AlchemyAPI™
An IBM Company

The Unstructured Data Problem

Why Organizations Should Not Ignore Unstructured Data

Written By Janet Wagner for AlchemyAPI, An IBM Company

@AlchemyAPI

AlchemyAPI™
An IBM Company

“We Are Drowning in Information But Starved For Knowledge.”

Author John Naisbitt first penned this well known quote back in 1982 in his book “Megatrends.” These words are as true today as when they were first written. The world is drowning in data, most of which has been generated only in the last few years. Many organizations are struggling to keep up with the vast streams of data generated from social media, websites, emails, smartphones and other internet-connected devices.

Much of the world’s data is unstructured making it difficult for companies to discover, collect and extract valuable business insights from it. While some companies understand the value of unstructured data, others choose to ignore unstructured data completely or simply do not know how to effectively use it.

“Unstructured data is the life blood of all businesses today, regardless of size. We’re all swimming in emails, instant messages, text messages, social media and news,” said Elliot Turner, Founder and CEO at AlchemyAPI, an IBM Company. “It’s how we communicate internally and externally and how we track our market and competitors. Companies are surrounded by this data, but they are not leveraging it holistically within automation frameworks.”

“Unstructured data is the life blood of all businesses today, regardless of size.”

Elliot Turner, Founder and CEO at AlchemyAPI

AlchemyAPI™
An IBM Company

What is Unstructured Data?

Unstructured data does not adhere to any formal pre-defined data model; it is data intended for human consumption and not designed for computers to process. Unstructured data is generated from many different sources including websites, blogs, social media, digital images, videos, mobile devices, wearable electronics and the list goes on and on.

There is a Vast Amount of Unstructured Data

According to the 2014 Digital Universe report by IDC/EMC, the amount of digital data in the world is approximately 4.4 zettabytes (4.4ZB). To put this in perspective, IDC calculated that this amount of data would fill a stack of iPad Air tablets (.29" tick, 128GB capacity) that reached 2/3 of the way to the moon (about 157,674 miles). According to IDC, 90% of this digital data is unstructured and can be found in many different types of file formats.

Types of Unstructured Data

- *Emails and text messages*
- *Tweets and social posts*
- *Chat and forum comments*
- *Support and CRM data*
- *PDFs and slide presentations*
- *Web site images*
- *News and blog articles*
- *Photo libraries*
- *....and more*

Unstructured Data is Growing Rapidly

The 2014 Digital Universe report also states that the digital universe is doubling in size every two years and the total amount of digital data in the world will increase from 4.4ZB to 44ZB by 2020. That's the equivalent of 6.6 stacks of iPad Air tablets or a total of 1,040,648.4 miles. In addition, the amount of unstructured data is growing 62% faster per year than structured data, according to IDC.

Thanks to the increasing use of internet-connected devices and the rise of the internet of things (IoT), the sources generating unstructured data are also rapidly expanding. There are now many more sources of data and file formats where unstructured data can be contained; text documents, audio, video, digital photos, emails, tweets, PDFs, sensors, smartphones, wearable electronics, GPS data and more.

