HURWITZ & ASSOCIATES

Expanding the Boundaries of Enterprise Content Management Systems

Content Analytics: Delivering Improved Business Insight and Performance

Dr. Fern Halper, Partner

A HURWITZ WHITE PAPER

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

Contents

Introduction

Nothing exasperates an executive more than asking a business critical question and not getting a good answer. An executive might ask questions like:

- Why are we losing customers?
- Why is our fraud rate so high?
- Why didn't we see this coming?

These types of questions often go unanswered if the company lacks a comprehensive Information Management strategy. The technology needed to implement an information management strategy will provide access to and analysis of systems data and the vast amount of unstructured data or "content" sitting in content repositories. To be successful, companies need to be able to leverage all information assets – structured and unstructured - so that the answers to these vital questions don't remain buried in the mountain of call center notes, claims forms, case files, and emails stored throughout the company. Very often it is this unstructured data that holds the key to unlocking the mysteries of fraudulent practices, serious product or service failures and the reasons behind customer dissatisfaction.

Even companies with comprehensive Enterprise Content Management (ECM) systems find that while they have excellent access to information about their business, it doesn't always provide them with the answers they are looking for. Traditional ECM systems are successfully used by many companies to help store, organize, classify and even search internal textual information. However, these ECM systems typically fall short when it comes to identifying patterns and trends in that information. This is problematic when an organization knows that something is wrong but isn't sure what the problem is. For example, there has been a drop off in sales without an explanation. The answer may only be found in customer dissatisfaction with a new company policy or an internal conflict with external business practices that is only surfaced in company correspondence like emails. Searching through highly structured and organized data is not likely to provide you with all the answers you need to make good decisions.

The technology needed to implement an information management strategy will provide access to and analysis of systems data and the vast amount of unstructured data or "content" sitting in content repositories.

Until recently, the ability to integrate this text and derive business insight wasn't commercially viable. Now, companies are considering content analytics technology as a powerful way to generate actionable information from the mass of unstructured data available to them. In fact, the technology is maturing out of the early adopter stage and becoming more mainstream. Factors fueling market adoption include a desire to strengthen competitive advantage and to make faster and more informed business decisions.

Now, companies are considering content analytics technology as a powerful way to generate actionable information from the mass of unstructured data available to them.

In this paper we examine content analytics technology and how it is transforming the way companies run their business. We will examine the approach to content analytics and provide some detailed use cases illustrating how content analytics works in the real world, and finally describe the IBM offerings in this space.

What is Content Analytics?

Enterprise Content Management is a mature technology that enables companies to leverage their unstructured information to support decision making. ECM capabilities historically have been employed to:

- Help manage, secure and provide access to all forms of unstructured data such as document images, emails, faxes, call center notes, electronic documents, eforms, and so on. Unstructured data is growing at a faster rate than structured data and this unstructured content accounts for more than 80% of the data that flows through a company.

- Improve corporate accountability, risk reduction, and regulatory compliance through content control and process visibility. Content management systems ensure that content is retained for the appropriate amount of time for operational and legal purposes and then destroyed in accordance with corporate policy.

- Streamline, automate and report on business processes to deliver operational efficiency and better visibility into how the business is performing. Combined with the power of business process management, ECM systems enable businesses to track, monitor, measure, report, and optimize business processes.

While ECM systems provide a great deal of value, what has been missing is the ability to analyze all this content to gain insight at a document level and across documents in order to understand important patterns and trends.

Let's take a look at an example. An automobile manufacturer stores warranty claims, reports, and technician notes in its content management system. Management wants to track issues related to unknown defects with cars as they enter the marketplace but the only structured information it has at its disposal is vehicle type, owner, part numbers and known defect codes. It is difficult to provide timely insight into problems with this kind of information. If the manufacturer could analyze the text found in the claims, reports and notes it stores in its ECM system it would have a much richer set of information about the status of its cars. In fact, the manufacturer might be able to detect an issue with a car before it becomes a major problem. For example, individual technicians might discover that some seat belts are locking without explanation. Each technician might think this is an isolated incident since there are no reports from corporate about such a defect. However, if company management had been able to analyze reports from technicians in the field, it would have become apparent that there was a pattern of unexplained seat belt problems.

While ECM systems provide a great deal of value, what has been missing is the ability to analyze all this content to gain insight at a document level and across documents in order to understand important patterns and trends.

