A Framework for Knowledge Management of I&C using a CMS with Ontology extension

Anne Dourgnon, T. Dang, B Dissert, S Reguigui

A Framework for Knowledge Management of I&C using a CMS with Ontology extension

A. Dourgnon-Hanoune,
T. Dang
EDF Research and Development
Performance Optimization of
Industrial Process Department
6, Quai Watier – BP 49
78401 Chatou Cedex
France
33-13087{8043, 7918}
{anne.dourgnon,
tuan.dang}@edf.fr

B. Dissert
JALIOS
58, rue Pottier
78150 Le Chesnay
France
33-139239281
benoit.dissert@jalios.com

S. Reguigui
ONTOLOGOS Corp.
PAE du Levray
6, route de Nanfray
74960 Cran Gevrier
France
33-450698837
samah.reguigui@ontologos-corp.com

ABSTRACT
In this paper, we describe a framework for knowledge management of Instrumentation and Control (I&C) using a Content Management System (CMS) with ontology extension. Such approach provides an efficient and meaningful search of non-textual and textual information corresponding to a consistent terminology, and a lifecycle management of documents. A innovative architecture of the framework has been developed using web services in order to integrate two Commercial-Off-The-Shelf.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Linguistic processing.*

H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *Information networks, User profiles and alert services.*

H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *Data sharing, Web-based services.*

General Terms
Design, Experimentation, Theory.

Keywords
Instrumentation and Control (I&C); Industrial IT; Knowledge Management; Competence Network; Collaborative Network; Ontology; CMS; COTS; Portal; Web services.

1. INTRODUCTION
Like other firms, which are characterized by their large and complex core business, Electricité de France (EDF) made the choice of collaborative training network where men and women from different horizons join for better working [1].
The Instrumentation and Control (I&C) competence network is one of these collaborative networks. It gathers competences those are responsible for the design and the implementation of instrumentation and control in power plants. The characteristics of our network can be summarized as follows: diversity of business and technological knowledge, diversity of the sources of information.

In this paper, we will describe the framework we use to address the problem of knowledge management of textual and non-textual information using a Content Management System (CMS), which is coupled with an ontology-based search module. Such approach offers on the one hand the power of a CMS to manage, through workflow, content during its entire lifecycle i.e. from creation through publishing, and on the other hand, an efficient (meaningful) search of textual information corresponding to a consistent terminology that represents the concepts used by our I&C competence network.

First, we will describe the functional architecture of our framework where the integration process of two Commercial-Off-The-Shelf (COTS) will be explained. Secondly, we will give examples of knowledge capture and structuring using this framework. Finally, some feedbacks are presented.

2. FUNCTIONAL ARCHITECTURE OF THE FRAMEWORK
The difficulties we encounter when we decide to develop a framework for managing I&C knowledge are double: the diversity of concepts (from level 0 to level 2 of the CIM - Computer Integrated Manufacturing – Model [2]) and the diversity of information sources (textual and non-textual information).

I&C concepts are pretty complex and diverse: at level 0, we have sensors and actuators. Sensors have different families (according to physical measure) and technologies (ex: wire and wireless…), which need to be classified according to different criterion (ex: safety, precision…). At level 1, we have PLC (Programmable Logic Controller) and DCS (Distributed Control System). These hardware and software components are used to acquire and transmit information to components at level 0. On the “top” level (2), we have supervisory hardware and software, which offer a global control of power plant by acquiring and transmitting information to equipments on level 1. Generally speaking, at every level, every business has its particular terminology. One of the challenges is to build a consistent relation between terminologies across different levels. Of course, the diversity of information sources creates another challenge concerning the indexation of embedded information in binary files (pictures, schematics…) or visual presentations.

When designing the framework, our requirements are following:

- Actor, a person of I&C competence network, may participate in several groups. He is called member of the groups and is characterized by his identity number inside the Enterprise, his preferred language, his function and role, his groups, etc.
- A group gathers actors who are interested in collaborating about a given technical field. Two leaders of the group moderate it. The leaders are responsible for the validation (through a workflow cycle) of documents proposed by members of the group.
- A document is a textual information that may be structured or unstructured. It may include non-textual information like picture, schematic or visual presentation. It has one of the following status: *private* (only visible by the owner); *submitted* (only visible by the owner and the leaders of the group to which the document is submitted, it is not modifiable during this stage); *to be discussed* (visible by members of the group and not modifiable); *rejected* (only visible by the owner and modifiable by this one); *published* (visible by members of the group only or by any actor; it depends on the decision of the leaders). When a document is published, it can be outdated or archived automatically according to the choice of the leaders.
- Full text search and voting/survey capability on documents.

