

The Business Value of Compound Term Processing and Automatically Generated Intelligent Metadata White Paper

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

info-usa@conceptsearching.com
<http://www.conceptsearching.com>
Twitter: [@conceptsearch](https://twitter.com/conceptsearch)
[Concept Searching Blog](#)

Martin Garland
President
+1 703 531 8567
marting@conceptsearching.com

March 2015

Abstract

This purpose of this White Paper is to explore the capabilities of the Concept Searching technology platforms and the options organizations may consider, and to illustrate the business benefits of the overarching enterprise metadata framework that can be achieved through the deployment of Concept Searching's compound term processing technologies, in combination with SharePoint on-premises, SharePoint Online, and OneDrive for Business.

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 his 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 expertise 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.

Thank You to Market Experts, Vendors, and Analysts

Market experts, vendors, and analysts have been referenced in this White Paper to provide additional insight. All references are noted. We certainly hope we have not crossed the boundaries, and sincerely appreciate their perspective and knowledge in contributing to this White Paper.

Feedback is welcome. Please direct any comments to Carla Mulley, VP of Marketing, at carlam@conceptsearching.com

Table of Contents

Abstract	1
Author Information	1
Overview	3
Why Concept Searching?	4
Why an Enterprise Metadata Framework?	5
The Problem with Metadata	5
Precision versus Recall.....	6
The Market	6
Applications	7
Information Governance	7
Enterprise Search.....	8
Security of Content in Context.....	8
Records Identification.....	9
eDiscovery, Litigation Support, FOIA	10
Migration	10
Collaboration, Business Social Applications, and Text Analytics	11
Summary	11
About Concept Searching	12

Overview

The organizational paradigm has changed. The fact that 80% of business decisions are made using unstructured content¹ is significantly increasing organizational risk, is costly, and impacts the agility of the organization to meet its goals. Content is the lifeblood of an enterprise, created because of business processes, changed through business processes, and ultimately drives the business processes. Enterprises should be looking for a way to not only improve search outcomes but to fundamentally improve business outcomes. The tentacles of unstructured data are now impacting search, records management, security, migration, eDiscovery, FOIA, litigation support, and even collaboration, business social, and text analytics.

Unstructured content for many, if not all organizations, has become overwhelming. The traditional use of SharePoint as a file server has gone by the wayside, as both companies and platforms have matured to the point that, for many, content is central to the organization. According to Microsoft, Office 365 is the future and although it has given rise to a new set of challenges, it is apparent that the cloud is the new standard going forward. Microsoft has recognized that for Office 365 to succeed, it must address the proliferation of content. This can be seen by its recent investments around compliance, legal de-duping, and clustering that will enhance the compliance in Office 365. DLP has been announced, representing an attempt to ease security fears with the cloud.

The metadata, auto-classification and taxonomy market is growing as organizations realize that they can no longer cope with content overload. The majority of vendors in this market focus their efforts almost exclusively on improving enterprise search, and although improved search does yield business value, organizations must adopt a more comprehensive approach and understand the quantifiable business value of an enterprise metadata framework.

The mindset reinforced by most vendors' sales strategy is to provide a metadata, auto-classification, and Term Store or taxonomy tool, as a single solution to improve search. A more effective approach is to adopt an integrated enterprise metadata platform, where metadata is a core component of the technology infrastructure and can therefore be applied to a variety of application challenges in the management of unstructured content. This metadata framework can then be used to support a complete range of intelligent metadata enabled solutions, enforce policy across the entire enterprise, and protect the organization's investment now and into the future.

The purpose of this White Paper is to explore the capabilities of the Concept Searching technology platforms, the options organizations may consider, and to illustrate the business benefits of the overarching enterprise metadata framework that can be achieved through the deployment of Concept Searching's compound term processing technologies in combination with SharePoint on-premises, SharePoint Online, and OneDrive for Business simultaneously.

“Concept Searching’s industry unique compound term processing enables the identification and the correct weighting of multi-word concepts in unstructured text, affording the rapid creation of semantic metadata, which can be used to apply organizationally aligned descriptors to content.”

