

**Connect ▶ Find ▶ Act**

# **SharePoint Search and Microsoft Search Server 2008**

**Connecting Search to Enterprise Systems**

January 2009

**Martin Muldoon**

VP - Marketing

## Abstract

This white paper covers the business value of connecting search to Line of Business applications, utilizing SharePoint Search and Microsoft Search Server 2008's ability to index Line of Business data, and explore **BA-Insight Longitude's** ability to extend SharePoint and Search Server's capabilities and enhance the appeal and reach of search services across the enterprise.

The information contained in this document represents the current view of BA-Insight on the issues discussed as of the date of publication. Because BA-Insight must respond to changing market conditions, it should not be interpreted to be a commitment on the part of BA-Insight, and BA-Insight cannot guarantee the accuracy of any information presented after the date of publication. This document is for informational purposes only. BA-INSIGHT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, AS TO THE INFORMATION IN THIS DOCUMENT. BA-Insight may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from BA-Insight, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

## Contents

<b>Enterprise Search Comes of Age</b>	<b>4</b>
The ROI of Enterprise Search	4
Why Most Search Initiatives Fail	5
<b>Technical Challenges and Basic Features Required to Connect Enterprise Systems</b>	<b>6</b>
Security	6
Advanced Meta Data Support	7
Pre-Built Connectors	7
Connecting to Custom Systems	7
<b>Microsoft Enterprise Search Technologies</b>	<b>8</b>
Microsoft Search Server 2008	8
SharePoint Server – Business Data Catalog	9
<b>BA-Insight’s Connectors for SharePoint and Microsoft Search Server</b>	<b>11</b>
BA-Insight’s Connector Security	11

## Enterprise Search Comes of Age

Companies and organizations of all sizes and industries have begun, or are planning to deploy, search technologies over the course of the next few years. The sudden widespread interest is based on a number of factors:

1. Technologies have matured and become more affordable. Companies such as Microsoft are releasing powerful new search products such as Microsoft Search Server 2008. These technologies are cost effective, and relatively easy to deploy and maintain.
2. Companies and Organizations have put in place personal productivity tools (Microsoft Office), communication tools (email), and in many cases Business Process Automation tools (ERP, and CRM). With these critical systems in place, companies can now begin to focus on helping users collaborate and find the information they need to be as productive as possible. Simply stated, with so many tools in place that produce content, it's only natural that finding content and leveraging it becomes an imperative.
3. Google has left little doubt in the mind of users of the importance of accessing information quickly and reliably.
4. The ROI of a search initiative is huge, and increases with each system connected to the search engine.

### The ROI of Enterprise Search

Many studies have been conducted that calculate the ROI of Enterprise Search. Sue Feldman, from the IDC's has conducted two extensive studies<sup>1</sup> which are probably the best known. To summarize her findings, the ROI of search is based on three factors. One tangible, and two that are less so:

1. Time lost searching for information
2. Time lost re-inventing the wheel
3. The benefit of having timely information to improve decision making.

The first is easy to understand, quantify, and argue to management when making the case for budgeting a search initiative. The average worker spends roughly 2.5 to 9 hours per week searching for information. Clearly during this time, nothing of value is being created, yet that worker is drawing salary. Simply multiplying the number of hours a year this occurs times the average hourly salary times the number of employees in the organization yields a quantified loss of productivity which is typically jaw dropping. The extent to which search technology can lessen this is arguable, but even conservative numbers yield a huge payback.

---

<sup>1</sup> The High Cost of Not Finding Information – Sue Feldman, IDC, 2001  
The Hidden Cost of Information Work – Sue Feldman, IDC, 2004

## Why Most Search Initiatives Fail

One of the biggest mistakes Enterprises and Organizations make when deploying a search solution is to equate the successful deployment of the search technology with project success. While this might sound reasonable, it's important to realize that search technology is different than other types of technology infrastructure which requires different success criteria. As an example, consider deploying an email system. After the system is deployed, people use it without question. They don't have a choice. Search on the other hand is different in this regard. If a user finds a search system to be poor they simply won't use it. They'll spend their 2.5 hours per week searching for information manually as they had done in the past. The key take away from the way in which the ROI of search is calculated is that for search initiatives, ROI is driven by use of the system. The more the system is utilized higher the return on the investment of the project

A logical question to ask is why would someone not use the search system? The answer may surprise you in its simplicity. If the data they use to do their job on a daily basis can't be found in the system, they will see little value in using it. This occurs if the data that's relevant to them is not being indexed by the search system.

