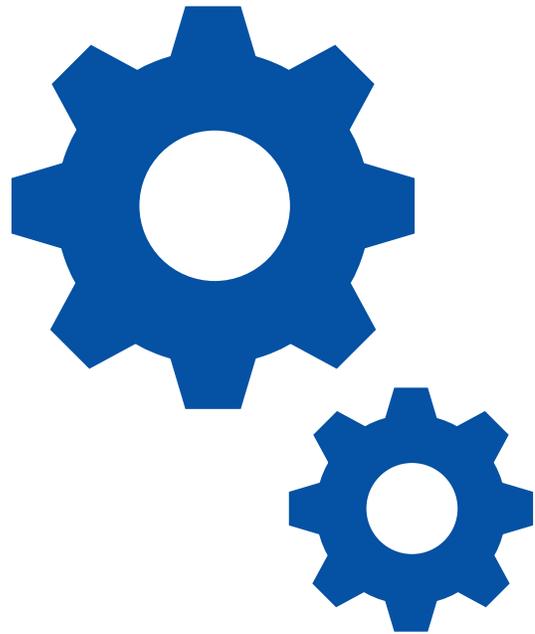


# Unstructured Data Analytics

Unlock the treasures of your unstructured content



## Executive Summary

An amazing level of detailed business intelligence can be found in an organization's unstructured content. This comprises about 80 percent of enterprise data according to many estimates. For decades, unstructured content has been recognized as a kind of gold mine too complex to exploit, but recent advances in analytics have begun to unlock its treasures. Deploying this knowledge can confer considerable competitive advantage to those companies that can do so skillfully. Despite obvious spurs to use unstructured content analytics, the price of the technology has limited the number of companies able to deploy it. Recently, however, some providers of the technology have begun to make it available through a Software as a Service business model. Using the technology, and the computing power behind it, services can drastically lower the price of admission to this powerful business solution.

**Unstructured content** is the gold mine that has been too complex to exploit. Until now.

## Business Advantages of Unstructured Data Analytics

Imagine a technology that produces rich and complex insight from the unstructured data you already have.

Extracting knowledge out of the information residing in databases is easy. The information has structure. You define your categories up front, so it is simple to evaluate the status of any given category, track trends in any category, and compare how any category relates to others across time, geography, or any other categories you define. But what about extracting knowledge out of the flood of unstructured content pouring through the Internet? With content such as text-and especially with media like video and audio-extracting knowledge can be much more difficult. No wonder Business Intelligence systems pay off so well when savvy users load them with volumes of structured data that the systems can analyze practically in real time. Text is the classic example of unstructured data. Words flow, uncategorized, onto the Internet in an endless river from billions of fingertips. How do you make sense of even a trickle of it? You could have hundreds if not thousands of friends on Facebook or followers on Twitter, but even if you could examine every post in which they mention your business, could you quantitatively understand what they, in aggregate, were saying about you? Not if you were reading with eyes, you couldn't. You couldn't keep up with it, even if you were using computers to count keywords. [what are the speeds at which the average human and average computer read text?] So imagine a technology that can scan through thousands of pages of text in a day, and produce structured information out of the unstructured content, reporting on how often people wrote nice, or negative, things about your business. Imagine a technology that could even report on which words people were becoming more or less popular to convey positive or negative emotions. With such tech-

nology, you could understand in a much richer and more complex way-compared to, say, standard marketing surveys-why people like or don't like what you do. And with knowledge like that, you could tailor your business closer to your customer's expectations, market your business more selectively to those who would most value it, and understand better what people are talking about within your organization.

Marketing companies could find tremendous value in this technology, unstructured data analytics, as the path to business success keeps changing. In the not-too-distant past, exposure was the golden metric of marketing. Organizations with something to sell would pay more for more exposure. How many eyeballs can you give me?" was the question that cut to the heart of a marketing campaign. But high tech has changed all that.