Text Analytics and Content Analytics Defined

There are numerous methods for analyzing unstructured data. Historically, these techniques arise from natural language processing (NLP), knowledge discovery, data mining, information retrieval, and statistics. Hurwitz & Associates defines text analytics as the process of analyzing unstructured text, extracting relevant information, and transforming that information into structured information that can then be leveraged in different ways. We are using the term Content Analytics to denote a layer above the actual extraction process that analyzes this information to understand trends and patterns in this content. Content analytics can be used with content from content management systems or used in conjunction with other unstructured data from any other corporate system or outside sources. This information can be combined with structured data to provide even greater insight.

Content Analytics Provides Actionable Insight

How does content analytics work? Consider this example...

How does content analytics work? Consider the following example: A marketing group at a telecommunications company is running a promotion for its phone/internet/cable TV bundle during the month of August. The promotion consists of a low monthly charge for first year of service for the TV and Internet piece of the plan and some premium stations for free for the first three months of service.

The marketing department launched the campaign and its structured data indicated that they were getting a good response—many customers were initiating the new plan. This structured information consisted of customer name, ID, the number of cable boxes, customer location, and revenue. In fact, the revenue seemed to be a bit higher than expected, but the marketing department just assumed that this was because people were purchasing on demand movies.

Four months later, however, the marketing team noticed that billed revenue for customers participating in the campaign had actually decreased and that customers were beginning to drop the service. The structured data did not explain why this was the case, but unstructured data did provide some clues. In reality, the company had billing problems that resulted in customers dropping the service. Figure 1 illustrates text from several call center records stored in the company's content management system.

Customer is annoyed. Says that she is being billed more than the promotion price for TV and Internet.
======

Customer responded to August 2008 promotion. He said that he is being billed the full price and not getting premium channels for free. He was thinking of dropping the service if company can't get act together.
======

Customer started service in August 2008. Says she is being billed for channels she wasn't supposed to be billed for. She is dropping service.

Figure 1

The underlined words provide the information the company needed in order to understand the "why." For example, the phrases August 2008 and the word promotion indicate that the reports mention the promotion. The terms billed more, full price, and free indicates a potential issue with billing. The terms TV and premium channels indicate that this might be the source of the billing problem. The words annoyed and dropping service provide insight into the customer sentiment, which in this case is negative.

The content analytics process uses algorithms to analyze the unstructured text, to extract information, and to transform that information into a structured data file...

The content analytics process uses algorithms to analyze the unstructured text, to extract information, and to transform that information into a structured data file such as the one in Table 1. The analyst can then drill down to learn more about each case, even accessing the unstructured text of the call records themselves.

Customer-ID	Event	Issue	Sentiment
125678	August Promotion	Billing	Negative
137642	August Promotion	Billing	Negative
1462789	August Promotion	Billing	Negative

Extracted Text
Table 1

If the company then analyzed this data, management can anticipate a potential problem before the company started losing customers.

Additionally, the company can also combine the data in Table 1 with other structured information in its data stores, for a deeper analysis of the unstructured information. For example, a merge of the extracted unstructured data with structured billing information (see Table 2 on the following page) uncovers the fact that the last two customers are "gold" customers based on phone bill alone, so it would be worthwhile for the company to make an extra effort to retain these customers.

Customer-ID	Name	Average Phone Bill	Plan Type	Retention (yrs)
125678	John Smith	$20	A	0.5
137642	Jane Thompson	$120	B	2.0
1462789	Susan Roe	$110	B	2.5

Existing Structured Data
Table 2

The overall content analysis process might include a set of processes as detailed in Illustration 1.

The overall content analysis process might include a set of processes as detailed in Illustration 1.

