

# The Business Value of Enterprise Search <sup>2009</sup>

A review of cost effective solutions for managers

Introduction written by: Martin Butler

July 2009

Business Value Series – No. 1

This Business Value report discusses the business and organizational issues a technology raises. We avoid buzzwords and unintelligible jargon wherever possible, and favour a straightforward discussion of costs, benefits and risks. If you would like to be notified of future reports in this series please go to our web site and register.

## Enterprise Search - Introduction

---

Only ten or fifteen years ago the task of searching for information was a fairly straightforward matter. Most information was held in transactional databases and various query languages could be used to search for pertinent records. A typical search might involve extracting details of all customers with particular demographics, or with a history of problematical payments. Most of the other information in the organization was stored in filing cabinets - contracts, memos, product descriptions and any other information that was needed for the organization to function correctly. Today most of this information is stored electronically and the problem of information search has grown with the diversity of information sources. The proliferation of email use has in itself created problems. Various regulatory requirements typically mean that emails must be archived and made available for interrogation. Word processing documents may contain anything from details of new product developments through to contracts with customers. Sensitive documents could be stored away under lock and key - today they typically sit on a disk drive allowing anyone to view them that has the relevant permissions.

From a purely economic point of view information search is becoming a more costly exercise, with plenty of scope for inefficiencies, security breaches, high maintenance overheads and unmanageable complexity. It is for these reasons that we stay focused on the basic economics of information search and avoid the three letter acronyms and unintelligible jargon that affects most IT markets. The goal of information search is simple enough - to be able to access information in a timely and efficient manner without incurring crippling costs. This is not as easy as it sounds, although if we maintain our focus on the central task before us we can avoid the unproductive detours that technology in itself might introduce.

In an ideal world information search would be justified using a simple cost/benefit approach. The costs will be relatively easy to estimate, whereas the benefits are much more difficult to assess. Technology simply gives us the tools to do the job, it is people who determine whether information search is a productive activity or an unmanaged free-for-all.

The key to successful enterprise search for many organizations will be simplicity. There are of course areas of every business where complexity cannot be avoided, but it is worth remembering that the people who use search technology simply want a button they can click in the applications they use, and for a meaningful search to be executed.

Estimation of the benefits to be gained from information search is a prerequisite for clear, unambiguous search activity. Some information is so important that lengthy, expensive searches are justified. Other information may have so little value that failure to find it during an initial search may signal the end of the exercise. It is the skill of the people involved that determines whether the benefits obtained from search exceed the costs. This is a management issue that is often overlooked, but as information sources proliferate the need for a managed approach becomes more important.

Some of the problems associated with information search are well illustrated by a typical Google search on the Internet. Most searches will produce tens of thousands of resulting references. The search term we use may have multiple meanings, and only after an initial search will we have the

context to refine the search further. Even so the new search may result in thousands of links, and determining which ones are most relevant becomes a long gruelling task. What we do after this depends on how valuable the information is to us (or our organization). The problem of search costs has been well understood by economists for decades. The classical problem of searching for the lowest cost supplier of an item essentially boils down to not spending more on the search than the savings that would be made as a result of the search. It's a tricky problem that calls for judgement and experience.

Some of us hope that the technology will become sufficiently sophisticated that it can accommodate nuances, ambiguity and significance. This hope is obviously encouraged by the Enterprise Search technology suppliers, but we should be cautious. Getting search technologies to do the donkey work is fine - expecting these same technologies to determine meaning and relevance is folly, and will result greater costs in the long run (as we make erroneous decisions and supply various authorities with flawed information). We cannot, and should not, exclude the people costs involved in information search, since these will typically be the major costs. Information sources have to be managed and maintained. Additional information costs are incurred through the need to establish high levels of security and add additional information that describes the information we are interrogating (metadata creation). The results of information search have to be manually processed and formatted in a way that is presentable to others both within and outside the organization. And to cap it all the people that carry out these tasks are usually highly skilled and expensive.