To fulfil the above requirements, we define the following functional architecture of our framework (see *Figure 1*):

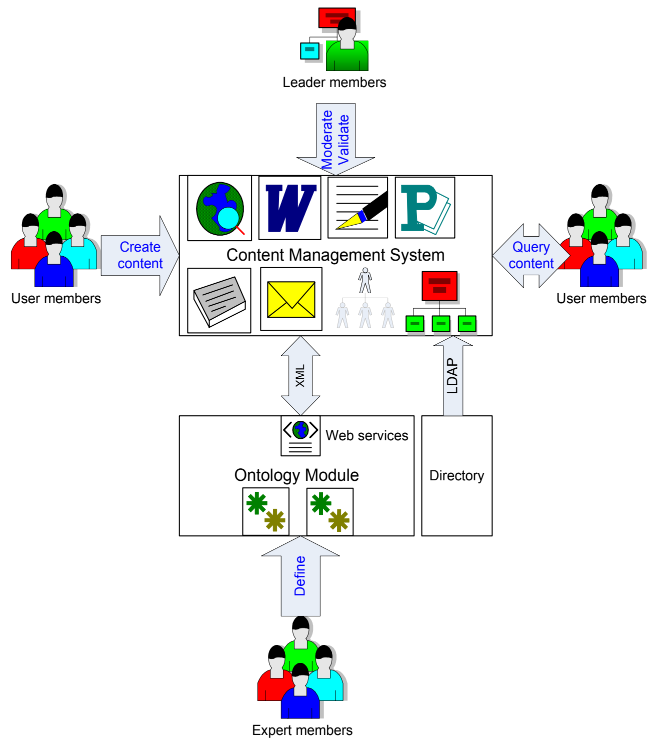

Figure 1: functional architecture of the framework

2.1 Role of the CMS

According to [3], CMS is "the collection of policies and technologies that guide and enable corporates to contribute, manage, and share their structured and/or unstructured information". In fact, CMS is a key point in the implementation process of knowledge management. In recent years, more and more enterprises become aware that their know-how in business and the knowledge of their personnel constitute a value and a capital for them. This is particularly true in industry where long-term knowledge is necessary to maintain operational process. Thus, more and more managers are promoted to titles such as "Chief Knowledge Officer", "Chief Information Officer", etc.

In practical terms, a CMS is a system used to organize and facilitate collaborative content creation. The systems also often include some sort of concept of the workflow for the target users, which defines how the new content is to be routed around the system. Perhaps the best example of such a system is the WikiWiki [4]. Wiki software includes logic that allows authors to edit the content of the site online - this is actually not that common, many systems require the articles to be constructed and edited in some other software. When you have completed your changes, the Wiki software tracks what changes you have made, and updates the page. Wiki also allows many more than one person to edit the target page at the same time, solving simultaneous write conflicts as they arise. There exist a large number of software products claiming to be CMS:

- phpCMS; PHP-Nuke; Plone; SPIP
- Envolution
- Nuxeo Collaborative Portal Server (CPS)

- Typo3…

None of them had a ready tool for customizing the look-and-feel of the user interface without programming effort. In most cases, programming knowledge is required to change the software functionalities and behaviours.

Our choice was based on JCMS from Jalios. In our framework, JCMS provides an integration infrastructure for content and user management, and other services. User's information is retrieved from the Enterprise Directory via LDAP. JCMS offers standard services like workflow and full text search functionality.

2.2 Role of the Ontology

As said earlier, we have chosen to define document as a unifying concept for textual information. Of course, document may contain non-textual information. In order to provide a relevant search of concept that may be found in non-textual information, we need to build a formal terminology that may be exploited by full text search engine (in JCMS in particular). In fact, full text search technique, web semantic or textual analysis do not provide an efficient approach for searching I&C concepts. Indexation by concepts allows retrieving relevant information according to the viewpoint adopted by the users. We observe that people often work and collaborate around the concepts they usually manipulated. This is confirmed by what we could see on the eCots (www.ecots.org) Internet site, where we organize an exchange network on the use of Commercial Off The Shelf (COTS). Such observations make us decide to construct an ontology of COTS [5].

