



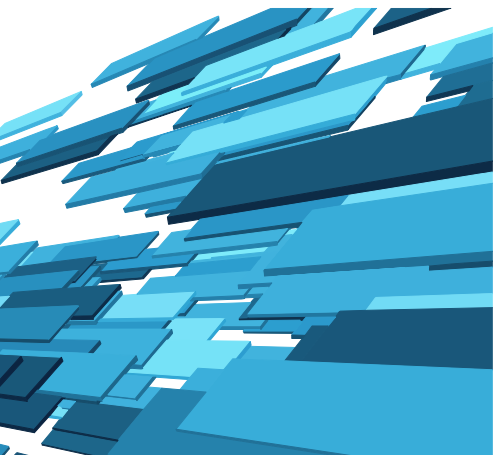
Smartlogic™

THE CONTENT
INTELLIGENCE COMPANY

In depth

Smartlogic Semaphore

The *what?* and *how?* of our Content Intelligence solution





Executive Summary

Enterprises no longer face an acute information access challenge. This is mainly because the information search market has matured in the last few years so access to information via search is common. However, enterprise search, by itself, is of limited value. Today's enterprises recognize the potential value of the information contained in unstructured information and they need tools to support their business and decision making processes.

To achieve the end goal of findability and determining governance such as content disposition, compliance, governance, records management, etc. search alone is not enough.

Smartlogic connects information with people, with processes and with context. We call this **Content Intelligence**. We address the difficult information management tasks: finding information efficiently; governing information effectively and putting information into context.

Smartlogic's Semaphore is the leading Content Intelligence software solution that augments traditional information management systems like enterprise search, content management and business workflow engines by capturing important topics, resources and people into a model (list, taxonomy or ontology) and then using this to classify content and enrich it with metadata to result in a complete enterprise information management experience. In collaboration with existing enterprise applications Semaphore provides a step change in search and content navigation for intranets and websites and ensures findability and appropriate content disposition, governance, data loss prevention, compliance and records management.

Semaphore enhanced systems allow:

- Re-use of valuable work, research and insights your organization has already completed.
- Improved access to information your colleagues have already researched.
- Rapid access to the total intelligence within your organization.
- Monetization of information and content into niche services for specialist audiences.
- Web-portal information findability delivering better, lower cost customer service.
- Greater insight by incorporating unstructured information into Business Intelligence applications.
- Business process efficiency through improved information linking.
- Tighter regulatory compliance via consistent classification.

"Enterprise information access technologies are maturing and now offer better indexing, querying, presentation and drill-down of results. However, the real value of information access technologies is in the upfront and ongoing efforts needed to establish effective taxonomies, to index, and to classify content of all kinds that must be accessed. By itself, the search function has limited value."

Tom Eid, VP Research - Gartner

Semaphore consists of four core modules:



- **Ontology Server & Editor** – allow multiple users to collaborate on the development and management of ontologies which capture the essential topics, resources and vocabulary for the business. Ontology Editor ensures the integrity and relevance of the models while dramatically reducing the effort to build them.
- **Semantic Enhancement Server** – a powerful, scalable and resilient web service delivering the ontology to any application that needs it.
- **Classification Server** - a rules-based semantic classification engine providing accurate metadata tagging of content in 26 languages. It can be used to identify entities using a range of vocabulary, to extract facts and information, or to determine the about-ness of content through the use of over 20 rule types. Rulebases are automatically generated from the ontologies managed within Ontology Editor — dramatically reducing set-up and management time.

Classification Server provides statistical output to identify relationships between topics, entities or nodes and supports over 280 file formats.

The optional Advanced Linguistics Pack provides text mining and entity extraction based on part-of speech tagging. It identifies over 30 different entity types (e.g. people, places, products, organizations, dates, facilities, URLs, etc.) which reduces the time to build ontologies and helps support accurate classification.

- **Search Application Framework** is a best practice, out-of-the-box Semantic search interface that can be deployed over search engines (e.g. Microsoft FAST, Lucene/Solr, Google Search Appliance, etc). It also provides published APIs and tag libraries for additional rapid user interface development.

The power of metadata

A content intelligence platform is a system that provides meaning and context to textual data. The ingredients that deliver context include:

- semantic models
- metadata
- navigation constructs

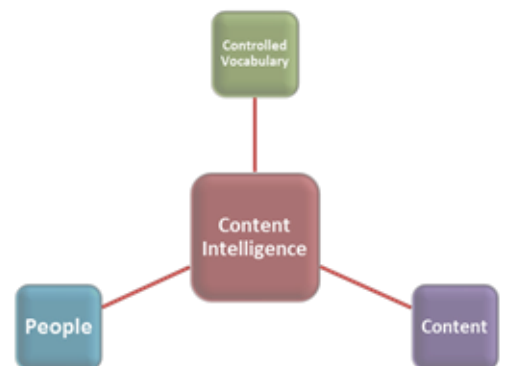
To apply metadata consistently, according to the enterprise standards, and in a way that can be demonstrated to be accurate and precise, the determination and application of metadata must be an automatic process.

This automatic process needs a reference model of the standards that drive the classification process. The model determines the metadata values for any particular piece of information.

These models are called semantic models as they refer to the science of meaning in language and contain the

vocabulary associated with the knowledge domain. Semantic models contain all types of controlled vocabulary from simple lists of terms, to taxonomies, thesauri to full ontologies.

Semantic models offer a way to organize knowledge and information. They make it possible to define a subject domain using a hierarchy of terms and the relationships between those terms. The model makes the subject clear to a user. These models illustrate to a person what they know, what they don't know and what they need to know. Combining





metadata, semantic models and enterprise search produces a contextual navigation in a way that turns a hit-or-miss search experience into rewarding information discovery.

Whether it's researching a company, finding the right product or identifying who to talk with, semantic search is a powerful tool that allows people to quickly find answers or stimulate problem solving.

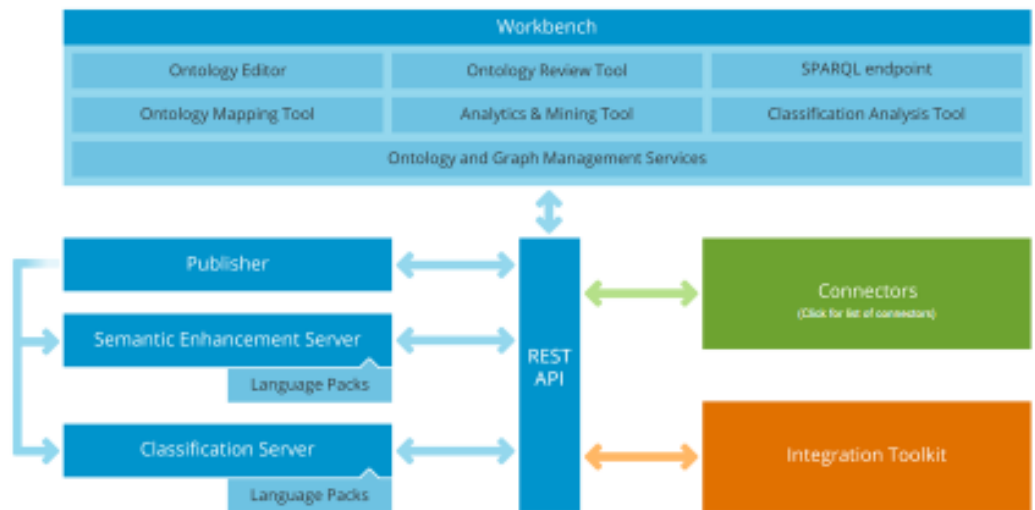
To deliver this context across the enterprise, the system must be manageable, scalable and robust and built using open standards that can integrate into any system that will benefit from enhanced content intelligence.



What is Semaphore?

Semaphore is the name for Smartlogic's software product suite. The software is comprised of core functional modules which provide:

- Complete lifecycle management of complex ontologies, taxonomies or thesauri (Semaphore semantic models).
- A Natural Language Processing engine which analyzes content and applies metadata tags. This process includes rule based classification logic that can classify content against the Semaphore model, entity extraction (companies, person names, places, etc.), phrase extraction (parts of speech) or fact extraction (return the first "thing" matching a specified pattern after a specified phrase).
- A highly scalable index of the Semaphore model that can ingest related information, such as the count of terms used in a search index, to drive any user interface components (facet filters, taxonomy selection trees, "did you mean" widgets, etc.).



