

Content Management Meets Text Analytics

Dr. Fern Halper, Partner





© Copyright 2009, Hurwitz & Associates

All rights reserved. No part of this publication may be reproduced or stored in a retrieval system or transmitted in any form or by any means, without the prior written permission of the copyright holder. Hurwitz & Associates is the sole copyright owner of this publication. All trademarks herein are the property of their respective owners.

□ 233 Needham Street □ Newton, MA 02464 □ Tel: 617 454 1030 □
www.hurwitz.com



Contents

Introduction.....	1
Text Analytics Overview.....	2
<i>What Is Text Analytics and How Does It Work?.....</i>	<i>2</i>
<i>Key Business Drivers for Text Analytics.....</i>	<i>3</i>
<i>How Are Companies Planning to Implement Text Analytics?.....</i>	<i>4</i>
Use Cases	5
<i>Voice of the Customer.....</i>	<i>5</i>
<i>Market/Competitive Intelligence.....</i>	<i>6</i>
<i>Product Design and Development.....</i>	<i>7</i>
EMC Products.....	8
<i>CenterStage Essentials and CenterStage Pro.....</i>	<i>8</i>
Conclusion	9

Introduction

Information is the lifeblood of any company. While organizations have made significant progress in analyzing their structured data (such as sales figures and number of complaints), the reality is that most company information is unstructured. This unstructured information includes claims, contracts, patents, call center notes, clinical trial records, and survey responses, which are often stored in enterprise content management systems. Content management systems are a treasure trove of information since there is a significant amount of insight that can be derived from this unstructured data. The flip side of this is that there is so much enterprise information that a knowledge worker (e.g. a product manager, marketing manager) can quickly become overwhelmed and not make the best use of it.

Let's examine one fictitious telecommunications company, GlobalTel, which has a customer retention problem. Company executives put together a small team of marketing managers, representing several divisions, to get to the root cause of the problem. The team knows that the answer to this question can be found in its call center notes, emails, and open-ended survey responses, which are stored in its enterprise content management system. These documents contain information such as customer questions, complaints and opinions that would be immensely useful in understanding the problem. Although the documents are stored by type (i.e. call center, email, etc.) and date, it is nearly impossible for the marketing managers to cull through all of this information manually. There is simply too much data and they cannot find what they need quickly and easily. The managers assemble teams to help read through all of this information, but eventually everyone gives up in exasperation. The problem was that these managers could not access and analyze the relevant information. While key word search could have helped to narrow down the number of documents that the team needed to read, it is not enough. Text analytics technology can help companies cull through the massive amount of unstructured information much more effectively and maximize the usefulness of their content management systems. Over the past few years, this technology has moved out of the early adopter phase to become mainstream.

In this paper we examine text analytics technology and how it can help companies effectively analyze and maximize the value of the unstructured information in content management systems. We first briefly examine what text analytics is and how companies are using the technology. We then provide some use cases illustrating the benefits of the approach. Finally, we describe CenterStage, a client for EMC's Documentum ECM platform, which incorporates this technology.

Content management systems are a treasure trove of information since there is a significant amount of insight that can be derived from this unstructured data.

Text Analytics Overview

What is text analytics and how does it work?

Hurwitz & Associates defines text analytics as:

The automated process of analyzing unstructured text, extracting relevant information, and transforming that information into structured information that can then be leveraged in different ways.

What does this mean? There are numerous methods for analyzing unstructured text. Historically, these techniques arise from natural language processing (NLP), knowledge discovery, data mining, information retrieval, and statistics. (See the sidebar on NLP on the following page.) This text can be extracted from the content found in unstructured sources and then iteratively analyzed to determine relationships and trends, look for clusters, and so on. The transformed information can also be combined with additional structured data (e.g., sales and demographic data) and analyzed using business intelligence or other analytical techniques.

In the GlobalTel example, the unstructured text in its content management system could be analyzed and information about various entities extracted. Figure 1, below, illustrates some of the unstructured text from call center notes that could be extracted and used to get at the root cause of the customer retention issue.

