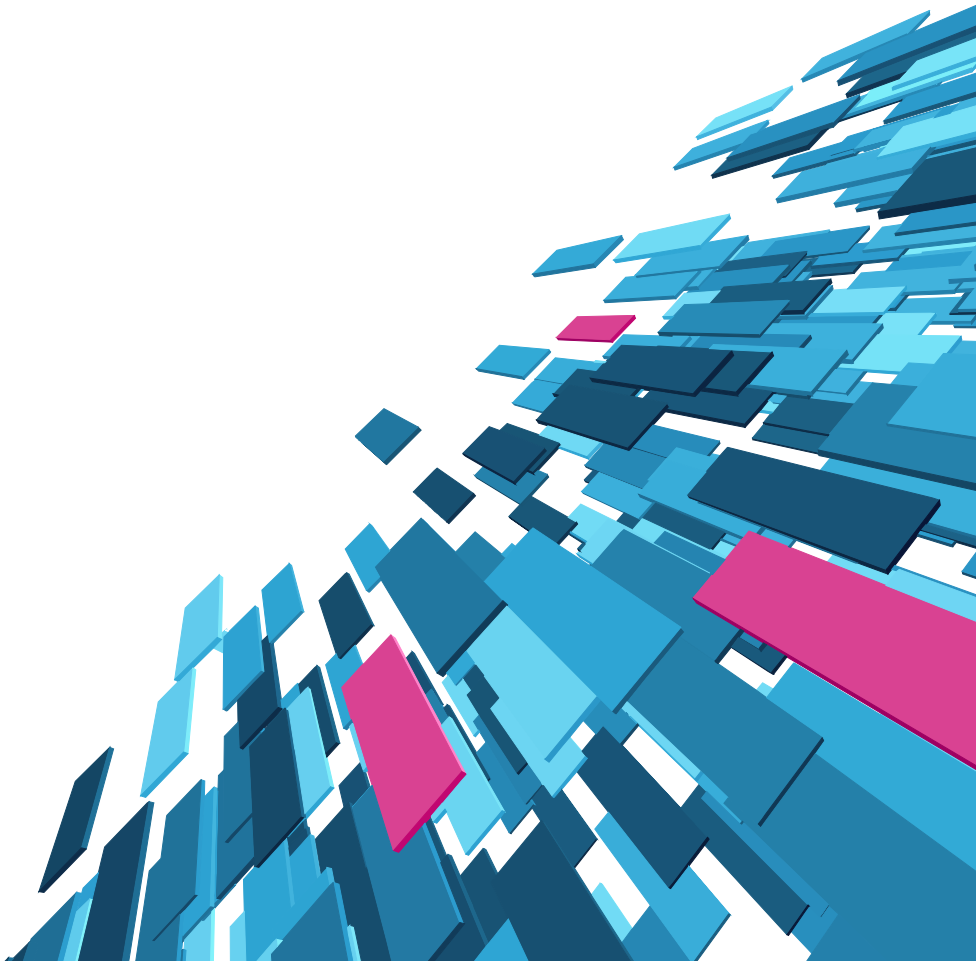
Smartlogic™ | THE CONTENT INTELLIGENCE COMPANY

White Paper

Why do I need content intelligence?

How Smartlogic Semaphore integrates with your enterprise to extract the value in your content



Why do I need content intelligence? How Smartlogic Semaphore integrates with your enterprise to extract the value in your content

> *There is huge business value locked away in content, because content contains the majority of the organization's intelligence. Organizations that unlock that value can outperform competitors in their market. Content augments the value that most organizations have already found in structured data - it is the untapped frontier of competitive advantage.*
>
> **Jeremy Bentley, Smartlogic CEO.**

Smartlogic Semaphore is the bridge between the human understanding hidden in content and computers.

Computers have a difficult time determining meaning in unstructured text – and there is too much of it for humans to cope with. Consequently, the vast majority of the enterprise's human intelligence – those man-years of experience that every organization has – is not being used to inform the business decisions that we are making today. Content intelligence is the solution to this problem.

