

## Enterprise Search Vendor Comparisons

### Introduction

The enterprise search market is not homogeneous and so a considerable amount of effort is required to find the most appropriate solution. This brief report profiles eleven of the most prominent search technology suppliers. Two more are included as a sample outside the usual names associated with this topic.

It is very difficult to compare these offerings and in fairness to the suppliers it would be misleading to do so. However some of the products do stand out in various ways and so we've highlighted them in this introduction.

Also included is an overview of why enterprise search might be of interest at all, and some strong statements on the value it can deliver and the pitfalls to avoid. Readers primarily interested in product can ignore this section.

### Comparison

Search performance is primarily dependent on two factors – the efficiency of the software and the tuning options that are provided. The following technologies stand out in this respect, and the order is indicative of what might be possible:

1. Lucid Imagination based on Solr/Lucene
2. Dieselpoint
3. Google Search Appliance
4. ISYS
5. Endeca
6. Brainware

Search sophistication primarily depends on the number of search techniques used and the ability to configure them as needed.

1. Endeca
2. Lucid Imagination
3. Autonomy
4. Vivisimo
5. Exalead

Presentation of search results requires variety of user interfaces and rich formatting options:

1. Endeca
2. ISYS
3. Vivisimo
4. Autonomy

Interoperability and support for diverse data sources:

1. Endeca
2. Autonomy
3. Vivisimo
4. Coveo
5. ISYS

It is fairly clear that strengths and weaknesses vary and that selecting the most appropriate solution will depend on needs. It should be added that performance is much more important than many appreciate. It is not uncommon for some vendor products to inflict heavy resource requirements with search response times measured in the tens of seconds (or worse).

The very high ranking of Endeca reflects the excellence of the solution. However this too will be dependent upon specific usage.

## The Case for Enterprise Search

Most large organizations have spent the last half century or so accumulating copious amounts of data on every conceivable aspect of their operation. The cost of this acquisition has been enormous - much more than many business managers realize. Every detail of every customer has been input by someone. Every contract has been assembled and processed by someone. The creation of every piece of marketing collateral has probably involved several people. These are all information costs, and for companies in service industries these costs may be as high as eighty per cent of revenue. The largest component of information costs is labour. Most managers do not think of it in this way. Labour costs belong to sales, finance, design, production and so on. The common thread however for all these labour costs is that the main activity involves the creation and manipulation of information, and as such they are predominantly information costs. Unless we can get at this information it has no value at all.

Just consider any large corporation operating in a service industry - take insurance for example. Labour costs associated with information creation and manipulation will be around seventy per cent of revenues (and this is being conservative). If this company had average annual revenues of ten billion dollars over the last ten years it is easy to see that its information assets cost seventy billion dollars to acquire, manage and manipulate. Surely it would be worth investing in people, methods and technology to get at and derive some value from this information?

We have had no problem at all spending on systems and technology to acquire and manage information. The four billion dollars a year that a large bank might spend on IT primarily supports the systems that acquire and manage information - very little of it will be spent on information search and deriving value from this very expensive commodity. There is a strange kind of logic here. We accept that acquiring and managing information will be expensive, and that this is justified because it supports day-to-day operational activity. On the other hand we have something of a miserly attitude to the other half of the equation - getting the information out and deriving some value from it.

We only do three things with information. We create it, We manage it. We search it. Search is at least a third of this equation, and the most important third because it concerns risk management and value. The accumulation of information has accelerated in recent years, and looks set to accelerate even more. It is time for organizations of all sizes to get serious about search if they do not want risks to escalate and their most precious of all assets to simply decay in value with the passage of time. To drive the message home just consider how useful the Internet would be if we could not search it. Hundreds of millions of people and organizations creating information, but with no means of access. Many large organizations have created a private Internet without a search engine.

As if this was not bad enough the situation is made worse - much worse, by the fact that individuals, groups and organizations carry knowledge and information around with them that they do not, and often cannot share. If the terabytes of data sitting in corporate databases are underutilized, the knowledge of possibly thousands of individuals in your organization is a resource that is not only untapped, but is acquired and utilized with great inefficiency. Knowledge management has been touted as a means of reducing these inefficiencies, but coercion is not an effective mechanism for getting people to share what makes them valuable to your organization. This situation can be eased by creating a common language which everyone can share, and providing mechanisms to interrogate explicit expressions of knowledge in documents, emails and any other files created by knowledge workers.

The economics of information search is simple enough. We need to derive more value from search than it costs us to facilitate and execute. The problem in all of this is putting value on the results of search. Many managers complain that they cannot access the information they need for sound decision making. Effective search will help managers avoid bad decisions - but putting a value on these decisions is usually very difficult.

Information workers spend anywhere up to thirty per cent of their time searching for information and many claim that a third of this time is totally unproductive. If effective search could reduce this we should see more productive workers and greater throughout - but who will measure it? Many organizations have been forced to do something about search because of the need to comply with various regulations. Document, email and database search is usually what is needed, and a search facility is installed to make sure that the organization does not fall foul of the various regulators - but it does little for productivity, decision making, intelligence gathering and opportunity identification within the organization.

So what does an organization need to do to get serious about search? A new or enhanced set of skills is needed. We need information professionals who understand how to categorize information and we need them to work with technicians implementing search engine technology. We also need structured dialogue with knowledge workers and management to establish a common understanding of terms so that search means the same thing to all people within an organization. Search really is not about just buying a black box and accepting whatever it throws up with gratitude - although this might be a good place to start just to experience the shortcomings and establish what else is needed.