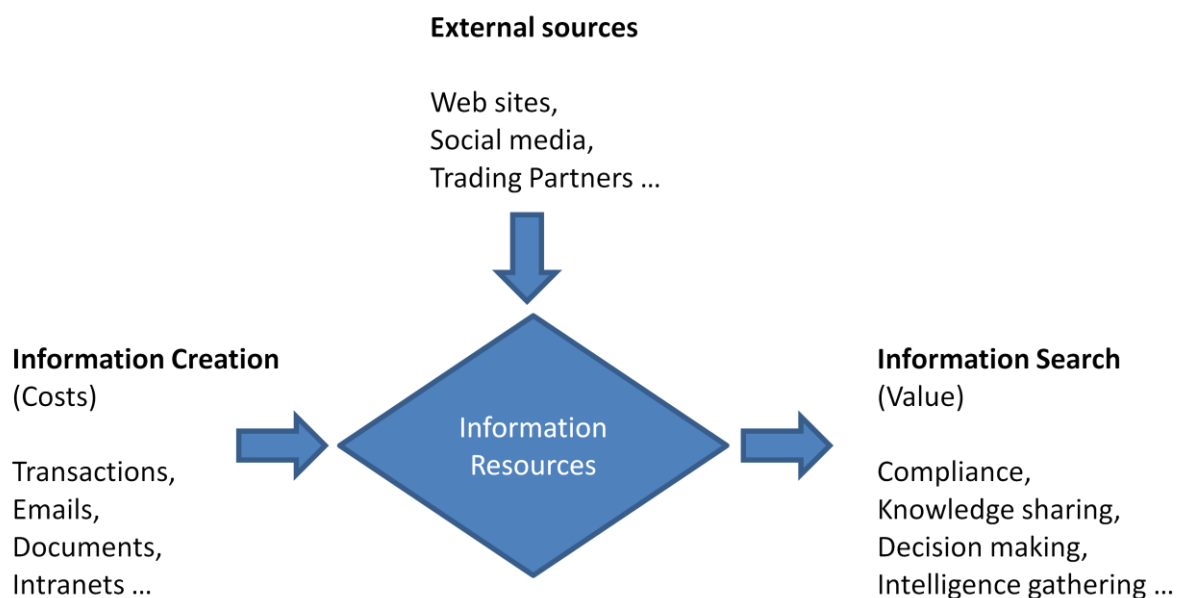
Unlike information creation, which usually occurs in a highly structured environment (data entry and document creation for example), the information search environment is highly unstructured and desperately in need of supporting tools to help professionals determine the efficacy of their search activities. One of the major predictions we make at MBR is the emergence of tools to help information professionals manage their working environments more effectively - they are sadly lacking at the moment.

The economics of information search will become a more pressing issue as information sources proliferate. Internal information sources include transactional databases, email, instant messaging, documents, scanned documents, intranets and even some multimedia files such as voice and graphics. It is however the external sources that are proliferating most rapidly. Organizations may wish to interrogate subsets of each others' information sources - when collaborating with suppliers and customers for example. Outside of this we have web sites and the new social media phenomena such as Facebook and Twitter. Some organizations are already interrogating Facebook as an additional resource in their recruitment process.

We should not expect this proliferation of information sources to cease any time soon. What we need is a sound approach to the costs and benefits associated with using these information resources. There are already signs that the sheer volume of information sources is causing serious inefficiencies to occur in the way we manage our time and resources. Information can be more of a distraction than a useful resource - technology alone will probably just compound the problem.

## Why Search?

When we look at the way organizations use information it is easy to see that information search is easily more than half the picture. While information creation (through data entry, document creation, writing emails etc.) is an unavoidable cost, information search is the primary mechanism for creating value from information. The fact that search has only recently become important demonstrates how little attention it has been given. As information sources proliferate so search will take on much greater importance.



Information search is a much larger part of the way we use information than many of us appreciate. At its simplest we may simply want to retrieve details of a customer given the name and post code. We take this level of search for granted, but even here there are efficiencies that can be realized, particularly when the action is repeated hundreds of times a day. More pertinent to contemporary needs however is text search. For many organizations this will account for eighty per cent or more of the searches they wish to perform. Text is held in documents, emails and databases - bringing these together under one seamless search interface is what most people want. The applications of search technology are as varied as the organizations that use it.

Pharmaceutical companies may wish to search documents, emails and databases to track performance of a drug. Financial services companies may wish to perform searches across multiple financial products for customers. Security agencies use search extensively to home in on suspect communications. Other users may not only want to interrogate internal information sources, but combine these with external sources such as web sites and the information resources of trading partners. This level of search is not easily achieved, and is more a dream than a reality for most organizations.

Individuals may need to search enterprise information resources on an ongoing basis, and if the organization depends upon the skills of knowledge workers, information search capability may be as necessary as a telephone system.

One of the major drivers of search technology has been the need to meet various regulatory requirements. This typically means that a complete, lengthy audit trail needs to be in place so that officials can gain access to information should they need it. The financial services industry has been particularly hard hit by regulatory requirements, and search technologies are widely used.