**A large number of employees spend their entire day working with data found in ERP, CRM, and database systems. If this data is not included in the scope of an Enterprise Search initiative, the search system will not be used by these individuals.**

When one thinks of an Enterprise Search initiative, typically the scope of the project focuses exclusively on document content. Content from Databases, ERP Systems, CRM Systems, and other sources containing structured content are often not included in the scope of the project. There are a number of reasons why this is the case:

1. Search initiatives are typically IT led projects. While IT owns File Systems and other ECM systems, ERP systems and the like are managed by a different group. In an effort to reduce project complexity and risk, IT only focuses on those systems they control
2. IT often intends to connect to Enterprise Systems as a Phase II initiative, but other projects take priority and Phase II never occurs.
3. A failure to recognize the importance of connecting to these systems.

The obvious implication of not indexing data from these systems is that the workers that need this data to do their jobs will see less value in the search system, if any at all.

## Technical Challenges and Basic Features Required to Connect Enterprise Systems

This section covers some of the technical challenges associated with exposing enterprise content through search

### Security

It goes without saying that Security is one of the most important features a search system can provide.

#### Index Level Security Trimming

In an Index Level Security model, the security settings of each piece of data indexed by the search system is read, and then stored in the index. This is the most commonly used way to handle security in search systems. The benefit of this approach is performance. When a user executes a query against the search system, the content matching the query is generated, and then culled down to the set that the user has permissions to see. Since the security is stored in the index, a secure query result can be generated in less than one second. The drawback of this security model is that security settings are only as up to-date as they were in the last crawl. Any changes to the security on a piece of data will not be reflected until the data is indexed again by the search system. Despite this shortcoming, Index Level Security is the only model that can produce Enterprise Level Scalability, with millions of records indexed.

#### Real Time Security Trimming

Real time security trimming overcomes the issue of outdated security between scheduled crawls by performing a real time security lookup against the content source rather than against the index at the time the query is executed. The current security setting on the data will be fully honored, and the search result item will be removed from the list if the user does not have access to it. The shortcoming of this security model is that it is not scalable when dealing with large sets of records. The real time security lookup will result in a noticeable delay in the time a query is executed until the result is returned.

#### Recommended Security Trimming

Combining index level security trimming with real-time security trimming address both the scalability and freshness requirements of security information. The bulk of search result items are pre-filtered based on index-level security information, and the most recent changes between scheduled crawls are picked up real time, with minimal impact on the content source and response time.

## **Advanced Meta Data Support**

Meta Data can significantly improve the search experience of a user by enabling the search engine to server up more relevant results, and provide the user with a means to refine the search result both before and after the search is run. Without metadata, a search system is severely limited because users typically enter only two or three keywords. A ranking algorithm, no matter how sophisticated, can do little with such a query, other than return the list of documents containing these keywords. In this situation, the search result will often contain many thousands of documents leaving little chance of finding relevant content.

The good news is mostly all modern search engines support metadata at some level. The most basic will index metadata associated with the document itself, such as author, last modified date, file type, etc.

More sophisticated systems have the ability to index metadata from different applications and connect any related metadata in each system. There are numerous scenarios where this occurs. In a law firm, for example, client information can be found in the document management system, the billing system, and the CRM system and needs to be merged into a single indexed record.

## **Pre-Built Connectors**

Creating a map between the search system and the enterprise system you wish to connect to can be very time consuming. Anyone who has worked with a database system that was created by someone else knows how big a challenge it is to understand the schema of the database. In more complex enterprise systems like ERP systems, for example, the system can consist of tens of thousands of tables. Ideally, the search vendor should supply connectors to common enterprise systems, or have partners who do this.

## **Connecting to Custom Systems**

Many if not all, organizations have developed custom applications internally that contain valuable data that should be searchable. The search system should make connecting to enterprise systems as easy as possible, through the use of graphic user interfaces.

## Microsoft Enterprise Search Technologies

Microsoft offers a number of different search technologies, ranging from desktop search, intranet, and internet search. The company is moving quickly to rationalize and integrate them. This paper focuses Microsoft current Intranet offerings and how they enable companies to connect to enterprise data from the search system.