A word of caution is needed here. It is just too easy to get side-tracked by long technology and methods detours. Someone needs to make sure that consulting, the setting up of various categories and structures, and the implementation of the search technology remains a pragmatic exercise with well-defined milestones. Attempts to create information categories that are all embracing or just too precise are a fruitless exercise that should be avoided.

Perhaps no other technology is so affected by smoke and mirrors as Enterprise Search. The jargon is formidable and the claims often outrageous. Not surprisingly many users of search technologies are dissatisfied and confused - sure signs of a technology that is nowhere near maturity. On the one hand we have the uber-high tech offerings from companies such as Autonomy who pour derision on the familiar Google type keyword search, and on the other hand practical, functional offerings from new startups that will do eighty per cent of what many organizations need with twenty per cent of the fuss. Clearly it all depends on what you want and what you believe the technology can deliver. Despite the aggressive claims of many of the vendors there are actually no free lunches. If you want more out of your search technology you will have to put more into it.

It is worth spending a few moments reflecting on how we got into this mess - because that is what it is. For decades we have merrily been accumulating data, and more recently this data has become richer and more complex. Structured databases, email, electronic documents, scanned documents, html, graphics, sound, video and whatever else the future delivers, all have different formats and programs to make them meaningful to human beings. The cost of producing and managing this data is formidable. Most organisations spend anywhere between two and fifteen per cent of their revenues managing data, and anywhere up to eighty per cent actually creating it. During the last ten years it has suddenly dawned upon us that maybe this expensive commodity can be used to create value and help manage risk. Unfortunately this was not thought of when we were busy creating data, and so the formats often have little or no information to help us make sense of the content. As a result we have to find ways to extract meaning from, and categorize data as an afterthought - so enterprise search is likely to be messy and less than satisfying if your expectations are too high.

Given the current state of the art organizations have three options. They can be summarised as:

- Do Nothing - and wait for the technology to mature and become cheaper.
- Buy a black box search engine and just accept that it will only do so much.
- Venture into the labyrinth of advanced search technologies with the attendant high cost and high risk, in an attempt to deliver something exceptional.

The first option will not be an option for many organizations, they have to comply with various requirements and a search technology makes such compliance feasible. There may also be a desire to reduce

the cost of searching for information - a cost that is rising exponentially as we create more of the stuff, and as formats become more diverse. If doing nothing is not feasible then we should look at option two, and while it may seem like a lazy man's approach to search it has many advantages. It is predicated on an understanding that the search technology will do the donkey work and people will have to do a certain amount of manual processing. Search technology will comfortably find all the documents and/or records with the words 'returned by customer', but finding out why might take some human intervention. This approach would largely avoid lengthy, costly technology detours into semantics, ontologies, taxonomies, tagging and other techniques to attach meaning, but it would also mean that more human effort might be required to process search results. It's a trade-off - but it might be a good place to start. Venturing into the labyrinth should only be done by the prepared. Expect longer time scales, mistakes, greater complexity, more cost and a high on-going maintenance overhead. On the plus side you might just deliver a real advantage to your organization - but the risks are much higher

## Autonomy

Enterprise Search

ESV004 - April 2011

### Strengths

- ✓ Very large integrated suite of search solutions.
- ✓ Unique search technologies.
- ✓ Ability to search all content.
- ✓ Contextual analysis of information.

### Weaknesses

- ✗ Significant infrastructure requirements.
- ✗ Expensive solution.

### Advisory

Autonomy is the undisputed 'big-dog' of enterprise search. This dominating position is a result of its early adoption of novel techniques (specifically Shannon information theory and Bayesian) which delivered functionality no other supplier could match. The company is also very good at marketing and leverages relentlessly on its 'meaning based computing' mantra. But this is changing.

Curiously enough some of the Open Source offerings are approaching Autonomy in terms of the core search technology (Lucene for example), and it is only a matter of time before others catch up. However Autonomy has created such a commanding presence that its position is unlikely to be challenged any time soon.

For many organisations Autonomy will be overkill, and an expensive one at that. Where interrogation of raw video, voice and graphics files is important then there are only a few alternatives (Nexidia for example). But for text based search there are many more efficient alternatives. We would suggest that organisations look at their real needs, and if they are primarily text based then Autonomy may simply be an expensive solution with considerable infrastructure requirements.

### What it Does

Autonomy's 'Intelligent Data Operating Layer' (IDOL) server is at the heart of the offering. This unique piece of technology is well regarded for its tight security, integration capability and the ability to extract meaning from almost any data source. Universal Search provides an end user search interface with various search tuning parameters such as weightings (date, synonyms, sources etc) and the ability to automate repetitive search tasks. A large variety of connectors are available (at a cost) and it is probably true to say that if it can be stored electronically Autonomy can access it.

Autonomy's real growth comes from its search applications for specific markets. The growth of regulatory and compliance demands has been a real boon to Autonomy and it offers a bewildering set of solutions in this space. These include ControlPoint for policy enforcement of content distribution, Interaction Control Element for real-time security, Supervisor to monitor communication between individuals, and more recently Social Media Governance to analyse the large volumes of data generated by social networks.

Archiving and records and records management also feature strongly in the solutions Autonomy offers and e-discovery is also well addressed. Solutions are offered for legal and professional services which are simply applications of the technology to these particular domains.

Finally Autonomy provides a web content management environment (from its acquisition of Interwoven). This is definitely not a best-of-breed solution, but many organisations will undoubtedly use it because they have made a 'strategic' commitment to Autonomy information management technologies.