For example:

- A constant stream of regulation creates new risks for companies and their executives - and the critical information needed to comply with regulations is locked away in content.
- Shared folders and enterprise content management systems (ECM) become increasingly chaotic under the huge load and associates spend more time looking for the information they need than they do using it.
- The human intelligence of the enterprise – embedded in the content - is hard to find, leading to wasted time and rework.
- The opportunity cost of not mining the enterprise knowledge in the content leaves organizations at risk from competitors that do.
- Even though the cost of raw storage is decreasing, the increased volume of content means that the cost of storage is continuously increasing. As it does, it eats into the IT budget leaving less funding for the new projects the business needs.

While tremendous strides have been made in the management and analysis of structured data over the last 30 years, content has languished in rigid hierarchical content management systems accessible only through primitive and brittle search capabilities. And because the content is "unstructured", which means that the formats, composition and even the words used to describe the concepts are not consistent, processing this volume of content and extracting its information by machine is extremely difficult.

In order to unlock this valuable information by machine (instead of having to read and codify it using human effort) we need to make the content 'self-describing'. Content Intelligence is the method and technology that makes this 'self-description' happen.

> Content Intelligence is the combination of semantic technology and information science that allows machines to model, interpret, describe, analyze and visualize the content of the enterprise in order to leverage the human intelligence locked in that content.

In other words, Content Intelligence is the process of automatically analyzing the unstructured content to identify what each item of content is, what it is about, how it relates to the business and what facts and relationships it contains. The process creates a rich layer of additional data describing the information in the content – this type of data is called metadata. Metadata is used by content management systems to drive governance processes and workflows, by enterprise search engines to improve "findability", and by business intelligence applications for analytics. Today, most metadata is manually applied to content – incompletely and inconsistently.

When you consider the massive volume increases in content and the speed at which the business demands that it be processed, it is clear that organizations can no longer rely on manual application of metadata. The ability to make the content "self-describing" is an economic imperative – not just a "nice to have".

Smartlogic Semaphore is the content intelligence platform that makes content self-describing – allowing organizations to:

1. Improve information governance
2. Improve findability
3. Create more revealing analytics
4. Create smarter case management and workflow

Improving information governance

Semaphore augments SharePoint and leading Enterprise Content Management Systems such as OpenText, Documentum, Filenet, Adobe AEM and Alfresco which rely on metadata to manage content. However, ECMSs rely on metadata provided by users, which is sparse and inconsistent. Absent rich metadata, the management capabilities are starved and cannot perform.

With Semaphore, existing content is automatically analyzed in background and new content is analyzed as part of the user workflow – all without burdening end users. Consistent and complete metadata is stored natively in the ECM application so that it can be used to apply governance policies.

Documents are analyzed to identify sensitive content such as Personal Identification Information (PII), Personal Health Information (PHI), trade secrets or trade restricted information. Such information can be automatically secured in restricted libraries. This reduces the risks of data or intellectual property loss.

November 10, 2014

US Postal Service data breach may compromise staff, customer details

Content that must be legally kept for a defined period can be automatically identified. This content is analyzed to determine the record type along with other facts that determine the date of disposition. Facts such as invoice date, agreement date or jurisdiction are extracted and stored as metadata. This information is often missing for paper documents that are digitized and may be erroneous for digital sources. Customer names can be extracted in case there is a legal hold on documents associated with a customer. These capabilities deliver defensible compliance and mitigate damage to reputation.

For content on shared folders or digitized from paper archives, Semaphore can be used to provide a granular audit of the content – not just how many files with each extension type – but also content type, sensitivity, relevant dates, customers or products mentioned. Storage and migration decisions such as what to archive, what to migrate, in which system, library and folder to organize the content and what security to apply become both intelligent and automated. Organizations can better organize the content they need and remove the content they do not.

Applying content intelligence to information governance enables organizations to reduce storage costs, mitigate litigation, reputation and regulatory risks and allows users to more quickly find the information, knowledge and expertise they need.

Improving findability

According to IDC knowledge workers spend a quarter of their time finding and analyzing information and only find the information they need 56% of the time.

Enterprise search engines are essential tools in making the knowledge of the enterprise findable and usable. But as the volume and variety of content grows the quality of search results deteriorates quickly leaving users frustrated.

