Proceedings of the World Congress on Engineering and Computer Science 2012 Vol II
WCECS 2012, October 24-26, 2012, San Francisco, USA

A Topic-map-based Framework for Resource Retrieval in an Industrial Context STMicroelectronics' Case Study

Sara Bouzid, Corine Cauvet and Jacques Pinaton

***Abstract*—With the increase of company resources for data processing, calculation and reporting for the control of business processes, retrieving these resources has become a real challenge in these companies. Existing retrieval systems don't enable to express the relation between company resources and the business activities and processes that use them. We propose in this paper a Topic-Map-based framework for referencing and retrieving a set of software resources used to control company processes. The characteristic of this approach is that it proposes a goal-oriented description of resources and layered knowledge maps which expose the link between resources and their business usage. In this way, Topic Maps enable to guide resource retrieval in a business context. This framework is being experimented on a repository of resources of the Manufacturing Process Control System within the STMicroelectronics Company.**

***Keywords:* Topic Maps, Process Control System, Business semantic, Driven search**

1. INTRODUCTION

A Process Control System (PCS) in a company regroups a set of resources (data reports, web resources, documents, etc.) which provide aggregate of measures -i.e. indicators- based on standard process control methods (SPC, APC, FDC, etc.)[1]. These resources enable the companies to monitor their activities, identify and analyze business tendencies and ensure process improvement. The resulting resources contain data related to the business process, processed with kinds of decision-support systems (figure 1).

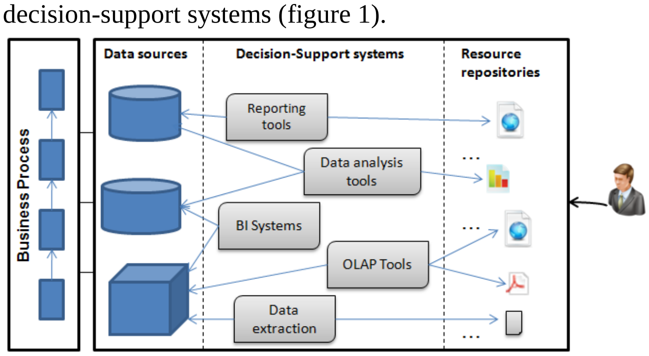

Figure 1. The resources of a PCS with their data sources

Sara Bouzid is with the Laboratory for Systems and Information Sciences (LSIS), University of Aix-Marseille, Marseilles, France (sara.bouzid@lsis.org; sara.bouzid@st.com).
Corine Cauvet is with the LSIS, University of Aix-Marseille, Marseilles, France (corine.cauvet@lsis.org)
Jacques Pinaton is with the STMicroelectronics Company, Rousset area, France (jacques.pinaton@st.com)

The main advantage of a PCS is that it centralizes the access to the data sources of a company. It enables then the end-users to use same data to produce reliable business indicators. However, this kind of system often entails a significant increase of the produced resources that provides these indicators, especially if there is no approach in the company to manage the resources. As a consequence, users can spend a lot of time in searching for the resources that can respond to their needs or they can instead re-develop same resources. An efficient PCS must enable the reuse of same resources by the users that have same needs in a business domain. The main difficulty denoted here is the lack of business knowledge in these resources. In fact, the resources of a PCS enable to respond to business objectives and the users use business knowledge to retrieve this kind of resources. However, the PCS resources contain data depicted in graphs, histograms, flowcharts, etc. These data are interpreted by the user as business knowledge but this interpretation is not explicit to any search system because it is held by the user only. Actually, the study of a real PCS implemented in a company like STMicroelectronics shows us clearly that there is significant distance between the resources of a PCS and how a user perceives his need.

Therefore, one of the challenges in establishing a search system in this context is to enable to bridge the gap between the resources and their business usage. Accordingly, we address in this paper two research questions:

- How to organize resources' semantics to improve their findability in an industrial context?
- and how to assist the user in resource retrieval starting from a business need?

To that aim, we propose in this paper a software framework based on the Topic Map standard. This framework provides a Topic-Map-Oriented knowledge-base for referencing these resources and a search system based on knowledge-maps visualization. The main outlines of this work have been set with the study of the PCS of the STMicroelectronics Company.