The Autonomy story is quite simple. The founders of the company applied some novel techniques to information search and jumped ahead of the competition. This in itself is not particularly significant, but the company then went on to apply the technology to every conceivable aspect of information search, and has been very astute in shadowing the various needs of large firms (compliance for example) as they arose. Any number of other search technology vendors could have done this, but it was Autonomy that sensed the various trends and they have profited accordingly.

### **What's Different**

Autonomy has differentiated itself for many years through the unique search algorithms it developed during its inception. In reality nothing has changed since then apart from the excellent marketing and clever anticipation of large corporate needs. In many ways Autonomy was the right product at the right time, although things are now changing.

The company needs to move on from the techniques which have sustained its advantage for more than a decade, since others are catching up (in fact there are better technologies out there now, and a blurring between enterprise search, business intelligence and predictive analytics). Autonomy is already trying to make moves into the broader information analytics space, but it will meet with stiff competition (and better technology). While most commentators continue to see Autonomy as unchallenged for some considerable time we beg to differ. Convergence of information analytics technologies will cramp Autonomy's space and its technology will not look so special after all.

### **About Autonomy**

Autonomy needs very little introduction and with revenues of around US\$900 million it is the 'big dog' of enterprise search. The list of customers is a who's who of global corporations and it will undoubtedly remain this way for some time to come.

[www.autonomy.com](http://www.autonomy.com)

## Brainware

Enterprise Search

ESV007 - April 2011

## Strengths

- ✓ Extraction of information from paper documents.
- ✓ Unambiguous productivity gains.
- ✓ Variety of vertical solutions.
- ✓ Novel (and very effective) search technology.
- ✓ Easy to embed in other applications (ERP for example).

## Weaknesses

- ✗ Native user interface is fairly basic.

## Advisory

It's always refreshing to come across technology with a simple, obvious productivity advantage. Brainware Distiller excels in the extraction of information from paper documents and unstructured electronic communications so they can be processed in an organisation's information systems (ERP, CRM etc). Unlike the 'smoke and mirrors' productivity arguments that infect much of the IT industry Brainware offers technology that is easy to use and comes with a variety of reporting mechanisms so productivity gains can be measured – it's very simple.

Brainware Globalbrain provides enterprise search capability based on some innovative and sophisticated search technology (more on this later). It allows organisations to search most data formats and is easily embedded into other applications. This is particularly useful when users do not want to access a separate search utility, but want search in the applications they use every day.

Brainware offers obvious productivity gains and straightforward search technology. It provides an uncomplicated route to exploit information resources and will appeal to many organisations.

## What it Does

Brainware Globalbrain is based on probability techniques called 'n-grams'. These provide a mechanism for predicting the significance of a word or phrase and its likelihood of occurrence. The net result is that Globalbrain will parse natural language so the usual Boolean operators do not have to be employed (which makes search easier to use). There are similarities here with the way Autonomy works since they both have their roots in Shannon's information theory.

While the front end functionality is fairly basic Globalbrain is really intended for use within other applications – it's an engine that can add value to almost any application. For many users this is preferable because search can then be conducted within familiar applications. To this end Globalbrain comes with a variety of mechanisms to facilitate integration including Java, COM, SOAP and .NET.

Brainware Distiller is primarily concerned with extraction of information from paper documents and unstructured electronic messages such as email. This is not simple OCR, but a complete system that knows the relevance and order of information that is extracted. The technology is quite capable of matching details on a processed document with master data (the name of a customer for example). Match rates are very high



because of the search technology the company has developed. There really is no excuse to have people input details from paper documents.

Brainware provide a number of discrete solutions to address invoices, orders, remittances and mailroom automation. It also offers vertical solutions in banking, biotech, energy, legal, manufacturing and several other industries.

Central to Distiller functionality is the reporting and analytics. These provide metrics such as extraction efficiency, throughput and other ad-hoc reports. This functionality gives a simple measure of the productivity gains that are being realised.

### **What's Different**

Brainware provides some very smart and disarmingly simple technology. The search technology uses novel techniques which result in simpler query formulation and much greater context relevance for search results. The embedding of search within day-to-day applications is something we like – more than anything it avoids time consuming, and unproductive context switching between applications.

The Distiller applications can be implemented in a very short amount of time (Brainware claim 2 days). Benefits are realised immediately. Ultimately it is capable of becoming an enterprise solution to the capture of all unstructured data – hard copy and electronic. If your organisation is still engaged in the manual processing of documents then Brainware is worthy of investigation.

### **About Brainware**

Brainware was formed in 2006 and originates from SER Solutions AG - a well-known content management software provider. The company is privately owned and so financials are not available. Customers include Airbus, Kimberly Clark and Sun Chemical.

<http://www.brainware.com>

## Coveo

### Enterprise Search

ESV001 - April 2011

## Strengths

- ✓ Excellent solution for knowledge intensive environments.
- ✓ Extensive array of administration and monitoring tools.
- ✓ Customer sales/support/call centre solutions.
- ✓ Multiple user interfaces.
- ✓ Usage analytics.

## Weaknesses

- ✗ Rich functionality means close attention to performance issues.

## Advisory

Coveo provides a mature enterprise-wide information search environment. The technology is well suited to knowledge intensive environments where a global view of information resources can deliver efficiency gains and enhance efficacy. Customer oriented solutions are also a main feature of the offering and Coveo has considerable experience in the customer support/ sales/call centre application of search technology. The user interfaces are worthy of special mention and these include rich dashboards with graphical representations, an Outlook sidebar, browser based access and mobile access.

The Unified Index sits at the heart of the product and this provides access to virtually all information resources including email, documents, social networks, CRM and ERP applications, databases, web site content and multimedia. This provides a true 360 degree view of the organisation and can revolutionise the exploitation of information.