Why Concept Searching?

The Concept Searching platforms are built on a unique concept extraction and recommendation engine that uses compound term processing, and they are the most mature, adaptive, and scalable technologies available in the market today. These technologies enable the automatic generation of semantic metadata to drive accurate auto-classification and deploy intelligent metadata enabled solutions across a broad spectrum of the enterprise. This is made possible because Concept Searching is the only statistical software vendor that has an embedded compound term processing engine. Please see the [Compound Term Processing White Paper](#) for further information.

Concept Searching’s industry unique compound term processing enables the identification and the correct weighting of multi-word concepts in unstructured text, affording the rapid creation of semantic metadata, which can be used to apply organizationally aligned descriptors to content. The tagging and auto-classification of that content can be aligned with business goals, and the semantic metadata generated can be easily integrated with any third party application or platform that can interface via web services.

Creating, tuning, and managing term sets and taxonomies that are optimized for the organization is challenging, as each participant in the process, and every end user, may have a different way of expressing the same or similar descriptors (metadata). This is important as a taxonomy must remain fluid as opposed to static, and must be managed in a way that easily facilitates change. The goal is not only to give people the right information, but to distil content from a variety of sources, creating useable knowledge.

Organizations must also consider management of the term sets and taxonomies, as well as the resources and associated costs, to adapt to organizational change and continuous improvement. Traditional taxonomy tools, and several competing products available today, require significant investments in time, expertise, and money to develop and maintain. The Concept Searching auto-classification component has the capability to automatically group unstructured content, based on an understanding of the concepts that share mutual attributes, while separating dissimilar concepts. The taxonomy manager component provides a hierarchical view of topics that have been grouped because they share the same quality or characteristic. Because of Concept Searching’s compound term processing, documents are automatically classified in the taxonomy based on their relationships and relevance based on concepts. Documents may exist in multiple categories as one document may contain multiple concepts.

Supporting both automatic and manual classification, Subject Matter Experts (SMEs) can utilize rich features such as node weighting, ability to see the ‘concepts in context’, ability to search the corpus, auto-clue suggestion for classification, and instant feedback on the impact of changes. The taxonomy provides the structure for the grouping of like documents and enables a more targeted, accurate, and efficient management and tuning tool that translates into reduced costs and improved productivity. Many of these features are unique in the industry.

Running natively and fully integrated with the SharePoint Term Store, on-premises or in SharePoint Online, the technology can consistently apply conceptual metadata to content and auto-classify to the Term Store, writing back the metadata into the managed metadata service, solving the challenge of applying the metadata to potentially thousands of documents and eliminating the need to depend on the end user community to correctly tag content. The taxonomy manager component communicates bi-directionally and simultaneously with the Term Store, when changes are made in the Term Store or in the taxonomy manager.

This added functionality helps to expedite the development of the metadata models, offers sophisticated refinement capabilities, and significantly reduces ongoing maintenance and risk, as the structure is held only in the Term Store, the classification clues stored only in SharePoint custom properties, and backup is done only in one place – SharePoint.

Why an Enterprise Metadata Framework?

Quite simply, the value is the ability to use one set of technologies to solve multiple application challenges across an enterprise. The framework provides the enterprise infrastructure for unstructured information governance that results in reduced risk and ensures compliance. The framework addresses information complexity, and delivers solutions that can effectively increase information transparency and provide the ability to manage content, regardless of where it resides. By providing immediate and actionable visibility of all relevant content, decision making can be improved, risk minimized, and costs reduced. This eliminates the need to purchase applications that address a specific application, removes interoperability issues, and reduces costs associated with manpower and misaligned technology. The optimal result is an enterprise metadata repository containing a rich set of intelligent metadata that reflects the unique terminology and vocabulary of the organization.

“Without the ability to identify ‘concepts in context’ the hierarchical structure contains little value and, more importantly, the metadata is rendered useless to other applications that could be improved.”

The Problem with Metadata

The challenge with metadata is both obvious and elusive. The ability to harness the meaning of content requires the utilization of tools that deliver the ability to manage and retrieve content at the same rate that it is being created, ingested, and distributed. The fundamental component is the quality of metadata, which is required by many applications within the organization.

