Smartlogic™

White paper

The Semantics of Semaphore

What is the Semantic Web?

In today's enterprise, unstructured information has become one of the most valuable assets of the organization and its rate of growth is staggering. The need to extract information from internal and external sources and integrate it into existing systems for use in key business decisions is critical to successful business outcomes.

The Semantic Web was supposed to help us solve this problem. By extending the principles of the Web to documents and data, all information would be available, required facts a click away and intelligent agents would locate things. Yet without a mechanism to apply context and meaning to information, the enterprise cannot integrate it into their systems and the Semantic Web fails. Semaphore, Smartlogic's Semantic AI platform, has been carefully architected to address this problem.

The underlying design principles of Semaphore are a precise combination of W3C standards and semantic best practices built upon Resource Description Framework (RDF), SKOS and SKOS-XL, which allow organizations to unify all enterprise information into a single information source to identify relationships, explore opportunities and answer questions they could not answer before. Semaphore uses AI, Machine Learning, Natural Language Processing and other sophisticated techniques to produce precise, complete and consistent results.

The Value of RDF

Resource Description Framework (RDF) is a basic building block of the Semantic Web. In its simplest form, RDF is used to make statements about resources or things in the world using a subject–predicate–object expression such as, Fred is 55 or Fred lives in Texas. These expressions are also known as triples; powerful, flexible and scalable structures that allow machines to draw inference and drive information discovery.

Different than relational models, triple stores represent a paradigm shift in the data organizational model motivated by:

- The inability of traditional, relational systems to ingest, manage and analyze Big Data
- The need for a flexible data model that can handle all information assets; structured, semi-structured and unstructured
- The ability to drive information discovery without having to pre-suppose questions – a requirement of traditional technology
- An organizational need to incorporate the valuable, unstructured information throughout the enterprise that is currently not part of the structured data set
- The ability to reuse and link to information that already exists in internal and external sources to promote interoperability and avoid duplication

The process of normalizing data, classifying content, building models and expressing the results as RDF using Semaphore, allow organizations to harmonize all enterprise information assets; structured and unstructured, internal and external, into distributed, decentralized metadata applications in an agile an interoperable way.

Linked Data

At the heart of the Semantic Web is Linked Data: the term used to describe a way to expose and connect data on the Web from different sources. In the same way the Web uses hypertext links to move from one document to another, Linked Data is the hyper-data link that allow people and machines to locate and link to related information regardless of where it resides.

In 2006, Tim Berners-Lee identified four basic principles of Linked Data:

- Information should be identified using Uniform Resource Identifiers (URI)
- Use the HTTP protocol with URIs so that information can be located (interpreted, dereferenced)
- Use open standards such as RDF and SPARQL to provide useful information about what a name identifies
- Refer to other information using its HTTP URI-based name when publishing data on the Web

When an enterprise can take advantage of Linked Data's value, economies of scale, scope and cost effectiveness are realized.

Incorporating Linked Data principles brings value to the enterprise by:

- Building vocabularies using Linked Data allows organizations to import or link to publically available vocabularies and expand their model to incorporate company or domain specific knowledge.
- Providing the capability to link internal and external information together - drives information discovery.
- Reducing or eliminating costly ETL processes - by linking directly to the data source, information (internal as well as external) can be repurposed across enterprise, community and domain boundaries.

Perception versus Reality

In spite of technology and standards, organizations continue to struggle with the volume, velocity and variety of data in the enterprise, the quality of public information is suspect, and the context in which that information was created is seldom obvious.

The real problem is description. It's difficult to extract value from information unless it is well described; the context, the topics it pertains to and the relationships between facts must be clearly identified in a way machines can interpret. Smartlogic's Semaphore platform 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 help solve the description problem.

The Semaphore platform incorporates the power of Semantic Web standards, Ontology mapping and Linked Data, which allow organizations to connect related information across internal and external boundaries to leverage their own knowledge as well as the accumulated knowledge in most problem domains. Semaphore provides the tools and technologies organizations need to harmonize their information assets and make them useful throughout the enterprise.

