

The Value of Descriptive Metadata to Improve Enterprise Search

Why making sure your employees can find information is critical to an organization's bottom line and how descriptive metadata, taxonomy, and tagging will save money and increase competitiveness.

May 1, 2013

A WAND, Inc. White Paper

By Mark Leher, COO, WAND, Inc.



P.O. Box 757
Englewood, CO
U.S.A.
www.wandinc.com

Phone: +1 (303) 623-1200
Email: info@wandinc.com

Bad search is costing you money

In the 1980's, legendary management consultant Peter Drucker predicted that "The factory of tomorrow will be organized around information rather than automation."¹ Drucker's vision is today's reality: information the most valuable asset most organizations have and effectively managing information assets are critical a company's success.

There are two main types of information that most organizations need to manage: 1) structured data, which is usually in databases and often quantitative in nature; and, 2) unstructured data, or text. Structured data is handled fairly well with business intelligence tools and data warehousing. According to the Aberdeen Group, unstructured data can account for up to 80% of the information available to a business.²



"The factory of tomorrow will be organized around information"

Peter Drucker

Despite the importance of unstructured information assets, much of its value is locked up because employees still cannot find the information they need. In fact, 60% of information workers say it is difficult to find the right information³ and information workers are spending an average of 8.8 hours every week searching for information. This costs organizations more than \$14,000 per employee each year,⁴ meaning 22% of employee productivity and salary cost is going down the drain.

Poor findability is not only a financial drain, but more importantly a strategic liability. The quantity of unstructured information inside every organization is growing dramatically and companies who aggressively leverage information will have a significant competitive advantage. Gartner says that "by 2015 organizations integrating high-value, diverse, new information types and sources into a coherent information management infrastructure will outperform their industry peers financially by more than 20%."⁵ Employees agree: "75% of information workers say that finding the right information is critical to the organization's success."⁶ Companies who have not begun taking steps to make information easily accessible and findable for employees are falling behind the competition.

¹ Huang, K.T. (1998). "Capitalizing on intellectual assets." *IBM Systems Journal*. Volume 37, Number 4. Available at <http://www.research.ibm.com/journal/sj/374/huang.html>.

² The Data Chain. "Getting the best from big data: lessons from the enterprise search world". Available at http://www.thedatachain.com/articles/2012/7/getting_the_best_from_big_data_lessons_from_the_enterprise_search_world

³ "Enterprise Search and Findability Survey 2012." Findwise.

⁴ Boyd, Angele. January 2011. "Managed Print and Document Services for Controlling today's - - and Tomorrow's - Information Costs." IDC Executive Insights.

⁵ "A Framework for Creating Value From information Assets: The Key to Information Management Success", Gartner Symposium Nov 11, Ted Friedman and Regina Casonato

⁶ "Enterprise Search and Findability Survey 2012." Findwise.

Why knowledge workers can't find documents

With so much information, how can an organization make sure the right information is quickly available to its employees? Companies have deployed a variety of tools to organize information. At the basic level, most companies have created shared drives where users can create folders to store documents. These shared drives eventually end up as a convoluted mess of folders nested within folders, duplicate documents, and confusion as to where documents reside. More advanced approaches involve enterprise content management (ECM) software paired with an enterprise search engine. This is a big step forward, but users still experience frustration finding information because keyword search results bring back too many results and users are left to sift through pages of results to find exactly the information they are looking for.

A major deficiency of document management in organizations is the lack of metadata. Most people have heard of metadata, but might not be sure what it means. Metadata is simply data about data. Examples of metadata can be found if you right click on any document and choose the "View Properties" option. Figure 1 shows some of the metadata for this document. Many of the values have already been populated automatically. Author, Date Last Saved, Total Editing Time, Program Name were all populated. These automatically populated metadata values are called "Flag Metadata". Other flag metadata values not shown in the screenshot might include file size, file type, location, and more. Flag metadata is easy to get, but, unfortunately, it does not help users find the document. Nobody searches for a document that is 10 kb and was last saved on 5/2/2013.