A unique selling proposition for Semaphore lies in the integration between these components and the resulting capability that allow an organization to build, deploy and maintain ontology-driven search and classification capabilities with realistic resources.

It's implemented with simple XML based web service interfaces and can be applied to any application that needs to be semantically enabled.

Semaphore offers out-of-the-box integration to SharePoint as well as other leading enterprise search engines.

Some of the key functional capabilities that differentiate Semaphore from other products are shown in the following table:

"At Smartlogic, we believe organizations can outperform others if they fully utilize the huge business value contained in unstructured content.

To realize value, you must understand your content, the information and knowledge it contains, and how it can be applied in the specific context of your operations"

Matthieu Jonglez, CTO - Smartlogic



Functionality	System			
	Semaphore	Direct Competitors	Enterprise Search	Content Management
Semantic Model Management				
Simple Reference List Management	✓	✓	✓	✓
Taxonomy Management (hierarchies, thesauri)	✓	✓	?	?
Ontology Management (concept classes, many relationships between concepts)	✓	?	✗	✗
Multi-Lingual Concept Management	✓	?	✗	?
Web-based model management (task-centric; model change governance)	✓	?	✗	?
Robust, scalable RESTful XML Web Service Interface (to the ontology)	✓	✗	✗	?
Complete Application Programming Interface (API) to ontology features	✓	✓	✗	?
Comprehensive import, export and reporting	✓	?	✗	✓
Automatic Content Classification				
Direct generation of classification logic from semantic model	✓	✗	✗	✗
"About-ness" tagging – assessing terms in context	✓	?	✗	✗
Entity Extraction – identifying places, names, etc.	✓	?	?	✗
Text Mining – highlighting phrases and terms	✓	?	✗	✗
Foreign language processing	✓	?	?	✗
Scalable and resilient for processing large content volumes	✓	✗	✓	?
Enterprise Search				
Indexing and full-text content retrieval	✗	?	✓	✗
Improved precision and relevance by consistent application of metadata	✓	?	✗	✗
Content Management				
Management and storage of metadata	✗	?	✗	✓
Sophisticated Natural Language Processing to determine metadata	✓	?	?	?
Semantic Web				
Help the management and population of triple stores	✓	✓	✗	✗
Provide human interface into SPARQL query systems	✓	✗	✗	✗
Out of the box Content Intelligence integrations				
Google Search Appliance	✓	?		
Apache Solr	✓	?		

? = Not in all cases, or not to a full extent

✓ = Typically available (definitely for Semaphore)

✗ = Typically not in this type of system



Requirement	Semaphore	Specific Products
<p>Manage semantic models that have become too big or complex for office or simple taxonomy management tools.</p> <p>Allow a team to control, edit and report on taxonomies, ontologies and thesauri with mechanisms for review and feedback from a broader audience.</p> <p>Should allow a great deal of flexibility in the model management such as term status, term translation, many user defined relationship, label and note types.</p>	Ontology Management Services	Ontology Server Ontology Editor Text Miner Ontology Review Tool
<p>Automatically analyze text, in any format or language and return metadata tags that classify the item or extract facts, events, relationships or information.</p> <p>The tags should be sourced from the semantic model, or provide useful mark up evidence by algorithmically extracting dates, people, companies place names, relationships, etc.</p> <p>A "Part of Speech" algorithm should provide feedback on the number and groups of phrases within a set of content and show how it relates to the semantic model.</p>	Content Classification Services	Classification Server Advanced Language Packs Classification Review Tool
Enrich existing Enterprise Search, Content Management and Business Intelligence systems with the improved information management and navigation capabilities provided by an ontology.	Semantic Enhancement Services	Semantic Enhancement Server SES SDK
<p>Provide a best-practice interface that exposes all ontology and tagging effort to deliver an exceptional user search and e-discovery experience with minimal development effort.</p> <p>Work directly with Google Search Appliance, Microsoft FAST or Apache Solr Search engines.</p>	Search Application Framework	Search Application Framework



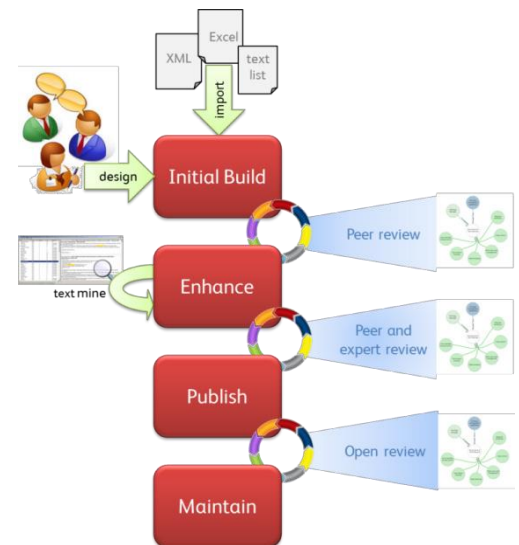
Ontology Management Services

Ontology Editor is at the core of the Semaphore Content Intelligence platform. It's a software tool designed to accelerate the time to build an ontology and maintain the integrity of a model as it matures. This allows more complex vocabularies to be developed than could be achieved using Microsoft Office or other taxonomy management products in a shorter time.

To do this the tool must support:

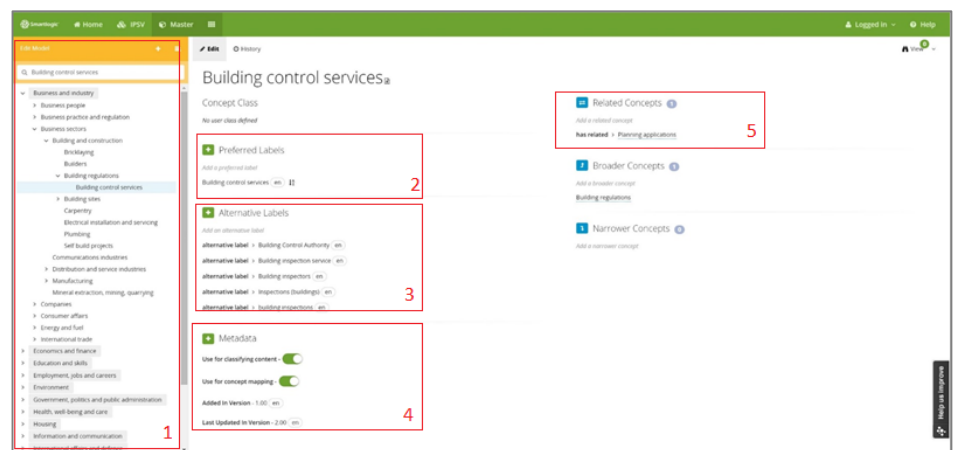
- The design and configuration phase - initial decisions about the ontology structure can change, new elements are added and sections deleted.
- The build phase – importing content in from other sources, Text Mining content for candidate concepts.
- The review phase – the ontology is reviewed by subject matter experts and their feedback incorporated.

Ontology Editor is designed for business analysts or information scientists who have editorial control over the model. It lies at the heart of the Semaphore solution driving the classification, search and discovery experience.

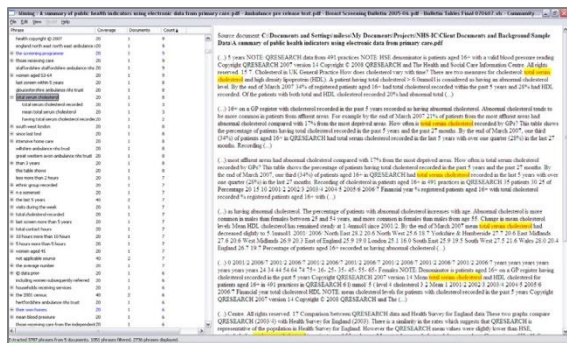


The screenshot shows *Ontology Editor* with some elements that drive the Semaphore experience outlined:

1. Hierarchical relationships can be viewed and edited - allows end users to navigate up and down the taxonomy
2. Preferred labels – the official name (representation) for a concept within an ontology.
3. Alternative labels – i.e. synonyms. Supports “Did you mean” capability and used in automated classification of content.
4. Additional metadata and attributes used to control search behavior, offer information about topics, Best Bets, A-Z’s and Scope Notes.
5. Related concepts - allows end users to browse related topics. Smartlogic supports unlimited relationship classes and behaviors to be defined.



