

# conceptSearching

## Compound Term Processing

### Delivering the Ability to Identify and Correctly Weight Multi-word Concepts in Unstructured Text

Business agility, informed decision making, improved business outcomes, and corporate governance are challenges becoming inseparably entangled in the inability to manage content. Enterprises continue to struggle with lending structure to content that can be tightly coupled to business processes.

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 Problem is Metadata

The challenge with metadata is both obvious and elusive. The ability to harness the meaning of content must utilize tools that enable the ability to manage and retrieve content at the same rate that it is being created, ingested, and distributed. The fundamental factor is the quality of metadata which is used by many applications within the organization.

What are the problems associated with metadata tagging?

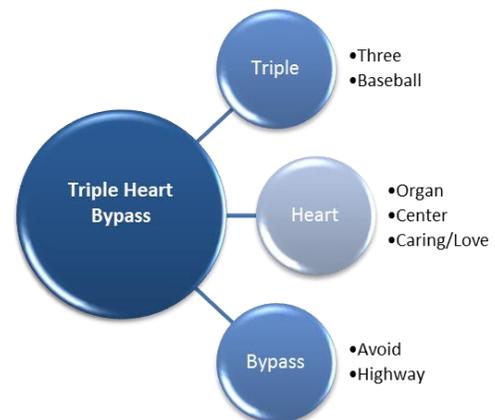
- Insufficient meta tags
- Ambiguous meta tags
- Subjective meta tags
- No meta tags
- Relying solely on system generated metadata tags
- Relying on the end user to consistently select the correct meta tags such as from drop down lists

#### What is Compound Term Processing?

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

For example, in a typical search scenario, many words have multiple meanings as depicted in the following graphic.

Results from an end user query, depending on the search engine, can return keywords only, the query words that appear close together (proximity), results based on other end user queries (boosting), and other search enhancement technologies. In this example using compound term processing, 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. 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, such as "heart attack" or "coronary artery surgery".



Compound Term Processing is a technology to identify and correctly weight multi-word concepts in unstructured text

#### What Are the Gaps Caused By Poor Metadata?

Taxonomies are not new. Yet many organizations do not even have a defined structure to organize unstructured and semi-structured content. Typically, many organizations have home-grown solutions based on departments, locations, and business function. In the past, although not optimal they worked. That is no longer the case. The resulting gaps lead to non-compliance, increased risk, and reduced organizational performance.

For solutions that use auto-classification the classification is either highly general (i.e. this document comes from Finance), or 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.

# conceptSearching

Where do these gaps occur?

- Poor search results, inability to collaborate and share knowledge
- Records Management
- Information Security
- eDiscovery and Litigation Support, FOIA processing
- Migration
- Text Analytics

## How Does Compound Term Processing Fill the Gaps?

The core Concept Searching technology is built around compound term processing technology that recognizes and identifies phrases that represent a concept. The supporting technologies that transform the metadata into improved business processes include auto-classification and taxonomy management.

Falling under Concept Searching's award winning Smart Content Framework™ these three technologies form the infrastructure component to support and manage unstructured and semi-structured content throughout the enterprise and for use by specific business applications.

The key enabler is the Intelligent Enterprise Metadata Repository, or taxonomy, from which improvements to any business application that requires metadata can be deployed.

## Enterprise Search

- Using any search engine enable higher order concept based search and browse capabilities
- Improve relevance and the end user experience

## Records Management

- Automatically identify and tag documents of record and route to the appropriate Records Management application for preservation and destruction

## Data Privacy

- Identify and secure any sensitive information in real time and automatically remove from search and access by unauthorized users

## Migration

- Automatically identify records that were not declared, as well as unknown data privacy exposures
- Identify concepts within content to classify to the appropriate structure
- Maintain security on migrated content
- Identify redundant content that should be deleted or archived
- Leave in place or migrate to the appropriate repository

## Text Analytics

- Leverage and identify high value content from terabytes of unstructured data

## Social Networking

- Improve collaboration and knowledge sharing by providing structure and leveraging existing knowledge within the organization
- Ability to identify and link new terms and vocabularies to provide consistency and preservation of historical terms

## The Benefits of Compound Term Processing

The key benefits of compound term processing are that it is an adaptive technology that can generate conceptual metadata at source, is not based on keywords, proximity, or algorithms that can't be changed, and can identify very specific criteria for business applications that use the metadata. The key differentiator is the ability to identify and correctly weight multi-word concepts in unstructured text. The end result is an enterprise metadata repository containing a rich set of intelligent metadata that reflects the unique terminology and vocabulary of the organization.

### Intelligent Enterprise Metadata Repository



Intelligent Metadata Generation as content is created or ingested  
Auto-classification to one or more taxonomies  
Automatic workflow and routing as defined by the business process



© 2013 Concept Searching

### Americas

+1 703 531 8567

info-usa@conceptsearching.com

### Australia

+61 (0)2 8006 2611

info-australia@conceptsearching.com

### Europe

+44 (0)1438 213545

info-uk@conceptsearching.com

### New Zealand

+64 (0)4 889 2867

info-nz@conceptsearching.com

### Canada

+1 703 531 8567

info-canada@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

www.conceptsearching.com