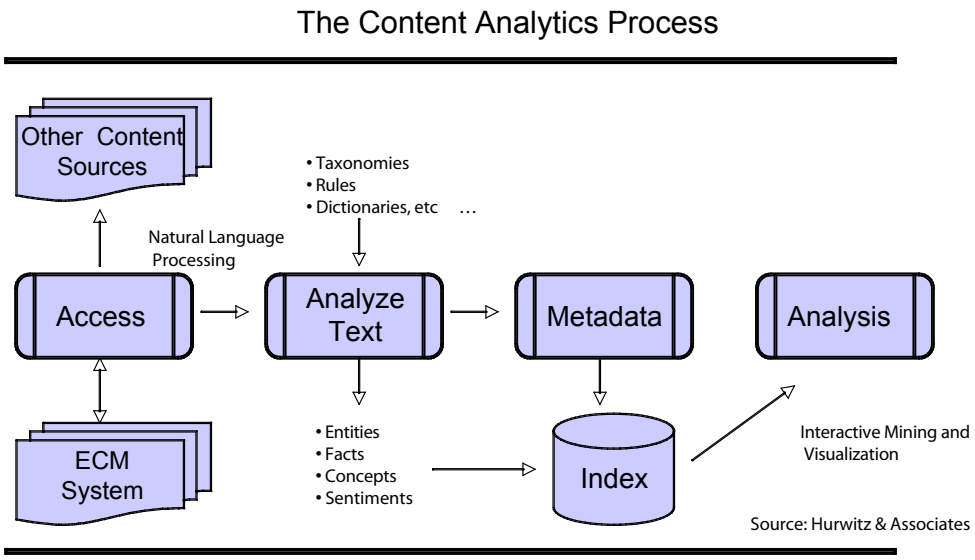

Illustration 1

Content Analytics Improves Business Processes

Content analytics can also be used to enrich business processes. In this case, as digital assets are added to a content repository, information is extracted, merged with other enterprise information and fed into the workflow process. For instance, in the call center example above, it would be possible to pull information out of the call center notes and link that information with the data about the customers, their bills, and information found in other systems. At this stage, it is logical to provide this information to a customer care agent. Now the customer care agent will understand that there is an urgency to contact the gold customers directly. If the agent is successful, the information is fed back into the system. This content lifecycle becomes part of the context of the process and helps define how proactive decisions are made.

Similarly, take the example of a claims process workflow. Unstructured information, which flows in and out of an insurance claim, along with metadata from the process workflow could automatically be analyzed. The results are then merged into the case file within the ECM system to provide a deeper understanding for the claims adjuster. Because the content analytics software integrates with the content repository, it can directly ingest documents and its metadata for streamlined analysis. Here again, the process changes the content and the content defines the context in which decisions are made. This interplay between content, content management and business process management can result in a number of benefits including higher productivity and efficiency, better decisions, and higher levels of innovation.

This interplay between content, content management and business process management can result in a number of benefits including higher productivity and efficiency, better decisions, and higher levels of innovation.

Content Analytics in the Real World

Content analytics is a powerful tool that can be used to help companies gain insight into the massive amounts of unstructured data found in ECM systems. In this next section, we spotlight three specific use cases to illustrate how the technology can be used in the insurance, state and local government, and in financial services industries. These use cases demonstrate how content analytics can help reduce costs, improve efficiency and productivity, and improve customer retention.

Case 1- Fighting insurance soft fraud using content analytics and predictive analysis

Fraud is a multi-billion dollar problem in the insurance industry impacting premium costs for consumers and impacting insurance industry profitability. Estimates by insurance industry research groups indicate that as much as 25% of all claims contain some degree of fraud. For example, the Insurance Information Institute estimates that in 2004 property/casualty insurance fraud alone costs insurers $30 billion or $200-300 per year in extra premiums. The National Insurance Crime Bureau estimates that fraudulent workers compensation costs the industry billions of dollars per year. The insurance industry has typically used specially trained investigation units to look for organized fraud groups. But, despite their best efforts, there are still enormous amounts of fraudulent claims that go undetected. Some insurance companies have begun to use highly sophisticated analytical and predictive modeling tools on insurance claim data to provide an automated approach to identifying claims with the greatest potential for fraud. These techniques include analysis of both structured and, more recently, unstructured information to increase an insurance company's ability to separate legitimate from potentially fraudulent claims. As a result, these companies are in a better position to catch potential problems early in the claim process and reduce over payments caused by fraudulent activity.

Some insurance companies have begun to use highly sophisticated analytical and predictive modeling tools on insurance claim data to provide an automated approach to identifying claims with the greatest potential for fraud.

There are two types of insurance fraud: hard fraud and soft fraud. Hard fraud is deliberate, criminal activity such as auto insurance crash rings where people deliberately crash their vehicles into others to fake medical injury and then collect insurance dollars and potentially sue for a large personal injury claim. Soft fraud is harder to detect than hard fraud because it often involves a relatively honest person who, for example, exaggerates a homeowner's claim to collect more in insurance dollars or claims disability to extend time off from work. Why does this soft fraud occur? Typically, people justify soft fraud because they believe that insurance premiums are too high or they pad a claim to get back their deductible. Historically, insurance companies often felt it was not worth the cost of pursuing a soft fraudulent claim because these kinds of claims are difficult and costly to prove.