Importing and Linking Models



Typically organizations don't build their ontologies from scratch. With Ontology Editor you can import proprietary vocabularies, link to public industry standard ontologies such as Medical Subject Headings (MeSH) and Financial Industry Business Ontology (FIBO) or choose starter vocabularies directly from Smartlogic and customize them to reflect the unique characteristics of your organization.

The process of importing and reusing vocabularies allow you to jumpstart the model building process, incorporate and promote model standards and save time, money and effort in your model building

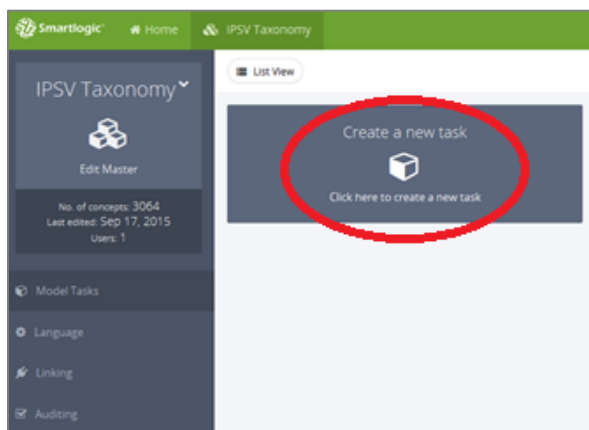
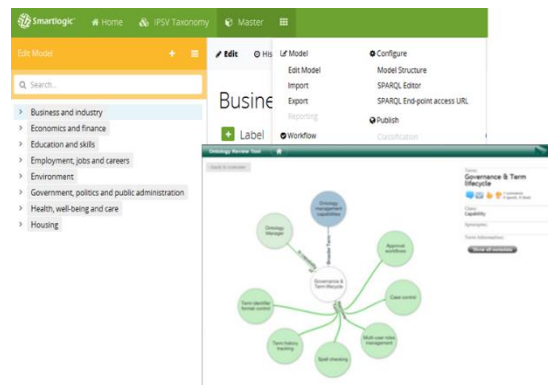
process.

Semaphore's *Text Miner* add-in is a "noun phrase extraction" tool that improves Information Scientist's productivity by focusing them on commonly occurring phrases within a sample document set and letting them view candidate phrases for inclusion in the ontology or thesauri.

Ontology Review Tool

Ontology Editor is typically used by a core team of information professionals who focus on the initial development of the ontology and curate the model over its lifetime. However, these users, while understanding the development process and applications for the ontology are unlikely to know all the terminology that it incorporates.

Ontology Review Tool is a component of Semaphore Workbench that helps gather the vocabulary from subject matter experts and users across the enterprise.



Throughout the development process, model builders can collaborate on one or more models simultaneously.

Creating a new task generates a "change set" of the model so that users can edit, review and approve model changes without affecting production. When changes are complete, they can be promoted for general use.

This iterative process of mining and review results in a precise and consistent model that represents the topics, concepts and unique characteristics of an organization in any subject domain.



Content Classification Services

Semaphore *Classification Server* is a sophisticated "Natural Language Processing" engine. It is a software service that sits and waits for any application in the enterprise - such as content management (e.g. SharePoint, Documentum); or enterprise search (e.g. FAST, Solr, Google Search Appliance), Business Intelligence or Business Workflow engines (including migration tools) – to send it a document, web page, PDF, or text. It analyzes the text and executes four distinct processes:

- Text mining
- Entity extraction
- Fact extraction
- Rule-based classification

Text mining uses "noun phrase extraction" to suggest candidate concepts to enhance an ontology.

Entity extraction uses a combination of algorithms and dictionaries to identify "entities" (person or company names, locations, job roles, phone numbers, etc.).

Fact extraction uses a "capture" rule to identify items of text, for example get the 10 digit code following the phrase "Project Reference" from the document header.


Rule-based classification is the assessment of which concepts in the available taxonomies match the content. How this match is made is determined by a classification strategy, for example, return the topic tag if the topic name is in the title AND at least 3 keywords are in the body AND at least one related topic is present.

Rule Based Classification

Semaphore uses rule-based classification logic (as opposed to Regular Ex, Bayesian Statistical or keyword search methods). Rule based classification results in the highest level of classification accuracy with complete transparency (useful for debugging, continuous improvement and compliance applications.)

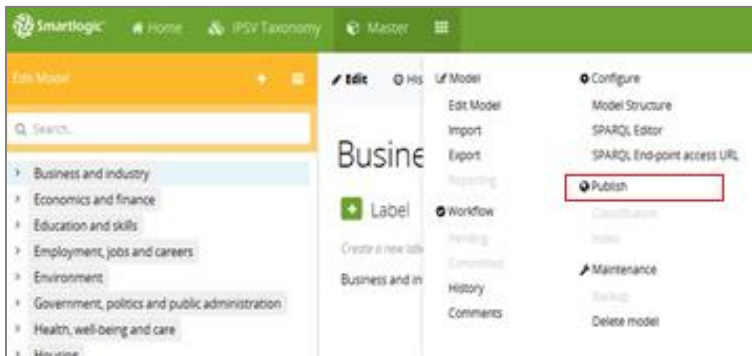
Rule-based classification can be resource intensive to implement, but not in Semaphore. The concepts, labels and relationships in the ontology are a great way to describe a concept. In Semaphore we take this set of language and apply a classification strategy that utilizes the sophisticated text analysis and processing capabilities of *Classification Server*.

A *classification strategy* applies weights to different elements which combine to give a total rule score for the document article. This score can be 0 (no evidence) to 100 (many pieces of matching evidence) and typically we set a threshold of 48 – terms below this have some matches, but not enough to be considered as document tags.



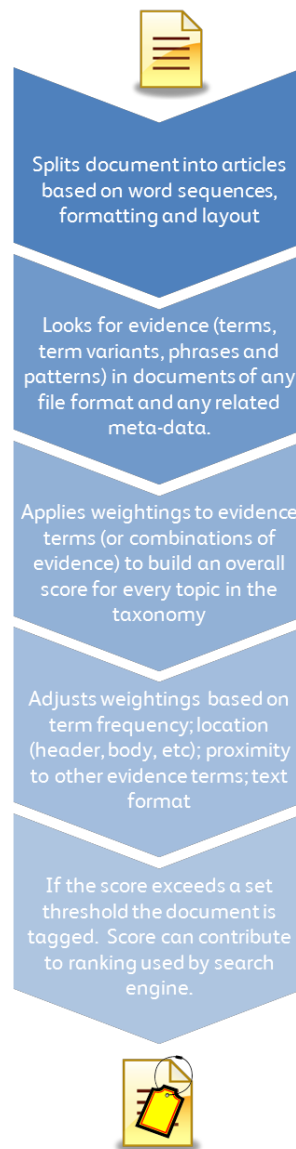
Phrase rule	Terms from document property fields	Exact phrase un-stemmed, matching case	Preferred terms (labels) in Ontology	Link rules to closely related associative terms
Near rule (terms in any order within n words)	Terms from titles	Phrase with stems	Terms flagged with "higher influence" attribute	Link rules to child terms
Terms in a sentence	Emphasised (bold, italic) terms	Single words un-stemmed	Non-preferred terms in ontology (synonyms)	Link rules to standard associative related terms
Terms in a paragraph	Terms in content body	Single word stems	Terms flagged with lower influence attribute	

The grid shows some of the factors that are encapsulated in the rule base template. This is combined with the language (concepts) in the ontology and published to classification server.



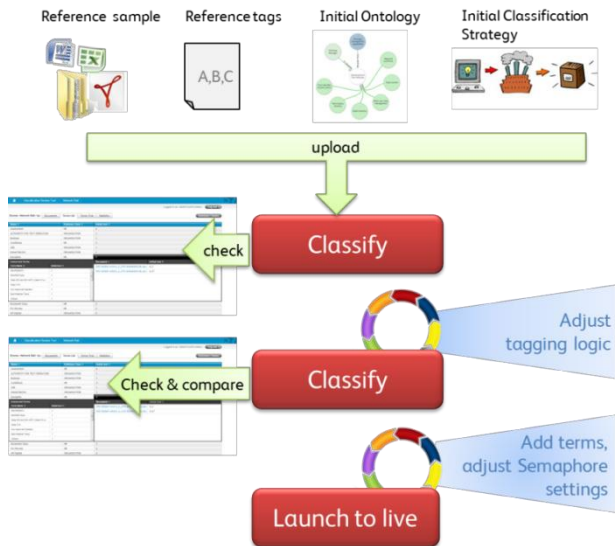
Once configured, the ontology is published from within *Ontology Editor* so that it can be incorporated into the classification process.

