KM WORLD

4 Capabilities for Becoming Information-Driven

May 1, 2018
Scott Parker
This article is part of the Best Practices White Paper Text Analytics

Sinequa helps our customers to become information-driven, which means having actionable information presented in context to surface insights, inform decisions, and elevate productivity, consistently and reliably.

Let's delve into the four key capabilities that enable organizations to be information-driven.

Incorporating Context

Being information-driven means not only having a comprehensive view of the information available, but also the best possible understanding of the meaning it carries. If being information-driven were all about fielding queries and matching on keywords, a simple indexing approach would suffice. The best results are obtained when multiple indexes are combined, each contributing a different perspective or emphasis. Of course, we at Sinequa believe the following indexes, designed to work in concert, provide the best results:

♦ A full text index contains statistics of key terms and descriptions about how those terms appear within documents. These statistics help identify the importance of these terms.

♦ A structured index contains the structured metadata of documents (titles, dates, authors, keywords, document version, etc.) whenever that metadata is available. This index also contains ACLs (Access Control Lists).

♦ A semantic index focuses on the meaning of the information in the text by leveraging advanced processing algorithms to extract concepts using a semantic dictionary in multiple languages.

Being information-driven also means having a solution that is enterprise-grade, meaning:

♦ Strong Security Controls—designed from the start to support the complexities and multiple security layers of today's enterprises.

♦ Contextual Enrichment—the ability to learn your vernacular so that it operates within the context of your industry and the language of your organization.

♦ Relevance Feedback–as searches are performed, information is captured about the system interacting with the end user and incorporated back into the system. All subsequent searches leverage the information captured to maximize the quality of responses returned. This way, the quality of information returned improves as the system learns what documents are most used and as users indicate the value of documents available in the system.

Connecting information along topical lines across all repositories allows information-driven organizations to surface the collective expertise of the organization and make it transparent. This is especially valuable in large organizations that are geographically distributed. By connecting people with expertise, the overall responsiveness of the organization increases. This means everywhere, from the folks driving innovation in Research and Development, to the Service and Support folks helping customers, to the Marketing and Sales folks bringing in new business. The results can be spectacular. Time to proficiency decreases as new employees, and even existing employees learning new skills, have ready access to the expertise needed to take things to the next level.

Surfacing Meaning

By connecting related information across dispersed applications and repositories, folks can leverage 360-degree views and have more confidence that they are getting holistic information about whatever topic they are interested in, whether it be a specific customer, a service being provided, a sales opportunity, or any other business entity critical to driving the business. A key to connecting information and surfacing meaning is Natural Language Processing (NLP), which performs several important functions, including:

♦ Automated language detection

♦ Lexical analysis (part of speech tagging, compound word detection) and syntactical analysis (disambiguation, lemmatization of nouns, verbs, adjectives)

♦ Automatic extraction of dozens of entity types, including Concepts and Named Entities like people, places, companies, etc.

♦ Text mining agents integrated into the indexing engine that detect regular expressions and/or complex "shapes" that describe the likely meaning of specific terms and phrases and then normalize them for use across the enterprise.

Self-Learning

As we know, Machine Learning (ML) and even Deep Learning (DL) are becoming increasingly critical to enhance and improve search results and relevancy. This is done during ingestion but also constantly in the background as humans interact with the system. The reason ML has become essential in recent years is that it can handle complexity beyond what's possible with rules. ML helps organizations become information-driven by:

♦ Analyzing and structuring content to both enrich and extract concepts such as entities and relationships, which can subsequently be used for further enrichment and human navigation

♦ Modifying results through usage, essentially incorporating human behavior into the calculation of relevance

♦ Providing recommendations based on both what is in the content (content-based) and by examining users' interactions (collaborative filtering)

Intuitive Presentation

Most organizations striving to be information-driven want to end up with an intelligent agent that employees can consult for institutional knowledge that can be readily applied to the task or situation at hand. However, even the most intelligent of agents will not be useful without a well-designed user experience. So, what is good design? At Sinequa, we know that good design is aesthetically pleasing and understandable in that it makes use of the user's intuition. We also know that good design is unobtrusive and perhaps most importantly, that good design is contextual to the user's goals.

These are the capabilities required by the information-driven organization, and unfortunately there are no shortcuts. At Sinequa, we believe these capabilities, intelligently combined, serve as the primary enablers for organizations seeking to become information-driven.

http://www.kmworld.com/