A Topic-Map-Based Framework for Decision Information Systems

STMicroelectronics’ case study

S. Bouzid\textsuperscript{1,2}, C. Cauvet\textsuperscript{1}, and J. Pinaton\textsuperscript{2}
\textsuperscript{1}Laboratory for Systems and Information Sciences (LSIS), Marseille, France
\textsuperscript{2}STMicroelectronics, Rousset, France

Abstract - With the increase of company components for data processing, calculation and reporting –known as Decision Information Systems (DIS), many companies seek today to enhance the sharing and the retrieval of such components among end-users to have effective returns on investment. The specificity of these components is that they highly support business requirements in a company with a set of indicators. We propose in this paper a methodological framework based on the Topic Map standard to integrate business semantic in a DIS. The aim of this work is to improve the sharing and the retrieval of DIS components among end-users in a company. The characteristic of this framework is that it provides a requirement-oriented description to DIS components and a methodology to integrate this semantic in an iterative approach.

Keywords: Decision Information System; Semantic Description; Topic Maps; Business Requirements

1 Introduction

A Decision Information System (DIS) is a component of Information System aiming at processing large amount of data in order to identify business tendencies and to support business-processes’ performance. The components of such a system consist of a set of data sources like company data bases and data warehouses supported with a set of decision-support applications for data collection, transformation and reporting. These components are intended for different profiles of end-users (i.e. managers, executives and operatives) enabling them to produce and retrieve aggregate of measures called indicators (figure 1).

DIS have acquired a wide importance in the business world nowadays because they are involved in many business purposes. Examples include the control of business processes such as manufacturing processes, healthcare processes, the management control, etc. The key success of establishing DIS in companies consists not only in how to design such systems according to business expectations, but also in how to efficiently use them by the end-users. In fact, many companies adopt DIS to support their core business for two purposes: (a) to enable end-users to use same data and consistent data sources to produce accurate indicators, and (b) to enable the sharing and the retrieval of these indicators by end-users so to avoid redundancies and the waste of time in the development of same needs. However, the second purpose is not really enough tackled by researchers comparing to the first one. Most research works devoted to DIS focus on approaches to design data-sources models and decision-support components [1] [2] [3] [4]. Even if some of these approaches are successful [5], users still complain about mismatching with their needs mainly because there are heterogeneous DIS components in the company and users don’t know how to retrieve existing indicators. Indeed, an indicator can be created with different types of components such as by querying with OLAP tools and BI systems or by using complex components for statistical data analysis. It is important that the users who have same needs can share and reuse same components. This process needs to be efficiently handled.

To that aim, we propose to use semantic description techniques, widely used by the scientific community to address sharing and retrieval issues. The challenge in using such techniques is to identify what kind of semantic knowledge must be represented and how to implement it in an existing system. We propose in this paper a methodological framework to integrate business semantic in a DIS using the Topic Map standard which is an ISO semantic web standard. This semantic allows retrieving DIS components by the end-users according to their business requirements. The characteristic of this framework is that (i) it proposes a requirement-oriented knowledge-map and (ii) provides a methodology to match business requirements with existing components of a DIS. This methodology is supported with a case study related to the control of the manufacturing process within STMicroelectronics. A DIS is

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Components_of_a_DIS.png}
\caption{Components of a DIS}
\end{figure}
implemented in this company to support the control of its manufacturing process with a set of indicators.

The second section in this paper presents briefly some related works. The third section presents the Topic-Map architecture of the proposed framework. The fourth section presents in detail the methodological framework with the STMicroelectronics’ case study. Finally, we present in the last section the resulting Topic-Map meta-model for DIS.

2 Related work

Semantic-description techniques consist of a set of tools and methods for knowledge representation using ontology paradigm and semantic web technologies. Two main categories of technologies are used: the W3C standards such as OWL and RDF, and the ISO Semantic Web standards such as Topic Maps. These technologies are used in many disciplines to address sharing and retrieval issues.

