What’s New in Search for SharePoint 2013

Top 5 Key Changes in Architecture and Search
Michael Himelstein
Director, Search Consulting

Mike Himelstein, has nearly ten years of practical experience developing, deploying, and architecting search-based applications. Mike has advised hundreds of the largest companies in the world and worked directly as a trusted advisor and architect for major finance, media, and healthcare organizations, helping them to shape their strategy around unified information access.

Prior to joining BA Insight, Mike was the Technology Solutions Manager in the Enterprise Search Group at Microsoft. Mike came to Microsoft through the acquisition of FAST, where he had been the Technology Solutions Manager for the NY Metro area since 2004.

Read Mike’s column on the search experts’ blog at www.DoMoreWithSearch.com
Email: Michael.Himelstein@BAinsight.com
Overview & Audience

Architecture & Topology Changes
One Search Engine

Crawl & Content Processing Changes
Anonymous Crawl
Asynchronous Content Crawl
Visual Metadata Extraction
Company Names Entity Extraction
Result Sources

Query Execution and Page Changes
Ranking
Query Spell Correction
Query Rules
Pre Query Suggestions
Post Query Suggestions

Search Results Interface Enhancements
Result Types
Refinement
Results Blocks
Content by Search
Office Web Applications

Looking Ahead
Should I Wait for SharePoint 2013, or Deploy on 2010 and Upgrade Later?
Buckle Up
Overview & Audience

This e-book is designed for IT professionals to provide them with insight into the upcoming release of SharePoint 2013. This release has powerful new features that enhance and simplify the overall search process and information rendering. The features outlined within this e-book are key components that can be used in the deployment of enterprise search applications.

In this e-book, we will review and explain five key feature areas that will have a dramatic effect on your ability to build, leverage, and extend the search features in SharePoint 2013. These features will radically change how you think about search and the SharePoint platform. These changes now provide a platform for new and exciting applications and enterprise search apps.

These new changes are as follows:

- Overall architecture/topology changes
- Crawl and content processing changes
- Query execution and processing changes
- Search results interface
- Office web applications

Microsoft has been adding 20,000 SharePoint users per day, every day for the last five years, according to Jared Spataro, Director of SharePoint Product Management. That means 7.3 million new SharePoint users every year.

—Mary Jo Foley for All About Microsoft, ZDNet, March 24, 2011
CHAPTER 1

Architecture and Topology Changes
One Search Engine

With the release of SharePoint 2010, and specifically FAST Search for SharePoint 2010, Microsoft announced its intent to lead the marketplace when it comes to collaboration and information access. The acquisition of FAST Search and Transfer in 2008 was regarded by the industry as a major step forward in taking the lead in the enterprise search marketplace. The incorporation of FAST within the overall SharePoint 2010 architecture allowed organizations to leverage enterprise class search capabilities in a platform that was within the cost and budget requirements of today’s enterprises.

Unfortunately, the acquisition occurred midway between release cycles which forced Microsoft to determine which features would be available in the wave 14 release (SharePoint 2010), and which features would end up in the next release.

With the release of SharePoint 2013, the full realization of Microsoft’s investment in FAST Search and Transfer is now evident. The feature set in SharePoint 2013 takes enterprise search to a whole new level.

The first and foremost change to search in SharePoint 2013 is there is only one search engine. The idea that you use the FAST engine for content and the SharePoint engine for people has been completely eliminated. In SharePoint 2013 there is a brand new search core that combines the best qualities and functionality of both SharePoint search and FAST search. The powerful indexing, linguistics, extraction, and query expressiveness are now evident throughout the platform. With this change comes another dramatic change in overall search architecture. There is now only one query language across...
all of SharePoint 2013, which combines the best features of FQL (FAST Query Language), and KQL (Keyword Query Language).

Less evident, but almost as crucial is the addition of both the Link database and the Analytics Reporting database. These two new databases in the overall SharePoint 2013 architecture provide for a great deal more personalization, analysis, and relevancy within the engine. The Analytics Reporting database has been added to keep track of all forms of analytics. Search Analytics capabilities analyze crawled items and how users interact with search results. These actions are stored in the event store within the Web Front End (WFE) server and are regularly pushed to the analytics processing component where the actions are analyzed and reconciled. They are pushed into the analytics reporting database and are then available to the query and processing components. This enables search to keep track of user actions, queries, and trends to provide the user with better search results and suggestions. This database now powers features such as personal and engine-wide query suggestions, favorites, and other search personalization components not found in any other enterprise search platform today.