Classification Server receives an XML request specifying the item to be classified and any specific parameters. For example, should it treat the content as a “Single Article” (the right approach for a web page) or split it into “Multiple Articles” (the best way to handle a large PDF). Classification Server performs the following tasks:





Classification Testing and Modification



Using Semaphore dramatically reduces the time and effort to deliver this accurate “tagging” capability. Unfortunately, language continues to be ambiguous, subtle, confusing and able throw out all the best laid plans for the classification system.

To help an organization assess the quality of classification, quickly identify anomalies and ensure consistent application of tags, Smartlogic has created the *Classification Review Tool*. This web application is a component of the *Semaphore Workbench* and works alongside the Classification and Semantic Enhancement Server.

The Classification Review Tool is designed to facilitate classification testing by capturing the classification results by document and concept for a representative sample of documents. The user can assess:

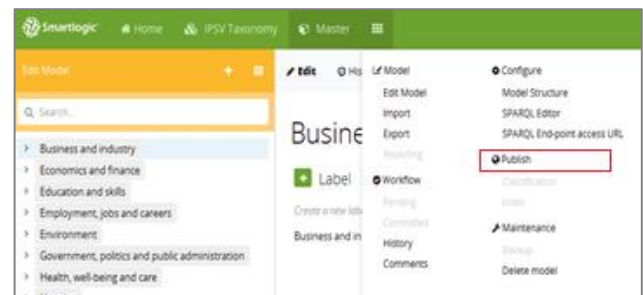
- Whether classification logic is working as expected?
- Are particular concepts over-firing? (Are these “false- positive”?)
- Why are some documents getting no tags? (Are more concepts needed in the vocabulary?)
- What concepts occur concurrently in documents? (Can we enrich the ontology with more relationships?).

Classification Step-by-Step

The Business Classification Scheme is created in *Ontology Editor*. This taxonomy or ontology is a collection of *preferred labels*, i.e. the label to be used for tagging and alternative labels, i.e. evidence for the classification process. When the ontology is sufficiently developed, it is published.

The publish process takes the *vocabulary* from the ontology and applies it to a classification template which holds the *logic* (classification strategy) used to tag content. Different classification approaches can be applied to different facets of an ontology– a people name list could use a fairly simple direct matching on proper nouns, while a subject branch could use a more subtle “about-ness” approach.

A *Review* is created and a set of sample documents loaded. If a reference set of tags that are expected for that sample are available these can be imported and compared.





The results can be reviewed in a number of ways:

Browse - Formula 1 Sample Set- by: Documents Terms List Terms Tree Statistics

Name ID	Initial Run
Australian GP.docx	15
Malaysian GP.docx	13
Brazil GP.docx	10
Bahrain GP.docx	8
Canadian GP.docx	8
German GP.docx	6
Korean GP.docx	5
Monaco GP.docx	5
Spanish GP.docx	5
Team management.docx	5
Turkish GP.docx	5
Belgian GP.docx	4
British GP.docx	4

By document

Browse - Formula 1 Sample Set- by: Documents Terms List Terms Tree Statistics

Term	Initial Run
Car Technology	8
Brakes	1
Engine	1
Safety Car	1
Steering	0
Suspension	0
Tires	3
option tyre	0
prime tyre	0
Wings	0
pit stops	0
Countries and Circuits	0

By term

Browse - Formula 1 Sample Set- by: Documents Terms List Terms Tree Statistics

Name ID	Subclass Class ID	Initial Run
Red Bull	F1	20
McLaren	F1	11
Mercedes	F1	11
Ferrari	F1	10
Lewis Hamilton	F1	9
Mark Webber	F1	8
Sebastian Vettel	F1	7
Felipe Massa	F1	6
Fernando Alonso	F1	6
Red Bull Racing	F1	6
Jenson Button	F1	5
Renault	F1	5
Force India	F1	4

By term within the taxonomy tree

In the *Classification Review Tool* the reviewer can drill in to find additional information. In the Formula 1 example below, selecting the driver "Jenson Button" shows the documents to which that tag was applied and the *tagging score* (indicating how much classification evidence was matched).

Also shown are *concurrent tags*. In this sample the driver, his team and co-driver are also found in the selected document set.

Jenson Button	F1	5
Concurrent Terms		
Term Name ID	Initial Run	
McLaren	3	
Renault	3	
Ferrari	2	
Lewis Hamilton	2	
Mark Webber	2	
Mercedes	2	
Nico Rosberg	2	
Team Brazil	2	
Document		Initial Run ID
Australian GP.docx		0.89
Bahrain GP.docx		0.65
Malaysian GP.docx		0.8
Singapore GP Post Race Conference Tran...		0.65
Turkish GP.docx		0.67

For any document details about how it was classified and what evidence within the document was used to derive the result.

F1_Race	True	15N29472576R0menhobwZNRMLGTs0L3	0.52	9589	highlight
<p>butted to pass Fernando Alonso. He apologised, but it summed up the Englishman's weekend. ¶</p> <p>Hamilton's visit to Australia had already been more than eventful and his race lived up to the hype. He made a good start from 11th on the grid and he was up to 8th by the first corner. A few laps later he seemed to have made a point by passing McLaren team mate Button-only for Button to pull off the inspirational change to dry tyres a lap later. ¶</p> <p>By the time Hamilton came out he was down in 8th place, six places behind his team mate. From then he nonetheless proceeded to battle through the field, making a number of stunning passing moves. He took Rubens Barrichello and Felipe Massa before an amazing pass round the outside of Nico Rosberg at turn 11 and 12. He was up to 3rd by half distance and reeling in 2nd placed Kubica when he was called into the pits for a new set of tyres. The two Mercedes had just done the same and the improvement in their pace convinced McLaren to make an ultimately poor call. ¶</p>					

Bulk Classification and RDF Output

The *Bulk classification utility* can be pointed at an area of a file system to classify the content. Output from this can be in a CSV format which can be analyzed in an Excel pivot table.

The output can also be raw XML and XML with an XSLT transformation. Smartlogic provides XSLT to generate RDF output – capturing triples that describe the content (Entity "xyz" > occurs in URI > Document URI)



Semantic Enhancement Services

Semantic Enhancement Server (SES) is a web service interface to the ontology. It is a technology that drives many interface components in search and content management systems.

Semantic Enhancement Server is a separate component to Ontology Editor which contains a snapshot of the ontology content – essentially holding a published “live access” view of the ontology allowing the base ontology to continue to be updated and modified in the background.

When working in conjunction with search engines such as Solr, FAST and Google Search Appliance, the SES Count Updater service will add the term frequency (number of URIs indexed with a particular ontology concept) into the index.

The server is optimized for high query per second throughput and linear scalability over multiple CPU cores as required.

The server is called with HTTP GET/POST or REST request and returns information in XML or JSON format. The core services include:

- Concept Information
- Concept Hierarchy
- Concept Mapping (think of it as free-text query over the ontology)
- Concept Prefix (search as you type)
- Concept Changes
- Facet List
- A-Z (concepts can be flagged with an attribute to drive an A-Z listing)

This service is used by many large web portals powering search, navigation and browse components that need to reference the business ontology. Some examples are shown here.

Search suggestions as you type

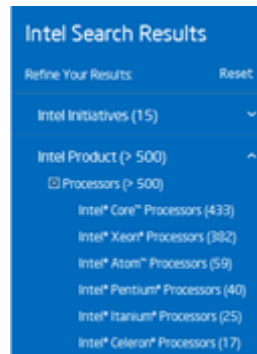
NHS Choices (www.nhs.uk) receives 6m visitors each month. Every character entered in the search box sends a query to SES to request terms.





Refinement panels

Allowing users to navigate a set of search results is common, but the experience is enhanced by putting the possible concept filters in the context of a facet and hierarchy. Calls to SES hierarchy service provide the ontology context.



