W3C® > Semantic Web Use Cases and Case Studies

Case Study: Enhancement and Integration of Corporate Social Software Using the Semantic Web

Alexandre Passant, Electricité de France R&D & LaLIC, Université Paris-Sorbonne

June 2008

General Description

Introduction

Electricité de France, the largest electricity company in France, recently introduced the use of social software within its R&D department, embracing the Enterprise 2.0 movement. The use of blogs, wikis, free-tagging, and the integration of external RSS feeds offers new possibilities for knowledge management and collaboration between engineers and researchers. Yet, these tools raise various issues, such as:

- Querying data across applications is not straightforward as different applications use different formats (database structure or output format) to model their data
- Knowledge created using wikis cannot be easily understood by computers. For instance, a user cannot run a query such as "List all companies working on solar energy and based in the US". The user would need instead to browse various pages to find the answer
- Free-tagging leads to heterogeneity and ambiguity which complicates the search for relevant content. For instance, a query about "solar" will not retrieve documents tagged with "solar energy" or "solar_energy"
- RSS feeds tend to produce a lot of incoming data, which, for example, makes it difficult to follow all information about a given company

The solution

To solve these problems and offer new and value-added services to end-users, we developed a solution that uses Semantic Web technologies and relies on various components that act together and provide a mediation system between those services and the users. This mediation system provides a common model for meta-data and for document content. It achieves this using ontologies, plugins for existing tools to create data according to these ontologies, a central storage system for this data, and services to enrich information retrieval and data exchange between components.

Since our first requirement was to provide a common and machine-readable model of meta-data for content from any service, we decided to that the model should be implemented in RDF. We then took part in the development of the SIOC (Semantically-Interlinked Online Communities) ontology which provides a model for describing activities of online communities in RDF. For example, SIOC can be used to describe what is a blog post, what properties a blog has, and how a blog post relates to a user and user comments. SIOC takes advantage of commonly used vocabularies such as FOAF (Friend Of A Friend) and Dublin Core. SIOC exporters and translators were added to our services so that wherever the data comes from (blogs, wikis, RSS feeds), it is automatically modeled in a common way, offering a first layer of unified semantics over existing tools in our mediation architecture.
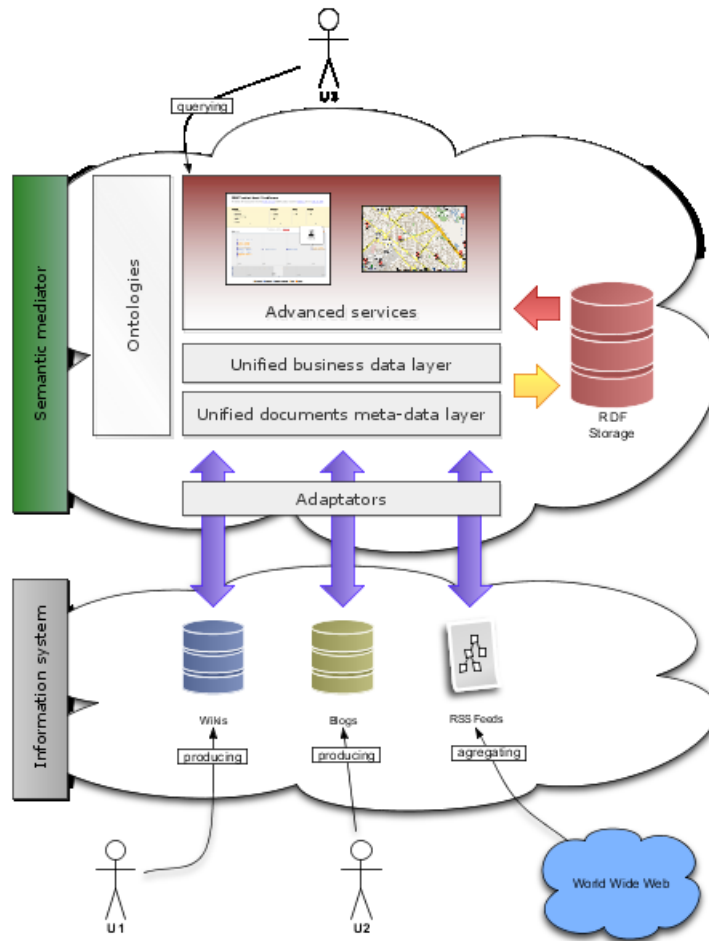
As much valuable knowledge is contained within our wikis, we extended the wiki server with semantic functionalities in order to model some of its content in a machine-readable way. To do this we created ontologies which model the concepts within the knowledge fields of our wikis. For example, we designed an ontology to model information about companies, their industry, and location. In order to benefit from existing models and data, our ontologies extend or reuse existing ones such as Geonames and SKOS (Simple Knowledge Organization System). Moreover, to allow users to easily publish and maintain ontology instances from wiki pages, our add-on provides the ability for wiki administrators to define form templates for wiki pages and to map them to the classes and properties of the ontologies. Thus, users create and maintain instances by editing wiki pages, which is as simple as what they were doing prior to implementing Semantic Web technologies. For instance, instead of writing that *"EDF is an organization located in France"*, a user fills in the template so that the following RDF triples will be immediately created when saving the page, thus providing a second layer of semantics for the mediator:

```
athena:EDF rdf:type foaf:Organization ;
    geonames:locatedIn <http://sws.geonames.org/3017382/> .
```

