

conceptClassifier for SharePoint 2010 White Paper

The Optimal Solution for Leveraging SharePoint to Manage Unstructured Content

Prepared by:
Concept Searching
8300 Greensboro Drive, Suite 800
McLean, VA 22102
USA
+1 703 531 8567

9 Shephall Lane
Stevenage
Hertfordshire
SG2 8DH
UK
+44 (0)1438 213545

Twitter: @conceptsearch
<http://www.conceptsearching.com>

Martin Garland
President
marting@conceptsearching.com

June 25th 2012

Abstract

The purpose of this white paper is to provide a broad overview of conceptClassifier for SharePoint and the key functionality available in the product that is still unique in the industry, and to explain why it is the optimal choice for managing unstructured content in SharePoint 2010. This paper explores the uses of conceptClassifier for SharePoint and how it is being used by Concept Searching clients as an information governance solution in SharePoint, which addresses search, records management, compliance, data privacy, intelligent migration, and Enterprise 2.0.

Author Information

Martin Garland has over 20 years' experience in search, classification and Enterprise Content Management within the broader information management industry. His keen understanding of the information management landscape and business acumen provide a solid foundation for guiding organizations to achieve their business objectives using best practices, industry experience, and technology. Martin's expertise has been instrumental in assisting multi-national clients in diverse industries to understand the value of managing unstructured content to improve business processes.

He has focused on sales, marketing and general management, and has experience in both startup and turnaround operations throughout Europe, the US and Asia Pacific. One of the founders of Concept Searching, Martin is responsible for both business strategy and North American and International operations.

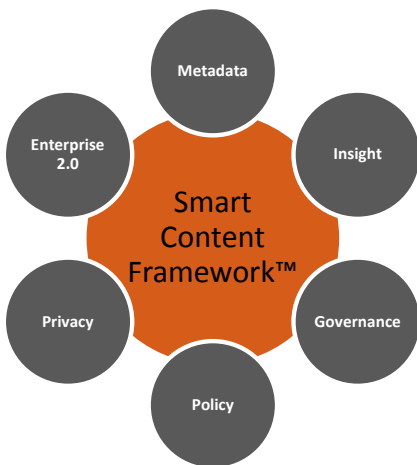
Overview

Traditionally, there has been a one dimensional approach to solving business problems with technology. Myriad solutions are available that solve a specific problem, such as search, records management or compliance, but they are typically narrowly focused and not adaptable outside the boundaries of the application. This leaves organizations in a continual state of flux, trying to realize incremental improvements to achieve specific business outcomes with multiple technologies.

This mindset appears to be slowly changing as there is a growing awareness in organizations to look beyond addressing a single business challenge and solve the fundamental problem that prohibits progress. The unrealistic expectation that users follow organizationally defined procedures to determine where a document should be stored, how long it should be preserved, who has access to the document and how it can be used has failed. The inability to find, use and protect content assets is now becoming a boardroom priority that carries financial repercussions and impacts business agility.

Smart Content Framework™ for Information Governance

Concept Searching's **Smart Content Framework™** is a toolset that provides the enterprise framework to mitigate risk, automate processes, manage information, protect privacy, and address compliance issues. The **Smart Content Framework™** is a multi-disciplinary solution delivered through Concept Searching technologies, which encompasses the entire portfolio of information assets, resulting in increased organizational performance. The framework has been proven to be a flexible solution to address recurring problems in organizations of any size or industry. Underlying the **Smart Content Framework™** are technologies that provide the ability to transparently tag content, deliver insight into search, classify it to organizational taxonomies, and preserve and protect information.



An enterprise *metadata* repository is the primary building block in the **Smart Content Framework™** that enables the proactive management of content. This first building block is an enterprise infrastructure component, tightly integrated with information governance and the management of the lifecycle of content. From this, enterprise search, compliance, records management, data privacy issues, and Enterprise 2.0 can be addressed and managed which are illustrated in the remaining building blocks.

This approach removes the barriers for organizations to manage unstructured content in a SharePoint environment and utilize the same technologies to address more than one challenge. The return on investment can be substantial, as the resources and costs are greatly reduced and the SharePoint investment can be leveraged. From a business perspective, content can be managed and used to improve business processes and ensure core business objectives can be met.