For solutions that use auto-classification, the classification is either highly general, for example the document comes from Finance, or is dependent on end user metadata or system defined metadata. Without the ability to identify ‘concepts in context’, the hierarchical structure contains little value and, more importantly, the metadata is rendered useless to other applications that could be improved.

Results from Concept Searching’s annual SharePoint 2015 Metadata Survey, currently unpublished, indicated that 93% of organizations were manually metadata tagging. To further substantiate our results, they matched those in an AIIM survey². They indicated that 93% of organizations still rely on manual tagging, which is the primary reason for the failure of search. Humans are inconsistent. In tagging habits, less than 50% of content is correctly indexed, meta tagged, or efficiently searchable³. Highly trained information specialists will agree on meta tags between 33% and 50% of the time⁴. If supplied with a dropdown list, end users will typically select the first option. In other words⁵, manual tagging is totally ineffective.

Unless the metadata has value, meaning that it can recognize concepts and relationships, the organization becomes limited in the ways that it can deliver and use the metadata to improve diverse applications. There are tools that will aggregate metadata and generate metadata using various technologies, but not automatically generate conceptual metadata. Conceptual metadata, consisting of multiple words that represent a concept, will be auto-classified, and the technologies can use and reuse the metadata across the enterprise and integrate into any environment, which gives organizations the opportunity to solve multiple business challenges.

Precision versus Recall

Concept Searching's technologies deliver both high precision and high recall. Precision and recall are the two key performance measurements for information retrieval. Precision is the retrieval of only those items that are relevant to the query. Recall is the retrieval of all items that are relevant to the query. The ideal goal is for the outcomes of these two actions to be balanced. Compound term processing has the ability to increase precision with no loss of recall.

The Market

Unstructured data and metadata are increasing at an average annual growth rate of 62%⁶. Despite the fact that Microsoft is filling gaps in its Office 365 product, such as Delve and Clutter, the SharePoint on-premises and SharePoint Online communities have not received the same attention, although the details of SharePoint 2016 have not yet been disclosed.

The SharePoint Term Store uses a manual process and does not include any auto-classification capability. Although Microsoft has made incremental improvements from SharePoint 2010 to SharePoint 2013, term set capabilities are somewhat limited. There are certain vendors, including Concept Searching, that provide an auto-classification capability to augment the Term Store and write back metadata into the managed metadata properties. But this does not eliminate manually intensive rule building, assumes the administrator understands the nuances of the all functional groups within the organization, and means that testing is cumbersome. Only Concept Searching provides a fully functional taxonomy component, to rapidly create and test rules, ensuring all term sets are accurate and easily managed. This is a significant opportunity to provide value to SharePoint, SharePoint Online, and OneDrive for Business for organizations that have been hesitant about using the Term Store.

The Concept Searching platforms are the only available platforms on the market that use compound term processing, delivering better accuracy and scalability. They are the only platforms that are integrated with all versions of SharePoint on-premises, SharePoint Online, and OneDrive for Business, enabling the management of unstructured content across hybrid environments. This is a significant advantage as organizations have the ability to tune and manage the term sets via the taxonomy manager component centrally, yet deploy specific term sets locally, regardless of platform, eliminating the manual and often subjective rule building tasks. For an organization to fully leverage metadata, the use of an ancillary tool is a requirement.

When compared with alternative products, Concept Searching's platform does not share the limitations of linguistic based tools. The concept identification works in any language and, because the technology is statistical, it works regardless of the vocabulary, for example vertical sector, or grammatical style. Linguistic tools have the problem of working consistently across different markets and in each language, whereas Concept Searching will adapt to the language, organizational nomenclature and the concepts found within the corpus.

The other issue with linguistic tools is the grammar used in the content. Since these technologies use parts of speech, content must be evaluated on the type of document, because a news article, a resume, a patent application, a legal contract, or a web page use very different styles.

“The Concept Searching platforms are the only available platforms on the market that use compound term processing, delivering better accuracy and scalability. They are the only platforms that are integrated with all versions of SharePoint on-premises, SharePoint Online, and One Drive for business, enabling the management of unstructured content across hybrid environments.”