AlchemyAPI™
An IBM Company

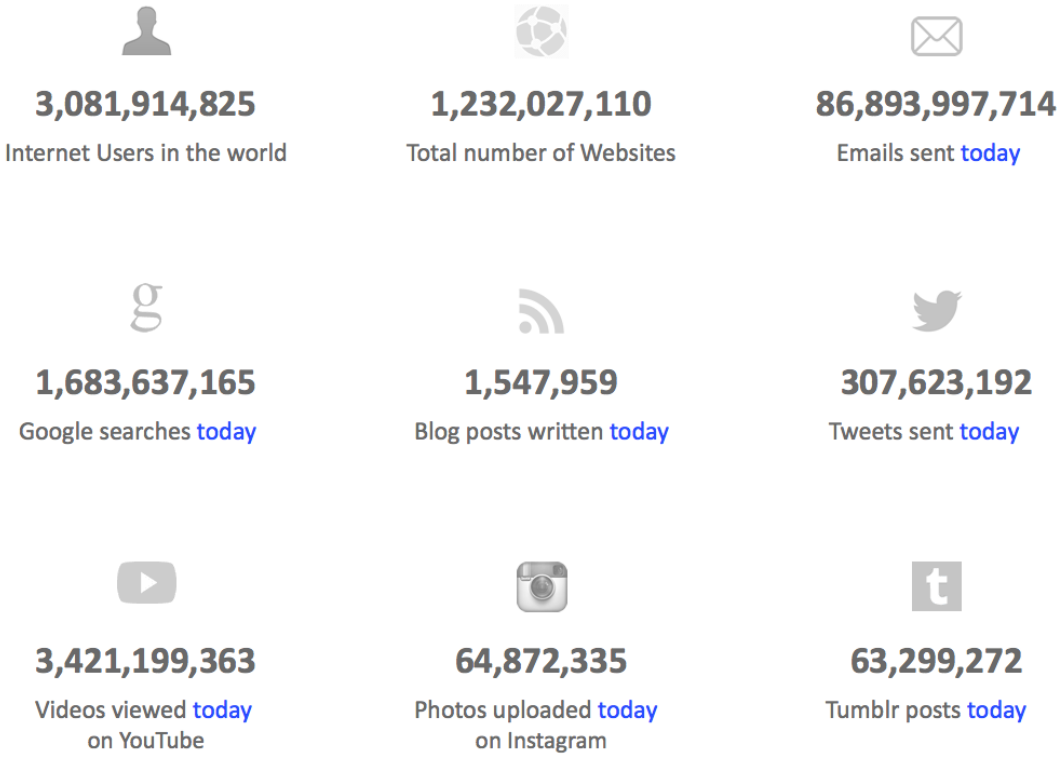
Unstructured Data Moves at Lightning Speed

Much of unstructured data is “data in motion,” a perpetual stream of data generated via the internet and internet-connected devices. The speed at which unstructured data is generated is almost inconceivable and the velocity of unstructured data is only increasing.

Media Type	Tweets	Instagram Photos	Emails
# Sent/Second	8,658	1,795	2,374,512

To understand the sheer magnitude and velocity of unstructured data, visit the Internet Live Stats website where you can see the amount of unstructured data generated every second, every day and every year. At the time of this publication, there are 8,658 tweets posted, 1,795 photos uploaded to Instagram and 2,374,512 emails sent every second.

Unstructured Data Does Not Stop



Internet Live Stats screenshot taken on 3/10/15 at 9:45am MST

AlchemyAPI™
An IBM Company

Unstructured Data Hides in Dark Places

There are many business leaders that believe unstructured data is generated strictly from external sources like websites, blogs, social media and the like. However, most organizations are their own source of unstructured data, much of it stored indefinitely, unclassified and never used. This unused data is referred to as "dark data" and is often hidden away in an organization's archives and stored in multiple data silos that contain vast repositories of unstructured data.

What is Dark Data?

Gartner defines dark data as "the information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes (for example, analytics, business relationships and direct monetizing)." Dark data is generated from an organization's emails, Word and Excel documents, PDFs, draft documents, presentations and more. An organization's dark data often grows because of governance and regulatory compliance, where data must be stored for a very long period of time.

Challenges

"The challenge is to draw insights and signals from the totality of an organization's unstructured data assets that can drive business strategy, as opposed to allowing individuals to make decisions based on emotions, whims or anecdotal evidence," said Elliot Turner, Founder and CEO at AlchemyAPI. "Ultimately, while all organizations have unstructured data, many are not leveraging it in this fashion. This is akin to keeping a bunch of stuff piled up in your garage. You don't really know what's there, and can't leverage those assets. An asset you cannot find or leverage is really not an asset at all. We want organizations to leverage their unstructured data to make better business decisions."

"The challenge is to draw insights and signals from the totality of an organization's unstructured data...An asset you cannot find or leverage is really not an asset at all."

Elliot Turner,
Founder and CEO
at AlchemyAPI

AlchemyAPI™
An IBM Company

Organizations Need to Effectively Manage Unstructured Data

Maintaining data is not the same as managing data, regardless of whether the data is structured or unstructured. Many organizations simply maintain their data ensuring that there is adequate storage space and infrastructure, but not utilizing the data to find key insights into their business or to gain a competitive advantage in the marketplace.

Most organizations store data far too long, holding on to data that is long past its shelf life and contains little, if any, value. Without some type of data management process in place unusable and unnecessary data will continue to accumulate taking up more storage space and making it even more difficult to find and process the data that is of value.

Data management is vital when it comes to storing sensitive information and confidential data. Without knowing how and where sensitive data is being stored, an organization is vulnerable to security breaches and costly information leaks. There were a lot of high profile data breaches in 2014 including Sony, Michaels, UPS, Home Depot and JP Morgan Chase.