Compound Term Processing automatically identifies the word patterns in unstructured text that convey the most meaning and uses these higher order terms to improve precision with no loss of recall. The algorithms adapt to each customer's content and they work in any language regardless of vocabulary or linguistic style. This capability is still unique in the marketplace.

conceptClassifier for SharePoint

Concept Searching's technologies are still unique in the marketplace. What makes conceptClassifier for SharePoint unique is the ability to integrate conceptual indexing, as opposed to keywords or proximity, into the SharePoint platform automatically.

The three primary components of the technology suite that provide an enterprise extensible framework are the ability to automatically generate semantic metadata, auto-classify content, and develop controlled vocabularies and taxonomies. The combined functionality provides the structure for on-going management and the capacity to rapidly adapt to organizational changes. This approach is successfully enabling organizations to improve search, records management and compliance; enforce governance; and dramatically improve a broad set of business processes, resulting in reduced organizational risk and achievement of quantifiable business objectives.

conceptClassifier for SharePoint Components

conceptClassifier for SharePoint is the only industry solution that delivers automatic identification and extraction of concepts from within content as it is created or ingested, provides intelligent auto-classification, and enables enterprise class taxonomy management fully integrated with the SharePoint server environment. The only solution that runs natively in the Term Store, conceptClassifier for SharePoint is optimally delivered as a complete platform with all standard features included.

conceptClassifier for SharePoint's intelligent automated classification and taxonomy management provide organizations with the ability to develop a consistent structure to more effectively manage their content assets and utilize these assets to improve business processes. Unlike traditional tools, where taxonomy planning and deployment take considerable time and resources, or fully automated approaches, where rich human knowledge capital and input is eliminated, conceptClassifier for SharePoint has combined the strengths of both approaches. conceptClassifier for SharePoint expedites the process by providing the rich multi-term metadata to rapidly build taxonomies through innovative auto-clue suggestion based on the concepts within the content. The key features of the product are detailed below.

Compound Term Processing

A comprehensive approach to generating metadata requires more than syntactic metadata, and end users adding rich metadata is haphazard and subjective at best. Even if selecting terms from the term store or from a taxonomy, the chances for accurate classification due to human subjectivity is greatly diminished.

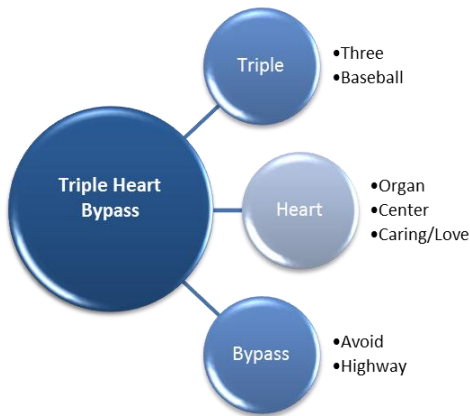
Unlike traditional metadata generators, Concept Searching uses *compound term processing* technology, which is unique in the industry. Compound term processing is a new approach to an old problem. Instead of identifying single keywords, compound term processing identifies multi-word terms that form a complex entity and identifies them as a concept. By forming these compound terms and placing them in the search engine's index, the search can be performed with a higher degree of accuracy because the ambiguity inherent in single words is no longer a problem.

As a result, a search for *"survival rates following a triple heart bypass"* will locate documents about this topic even if this precise phrase is not contained in any document. A search using compound term processing can extract the key concepts, in this case *"survival rates"* and *"triple heart bypass"*, and use these concepts to select the most relevant documents.

Identification of concepts within a large corpus of information removes the ambiguity in search, eliminates inconsistent meta tagging, and automatic classification and taxonomy management based on concept identification simplifies development and ongoing maintenance.

Semantic Metadata Generation

Since conceptClassifier for SharePoint is not restricted to keyword identification, compound term metadata - *concepts in context* - can be automatically generated either when the content is created or ingested. The generation of metadata based on concepts extracts compound terms and keywords from a document, or a corpus of documents, that are highly correlated to a particular concept. By identifying the most significant patterns in any text, these compound terms can then be used to generate non-subjective metadata based on an understanding of conceptual meaning. The compound terms, keywords, and acronyms identified are then used by the taxonomy developer to rapidly deploy enterprise taxonomies.



As shown in the diagram, the term “triple heart bypass” consists of three words, each of which has several meanings. conceptClassifier for SharePoint will generate ‘conceptual’ metadata, by extracting multi-word terms that identify “triple heart bypass” as a concept as opposed to single keywords. The search will return results based on the concept even if the exact terms are not contained in the document, for example “coronary artery surgery” or “heart surgery”. This metadata can be used by any search engine index, or any application/process, that uses metadata.