“Using compound term processing, auto-classification, and taxonomies incorporated into an enterprise metadata framework, organizations can significantly reduce costs, improve business processes, reduce risk, and improve the bottom line.”

Many vendor products typically require a document training set for every term. This represents a high initial cost as typically more than 50 documents are used to train the system. Ongoing management of the term sets and taxonomies can be a resource intensive activity. These rules-based training set platforms have been generally discredited for use within large scale enterprise accounts for these very reasons.

Other products require the organizational staff to learn a new programming language, or depend on consultants, and ongoing cost if the taxonomy is to be kept up to date. Vendors also provide ‘pre-packaged’ taxonomy structures, for specific industries, but lacking classification rules. These often do not have the flexibility to be customized to the unique nomenclature of the organization. Machine learning, or artificial intelligence, can teach itself to change or grow as new data is added. The downfall is, instead of extracting data for human comprehension, as is the case in data mining applications, machine learning uses that data to improve the programs’ own understanding, as opposed to human understanding. Semantic networks are expressed visually, but they are intractable for large domains, and do not represent performance or meta-knowledge very well. Some properties are not easily expressed using a semantic network, for example negation, disjunction, and general non-taxonomic knowledge. Expressing these relationships requires workarounds, such as having complementary predicates and using specialized procedures to check for them, but this can be regarded as less elegant.

Applications

Improving search is a highly worthwhile and valuable endeavor. Search is not the only application that requires the use of intelligent metadata. Using compound term processing, auto-classification, and taxonomies incorporated into an enterprise metadata framework, organizations can significantly reduce costs, improve business processes, reduce risk, and improve the bottom line.

Information Governance

The term ‘information governance’ is bandied about by the media and analysts. Unfortunately, it is easier said than done. Information governance is a complex set of processes that requires participation of all levels of staff, workflow and documented processes, and strategies for enforcement of policies, as well as the appropriate technology to deploy actions and expected outcomes based on the decisions. Often overlooked is the information governance of unstructured and semi-structured data.

Organizations have taken the concept of information governance to heart. Concept Searching’s 2015 Metadata Survey, currently being compiled, shows this has become a high priority for 20% of respondents, 25% are actively improving information governance, and 26% plan to focus on it within the next twelve months. Without an enterprise metadata framework, efforts will fail or be extremely challenging, as each application will need to be addressed separately, despite the fact that much of the metadata is shared and reused enterprise wide. This limits the ability to deploy workflow procedures to enforce policy. In the scenario of SharePoint, SharePoint Online, and even OneDrive for Business, information governance becomes too costly, problematic, and manually intensive to apply to all environments, let alone all applications that use metadata.

Enterprise Search

Enterprise search, defined as delivering relevant and accurate results, is a failure. According to AIIM, for 71% of organizations, search is vital or essential, yet only 18% have cross-repository search capabilities, 38% of organizations have not tuned or optimized their search tool at all, including 8% that have not even turned it on, 47% of organizations feel that universal search and compliant eDiscovery is becoming nearly impossible, and 53% of organizations admit that their legal discovery procedures are ad hoc, manual, disruptive, and expensive⁷.

Responses to Concept Searching's 2015 Metadata Survey showed that improving enterprise search was important to most organizations, 15% of respondents consider search a high priority, 30% indicated they were actively improving search now, and 26% planned on improving search within the next twelve months.

This harbinger of change on the horizon indicates organizations know they have a problem, and some are investing the time and resources to improve it. SharePoint does provide tools to tune the search engine, involving resources, testing, reiterative tasks, and ultimately increased costs. The fundamental problem with business users spending 2.5 hours per day searching⁸ remains. Erroneous and often meaningless metadata populating the search engine index will always result in irrelevant results. Without a third party tool, SharePoint does not currently have the ability to automatically tag content, nor auto-classify it to the Term Store or a taxonomy product.

“Compound term processing can also be deployed to drive ‘push features’ within SharePoint, such as the display of relevant people, news and recent documents.”