There are several problems:

1) Search engines know nothing of the concepts and vocabulary important to a company and its industry. For example, in a healthcare context, search engines know nothing about Chronic Fatigue Syndrome. They do not know that it is referred to in a number of different ways (e.g. myalgic encephalomyelitis, ME, CFS) nor do they understand types, symptoms, diagnosis or treatments.
2) Users' queries are usually short, ambiguous and an approximation of their real need. This may work fine on the internet where companies invest in SEO to make their content findable – or where any of 5 million Tarte Tatin recipes will do – but it doesn't work in the enterprise when someone needs a specific customer proposal.
3) Authors of content often use different vocabulary than those searching for the content – meaning key results are missing from the first page where users look.
4) Search engines provide very little in the way of filtering, navigating, exploring and recommending content which users need to refine results, learn about topics and discover important relationships.

September 2012

IDC estimates cost of hidden information is $20,000 per year per knowledge worker

Content Intelligence augments enterprise search engines such as Google Search Appliance, Microsoft Fast and SharePoint, MarkLogic, Solr and IBM Watson. Users intuitively select concepts suggested by the content intelligence engine as they type in their query – removing any ambiguity. The relationships in the ontology enable users to explore and refine search results to get to the information they need quickly and to drive user journeys through the content with successful outcomes.

The metadata generated by Semaphore associates the content with the very same concepts that the user selected during their query – overcoming the variety of language used in content and the differences between author and searcher. Concepts are also associated with people making it easy for users to find people with expertise, knowledge or similar interests.

The user experience combines more accurate first-page results with powerful ways to explore and discover.

"Field service engineers at Applied Materials have reported a 50% reduction in time spent searching for information, reducing time for service calls and increasing customer satisfaction."

More revealing analytics

May 2011

Insurance company increased fraud detection rate to 88% by analyzing historical claims reports and identifying fraud predictors

In most organizations today business intelligence has focused on the 20% of information that is structured data – and this has provided many strategic and operational insights about what has happened or what is going on. While the structured data reveals what happened, the unstructured data goes on to reveal why it happened.

Content Intelligence provides the ability to augment the existing analytics infrastructure – expanding it into the unstructured content domain that provides deeper insight into why things are happening. The Big Data promise. Turning attention to harnessing the value of content-based information in new ways adds a further dimension of value to Content Intelligence. Gartner have stated that 80% of an organization's information is contained in content; today that information is beyond the data horizon from an analytics perspective. But Content Intelligence can bring that information into view - unifying unstructured with structured data under a consistent set of semantics, allowing the organization to discover new insights, develop competitive strategies and solve problems more effectively.

Ontologies are used to create an atlas of the content and data that exists within the enterprise and that is accessible beyond it. The ontology describes the data, the vocabulary it uses and how to map it. This is a far more dynamic and cost effective approach compared to traditional Extract-Transform-Load (ETL) methods.

Semaphore analyzes the data and content itself to normalize data and to extract facts from text.

Examples of facts extracted from text could be:

- Products that failed, the reason for failure, product age and circumstances drawn from warranty reports and trouble tickets.
- Products that customers discuss and the opinions stated drawn from call center reports or online social media.
- Equipment failures including symptoms, error codes, environment and diagnosis drawn from service reports.

- Medications taken including their dosage, frequency, side effects and outcomes from clinical trials or medical notes.
- Acquisition deals including deal size, date, target, acquirer, regulator and advisor from news articles.
- Well locations, drilling rights, term and fees paid drawn from legal contracts.

The mapped data and extracted facts are made available as structured data to analytics platforms such as Tableau so that users can explore and graph the data. The ontology gives users a powerful way to explore the relationships in the ontology to find new connections in the content.

Smarter Case Management and Workflow

January 2015

Ineffective manual processes drive financial losses, regulatory risk and lost opportunity

Some business processes rely on collections of documents which in turn contain many facts that determine how the business process should progress.

For example a bank on-boarding a new customer will collect hundreds of data points that must be cross referenced to ensure compliance with anti-money laundering regulation. Any anomalies must be investigated and resolved. Certain facts may drive additional processes – for example a director domiciled abroad may necessitate the capture of supplemental data. The extracted facts in turn create key reference data for the organization.