Auto-classification

Automatic and/or manual classification is included. Knowledge workers with the appropriate security can also classify content in real time, provided through a ribbon bar in Microsoft Office. Content can be classified not only from within SharePoint but also from diverse repositories including file shares, Exchange Public Folders, and websites. All content can be classified on the fly, in real time and classified to one or more taxonomies.

Taxonomy Management

The taxonomy component is easy to use and designed for Subject Matter Experts. There is no need for extensive training, and basic training can be completed in approximately one day. The taxonomy runs natively in the SharePoint Term Store and functions bi-directionally so all changes, regardless of where they are made, are immediately available.

The tool is extremely powerful and includes features that are not available in any commercial products. These features have been proven to reduce taxonomy development time by up to 80%. This represents a substantial reduction in the cost and manpower needed to develop and manage enterprise taxonomies which the total cost of ownership is typically very high.

Summary of Key Features and Capabilities

“The average cost to meta tag one document is \$4.00.”

Hoovers

Feature	Taxonomy Component	Explanation
Ability to build taxonomies or ontologies	Yes	
Auto-classification	Yes	
On the fly classification of internal and external content	Yes	Can also be scheduled if preferred.
Automatic generation of compound terms or phrases to build the taxonomy	Yes	This feature eliminates developing on a keyword based platform with rules bases to develop the concepts. This significantly reduces time, costs and the need for highly trained specialists.
Automated Taxonomy Load	Yes	Can be used to add or delete taxonomies from and existing index. This feature is also used to import organizational taxonomies and supports OWL and MeSH.
Hard coded application	No	conceptClassifier is highly adaptable and extensible delivered as APIs that work in any industry and/or market.
Synonyms Supported	Yes	
Synonym Definitions Required	No	Synonyms are supported but not needed as the technology will identify 'like' topics from the organization's own content as opposed to being created manually, which is very subjective.
Controlled vocabularies	Yes	Typical solutions do not support controlled vocabularies using the client's own content and have to be purchased and applied in isolation of the organizations own content, which is a unique organizational nomenclature. Without this capability, the controlled vocabularies must be created manually.
Scalability	Yes	The product can scale in an enterprise environment to handle petabytes of data. When using a statistical modeling approach the fewer the documents the less accurate the modeling.
Native Integration with SharePoint 2010 Term Store	Yes	
Integrated with SharePoint 2007, Microsoft Office, FAST, SharePoint Search, Windows Server 2008 R2 FCI	Yes	
Integrated with the Refinement Panel	Yes	
Managed by Subject Matter Experts	Yes	Does not require specialists or extensive training. This is very helpful within an organization, as each business unit may have different nomenclature to describe the same topic and the knowledge worker is most likely to identify correct terms for the taxonomy.

IDC estimates that return on investment for extending an organization's knowledge infrastructure ranges from a minimum of 38% to as high as 600%.

Instant Feedback on Changes	Yes	Instant Feedback is used to immediately tell the taxonomy developer the impact of their changes to classification rules. Any documents can be selected and re-classified in real time to show the effect of the changes. If the changes are not satisfactory they can be rolled using the Rollback Features. <i>Unique to conceptClassifier for SharePoint</i>
Suggest Clues for Class Clues are used to describe the language found in documents that make them 'about' a particular topic.	Yes	Clue suggestion is based on the statistical modeling of the entire corpus of documents. The larger the corpus the better the modeling. <i>Unique to conceptClassifier for SharePoint</i>
Audit and Rollback	Yes	
Distributed Taxonomy Management	Yes	For large enterprises that have a staff of professionals managing the taxonomies, this feature allows multiple servers to be used simultaneously without impacting performance.
Automatic Content Type Updating	Yes	<i>Unique to conceptClassifier for SharePoint</i>
Taxonomy Workflow Capabilities	Yes	<i>Unique to conceptClassifier for SharePoint</i>
Records Management conceptClassifier can automatically declare a record based on the client file plan, change the content type in SharePoint and route to the appropriate repository for disposition.	Yes	<i>Unique to conceptClassifier for SharePoint</i>
Data Privacy The feature works in the same as above, except it is identifying organizationally defined descriptors and vocabulary that indicate it is confidential information or a data privacy asset that must be secured.	Yes	<i>Unique to conceptClassifier for SharePoint</i>
Search Integration	Yes	Does not need a separate index when integrated with FAST or SharePoint search, or any search engine.
Search Improvements	Yes	Can be integrated with the Refinement Panel to capitalize on the underlying taxonomies to improve navigation as well as deliver 'relevant' information, reducing the search time.

