comparison, Vivisimo's \$4 million infusion a couple of years ago looks modest.

Start ups signal their transition from developing product to selling a product when the firms put in place a "real" sales and marketing staff. Almost immediately upon accepting the job, the marketer creates a legend for the products' features and functions. More search vendors are writing blogs like Coveo, a search vendor pushing into quite diverse market sectors. Others use Facebook, LinkedIn, and Twitter to create buzz and a presence. The major hurdle is for newcomers and established search system vendors to differentiate themselves. Autonomy uses "meaning based computing" and Exalead relies on "search enabled applications". Other tags are "instantly search terabytes of text" (dtSearch) and "extreme speed, extreme insight" (Exegy). The remarkable similarity of marketing collateral from search vendors makes most vendors like peas in a pod. Maybe the pod knows the differences, but I often do not.

Every vendor works overtime to generate revenue. My research indicates that most search vendors want to hit Autonomy's revenue of US\$800 million or Endeca's estimated \$130 million. The reality is that the majority of the search vendors hit a ceiling, often in the \$6 to \$12 million range, although there are exceptions like Exalead. If these figures seem out of step with other enterprise software revenues, they are. Search has a handful of companies that get the lion's share of available revenue. Most search vendors scramble for any deals that do not go to one of the best known brands.

At any point in the search vendor's life cycle, the vendor may reposition itself. Examples include Autonomy's move from search to infrastructure. The "meaning based computing" tag and the IDOL system provide continuity, but Autonomy now spans eDiscovery to collaborative systems. Other vendors add a feature like a connector to index Twitter content and describe themselves as a social search vendors. Social is hot, and I have a difficult time determining who has a fully functioning, robust system and who has marketing collateral with the words "social" in it. A similar problem surfaces in the hard-to-define niche of business intelligence. Some vendors do have business intelligence systems. Examples range from SAS to SPSS to SAP Business Objects. Other vendors are in the search space and use flashy report outputs or visualizations to show their business intelligence functions. Some vendors appear to be oblivious to the confusion the shape shifting creates. What is OmniFind? What is Microsoft Fast Search Server? What is Attensity? What is the Google Search Appliance? These are difficult questions to answer because the search system can deliver many different functions. Figuring out what vendor does which function best is a non-trivial task.

These five characteristics of search vendors are not unusual in the abstract world of digital information. Think about the cost, time, and effort an organization must invest to acquire a search system. The problem, I want to point out, is caused by the interaction of vendors with customers. The company, procurement team, or manager who wants to license a search

system often does not know what he or she needs. The vendors offer a smorgasbord of options, usually mixing and matching trendy terms with jargon like latent semantic indexing, cloud based solution, and discovery interfaces, among other phrases. The prospect is not sure what he or she wants and the vendor is tossing out concepts in order to increase the chances of making a sale.

This is a toothpaste problem. Hundreds of choices. No easy differentiation.

In my experience, name brand search systems become the safe choice. Oracle offers its Secure Enterprise Search 11g system when an Oracle database customer requires enterprise search. The licensing organization may want to select the SES 11g system because the Oracle database administrator, the chief financial officer, and the technical team are familiar with how Oracle works as a vendor. SES 11g becomes the safe choice. The same approach works for IBM-centric organizations. OmniFind is based on Lucene, but because IBM stands behind the search system, the IBM shop either licenses OmniFind or selects from one of the search systems that comply with IBM's standards; for example, Endeca. Microsoft has done a great job of placing SharePoint into organizations. When search is required, Microsoft Fast Search Server is the obvious choice. For SharePoint licensees who want an alternative, Microsoft offers certified vendor solutions which "snap in" and work. Examples include Fabasoft Mindbreeze, SurfRay Ontolica, and BA Insight. The confusion created allows some organizations to make what is called the "logical decision."

In open competitions, the prospect faces the daunting task of narrowing the choice of potential search vendors. Some firms rely on pricey analyst reports that offer simplified descriptions of each vendor's system. Most of these reports provide the potential licensee with a quick and easy way of narrowing a procurement to a handful of vendors. The potential licensee does not have the time, knowledge, or patience to consider more than a handful of solutions.

Special problems do arise. Dissatisfaction with incumbent search systems often triggers the cyclical hunt for a better, faster, and cheaper findability solution. Another source of annoyance with enterprise search is its cost.

What causes cost overruns occur in a search implementation?

There are four factors that contribute to the bad rap enterprise search often has among information technology professionals. There are other factors, but I want to highlight these four items.

First, because the client may not know exactly what function the search system must perform, the search vendor may have to find a work around or write new code. Some search vendors sell using buzzwords, and then those vendors trust that the programmers can code what is needed to meet the client requirements. The approach is fraught with risk, because the client may be unwilling to pay for changes or unwilling to wait until the function is implemented. As delays and costs rise, the licensee is stressed. The best case is that the vendor can work up a fix without additional cost or adding time to the project. In the worst case, the licensee refuses to pay and may renege on the deal. Believe it or not, this approach is common and has

created problems for some of the highest profile vendors in the search sector in the last five years.

Second, because the licensee was unaware of the content types in his or her own organization, the search vendor may discover that connectors, filters, or transformation routines are required that were not part of the original estimate. Licensees of search systems often assume that a vendor can handle the information types within an organization. Search vendors can handle common file types and may have connectors to some specialized systems like Lotus Notes or a Documentum repository. Other files types such as the i2 Ltd. ANB file cannot be easily processed by a third party system. A company that tries to reengineer a connector to handle certain file types may find itself a defendant in a legal matter as Palantir did before the dispute was settled out of court in February 2011. A price tag for a connector may be acceptable to some licensees. For others, the search vendor will have to find a solution. Tensions often rise over what at first glance seems a trivial matter, accessing content on a server in a third party program.

Third, the performance of the search system is unsatisfactory after launch. There are numerous reasons for performance problems. The first reason is that the volume of content to be processed and the index update or refresh bog down the system. Some enterprise search systems require that an index be rebuilt for each update. Other vendors have indexes that support incremental updates. Regardless of approach, machine and network resources must be sufficient to keep the content processing in sync with user expectations for freshness of the index. It is easy to say “real time indexing” but very difficult to implement in most organizations without necessary resources. The second cause of performance issues is query response time. Users are often conditioned to the response times of public search systems like Bing, DuckDuckGo, or Google. Enterprise search is quite different from Web search. Often the only way to improve the query response time is to make significant changes in the hardware or alter some of the value-added processing that displays suggested content, mashups of information of different types, or other types of triage. Licensees do not understand these issues, and the search vendor may not have many degrees of freedom unless the licensee agrees to additional expenditures.

Finally, search systems crash. When a search system is deployed, most licensees assert that the information technology department has a back up and restore system in place. In actual practice, search systems require redundancy plus back up systems that conform to the methods used by the search vendor to keep indexes current. Google encourages licensees of the Google Search Appliance to consider redundant appliances. Other search vendors recommend different solutions, but some vendors accept the licensee at his or her word. When a search system fails, it usually costs money to get back up and running. Search systems are not a single process operation. Search requires distinct subsystems which work together. A failure in one component can cascade to other subsystems, making troubleshooting for the licensee difficult.

Compared to enterprise search, selecting toothpaste is trivial. Search and content processing choice is a more complicated affair. In my opinion, the number of options is putting search and content processing in the line of fire. I want to echo the Wall Street Journal, “Does search brand matter?”

Stephen E Arnold, February 28, 2011

Mr. Arnold is a consultant. More information about his practice is available at www.arnoldit.com and in his Web log at www.arnoldit.com/wordpress.