The most important metadata for search and findability is "descriptive metadata". Looking back at Figure 1, the descriptive metadata (Title, Subject, Tags, and Categories) are all missing. Descriptive metadata is related to the "aboutness" or the subject/topic of the document. Descriptive metadata fields don't get populated automatically – it's up the user to add that data in and the overwhelming majority of people don't know how to populate these fields and even if they do, don't take the time to do so. Descriptive metadata is the most important metadata for effective search and information management, but typically, it is exactly the type of metadata that is missing.

Imagine somebody snuck into your office and removed all of the labels off of the folders in your filing cabinet. It would be nearly impossible to find anything. This is the situation most companies have with electronic information.

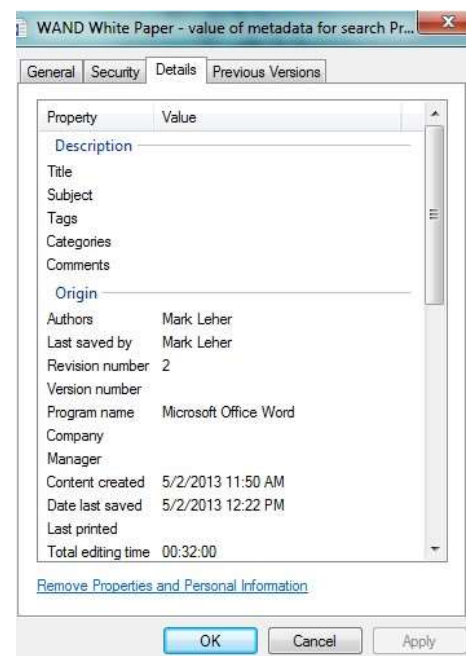


Figure 1

How descriptive metadata makes search better

We know that most documents are missing descriptive metadata, but how would having descriptive metadata make it easier for a user to find information? Descriptive metadata helps users find information in two primary ways.

First, Descriptive metadata helps highlight the most important concepts in a document for the search engine. If you imagine all of the keywords in a document, the descriptive metadata fields are a way to tell the search engine that certain keywords or concepts are more important than the rest. This gives users greater relevance for keyword searches.

Second, descriptive metadata can be used to populate left hand search result refinement capabilities so that users can take a large keyword results set and narrow down those results based on highly relevant descriptive metadata associated with the result set. When using Amazon, searchers can filter a search for “televisions” based on brand, price, color, and screen size to narrow down the result list to just

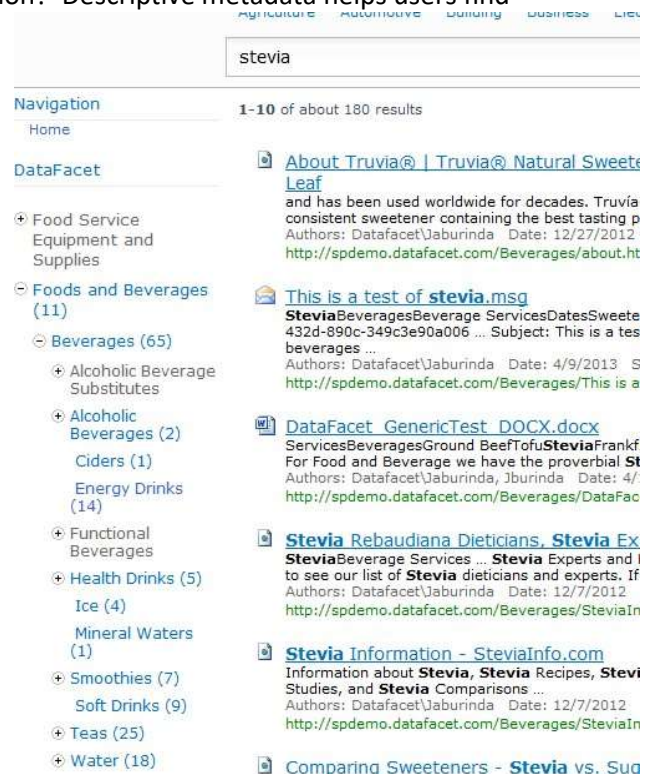


Figure 2

the televisions with only the desired features. Descriptive metadata allows a knowledge worker