Insurance claim data stored in structured databases tells only part of the story

Special investigation units have made progress dealing with hard fraud by leveraging insurance company databases filled with information from prior claims.

Special investigation units have made progress dealing with hard fraud by leveraging insurance company databases filled with information from prior claims. These data might include vehicle identification numbers of stolen cars, social security numbers of previous claimants including those of deceased persons, and names of licensed physicians, that insurers can access as part of the claims investigation process. These databases of structured information can be great sources of information. The investigation units use technology to set up rules to help look for signals or triggers that indicate organized fraud.

Companies have also begun to deploy more sophisticated analytical techniques, such as data mining, to help identify fraud. These analytical techniques are being used in conjunction with what insurance claims experts refer to as "gentle techniques" such as telling the claimant that the phone conversation is being recorded or calling the claimant for additional information in order to help to deter soft fraud. But, there is a lot of important information about prior and current claims that is not contained in the structured databases of claim data. While applying predictive modeling to structured information to analyze potential fraud has been very helpful, the reality is that much of the important content regarding a claim is found in the unstructured text contained in claim notes. Finding some of this information by manually reading through the notes is enormously time consuming and it is almost impossible to search through large amounts of data to identify patterns that might indicate fraud.

Data Mining

Data mining is the process of exploration and analysis of large amounts of data in order to discover meaningful patterns and rules. This is often achieved by automatic and semi-automatic means. Typically, historical information with known outcomes (e.g. fraud, not fraud) would be used to drive the system. Other predictive models use techniques such as logistic regression to assign a score to a claim that is possibly fraudulent. Any time information in the claim changes, the score may change, as well.

How content analytics can help

Some insurance companies are beginning to use content analytics to gain additional insight about potentially fraudulent claims from unstructured information. This type of information might include notes and comments related to a claim such as eye - witness accounts, the claimant's account of an incident, physician's notes, and law enforcement notes.

For example, consider the case of a person filing for workers compensation claiming a non-work injury as work-related. The typical structured data captured for this kind of claim would include name, address, gender, age, date of injury, date of employment, physical location of injury, part of body injured, type of injury (e.g. slip and fall, repetitive motion injury), prior injuries (code), and prior medical conditions (code). This information might be stored in the company's database. This is all useful information, but the data quality might be poor and there might not be enough detail to be able to accurately predict fraud.

The valuable data buried in claim notes or the description by the claimant on the claim form stored in the insurance company's ECM system, however, might contain the key to determine whether or not the claim is suspicious. For example, the adjustor's notes might contain employers' notes that indicate that the claimant was on probation for misconduct or about to be laid off due to consistently poor performance. Or, the claimant was inconsistent in the activities and description leading up to the injury. This type of information would be very useful in assessing the validity of a claim. The issue, however, is that it is difficult for a person to manually cull through these dense notes to get at the desired information. Workers compensation claims may involve hundreds of claims handler notes through the life of the claim. Additionally, claim adjustors typically handle many claims simultaneously and may inherit claims, all further making it difficult to determine whether a case is fraudulent or not.

The content management system provides incredibly valuable data since it stores the history of the claims that the company has received. These might be classified according to type – neck, back, wrist and so on. Additionally, the company has tagged whether the claim is fraudulent or legitimate. Information from these claims can be extracted using text analytics technology. This information might include state of mind of the clamant, work history, medical history, personal history and so on. The data is then used to build a model and

The content management system provides incredibly valuable data since it stores the history of the claims that the company has received.

train that model based on known outcomes; i.e. fraudulent claim vs. legitimate claim. This model, which combines the structured and previously unstructured information, can then be used in the field to score new claims that come into the insurance company. If the claim is suspicious, the Special Investigation Unit works hand in hand with the claims handler to determine whether or not the claim is legitimate or not. Through time, as new information becomes available, the model can be updated and improved.

Content analytics can improve the efficiency with which fraudulent claims are identified because unstructured information is now being effectively analyzed.

Insurance companies that are making use of content from claims report that the results have, in many instances, exceeded expectations. Content analytics can improve the efficiency with which fraudulent claims are identified because unstructured information is now being effectively analyzed. Claims adjusters also spend less time reading through all of the claims and so they can be more productive. The net result is that by using content analytics, insurance companies can save time and cost associated with identifying fraudulent claims as well as potentially decreasing payouts for fraudulent claims.