Optional Components

There are two optional components for conceptClassifier for SharePoint that can be used to more fully leverage the organizations SharePoint 2010 investment.

conceptContentTypeUpdater

conceptContentTypeUpdater works with conceptClassifier for SharePoint to bypass manual processes with the SharePoint 2010 Content Organizer and automatically apply correct content types based on managed metadata properties.

The add-on product is deployed at the operational and tactical levels to provide site collection administrators with the ability to independently manage access, information management, information rights management, and records management policy application within their respective business units and functional areas, without the need for IT support or access to enterprise wide servers.

Automatically generated semantic metadata automates the tagging of content and triggers the content type update, which in turn applies actions on the content, thereby automating and enforcing the application of policies aligned to the organizational goals.

conceptTaxonomyWorkflow

conceptTaxonomyWorkflow is an optional Concept Searching component that can perform an action on a document following a classification decision when the criteria are met. conceptTaxonomyWorkflow is not a general purpose workflow engine, so does not compete with Microsoft's Workflow Foundation or K2 blackpearl/blackpoint. Unlike Concept Searching's conceptContentTypeUpdater, the workflow source type works in the SharePoint 2007 and 2010 as well as for all document types, FILE document types, and HTTP document types.

conceptTaxonomyWorkflow is typically used as a strategic tool for managing migration activities and content type application across multiple SharePoint farms. The module delivers workflow capabilities that enable intelligent automatic classification decisions during and after migration. These decisions enhance organizational performance and drive down costs, but more importantly enforce corporate and legal compliance guidelines.

Native SharePoint 2010 Term Store Integration

With the Term Store functionality in SharePoint 2010, organizations can develop a metadata model using out of the box SharePoint capabilities. An industry unique feature in conceptClassifier for SharePoint is the native integration with the Term Store and the Managed Metadata Services, where changes in the Term Store will be automatically available in the taxonomy component and any changes in the taxonomy component will be immediately available in the Term Store.

A compelling advantage is the ability to consistently apply semantic metadata to content and auto-classify it to the Term Store metadata model, solving the challenges of applying the metadata to a large corpus of documents and eliminating the need for the end user community to correctly tag content. Utilizing the taxonomy component, the taxonomies can be tested, validated, and managed, which is not a function provided by SharePoint 2010.

Why is Native Integration with the Term Store Important?

Since the taxonomy component functions bi-directionally with the Term Store, the additional functionality assists in expediting the development of the metadata models - taxonomies - offers sophisticated refinement capabilities, and significantly reduces manual development time and on-going maintenance.

Advantages	SharePoint 2010	conceptClassifier for SharePoint
Ability to automatically create and store classification metadata	No	Yes
Auto-classification capabilities	No	Yes
On the fly or scheduled auto-classification	No	Yes
Provide semantic metadata to any search engine index	No	Yes
Ability to tag content with vocabulary or retention codes for records management	No	Yes
Ability to automatically update the SharePoint content type for records management	No	Yes
Ability to automatically update a content type for identification of privacy information	No	Yes
Ability to test, validate, and manage taxonomies	No	Yes
Ability to eliminate manual tagging	No	Yes
Support of polyhierarchies	No	Yes
Support of ontological features	No	Yes

conceptClassifier for SharePoint Solutions

The strength of conceptClassifier for SharePoint is the flexibility and adaptability of the technology to address any application that requires the use of metadata. Based on the **Smart Content Framework™**, organizations have the ability to solve other challenges they may be facing. At the same time, they are leveraging the strengths of SharePoint 2010 and utilizing one set of technologies to improve business processes.

Search and Information Transparency

Many organizations continue to evaluate search solutions that deliver relevant results, are easy to use, and enhance end user productivity. This approach continually results in search engines that essentially don't work as promised. conceptClassifier for SharePoint provides the framework for organizations to improve access to information by enhancing the overall search experience. conceptClassifier for SharePoint provides the rich semantic metadata for any search engine index. By identifying every meaningful concept in every document, end users are no longer restricted to keyword searches and can eliminate iterative searches that typically don't provide relevant results. Since the taxonomies are developed based on the organization's own content, vocabulary normalization and controlled vocabularies provide results in the unique industry or organizational nomenclature.

conceptClassifier for SharePoint processes documents at index time and the resulting metadata is stored in the search index. This is much more efficient and accurate than any dynamic clustering approach. In addition, conceptClassifier for SharePoint allows the user to browse all of the content using the taxonomy structures without the need to specify any query text. Combining it with the FAST or the SharePoint search platform brings the ability to dynamically cluster, filtering on all conceptual metadata plus the user ability to browse all documents related to a particular subject resulting in significantly improved search precision and end user experience.