Unstructured Data Lives in Many Locations

Many organizations these days have stockpiles of data (structured and unstructured) that grow larger every day and are stored in multiple locations such as on premises servers, cloud storage accounts and log files. In addition, many employees are now essentially “digital hoarders” storing data not only on company servers but on employee hard drives, USB and cloud storage (Box, Dropbox, Google Drive, etc). Unstructured data can also be found in layers of file formats; a text document attached to an email for example or text included in a PowerPoint presentation. Without a data management system in place, unstructured data continues to grow unchecked in multiple locations making it difficult to find and use.

AlchemyAPI™
An IBM Company

Unstructured Data Can Undermine Your Organization

Security breaches and information leaks are not the only ways that unstructured data can be used to undermine an organization. Public sources of unstructured data can be used for "competitive intelligence," techniques that allow an organization to gain key insights about the competition and gain a tactical advantage in the marketplace. There is a vast amount of public information available about companies, products, brands and customers; information that can provide insights as to how the market judges your company's strengths and weaknesses as well as your competitor's. Businesses need to take advantage of publicly available unstructured data before their competitors do.

"Leveraging unstructured data enables a business to become more nimble and more responsive to rapidly-changing markets, customer expectations and media landscapes. It enables an organization to go "beyond the focus group" and leverage the collective opinions of millions of customers, versus just a few," said Elliot Turner, Founder and CEO at AlchemyAPI. "This leads to better strategic decisions that more accurately reflect what the market wants, avoiding the bias that is evident in many traditional approaches such as focus groups or individual customer interviews."

Organizations Are Facing an Information Crisis

Gartner predicts that 33 percent of Fortune 100 organizations will face an information crisis by 2017. This upcoming crisis is largely due to the rise of big data, social networking and mobile devices as well as the rapid growth of structured and unstructured data. An organization's inability to correctly manage, value and trust their enterprise information will also add fuel to the fire.

"There is an overall lack of maturity when it comes to governing information as an enterprise asset. It is likely that a number of organizations, unable to organize themselves effectively for 2020, unwilling to focus on capabilities rather than tools, and not ready to revise their information strategy, will suffer the consequences," said Andrew White, research vice president at Gartner. "In a digital economy, information is becoming the competitive asset to drive business advantage, and it is the critical connection that links the value chain of organizations."

AlchemyAPI™
An IBM Company

Don't Wait. Start Looking for Smarter Solutions Now.

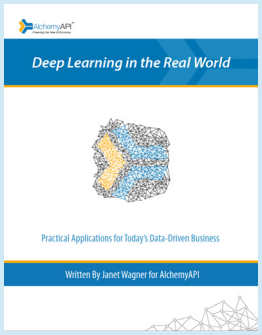
It is imperative that organizations start looking now for smarter solutions to the problems associated with unstructured data. There are many issues related to unstructured data that need to be immediately addressed such as storage, discovery, organization, tagging and deduplication. The most important issue related to unstructured data is finding and discovering key business insights as quickly as possible, preferably in real time, to gain a significant competitive advantage.

"There is value in both structured and unstructured data sources. Structured data can often tell you the "what" (e.g., sales went up or down last month), but not the "why" (which is often hidden within a multitude of customer communications, support requests, chats and emails)." said Elliot Turner, Founder and CEO at AlchemyAPI. "Unstructured data is key to understanding the "why" within a business. "Do people like our latest product release? Why is competitor X winning more deals this quarter? How are we being received in the press?"

Deep learning APIs and other advanced technologies can be used to address the problems associated with unstructured data and to discover valuable business insights.

Don't wait. Start using deep learning APIs and smart solutions today to harness the power of structured and unstructured data.

Learn How Businesses Integrate Deep Learning Into Data Analysis Strategies



There are many ways for you to utilize deep learning to solve your unstructured data challenges.

Get inspiration from 6 real world examples.

Download eBook

About AlchemyAPI, An IBM Company

AlchemyAPI's mission is to power smart applications that understand human language and vision by making breakthroughs in deep learning-based artificial intelligence available to everyone. AlchemyAPI is used by more than 40,000 developers across 36 countries and a wide variety of industries to process over 3 billion texts and images every month. For more information, visit our website at alchemyapi.com.

@AlchemyAPI