Beyond fraud

There are many other applications beyond fraud that insurance companies can use content analytics. For example, insurance companies are faced with managing the high cost of medical care associated with legitimate workers compensation claims that ultimately become catastrophic claims. In fact, some insurance companies estimate that 75% of the dollars paid for legitimate workers compensation come from 5% of the cases. There are several reasons for this. First, there is no limit on the amount that can be paid on workers compensation claims. Second, there is no statute of limitations on the claim. This means that insurance companies are still paying out claims from past decades. It is in both the claimant's and the insurance company's best interest to keep the claimant as healthy as possible.

In the past, information that could help predict whether a patient might benefit from a certain medical procedure was difficult to determine because critical information was buried in physician's notes. The insurance company could access this information from claims forms and physician notes to determine the potential of success from various procedures. This enables the company to be make better decisions based on outcome and costs.

Case 2: Using Content Analytics to help improve foster care

The foster care system in the United States provides temporary safe housing for neglected or abused children. In 2005, over 800,000 children moved through the foster care system with 500,000 placed into alternative homes at any point in time according to the U.S. Department of Health and Human Services. However, there are additional complicated issues. For example, nearly 25,000 reach the age where they are no longer eligible for support. These older children might live with relatives, in group homes, or other institutions. While some of these children make the transition well, others are in need of a different level of support.

The individual states within the United States all utilize... [NCANDS]... as well as information contained in their own data warehouses to answer basic questions about foster care...

Determining whether these youth are getting the appropriate quality of care can be difficult for a number of reasons. First, case managers are often asked to juggle too many cases at one time. Recommendations from The Child Welfare League state that a social worker should handle between 17-20 cases at any time, but often the caseload is much higher. In many instances caseworkers are taking on double this caseload. Second, caseworkers generate a lot of documentation about these children that provides valuable information. And, often, case managers inherit cases from other workers. With such a heavy workload, it is often difficult, if not impossible, to read everything about a particular child. These factors can contribute to poor quality investigations and issues that fall through the cracks. If caseworkers are simply struggling to keep up with their caseloads, they do not have enough time to spend out in the field with the children. Needless to say, the quality of care that foster children receive may also be jeopardized.

There are several data sources that the states use to report on foster care. The Adoption and Foster Care Analysis and Reporting System (AFCARS) collects information on all children in foster care. The National Child Abuse and Neglect Data System (NCANDS) is a voluntary national data collection and analysis system that is also administered by the Children's Bureau, Administration on Children, Youth and Families. The individual states within the United States all utilize this information as well as information contained in their own data warehouses to answer basic questions about foster care such as the number of children entering the program, broken down by age of the child or placement type or the ratio of teenagers in homes vs. other living environments. This data is used in many different ways. It is used to conduct trend analysis

such as foster care placements for teens, or placements in psychiatric care. The data can also be used to track metrics such as the number of in-home visits conducted and case plans completed on time. While this information is very useful, it does not necessarily help the caseworker perform their jobs more effectively or avoid a crisis situation before it happens.

Some states have been able to implement content management systems that help caseworkers to create and manage their cases more effectively so they can spend more time dealing with clients. While a content management system can help caseworkers build and organize their files, it doesn't necessarily provide any insight into the information stored in these files. Content analytics in conjunction with content management systems can help state agencies improve their understanding of their cases and find patterns in the data that might elude a single caseworker. Content analytics can also be useful in helping new caseworkers by quickly providing insight into past trends and events. The result is that caseworkers spend more time where they are needed –with the children.

Some states have been able to implement content management systems that help caseworkers to create and manage their cases more effectively so they can spend more time dealing with clients.

Content analytics can provide valuable insight

Let's examine several specific situations where content analytics can be used to help improve foster care analysis. For example, consider the case of young adults aging out of the foster care system. New programs have been put in place to help youth transition to adult living including educational vouchers and transitional support and training services. Caseworkers still visit with these young adults on a monthly basis. These caseworkers might be noting that these youth are not taking advantage of the training support services available to them. Each caseworker notes this in his or her case file along with the reason why the youth is not utilizing the services. For example, one teen might say that she couldn't get a ride. Another might say that it was too icy out to walk to the facility. A third might say that he missed the bus. The underlying reason for this might be that they had no transportation to the facility. However, the case worker has no way of knowing that many other teens aren't getting to the programs because he or she doesn't have access to, or the time to read through all of the cases of the other case workers in the area. So, what might seem like an isolated incident is actually a significant issue that might impact the support services that this individual and others are receiving. Perhaps these people live in rural areas and timely and reliable transportation is not currently available to them. If the case notes were analyzed using content analytics, these types of

patterns would emerge. For example, in the case of the transportation issue, words such as ride and bus might be picked up from multiple case files and a trend would become apparent.