Compound term processing can be deployed to improve search and the end user experience by automatically offering a ‘search as you type’ feature, guiding the user using relevant extracted concepts. Automatic query expansion and the ability to retrieve relevant documents even if they do not contain the actual words used in the query text, can significantly increase search precision and the end user experience. Compound term processing can be used for predictive coding applications, based on its ability to extract the key concepts found in one or more documents and then use this list of concepts to deliver a ranking of all available documents based on their relevance to the training set. Finally, compound term processing can also be deployed to drive ‘push features’ within SharePoint, such as the display of relevant people, news and recent documents.

The ability to capture concept based metadata and retrieve relevant search results from within an organization and diverse repositories is the real currency of interoperability. Providing syntactic as well as automatically generated metadata enables the meaning of content to be represented and shared in an unambiguous and transparent manner.

Security of Content in Context

One only has to read the daily news to be aware that hackers and security breaches, both internal and external, are on the rise, and alarmingly so. The statistics are quite high for internal data leaks and breaches. According to the Ponemon Institute, 69% of companies reporting serious data leaks responded that their data security breaches were the result of either malicious employee activities or non-malicious employee error. In fact, the number one leading cause of data security breaches resulted from non-malicious employee error (39%). The Ponemon Institute concludes that these breaches are typically the consequence of complacency or negligence, from lax or insufficient access controls to sensitive or confidential data. Only 16% of serious data leaks were linked to hackers or penetration⁹.

Despite the fact that 65% of employees believe it is their responsibility to ensure that sensitive company information is not leaked, only 32% of knowledge workers always clean files of hidden sensitive data before sharing, 70% of those who forward emails with attachments without reading them first do not remove sensitive data before sending, and 80% of employees use unsecure file sharing methods, putting corporate data at risk¹⁰. If we consider unstructured data, it becomes highly likely that there would be a breach of confidential information. The solution of course is to ensure security and access controls are in use, up-to-date, and enforced. But that is not the real problem. Organizations need to evaluate the risk of ignoring the potential of a data breach based on privacy or confidential information being exposed from within content, either internally or externally.

The Impact of Security in the Cloud

The security factor is the primary concern for organizations that are contemplating moving to a cloud environment. Shadow IT is emerging as a critical problem. Shadow IT is the result of cloud IT not being able to deliver business users with the services and apps they need in order to achieve a maximum degree of productivity. If they cannot get what they need from the corporate IT department, business users simply swipe their own credit cards and obtain the required IT resources elsewhere¹¹. In addition, a growing number of workers are storing confidential company information via unauthorized sharing and sync applications such as Dropbox or Box¹²

Since Concept Searching is not limited to improving search results, compound term processing can be applied to identifying and protecting privacy and confidential information. Many security products have some of this capability, such as identifying a social security number or a credit card number from unstructured data. Concept Searching provides the ability for the organization to define what is privacy and confidential information, using the unique terms of the organization's nomenclature. The extraction of multi-term concepts and descriptors, when content is created or ingested, enables the organization to proactively identify and protect content, remove it from unauthorized access, and prevent portability. One military organization has a number of locations using Concept Searching products to identify potential data breaches, and for the past nine years since deployment has not had a single incident. Since the technologies are not limited by where the content resides, on-premises or in the cloud, content that needs to be secured can reside in diverse repositories and environments.

The elimination of end user tagging provided by automatic conceptual metadata generation capabilities represents a significant step forward in records management, provides high value, and reduces organizational risk.”

Records Identification

With compliance requirements changing rapidly and end user adoption cited as the biggest source of failure in records management, there is a great need to simplify and manage the process. Non-compliance can represent monetary loss as well as disrupt an organization's business in a global world. More than 100,000 international laws and regulations are potentially relevant to Forbes Global 1000 companies, ranging from financial disclosure requirements to standards for data retention and privacy. Additionally, many of these regulations are evolving and often vary or even contradict one another across borders and jurisdictions¹³. The elimination of end user tagging provided by automatic conceptual metadata generation capabilities represents a significant step forward in records management, provides high value, and reduces organizational risk. It is simply not realistic to expect broad sets of employees to navigate extensive classification options while referring to a records schedule that may weigh in at more than 100 pages¹⁴.

