CITOM: An Incremental Construction of Topic Maps

Nebrasse Ellouze, Nadira Lammari, Elisabeth Métais

PII: S0169-023X(12)00020-1
DOI: doi: 10.1016/j.datak.2012.02.002
Reference: DATAK 1370

To appear in: *Data & Knowledge Engineering*

Received date: 9 September 2010
Revised date: 1 February 2012
Accepted date: 23 February 2012

CITOM: An Incremental Construction of Topic Maps

Nebrasse Ellouze, Nadira Lammari, Elisabeth Métais

Laboratoire Cedric, CNAM
292 rue Saint Martin, 75141 Paris cedex 3, France
ellouznebrass@yahoo.fr
{metais, lammari}@cnam.fr

Abstract This paper proposes the CITOM approach for an incremental construction of multilingual Topic Maps. Our main goal is to facilitate user's navigation across documents available in different languages. Our approach takes into account three types of information sources: (a) a set of multilingual documents, (b) a domain thesaurus and (c) all the possible questioning sources such as FAQ and user's or expert's requests about documents. In this paper we present the different steps of the proposed approach to construct the Topic Map and the pruning process of the generated Topic Map. We validate our approach with a real corpus from the sustainable construction domain.

Keywords: Topic Map (TM), incremental construction, multilingual documents, information retrieval, thesaurus.

1 Introduction

Nowadays, besides information processing, current information systems (IS) are brought to manage information resources. The latter are, for many companies, one of the most important resources. So, they must be accessible to all IS stakeholders handling these resources. To enable this access, it is necessary to model and organize the knowledge contained in these resources. The Topic Map model is one of the existing models dedicated to knowledge representation and organization [1]. It allows structuring contents and knowledge provided by different information sources and different languages. It is intended to enhance navigation and improve information retrieval in these resources.

The information resources available in the present IS are voluminous and are continuously enriched. It is therefore impossible to envisage a manual creation of the Topic Map representing them. Several research studies have addressed this issue. Many proposals have focused on the construction of Topic Maps from text documents [2]. However, none of them can handle multilingual content. Moreover, although Topic Maps are dedicated to user's navigation and information retrieval (usage oriented), none of them takes into account user queries in the Topic Map building process.

In this paper, we propose an approach called CITOM (a French acronym for Construction Incrémentale de Topic Map), an evolutionary and incremental construction of a multilingual Topic Map. The resulting Topic Map gives a user the possibility to acquire knowledge from documents written in languages different from his native language.

In addition to a multilingual content composed by textual documents, CITOM uses two other information resources which are: a domain thesaurus and all possible query resources such as FAQ, potential queries of experts (or users) about the content, etc.

CITOM aims to provide a Topic Map allowing us to semantically structure concepts from various languages: it supports the specificity of the multilingualism which is the eventual missing of semantically equivalent translation of concepts. This is quite frequent when we consider documents provided from different cultures.

CITOM proposes to annotate the topics by meta-properties initialized during the creation of the Topic Map. The values of these meta-properties for a given topic will indicate the relevance of this topic and then could be used for the Topic Map evolution management or for the dynamic pruning that could be done before the Topic Map visualization in order to overcome its size problem.

The remainder of this paper is organized as follows. Section 2 presents our Topic Map model and describes its particularities. Section 3 focuses on our approach of Topic Maps construction and its different steps. Section 4 is dedicated to the pruning process of the generated Topic Map. A case study illustrating our approach is described in Section 5. Section 6 is devoted to a presentation of the related literature. Section 7 concludes and discusses further research.

2 CITOM Topic Map Model

A Topic Map Model meets the need of organizing contents from various resources (documents, databases, videos, etc.) and from different languages. It is an ISO standard (ISO 13250) [1] included into the ODM (Ontology Meta-Model Definition), by the OMG community [3], in order to provide a standard TM-UML model.

Figure 1 describes an extract of this model. As shown in this figure, a Topic Map is organized around subjects, called *topics*, representing subjects that the creator wishes to describe and for which resources are available. A topic may have a *base name* and *variant names*. It may be linked to one or more information resources that are deemed to be relevant to the topic. Such links are called *occurrences* of the topic. An *association* is a link representing a relationship between topics. They are specified by the creator of the Topic Map according to the knowledge required and according to the application to which the Topic Map is intended. They allow browsing the Topic Map and enable the interconnection of resources. The role played by a topic in an association (*association Role*) is also one of the characteristics of this topic.