## The Search Cost Spiral

As with most things in life eighty per cent of the benefits typically derive from twenty per cent of the functionality. This is certainly true of enterprise search technologies, and organizations would be wise to consider the cost spiral that can manifest as greater sophistication is required. The basic trade off is this - greater sophistication should make the search process more efficient, but the sophistication should not cost more than the benefits it delivers. To understand this point we have to use some of the terminology associated with search technology.

An ontology is essentially a dictionary that conveys the meanings of terms used within an organization. The word 'track' will have a different meaning if you are in the music industry than it will if you are a sports equipment company. Creating and maintaining an ontology is a labour intensive process - and it is always incomplete. Taxonomies are also used widely by search technologies, and these are structures for classifying information. A variety of names used in most organizations will refer to products, and in the organizational taxonomy these names would appear under the 'product' classification. Once again this is a labour intensive process that is ongoing and never completed. We should also consider the cost of tagging information. For effective document search we might need to create keywords, or abstracts.

Multi-media files are often interrogated on the associated tagging and not by interrogating the file itself (which is unintelligible to most search technology). Once again someone has to do this. The process of adding meaning to information can be considered a process of adding semantic content.

We should also consider security as one of the major cost overheads associated with search technology. Every document, email, item of data in a database, and all other information instances should have permissions set up. The classic test to see if this has been done in your organization is to execute a search on the term 'confidential' - but we didn't tell you to do that.

All these costs are real and significant, and will dwarf the actual cost of the technology. It is for management to decide how much value they expect the availability of sophisticated search to deliver and then determine the costs they are prepared to bear.

For most organizations the best efficiencies will be realized through a fairly 'crude' use of the technology. This will entail greater manual post processing of search results, but will avoid costly diversions into realms such as semantic meaning. The security issue always needs addressing, but the most efficient way to implement security is simply to make sensitive documents unavailable, and a resource that must be specifically requested.

## The Mediocre Search

Search results can be broadly categorized as failed, successful and borderline. We should have no problem with the first two. If a search fails we can try a new strategy or simply abandon the task. A successful search is what we all aim for and is a conclusive positive outcome. The mediocre search outcome is quite easily the most problematical. This would be characterized by just enough search results to indicate value in the search process, but insufficient results to compile the necessary report or reach a firm decision. The problem with this type of outcome is that it does not warrant abandonment of the process or a new strategy, but neither does it lead us to a conclusion. It is here that costs can escalate well beyond the value that might be derived from a search.

The mediocre search should be acknowledged for what it is with an extra keen focus on weighing the ongoing costs against the possible value. Of course the natural tendency is just to keep going, in the hope that we find some new thread that takes us to where we want to be. Mediocre searches tend to remain just that, and only exceptionally will they deliver the results that are needed. Once again this is a matter for human judgment and experience - no amount of technology will tell you that a mediocre search is unlikely to lead anywhere.

## What does the Technology Do?

We've all used Google search and are familiar with the interface. Enterprise search technologies typically provide a similar interface where phrases can be entered and the search executed. The results are often presented in a similar manner to Google - a list of references ordered by relevance. But that is where the similarity ends. Users of enterprise search technologies may be disappointed by the results a search delivers. Google has access to billions of web pages, whereas a search within the confines of the enterprise may offer a relatively small number of references. We should also remember that the information sources we access through enterprise search are not linked in the same way that web pages are - it's a wholly different environment. Where one search on Google may suffice, we may have to conduct several searches within the enterprise to get what we want.

Despite all the hype and complex terminology, what we would all ideally want is an enterprise Google. Enter a search and browse a rich set of references. Enterprise search emulates this model, but it is still early days for the technology, and because of the diverse nature of the information we need to search, the results are less reliable.

## Sharing Knowledge

The term Knowledge Management is not as widely used today as it was a few years ago. The basic idea was that individuals would freely share their knowledge via some repository, and the organization as a whole would benefit. This 'managed' approach never really worked, people are much more protective of their knowledge and experience than we often assume. Enterprise search on the other hand does allow knowledge workers to dig around for available information and share knowledge in a more informal way. This will be a significant benefit for some organizations and may become the single most important benefit, particularly where knowledge and information sharing is crucial to the welfare of the organization.

It is worth pointing out that knowledge sharing 'by schedule' has been shown to be particularly inefficient. Opportunity costs and the cost of setting up a more formal environment are often greater than the benefits derived. Where knowledge sharing and collaboration happen on an 'as needed' basis the benefits easily outstrip the costs.