The backbone again becomes the enterprise metadata repository. The approach is to develop a taxonomy that mirrors the file plan where content will be auto-classified by identifying and assigning correct record identifiers and other organizationally defined descriptors, optionally apply an appropriate content type, then either in-place or automatically route to the organization's records management platform. Intelligently automating the records identification process is the precursor to an effective records management deployment, and its absence often results in failure.

A successful records management system must suit the organization's workflow, be easily adaptable by users, and be integrated into their daily activities, ideally transparently, as most users will not be able to remember, nor care, about what constitutes a document of record or the correct tags to apply.

From a management perspective, the system must achieve the information lifecycle management goals of the organization and effectively integrate governance, risk, and compliance into a single corporate wide objective that can realistically be achieved. Concept Searching does both.

eDiscovery, Litigation Support, FOIA

Compound term processing identifies how data items are related, as well as the meaning of the content. For example in a typical search scenario, many words have multiple meanings. If the search was for "triple heart bypass", each word in isolation has several meanings. Using compound term processing, a search for "triple heart bypass" will locate documents about this topic even if this precise phrase is not contained in any document. The results will extract the key concepts, in this case "triple heart bypass", and use these concepts to retrieve relevant documents containing concepts such as "heart surgery", "coronary artery bypass", or "open heart surgery". With the identification of these highly correlated topics in the form of keywords, multi-word fragments, and compound terms, the result is automatically generated intelligent metadata that is unique to the organization.

For legal or FOIA professionals, the teams can see the concepts within the context of the document and appropriately decide if the document is relevant. For eDiscovery, the identification of potential keywords to aid in the negotiations at the 'meet and confer' conference can be simplified. This ability to capture the relationships between multiple documents, automatically understand the concepts, and group the results together not only reduces efforts for legal professionals, and time expended on the identification, collection, and review phase, but also provides a comprehensive approach to better manage unstructured information within the organization and more readily facilitates information governance, which is a high priority in the legal profession.

Migration

Organizations are requiring more sophisticated techniques to ensure compliance objectives are met, and a typical loop hole experienced is in the migration process. It is estimated that 69% of the data most organizations keep can, and should, be deleted¹⁵. Simply moving documents from one repository to another is not enough, as content that was typically unmanaged will remain unmanaged, continuing to expose an organization to risk. Information cannot be managed from inception to deletion without comprehensive metadata associated with the content, and incorporating the multiple channels and origination points from which content was received.

"Compound term processing identifies how data items are related, as well as the meaning of the content. With the identification of these highly correlated topics in the form of keywords, multi-word fragments, and compound terms, the result is automatically generated intelligent metadata that is unique to the organization."

“Simply moving documents from one repository to another is not enough, as content that was typically unmanaged will remain unmanaged, continuing to expose an organization to risk.”

With an intelligent metadata enabled solution, migration tasks are simplified and the accuracy of the migration is greatly improved. As content is migrated, it is analyzed for organizationally defined descriptors and vocabularies, which will auto-classify the document to the appropriate repository for review and disposition.

General migration tools cannot safeguard document confidentiality because they do not make intelligent taxonomy workflow decisions based on the text content of the document. If this security profiling is not performed during migration, then many of these documents will be retrievable during the search process, breaching the relevant document security obligations.

Collaboration, Business Social Applications, and Text Analytics

The ability to generate compound term metadata is not limited to only business critical applications. The technologies can readily be used to improve collaboration, provide a framework for business social applications, automate social tagging, and to some extent provide basic text analytics capabilities.

Summary

Organizations can no longer cope with growing challenges in search, the possibility of data breaches, compliance, and the typical issues associated with migration. Topics such as information governance, text analytics, collaboration, social intelligence, and auto-classification are now creeping into corporate jargon. Yet, for most organizations, these are not viable options until an enterprise metadata framework is in place to exploit the inherent value in unstructured content, making it useful to a variety of stakeholders for different purposes, and to the enterprise as a whole.