Coveo is an ideal solution for large organisations with dispersed, disparate information resources that are difficult to unify. The customer oriented solutions will also be attractive to many businesses, and knowledge intensive environments will be particularly well served by the technology.

## What it Does

At the heart of Coveo is the Unified Index. This monitors content in real-time so that new additions and modifications are always available. It accesses most forms of content and implements a security model so that confidential and sensitive information is protected from access. The 'concept and summary indexing' capability adds to the ease with which information can be accessed.

Users of Coveo are well served by a variety of interfaces. The dashboards are particularly powerful providing an opportunity to customise the way information is viewed with useful graphics and the juxtaposition of related information. The Outlook Sidebar provides information access capabilities that are integrated into the Outlook interface. This is a highly functional utility for intensive Outlook users. A Desktop Floating Searchbar provides an 'always available' interface to the search capabilities of Coveo, allowing users to get to the information they need regardless of the application they might currently be using. Finally Coveo can be accessed via the Blackberry with obvious uses in sales, support and other field based activities. It is not clear whether Coveo intend to support the iPhone or Android devices.

Coveo categorise the application of their technology into four main areas: Collaboration and knowledge management, Compliance and Internal Investigations, Expertise finding and customer service. This misses the general need for information which is always present and is well served by Coveo. Someone in purchasing might need information on a supplier for example, and this might involve emails, database, social data and other types. Coveo should not be viewed in a limited context.

Administration and analysis of enterprise search facilities is often given scant attention. However Coveo excels in this area offering various deployment, tuning, optimisation and usage tools. These address issues such as availability of search capability to users, tuning of ranking parameters, creation of routine backups, analysis of index usage, management of the thesaurus and so on. We particularly like the analytics which allow administrators to optimize search paths and navigation. The functionality is broad with reports on top queries, popular documents, failed queries and these capabilities are often used in troubleshooting.

Coveo will connect with most data and information sources. The connectors will often embrace the native security model of an application and make use of specific functionality.

Organisations with strong customer facing activities will benefit from the Customer Information Access Solutions offered by Coveo. The company provides many example of productivity improvements that come from efficient, functional access to information resources. The main application areas are customer self-service, call centres and what Coveo call Customer 360 where managers and executives can get an overall picture of customer facing operations.

## **What's Different**

Coveo distinguishes itself by providing ubiquitous access to enterprise information resources to those who need them. The various interfaces make this particularly easy. It also proves its enterprise credentials through the extensive administration and analysis tools it offers, which are absolutely necessary in a large enterprise setting.

This is a very pragmatic solution to enterprise search needs, but with sophisticated functionality. Knowledge workers will be particularly well served with easy access to information resources and identification of subject experts.

## **About Coveo**

Coveo is a privately held company founded in 2005 and has headquarters in both USA and Canada. Customers include Computer Associates, Unilever and Panasonic.

[www.coveo.com](http://www.coveo.com)

## Dieselpoint

Enterprise Search

ESV002 - April 2011

### Strengths

- ✓ Industry leading performance.
- ✓ Small footprint.
- ✓ Excellent navigation functionality.
- ✓ Java based technology with clean API
- ✓ Highly configurable.

### Weaknesses

- ✗ Poor business oriented description of the platform.

### Advisory

Dieselpoint is something of a Porsche in the Enterprise Search space. It is very fast, well-engineered, doesn't carry much excess weight, and its text based searching technology can be made to do almost anything. The user interface is also very functional providing navigation via dynamically created menus and new features such as taxonomies.

IT vendors can usually be classified as overtly marketing biased or technology biased. Dieselpoint belongs to the latter camp and the success of its technology (it has some very impressive customers) is proof of the quality of the offering. This is not a smoke and mirrors supplier.

For organisations with demanding search requirements and a need for sub-second responses, even when hundreds of searches are being executed every second, Dieselpoint is a platform that should be considered. The footprint is small and the technology moderately priced. The spectre of burdensome search infrastructure is very unlikely with Dieselpoint, and for many users there will even be no need to run the technology on a separate server.

### What it Does

Dieselpoint is essentially a single product that can be configured to satisfy virtually all text based searching needs. From a productivity standpoint Dieselpoint has the potential to revolutionise information search efficiency and efficacy. This is no small claim and from the user feedback we have seen the combination of lightning fast search with highly functional navigation does indeed transform information search productivity. The product was engineered from the ground up in Java to provide the performance, extensibility and integration many organisations need. Utilising the technology from other systems is a fairly straightforward affair, allowing powerful search functionality to be embedded into other applications.

Most common file and database formats can be accessed by Dieselpoint, including HTML, XML, PDF and Microsoft Office documents. It also supports various flat file formats and commonly used databases such as Oracle, MS SQL, My SQL and DB/2.

The navigation features distinguish Dieselpoint from the crowd in many respects and these include dynamically generated menus that are created as a direct result of a search. Other features include the taxonomy data type which supports the creation of hierarchical menus, and several interface classes such as lists, trees, dropdowns and dynamic ranges for high-cardinality data.

Text search functionality supports many options, some of which are common in other products and some of which are not. Automatic relevance ranking is obviously de facto, but Dieselpoint adds a forced ranking option which ranks on the presence or absence of keywords in a query. Document summaries can be displayed and a full query language is supported with the usual Boolean operators. The list of features is fairly long and without covering all of them (the full list can be found on the Dieselpoint web site) it is obvious that any search need can be accommodated.

A variety of application development and administration capabilities are offered. Indexes can be created, edited or deleted using an administration application and the in-built thesaurus can also be maintained.