Semaphore Knowledge Model Management begins with a model that captures the topics, concepts, market segments and organizational structures that uniquely reflect an organization. Model building can be streamlined using Linked Open Vocabularies, linking to internal enterprise vocabularies or by mining enterprise content. These models can then be enriched with organization specific information to ensure the model accurately reflects the enterprise and use case.

The model is published and rule bases are generated and combined with Natural Language Processing, entity and fact extraction and subject, topic and thematic classification strategies in the auto-classification process to result in precise and consistent metadata.

When content is tagged and metadata is expressed in RDF, organizations can automatically harmonize their information, aggregate and relate facts from disparate data sources and combine them with graph-based technologies to identify patterns and look for relationships.

The Semantic Building Blocks of Semaphore

Semaphore's technology stack supports RDF, RDFS, SKOS, SKOS XL and OWL where each technology builds upon its predecessor to result in a robust semantic platform to drive model management and classification processes.

Semaphore's unique ability to export both the model and classification results in a standard RDF format allow it to be stored and subsequently explored using graph technology. Each RDF triple is identified by a URI, which connects concepts found in the model to the related information found in the documents.

Resource Definition Framework Schema (RDFS)

RDFS is used to define the fundamental structure — classes and properties -- of Semaphore models; for example, what associative relationships are valid with what types of concepts. A geography model may have a structure, which shows that the "has capital" relationship property is only valid between a country or state and a city. This is defined using the RDFS construct of a property's domain and range.

Simple Knowledge Organization System (SKOS) and SKOS XL (eXtension)

SKOS, a W3C standard, provides a standard way to represent concepts within knowledge organization systems such as taxonomies and thesauri, using the RDFS Plus framework. Because SKOS is based on RDF, concepts are machine-readable and can be exchanged between software applications, published on the Web, linked with other data on the Web and integrated into other concept schemes. SKOS is designed for the labeling and identification of concepts, its purpose is not for reasoning.

This process of labeling and identification allow you to easily manage internationalized vocabularies. For example, a single concept with an American preferred label of "cat" can also have a German preferred label of "katze", a French preferred label of "chat" and an American alternative label of "kitty cat" - all of these labels refer to the same concept and can be used in the classification process to increase precision and consistency.

SKOS's extensibility allows you to attach all the metadata you want to a particular concept, it does not however, allow you to attach metadata to items defined as labels (i.e. preferred labels and alternative labels) because they are nothing more than strings. To assign metadata to the labels themselves, such as the name of the person who added a particular label, or the date it was last updated Semaphore uses SKOS XL. SKOS XL allows us to identify an unlimited number of additional properties, which enrich classification results.

Web Ontology Language (OWL)

OWL is a family of knowledge representation languages for authoring models and is designed to be used by software developers who are writing semantic web applications. OWL plays an active role in enforcing the integrity and design of a model through description logic, inference and building formal class definitions.

OWL's distinguishing feature - expressivity - lets you model extremely complicated and subtle ideas about your data with a rich collection of operators to describe properties and classes. And OWL is flexible; all data modeling statements are RFD triples, modifications after the fact can be easily accomplished by modifying the relevant triple.

While OWL principles are used to ensure good model design, is expressive and flexible, it is extremely complicated. Many things cannot be expressed in OWL that can be in SPARQL and so we use SPIN (SPARQL Inferencing Notation) for constraint management.

Semaphore Core

Semaphore core is described in OWL and incorporates all of the design principles that govern how Semaphore models should be built. The Semaphore core design constructs are available as a public resource, from Smartlogic. The Semaphore core is imported in place of SKOS and SKOS-XL and in turn imports them as required.

Smartlogic's technology stack results in a sophisticated semantic platform that turns content into a self-describing resource. Semaphore leverages Linked Data standards and graph database technology to unify all information within the enterprise to be used for analysis, to gain insight and to answer questions they could not answer before.

Semaphore and Semantics; Improve patient care and reduce costs

Each year one in three individuals over the age of 75 will experience a fall. Four out of ten falls will result in hospitalization, a long period of recuperation, surgery or death. As the population ages this problem becomes more acute and more expensive.