### **Engagement is the new exposure, and to measure engagement, you need unstructured data analytics.**

Engagement metrics suss out the reactions people have to your marketing message. They tell you the degree to which your message is being received in the way you intended. Text analytics can examine online content and identify trends, sentiments, and connections in a way that traditional marketing surveys never could. Retail and hospitality companies can use text analytics to understand in a deeper way than previously possible what people like and don't like about their businesses and the way they operate them. The vehicle that carries this understanding is social media. Social media is not just another channel for delivering your message, it also is a channel for you to hear what the outside world is saying about you. Text analytics can mine the value from feedback such as: "We loved the décor of our hotel suite, but there was an odd smell in the bathroom," which might have been scored as a neutral statement by less-advanced technology.

## **Uncover Value Across The Enterprise Landscape**

Understanding to a finer level of detail what your customers like and don't like, and why they have those opinions, enables you to respond with more agility to provide them with more of what they like, less of what they don't like, and find more customers who are similar to your best customers.

Text analytics can also give you new eyes to what is happening within your organization. Analysis of internal communications such as email and instant messages can reveal issues of concern that might have otherwise gone undetected.

Text analytics provides a fine comb for fraud detection and better customer profiling.

Currently, most automated antifraud technology can access only structured content, which represents only about 20 percent of enterprise data. While auditing structured data such as transactional databases and general ledgers is certainly a necessity, text analytics can raise red flags regarding the who, what, when, and why of business risks and fraud.

Text analytics enable businesses to monitor who is talking to whom, about what, during what period of time, and how they feel about what they are talking about. What trends occur just before the end of a quarter or fiscal year? What might a spike in communications between executive management and the stock-options administrator in HR signify?

## **Four Features Of Effective Text Analysis**

### **Centralized Scalable Repository**



Upload to a central repository dedicated to your content, simplifying analysis. By taking advantage of Amazon Web Services (AWS), storage is scalable and immediately responsive to peaks in demand.



### Automatic Uploads from Multiple Sources

A solution that provides automatic uploading of content from multiple sources (email, social media, text messages, etc.) to the repository ensures your content analyses reflect the most current information available.



### Automatic OCR Scanning

Optical Character Recognition (OCR) reads scanned documents or photographs enabling the analytics tool to search for particular words and phrases even if that information is only available visually.



### Analytics Reporting Dashboards

Develop specialized tools especially for your content and answer questions you actually have. Tailor analytics to benefit brand reputation, grow competitive intelligence, and gain customer service insight.

While unstructured data analytics has been around for about a decade, the price to acquire the technology, and its learning curve, have been steep. For example, the user license for an unstructured data analytics solution from one of the best-known names in computing can run to many tens of thousands of dollars. Another problem is where to start. Unstructured data can often represent as much as 80 percent of an organization's data. Estimates indicate that the volume of unstructured data is growing at more than 90 percent annually. Document silos typically contain multiple copies of documents, 19 is an often-cited number of copies per document. The sheer volume of information to wade through is daunting. No wonder 90 percent of documents in enterprise storage are never viewed again once they have been in storage for 90 days, or that knowledge workers can spend 50 percent of their time searching for the right documents.

If there were a way to affordably access advanced analytics for unstructured content, an enterprise could emulate the Oakland A's of the late 1990s and early 2000s, when General Manager Billy Beane leveraged advanced baseball analytics, popularly known as sabermetrics, to recognize and then acquire undervalued players. As told in Michael Lewis's bestseller, *Moneyball*, the A's under Beane's sabermetric-guided management became the most successful team in Major League Baseball in terms of games won per dollar spent on talent. The key is finding a way to make unstructured data analytics affordable and user-friendly.

One way that some businesses affordably use technology that typically carries a high price tag is to access the technology as a service, rather than as a licensed product, an arrangement known as Software as a Service, or SaaS. Using SaaS enables an organization to begin deriving benefits from the technology without having to first clear a high financial hurdle. Taking the concept one step further upstream, an SaaS provider could keep its costs low, and still have the scalability to satisfy fluctuating demand, if its costs for computing power tracked the amount of computing power it needed currently, rather than the most it expected to need. In other words, raw computing power itself can be accessed as a service, obviating the need to develop and maintain storage farms.

A small number of companies now offer text analytics as a SaaS basis for specialized and general enterprise users. These are the four features you should look for in a text analytics solution.

# filelytics

[www.filelytics.com](http://www.filelytics.com)

206.826.2216

[experience@interstack.com](mailto:experience@interstack.com)