to narrow down his search results to just include the documents with the desired subjects in the very same way that the user can narrow product search results on Amazon. Figure 2 shows an example of how metadata filters on the left hand side give a searcher the options to narrow search results and increase relevance.

How to get started with descriptive metadata

There are two key ingredients to bring descriptive metadata into an organization so knowledge workers can begin to find the information they need more quickly.

First, an organizational taxonomy is necessary to establish the set of terms that can be used as descriptive metadata values. Figure 3 shows an example of a taxonomy for sales and marketing. Each term in the taxonomy is a potential candidate to be tagged to a document as descriptive metadata.

Second, a strategy for tagging documents with descriptive metadata must be selected. One possible

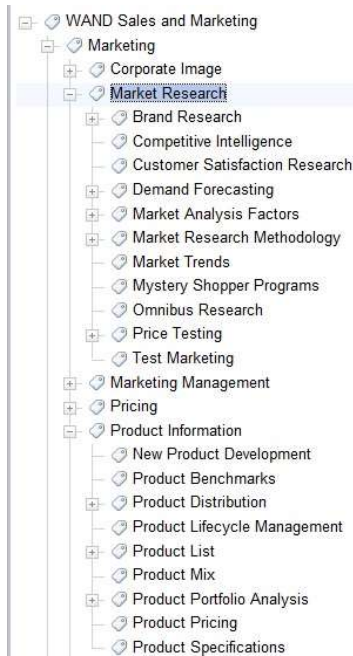


Figure 3

approach would be to require employees to add descriptive metadata when creating a document. The problem with this approach has already been defined above: employees do not like to add metadata to documents and they often do the job poorly. Instead of burdening employees with manual tagging, many companies are adopting an automatic tagging approach. Automatic tagging is becoming a widely accepted best practice as demonstrated by Gartner’s recommendation that “Metatagging (should be) automated as much as possible and (should) become a normal experience for users.”⁷

The organizational taxonomy and an automated tagging engine work hand in hand together to quickly label each document with appropriate and highly relevant business terminology. With both tools in place, a business can build a scalable, governable process for tagging documents so that knowledge workers experience higher search relevance than ever before.

Conclusion

As the amount of unstructured information continues to grow, organizations who have not addressed search relevance will continue to be at a competitive disadvantage. It’s more critical than ever to reduce the amount of time employees are wasting searching for information and, more importantly, to give knowledge workers a complete picture of the information the organization has so the best business decisions can be made. Enabling descriptive metadata with an organizational taxonomy should be investigated by every organization that is serious about solving this critical business issue.

How WAND can help you achieve ROI on improve search of unstructured information.

WAND has a library of pre-built foundation taxonomies covering a wide variety of business and industry vertical topics that can be used to jump start the process of putting an organizational taxonomy in place. The WAND Taxonomies are built to get your organization a taxonomy that is a 75-80% fit for your organization. From there, the taxonomy can easily be customized with terms that are specific to your organization such as departments, products, geographical

⁷ Tay, Gavin and Kenneth Chin. June 13, 2011. “Maturity Model for Enterprise Content Management.” Gartner. <http://www.gartner.com/technology/media-products/reprints/emc/213197.html>

locations, customer segments, or others as appropriate. Starting with a foundation taxonomy approach can save up to 12 months over building an organizational taxonomy from scratch.

Contact WAND today to learn how our solutions can help you deploy descriptive metadata in your organization to reduce your costs and increase your strategic competitiveness.