Other technical features include a built in webserver and servlet container, an autoindexer which will crawl through websites and directories, an item pre-processor which will process items before they are added to an index, and a particularly powerful Java based API.

Finally it should be added that Dieselpoint comes with a search –enabled product catalogue application. If this is a likely application in your business then it is definitely worth looking at.

### **What's Different**

It should clear by now that the main differentiator of Dieselpoint is the speed of search execution, the navigation functionality and ease with which the technology can be called from, or embedded into other applications. These technical features can have a profound impact on productivity and simple arithmetic will demonstrate the cost of users having to wait tens of seconds for a search to execute. Dieselpoint implementations in high throughput scenarios typically deliver sub-second search execution times.

The only way we can fault Dieselpoint is in its lack of relevant documentation for the business manager. The impact of sub-second search execution needs to be explained in terms of productivity and efficacy.

### **About Dieselpoint**

Dieselpoint was formed in 1999 and is a privately held company, so financial details are not available. It is based in Chicago USA and customers include McGraw-Hill, PSS WorldMedical and OHSU Health.

[www.dieselpoint.com](http://www.dieselpoint.com)

## Endeca

### Enterprise Search

ESV006 - April 2011

## Strengths

- ✓ Combined business intelligence and enterprise search.
- ✓ Very advanced search and analytics capability.
- ✓ Highly productive 'templates' for business users and analysts.
- ✓ Industry leading retail solutions.
- ✓ On-demand version for smaller organisations.

## Weaknesses

- ✗ Cannot directly interrogate raw video and sound data.

## Advisory

Endeca is well ahead of the general enterprise search market in many respects. The business intelligence (BI) and search spaces must inevitably converge as BI seeks to access richer content and search technologies evolve to offer greater analytical capabilities. Endeca already offers this, and somewhat unusually we have even received begrudging acknowledgement of this lead from a few of the company's competitors.

The technology accesses virtually all the forms of electronic storage known to man – databases, web sites, documents, email and so on. If there is a weakness it is simply that rich media such as video and sound cannot be interrogated directly – but other specialist technology suppliers can fill this gap (eg Nexidia).

We would advise any large organisation considering enterprise search and/or business intelligence solutions to look at Endeca – it will set a benchmark. And for retail businesses the Commerce Suite will probably offer possibilities that have not even been considered.

## What it Does

At the heart of Endeca is the MDEX engine – quite simply one of the best search and analytics engines available. The faceted search capability is worthy of special note. Faceted search allows categorisation and ranking based on attributes (name, age, title etc) and while many vendors offer this capability very few do it well for very large data sets. The flexibility of this engine is due in large part to the fact that each data record carries its own metadata, allowing search and analysis structures to be constructed 'on-the-fly' – and with great speed.

The capabilities of the MDEX engine are then exploited through the Discovery Framework. This is essentially a tool for building search and analytics applications using standard components. These break down into navigation tasks, creating rich visual interfaces, personalisation functionality (eg bookmarks), layout and style components and when the standard components are insufficient a SDK is provided for highly bespoke requirements.

The Commercial Suite applies the MDEX engine and development tools to retailing. A central feature of this solution is provision of tools for non-technical users to create web pages and design interactive environments for customers. Rule driven dynamic merchandising is a particularly powerful capability and rich sources of content can be easily integrated into the customer experience. The Merchandising Workbench

provides a 'retail design' environment with creation of conditions for upselling and cross-selling opportunities. Finally Endeca provides Social Commerce functionality allowing users of Facebook to view and purchase goods directly, and for user generated content to be incorporated into the customer experience.

Mobile commerce channels can be exploited through the mobile design tools supplied by Endeca. Once again business users and analysts are presented with a set of pre-built templates to create mobile commerce applications that support most mobile device platforms (Android, iPhone ect).

For small and medium size organisations Endeca offers an on-demand solution which provides pretty much the same functionality but on a per user charging basis.

In some ways the Commerce Suite can overshadow the more general search and analytics capability Endeca offers, and the general applicability of the technology is well illustrated by the diversity of large organisations using the technology. This includes Boeing, Raytheon, Ford and Nasa – not particularly well known for their retail operations.

## **What's Different**

Endeca is a leader in the integration of business intelligence and search technologies. This is a big deal and should at least be considered by most large organisations. Bringing rich content into the analytics arena multiplies the effectiveness of many analysis activities, while providing high speed access to structured data enhances search tasks in a similar manner.

What also differentiates Endeca is the highly productive development environment the technology offers. Users, managers and analysts need to be empowered to find the information they need when they want it – not when a technician is able to program the necessary functionality or a search specialist available to design a taxonomy. To this end Endeca provides what may well be the most productive end user search and analytics environment available.

## **About Endeca**

Endeca headquarters are in Cambridge MA USA and there are several other offices both in the USA and worldwide. The company is experiencing rapid growth (hardly surprising) and revenues for FY 2010 look to be around US\$150 million. The customers read like a Who's Who of retailing and any number of other large corporations including IBM, Office Depot and Panasonic.

<http://www.endeca.com>

## Google Search Appliance

Enterprise Search

ESV005 - April 2011

### Strengths

- ✓ Low cost, high performance search engine.
- ✓ Connectors to most file, document, database and content types.
- ✓ Ideal solution for unsophisticated needs.

### Weaknesses

- ✗ Poor user interface for enterprise search needs.
- ✗ Federated search can be problematical.