Within the FAST environment, conceptClassifier runs natively as a FAST pipeline stage, eliminating integration and customization issues. The taxonomy component can also be used to import FAST Entities to fine-tune them with metadata generated from the corpus of information, tuning it to the organization's own nomenclature.

“Taxonomy navigation reduces search time by 36%-48%.”

Hao Chen & Susan Dumais,
‘Optimizing Search by Showing Results in Context’

Compliance

Organizations are facing increasing pressure to ensure compliance with information management directives and protection of data, whether confidential customer, employee or business information.

Compliance is a requirement facing all organizations, but the healthcare industry will have compliance requirements that differ from those in the financial or energy sectors. Regardless of the industry, managing data effectively is a critical initiative in many organizations as the mismanagement of information can lead to severe repercussions, both financially and legally.

Many offerings focus simply on findability, whereas concept **Classifier** for SharePoint moves beyond findability, enabling automatically generated metadata to drive application of policies, delivering true ROI for enterprises. Clients are able to automatically secure sensitive information, declare documents of record, automate workflow to comply with federal regulations and reduce organizational risk. In most organizations that use SharePoint, documents are often incorrectly filed and have inappropriate metadata applied, and there are no measures to control access and rights management for individual data assets. Many 'compliance' solutions exist that do a very good job of assessing risk and applying third party tools to minimize that risk. The challenge is to eliminate the risk before it occurs.

Both concept **TaxonomyWorkflow** and concept **ContentTypeUpdater** can be used to enforce compliance policies across the organization. concept **TaxonomyWorkflow** serves as a strategic tool managing migration activities and content type application across multiple SharePoint farms. While concept **ContentTypeUpdater** deploys at the operational and tactical levels, to provide site collection administrators with the ability to independently manage access, information management, information rights management, and records management policy application within their respective business units and functional areas, without the need for IT support or access to enterprise wide servers.

By effectively and accurately applying policy across applications and content repositories, Concept Searching's technology enables organizations to significantly improve their compliance and information governance initiatives.

"A 1,000 person organization can spend \$170k per month in lost productivity due to manual records declaration."

Cohasset Associates

Records Management

Many organizations that have implemented a records management solution have identified end user adoption as a key barrier to success. Difficult interfaces, complex processes, and disparate tools make it hard for users to embrace the records management process. End users are often reluctant to change their work habits, which can pose a challenge in consistently maintaining accurate records. As a result, too many documents are never subjected to enterprise policies, resulting in widespread noncompliance issues. Automating the records management processes to facilitate end user acceptance removes much of the burden from the staff. Although a records management solution can provide the method for document retention, workflow, archiving and maintenance, the end user is ultimately responsible for the assignment of documents with the appropriate descriptors, and is burdened with the task of providing sufficient information that provides corporate meaning to the entity.

Content types in SharePoint 2010 enable organizations to take advantage of the workflow capabilities that can enhance organizational performance while driving down costs. The only obstacle with content type applications is that individuals have to decide which content type applies to every document ingested by SharePoint.

“9 out of 10 breaches involved some type of ‘unknown’ which was data that was not known to be on the compromised system.”

2008 Data Breach
Investigations Report

To address this issue, concept **ContentTypeUpdater** works with concept **Classifier** for SharePoint to automatically apply correct content types when organizationally defined descriptors and vocabulary reside within documents. As a result, the combination of the technologies delivers a unique and powerful solution, leveraging SharePoint content types.

concept **Classifier** for SharePoint is used for records management by creating a taxonomy that mirrors the file plan and as documents are created or ingested, based on the vocabulary and/or record retention code, the documents are declared records according to the records management file plan. Using the concept **Classifier** for SharePoint taxonomy manager component, organizations can rapidly develop and deploy taxonomies and controlled vocabularies associated with business processes.

concept **ContentTypeUpdater** uses records retention codes, semantic, and security metadata associated with data assets to drive the automatic application of content types. Once documents have the appropriate content type, based on natural language and automatically applied metadata, workflows can be started.