The second section in this paper presents the context of work. The third section tackles the Topic-Map-based framework and explains how it can respond to the identified problematic. The fourth section presents the search system for resource retrieval with some technical aspects. Finally, the last section presents some related works.

ISBN: 978-988-19252-4-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

2. THE STMICROELECTRONICS' CONTEXT

STMicroelectronics is a French-Italian company specialized in the manufacturing of electronic chips. The process control in the company ensures that at any time front-end manufacturing processes and equipments meet the requirements for products. By studying the PCS of this company, we noticed many difficulties related to the retrieval of its resources and their use.

Lack of semantics:

Most of PCS resources contain figures (flow charts, histograms, etc.) and numerical data which are understandable by human users only. The lack of business semantics in these resources makes difficult their findability. In addition, STMicroelectronics has several standard management systems for referencing and retrieving diverse types of resources, but most of these systems fail in finding the PCS resources. In fact, the weakness of such systems is the type of the meta-data used to describe the usage of the resources. We noticed that if these meta-data don't reflect the business need of the user, this one fails to find the appropriate resource to his need. We did an experience on a panel of users within STMicroelectronics to confirm these conclusions. We proposed to a panel of 40 users which have different business functions to try to find three different resources. These resources corresponded to three basic needs required for the core business activity. The results show that only 3 users were able to find one appropriate resource among the three requested ones, 10 users found many duplicated resources and were hesitant in their choice and the rest of the users didn't find them or found non-appropriate resources.

Problem of resource heterogeneity:

We also denoted that an indicator provided by a resource can respond to several business purposes depending on the existing business functions in the company. This second problem explains the difficulty of the users to find the resources. Actually, the STMicroelectronics core activity is organized into business areas. Each area is specified in a step of the manufacturing process. We can find until 35 areas in one STMicroelectronics' plant. Furthermore, there are several decision-support systems in the company which can create the same resources. We found within STMicroelectronics around 20 kinds of decision-support systems for the core business. These systems coexist mainly because they were purchased to satisfy particular needs over time. Currently, there is an average of 100 to 300 created resources by each system, which means that we can find more than 100 000 resources related to the PCS of one STMicroelectronics' plant. Note that this number doesn't reflect the quantity of needs because there isn't one resource for one single need.

Lack of a structuring approach:

STMicroelectronics tries to organize the access to the PCS resources by business area through a set of web pages. This approach improved the reuse of existing resources by the users, mainly because the resources were referenced with business semantic. Nevertheless, there was a lack of an approach for structuring this semantic.

Finally, these difficulties lead us to think about a new approach to support the integration and the structuring of resource semantics on one hand, and the process of resource retrieval by a user on other hand. Thus, we choose to use the Topic Map standard because of its suitability in managing many types of resources and knowledge with high expressiveness.

3. THE TOPIC-MAP-BASED FRAMEWORK

A. *The Topic Map standard*

The Topic Map standard is an ISO semantic web standard usually used to improve information retrieval in the web. The key concepts of a Topic Map consist of topics, associations, occurrences and resources (figure 2).

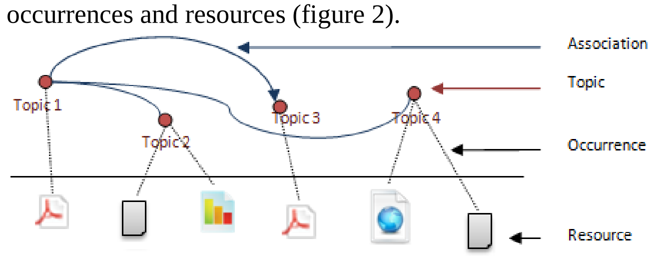

Figure 2. Core concepts of the Topic Map standard

A topic is a symbolic representation of a subject where a subject is a concept from a real world [2]. An association expresses a relationship between topics. An occurrence is what links an information resource to a topic and finally a resource is any technological support that handles information. It could be a document, a web page, software, a Database, etc. We chose to use the Topic Map standard for several reasons. A Topic-Map can represent any subject from the real world with any desired level of granularity by typing the topics, the associations and the occurrences and regardless of resource type and location. Furthermore, Topic Maps are highly oriented towards human users comparing to the W3C standards such as OWL and RDF [3]. In fact, Topic Maps are optimized for findability [4][5]. The key concepts of the Topic Map paradigm enable to organize the way of navigation among resources so to effectively improve the retrieval of these resources by human users. Finally, one other advantage of using this standard is that there are many tool-support that deal with Topic Maps. According to our purpose, we use the main concepts of the Topic Map standard to ensure the integration and the structuring of resource semantics.