Customer wants to know if we have a calling plan like FamilyOne from LocalTel. It is cheaper.

Customer wants to know if we have a family plan. She says this would save her money. Explained GlobalTel doesn't yet have this type of plan.

Customer wants to understand feature bundles. Customer says LocalTel is having a great promotion on unlimited plan with different features.

Figure 1. Portion of GlobalTel Call Center Log

The underlined words in this simple example provide the information the company needs in order to understand the “why.” For example, the phrases calling plan, family plan and feature bundle indicate that customers are calling in with questions about plans. These three phrases (and others) could be grouped under the entity “Plan.” The company LocalTel is a competitor of GlobalTel and appears in a number of call center records that also mention plan. The text analytics process uses algorithms to analyze the unstructured text, extract this information, and

The text analytics process uses algorithms to analyze the unstructured text, extract this information, and utilize this information as part of an index in the content management system.

utilize this information as part of an index in the content management system. When the manager accesses the system, he or she would see these entities (“plan” and “LocalTel”), investigate them by reading the call center notes, and determine that its competitor is offering plans that its current customers want. Of course, GlobalTel might receive hundreds of thousands of calls a month and this particular type of call might account for less than 10% of the call volume. Individual agents would not necessarily pick up on the trend. And the managers certainly could not read through all of the call records. Even if they did attempt to do so, if the records were divided between them, they might not have spotted this particular trend. Similarly, if the manager simply searched on calling plan, he or she may have only seen a fraction of the notes relating to calling plan since, as we have just seen, there are multiple ways to refer to a calling plan.

In this example, text analytics could have also been used to extract more complex relationships such as customer sentiment. Additionally, the technology could have utilized other content sources, outside of the content management, system such as blogs and message boards to understand what consumers are saying.

NLP

In general, text analytics solutions use a combination of statistical and natural language processing (NLP) techniques to extract information from unstructured data. NLP is a very broad and complex field, which has developed over the last ten to twenty years. The goals of NLP are to derive meaning from text. NLP generally makes use of linguistic concepts such as grammatical structures and parts of speech.

Key Business Drivers for Text Analytics

The simple GlobalTel example illustrates how text analytics helps companies to gain more insight from their content. This is one of the key drivers for text analytics. In a 2007 study, Hurwitz and Associates asked approximately 100 large companies that were either deploying the technology or planning to deploy it in the next year what the key business drivers were for using it. The top five drivers are illustrated in the chart, below. They indicate that companies are clearly deploying text analytics for important business reasons such as making better decisions and gaining competitive advantage. Companies also realize the value in automating a part of the analysis process in order to reduce manual processing time. Several of these drivers are explained on the following page.