### Advisory

The Google Search Appliance (GSA) is essentially 'Google' in a box and for serious enterprise search requirements this is not really adequate. Many of the vendors we look at in this space (ISYS, Coveo for example) have sophisticated user interfaces with functionality that supports very smart categorisation and filtering of information. The engine is just a starting point, and provided it delivers adequate speed the real differentiators in enterprise search are more concerned with the user interface and information delivery mechanisms (i.e. dashboards, mobile devices, embedded search).

Medium size organisations with modest needs will find that the GSA provides a low cost, high performance option. Most file types known to man can be accessed by the GSA and integration with enterprise identity and access management solutions means that confidentiality and security is fairly easy to implement.

We would suggest that organisations see what is on offer from more mature enterprise search vendors before looking at the GSA. It is likely they will find the user interface lacking when compared to other offerings. Even so if needs are fairly straightforward then the GSA may provide a low cost, high performance solution.

### What it Does

The GSA is a piece of hardware that is plugged in to an organisation's network and is then left to get on with content discovery, indexing and information delivery – it's that simple. Users access the GSA through a familiar Google type search interface where search results are ranked and listed in the way most of us are familiar with. It should be added that the GSA will search intranets, websites, portals and cloud based applications (e.g. Twitter, Google Apps) in addition to the usual content held within most organisations.

Security is well catered for by supporting a variety of authentication and sign-on methods. These include LDAP, NTLM, PKI, Kerberos and Windows Integrated Authentication. Late binding and early binding options are also supported.

Google has enhanced the user interface recently with dynamic navigation filter and sidebar elements. Dynamic navigation allows metadata attributes to be used to narrow a search and side bar elements format



presentation so that different types of information (internal databases, external web sites, information on people etc) are separated.

At the heart of the GSA is a self-learning algorithm that refines the search process as it learns from the behaviour of users. A spell checker also assists users by checking spelling and making suggestions (supports multiple languages).

The architecture of the GSA is advanced and is capable of supporting searches on hundreds of millions of documents. The GB-7007 supports installations up to 10 million documents and the GB-9009 provides theoretically unlimited capacity. GSAs can be connected in a group to improve throughput and provide resilience. Load balancing is a feature of the group architecture and information is shared so that crawling is a one-off activity.

All-in-all the GSA provides huge scalability – it just needs something smarter at the front end. Maybe Google should buy a suitable enterprise search vendor.

### **What's Different**

Google is clearly trying to commoditise the enterprise search space with low cost hardware and search technology. If enterprise search was the same as internet search they would have no real competitors – but it isn't. Content on the internet is linked and voluminous – content in the enterprise is usually poorly tagged, unlinked and instead of millions of search results a user might simply get a few dozen or less. It's a different game and Google needs to enhance the presentation front-end.

### **About Google**

Need we say anything.

<http://www.google.com/enterprise/search/gsa.html>

## ISYS

Enterprise Search

ESV003 - April 2011

### Strengths

- ✓ Extremely productive user interface with automatic classification of content.
- ✓ Embedded into technology from EMC<sup>2</sup>, HP and Sybase.
- ✓ Small footprint – high performance engine.
- ✓ 20+ years search experience - and it shows in the product.
- ✓ Timeline refinement and document filters for e-discovery.

### Weaknesses

- ✗ Low key marketing is at last giving way to a higher profile.

### Advisory

ISYS provides what is rapidly becoming recognised as the “professional’s choice” in enterprise search and e-discovery technology. Over a period of twenty years the company has developed a state-of-the-art set of search technologies that are becoming widely recognised by other players in the IT industry – the most recent being Sybase and its decision to use ISYS document filters in its own technology.

For enterprise users ISYS offers a number of significant benefits. More important than many initially appreciate with enterprise search, is the efficiency of the technology. It is not uncommon to find organisations supporting proliferating infrastructure to serve search needs with top-heavy search technologies. ISYS indexing algorithms are very efficient and some of the largest applications handle four million emails a day with less than ten indexing servers – the footprint is modest to say the least.

The user interface has to be tried to be appreciated. It features a refinement panel that makes drilling down into content particularly easy, and the timeline refinement will be particularly useful in e-discovery applications. More recently ISYS has developed its document filter technology so that a diverse set of content can be interrogated with full extraction and rendering.

Overall it is really quite difficult to fault the offering and in our 2009 comparison of enterprise search solutions we gave it lead position. Enterprise users should at least take a look at the technology to set a benchmark for alternatives they may be considering.

### What it Does

ISYS provides a complete suite of enterprise search technologies which facilitate access to over 200 content formats. This is complemented by the mobile offering which, in typical style, delivers full search capability to mobile users – a particularly useful feature in collaborative environments. The more recently developed document filter technology has been selected by Sybase to enhance its own technology. This provides in-situ extraction of content and high resolution rendering – ideal for e-discovery applications.

The Enterprise Server conducts federated search over most file, container and email formats. The server delivers excellent performance and scalability, and has modest resource requirements. Applications for the technology are diverse and include e-mail archiving, process (BPM) analysis, e-discovery and search embedded into other applications. Connectors are provided for access to Documentum, Interwoven, Symantec

Enterprise Vault, Lotus Notes and Microsoft SharePoint content. The efficiency of the technology means that multiple terabytes of data can be processed within a single index. Over sixty languages are supported.

ISYS Anywhere is the mobile offering and it provides mobile workers with access to all enterprise content from mobile devices – and the important word here is ‘all’. This is ideal for field workers in sales and support, and for collaborative teams who might need to access to information while on the move.

The ISYS Document Filters are rapidly becoming recognised as an industry leading solution and are ideally suited to compliance, e-discovery and text analysis applications. Support is provided for open source search solutions such as Lucene and it would seem that implementation is very rapid (a matter of hours) and fairly straightforward.