Visualization

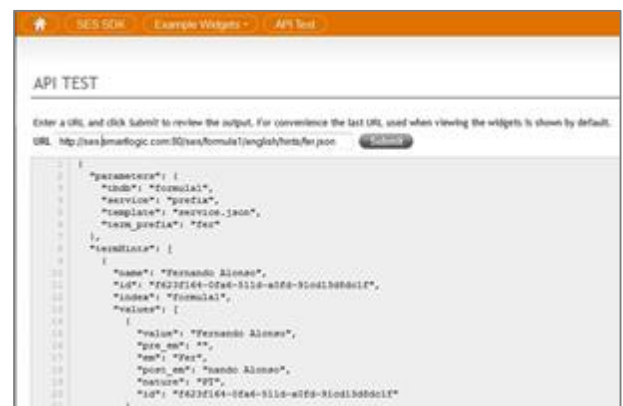
Ontologies are knowledge maps of a domain or subject area. The visualization here is a dynamic Flash component where users can click anywhere to browse the model. This component is used in a number of SharePoint based research portals, where users browse content using the visual map and associated search results are displayed for each node. Every click requests all the concept information from SES.









SES SDK

The SES SDK is part of *Semaphore Workbench*. It illustrates a number of ways of exploring a model interactively. A number of example widgets are provided which can be re-used to form the basis of integration projects. The *API test* section offers an inspection tool to help troubleshooting during integration projects.

The SDK provides sample “widgets” that can be quickly embedded into custom applications and interfaces.



Start typing in the box below and use the up and down keys to select a term. Click or press the enter key to make your selection.

Lotus Renault GP  Red Bull (Sponsor)                 

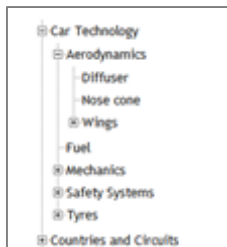
Fernando Alonso
in **Key People**

Ferrari
in **F1 Teams**

Ferrari (Engine)
in **Related Organizations**

Tony **Fernandes**
in **Key People**

Toro Rosso (Tony Rosso **Ferrari**)
in **F1 Teams**



Tree Browser

Type in or paste a user query then press enter or click the Submit button

red

Did you mean one of the following term(s)?

- Red Bull [5d1e7a6e-bf4e-57b4-9fb3-ae26fe31f9ca]
- Red Bull [sponsor] [3743d453-f69a-524b-9b2f-48b3163a35f5]
- Race start [0b00cb77-c208-431b-8c7c-f74420659291]
- Super-soft tyre [406c3447-e975-468f-a84a-8a6c4e7ad5c6]

Concept Mapping

The visualization code (previously illustrated) is also available on the SES SDK.

Search Application Framework

Semaphore Search Application Framework delivers a best practice search interface that combines the power of Semaphore with an Enterprise Search platform (i.e. Google Search Application, Microsoft FAST, and Apache Solr). A great search experience is provided out of the box and can be customized to meet an organizations specific requirements with minimal development effort.

The application comprises a collection of web services and tag libraries that expose core functionality like search results, relevance/date ranking, etc. with the advanced Semaphore capability of taxonomy or entity- based facet navigators and topic maps and filters. It is implemented on a technology platform that sits as a layer abstracted from the search engine itself. The code is provided as a fully functioning application that uses one or more of the applications processors to hook the interface into the native workflow of several search engines.



Buy versus Build?

The creation of a best practice search portal requires a mixed skill set of user experience and search savvy technologists to build a system that appears intuitive and seamless to a user looking for the right information.

The elements that make up good search systems are not well understood. For example, there is no common web model that indicates if selecting a search navigator should filter the current results or search again using

A membership society with 150,000 members and supported by web-savvy business analysts installed the Semaphore Search Application Framework on a Friday and went live with a new interface over their Solr index on Sunday.

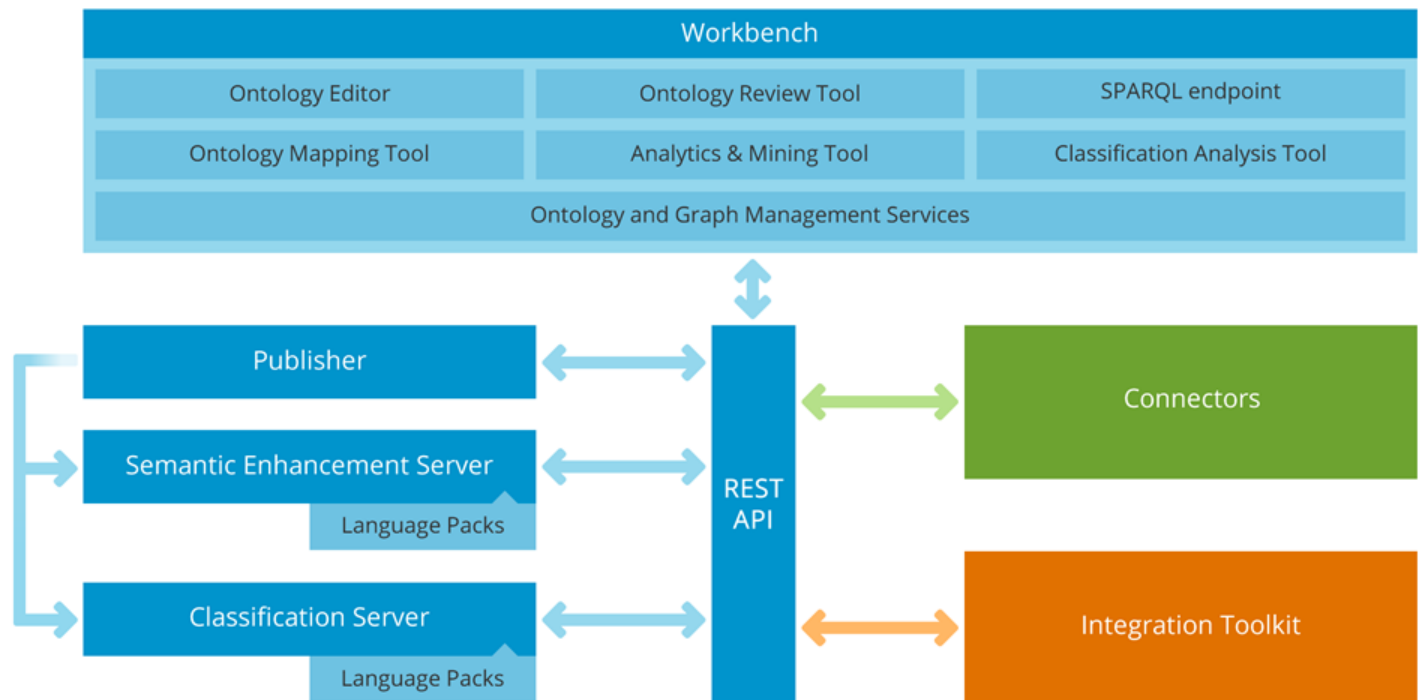
a new concept. The flow and positioning of search refiners is familiar to users on e-commerce sites, but less well understood on informational site.

Smartlogic supported a large government organization, who built their intranet using the Google Search Appliance. A dedicated team of four developers, very knowledgeable in GSA APIs, worked for 3 months to build the search interface.

Every organization has different mixes of skills and resources. Search Application Framework provides a portable, best practice solution to those who do not want to become Search Interface Design experts.



Technical Architecture





Semaphore Integrations

Smartlogic provides integration components for:

- Microsoft FAST for SharePoint / FAST ESP
- Google Search Appliance
- Apache Solr
- Microsoft SharePoint 2007, 2010, 2013.

Our partners and clients have integrated with other types and makes of system. Some of our integrated solutions are described here.

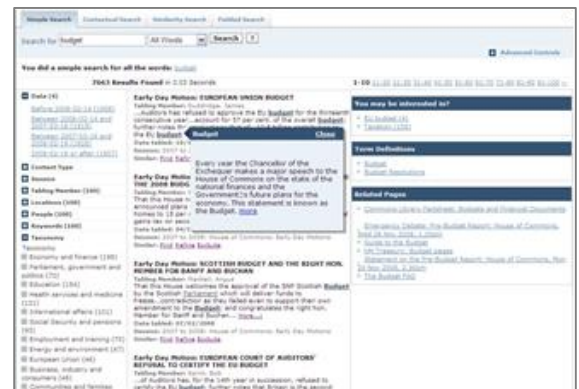
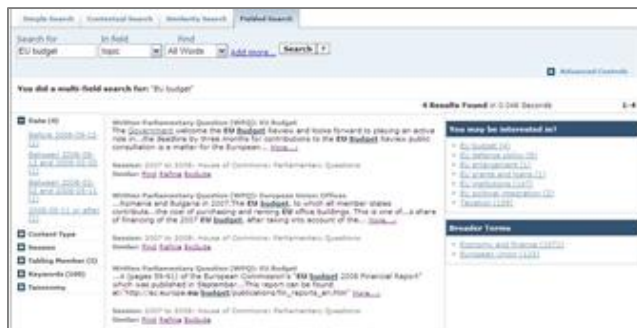
Semaphore for FAST

A model based view of information is a complimentary way to view and filter information and provide an additional means to the 'corpus' view that search engines offer. The Semaphore model view and the FAST corpus view coexist offering the highest levels of findability when used together.

