conceptSearching

Empowering Knowledge in Professional Services 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

Martin Garland
President of Concept Searching
+1 703 531 8567
info-usa@conceptsearching.com

March 12, 2012

conceptSearching

Abstract

Perhaps no other industry is more dependent on maximizing the use of information assets and knowledge capital than professional services and consulting organizations. Knowledge, whether tacit or explicit, is the single most valuable asset of a professional services firm. The inability to effectively manage knowledge and utilize the knowledge assets results in a high cost to not only the firm but also to clients. Improving the access and use of knowledge as well as combining access to the highly specialized expertise of consultants can result in improving organizational returns linked to that knowledge.

Concept Searching has several global professional services clients who have deployed the technologies to achieve the management of their unstructured knowledge assets. Concept Searching's **Smart Content Framework**™ addresses many of the unique issues in the professional services industry. The **Smart Content Framework**™ is a multi-disciplinary solution delivered through the Concept Searching technologies that encompasses the entire portfolio of information assets resulting in increased organizational performance and agility. The framework has proven to be a flexible solution to address recurring problems in organizations of any size or industry.

Utilizing the framework are technologies that provide the ability to transparently tag content, deliver insight into search, classify it to organizational taxonomies, preserve and protect information through the automatic identification of records and privacy data, and as a migration tool.

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 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.

Overview

Perhaps no other industry is more dependent on maximizing the use of information assets and knowledge capital than professional services and consulting organizations. Knowledge, whether tacit or explicit, is the single most valuable asset of a professional services firm. The inability to effectively manage knowledge and utilize the knowledge assets results in a high cost to not only the firm but also to clients. Improving the access and use of knowledge as well as combining access to the highly specialized expertise of consultants can result in improving organizational returns linked to that knowledge.

With the ever increasing mobility of consultants across dispersed geographies the challenge is to make knowledge sharing and distribution an easier and more transparent process creating a holistic view of knowledge assets regardless of where the information is stored or the location of staff. Although not restricted only to consulting firms, enterprises still suffer from silos of redundant client and prospect data located in a variety of applications and diverse repositories designed to support the individual needs and requirements for a specific community of users. This inability to find knowledge assets for re-use and to spur thought leadership can result in a loss of clients, decreasing returns, and ultimately a loss of revenues.

The Challenges

Metadata management is a critical process that can address the inconsistencies and ambiguities in data to improve organizational and individual performance. The management of metadata is a significant challenge and encompasses not only the quality but also the efficiency of the metadata. Without meaningful metadata, content remains unusable and remains difficult or impossible for individuals to identify and retrieve relevant content.

Metadata can only be meaningful based on the ability to clearly identify what the content contains. Unfortunately, many organizations do not differentiate between metadata management and search technologies, and even with a robust search tool the individual is still required to ask the right question or match the appropriate keywords based on the assumption that if the right word is used, the right content can be found.

Compounding the problem is information overload with the retrieval of information that is not relevant and time consuming to evaluate, often resulting in reinvention of information instead of re-use. To overcome findability challenges often complex processes to enforce governance on the individual to accurately apply metadata are deployed resulting in a haphazard and subjective approach that does not effectively address the problem.

According to Gartner Group, over 80% of business is conducted on unstructured information and doubles every three months. A company with only 500 employees will lose a minimum of $2.4 million dollars per year due to inefficiencies associated with managing and finding content. In addition, for every $1 spent creating a document $10 is spent managing it.

The issues associated with managing content revolve around the ability to find it when needed, reuse it, and repurpose it.

- 85% of documents are never retrieved during a search
- 50% of documents are duplicated
- 60% of stored documents are obsolete
- Keyword search captures only 33% of relevant information
- Knowledge workers spend 25% of their time searching for information and 15% of their time duplicating information that already exists

Solving Findability and Content Management Challenges

Concept Searching's proven approach incorporates the **Smart Content Framework™** that was developed as 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 the Concept Searching technologies that encompasses the entire portfolio of information assets resulting in increased organizational performance and agility. The framework has 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, classify it to organizational taxonomies. The framework is being used to improve search, in records management, enterprise metadata management, compliance, and governance. Providing a complete solution including automatic semantic metadata generation, automated classification, and taxonomy tools results in a flexible approach that significantly improves management and access to unstructured content.

The framework building blocks consist of Metadata, Insight, Governance, Policy, Privacy, and Enterprise/Web 2.0.

Building Block #1: Metadata
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 the management of the lifecycle of content. From this, enterprise search, compliance, records management, and data privacy issues can be addressed and managed.

Building Block #2: Insight
Whether the enterprise search solution is SharePoint, FAST, or any enterprise search engine, the delivery of meaningful results depends on the ability to effectively index and classify content and to develop taxonomies to better manage the content. The search engine itself provides the features, functions, and interface while the classification structure delivers relevant results.

Building Block #3: Governance
The overarching enterprise governance structure allows staff to work in the most efficient and effective way possible by giving them access to information assets in a controlled and secure manner. This building block consists of tools that ensure information quality, maintains the lifecycle of information, addresses the retention and disposition of records, secures and protects privacy, and establishes standards when dealing with information assets.