The first and foremost change to search in SharePoint 2013 is there is only one search engine.
CHAPTER 2

Crawl and Content Processing Changes
SharePoint 2013 introduces some exciting capabilities when it comes to the ingestion of content into the search framework components. A lot of concepts originally introduced in SharePoint 2010 have been extended and improved.

Here is a quick list of some of those improvements with deeper elaboration on truly revolutionary new features and concepts:

**Anonymous Crawl**
This adds a component to the crawling feature to support anonymous crawling. Prior to this change, all crawls needed to be associated to a user account or leverage other methods of authentication.

**Asynchronous Content Crawl**
Allows for crawling asynchronous, dynamic content web parts.

**Visual Metadata Extraction**
Support for extraction of metadata using high-performance format handlers that allow for extraction of titles, authors and dates from HTML, DOCX, PPTX, TXT, Image, XML, and PDF formats. These format handlers look at font type and size, text alignment, capitalization, and other visual cues that we ourselves will generally use to determine the title and author of a document.

**Company Names Entity Extraction**
First introduced in SharePoint 2010, company name entity extraction has undergone a serious facelift. Instead of extraction dictionaries being managed in XML files on the file system like in 2010, SharePoint 2013 now managed inclusion and exclusion dictionaries from within the term store of the Managed Metadata Service. This greatly simplifies the management and extension of this capability.
Result Sources

Formerly known as scopes in SharePoint 2010, the result sources tool in SharePoint 2013 now combines the 2010 concepts of scopes and federation into a new, powerful tool.

One of the most significant features within result sources is the support for Remote SharePoint Index federation. While simple on the surface, this functionality fills a serious gap that existed in the overall scalability of SharePoint 2010. FAST and SharePoint were criticized in the marketplace for not having a global systems architecture. The approach was to tell users to centrally index all content in a large central farm, if the latency allowed.

Remote SharePoint Indexing addresses this problem by allowing federation with interleaving between local and remote SharePoint indices. This gives SharePoint 2013 a true global architecture solution that can be redundantly meshed to provide a scalable, fault tolerant architecture.

To enable this distributed indexing methodology in a secure fashion, a trust must be established between the farms using a new method called oAuth which allows the passing of the current user’s claims to the remote farm when making the search request. This is similar in concept to establishing a trust between servers to consume service applications. oAuth replaces Kerberos shared authentication.

The bottom line: SharePoint 2013 is better than the last version — which I loved.

- J. Peter Bruzzese, InfoWorld, September 19, 2012
CHAPTER 3

Query Execution and Page Changes
There are a number of new features in the query execution and processing components within SharePoint 2013. We will briefly review some of these components and go into greater detail on revolutionary features such as Query Rules and Query Suggestions.

**Ranking**
Ranking models are still configured via PowerShell like in SharePoint 2010, but in 2013 site collection administrators now have the ability to call a specific ranking model defined by the SSA admin from within query components at the site level.

**Query Spell Correction**
In SharePoint 2010, spell correction was implemented as a series of XML files that defined inclusion and exclusion items for the dictionary. In SharePoint 2013, Query Spell Correction is managed from within the term store of the Managed Metadata Service. Within the term store, Query Spellcheck Exclusions and Inclusions are nodes within the term store. Dynamic dictionary creation is still supported, but is now managed from within the term store.

Within the user interface for search Query Spell Corrections can be configured to use “Did You Mean” type functionality for query transforms.
Query Rules

Query Rules are a new feature in SharePoint 2013. Query Rules are designed to allow you act upon the intent of the query.

The Query Rules framework is composed of three top level components:

**Query Conditions:**
(Does the query meet a rule condition?)

These are rule sets that are meant to determine the intent of the query. Some potential use cases include:

- Query contains a specific word or words
- Query contains a word in a specific dictionary
- Query contains an action word that matches a specific phrase or term set
- Query is common in a different source (like Videos Result source)
- Results include a common result type (like file type)
- Advanced rule – match across a set of terms, dictionary, regular expression, etc.
- If the query is against a particular result source (see result source section above) or category, result source conditions can also be applied
- If the Query Condition is met, Query Actions can then be triggered
**Query Actions:**
*(What to do if the rule condition is met)*

Once a query condition is met, a series of actions can then take place. These actions include:

**Assign a Promoted Result** — This includes the “Best Bet” and former Fast Search for SharePoint 2010 feature known as “Visual Best Bets”. The configuration of the promoted results allow you to specify if the returned action should be treated as a best bet (hyperlink) or as a fully formatted HTML block (Visual Best Bet).