In this screen-shot, the semantic model is a glossary of terms. This list, with its definitions and metadata (such as Related Pages) is managed in Semaphore Ontology Editor and made available to the FAST search interface via our Semantic Enhancement Server.

In this example, the glossary term "Budget" has been highlighted in the search result dynamic summary, and a Semaphore pop-up displays the definition of this concept – a *scope note* in the model. The left hand navigation exposes the facets dynamically created by the FAST ESP engine and delivers a "corpus view" – i.e. relationships FAST has determined from the available content set. Included in those is the subject taxonomy facet. This idea of "concepts" is a key difference between full-text retrieval and classification. Even though a document *mentions* words that relate to a specific concept does not mean that the document is *about* that concept. Algorithms and Dictionary lookup still do not provide good enough results to make the facet useful. Semaphore's Content Classification Services add that intelligence to the process to improve the end result for all users.

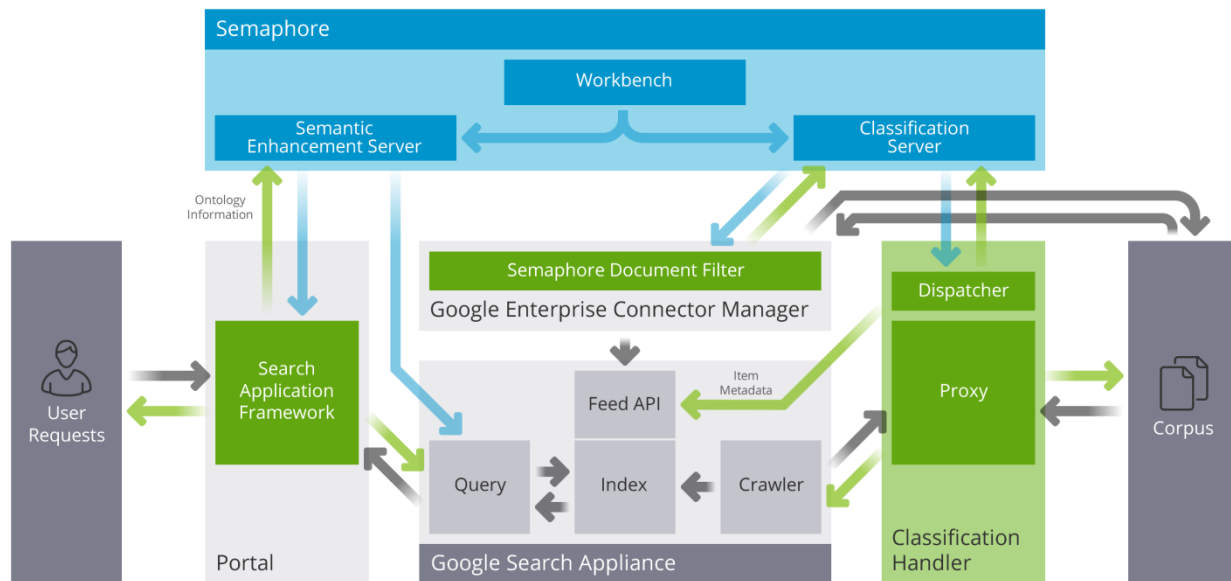
The right hand navigation exposes Semaphore driven navigation and shows a "model view", displaying links to other related topics that an information scientist has decided are relevant and has encapsulated in the under-lying Semaphore model. The FAST facets have no understanding of these valuable topic links.



In this second screen-shot, the Semaphore navigation delivers increased capability – the ability for a user to "drill up, down or across" using the concepts defined in the model and the metadata added by the classification process.



Semaphore for Google Search Appliance (GSA)



Semaphore for Google Search Appliance delivers higher quality search results through taxonomies, ontologies and automated content classification. The addition of Semaphore provides the following benefits to GSA:

- Accurate and complete search results with the most relevant content at the top by applying complex classification routines to add metadata to the GSA index. The latest GSA allows biasing of search ranking to these metadata fields.
- Natural Language Processing capabilities that complement the GSA entity extraction routines.
- Improved findability by enhancing Google enterprise search experience: users can filter and navigate topics, break down results, identify areas of interest, explore related topics and locate expertise, and resources.

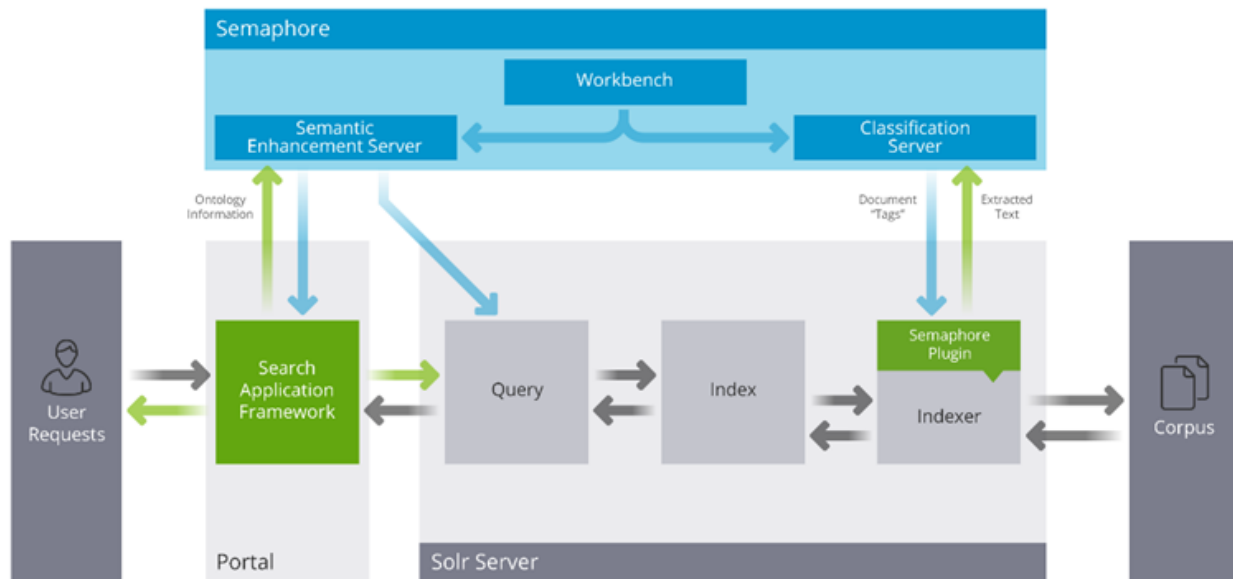
Semaphore's Content Intelligence Platform adds the power of advanced "aboutness" Classification to the GSA. A tried and tested integration works with any GSA ensuring all content indexed by the Google Search Appliance is passed to the Semaphore Classification Server for tagging. The Semaphore system scales to match the throughput required for the largest of appliances. Utilizing the Onebox API and tight integration techniques to the indexing pipeline, Smartlogic provides additional functionality to the powerful search capabilities of the GSA.

Semaphore Search Application Framework provides a portal interface with configurable widgets that are designed to work with and best utilize the Semaphore capabilities via the Semantic Enhancement we service, making the ontology detail available to the largest of portal applications.

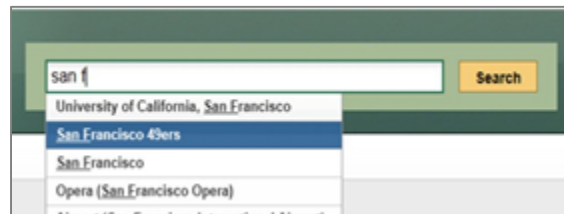




Semaphore for Apache Solr



Smartlogic turns Solr into a powerful semantic search platform. Our integration components inject additional metadata into the Solr free-text index. That metadata might include classifications against an ontology or other classification scheme, entities such as place, people or company names and facts such as project codes or NI number. Solr is a flexible engine and lends itself to make use of the additional metadata for refinement filters or as factors to boost the relevancy ranking model.