Building Block #4: Policy
The application of policy must be deployed from an enterprise perspective and address the entire portfolio of information assets. A twofold approach, the policies must be created and then followed. Inhibiting success has been the inability to monitor for compliance based on the enterprise policies. Policy goes beyond the identification of concepts, records, and privacy but utilizes the technologies to effectively process the information and disposition through automatic application of policy.

Building Block #5: Privacy
The demarcation of who is responsible for the protection of privacy data has become blurred. Each business function may have a unique view of what is confidential, such as legal, human resources, and product development. It is the responsibility of the organization to set the policies and the responsibility of the accountable stakeholders and ultimately the end users to protect and hold confidential certain information assets. This building block addresses the end user privacy processes through elimination of end user participation, and the identification and routing of confidential and/or privacy data to the appropriate secure repository.

Building Block #6: Enterprise and Web 2.0
Social networking tools that encourage collaboration can link employees, partner, suppliers, and customers to share information is becoming a useful tool for business communication and business to consumer. There are several concerns such as security, unauthorized use, and communication noise. One of the biggest weaknesses is losing control of content and potentially impacting the corporate brand. The framework adds structure to chaos and injects control over the collaboration processes.

Semantic and Compound Term Metadata Generation

The metadata generation issue is increasingly a growing concern in professional consulting services firms. A comprehensive approach requires more than syntactic metadata and requiring end users to add rich metadata is haphazard and subjective at best. Concept Searching's unique *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 unique organizational content and work in any language regardless of vocabulary or linguistic style.

Instead of identifying single keywords, compound term processing identifies multi-word terms that forms a complex entity and identifies them as a concept. By forming these compound terms and placing those 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 concept 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.

Compound term metadata can be automatically generated either when the content is created or ingested. 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.

Automatic Metadata Generation at the Desktop

Defining the relationships within content has traditionally been done by less than perfect automatic classification systems, resulting in less than perfect metadata. Forcing the end user to enter meaningful metadata is usually unsuccessful. conceptClassifier is fully integrated with Microsoft Office and can automatically generate semantic metadata based on the concepts found within the content. Optionally, the Subject Matter Expert (SME) can manually classify documents and modify the results from within the traditional Microsoft Office interface. The automatic classification returns not only single words but identifies concepts within the document to assist the knowledge worker in the classification process. This guided approach enables the knowledge worker to precisely and accurately classify the document for reuse and retrieval. Placing the ability to classify documents into the hands of knowledge workers results in rich and comprehensive metadata, significantly improving the organization's ability to leverage their information capital.

Automated Classification

The automated classification function classifies content that has been tagged with the highly relevant metadata associated with the organizational taxonomies. This eliminates all costs and human intervention associated with manually tagging documents for classification and results in information that is categorized in real-time. This speeds the identification and collection of business, research, client and prospect information, legal, and regulatory records from multiple sources and enables the identification of new information to be captured and classified as content is created or ingested making it immediately available to authorized users.

Flexible and Feature Rich Taxonomy Development

Due to the explosion of electronic content, organizations need to develop a comprehensive approach to identify, organize and retrieve content assets. Internal communities need to find and reuse content rather than recreate it or make do without it. Implementing a framework that enables identification of content for reuse provides a return on investment for the time and effort spent to originally produce the material as well as achieve productivity improvements through information reuse and repurposing.

Consulting firms have a wide variety of stakeholders that need access to content to meet different needs. Internal vocabularies are often specific to that community and may not be easily translated by personnel outside of the community rendering the content unusable. Further complicating matters is that within the organization there may be varying solutions for identifying and storing electronic documents. The inconsistency of these systems hampers the ability of users to find relevant information, specifically when searching across multiple silos of content within the organization. Although knowledge workers need unified and universal access to information, at a more granular level they need to be able to find exactly and only the content they need.

Concept Searching's robust automatic classification and taxonomy management tools are designed to provide as much depth or hierarchical granularity that is often needed. Since the automatic semantic metadata generation has the ability to identify concepts as opposed to keywords, documents with the same concept can be classified against multiple nodes within a taxonomy or multiple taxonomies. From an end user perspective, knowledge workers can locate pertinent information from his or her own individual viewpoint without knowing the exact search terms to use. The easy to use taxonomy and automatic classification features function as a labeling mechanism to quickly create the foundation that can be altered to suit the requirements of the organization.

Taxonomies by nature are organic as they reflect the current state of knowledge by an organization as content is continually changing. Concept Searching's taxonomy development tools can address the fluidity of content changes to ensure that the taxonomy remains current and is easily managed. Providing both automatic and manual classification, Subject Matter Experts (SME's) can utilize rich features such as node weighting, ability to see the 'concepts in context', ability to search the corpus, auto-clue suggestion for categorization, and instant feedback on the impact of changes. Traditional taxonomy tools often require significant investments in time, expertise, and money to develop and maintain. Concept Searching's taxonomy management tool has been proven to reduce the time to build and subsequently maintain taxonomies by up to 80%.