In software engineering, the semantic description is used to enhance the retrieval of software components between application developers. Among the existing works, [6] propose an ontology-driven paradigm for component representation and retrieval. The authors propose a Component Ontology composed of five facets: Provider, Environment, Application Domain, Function and Interface. These facets can be adapted to new user requirements. They also define a retrieving algorithm based on ontology query and reasoning to enable a better syntactic and semantic matching between user requirements and component description. In [7], the authors propose a similar idea based on a description model for component ontology using OWL. This model classifies components’ entities according to function entities and environment entities. Otherwise, in [8], the authors propose an RDF descriptor to capture the components’ meta-data at a semantic and a syntactic level. They use an ontology model that includes source-code ontology, component ontology and a domain-specific ontology. What we can point out from these techniques is that the semantic description focuses on the functionalities and the services provided by the components. In addition, the used semantic description is flat and only provides a classification of components according to their meta-data. In the service-oriented approaches which are a related field, semantic services are proposed to add semantic to software and web services. The DAML-S technology [9] created for this purpose enables a less flat semantic description to services but this description remains also more related to a functional usage than to a business usage.

In information retrieval as well as in resources’ retrieval on the web, semantic-description techniques are usually used to expose the subjects contained in the resources. Main existing approaches aim at classifying resources in grouped subjects [10] [11] [12], besides that few of them propose methodological approaches. For example, in the HyperTopic approach [13], the authors use the concept of point-of-view and entity to improve the access to resources. The point-of-view corresponds to a vision of a user or a group of users related to a requirement. The entity refers to the main subject of a resource. A collaborative-building technique is used to reference all used resources by the users. In [12], the authors propose a layered framework extended from the Topic-Map paradigm to organize knowledge with resources on four levels: the cluster level (provides the effective navigation and browsing mechanism for end-users), the topic level (provides the main subjects contained in the resources), the knowledge-element level (provides more detailed knowledge information) and the resources level related to the physical resources. The construction of the framework is supported with a merging mechanism between these levels. In [14], the authors propose an ontology-based framework to enhance the retrieval of information resources. They propose a domain ontology to reference main concepts of a domain using RDF or OWL and with the help of domain experts. They also extract resources’ meta-data which contain several concepts of the domain. The system tries then to match these meta-data with the concepts of the domain ontology using a degree of mutuality.

Thus, even if a large amount of techniques and approaches exists to support the semantic-description field, they are not really adapted to the context of DIS for two main reasons. The first one is that existing techniques provide functional descriptions or even content-oriented descriptions whereas a DIS is intended to satisfy business requirements. The second one is that most techniques do not provide a methodological support to capture and structure semantic description of components. According to these findings we propose a methodological framework to build and integrate semantic description in a DIS using the Topic Map standard. This framework is based on a requirement-oriented description for DIS components on one hand, and on a methodology to integrate the semantic description in such a system on the other hand.

3 The topic-map architecture

As introduced in section I, we use the Topic Map standard, to technically support the integration of the semantic description in a DIS. The Topic-Map concepts are used in a layered description to structure different knowledge related to both the business domain and the DIS.

3.1 The Topic Map standard

Topic Maps are an ISO semantic web standard usually used to improve information retrieval and the navigation in web resources. The key concepts of a Topic Map consist of topics, associations, occurrences and resources (figure 2).
A topic is a symbolic representation of a subject where a subject is a concept from a real world [15]. 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. The main concepts provided by the Topic-Map Meta-Model enable to describe heterogeneous resources with high semantic abstraction and regardless of their type and their location. A Topic Map can therefore represent any subject from the real world with any desired level of granularity by typing topics and associations. Furthermore, one main characteristic of Topic Maps is that they are highly oriented towards human users comparing to the W3C standards such as OWL and RDF. These ones are more devoted to machines and interoperability between applications [16] [17]. In fact, Topic Maps are optimized for findability. The key concepts of the Topic Map paradigm enable to organize the way of navigation among resources and to effectively improve the retrieval of these resources by human users. In addition, Topic Maps are supported with many tools for topics’ structuring and search.

According to our purpose, a Topic Map is used to integrate a layered semantic description in a DIS starting from business requirements and using typed topics and associations. The retrieval of such components by end-users is then driven by this description.