## Confidence

A common flaw in the use of information search technologies is overconfidence in the results. There is, unfortunately, no substitute for understanding and experience. Executing a search on a topic that is unfamiliar will inevitably generate results that are at best misleading and may be simply inaccurate. An experienced eye would spot the mistakes, but blind faith in the content referenced by a search will result in erroneous conclusions. It is best to have the attitude that the search technology can automate the process of sifting through vast electronic 'filing cabinets', but that the results need to be checked by someone with domain experience.

## Summary

Information search is a generic function in all organizations that is worthy of planned implementation and managed usage. As information sources proliferate, so the need to present a single search interface will become more important. While the search function has often been treated as an afterthought, it is deserving of at least as much attention as information creation processes (data entry, document creation etc.) and is the primary mechanism for creating value from an organization's information resources.

We expect information search to become a major issue in all organizations as the need to reduce information costs becomes a more pressing issue. Only careful analysis of search needs will result in a level of investment in the semantics of the information an organization possesses that is appropriate. Some information may need very little preparation, while other information may be central to the way the organization works and require lengthy processing (creation of tags, taxonomies, ontology, abstracts etc.)

The guiding principle is to always ensure that the benefits derived from information search exceed the costs associated with the search. This may seem obvious, but it is quite common for information workers to spend inappropriate amounts of time searching for information that has little real value. Recent research on the benefits of knowledge sharing reveal that costs can quickly exceed benefits.

Expect information search to become a much larger component in your organization's IT armoury, and with it will come a need for constant awareness of cost and value.

# Industry Perspective Section

# Oracle Secure Enterprise Search

---

## Addressing the Top 10 Issues in Enterprise Search

With trends in distributed enterprise structured and unstructured content, storage grids, and virtualization, enterprise search takes on critical importance as the unifying engine for content and process exploration. Search provides improved visibility across diverse data sources and applications, acceleration of business processes, timely access to relevant data points, and the ability to analyze complex situations more effectively.

Enterprise search technology is growing in importance as corporate repositories and intranets grow in size and complexity. Across most industries, corporate content is more than doubling each year. Important and sensitive content can now be found anywhere – in email attachments, on dedicated content management servers, across distributed file systems or within private workspaces – as well as on internal and public facing web sites. Effective enterprise search is now required to allow users to convert distributed pieces of information into operational advantage.

However, by providing a new means of accessing content, enterprise search capabilities can potentially conflict with increasingly strict information management mandates implemented by organizations to comply with governmental regulations and corporate governance policies. Financial information, trade secrets, confidential customer records and sensitive personal information all must be safeguarded. As a result, many organizations are seeking to establish more consistent and reliable system for controlling information access and retrieval. Because search technology makes information more accessible it can quickly expose underlying deficiencies in how organizations manage sensitive content.

Oracle Secure Enterprise Search is a stand-alone enterprise search product incorporating best-in-class indexing, crawling, and security capabilities to create a reliable, comprehensive search solution to meet organizations' needs for information access and enterprise content security. Oracle Secure Enterprise Search uniquely addresses ten critical enterprise search issues:

1. Relevance
2. Comprehensiveness
3. Security and Privacy
4. Policy and Compliance
5. Access Control
6. Personalization
7. Federation
8. Search As A Service
9. Extensibility
10. Support

## 1. Relevance

Relevance measures how closely search results match user expectations. A search with high relevance will successfully return the documents the user intended when specifying the search term. When considering relevance, it is critical to understand that enterprise data systems are fundamentally different from the public Internet. Enterprise data systems are typically developed privately and in parallel with each other and do not have a rich topology of links across systems. Thus an enterprise search engine must determine relevance without some of the structural clues available to Internet search tools. This requires a much more powerful capability to determine relevance.

Enterprise content also can have unique meanings for terms that vary from department to department, or division to division, or even person to person. So, a search must also be tunable – to ensure that the right results reach the user first. In the end, relevance is a combination of pure calculation plus corporate experience – helping to surface the right content with each search. In addition, enhanced search capabilities can cluster results in context, and use metadata such as names, dates, and content location to sort or filter results. These capabilities help a user find the right results set from a very simple search.

## 2. Comprehensiveness

One of the primary benefits of enterprise search is search consolidation. Rather than searching individual repositories, a user can search one, some, or all enterprise content through a single search. Enterprise search must provide comprehensive access to content in servers and databases. It must also be able to reach the underlying content in complex business applications, including hosted applications. And for content security, your enterprise search solution must be able to integrate with each repository's authentication scheme.

An enterprise search solution needs to reach each enterprise repository and index its content. In addition to content location, search must be able to index critical content types. This includes files in file systems, documents and content management systems, structured data in databases – even business data in business applications. Specific file types to index include text files, databases, desktop applications output, and compressed files.