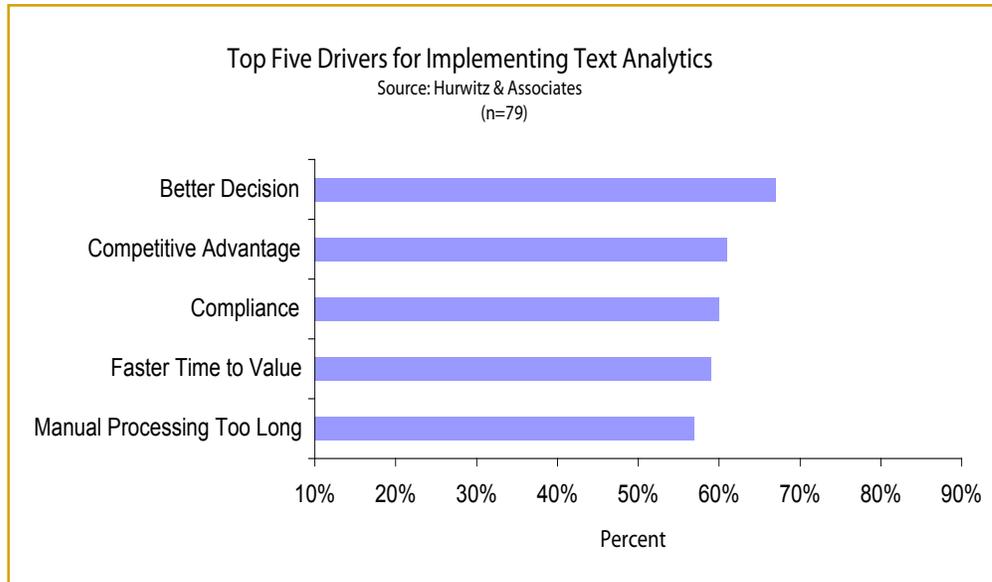


Figure 2. Top Five Drivers for Implementing Text Analytics

Better Decisions. Companies believe that the data to be found in unstructured sources will reveal information and relationships that have eluded them before but will help them make better business decisions. For example, the text from call center notes can help a company find problems in a product quickly enough to avoid losing customers.

Competitive Advantage. Companies believe that they can gain competitive advantage from the information they extract. Gathering information about competitors can help a company understand more fully what is happening in the market and manage its brand more effectively. It can give the company a good sense of its brand image and alert the company that a certain event has affected its image.

Faster Manual Processing. Organizations sometimes employ staff to manually sift through unstructured documents and pull out important information—a very time-consuming process. In discussions with customers they have told us that reading through this information is cumbersome. However, if someone doesn't read it, the information is lost. Text analytics technology helps to automate this process.

How Are Companies Planning to Implement Text Analytics?

In our 2007 study, we asked these companies how they were planning to implement their text analytics solution. A large majority of companies (62%) stated that they plan to implement text analytics in conjunction with their content management systems.



Content Management Meets Text Analytics

These companies realize that critical information is found in these systems that they need to utilize to their advantage. It behooves companies to utilize text analytics as part of a content management solution.

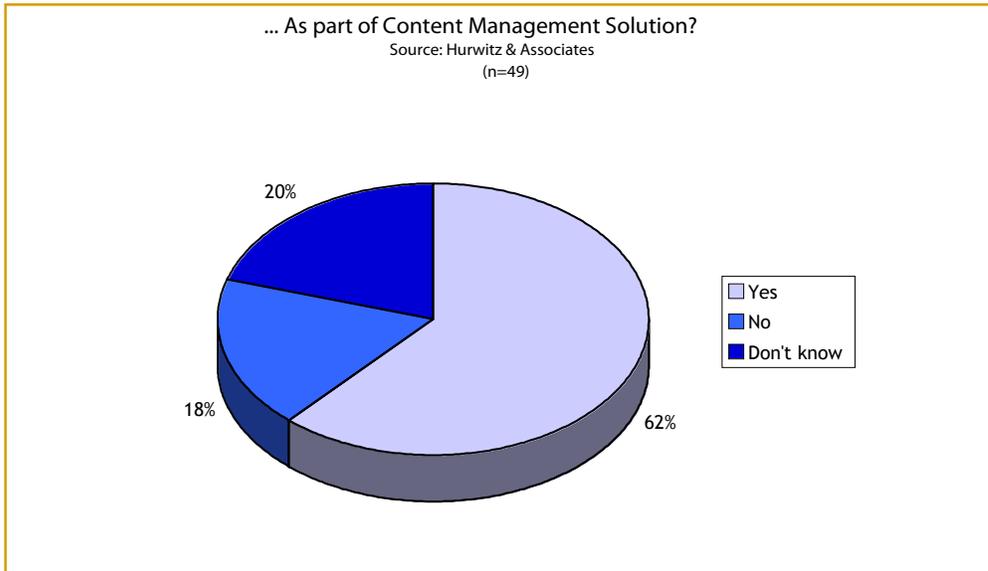


Figure 3. Text Analytics Used in Conjunction with Content Management

Use Cases

There are numerous situations where text analytics can help knowledge workers uncover relevant information that might have otherwise gone unnoticed. Voice of the customer, Competitive intelligence, and Product design and development are three areas where companies are using text analytics successfully. These three somewhat related applications make extensive use of end-customer information such as that found in emails, call center notes, surveys, warranty or claim forms. They can also make use of external information such as that found in news feeds, blogs, and message boards.