Content analytics ... can help surface issues that might not have been apparent.

Or, consider the case of a youngster placed in a group home. The caseworker notes that the child is frequently missing school due to illness and says she is shivering. Another child at the home might state that it is cold at the facility at night. The caseworker might speak to the person in charge who might say that the heat went out one night and that there isn't a problem. A new caseworker might come on board and be told that it is cold in the home at night. However, since she didn't read the files she doesn't know that this is a recurrent problem. The same types of issues might be occurring across the region and the state notices a 20% increase in foster children out ill from school. The reason might be that the high cost of heating oil has made the care providers lower the heat at night. Or, foster parents might be using the subsidies the government provides in another, potentially fraudulent manner. Perhaps, the increase in illness might have also caused a cascading effect such as an increase in doctor or hospital visits. Content analytics would have picked up the words cold, shivering and might have spotted the trend. Other caseworkers could have been alerted to look out for this issue, as well.

Finally, using content analytics, a person in the state social services department might develop categories such as "at risk youth" or "on target youth " using the information contained in the case records. For example, an at risk youth might be one who is not making use of the transitional services provided by the state. These categories could be explored further to better understand the characteristics of these two groups. Users could drill down into the text of the case files associates with these two groups to gather even more insight. Content could also be filtered to determine top issue categories facing foster parents as well as foster children. Some of these issues may be known, but others might not be.

Content analytics, then, can help surface issues that might not have been apparent. For example, in the first situation, if the issue is circulated as part of a report or discussed in a meeting, then other caseworkers would become aware of this potential problem when working with other youths and can be armed with alternative transportation arrangements. By improving the transportation issue, more people will show up for the program, the program will be better

utilized and have a better chance of success. In the second example, the net result could be a savings associated with additional problems that arise because of this issue. Alerting caseworkers to this and other issues also saves them time in their investigations. In all of the cases, improved analysis increases the likelihood of a better outcome for the children.

In all of the cases, improved analysis increases the likelihood of a better outcome for the children.

Case 3- Improving Customer Retention in Banking

Optimizing the customer experience and improving customer retention are dominant drivers for the financial services industry. Customer attrition rates for banks in the United States are estimated to be as much as 30 percent. Even prior to the 2008 crisis of confidence in the financial industry, banks were struggling to retain and grow their customer base. Banks realized a number of years ago that they needed to establish a strong customer relationship strategy in order to retain customers and differentiate themselves. Regardless of the customer interaction channel - online, face-to-face, call center, etc banks have realized that they need deep customer insight across all customer touch points in order to improve customer service levels. This makes sense since it costs at least five times as much to acquire a new customer than to retain an existing customer. It is no wonder that stopping customer attrition is a top priority with US banks.

Analytics and Customer Interaction Management

Financial Institutions have been early adopters of advanced Business Intelligence (BI) tools that help them in marketing and business analysis efforts. Many banks have built data warehouses and are using this data to help better understand customers and predict their behavior. There are numerous examples of how banks have used data mining against structured information to help improve marketing and customer relationships. These include:

- Profiling and segmenting customers to understand high value vs. low value customers. These groups could be offered marketing programs tailored to their potential needs. Some banks even provide relationship managers for their highest value customers.

- Analyzing buying behavior of credit card customers to offer them other products based on what they have bought.

- Predicting churn of credit card customers. For example, if a model picks up that a customer is suddenly paying off his credit card balance, this may be a signal that the person is considering closing out their account and this can trigger some proactive, personalized marketing offers.

Data mining using structured information has been quite successful. However, in some cases, this kind of analysis can tell us what happened but not necessarily why it happened. Including the analysis of unstructured information can provide better insight into customer behavior. For example, unstructured information from call center notes could provide insight as to why the customer is considering closing out the account. Additionally, while historical analysis of information is very important, it doesn't necessarily help to understand current customer concerns.

Data mining using structured information has been quite successful. However, in some cases, this kind of analysis can tell us what happened but not necessarily why it happened.