Similar processes exist for insurance claims processing, health care customer onboarding, mortgage applications and so on.

Documents may arrive by mail or fax (which are scanned and digitized using optical character recognition tools like Kofax and ABBYY), by email or via on-line forms. Semaphore can analyze each document to identify what kind of document it is, what data needs to be extracted from that type of document and go on to extract each of those facts from the unstructured text. Examples include the names of directors and their shareholdings, details of employment and references, events leading up to a claim, past medical history, the contract term and more.

The extracted information is fed into case management or workflow applications such as EMC xCP, SAP or Tibco.

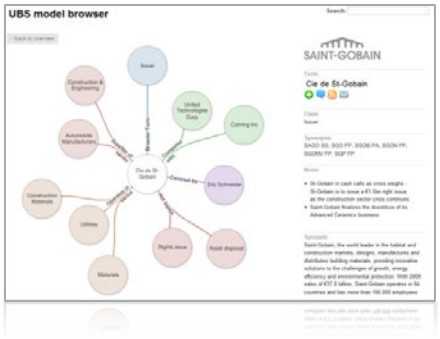
Processes become less labor intensive, faster and show fewer errors. The quality and consistency of reference data is improved and organizations can quickly work through existing cases to ensure complete and accurate data is on file.

"KPMG have seen a productivity improvement of 50% or better in over 40% of its onboarding cases"

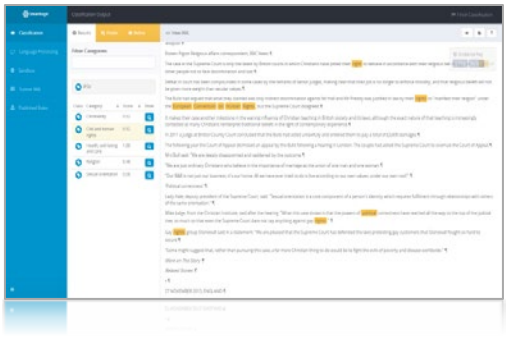
What is Semaphore?

Smartlogic's Semaphore is an enterprise-class Content Intelligence platform that augments existing enterprise applications and:

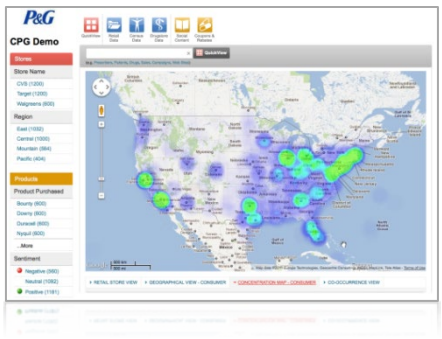
1. Efficiently captures what is important to an organization using an ontology (the concepts that are important to the enterprise and the relationships between them. For example a healthcare provider ontology would model conditions, treatments, services, facilities, physicians and so forth. A high tech manufacturer ontology would model products, customers, plants and equipment and processes. This enables all the content to be put into the context that users care about.
2. Automatically analyzes content of all kinds to identify what type of content it is, what it is about (based on the concepts in the ontology) and the facts and relationships contained in the content. Examples of facts in the healthcare provider's content might include a patient's medical conditions, and the type and volume of medications taken. The high tech manufacturer would be interested in the number of times a device failed and the most common symptoms before failure.
3. Provides a set of services and user interface components which can be easily added to enterprise applications to leverage the ontology and metadata to deliver the most compelling and intuitive way for users to find, discover, analyze and govern content and knowledge.