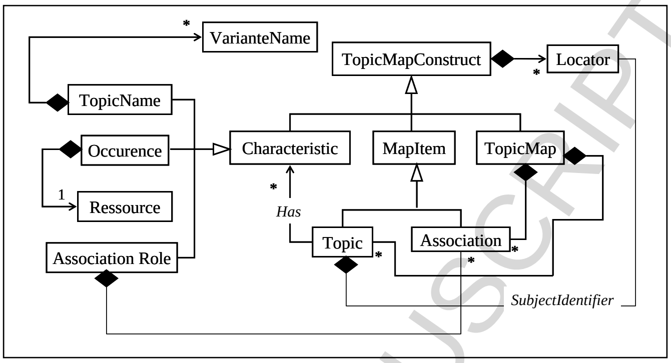


Fig. 1. An extract of the ODM Topic Map Model

The Topic Map standard provides also the concepts of *scope* (context) and *facets*. The scope is defined as a descriptor to contextualize topics corresponding to a vision of certain users. It provides the mean for indicating when each topic name, role and occurrences are appropriate. Facets are metadata associated to occurrences.

To achieve our goals, we have extended the Topic Map standard model. In the following, we describe the added concepts.

2.1 The CITOM Topic Map Model

Our Topic Map model gathers all the basic Topic Map standard model concepts (topics, occurrences, associations). A topic could be a term or a theme. It has at least one name. There may be several, one per language provided it has a name in that language. We also use the facet concept in order to filter documents according to their language. So a facet assigned to an occurrence is, in our context, a metadata describing the language of the document linked to this occurrence.

As shown in Figure 2, in addition to these concepts, we propose to assign meta-properties (metadata) to topics. They will contribute to the pruning of a Topic Map. We also define thematic segment within a document in order to refine information retrieval. Finally, we enrich the set of possible associations by a usage association that will help us to model sample queries. These three contributions are detailed in the following paragraphs.

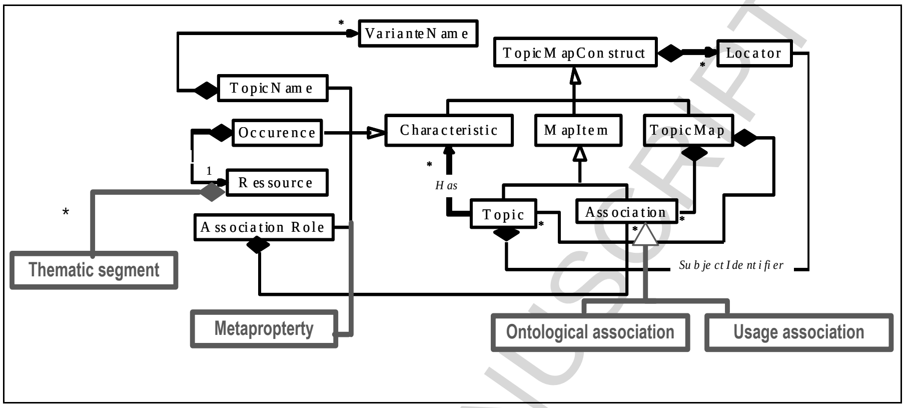


Fig. 2. Our extended Topic Map model

2.2 Classification of Topic Map links

Semantic links between topics allow navigation in the Topic Map structure. As mentioned in [4], there is no limitation in the definition of associations in a Topic Map. They are specified, by the Topic Map designer, according to its requirements, the knowledge to be represented and the associated domain application. However, the standard Topic Map model doesn't make the difference between association types.

Several works have been proposed to classify semantic links between concepts. For example, the work proposed by ANSI (American National Standards Institute) in ANSI/NISO Z39.19-2005, defines three types of links: (1) equivalence links like synonymy, (2) hierarchical links (generalization/specialization) and (3) other links called "associative links" such as "cause/effect link". In [5], the authors propose a multi-layered ontology to classify the semantics of relationship verb phrases.

In our extended Topic Map model, we propose to differentiate between links, since our approach produces, besides occurrences and ontological or structural links, usage links. Let us note that ontological and structural links are defined as specialization/generalization links, composition links and associative links as described in the ANSI/NISO Z39.19-2005 standard.

Usage links, called "is an answer to", are hyperlinks (hyper link questions/answers) between the question and all the associated answers. We propose also to link a question to all the keywords that compose this question using hyper links called "is composed of". Questions as well as answers and keywords are topics.