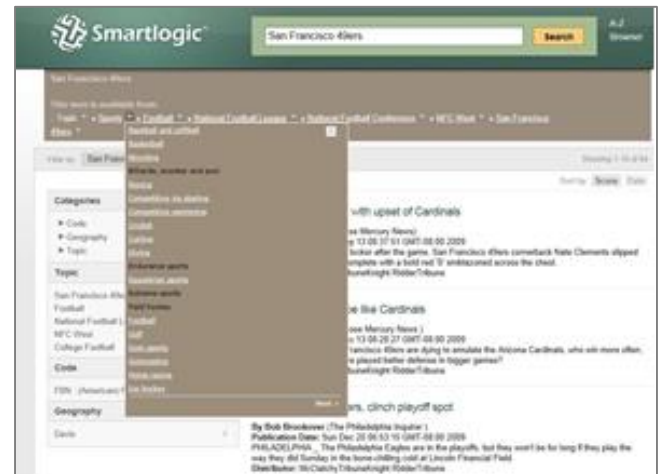


Semaphore search suggestions (sourced from an ontology of business vocabulary concepts)

▼ Topic	84
▶ Human Interest	1
▼ Sports	84
▶ Baseball And Softball	4
▼ Football	84
▶ College Football	5
▼ National Football League	84
▶ American Football Conf...	5
▼ National Football Conf...	84
▶ NFC East	2
▶ NFC North	2
▼ NFC West	84
Arizona Cardinals	2
San Francisco 49ers	84
Seattle Seahawks	2
St. Louis Rams	1

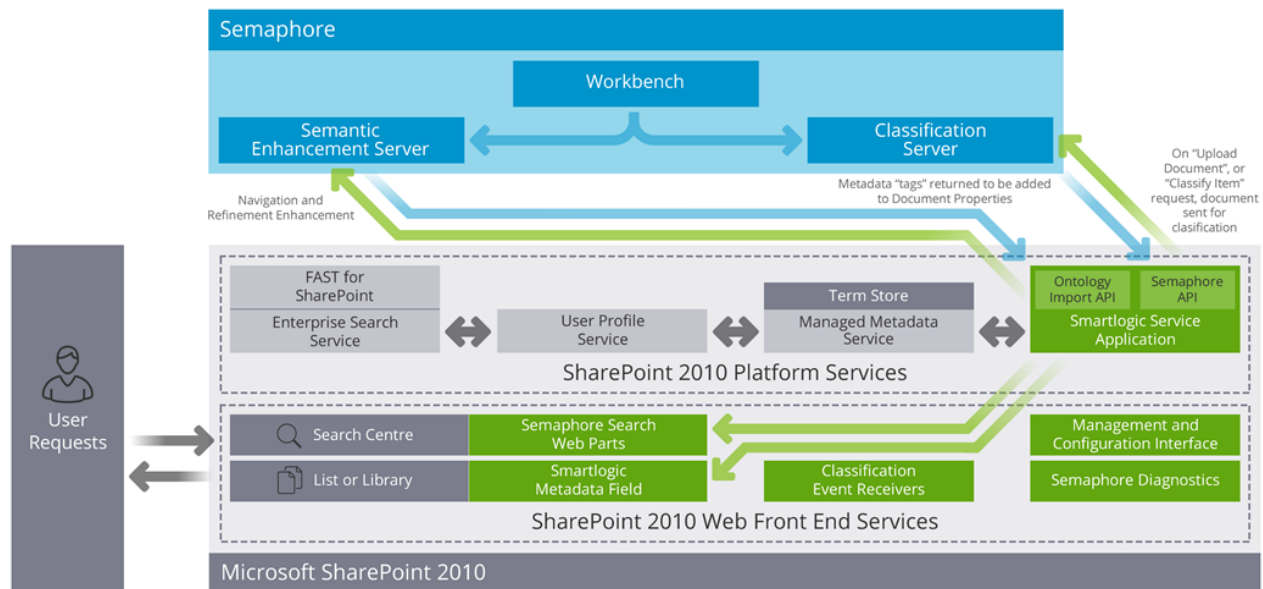
A refinement panel can offer choices in a flat list, by facet, or concepts can be put in better context by showing the search result filters in the context of an expandable hierarchy.

Quality metadata and semantic navigation components drive a rich user find and discovery experience in addition to all the benefits of the Solr engine.





Semaphore for SharePoint



Smartlogic provides a SharePoint Solution that integrates Semaphore's Content Intelligence platform with Microsoft Office SharePoint Server (MOSS 2007), SharePoint 2010.¹ SharePoint 2013 and O365.

In any corporate SharePoint implementation metadata is critical: if you know what something is about you can put processes in place to *manage* it and improve the ability to find the information.

The problem is that users are put off by completing "Add Properties" forms each time they load a document into a library. Often SharePoint is a change in the way they normally work and adding long selection lists may be the last straw. On the other hand, if the metadata is too simplistic all of your content will reside in a few large metadata buckets – which doesn't help you find or manage your content.

In brief, the integration provides:

- Automatic or user prompted "tagging" of content uploaded into libraries from a taxonomy
- The ability to classify whole libraries from the interface or the command line
- Semaphore Search and Navigation web parts that exploit the metadata in the search index of SharePoint Server Search or FAST for SharePoint
- Content Type Updater rules that work with the DropOff Library for document routing.
- Farm level services and solution deployment and administration interfaces into web applications, site collections and libraries.

¹ There are different levels of functionality, web-parts, etc. available for the SharePoint 2007 and 2010 releases. This document illustrates our Semaphore for SharePoint 2010 product.

Automatic and Assisted Classification

Semaphore classifies content as it is loaded into the system and suggests “tags” to the user. In *assisted tagging mode* the user can search for other terms using a search-as-you-type process or browse the ontology in a tree view.

The benefit in adding the tags comes through aggregation of content in the SharePoint interface and search. Views can be created on the Semaphore metadata which means users have an alternative way of finding documents – reducing the imperative to “file” in the right place and the pain caused as documents can be about many things, but can only be *stored* in one place.

Semaphore metadata is stored as a Site Column in a Content type – i.e. working with or extending standard SharePoint features. This means it can be used to drive workflows or absorbed into the standard SharePoint Search or FAST for SharePoint index from which can provide a compelling user findability experience.

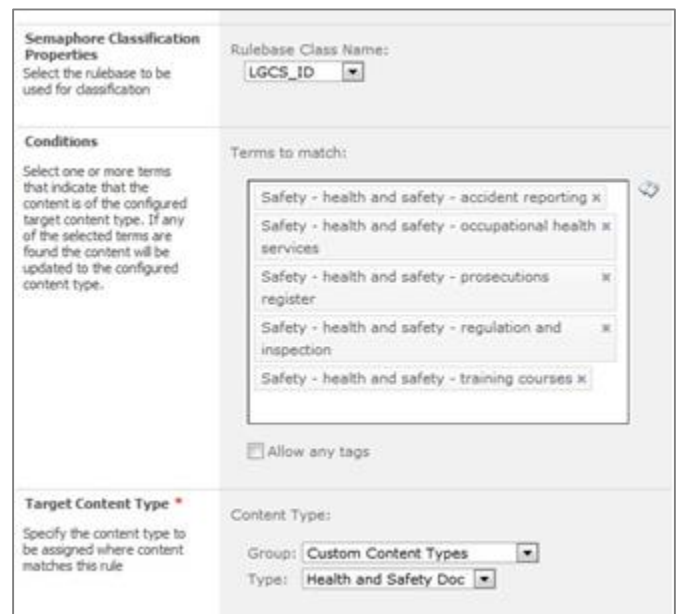


Content Type Updater

Applying the correct “Content Types” can be critical to managing your SharePoint content – for example Content Types control the record retention policy applied.

Once again, while the Information Architects that define the solution may know which of many Content Types to apply, a user might be less certain.

The Semaphore Content Type Updater works off the tags returned by automatic classification. This can look for a document’s subject, or user clues in the title, metadata, header/footer and body of a document to determine a logical content type, security marking, etc. Based on rules set up in the site collection Semaphore can apply a content type on upload, or reset existing content types.