Unstructured information can help to understand the voice of the customer

At best, unstructured information is managed within the financial institution's ECM system(s). Often, however, it is stored in a myriad of servers and desktop systems. As a regulated industry, much of the content must be retained and managed according to various regulatory compliance requirements. For example, financial services institutions are required to store information offsite on non-writeable media that is indexed and easily retrievable. This requirement includes any information used to process a transaction. The information is stored in order to be available for review and to better protect customers. Forward thinking banks are starting to make use of this information to gain a better understanding of their customers and their business. Some typical questions these banks are interested in answering from this unstructured information include:

- What are major areas of complaints by account holders and how are these changing over time?

- What is the level of satisfaction of customers with specific services?

- What are the most frequent issues that lead to customer churn?

- What are some key customer segments that provide higher potential up-sell opportunities?
- Are there early warnings of compliance related issues in customer contacts through email or call center?
- What is the expertise available in-house and how can this be tracked from the email exchanges?

Had the company proactively used its content analytics to indentify a pattern, it might have been able to save a valuable customer.

Information such as email, hold a wealth of information about whether compliance policies are being met, as well as information about market conditions, customer concerns and sentiment and potential fraud. But there are more subtle uses of information that banks can use to keep important customers happy. For example, take a customer who has many accounts with the bank. This very loyal customer has also purchased auto loans from the bank. The customer has become reliant on the bank's advanced online services. When the bank suddenly changes to a new online bill payment system, it upsets the customer since so many processes have changed. When the customer calls the call center to complain nothing is done to satisfy his problem. In frustration, the customer moves to a bank that has an online system that is more like the one he had used before. Had the company proactively used its content analytics to indentify a pattern, it might have been able to save a valuable customer.

Banks are using content analytics to cull through this information. The technology can pick up the fact that there was activity around "online accounts." It can pick up on phrases such as "don't like" "not happy with" "change" across call center notes and emails and can use a customer sentiment algorithm to determine that the customer was not happy about the change. The user sets up an alert if a negative sentiment around online banking exceeds a certain threshold. Additionally, this data that was extracted from call center notes or emails can be merged with structured data about the customer to determine the characteristics of the customers that are complaining about this particular feature. If there are enough complaints and the value of the customer is high, then the cost to make this change for this set of customers might be worth it. This information is then relayed back to the online product team and a change could have been made and a better customer relationship maintained.

Customer sentiment concerning service is just one example of how unstructured information is being used by banks. Customers can call to ask why rates have increased on credit cards, or why they are being charged such high wire transfer fees. Again, based on the customer profile, the bank may decide to cut charges or eliminate high fees for certain customers because they are profitable. This would increase customer satisfaction and reduce churn. And aside from customer concerns and complaints, product teams are deriving very useful information about the types of products and services customers are looking for by mining the email inquiries that customers send to the bank via its online system or from another source. This information is being used to help develop new products and services that might attract new customers or help retain existing customers, improving customer retention. Users are also filtering content to look for emails that mention a competitor, which will provide additional insight to the product and marketing teams.

Content analytics can help improve operational processes including customer care because it provides agents with a more complete view of the customer and with more accurate information about them.

Content analytics can help improve customer retention in numerous ways. It can help identify and address causes of customer dissatisfaction in a timely manner. It can help improve service image by detecting common problems early and enabling a bank to proactively address them. Content analytics can help improve operational processes including customer care because it provides agents with a more complete view of the customer and with more accurate information about them.

IBM Solutions

IBM offers a number of products to help companies gain better insight from their unstructured information. These include IBM ECM repositories that help store and manage content as well as IBM Content Analyzer, which enables customers to analyze unstructured information. These products have also been integrated to provide customers with an efficient way to both store and analyze content. The goal is to enable customers to more efficiently and effectively store, classify and analyze the unstructured information sources within the organization to provide valuable insight that can help reduce costs, improve productivity, increase sales and improve customer retention.

IBM FileNet Content Manager

IBM FileNet P8 is an integrated platform that provides interoperability to a wide range of database, operating system, storage, security and Web server environments.

IBM FileNet P8 is an integrated platform that provides interoperability to a wide range of database, operating system, storage, security and Web server environments. The FileNet P8 platform serves as the core content management, security management and storage management engine for the IBM FileNet P8 family of products. It provides a content lifecycle and document management capabilities for digital content. FileNet P8 combines document management with workflow and process capabilities to automate and drive content-related tasks and activities and help address compliance requirements.