B. *Approach overview*

Our approach deals with three main requirements to address the research problematic:

- It identifies the types of knowledge to semantically describe a set of resources in an industrial context

ISBN: 978-988-19252-4-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

- It provides a methodological framework to capture and structure the identified knowledge with resources
- It provides a software framework to assist a user in the resource-retrieval process

Knowledge typology:

We can distinguish three types of knowledge for the description of process-control resources:

- ***The why** of the user need*: the resources of a PCS seek to support business needs achievement, it is then important to expose why they were created. The why must handle any information related to the business needs of the user such as goals, tasks, business activities and so on. Thus, using knowledge related to the why of the resources according to a business point of view will make significant the resources' usage to the user.

- ***The what** of the user need*: a user has to know what a resource provides as functionalities, process control methods, outputs and so forth. Any description related to the services provided by a resource can be decisive in choosing the most appropriate resource to a business need.

- ***The how** of the user need*: this information can handle any description related to the decision-support applications that provides the resources. In this way a user can have a whole description of his need and how it can be realized.

The methodological framework for knowledge capture:

The methodological framework aims at identifying and gathering knowledge that will constitute the resources' semantic. The main outlines of this approach are depicted in figure 3 and slightly described in [6].

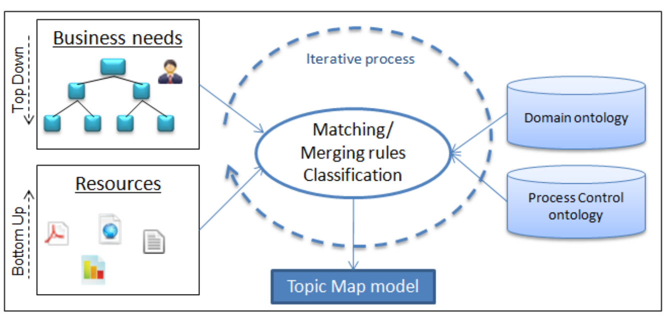

Figure 3. Overview of the approach

The business needs of the users are captured using a requirement model. This model is based on a goal decomposition mechanism. We extract on other hand the resources' meta-data. Afterward, the matching process between these two steps is enhanced with a process control ontology and a domain ontology. For example, in the case of STMicroelectronics, the domain ontology is related to the business process to control, i.e the manufacturing process. These ontologies are used to find implicit relations between business concepts and the resources' meta-data. This approach is being implemented within STMicroelectronics step by step, especially that it is difficult to gather large amount of needs in one step in a big company like STMicroelectronics. At the end of this process, a Topic-Map model is created referencing the PCS resources with their business usage. This model is a knowledge base that can be stored in different formats (flat files, databases, etc.).

4. SOFTWARE FRAMEWORK DESCRIPTION

The proposed framework is based on a semantic interface which guides a user in resource retrieval through a layered Topic Map. This Topic Map is created dynamically through a search application. The used semantic is provided by a Topic-Map-Oriented knowledge base (figure 4).

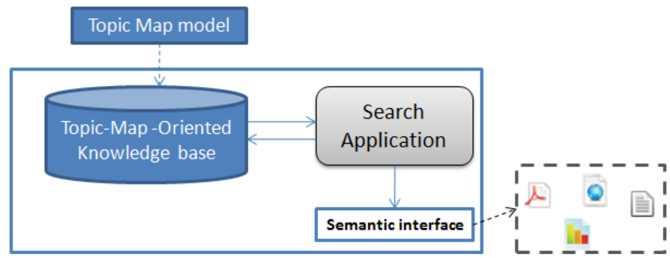

Figure 4. The software framework

A. *The semantic interface*

We propose to structure the captured semantic in three layers: goal layer, service layer and resource layer. These layers correspond to the types of knowledge introduced in the approach overview. The resulting layered structure is a knowledge map dynamically created to respond to a user need. In fact, this semantic will be exposed to the user in a way to guide him in the discovery and retrieval of existing resources. As a result, the user will access to the resources through a semantic interface as we can see in figure 5.