Build and manage semantic models



Automate or assist metadata enrichment

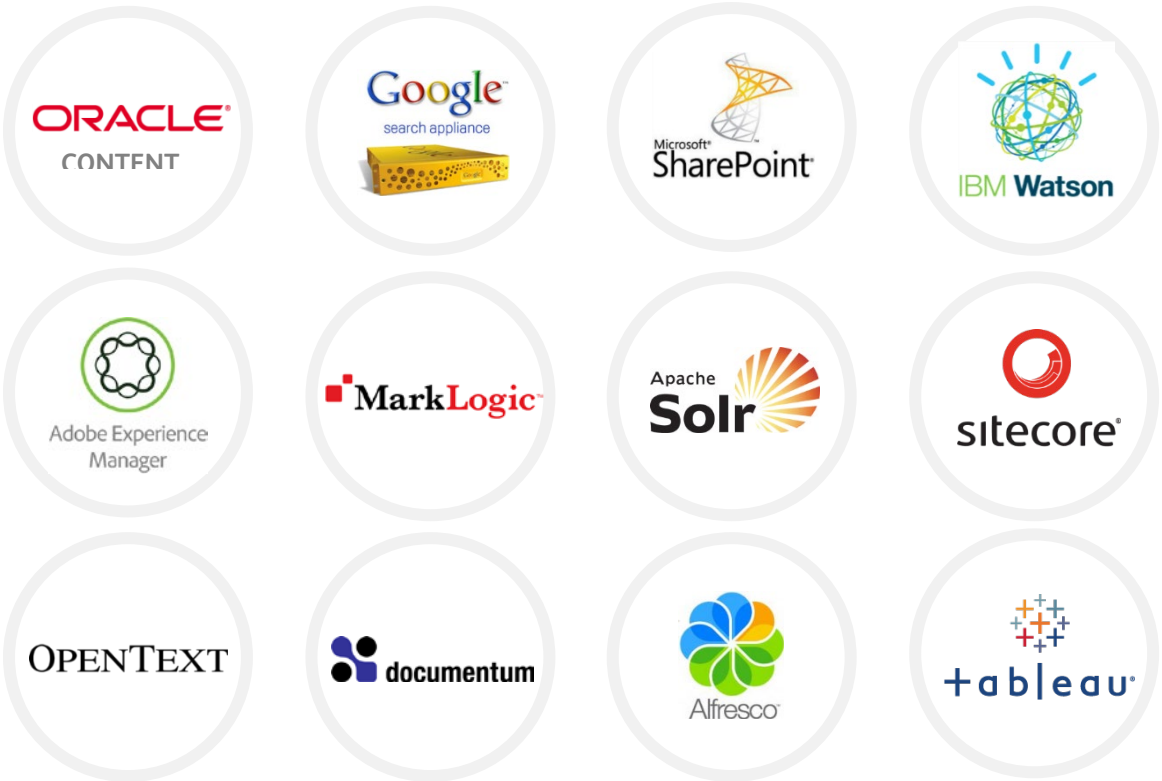


Drive a contextual user experience with metadata

Semaphore is deployed to solve enterprise-sized problems in some of the world's largest companies because:

- Semaphore is scalable:
  - Supports large and complex ontologies with millions of terms and synonyms and relationships
  - Processes large volume of content at speeds higher than any other platform and with 80-90% accuracy
  - Is deployed in environments with hundreds of thousands of users drawing on the services
- Semaphore supports the world's languages:
  - 20+ languages, including European languages, Chinese, Japanese and Korean, and Arabic, with new languages being added with every release
  - Sophisticated natural language processing techniques, including morphological stemming, to provide high quality classification, and fact and entity and sentiment extraction
- Semaphore integrates with a wide variety of enterprise systems including SharePoint 2007, 2010, 2013, Open Text, Documentum, EMC xCP, Tableau, MarkLogic, Oracle, Solr, Google Search Appliance, Adobe AEM, Sitecore, Alfresco, Watson and many others
  - Open APIs for easy integration with workflow engines and other enterprise applications
- Semaphore provides complete editorial control of both the model (ontology) and classification rules, to drive highly accurate results and support any classification strategy
- Semaphore is easy to use, with both the model and classification rules easily accessible to subject matter experts and others without advanced training in information science
- Semaphore has world-wide 7 by 24 hour support

Semaphore integrates with:



Smartlogic™ | THE CONTENT INTELLIGENCE COMPANY

SMARTLOGIC US
560 S. WINCHESTER BLVD, SUITE 500
SAN JOSE, CALIFORNIA, 95128
TEL: 408 213 9500
FAX: 408 572 5601

SMARTLOGIC UK
200 ALDERSGATE
LONDON, EC1A 4HD
TEL: +44 (0)203 176 4500
FAX: +44 (0)207 785 7014

WWW.SMARTLOGIC.COM
INFO@SMARTLOGIC.COM