Data Privacy

An often overlooked issue and a growing problem is the inability to identify confidential/privacy information that falls outside of the boundaries of even the most comprehensive compliance and security processes. With confidential information being communicated in emails, faxes, unstructured content, and traditional paper forms, the chances of unprotected privacy information becoming a compliance liability is inescapable. This is a growing concern in global organizations where specific government regulations specify where the content can be stored, as well as the need to address unique privacy safeguards.

With the ever changing landscape of regulatory compliance, and the assurance that more legal requirements to protect data will continually be mandated, organizations must deploy more proactive oversight for all potential data exposures to minimize organizational risk. concept **Classifier** for SharePoint and concept **ContentTypeUpdater** augment traditional products and compliance processes within an organization, by discovering where unknown privacy data exists. Content containing any type of organizationally defined confidential or privacy data is automatically tagged, and secured, and made available to selected users using Windows Rights Management for further disposition and analysis.

Fully customizable to identify unique or industry standard descriptors, content is automatically meta tagged and classified to the appropriate node(s) in the taxonomy, based upon the presence of the descriptors and/or vocabulary from within the content. Once tagged and classified the content can be managed in accordance with regulatory guidelines. The innovative solution is fully customizable to address all aspects of data privacy that are unique to the organization. Once identified, the organization can have the potential exposures automatically routed to specific or multiple locations. Privacy data can be identified from within scanned documents, faxes, Microsoft Office applications, email, servers and PCs, and websites.

Migration

Migration to SharePoint 2010 presents various challenges at various levels: bulk migration of well-managed, high-value documents; bulk archival or disposition of old or extraneous documents; and finding a way to deal with the poorly classified middle ground. These middle ground document collections contain a variety of documents – some are important, others are ‘nice to keep’ and many are usually irrelevant. Dealing with the middle ground is where many large archive vendor solutions are unable to deliver an intelligent approach and where **conceptClassifier** for SharePoint and **conceptTaxonomyWorkflow** add value.

To migrate these document collections effectively and with intelligence the text content of each document needs to be searched to determine its value. This classification must be done before an intelligent decision can be made about how to relocate items during the migration process. This is accomplished through the creation of taxonomies that are based on descriptors and vocabulary that indicate what type of information is contained in each content asset. This cannot be done manually as the volume is too high and the consistency of human review and decision making is unreliable and too costly.

This also addresses the security issues of the documents as they are moved to their new location. **conceptClassifier** for SharePoint will respect the existing security status and apply the same security in the new location and will also identify sensitive documents that may not currently be in a secure location. **conceptClassifier** for SharePoint will identify the security profiles of these documents based on their content and the relevant taxonomies, such as PII, PHI, ITAR. If a document is automatically classified against one or more of these taxonomies it will be assigned the appropriate security profile.

Concept Searching Qualifications

Concept Searching is the only globally managed Microsoft Independent Software Vendor (ISV) partner that develops technologies that run natively in the Microsoft suite of products. Working closely with Microsoft product teams, we have access to all early betas ensuring integration and upgradeability within the product stack. Focusing solely on developing innovative technologies for Microsoft, Concept Searching is the only statistical metadata generation and classification software company in the world that uses concept extraction and compound term processing to significantly improve access to unstructured information running natively in the Microsoft SharePoint environment.

Concept Searching’s clients, such as Parsons Brinckerhoff, the [NASA Safety Center](#), [OppenheimerFunds](#), [Perkins+Will](#), the US Army Medical Command, the [US Air Force Medical Service](#), the [US Army – Records Management and Declassification Agency](#), the [Development, Concepts and Doctrine, Ministry of Defense Centre \(DCDC\) UK](#), and Abbott Laboratories, have all realized the value of the **Smart Content Framework™** and **conceptClassifier** for SharePoint to control their content assets and maximize the value of unstructured content to meet business objectives

Summary

conceptClassifier for SharePoint is an optimal solution for any organization that needs to manage unstructured content to maximize business benefits. Going far beyond improving search outcomes, the **Smart Content Framework™** and the combined technologies capture knowledge assets and extend business processes, to leverage unstructured content assets and SharePoint to derive quantifiable results.