The main asset of the framework focuses on this layer because it gives a requirement-oriented description for a DIS. In fact a DIS provides a set of indicators and metrics to satisfy requirements in a business domain. Therefore, the semantic description must be enhanced with business requirements so to enable to better meet users’ needs in terms of indicators. The topics in the requirement layer are then typed as requirement topics. In addition a requirement related to a business domain can be complex; hence a refinement mechanism will be used to refine a requirement until the indicators definition. Afterwards, each indicator can be provided by one or several decision-support components. The service layer with the resource layer aims at describing these components so to meet the users’ requirements.

### 3.2 The Layered Topic Map for DIS

The Topic Map standard is used in our framework to provide a semantic description for the components of a DIS. We mainly use the concepts of topic and association to handle this semantic. We also use the concept of resource to represent the components that produce the indicators. At last, we consider that this semantic constitutes a part of the DIS because it normally must be integrated in such a system during its design and implementation.

Moreover, in current implemented systems, a user must know the services provided by DIS components to obtain the needed indicators. By using the semantic description, the main access entry to a DIS will be the semantic interface which results from the methodological framework. Therefore, the user can retrieve an indicator without necessarily knowing the components that provide it. As depicted in figure 3, the Topic Map constitutes the semantic interface for the access to a DIS. The Topic Map is organized in three layers: the requirement layer, the service layer and the resource layer.

- **The requirement layer:**
  The main asset of the framework focuses on this layer because it gives a requirement-oriented description for a DIS. In fact a DIS provides a set of indicators and metrics to satisfy requirements in a business domain. Therefore, the semantic description must be enhanced with business requirements so to enable to better meet users’ needs in terms of indicators. The topics in the requirement layer are then typed as requirement topics. In addition a requirement related to a business domain can be complex; hence a refinement mechanism will be used to refine a requirement until the indicators definition. Afterwards, each indicator can be provided by one or several decision-support components.

- **The service layer:**
  The service layer aims at exposing the main services provided by a DIS. As introduced in the beginning of this paper, a DIS aims at processing large amount of data to produce aggregate of measures. Thus, a DIS provides a set of outputs which are the indicators sought by the end-users. The used topics in this layer are typed as service topics.

- **The resource layer:**
  This layer enables the real linking to DIS components. The concepts of occurrence and resource of the Topic Map paradigm are used for this purpose. A resource represents a component from the DIS. The occurrence links a resource to a service topic with a web address or a physical address (URL, DNS, etc.). In addition an occurrence can be typed according to the types of DIS components.

![Figure 3. The DIS with the layered Topic Map](image-url)
4 The methodological framework

4.1 Overview

Our methodological framework provides an environment to build and integrate semantic description in a DIS using the Topic Map standard. We use a meet-in-the-middle approach which combines a bottom-up building technique of the topic map (i.e. from DIS components) and a top-down building technique (i.e. from business requirements). The main outlines of our methodology are summarized in figure 4.

In the first step, the structure of the indicators is formalized and referenced in an indicator ontology. The indicators’ structure is retrieved from the DIS of the company and is used afterward as support for the construction of the service layer. The construction of the requirement layer is supported with a requirement elicitation step basing on a requirement meta-model. After each requirement elicitation, the requirement model is transformed into requirement topics in the requirement layer. If same requirements are identified, some merging rules are applied. Afterwards, the requirement layer is linked to the service layer using the Topic-Maps matching rules. We note that this process aims at meeting as best as possible each requirement with an existing component.

Our methodology is based on an iterative process to enable the integration of the semantic description in a consistent manner, besides that it is difficult to gather large amount of requirements in one step in a big company like STMicroelectronics. This iterative process also enables to enhance the semantic during the construction of the Topic Map.

4.2 Construction of the service layer

The challenge in integrating the semantic description in a DIS mainly resides in the heterogeneity of the existing components. A DIS integrates a large variety of data sources and applications and the produced indicators represent heterogeneous information. For this reason, in the first step of the methodology, we propose to formalize and normalize the indicator structure in an indicator ontology. The goal of this ontology is to identify all the elements that constitute an indicator in a DIS. In addition, this ontology will enable to unify the terminologies used to express a requirement in terms of indicators. As a result, the matching between a requirement and the potential DIS components that can satisfy it will be better specified. A generic ontology for the indicator structure is given in figure 5.