The FileNet P8 platform includes three core components:

- **The Content Engine** provides software services for managing business content such as documents, emails, contracts, claims, etc. It includes an active content capability to automate business tasks. Content Engine uses the latest J2EE technology standards and is deployed inside of an application server that spans a Java Virtual Machine.

- **Process Engine** enables users to create and modify and manage automatic business processes. So the process engine might be responsible, for example, to release a claim to an investigator once all the parts of the claim are in order.

- **The Application Engine** is the presentation layer for both content and processes. It also handles authentication so that only certain users can access the content.

The IBM FileNet system provides the following features:

- **Document lifecycle management capabilities:** FileNet P8 provides versioning and parent-child document management capabilities. It also provides approval workflows and integrated publishing support. FileNet P8 manages and controls the content lifecycle in order to help ensure compliance in content-related tasks such as publishing, expiration and retention. It also maintains a complete audit trails to meet regulatory mandates.

- **Active content:** FileNet P8 also provides the capability to automatically move content and content related business tasks through a business process without requiring human initiation.
- **Transformation and rendition services:** FileNet P8 provides automated content publishing into multiple universal formats such as HTML and PDF.
- **Unified repository and metadata model:** FileNet P8 provides a single, unified content repository and metadata model regardless of the type of digital content being managed. It utilizes a common metadata model to enable streamlined search.

IBM Content Analyzer, formerly IBM OmniFind Analytics Edition, uses linguistic understanding and trend analysis to allow users to search, mine and analyze the combined information from their unstructured content and structured data.

IBM Content Analyzer

IBM Content Analyzer, formerly IBM OmniFind Analytics Edition, uses linguistic understanding and trend analysis to allow users to search, mine and analyze the combined information from their unstructured content and structured data. Content Analyzer consists of two pieces: a backend linguistics component and a visualization and analysis text mining UI.

- **A backend linguistics component:** On the back end, IBM provides a number of ways for unstructured information to be ingested and then indexed. Unstructured information can be ingested from any content source. Content Analyzer has a direct integration with FileNet P8, which means Content Analyzer understands FileNet formats and that the integrity of any information from the FileNet system is maintained as it moves into Content Analyzer. Once the unstructured information is ingested into Content Analyzer, linguistic analysis is performed. Content Analyzer provides natural language algorithms to perform tagging and named entity extraction. IBM also provides the software to create a synonym dictionary and a classification dictionary. The extracted information is then stored as metadata and is indexed to disk. For example, if a user created a category called negative sentiment this would be indexed along with the relevant text from the unstructured information that could be classified as negative sentiment.

  Additionally, if FileNet P8 is part of the solution, then categories created in Content Analyzer can be written back to FileNet to further improve content metadata.

- **A front end visualization component.** IBM Content Analyzer provides a standalone J2EE application called Text Miner that can help analyze the indexed source dataset. Text Miner provides facilities for real-time statistical analysis on the index for a source dataset. It allows the users to analyze the processed data by organizing the data into categories, applying search conditions, and further drilling down to analyze data from different perspectives such as over a period of time, changes in data, changes in topic, or as a matrix of two categories. The analysis leads to identifying leading trends, early detection of problems, or understanding the voice of customers. For example, if the user has created an entity called brand and a category called problem, he or she could examine the kinds of problems associated with a specific brand. The user could also drill down to the actual document level and examine the text associated with the problem for that brand.

Hurwitz & Associates expects to see increasing numbers of companies across many industries using content analytics to derive value from their unstructured information.

Content Analyzer uses the UIMA standard. UIMA, Unstructured Information Management Architecture, is an open framework and Software Developer's Kit for developing applications that utilize unstructured information. The UIMA standard enables developers to build UIMA compliant components, called annotators, that contain the logic to analyze this unstructured data. For example, an annotator can be the UIMA compliant logic needed to extract a person's name from a text document. These common sets of interfaces enable text analytics components to be integrated into a broader set of solutions.

Conclusion

Enterprise content management systems contain a wealth of valuable unstructured information in the form of call center notes, emails, case files and so on. This content represents a large portion of the information that companies manage. However, until recently has been totally underutilized because it has taken too long to manually cull through all of this information and derive effective insight. Content analytics used in conjunction with an enterprise content management system and other data stores can help organizations make better decisions, drive down costs, improve customer satisfaction, and make workers more productive. This technology is rapidly moving out of the early adopter phase and becoming mainstream. Hurwitz & Associates expects to see increasing numbers of companies across many industries using content analytics to derive value from their unstructured information. The value proposition is simply too compelling to ignore.

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.