### Microsoft Search Server 2008

Search Server is based on the same technology found in Microsoft Office SharePoint Server (described in the next section). One of the key features of Search Server is Federated Search which enables Users to access enterprise content from within a Search Server search result.

#### Search Server's Federated Search

Federated Search enables a user's query to be forwarded to an external content repository where it is processed by that repository's search engine. The repository's search engine then returns the results to the search server. The search server formats and renders the results from the external repository within the same search results page as the results from search server's own content index. By using federated search to return search results:

- You can take advantage of a repository's existing search engine. For example, you can federate to an Internet search engine to search the Web.
- You can optimize the content repository's search engine for the repository's specific set of content, which might provide better search performance on the content set.
- You can access repositories that are secured against crawls, but which can be accessed by search queries

To leverage Federated Search a connector must be created for the systems search engine that you wish to search against. This involves exposing the systems search result as a web service. A detailed description of how to build connectors, can be found on the Microsoft website<sup>2</sup>

#### Federated Search Security

Federated Search queries for data against an existing search system. The security of the result is completely dependent on that of the connected system.

#### Advanced Meta-Data Support

Microsoft Search Server does not have advanced metadata support.

#### Pre-Built Connectors

Microsoft is relying on partners to support it's the companies connector technology. A list of partner offerings can be found on the Microsoft Website<sup>3</sup>

---

<sup>2</sup> <http://msdn2.microsoft.com/en-us/library/bb931083.aspx>

## Connecting to Custom Systems

Creating connectors to custom systems is a coding effort. Microsoft does not currently offer a graphical user interface to do so.

## SharePoint Server – Business Data Catalog

Microsoft SharePoint Server contains a search engine that is more sophisticated than that of Search Server 2008. The component within the search system that enables SharePoint to connect to enterprise systems is the Business Data Catalog, or BDC which is what it is typically called.

The components the BDC provides are multiple configurable Web Parts for listing and displaying the data, a search indexer (protocol handler) and a search query security trimmer. The BDC operates off of xml definition files which are used to define the connections and entities in the LOB system. The BDC is the perfect choice for indexing and working with small data sources with minimal security requirements and only structured data.

## SharePoint Security Model

The SP Search system implements two forms of security by which search results are trimmed:

The first form of security is by standard **ACLs (Access Control Lists)** which is the most familiar as it is how the Windows file system determines if you have access to a document or not. During the crawl process the ACL's of items that are being added to the index are determined and added along with the item. The query engine uses these ACLs that are stored in the search database to determine quickly if a user should be allowed to see that item in the results.

This security method has been the standard for a while in the MS Search products and is very fast. When connecting new systems (like Documentum) a custom protocol handler would be created that knows how to map the security in Documentum to standard Windows ACL. For instance if a user in Documentum is allowed to access a particular document then the protocol handler will need to map that user id to a valid Active Directory user and create a read privilege ACL for that user and add it to the items security ACLs.

All users and groups would need to be mapped also and added for each item to ensure proper security. Note: This security model has a flaw in that if the security changes it will not be picked up until the next incremental crawl which may not happen for hours or days. Also the BDC (Business Data Catalog) does NOT support ACL based security and prior to the addition of the next form of security, the real time Security Trimmer, the BDC had no security for its search results.

---

<sup>3</sup> <http://www.microsoft.com/enterprisesearch/connectors/federated.aspx>

The second form of security is called **Real-Time Security Trimming** and is completely separate from the ACL based security above. It can be applied in conjunction with ACL security to provide an added check to ensure that changes to security since the last crawl are adhered to, or as in the case of the BDC it provides the primary and only means of security. Basically after search results are compiled during a query the items are individually compared to a set of rules to determine if they have a Security Trimmer registered. The ones that do have a security trimmer registered (as in BDC items) are grouped into an array and passed into their respective Trimmers.

The Security Trimmers validate the security of the items and return back an array with a simple true or false for each item indicating whether they are allowed or not. Depending on how the Security Trimmers were written they can be a source of performance contention, as in the case of the BDC one where each and every item is individually validated which could mean hundreds of database queries. As search results can have combined results from multiple sources there may be more than one Security Trimmer involved in each query.

### **Advanced Meta-Data Handling**

The BDC does not offer advanced meta-data handling, merging multiple metadata across multiple data sources into a single index record, or merging metadata with documents.