In order to provide a bridge between the advantages and openness of tagging, and the powerful but complex use of ontologies and semantic annotation, we developed a framework called MOAT (Meaning Of A Tag). MOAT allows users to collaboratively provide links between tags and their meanings. The resources (classes or instances) of the ontologies in our system define the meaning. Thus, users can keep using free-keywords when tagging content, since this layer helps to solve ambiguity and heterogeneity problems, as different tags can be related to the same resource (eg: *"solar"* and *"solar_energy"* linked to athena:SolarEnergy). Moreover, users can browse a human-readable version (using labels instead of URIs) of the ontology in case they want to add a new link or if the tag leads to ambiguity and they must chose the relevant resource when tagging content. Furthermore, when saving tagged content, the links between content and resources are exported in a RDF export using SIOC and MOAT.

Each time a service produces a new document, the storage system is notified by the plug-ins of our mediation architecture, saves its RDF data instantaneously and merges it with other data. This allows us to benefit from a unique view of the many integrated data sources (e.g. blogs, wikis, RSS) and to have access to

up to date information. Then, using the SPARQL query language and protocol, we can query across the many data sources, and services can be plugged on top of the central storage system.

U3

querying

Semantic mediator

Ontologies

Advanced services

Unified business data layer

Unified documents meta-data layer

RDF Storage

Adaptators

Information system

Wikis

Blogs

RSS Feeds

producing

producing

agregating

U1

U2

World Wide Web

(a larger version of the screenshot is also available)

The most beneficial service we developed is a dedicated semantic search engine, which allows users to find information by concept, from a given keyword, using all available sources of information. When a user searches for *"France"*, the system will suggest all instances containing that word whether it is used in the label or the tag. For example, the user will retrieve *"Association des Maires de France"*, *"France"* and *"Electricité de France"*. This approach allows the user to define precisely what they are looking for and to then displaying related information from wikis, blogs, and RSS feeds. In addition, our system can also reuse resources labels to provide a first step of semi-automatic indexing of incoming RSS feeds. The system also proposes to extend search regarding relationships between concepts that exist in our ontologies, for example, suggesting *"Solar energy"* when searching for *"Renewable energies"*.

Résultats de votre recherche sur "AMF - Association des Maires de France", associé aux mot-clés:

- AMF
- association des maires de France

Voir l'URI

Page wiki hpedia principale pour "AMF - Association des Maires de France"

- **AMF - Association des Maires de France**

Page(s) wiki(s) hpedia référençant "AMF - Association des Maires de France"

- **Gourault Jacqueline**
- **Laignel André**
- **Pelissard Jacques**

Page(s) wiki(s) RDpedia annotées "AMF - Association des Maires de France"
Pas de résultat. Souhaitez-vous crér une page RDpedia à ce sujet ?

10 derniers billets annotées "AMF - Association des Maires de France"

- **Bras de fer en vue sur le volet Bâtiment du Grenelle (voir)**
- **nouveau billet (voir)**
- **Association des Maires de France : Quelles conséquences pour les collectivités suite au Grenelle ?**

Another interesting component we can now provide due to the technology is wiki content geolocation. Since we primarily use the Geonames ontology to model location, we can reuse the freely-available data from the Geonames project to build geolocation services at zero cost to the company.

Key Benefits of Using Semantic Web Technologies

- unifying Web 2.0 and the Semantic Web with lightweight ontologies;
- common semantics to model meta-data of existing Web 2.0 services with SIOC;
- advanced and collaborative knowledge modeling using wikis;
- interlinking tags, ontology instances and tagged content with MOAT;
- reusing ontologies and RDF data available on the Web;
- ability to merge and query data from various services using a central storage system;
- ontology based querying;
- suggesting related content thanks to relationships in the ontology;
- evolution of query services thanks to the SPARQL language and protocol;

Finally, one of the most important points of our system is that most of the semantics are hidden from end users who do not need to struggle with complex Semantic Web modeling principles to benefit from the services.

> *The research center of EDF is a very "multicultural" place, from chemists to linguists, from material aging engineers to trading experts, … It may seem weird but these people often have to share information and knowledge. The problem is that they do not belong to the same communities so they do not speak the same language. The Semantic Web is one of the tools we have experienced that creates bridges between communities and it does the job provided that it remains almost invisible thanks to a smart and user friendly interface. I do think that the Semantic Web will be a means to encourage our researchers to share more and more knowledge and that it will be easier and easier to use. We do need this.*
>
> François-Xavier Testard-Vaillant, Senior adviser for corporate collective intelligence, EDF R&D