More information about the benefits of the ontology to the end user can be found in [10], where detailed comparison is discussed.

In the I&C domain there are a lot of existing concepts, which can be found in documents and references. These concepts need to be structured and completed with missing and scattered elements in a structure, which is not interpretative. Only a logic structure could answer this requirement. That is why we need to use technique such as ontology as a basis for expressing and structuring concepts. Moreover, the ontology can assist the process of identifying existing requirements and defining a specification for an IT system [6].

2.3 Integration process

As depicted in *Figure 1*, our framework is composed of two main components:

- the CMS is used for enabling users to manage contents, files and concepts;
- the Ontology tools for creating associations on concepts and manage these associations.

We have chosen an architecture and an integration process in which, each of these components can work without the other. Such architecture allows us to minimize the dependency of one component on the other when one of these evolves.

The first difficulty we encountered in the integration process is that the items/objects manipulated in each component are very different. The second one is that we have different programming languages on each of these components that works on different operating system: Java based CMS on Linux and Smalltalk based ontology module on Microsoft Windows.

To overcome the first difficulty, we have chosen to create a wrapper in the ontology module that wraps around the items created in the CMS without changing the identifier of the items used in the CMS. So the ontology tool manages the composite objects, which refer to objects/items in the CMS.

To resolve the second difficulty, we decided to use Web services since the API (Application programming interface) exist in both Java and Smalltalk to call Web services.

The API of the CMS make it possible to invoke a specific Web service for each operation in the CMS (creation, modification, deletion) on each type of item of its data model.

When creating an item in the CMS, the ontology tool is aware that a wrapper has to be created for this new CMS item. The ontology tool administration interface is then used to make possible associations on this wrapper.

Security considerations are also taken into account in our framework. Since user profile and rights are stored in the CMS, a CMS Web service is called by the ontology tool to check the rights of a given member so it presents only the authorized items to him.

Of course, the CMS is used as a knowledge management portal to present all the contents. So the search results generated from the ontology module are shown in the CMS as if it was embedded in the CMS. In fact, the ontology module is called via its Aurele Web services.

The Aurele Web services offer the semantic search engine and also the visualization of the ontology in different forms (treeview, eyetree or a simple list of terms). These Web services also deal with the deletion of the links between documents and concepts within the ontology, which are generated by the semantic annotation.

In the next paragraph, we will describe more about the construction of I&C ontology

3. KNOWLEDGE CAPTURE AND STRUCTURING

As said earlier, there are a lot of concepts in I&C domain. In order to capture these concepts rigorously, we need to build an ontology.

Building ontologies [8] is not a simple task. In order to build good/consistent ontologies we need methods and tools. So the I&C ontology has been developed using the OK Model [9] following the Ousia methodology of the Ontologos Corp. company that is characterized by five main steps. First, we determine the domain and the scope of the ontology. Secondly candidate terms are extracted from a corpus, that is, a set of documents representative of the given domain. As the corpus does not contain all the necessary knowledge for building an ontology (e.g. in general, knowledge on difference that may exist between concepts does not appear in text), we need experts of the domain to achieve consensus. That is why the extracted vocabulary must be validated by experts. It implies that only terms judged as suitable to domain terms will be retained from the generated lexicon. Finally, the validated vocabulary can be used to build the ontology: relevant concepts and relations are identified and organized.

Once the I&C ontology has been built, it is integrated in the Aurele ontology-module of the CMS where users can, among other things, visualize it in different forms as said earlier.

Aurele ontology web services allow users, according to their rights defined in the CMS, to enrich ontology with additional information. For each concept within the ontology, an encyclopaedic definition, a multi-lingual set of synonyms, a list of contacts or comments can be added through the ontology tool.

As soon as we get new document on the CMS, Aurele web services are asked to automatically create a link to this new item from a CMS call. Once it is done, the user can associate manually or automatically this new document to concepts of the ontology (i.e. a semantic annotation).

The I&C ontology is a common multi-lingual vocabulary which allows people to share knowledge. Then, members can submit queries in natural language and use the conceptual structure of the I&C ontology to refine these queries (*Figure 2*). When a term in the query is ambiguous (i.e. when it denotes several concepts), Aurele suggests the user all its corresponding concepts. The results are all the relevant documents, displayed according to the member rights, with their relevancy percentage and with their semantic code.