### **Pre-Built Connectors**

Microsoft offers pre-built connectors to a number modules in SAP and Siebel. Partners such as BA-Insight<sup>4</sup> provide connectors to all major enterprise systems.

### **Connecting to Custom Systems**

Microsoft does have a graphic user interface to create custom BDC definitions to connect to custom systems.

---

<sup>4</sup> <http://www.ba-insight.net>

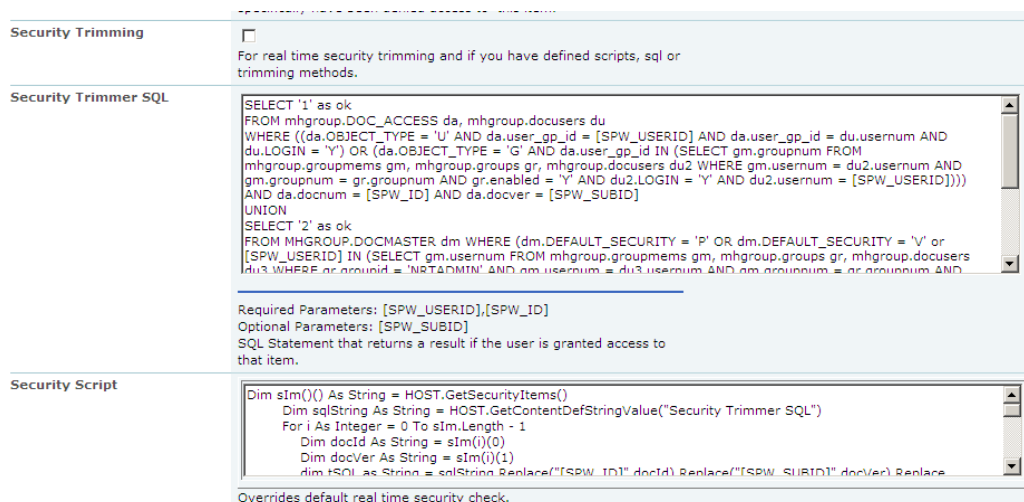
## BA-Insight's Connectors for SharePoint and Microsoft Search Server

This section briefly describes the capabilities of Longitude for SharePoint and Search Server as it relate to connecting to enterprise systems.

### BA-Insight's Connector Security

BA-Insight Connectors provide the Administrator with unparallel levels of security by enabling both index level security and real time security trimming models if desired. With both index and real time, administrators are assured that users only see what they have access to regardless of when SharePoint finished its last crawl.

Unlike the BDC, which initiates a query for each record its performs a security check on, BA-Insight implemented the security check in batch, which is to say, hundreds of records can be checked with a single query, resulting in superior performance, and negligible impact to the user running the search.



**Screenshot:** security trimming can be defined as an SQL statement and/or VB Script

### Advanced Meta-Data Handling

Longitude enables meta-data to be mapped between any number of systems. Mapped meta-data is stored in the index for optimal performance. With additional metadata describing the search record, users have the ability to refine search results in a number of different ways that are more intuitive to them. For example, they can use the Faceted Search capabilities of Longitude to drill into search results after executing the search. Or, prior to running the search, they can use the SharePoint Advanced Search Page or Longitude Parametric Navigation features.

Longitude can also merge metadata with documents into a single index record. Many custom intranet or document management systems are built on top of SQL relational databases, with files typically stored on a file system.

**Generate Metadata**

Handler: Class (SPWorks.Search.Common.SQLSource) found for content type.

Auto create managed Property Prefix

Property	Metadata Description	Content Type	Active	Managed	Alerts
delete SPW_SICKLEAVEHOURS	SickLeaveHours	Integer Metadata	Yes	No	
delete SPW_SALARIEDFLAG	SalariedFlag	Boolean Metadata	Yes	No	
delete SPW_HIREDATE	HireDate	Date/Time Metadata	Yes	No	
delete SPW_GENDER	Gender	Text Metadata	Yes	No	
delete SPW_MARITALSTATUS	MaritalStatus	Text Metadata	Yes	No	
delete SPW_BIRTHDATE	BirthDate	Date/Time Metadata	Yes	No	
delete SPW_TITLE	Title	Text Metadata	Yes	No	
delete SPW_FIRSTNAME	FirstName	Text Metadata	Yes	No	
delete SPW_MIDDLENAME	MiddleName	Text Metadata	Yes	No	
delete SPW_LASTNAME	LastName	Text Metadata	Yes	No	
delete SPW_EMAILADDRESS	EmailAddress	Text Metadata	Yes	No	
delete SPW_EMPLOYEEID	EmployeeID	Integer Metadata	Yes	No	
delete SPW_EMAILADDRESS1	EmailAddress1	Text Metadata	Yes	No	
delete SPW_CONTACTID	ContactID	Integer Metadata	Yes	No	
delete SPW_EMAILADDRESS2	EmailAddress2	Text Metadata	Yes	No	
delete SPW_MANAGERID	ManagerID	Integer Metadata	Yes	No	