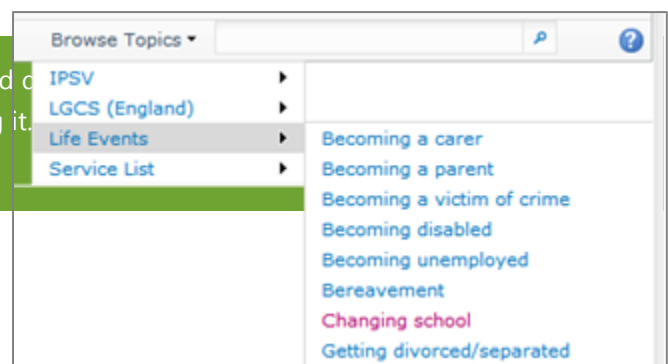


Semaphore Web Parts

Semaphore for SharePoint automatically classifies content and displays tags. This allows users to see the real benefits of metadata without the effort of adding it. It helps organizations to save time and money.

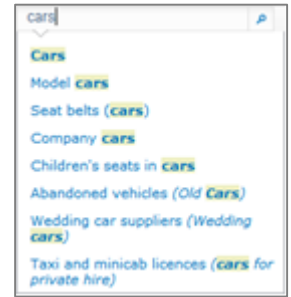
Browse Topics

Browse Topics Pop-up displays the top level of an ontology to give the user a starting point for search.



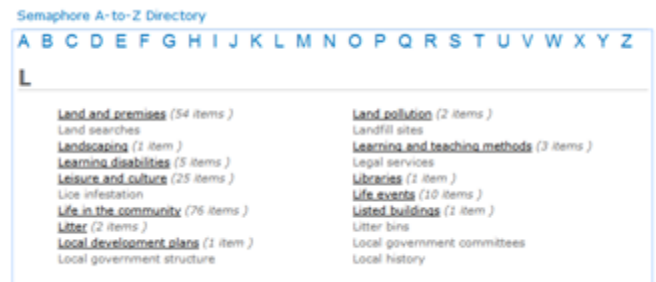
Search-as-you-Type

Search-as-you-Type web part suggests topics from the ontology as the user enters the search string. This can span many facets of the taxonomy and uses both the preferred and alternative labels. Unlike other auto-suggestion solutions that only match if the word is at the beginning of the search text, this web part will match where the search word is anywhere in the word.



A to Z Listing

A to Z Directory uses an A-Z flag set in the ontology on the concepts you want to show. There are options to exclude concepts where there are no tagged documents or to show the count of documents for each concept. This all obeys security trimming and provides an easy to understand way of accessing search results.



Tag Cloud

Based on the popular animated WordPress design, the Smartlogic *Tag Cloud* shows all the tags added to content by Semaphore and users. Completely integrated to the SharePoint social tagging aspects, this control links to search or the user profile service. With this a user can "follow" topics by adding tags into their profile.



Taxonomy Browser

The Taxonomy Browser shows the content organized by subject. This means that users researching a topic do not need to know where the information is located. From this web part they can simply select the topic of interest to access all items tagged.





Best Bet

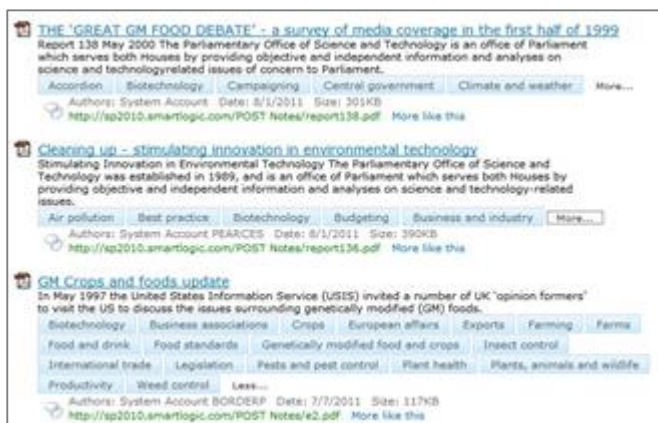
The *Best Bet* web part uses information held with the concept in the ontology to provide more information, e.g. a glossary description, or to point the user to a resource on the subject such as the web site shown in this example.



Search Results

Semaphore *Search Results* can be enhanced by displaying the tags applied for each result. Selecting a tag will re-execute the search returning all items that have this tag.

A “more like this” option runs a search for all of the document tags and returns similar content.



Ontology Visualization

The *Visualizer* web part shows a graphical and animated view of the ontology. The user can select a concept “bubble” to re-orientate the model around that topic.

The search results for the central topic update as the user browses. This provides a whole new way to browse your organization’s content.



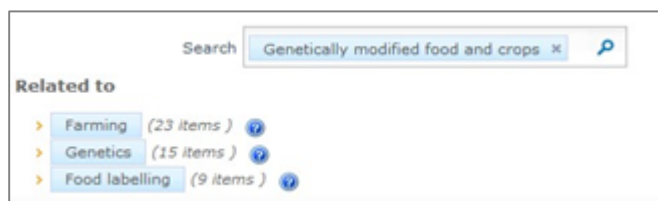
Concept Mapping

Concept Mapping post-search helps a user that has entered a free-text search and received no results or too many results. This web part offers suggestions from the ontology to help them find the right search term.



Related Terms

The Related Terms web part offers the user suggestions of topics related to the tag they previously selected.





Topic Browser

The Topic Browser aims to put the selected concept in context by showing where it sits in the ontology path in a breadcrumb style display. Semaphore adds “poly- hierarchy” support which is not available in the standard Term Store capabilities.

Environment (100+) > Farming (23) > Crops (19) > Genetically modified food and crops
Science, technology and innovation (100+) > Biotechnology (14) > Genetically modified food and crops
Health, well-being and care (100+) > Food and drink (31) > Genetically modified food and crops

Navigator

The *Semaphore Navigator* brings together a number of our web parts into a tabbed container. This can be used underneath the search results, the idea being a user failing to find what they need on the first page of results will have all the help Semaphore provides available to them.



Refinement Panel

The Semaphore Refinement Panel extends the standard refinement panel web part. Semaphore tags can be displayed alongside other site metadata.

The tags can be displayed in multiple layouts:

- a flat list
- a flat list grouped by facet
- a semi-hierarchical list (as shown)

This web part displays the concepts available in the current result set. Selecting one filters the current results down to only those with the selected tag.



Summary

Smartlogic works with our clients to:

- Add semantic model management, automatic classification, text mining and enhanced navigation to any application
- Allow users to break down, explore and learn about topics
- Analyze content to support Business Intelligence, Big Data and Workflow applications
- Improve findability, precision and completeness
- Assist governance by identifying personal information or other sensitive text

Semaphore provides the capability to:

- Build, import, link to, manage and deploy taxonomies and ontologies across multiple websites and applications
- Organize content according to your ontologies through accurate and controllable automated classification



- Extract entities and facts from content
- Provide greater control over the search and navigation experience.
- Allow users to analyze, organize and refine results, explore related content and topics, locate useful resources and break down topics

Semaphore's Content Intelligence platform is a robust, scalable enterprise application with many proven installations and references. The software has open API and Web Service interfaces allowing organizations to add semantic enhancements to any system. Smartlogic provides packaged integrations for:

- Apache Solr
- Google Search Appliance
- Microsoft Fast for SharePoint
- Microsoft SharePoint



About Smartlogic

Smartlogic's Semaphore is an enterprise grade Content Intelligence platform that complements an organization's investment in enterprise search, workflow, analytics and content management systems. Semaphore is the bridge between human language and computers. It extracts the human intelligence from information assets and applies precise and consistent metadata to harmonize information and drive business decisions.

Global organizations in the financial services, oil and gas, healthcare, government, life sciences, high tech manufacturing, media and publishing, and retail industries use Semaphore to manage ontologies and taxonomies, drive automatic classification and provide contextual navigation. With Semaphore, they harmonize data, improve search and retrieval, drive workflows, secure sensitive information, comply with governmental regulations and replace rigid data warehouse systems with flexible solutions in a fraction of the time and cost to improve operations and gain competitive advantage.

SMARTLOGIC US

111 N MARKET ST SUITE 300
SAN JOSE, CALIFORNIA, 95113
TEL: 408 213 9500

SMARTLOGIC UK

200 ALDERSGATE
LONDON, EC1A 4HD
TEL: +44 (0)203 176 4500

WWW.SMARTLOGIC.COM
INFO@SMARTLOGIC.COM

© 2015 SMARTLOGIC SEMAPHORE LIMITED