Voice of the Customer

Voice of the customer (VoC) solutions use text analytics to cull through information from customer interactions - surveys, emails, call center notes, payment histories, regulatory filings, and a myriad of content in a content management system - in order to understand what customers are saying about the company's products and service. Here, the company is usually trying to determine customer satisfaction with a product



Content Management Meets Text Analytics

or service. An important consideration in this use case can often be determining customer sentiment—both good and bad. Text analytics can pick up on words such as “unhappy” or “problem” to determine when a customer is less than satisfied about a particular product or service. We’ve already provided one example of this kind of use case in the beginning of the paper.

Of course, many industries use VoC applications to help understand what customers are thinking. For example, banks are very customer satisfaction oriented since it is fairly easy for customers to move their money from one bank to another. Financial services companies utilize email and call center notes to understand customer concerns. In addition to addressing retention issues, VoC can also be used to determine qualified leads for a marketing campaign. In the example above, if GlobalTel decided to provide some new calling plans, it could target those customers who were asking about the different plans as very good leads (especially if they have not already abandoned the service).

Customer feedback information is vital for helping companies improve customer satisfaction and retention rates which is critical in today’s economic environment. It can also provide information that can help companies be much more effective in marketing campaigns and other business activities.

Market/Competitive Intelligence

Companies need to be aware of how their market is changing. A big part of this is to understand what competitors are doing in terms of products and services as well as how these competitors are being perceived by the market. Good sources of information for this include content that would be available in the company’s content management system as well as content from external sources. This would include call center notes, email queries, survey responses as well as news feeds, patent filings, blog postings, and message boards. This information could be culled to determine new patents, product announcements, product advancements, and customer feedback about these new products.

For example, an electronics company might be interested in analyzing how competitors’ products are perceived by the market. Its market research managers utilize some of the company’s internal call center and survey verbatim to understand what its customers are saying about competitor products. It also makes use of news feeds to determine the type of coverage its competitors are getting and what is being written about the market in general. Finally, it makes use of blogs and message boards to understand customer sentiment regarding its competitors. Putting all of these sources of information together gives the electronics company a good indication of

Text analytics can pick up on words such as “unhappy” or “problem” to determine when a customer is less than satisfied about a particular product or service.



Content Management Meets Text Analytics

how its competitors are faring in the market. It provides insight into new product offerings, customer reactions to new product offerings, problems customers are having with company products, as well as a general sense of competitor brand reputation. All of this information is invaluable in helping the company to make the right decisions in terms of competing against these companies. Additionally, this information can also be used to help determine enhancements to existing products and the design for new products.

Companies use text analytics to aid in new product design and development.

Product Design and Development

Companies use text analytics to aid in new product design and development. Information from internal sources as well as external sources can help product managers determine what products and features to develop based upon known customer preferences and competitive differentiators. For example, a product manager and R&D group who are looking to design a new energy saving window might utilize multiple sources of unstructured data to determine features. The product manager might try to systematically track a competitor's patent filings and new product introduction in order to determine what these companies are bringing to market and how the market is responding to the new products. The R&D staff might be interested in culling through the literature to understand new advancements in energy saving technology. The team would also utilize internal information from call center notes, email queries, and warranty claims to determine customer preferences along with any product problems that need to be fixed. Survey notes might also be used to understand any other feedback that customers have provided in the open-ended portion of the survey.

Utilizing this information, the product development team would have a much better understanding of what features customers are looking for in a new product. Since the company has better insight into competitive offerings, it can use this information to help differentiate its own products. Text analytics can help companies get a much better handle on the right products to deliver to market.

In addition to the horizontal application areas listed above, companies are also using the technology for specific industry needs. For example, pharmaceutical companies are using text analytics as part of the discovery process. These companies are mining literature for new advancements, clinical notes for drug interactions, and so on. They are basing their decisions on what lab projects to undertake, in part, on this information. Insurance companies are using text analytics to help sift through massive amounts of insurance claim information in order to find patterns in the claims to help reduce fraud. Law enforcement is using text analytics to help detect criminal patterns. And the list goes on.