ISYS Workgroup provides access to desktop content and shared repositories, and WorkGroup Web adds the ability to facilitate search for partners.

Finally ISYS QuickSite supports the search activities of web site visitors and again is very easy and quick to install. A variety of administrative functions allow an organisation to modify search behaviour according to the habits of visitors.

ISYS products are available for individual users (a convenient way to explore the technology), workgroups and the enterprise.

## **What's Different**

Experience is what differentiates ISYS from most of its competitors. Having been in the enterprise search market since 1988 the company has continually refined its product set until it now represents a best-in-class offering. Industry leading performance is the result of experience, and the powerful user interface is clearly a result of continual evolution based on the needs of users. In an industry where product excellence is typically uncorrelated with product success it is refreshing to see a leading technology gaining a prominent position in its market.

The considerable experience of ISYS will be particularly useful for firms in financial services, legal, government, technology, energy and healthcare. Other industry experience includes media, transport and engineering.

## **About ISYS**

ISYS Search Software was formed in 1988 and is a privately owned company – and so financials are difficult to get hold of. We do know that ISYS has around five thousand corporate customers and a growing number of IT industry customers who embed its technology into their own. The list of customers is impressive and includes Exxon, Visa International, Boeing and Singapore Airlines.

[www.isys-search.com](http://www.isys-search.com)

## Lucid Imagination

Enterprise Search

ESV008 - April 2011

### Strengths

- ✓ Extremely flexible and highly configurable search technology.
- ✓ Industry leading search methods.
- ✓ A full toolkit for building search applications.
- ✓ Large, highly skilled ecosystem of technicians and consultants.
- ✓ Open source – access to source code for highly bespoke applications.

### Weaknesses

- ✗ Not a 'plug-and-play' search technology.

### Advisory

Lucidworks Enterprise is a commercial search technology solution based on the Solr open source search offering, which in turn is based on the widely acclaimed Lucene search library. The bottom line is that businesses of all sizes now have access to a commercial grade, open source enterprise search offering that is undisputedly among the best in class.

While Lucidworks Enterprise can be used 'out of the box' to some extent, that really is not its purpose. Organisations with specific search requirements need a rich search development environment to satisfy their needs, and this is exactly what is being offered. Businesses such as ebay and Raytheon use the technology to satisfy very specific, highly bespoke requirements.

Lucidworks Enterprise presents the opportunity to do something different with search and to realise commercial advantage from it. It is essentially a very large toolbox of search technologies that can be tailored to specific business needs. Fortunately there is plenty of help at hand with experienced technicians and consultant to explore possibilities – and many of them have worked on highly tailored search solutions in other organisations.

The solutions provided by Lucid Imagination are not expensive. The company reckons that a solution will typically come in at around 25% of the cost of most commercial alternatives – although we have no proof of this. It should also be stressed that the cost of the technology is not dependent on database size or the number of documents searched. We would advise that where a more aggressive approach to search is needed that Lucid Imagination might be a useful port of call.

### What it Does

The base technology supporting Lucidworks Enterprise has some fairly special features. First it is not particularly resource hungry, which means a small technology footprint. Indexes are typically 20-30% of the original text size – this is not unique, but it is up there with the leaders. Second it delivers lightning fast performance and can be tuned extensively to meet specific needs. Tuning is often overlooked, but just as with databases, any database administrator will tell you without tuning performance will be orders of magnitude slower than it might be. Lucidworks Enterprise offers a bewildering number of tuning options and it seems likely that even the most obtuse requirements will be satisfied with the same excellent performance.

Functionality options are also very rich and include features such as alerts. These can be set to trigger at regular intervals should new content appear that is relevant to a specific search task. Many search users have tasks that are conducted on a regular basis. A product manager for example might want to see all new content that relates to the products she/he is responsible for. The alerts would provide this information automatically. Implied in this is the ability to create standard searches and have them execute when new content is available.

The manner in which information is delivered is also very flexible. It might be preferred that search results are ordered by their popularity, or categorised as facets, or simply presented in date order. Lucidworks Enterprise is capable of delivering in all these formats and many more – it depends purely upon need.

### **What's Different**

The technicals for Lucidworks Enterprise are about as good as it gets, but this is no plug-and-play solution. The technology is intended for well thought out search applications that are designed to deliver productivity and commercial advantage. A quick scan of the customers will show that commercial advantage is indeed what is often realised. This is no quick and dirty solution to enterprise search – you need to know what you want.

### **About Lucid Imagination**

The company was founded in 2007 and is headquartered in San Mateo, California. It is largely owned by a number of private investors and so financials are not available. Customers include Macys, Getty Images, The Motley Fool, ebay, Sears ...

<http://www.lucidimagination.com>

## Vivisimo

Enterprise Search

ESV009 - April 2011

### Strengths

- ✓ Very productive, easy-to-use interfaces.
- ✓ Wide variety of search methods (clustering, facets, metadata etc).
- ✓ Broad set of information source connectors and integration options.
- ✓ Social search and content alerts.
- ✓ Ability to identify topic experts.
- ✓ Best suited to knowledge intensive work environments.

### Weaknesses

- ✗ No real information on footprint and performance.
- ✗ Less suited to high throughput production environments.

### Advisory

Vivisimo provides particularly user friendly enterprise search technology with support for advanced search techniques. The company boasts 'out-of-the-box' functionality, allowing administrators and search users to gain advantage from the technology very quickly. The Velocity Platform provides the underlying search functionality and this is very rich in configuration options, connectors, search methods and information delivery formats. The real strength of Vivisimo in our opinion is the provision of sophistication wrapped in an easy-to-use, highly productive set of tools.