The *goal layer* regroups the set of topics that capture the business needs of the user. In fact, our experience within STMicroelectronics reveals us that the major difficulty in existing resources' description is the gap between how the users perceive their needs and how they describe them. Using a goal orientation as a starting point in resource description seeks to fill this gap on one hand and to guide the user in choosing a resource on other hand. Moreover, a need can be complex; hence a refinement mechanism is used to refine it until the service layer.

The *service layer* regroups the topics that capture the services provided by the resources. We note that the term service is used here to refer in general to any information related to what a resource provides as functions, functionalities, methods and so forth. In the PCS context, the services would be the produced indicators and the used business methods for data processing and calculation.

The *resource layer* exposes some useful meta-data about the physical resources such as the used decision-support systems, the used data sources, the type of resources, etc.

To sum up, the defined layers aim at clustering topics of same type in a logical structure. This structure is a Topic Map created dynamically when a user search for a need. It provides the useful navigation that will help the user in discovering and understanding the existing resources. The user can see then if there are resources that meet his need and can choose in this way the most appropriate one to his need.

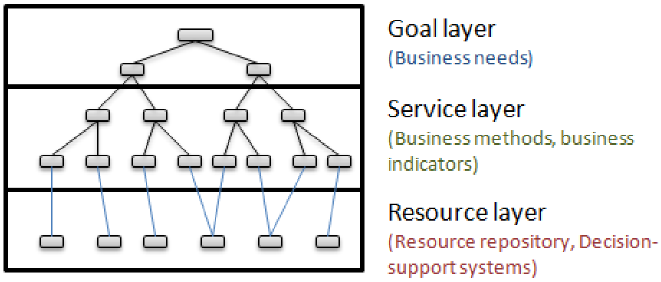

Figure 5. The semantic interface for the PCS resources

Note that Topic Maps are used here in a context different from their usual usage. Topic Maps are usually used to improve information retrieval by exposing the subjects contained in resources. In our context of work, Topic Maps are not content-oriented, but rather, knowledge-oriented.

B. The Topic-Map-oriented knowledge base

Figure 6 depicts the meta-model of the Topic-Map model used to implement the knowledge base of the framework. This meta-model relies on the basic concepts of the standard Topic-Map model, known as the TAO model (Topic, Association, Occurrence) [7]. We integrate in this model two more concepts related to this standard: the resources and the variants. A resource is any addressable physical object. A variant refers to any variant of a topic name. It can be used to represent alternative orthographies, synonymies, transliterations, abbreviations and so on [2].

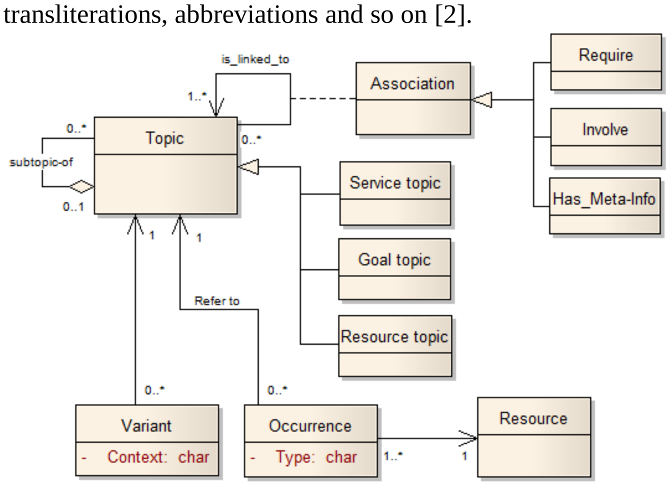

Figure 6. The Topic-Map Meta-Model

The proposed Topic-Map model has been designed in a way to enable high expressiveness of a business context with flexibility in use and easiness in maintenance. As we can see in figure 6, this meta-model allows referencing resources to one or many topics through the occurrence concept.

The Topic regroups all types of topics that can represent a subject of a business context. In addition, each topic can be a sub-topic of another topic.

The Association refers to a semantic relation between two topics. The interesting aspect of using the associations is in typing them following the meaning we want to infer. Concretely, the associations enable to expand a hierarchy of topics into complex topic networks.