The organizations that will succeed will be those that recognize that unstructured content can no longer be ignored. Concept Searching’s unique compound term processing capability and automated classification enable organizations to address a wide range of challenges and improve business processes. Although the technologies are platform independent, the native, real-time integration with the SharePoint Term Store and feature rich set of taxonomy tools lower costs, manpower, and reduce development.

Concept Searching’s platforms deliver the only metadata generation auto-classification and taxonomy management solutions for improved search and information governance, integrated with SharePoint on-premises, SharePoint Online, and OneDrive for Business. Organizations can rapidly achieve an enterprise metadata framework using the most sophisticated technology available that will address information governance enterprise wide, enforce policies, and improve a host of applications.

Delivering a quantifiable return on investment, enterprises are using the technologies to provide an enterprise metadata repository that is consistent, scalable, and manageable; protect the organization where compliance is mandatory; reduce the costs associated with poor findability in search; eliminate manual tagging; and facilitate the deployment of intelligent metadata enabled solutions, in any environment.

Concept Searching offers the only viable and comprehensive choice for transforming unstructured content into knowledge assets that improve the bottom line. This is an immediate and long-term strategy that will consistently provide a significant return on investment.

About Concept Searching

Concept Searching is the industry leader in advanced semantic metadata generation, auto-classification, and taxonomy management. Its award winning products are the only statistical metadata generation and classification technologies that use compound term processing to generate intelligent metadata from unstructured and semi-structured data. Compound term processing, or identifying ‘concepts in context’, solves a variety of business challenges. Using the concept identification capabilities, organizations can transform content into business assets to improve performance.

Concept Searching’s Smart Content Framework™ for information governance is a combination of best practices and underlying products that encompass the entire portfolio of unstructured information assets, resulting in increased organizational performance and agility. The output from the Smart Content Framework™ delivers intelligent metadata enabled solutions that are being used to enable concept based searching, automatic declaration of documents of record, identification and protection of privacy and confidential data, intelligent migration, content management, granular identification of content for text analytics, and improved delivery of social content. The solutions are deployed in diverse industries, Fortune 1000 companies, and smaller companies that need to meet strict compliance, data privacy, and information governance regulations.

Concept Searching has a Microsoft Gold Application Development competency and is a participant in the global Business-Critical SharePoint program. Although platform independent, the Concept Searching Microsoft suite of products uses a single code base, supporting all versions of SharePoint, SharePoint Online, and OneDrive for Business, providing clients with the choice of on-premises, cloud based, or hybrid environments to best meet their needs.

Headquartered in the US, with offices in the UK, Canada, and South Africa, Concept Searching solves the problem of finding, organizing, and managing information capital. For more information about Concept Searching’s solutions and technologies please visit www.conceptsearching.com and follow on [Twitter](#) and [LinkedIn](#).

¹ [IBM, Improve Decision Making with Content Analytics](#)

² [AIIM, Connecting and Optimizing SharePoint – important strategy choices](#)

³ [IDC, Hidden Costs of Information Work, Sue Feldman](#)

⁴ [C. Cleverdon](#)

⁵ [Sarah Currier, Perspectives on Metadata](#)

⁶ [IDC, Hidden Costs of Information Work, Sue Feldman](#)

⁷ [AIIM, 29 Warning Signs of Digital Disruption, John Mancini, President](#)

⁸ [IDC, Hidden Costs of Information Work, Sue Feldman](#)

⁹ [Ponemon Institute, 2014 Cost of a Data Breach](#)

¹⁰ [Workshare, 68 Percent of Employees Expose Critical Corporate Metadata by Mistake](#)

¹¹ [ASG Software Solutions, Torsten Volk, VP Product Management](#)

¹² [M-Files, Infographic: Risks of File Sharing](#)

¹³ [Lorrie Luellig, Of counsel, Ryley Carlock & Applewhite, PC](#)

¹⁴ [Forrester Research/ARMA International Survey](#)

¹⁵ [‘2012 Compliance, Governance and Oversight Counsel \(CGOC\) Summit](#)



Microsoft Partner

Gold Application Development

© 2015 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



www.conceptsearching.com