**Screenshot:** Longitude creates metadata in SharePoint Search and Search Server automatically

### Pre-Built Connectors

BA-Insight’s product Longitude provides connectors to the most popular enterprise systems. These include:

ERP, CRM, Databases	ECM
SAP	Microsoft Exchange Private Mailboxes
Oracle	EMC Documentum
PeopleSoft	Interwoven Worksite
JDEdwards	Open Text Hummingbird
Siebel	IBM FileNet
Custom Applications	

### Connecting to Custom Systems

Longitude enables administrators to connect to custom applications via a simple, intuitive, graphical user interface. Debugging/Testing

Longitude enables administrators to test their connectors before deploying them to a SharePoint Search or Search Server environment with full simulation capabilities. The administrator sees exactly what data and security information the connector is pulling and may adjust the connector instantly without having to perform a full crawl. By contrast, the SharePoint Search and Search Server crawl logs are cryptic to understand, which makes the development of a connector without the Longitude connector framework quite challenging.

**Indexer Test Bench**

Handler Class (SPWorks.Search.Common.SQLSource) found for content type.

Maximum Returned Results:   Include Trace  Test Incremental  Do Security Trimming

CrawlStarted:12:54 PM For Item:30

Items Found:5

More Exist:False

Returning Results:

Items Found:5

More Exist:False

Returning Results:

Item:1

```
FilePath: E:\samplefiles\downloads\b094db6f-8e1c-4456-9122-dbfcd761b20a.pdf !!!DOES NOT EXIST or Not running local
isPublic: True
extension: pdf
flagDelete: False
Last Update: 11/30/2007
Property: spworksbase_class Type: STRING Value: SIMPLEDOC
Property: spworksbase_id Type: STRING Value: 1
Property: spworksbase_extension Type: STRING Value: pdf
Property: spworksbase_contentid Type: INTEGER Value: 30
Property: spworksbase_public Type: BOOLEAN Value: True
Property: 9 Type: STRING Value: http://www.corel.com/futuretense_cs/ccurl/2276-151_WPO11_CaseStudy_LawFirm_STARS_ENG.pdf
Property: 2 Type: STRING Value: A busy family <b>law</b> <b>firm</b> relies on intelligent <b>software</b> to draft <b>...</b>
Property: SPW_URL Type: STRING Value: http://www.corel.com/futuretense_cs/ccurl/2276-151_WPO11_CaseStudy_LawFirm_STARS_ENG.pdf
Property: SPW_RESOURCECENAME Type: STRING Value: 2276-151_WPO11_CaseStudy_LawFirm_STARS_ENG.pdf
Property: SPW_TITLE Type: STRING Value: A busy family <b>law</b> <b>firm</b> relies on intelligent <b>software</b> to draft <b>...</b>
Property: SPW_DOCUMENTFILENAME Type: STRING Value: b094db6f-8e1c-4456-9122-dbfcd761b20a.pdf
Property: SPW_DOCUMENTTYPE Type: STRING Value: PDF Document
Property: SPW_DOCUMENTSIZE Type: INTEGER Value: 62794
Property: SPW_DOCUMENTPAGECOUNT Type: INTEGER Value: 2
ACL: Everyone(S-1-1-0)=True,
```

**Screenshot:** result of test simulation for 5 indexed records



**United States**

445 Hamilton Avenue, SUITE 1102  
White Plains, New York 10601  
United States of America  
Tel | 914.220.8395

**Europe**

Vester Voldgade 106,  
DK-1552 København V  
Denmark  
Tel | +45 2161 6500

Web | [www.BA-Insight.net](http://www.BA-Insight.net)  
Email | [info@ba-insight.net](mailto:info@ba-insight.net)