We particularly welcome the people oriented aspects of Vivisimo. Experts can be found based on the documents they have created and information they have tagged. Social search is also supported with the ability to share search results and create groups.

The highly productive environment provided by Vivisimo includes automatic classification of content, the ability to tag large sets of content in a single action, notification of new content via alerts and the very rich search interface. Vivisimo is ideal for organisations with a diverse set of content, disparate information sources (federated search is supported) and a knowledge intensive work environment. While Vivisimo functionality can be embedded into production applications this is not its real strength.

### What it Does

At the heart of Vivisimo is the Velocity Platform and a number of add-on modules. This provides the necessary connectors to access most types of structured (ie databases) and unstructured information. It also supports excellent REST/SOAP APIs and SOA methods for accessing other applications. Security is also well catered for through support of most of the common authentication and authorisation standards.

The end-user functionality provided by Vivisimo is about as good as it gets. Various forms of navigation are supported including clustering, facets, metadata, by source and combinations of these. Relevancy can be specified by word proximity, synonyms, date and again combinations. Users can set alerts and be notified of new content that is relevant to a particular search, and experts can be found based on the content they produce and the documents they tag.

The Velocity Platform comes ready for 'plug-and-play' with pre-built connectors, reports and display formats. This can be enhanced through a number of additional modules. The Discovery Module adds auto-classification of content and express tagging, where a content set can be tagged in a single action. Vivisimo also offers modules to integrate the desktop into enterprise search and provide search capability on mobile devices.

A number of search solutions are provided for customer experience, supply chain, R&D and Government.

### **What's Different**

Vivisimo has successfully combined ease-of-use with technology sophistication. This means users can familiarise themselves with the technology on an incremental basis while still performing productive search activities. This is a very useful solution for knowledge intensive work environments and the social search capabilities will prove to be particularly relevant.

### **About Vivisimo**

The company was founded in 2000 and is headquartered in Pittsburgh USA. It also has other US offices and offices in the UK, Germany and France. Customers include Airbus, Bupa, Procter and Gamble and Schering-Plough.

<http://www.vivisimo.com>

## Connexica

Connexica targets the end user who might want to search and analyse large amounts of data in a fairly straightforward manner using a natural language interface. Their CXAIR technology provides rich information presentation formats and good performance although we have not been able to get an insight into the techniques used within the search engine.

The Google-type query look and feel will be adequate for many users, but (read our review of Google Search Appliance) it is not particularly sophisticated and may fall short for more complex requirements (for example the ability to specify a faceted search may be necessary in some cases).

We like the Venn diagram presentation of search results, allowing complex relationships between data to be represented graphically, and the dashboards provide a particularly rich user interface.

Connexica is a UK based company founded in 2006.

[www.connexica.com](http://www.connexica.com)

## Exalead

Although Exalead can be used to provide an out-of-the-box search solution, it's true strength is to provide an information access platform for creating highly customised search-based applications. For organisations looking at embedding search functionality into new or existing applications Exalead should be high on the agenda. It's patented results navigation interface, which uses advanced clustering and refinement techniques, also adds to it's pedigree.

Not only does the CloudView platform enable organisations to create custom search applications for unstructured information such as web pages, documents and emails etc, it can also include structured data sources such as databases, CRM and ERP applications for example. This enables organisations to use a search-based approach to querying these sources that is more flexible and user friendly than traditional database queries using SQL commands. In high intensity data applications, this can be used to offload the query process and therefore reducing hardware requirements. The comprehensiveness of Exalead's approach extends its reach beyond search and into areas such as master data management (MDM), information lifecycle management (ILM) and compliance.

Formed in 2000, Exalead is a subsidiary of Dassault Systèmes. It reports over 300 customers including HM Revenue and Customs, Thomson Reuters and The Scottish Government. The Exalead platform also has a strong reputation for powering some very high profile consumer sites such as RightMove, AutoTrader and Friendster.

This is a short review simply because Exalead failed to respond to requests for information.



## Nexidia

Advanced high performance audio and video search including real-time contact centre suite.

Most Enterprise Search solutions on the market today enable the searching of a wide array of structured and unstructured information formats. Unlike their machine readable counterparts, the searching of audio and video content is often not available. If it is, it is most likely to rely only on tags and meta-data associated with the content. Nexidia's strength lies in its ability to index the entire speech directly from audio and video content enabling it to be searched.

Specialist applications of Nexidia's technology can be found in areas such as media, legal, government and intelligence. What we find most interesting, and which will appeal to organisations with contact centre operations, is Nexidia's Enterprise Speech Intelligence Suite (ESI).

When integrated into a contact centre's CRM and telecoms system, ESI converts the speech from all conversations into a text format which can then be searched and analysed. The benefits can be wide ranging such as monitoring and improving agent performance, improved compliance monitoring and competitor intelligence. The real-time processing capabilities of the Nexidia platform can also provide agents with live on-screen prompts during calls. In a sales and marketing environment these prompts may highlight additional products or services to offer the caller to maximise revenue opportunities. In a support or customer service environment prompts to resolutions in a knowledge base can reduce time to resolution and customer satisfaction.

Large contact centres are often found in highly competitive markets such as finance, telecoms and utilities where efficiency, competitiveness and compliance is critical. Nexidia should at least be evaluated by companies operating in these areas.

Nexidia is a privately held company based in Atlanta, USA with a global presence.

[www.nexidia.com](http://www.nexidia.com)