**Create and Assign a Results Block** — When a condition is met, one or more results blocks can be triggered. Result blocks specify an additional query to run and how to display results. This feature includes a full query designer so you can build and test queries before finalizing them. You can include the results above those returned by core results, or interleaved by ranking. You also have the ability to choose custom display templates for the result or results block.

**Change the Ranked Results by Changing the Query** —
This allows you to assign (additional parameter and weighting (XRANK Boosts) values to the query (Query Transforms for those familiar with FAST). For example, if the condition of the rule is met, apply XRANK constant boost of x number of points. XRANK is a FAST component that allows you to override default relevancy ranking by boosting the relevancy score for particular results at query time.

**Publishing Options:**
*(When to do this?)*

These determine when a query rule is active (start date, end date) or always active (by default). You can also configure a review date (triggers an email reminder to review this rule).

The power of query rules lies not only the flexibility they provide, but also the richness and complexity that can be derived from them. Imagine a single Query Condition being met which then triggers a visual best bet, a results block from a remote SharePoint site, a results block from a cloud source, and a query transform that will boost results coming from the cloud.
And these actions are only taken between November 25th and December 26th. So a great example would be in an intranet scenario where you could have a query rule that was active only during insurance open enrollment windows.

**Pre Query Suggestions**

- **Items you have clicked on before**
- **List of queries from other users**

Pre-query suggestions happen prior to the query being executed. The goal of pre-query suggestions is to aid the users in selecting a query and to help them to write better queries.

These suggestions are provided in two forms:

1. A list of items that others are typing for their queries
2. A list of items you have clicked on before from your personal query log
One of the key features is that they will never provide a suggestion to a search that did not yield a click-through (someone clicking on the document) or if the results would lead to a dead end (zero result query).

**Post Query Suggestions**

The second form of query suggestions is a post-query suggestion. Post query suggestions are suggestions that are provided after the query is executed and when results are displayed. These suggestions are based upon the results that you have clicked on at least twice. They provide a quick means to go back to a document that you regularly review or select.

Suggestions can be tuned (inclusions, exclusions) within the Service Application Admin Pages. It is also important to note that these are not tuned at the site collection level, but only at the SSA level.
Search Results Interface Enhancements
Result Types

Result types provide a means of determining how search results will be displayed, and a framework for developing rules about how a set of results should be displayed. The framework is composed of three parts:

- **Rules Engine** — List of rules to determine if the result type should be triggered.
- **Property List** — Property list associates the rule to document type, content type, or other managed property within SharePoint search.
- **Rendering Template** — The rendering template defines how a particular result will be displayed.

The power of Result Types really becomes evident when looking at a real-world scenario. In this scenario you have multiple documents that have been assigned content types (i.e. specification documents, data sheets, etc.)

Within Result Types you can specify a rule based upon specific criteria. The rules can contain fairly advanced features such as Boolean logic (i.e. AND OR NOT), equality (i.e. = or !=) or comparison ( < OR > ). These rules can be applied to managed properties. So for our example the rule might be ContentType=“spec documents”).

![Image of SharePoint settings](image-url)
Once the rule conditions have been met, you can specify which managed properties you would like to have returned. You must specify at least one managed property before it can be used in a rendering template.

The Rendering template consists of a template that is composed of HTML and might contain JavaScript. Within this simple to edit template (not like editing XSLT in SharePoint 2010) you can specify where you would like the requested property list items to be displayed using a tagging convention as follows (-#= contenttype =#-). This rendering template can call specific graphics (icons, etc.) and be stylized in any way that you would normally stylize HTML. Below is an example of a completed rendering template for a text document.
Refinement
With search results the refinement data works essentially the same as SharePoint 2010. There are two different modes for the refiner web part:

Search Results
Refinement is different with SharePoint 15 in that you can define display templates to use for rendering different kinds of refinements.

Faceted Navigation
Faceted navigation uses a term from the term store to filter what kind of data should be displayed (explained on next page). Faceted navigation is used in conjunction with term sets that are used for navigation. With each term you can select which managed properties should be used as refiners with that term.

Within the managed property you need to configure it as “Refinable.”
Examples:

- You have term store terms *Camera* and *Laptop*
- You have managed properties *Megapixel Count*, *Color*, and *Manufacturer*
- For *Camera* term, you add refiners for *Megapixel Count* and *Manufacturer*
- For *Laptop* term you add refiners for *Color* and *Manufacturer*

Another Example:

- *Department* is the term store term, *Customers*, and *Projects* could be refiners
Results Blocks
Results Blocks are blocks of search results that are managed as a block. So if a query generates more than one query (i.e. Query Rule), the resulting sub-queries and results can be treated as blocks. The value/power of results blocks comes when you utilize Query Rules and Result Sources. For example, you can use this feature if you have data in the cloud (i.e. Office 365) as well as on premise, and you wanted to display each group independently (as blocks).