An indicator is composed of a set of three main elements: Business fact, Dimension and Output-form type.
- **Business fact**: is a measurable business concept from the business domain. It is composed of an attribute which represents the central business concept, a unit of measure and eventually a description. This description can be used to contextualize the attribute in the business domain. We propose then in this description to use a controlled vocabulary from the business domain. A process fact can be analyzed according to one or many dimensions.
- **Dimension**: represents an analysis axis for a business fact. A dimension is described by a type and a set of values related to this type.
- **Output-form type**: is the type of display of data such as the type of a graph, etc.

At last, an indicator is defined by the triplet (BusinessFact, Dimension Output-formType) where a BusinessFact is expressed with a unit-of-measure and an

![Figure 4. Outlines of the methodology](image)

![Figure 5. The generic indicator ontology](image)
attribute. At last, this ontology enables to support the construction of the service layer. To that aim, we use a Topic Map to create and integrate this ontology in the service layer. We use the Wandora tool, a free software environment for Topic Maps creation and visualization. One interesting characteristic of this tool is that it enables to create layers of Topic Maps and merge them in a single Topic Map. It also enables to extract topics from different sources (XML, RDF, SQL databases, text documents, etc.). In the context of the DIS of STMicroelectronics, these techniques can be used to extract the concepts related to the indicators from the existing components. These concepts must be after structured according to the defined indicator structure. Indeed, a component takes data related to a business fact and a dimension from the business domain and produces an indicator in a type of output form, all specified by the user during the creation of the indicator. Thus, we use the service topic in the Topic Map to expose a component -referenced by its name in the service topic- and its outputs which are the produced indicators. We use three types of associations in this layer.
- Provided by (I, C): means that an indicator is provided by a component (the component name is used in this case).
- Analyzed by (BF, D): means that a business fact is analyzed by a dimension
- HasOutput (BF, OF): means that a business fact has a type of output form for display

At this step, a Topic Map for the service layer is produced and linked directly to the resources using the occurrences. Figure 6 shows an extract of the Topic Map that is being implemented within STMicroelectronics using the tree visualization in wandora.

We have in figure 6 the triplet (%_OOC, Daily, ControlChart) as indicator used within STMicroelectronics for the control of equipments. An OOC (Out of control) is detected when the measures used to perform the products with equipments have exceeded the limits fixed by the production engineer. Thus, the concept OOC is specific to the function equipment control from the business domain. We use then this description for the attribute OOC. The main purpose of a description here is to enable a better referencing and retrieval of the indicators according to the business context. This description doesn’t reflect the user requirement; it only gives the main function of the indicator according to the business domain. It helps the end-user to choose an indicator instead of another for some specific needs.

4.3 Construction of the requirement layer

An indicator is required by a user in a business domain to satisfy a business requirement. The nature of business requirements can differ following the business context of using DIS but the concept of goal that can be refined following its complexity is common to any company that has a DIS. Thus, to assist the elicitation of business requirements in the requirement layer, we propose in figure 7 a generic requirement meta-model related to a business domain. This meta-model can be adapted and enriched according to the business context of the company.

In a business domain, business goals like corporate or strategic goals are achieved by a set of quantified goals known as objectives. In the Business Motivation Model [18], an objective is attainable, time-targeted and measurable. If we take as example the control of the manufacturing process within STMicroelectronics, one of the strategic goals of the company is to Improve the Wafer Fab Yield (WFY).

This goal is quantified in each STMicroelectronics’ plants. For example, in the Rousset (France) plant, this goal is translated into the objective Reducing WFY loss of x%. An objective is achieved with a set of tactical goals.

![Figure 7. The generic requirement meta-model for DIS](Image)

A tactical goal defines the approach to implement and achieve an objective. This type of goal can be complex and then refined into sub-goals using a goal decomposition mechanism such as the AND/OR tree [19]. If we take again the goal of reducing the WFY loss, the actors involved in this process must reduce the scraps and prevent them by a set of control methods (a scrap is a wafer that is not compliant to the quality and the reliability requirements).