EMC Products

EMC has recently incorporated text analytics technology into its CenterStage offerings. EMC CenterStage is a new client built for Documentum. The goal of CenterStage is to reduce the barriers to adoption of content management and enable knowledge workers to use enterprise content more effectively. EMC currently provides two products in this space: CenterStage Essentials and CenterStage Pro. Both of these offerings sit on top of the Documentum platform and therefore can leverage important the content management services such as retention policy services, media transformation services, intelligent classification and so on. Both utilize a consumer-based navigation model.

CenterStage Essentials and CenterStage Pro

CenterStage Essentials is a rich Internet based client that provides users with access to the Documentum repository. Knowledge workers can use this interface to access a series of workspaces, according to job function or project team. Selecting a certain workspace will allow users to access documents that they have permission to access. CenterStage Pro builds on what is offered in CenterStage Essentials and provides more Web 2.0 type features including the idea of community spaces to share information via blogs, RSS feeds, and wikis. It also allows users to customize their pages and save them as templates.

CenterStage incorporates federated search capabilities. This means that users do not have to be limited to searching the Documentum repository to find the relevant information they require. In addition to searching Documentum, CenterStage's federated search capabilities allow users with the appropriate security permissions to discover information residing in corporate file shares, email archives, other content management repositories, SharePoint sites as well as the public web and private web sites. Additionally, the worker can use a series of filters (document type, size, date, etc.) that EMC provides as part of the user interface to narrow down the documents they are interested in accessing.

CenterStage also makes use of EMC's Content Intelligent Services (CIS), which is an extension to Documentum. CIS automatically categorizes documents into folders called "topics" by adding metadata to unstructured content. It does this by extracting information from the document and mapping it against a pre-existing taxonomy. With CenterStage, EMC has incorporated text analytics into the Documentum platform, enabling search queries, for example, to automatically extract entities such as company name, place, date, product name from the content stored in the platform. Those search results can be clustered dynamically by topic, date, or a number of

The goal of CenterStage is to reduce the barriers to adoption of content management and enable knowledge workers to use enterprise content more effectively.



Content Management Meets Text Analytics

facets. In the GlobalTel example, this capability would have allowed managers to view the entities *plan* and *competitors* in CenterStage and easily navigate to these documents to explore the information further. In addition, these entities could have been cross-referenced against any topics that were already tagged (e.g. current customers, lost customers) to further narrow the search and get to the root cause of the problem faster.

Conclusion

Today's reality is that we are inundated with information and the amount of information will only continue to grow. This means that the need for technologies, such as text analytics, that can help to derive insights from this unstructured information will also increase over time. Text analytics is an important technology that can help companies automatically extract and analyze relevant information from the mountain of documents stored in content management systems; this saves time, which knowledge workers can use in other ways.

Companies deploying the technology are reporting significant benefits. In our 2007 survey, the vast majority of those customers who were already using the technology felt that the business value was critical or high. Hurwitz & Associates believes that companies investing in content management systems need to consider whether these systems can support the analysis of the unstructured text. By leveraging a content management system that incorporates text analytics capabilities, companies are much more likely to foster innovation across their own organization – and throughout the extended enterprise - and thereby provide valuable insight into the collective intelligence stored in unstructured content. By not using analytic capabilities on a content management repository, organizations fail to unlock the full value of their unstructured information assets.

By leveraging a content management system that incorporates text analytics capabilities, companies are much more likely to foster innovation across their own organization - and throughout the extended enterprise...



About Hurwitz & Associates

Hurwitz & Associates is a consulting, research and analyst firm that focuses on the customer benefits derived when advanced and emerging software technologies are used to solve business problems. The firm's research concentrates on understanding the business value of software technologies, such as Service Oriented Architecture and Web services, and how they are successfully implemented within highly distributed computing environments. Additional information on Hurwitz & Associates can be found at www.hurwitz.com.