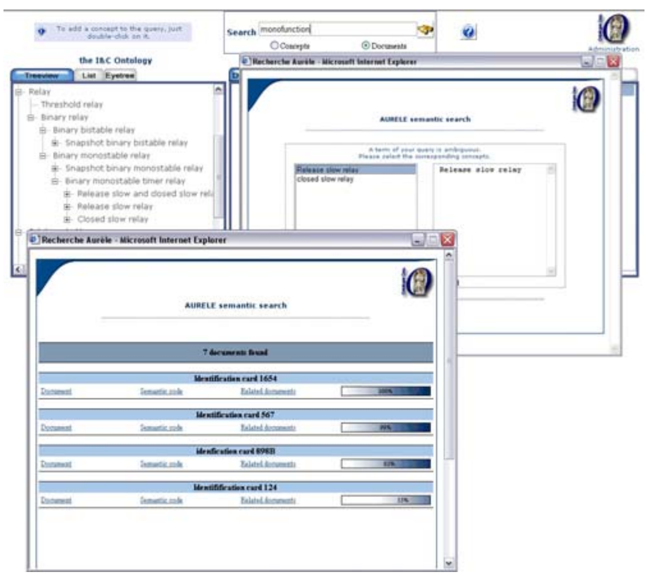

Figure 2: example of Aurele semantic search engine

The formal approach used here has enabled us to obtain a consistant and consensual ontology of level 0 concepts. This ontology is included in a CMS, based on documents and texts, not concepts.

4. CONCLUSION AND FUTURE WORKS

When developing this framework, our motivation was to capitalize the knowledge on the I&C domain. In fact, one of the objectives of our I&C competence network is to preserve the know-how we have acquired on the design, the commissioning and the operational maintenance of the industrial information systems in our power plant. The objectives we pursued were reached in terms of capitalization of documents and concepts in the I&C domain. Indeed, we are currently studying the generalization of such approach to other domains inside the EDF Group.

Our next work will be on levels 1 and 2 : linking the ontology with UML2/SysML description, giving the possibility to access diagram, data and source code, if necessary, directly from the ontology (or with the search engine).
We also plan to integrate this framework in an Enterprise Architecture Framework in order to enhance the description of our business activities and processes. Of course, integration in Business Process Management (BPM) tools is among our future works.

5. ACKNOWLEDGMENTS

We express our gratitude to our EDF engineering colleagues from CIPN, UTO, SEPTEN, CEIDRE and SFP divisions for their fruitful discussion. Many thanks to the SIDONI team: Nadège Vignol and Fabrice Vinet for their support and help in the integration process.

6. REFERENCES

[1] J-F. Ballay, *Firme et individu : qui apprend le plus ?,* L'expansion Management Review, sept. 2004.

[2] Purdue Reference Model for CIM, http://www.pera.net/Pera/PurdueReferenceModel/ReferenceModel.html

[3] Stephan E. Arnold*, Content Management: Role and Reality*, http://www.arnoldit.com/articles/cmkoenigJuly2002.pdf.

[4] *WikiWiki,* http://www.wikimedia.org/.

[5] A. Réquilé-Romanczuk, A. Cechich, A. Dourgnon-Hanoune, J-C. Mielnik*, Towards a Knowledge-Based Framework for COTS Component Identification*, MPEC 2005.

[6] M. Uschold and R.. Jasper, *A Framework for Understanding and Classifying Ontology Applications*, IJCAI99 Workshop on Ontologies and Problem-Solving Methods, Stockholm, Sweden, August 1999.

[7] A. Dourgnon-Hanoune, E. Mercier-Laurent, Ch. Roche, *How to value and transmit nuclear industry long term Knowledge*, ICEIS 2005.

[8] C.Roche, *Ontology : a Survey*, 8th Symposium on Automated Systems Based on Human Skill and Knowledge. IFAC. September 22-24, 2003, Göteborg, Sweden.

[9] C.Roche, *The 'Specific-Difference' Principle : a Methodology for Building Consensual and Coherent Ontologies,* IC-AI'2001 2000 : Las Vegas, USA, June 25-28 2001.

[10] A. Dourgnon-Hanoune, P. Salaün, C. Roche, *Ontology for long-term knowledge*, IEA/AIE'06 : Annecy, France, June 27-30 2006