## 3. Security and Privacy

While almost every user is focused on delivering value to their organization, there are always exceptions. The instant results offered by search can make curious and ill-intentioned forays into sensitive data more tempting. Of course, sensitive data sources either offer their own security or are integrated with access control systems. It is not the role of enterprise search to set control access policies but search must ensure that its activities enforce those policies. This level of enforcement is critical to ensure corporate security and the privacy of individuals.

Also consider the search index. Search is a matching task that compares a user request with a text index and delivers results. How is this index built, where is it stored and how is it protected? A large search index stored in a file system is an obvious target because it centralizes information from many locations. To ensure security, you must control access not only to source documents but also

to the search index that summarizes them. Otherwise, search can become a weak link in your IT security chain.

#### 4. Corporate Policy and Compliance

Today's information access policies must ensure compliance with privacy, piracy, or other laws and regulations. How can you tell if your systems are in compliance with your policies?

A well-designed enterprise search provides managers with the tools to separate public data – accessible to anyone inside the firewall, as well as private data that only named users can access. A regular sweep for sensitive content can tell quickly if critical content about sales or acquisitions is being made accessible in unsecured locations. Enterprise search can also ensure that enterprise resources are not being used to store inappropriate or unlicensed content.

#### 5. Access Control

To control user access, it is essential to ensure that users authenticate before using the full power of enterprise search. Your enterprise search system must be able to link to one or more authentication schemes available inside your enterprise, and ensure that no information is presented until the user has authenticated to each repository. To speed access, you can choose to use a single authentication system as the signon and link to others after a user has logged in. You can also limit the scope of search results to public-facing content – using the same search engine for both private and public searches – so long as you are clear on how the search system distinguishes between unauthenticated and authenticated users.

#### 6. Personalization

On the public Internet, most people prefer anonymity when searching; however, within the enterprise your identity matters. You have an employee ID and login. You are affiliated with a management structure, with subordinates, managers, and peers. You are affiliated with projects and activities, customers and partners. Your affiliations help identify your interests – aligning your need for information with the needs of your network.

Enterprise search results can be improved greatly by incorporating personal information into the results process. By relating your searches to the searches of your network, you can gain the benefits of automatic relevance boosts when your peers select specific information and search terms. At the same time, your history of choices can guide the current results list.

#### 7. Federation

Who wants to throw out something that works? Every enterprise already has a working search solution of some sort. It may be tied to a specific repository. It may be accessible only to a specific set of users. But, it is designed for information retrieval for that repository, and it works today. To gain the benefit of a unified search, you want a single search to access all repositories – even if you use an incumbent search tool instead of the enterprise search solution. Federated search can be

quite complicated, requiring capabilities such as advanced authentication, ranking of relevance across multiple repositories, and disaggregation of results from repositories with unique content.

## 8. Search As A Service

Enterprise search can extend the search paradigm to create a service for automated processes to retrieve information. Just as SQL has provided this capability for data in databases, search can enable processes to access and deliver unstructured and structured content from a variety of sources. As data and files are linked into workflow, enterprises have been challenged to find tools to retrieve both database data and structured content. Business processes will increasingly use enterprise search as a programmatic tool to retrieve business information.

## 9. Enterprise Scale and Scope

Most enterprises have begun using search to access information in specific repositories.

However, a number of technologies provide content aggregation. Portals can link to content from multiple repositories. RSS feeds also represent multiple disparate data sources. Be prepared for the scope of your enterprise search problem to grow, and for more uses and users to surface. This issue also relates to scale – an enterprise search solution must be able to scale to the needs of your enterprise. If it cannot, your desire to create a unified search will be replaced with the cost of supporting multiple redundant searches. Finally, if your organization is global and/or multilingual, your search solution must address the needs of speakers of specific languages.

## 10. Support

As you consider enterprise search, you now need to consider a search vendor that can address the requirements of multiple repositories, multiple authentication schemes, security, reliability, and results management. Where enterprises have deployed a unified search, they have discovered that this capability grows in importance and value to becoming a mission-critical component of their success. You need a vendor that can deliver support for your search implementation how and when you need it, to ensure that your enterprise search is always secure, reliable, and available to your users. Oracle provides this caliber of support.

## CONCLUSION

The need for a single, comprehensive enterprise search solution will only grow as users continue to generate and store content in various file servers, databases, workspaces, emails, and a host of other repositories. Oracle Secure Enterprise Search delivers a comprehensive, secure, and fast solution designed for the needs of organizations large and small that want to find information quickly and improve their productivity.