As we introduced before, the variant represents a variant name of a topic name so to express several rewrites or meanings of the same topic. This concept responds at first to terminological and grammatical problems related to a topic. We add to this concept an optional property called context. The context links a variant to the business context in which it is used. It could be a business function, a business area, a domain, a business profile and so on. The aim of adding this property is to respond as best as possible to users' needs regardless of the used terminologies.

If we refer to our Topic Map meta-model (figure 6), the resource layer of the semantic interface will correspond to the resource class linked to the occurrence class. The association class will link the goal layer to the service layer. Each layer corresponds to a set of topics of same type (i.e goal topics and service topics). We actually defined three types of associations for the STMicroelectronics case study:

- *Require (G1, G2)*: means that the goal G1 requires the goal G2 to be achieved (enables to link goals that do not belong to same goal hierarchies)
- *Involve (G, S || S, S)*: this association is used to link goals with services or services with services.
- *Has_meta-info (S, R)*: this association is used to link a service topic to a resource topic

As a result, all the associations in the goal layer are typed as *Require*, the associations of the service layer are typed as *Involve*, and the associations of the resource layer are types as *Has_meta-info*.

We choose to implement the knowledge base in a relational database to respond to performance issues in case of large volume of data within STMicroelectronics. We note that in the first step of this work, the approach has been experimented on a small repository of resources identified with a group of expert users within STMicroelectronics. The semantic interface of the framework is implemented with a web portal. This portal displays the layered Topic Maps that meet users' needs. The programming is done with the Php technology. The Graphviz[1] API has been plugged in the Php application to allow the visualization of the knowledge maps.

C. The Topic-Map-driven search

The search application:

The search application is based on Topic-Maps display. The user here expresses his request with one or a set of key words. Figure 7 shows an example of a user request. We take the example of a user in charge of controlling excursions within STMicroelectronics in order to improve the *Wafer Fab Yield (WFY)*. An excursion is an important accident that happens to a product during the manufacturing process. STMicroelectronics' product consists of one or some lots where each lot contains a set of wafers. A wafer is a silicon plate used as support for the construction of electronic chips.