Because an objective is measurable, it has metrics based on unit-of-measures. These metrics are the indicators. Following the decomposition of the tactical goals, each goal here can involve several indicators. These indicators

![Figure 6. Extract of the Topic Map of the service layer](Image)
represent the core of user requirements because an actor can know with the indicators if he reached business requirements. Afterwards, the indicators’ definition for each low-level requirement is done using the indicators’ structure defined in the indicator ontology. The requirements’ elicitation is currently done within STMicroelectronics step by step by interviewing groups of business functions and analyzing their requirements. The requirement layer of the Topic Map is realized iteratively following the advancement of the requirements’ elicitation. If same requirements are identified, they are then merged in the Topic Map. The associations between the requirements must also be identified to enable a consistent linking between the topics of the requirement layer. The final purpose is to obtain a single and unified model of business requirements with the required indicators for the business domain. We can see in figure 8 the example of reducing WFY loss within STMicroelectronics with the topics of the requirement layer. We identify two types of requirement topics: goals and indicators. Thus two types of binary associations are used to link these topics:

- **Require (G1,G2)**: means that a goal G1 requires the goal G2 to be achieved (refers to the decomposition of goals).
- **Involves (G1)**: means that a goal G involves an indicator I to be satisfied.

The associations analyzed by and hasOutput are also used in this layer to specify the triplet (*BusinessFact, Dimension, Output-Form_type*) for the indicator structure.

Afterwards, a matching process is used to meet each indicator (requirement topic) with an output (service topic). One of the meaningful merging rules of the wandora tool is the syntactic matching between two topics using statistical techniques. For each indicator captured with the triplet (*BusinessFact, Dimension, Output-Form_type*) the system selects the component(s) that realize(s) the best matching with the required indicator structure. One important point in this matching is that the business fact is mandatory to satisfy at least a part of the requirement. For example, consider the indicator (% _OOC, Daily, ControlChart), if the % _OOC is given by a component but without the required dimension and output form, the system proposes anyway to link this component to the indicator. It means that the user must afterward adapt this component to his need. The main purpose of using this matching strategy is to enable reusing as much as possible existing components. At the end of this process the whole Topic Map is generated with the specified semantic. Figure 9 illustrates a part of the example reduce WFY Loss within STMicroelectronics. The resulting Topic Map enables an access entry to DIS components starting from any level of requirements. This Topic Map provides a semantic search to a user using the goal decomposition mechanism. This mechanism also assists the user during the definition of new requirements that involve indicators.

**The resulting Topic-Map meta-model**

According to the main outlines of the proposed methodological framework, new concepts have been created and used for the integration of the semantic description in the DIS. As a result, the standard meta-model of Topic Maps has been extended to include the concepts of our framework. Figure 10 presents the resulting meta-model. The standard concepts in a Topic Map are depicted with yellow color. We extended from the concept topic two types of topics as explained before: the requirement topic and the service topic. The associations used are extended from the association type. We note that these associations can be typed as well as needed. Finally, the proposed layered semantic description in

![Figure 8](image1.png)

![Figure 9](image2.png)

**Figure 8.** Extract of the Topic Map of the requirement layer

**Figure 9.** Extract of the resulting Topic Map

**5 The resulting Topic-Map meta-model**

The set of indicators defined in the requirement layer are matched with the outputs of the components at this step. To associate an indicator to a service topic from the service layer we use the business fact of the indicator structure.
this framework leads us to create a new type of Topic Map. In fact the concept of layer doesn’t exist in the Topic Map paradigm. Even if some Topic-Maps-based approaches try to propose layered descriptions [11] [12], these layers only refers to a group of topics, they are not implemented in practice. We propose then in this meta-model a new type of Topic Maps for DIS description (DIS TM in figure 10).

6 Conclusion

We proposed in this paper a Topic-Map-based framework to support the integration of business semantic in a DIS of a company. This business semantic aims at enhancing the retrieval of DIS components by end-users for further reuse. One main characteristic of the proposed framework is that it provides a requirement-oriented description for DIS components. In addition, the Topic Map paradigm provides a meaningful technological support for semantic sharing and retrieval of these components.

The study of business requirements within STMicroelectronics confirms us our findings and helps us to propose a suitable methodological framework to gather and integrate progressively the required semantic for a DIS. Current works try to validate the main steps of the methodology and try to study potential inconsistencies in the application of the methodology in the company. A specific search application based on the Topic Map standard is also planned to enable the effective use of the Topic Map of the framework by the end-users within STMicroelectronics.

7 References