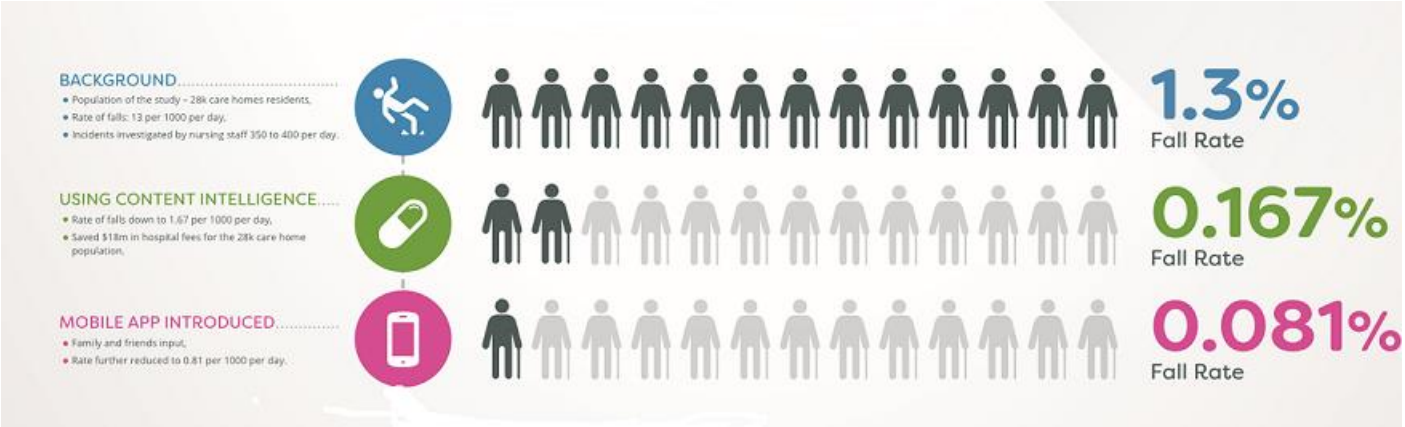
A senior care agency applied the power of Smartlogic's Semantic AI platform Semaphore to their disparate data sources to identify individuals who were at risk of falling in order to improve patient care and reduce costs.

The available information such as patient records, at home visiting nurse's notes and social workers observations, were stored in different systems within different organizations and varied in structure, document type and nomenclature.

The organization leveraged the U.S. National Library of Medicine's Medical Subject Headings (MeSH) public ontology as a starting point and built out additional detail for their area of interest, adding concepts that corresponded to potential incidents and environmental hazards. They created relationships between drugs and drug families and the risk of someone falling using a specific "risk increasing drugs" relationship. Similarly, they associated environmental hazards identified using Semaphore's Text Miner tool, such as loose carpet, steep stairs and missing handrails with the risk of falling.

This detailed model was used by Semaphore's Classification and Language Services to extract facts from all information assets including doctors' records, social worker reports and comments from care givers and visiting nurses. Sophisticated Natural Language Processing techniques within Semaphore ensured a high degree of accuracy in the created metadata, which was expressed as RDF triples.

Using next-generation graph-based search tools, they could explore the content, identify patterns, look for relationships and arrange preventive intervention before damaging incidents occurred.



Their ability to quickly identify patients at risk and take action to eliminate hazards reduced the rate of falls from 13 per 1000 to 1.67 per thousand – an 87% decrease and approximately $18 million dollars in savings. But every fall matters and they believed there was still an opportunity for additional improvement.

After some research they identified a missed opportunity to leverage what relatives may spot during a visit or hear over the phone. They developed a mobile feedback app where relatives could leave messages such as “Mom mentioned she was feeling dizzy yesterday” for nursing staff. These messages are incorporated into the analytics system and provide a new dimension to the prevention campaign.

Since the introduction of the mobile app, the rate of falls has again decreased from 1.67 to .81 per thousand resulting in additional savings and improved care.

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

SMARTLOGIC – EUROPE, MIDDLE-EAST AND AFRICA
200 ALDERSGATE
LONDON, EC1A 4HD
TEL: +44 (0)203 176 4500

WWW.SMARTLOGIC.COM
INFO@SMARTLOGIC.COM