Search and Navigation

Traditional search products force the end user to ask the right question, using the right combination of keywords, and repeating that iteration until the content is found. This not only hampers the user ability to find relevant information and reduces their productivity, but also impacts the organization financially. Information can't be used because it can't be found, information that can't be found has no value, and information that is hard to find is rarely used.

Utilizing the technology framework, end users can now search on concepts, delivering a multi-dimensional view of relevant information and the relationships between content assets that may not have been found. Offering taxonomy navigation and faceted navigation the interface can be customized to deliver relevance ranking, dynamic summarization, and search refinement.

Expert Identification

The ability to leverage existing knowledge and link the information to the experience of Subject Matter Experts enables the organization to derive value from human information capital.

The flexibility of the technology can be used to build a knowledge base of expertise within the organization. Providing the ability to aggregate information regarding an individual's experience from disparate data stores and return the results based on any number of parameters such as projects worked on, resume, work performed, work products produced, corporate profile, biography, and even documents that were authored that contain the search criteria entered by the end user.

Integration with Document Management Systems

Document Management Systems (DMS) manage content at the macro level. The true usefulness of the content however may reside at the word, sentence, or paragraph level. To overcome this limitation, Concept Searching provides connectors to the most widely used document management systems and the inherent security of the DMS is fully respected as the document access control lists and file information are also tagged with metadata. When a user selects a link to open a document, the DMS server can again authenticate the user credentials.

Connectors

Connectors are available to EMC Documentum, to websites, file systems, Microsoft SharePoint, Interwoven Worksite, Hummingbird, Lotus Domino.Doc, and Microsoft Exchange.

Microsoft Integration

As the only Microsoft managed partner in the taxonomy and auto-classification space, the Concept Searching Microsoft suite of products are unique in their native integration with the SharePoint 2007, SharePoint 2010, Full Term Store Integration, FAST Search, Windows Server 2008 R2 FCI, Office 365, and Microsoft Office applications.

Within the SharePoint 2007 and 2010 environment, our clients have deployed conceptClassifier for SharePoint to address a wide range of business issues including: increased accuracy of search results, records management, compliance, governance, enterprise metadata management, sensitive information identification and protection, and migration. conceptClassifier for SharePoint is the only automatic document classification and enterprise class taxonomy management solution that natively integrates with SharePoint 2010 and the Term Store.

Why is SharePoint Integration Important?

- SharePoint 2010 has no ability to automatically create and store classification metadata

- SharePoint 2010 has no taxonomy management tools to manage, test, and validate taxonomies based on the term store

- SharePoint 2010 has no auto-classification capabilities

- SharePoint 2010 has no ability to generate semantic metadata and surface it to the FAST or SharePoint Search engine, or any search engine, to improve search results

- SharePoint 2010 has no ability to automatically tag content with vocabulary or retention codes for records management

- SharePoint 2010 has no ability to automatically update the content type for records management or privacy protection and route to the appropriate repository

conceptContentTypeUpdater

Organizations can rapidly develop and deploy taxonomies and controlled vocabularies associated with business processes using conceptContentTypeUpdater along with the conceptClassifier for SharePoint taxonomy manager component. For example, creating a taxonomy that mirrors the records management file plan or one that addresses vocabulary and descriptors found in confidential and privacy data. conceptContentTypeUpdater uses records retention codes, semantic, and security metadata associated to data assets to identify and process the automatic application of content types. Once documents have the appropriate content type, based on natural language and organizational descriptors, workflows can be initiated.

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.

Technology

The technology is SOA compliant and delivered as Web Parts. Due to the SOA compliance and Web Services architecture we also support applications developed on any platform. The API is based entirely on Web Services and all information is exchanged in XML. The taxonomy formats are based on Web Ontology Language (OWL). Since the server is stateless is also works with all failover and load balancing hardware and software.

Transforming Knowledge Assets into Improved Business Outcomes

Drive Innovation

- Identify new relationships within content and discover new insights
- Encourages through leadership instead of intervention
- Intuitive and requires no training

Improves Business Agility

- Knowledge and expertise can be shared
- Improves collaboration and communication
- Increases productivity through integration of content, people, and external sources
- More rapidly assemble engagement methods and intellectual capital from previous building blocks

Reduce Organizational Risk

- Ensures all content is automatically meta-tagged
- Enables the retrieval of relevant information and identification of highly correlated content that normally would not be found
- Reduces time and cost associated with client engagements

Improves Productivity

- Reduce time spent finding information
- Ability to re-use and repurpose existing content
- Expedites access to real-time information

Organization Learning and Collaboration

- Harvest and reuse intellectual capital within and across engagements
- Identify and capture ideas for new services
- Increase efficiency and effectiveness of delivery methods

Increases Competitive Advantages

- Maximizes use of both tacit and explicit knowledge capital
- Provides insight into new service offering requirements and organizational capabilities

Improving findability, better managing content, as well as improving outcomes in research and discovery, time-to-market, engagement and practice management, and organizational learning and collaboration, Concept Searching's flexible and innovative **Smart Content Framework**™ and unique technology solutions deliver rapid and quantifiable results.

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

Microsoft Partner

Gold Independent Software Vendor (ISV)
Silver Content Management
Silver Search
Silver Portals and Collaboration

© 2012 Concept Searching