[1] http://www.graphviz.org/

ISBN: 978-988-19252-4-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

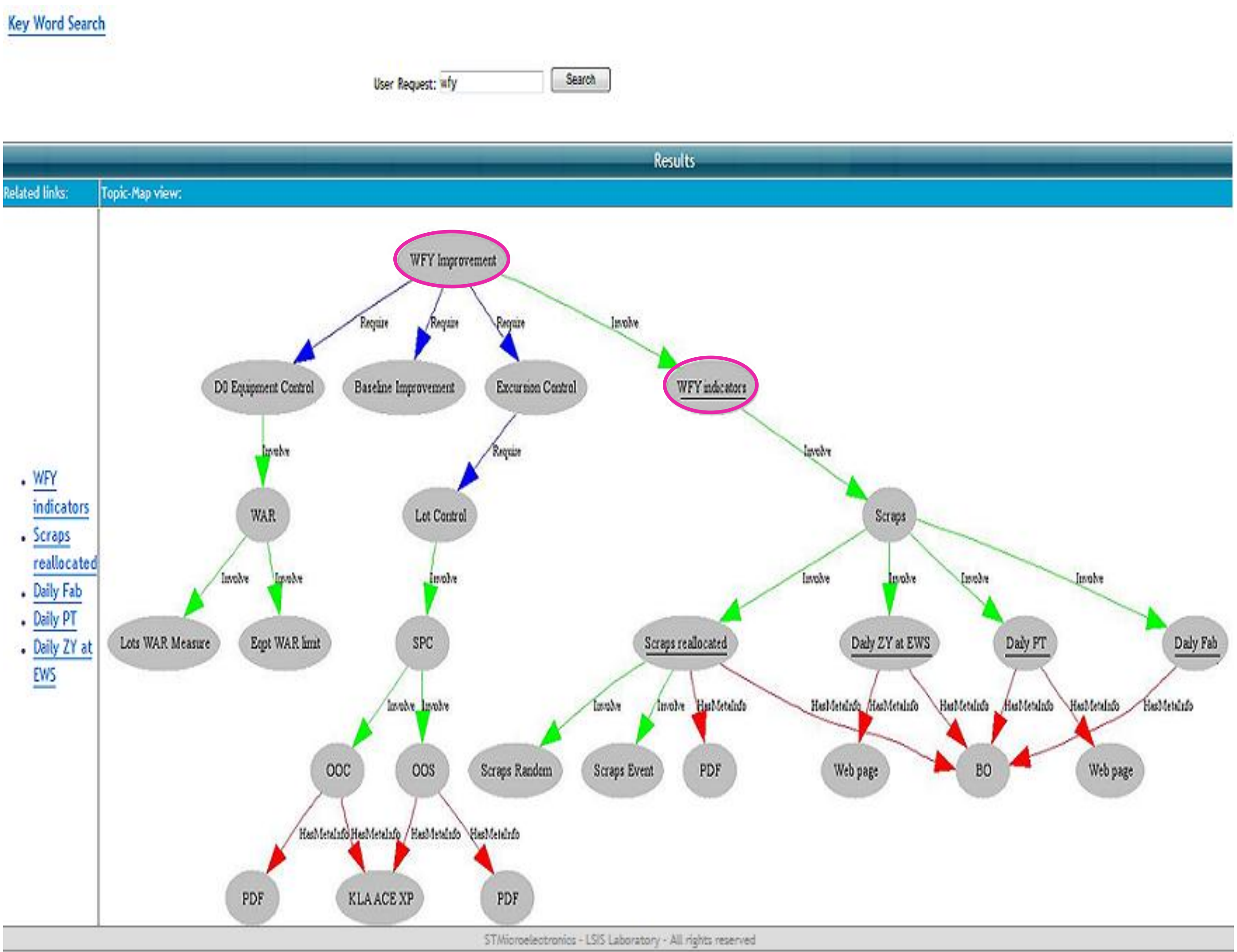

Figure 7. The Topic Map of the WFY example

The *WFY* refers to the manufactured quantity of wafers. We suppose that the user needs a specific indicator about equipments in relation with his business scope. Hence, by typing *WFY improvement* in the search system, the system returns all the topics that contain at least this word with all its related topics. In this case, the resulting Topic Map shows that the indicators related to equipments and close to the user assignment are the ones about *D0 Equipment Control*, because this topic refers to *equipments*. The user can then choose an indicator among the ones related to this topic.

As we can see in figure 7, each topic found in the hierarchy is displayed with its sub-hierarchy until the end of the tree. The resources are displayed both in the graph and in the left frame of the interface to allow the user a quick access when he knows exactly the resources he needs. In the other case, the created Topic Map seeks to guide him in choosing a resource. In fact, for each expressed need by the user, the system creates a specific Topic Map with the relevant topics to his need. In addition, the goal layer plays a major role in this process because it represents the starting point in the map. It places the user in a real business description; the user can then understand why a given resource is related to his need.

The process of Topic-Maps creation:

The basis of the search algorithm relies on nested recursive algorithms starting from the first topic found with the user query. When a user expresses a need, the system checks at first in the Topic table if there are concepts that matches. Otherwise it will checks in the Variant table. According to this step, each topic that matches with the user query represents a starting point for the creation of a Topic Map. Hence, for each topic found, the system selects its direct sub topics and its related associations with the two types "*Require*" and "*Involve*". This process is repeated while each ongoing topic has sub topics (recursive selection). The pseudo code of this process can be summarized as follow:

```
Foreach (topic (i))
    If (topic (i)) hasChild())
        While hasChild (topic(i))
            Select subTopics (topic(i))
            Select topicsAssociated(topic(i), "Require")
            Select topicsAssociated(topic(i), "Involve")
```

```
            End while
      End if
End for
```

The function hasChild() checks if each topic found is a parent topic so to select its sub topics. Afterwards, the system tries to find for each topic its related associations of type *Require* before the associations of type *Involve* so to create the goal layer before the service layer. For each association found between two topics, the hierarchy of the target topic in this association is selected. Thus, for each target topic in an association the system must also retrieve its hierarchy in a recursive way. The pseudo code of the function *topicsAssociated()* for the construction of the associations around a topic is as follow:

```
Foreach (topic(i) In association (a))
      If (topic(i) hasChild())
            While hasChild (topic(i))
                        Select subTopics(topic(i))
                        Select topicsAssociated(topic(i),a))
                        Select occurrences(topic(i))
            End while
      End if
End for
```

To sum up, the proposed search process retrieves the topics, their associated topics and their occurrences according to each topic found in a selected hierarchy. Note that the integration of the variant topics is done either in the beginning of the search algorithm to improve the matching process, or in the end at the display of the topic Map in order to create the appropriate map for a given business profile. As a result, this process tries to respond as best as possible to a given business need of a user.