2.3 Adding meta-properties to the Topic characteristics

We propose to extend the Topic Map standard model by providing topics with meta-properties that are initialized when creating the Topic Map. They can be of two kinds. The first type gathers usage-based metadata. The latter inform us about the importance of the Topic Map elements and the use made during the exploitation of the Topic Map. They can be used to evaluate the quality of the Topic Map, for the management of its evolution and, for the display, by dynamically pruning topics or links in order to overcome the problem of volume often encountered in the Topic Maps. The second kinds of meta-properties are structure-based ones. They can contribute to the organization of the Topic Map into layers.

At the present time only those concerning topics are considered. In fact, we consider two structure-based meta-properties. The first one organizes the Topic Map into two abstraction levels: the upper level or knowledge level represented by topics describing themes and a lower level or taxonomic level where topics represents terms of documents. The second one offers another classification of topics: those composing question/answers couples and those that not participate into question/answers couples. Each of these two types of metadata will provide two types of navigation and can then contribute to reduce the set of topics to display.

We also consider three usage-based meta-properties attributed to topics: the one indicating the number of documents associated to a topic, the one that stores the number of user's consultations for a topic and finally the one representing the number of FAQs referring to a topic. Values assigned to a topic for each usage-based meta-property, if combined, may reflect the popularity level of this topic and, therefore, can compel temporarily its display or influence the decision to keep it in the Topic Map.

2.4 Document Fragmentation

One of the main originality of our extended model is the ability of linking a subject to a document fragment, rather than to link it to the whole document. Although this aspect contributes to provide a user with a more accurate and concise information and facilitates its retrieval within a document, it is not supported by the standard model. Moreover, to our knowledge, no existing extended Topic Map model allows binding a topic to a document fragment.

Figure 3 gives a synthetic presentation of the architecture of a Topic Map generated according to our extended model.

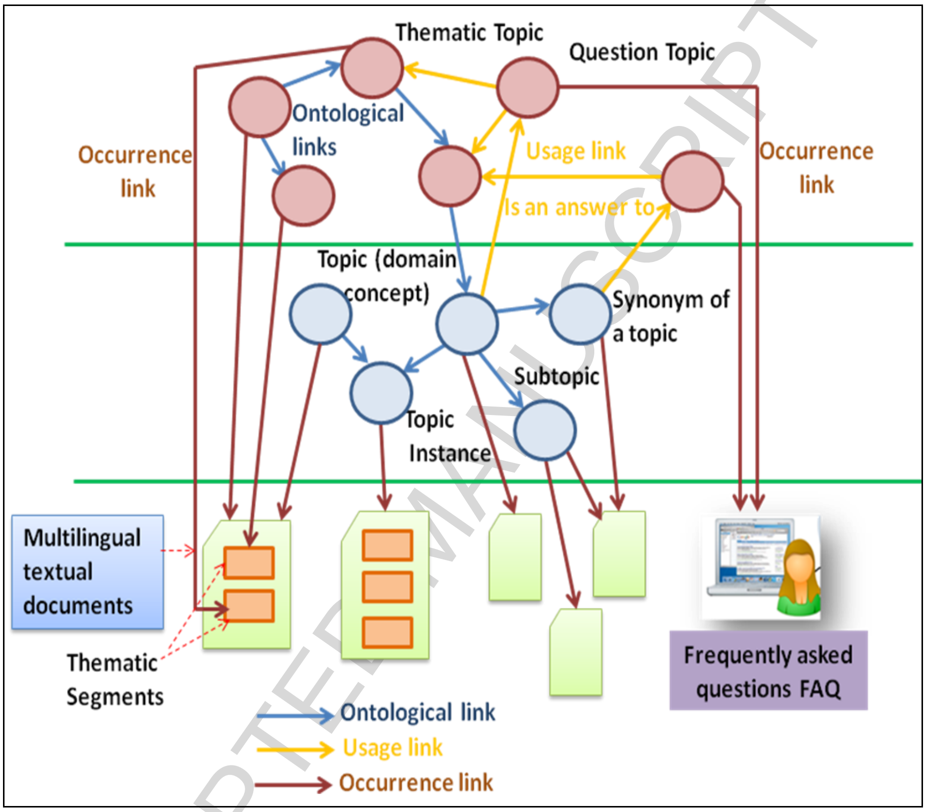


Fig. 3. General Architecture of the Topic Map according to our Topic Map extended model

3 Our Approach

Starting from a content composed of multilingual textual documents, our CITOM approach takes into consideration two other secondary sources which are: a domain thesaurus, a general ontology and all the possible questioning sources such as FAQ, user or expert requests related to the source documents, phone discussions and consultations with people working in the domain.