Appendix A: Search Options

Concept Searching has an enterprise search engine conceptSearch, but also integrates fully with SharePoint 2010 Search as well as Fast Search for Internet Sites, or any enterprise search engine. A brief comparison is shown using SharePoint or FAST search engines and the options available when using conceptClassifier for SharePoint.

The primary feature to note is that conceptClassifier for SharePoint provides the capability to generate multi-term words, which neither of the other technologies can do. Therefore, where features are equal, the conceptClassifier for SharePoint search results are far more robust and granular.

Feature	SharePoint Search	FAST Search	conceptClassifier for SharePoint
Keyword Search	X	X	X
Concept Search	N/A	N/A	X
Complex Ranking Algorithms	X	X	X
Federated Search	X	N/A	X
People Search/social tagging ranking	X	N/A	Yes, in conjunction with SharePoint
Property Extraction	Limited	People names, locations, company names	Unlimited – multiword terms
Similar Results	N/A	X	X
Basic Refiners	X	N/A	X
Deep Results Refinement	N/A	X	X
Keyword based document boosts	N/A	X	Based on concepts
Best Bets	X	X	X
Previewers	N/A	X	X
Taxonomy Search Hierarchy	N/A	N/A	X
Rich Web Indexing Support	N/A	X	X
Dynamic Clustering	N/A	Limited	X
Taxonomy	Limited	N/A	X
Taxonomy Management & Maintenance	N/A	N/A	X
Manual Metadata Tagging	X	X	X
Automatic Semantic Metadata Tagging	N/A	N/A	X
Auto-classification as content is created or ingested	N/A	N/A	X
Classification Rules	N/A	N/A	X

Feature	SharePoint Search	FAST Search	conceptClassifier for SharePoint
Vocabulary Normalization	N/A	N/A	X
Identification and protection of confidential information	N/A	N/A	X
Automatic tagging of documents applying retention codes for records management	N/A	N/A	X
Automatic Content Type updating based on organizationally defined descriptors and vocabulary	N/A	N/A	X

Additional Resources

Website Links

[Customers](#)

[Case Studies](#)

[Product Webcasts](#)

[Press Releases](#)

SharePoint White Papers and Collateral

[Smart Content Framework™](#)

[The Case for Going Native](#)

[When and When Not To Use Third Party Tools](#)

[Enhancing SharePoint Through Information Governance](#)

[conceptClassifier for SharePoint Product Overview](#)

[Transforming Content into Business Value](#)

[8 Reasons to use conceptClassifier with SharePoint](#)

[conceptContentTypeUpdater Product Overview](#)

[conceptContentTypeUpdater Solution Overview](#)

[conceptTaxonomyWorkflow Solution Overview - Migration](#)

About Concept Searching

Founded in 2002, Concept Searching provides software products that deliver conceptual metadata generation, auto-classification, and powerful taxonomy management from the desktop to the enterprise. Concept Searching, developer of the **Smart Content Framework™**, provides organizations with a method to mitigate risk, automate processes, manage information, protect privacy, and address compliance issues. This information governance framework utilizes a set of technologies that encompasses the entire portfolio of information assets, resulting in increased organizational performance and agility.

Concept Searching is the only platform independent statistical metadata generation and classification software company in the world that uses concept extraction and compound term processing to significantly improve access to unstructured information. The Concept Searching Microsoft suite of technologies runs natively in SharePoint 2010, FAST, Windows Server 2008 R2 FCI, and in Microsoft Office applications.

Headquartered in the US with offices in the UK, South Africa and Canada, Concept Searching solves the problem of finding, organizing, and managing information capital far beyond search and retrieval. The technologies are being used to drive intelligent migration, enable effective records management, identify and lock down sensitive information and enhance governance. For more information about Concept Searching's solutions and technologies please visit <http://www.conceptsearching.com>



© 2012 Concept Searching



Americas
+1 703 531 8567
info-usa@conceptsearching.com

Europe
+44 (0)1438 213545
info-uk@conceptsearching.com

Canada
+1 703 531 8567
info-canada@conceptsearching.com

Australia
+61 (0)2 8006 2611
info-australia@conceptsearching.com

New Zealand
+64 (0)4 889 2867
info-nz@conceptsearching.com

Africa
+27 (0)21 712 5179
info-sa@conceptsearching.com

Marketing and PR
International: +1 703 531 8564
Europe: +44 (0)1438 213545
marketing@conceptsearching.com

Follow us on Twitter
[@conceptsearch](https://twitter.com/conceptsearch)

www.conceptsearching.com