5. RELATED WORKS

Topic Maps mainly proved their usefulness in e-learning applications where they are used to improve navigation in documents and courseware [8] [9] [10] [11] [12]. On the other hand, we can notice that Topic Maps aims at exposing the subjects contained in resources. In industrial contexts, exposing resources' subjects is not really useful to business experts, because a resource in this context is mainly created to respond to a business need. The used techniques in e-learning provide content-oriented search with Topic Maps whereas the PCS resources are intended to satisfy business needs. There are few recent approaches that try to tackle the use of Topic Maps in industrial contexts such as the hypertopic approach [13] and the ETM approach [14]. However, none of existing approaches has yet proposed a goal orientation as a starting point in a Topic Map to guide a user in resource retrieval.

6. CONCLUSION

We propose in this paper a novel framework to improve resource retrieval in an industrial context. This framework allows creating layered Topic Maps according to a business need. The resulting Topic Maps provide meaningful semantics to the PCS resources of a company. They also guide the user in discovering these resources. Furthermore, the main asset of this framework resides in the genericity of the Topic-Map-based solution which can be implemented for any business context.

Current works tries to assist the user in formulating his need so to obtain lower and precise results in the resulting Topic Maps. In future work, we are trying to experiment the framework with its approach on a large volume of resources. This next step will help to detect potential problems related to the performance of the search algorithm or the visualization of Topic Maps.

REFERENCES

[1] D. . Wheeler and D. . Chambers, *Understanding Statistical Process Control*. p. 1992.
[2] S. Pepper, "Topic Maps," *Encyclopedia of Library and Information sciences*. 2010.
[3] Y. Ding, M. Stollberg, and D. Fensel, "SEMANTIC WEB LANGUAGES – STRENGTHS AND WEAKNESS," *Language*, 2001.
[4] S. Pepper, "Topic Maps and All That." [Online]. Available: http://topicmaps.wordpress.com/2008/05/11/topic-maps-and-the-semantic-web/.
[5] S. Bouzid, C. Cauvet, and J. Pinaton, "A Survey of Semantic Web Standards to Representing Knowledge in Problem Solving Situations," in *CAMP12*, 2012.
[6] S. Bouzid, C. Cauvet, and J. Pinaton, "A Topic-Map-Based Framework for Decision Information Systems," in *Infomation Knowledge Engineering*, 2012.
[7] R. Kannan, "Topic Map : An Ontology Framework for Information Retrieval," *Knowledge Management*, 2010.
[8] C. Dichev and D. Dicheva, "Contexts as Abstraction of Grouping," *American Association for Artificial Intelligence*, 2005.
[9] S. Lavik and T. W. Nordeng, "Brainbank learning – building topic maps-based e-portfolios," in *XML 2004 Proceedings*, 2004.
[10] C. Dichev, D. Dicheva, and L. Aroyo, "USING TOPIC MAPS FOR E-LEARNING," in *Proceedings of the Latest International conference On Computers and Advanced Technology*, 2003.
[11] B. Sridharan, H. Deng, B. Corbitt, and B. Information, "An Ontology-Driven Topic Mapping Approach To Multi-Level Management Of E-Learning Resources An Ontology-Driven Topic Mapping Approach To Multi-Level Management Of E-Learning Resources," in *Europeen Conference on Information System (ECIS)*, 2009.
[12] M. Ouziri, "Semantic integration of Web-based learning resources : A Topic Maps-based approach," *Learning*, pp. 0-4, 2006.
[13] L. H. Zaher, J.-pierre Cahier, M. Zacklad, and I. C. Delaunay, "The Agoræ / Hypertopic approach," in *Indexing and knowledge in Human Sciences (Sdc)*, 2006.
[14] H. Lu and B. Feng, "An Extended Topic Map-based Distributed Knowledge System," *Information Systems*, vol. 5, pp. 1621-1629, 2010.