Content by Search
The Content by Search functionality is a new capability specifically added to SharePoint 2013 to bolster Web Content Management capabilities. Similar to the Content Query Web Part, the new Content by Search Web Part works against indexed content in SharePoint 2013.

This functionality allows for some very interesting possibilities when it comes to not only content spotlighting, but also search driven site pages (Topic Pages). The Content Search Web Part takes the input of a query term or terms, and using a new feature called the Search Query Builder, you can specify exact queries that will drive content into the Web Part.
You can define display templates, and using simple HTML and JavaScript define an entire product catalog totally driven by the search engine.
Office Web Applications

Document Preview

One of the most exciting new features added to SharePoint 2013 is the integration of the ability to preview documents within search results. This technology leverages a new standalone server that hosts Office web applications. Office web applications provide a way to open the document in a web client environment with reasonably high fidelity (same format, fonts, sizing, etc.).

One key component within the document preview display is the “take a look inside” functionality. This provides a synopsis of the document and the ability to jump specifically to a relevant section of the document. It is important to realize that this is a 1.0 version of this technology. It is missing functionality that other preview technologies take for granted. This includes search term hit highlighting, immediately jumping to the most relevant page of the document, and copy and paste functionality from within the preview.

It is also important to realize the intention of this preview technology was not designed for documents to be consumed via this interface, but rather to determine if this is the particular document that you have been looking for.

The other key point to bring up in relation to Office web applications and specifically Document Preview functionality, at the time of this publication, there has been no announcement or public release of information regarding pricing, tiers, etc. It is unclear if this functionality will be part of the standard CAL or Enterprise CAL. This type of functionality typically garners a premium from other vendors, and I wouldn’t expect Microsoft to differ from this model.
Conclusion: Looking Ahead
**Should I wait for SharePoint 2013, or deploy on 2010 and upgrade later?**

One of the most common questions we are asked on a regular basis is should I deploy my site with SharePoint 2010 and FAST Search for SharePoint, or should I wait for SharePoint 2013? To be honest, the answer is a bit vague. It truly depends upon what you are trying to do. If you are planning to use SharePoint outside the firewall, I would recommend you wait for 2013. If you are planning to use SharePoint for e-discovery, I would wait. If you are just looking for better search results and none of the features mentioned in the e-book resonated with you, I would deploy SharePoint 2010 and FAST Search for SharePoint 2010, and plan for a migration in the next 12-18 months.

All the buzz around SharePoint 2013 has been about cloud, applications, and social capabilities. Sliding under the radar, but with core business changing functionality is Search. Search is pervasive in SharePoint 2013. To understand how pervasive it is used in SharePoint 2013 workloads, take a quick look at some of the workloads search powers in SharePoint 2013. These include: Content By Search Web Part (WCM), My Tasks (Social), My Site View (Social), “What’s new” feed (Social), E-Discovery Console, Topic Pages (WCM), Recommendations (WCM), Customer Engagement Portal, Image Library search, Exchange 15 search, and Web Content Management. Just to name a few.
**Buckle Up**

We’re excited about the upcoming release of SharePoint 2013 and the enhanced search functionality in SharePoint 2013. When Microsoft acquired FAST Search and Transfer in 2008, there was skepticism about how the functionality would come together. Search in SharePoint 2013 shows the full realization of the investment Microsoft made in acquiring FAST.

We’re also excited to see the many ways in which people will take this new functionality and use it to change their businesses and intranets. It will also be interesting to see how internet facing sites will leverage the web content management capabilities to transform their online properties and applications.

It is going to be an interesting year with lots of changes so…fasten your seat belts, it is going to be a fun ride!
BA Insight is a leader in agile information integration, enabling business to drive innovation by leveraging all knowledge and data across the enterprise. Offering new generation, cost effective alternative to expensive systems integration, the company’s award-winning technology provides a scalable foundation for liberating enterprise data, both structured and unstructured. Microsoft’s go-to partner for advanced search technologies, BA Insight enables customers to leverage their investments in SharePoint, FAST and other enterprise systems, and extend them with an overlay of easy-to-assemble, highly targeted business applications. Since 2004, more than three million users around the world have relied on BA Insight for low-cost, on-demand access to the information they need.

To learn more about BA Insight, visit www.BAInsight.com.