The main idea of CITOM is to build, in an incremental way, a Topic Map TM_i corresponding to a set of documents $D=\{d_1, d_2, . . ., d_i, . . .\}$ by enriching the Topic Map TM_{i-1} associated to the set of document $D-\{d_i\}$. This enrichment of TM_i is realized by integrating the Topic Map associated with a document d_i into TM_{i-1}. Each phase allowing the construction of the Topic Map associated with the document d_i, uses, as input, in addition to the document d_i, a domain thesaurus, a general ontology like WordNet and a set of questions related to the document and extracted from the questioning sources. Figure 4 provides an overall description of our approach.

Before starting to build a Topic Map, we propose to apply a thematic segmentation algorithm on these documents repository since, in most cases; documents may deal with various themes at the same time. In this case, it is, in our view, interesting to target, with the Topic Map, document segments as well as entire documents.

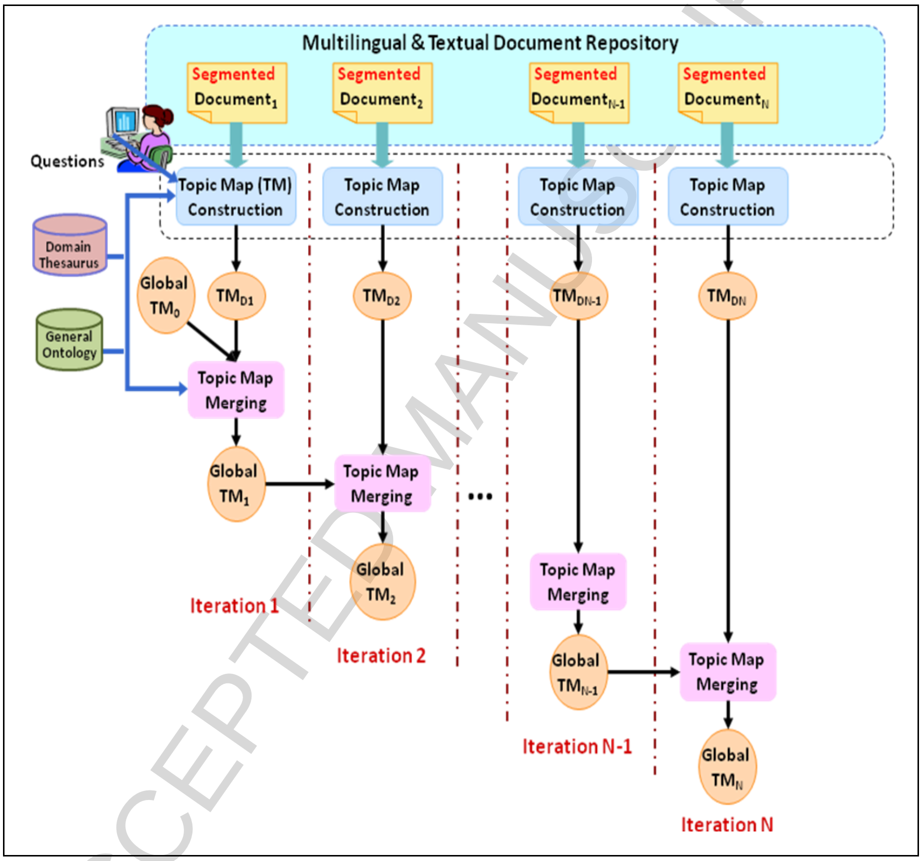


Fig. 4. CITOM approach

CITOM aims at providing a global Topic Map as a semantic structure that organizes concepts into various languages taking into account cultural peculiarity of the documents representing them. In fact, it may occur that a word in a source language may not have a correspondent word into a target language. This is very common when documents are from various cultures.

CITOM is an incremental approach in the sense that it produces a Topic Map that has been evolving gradually, during the construction process thanks to new documents or new questions introduced in the repository. Thus, our approach may contribute to the management of the evolution of Topic Maps.

The general algorithm to build the global Topic Map is the following:

Algorithm 1. The general algorithm of our approach

```
Inputs:  A repository composed of multilingual textual documents, a domain
         thesaurus and all questioning sources related to the source documents
         (experts questions, users requests, FAQ, phone discussions and interviews
         with people working in the domain, etc).
Output: A global Topic Map
    Action 1. Build the root of the global Topic Map. We mean by root the topic
              which has the domain name in different languages.
    Action 2. Treat questioning sources and provide, for each document of the
              repository, a set of potential questions.
For each document $D_i$ of the repository do:
    Action 3. Apply thematic segmentation on $D_i$ to get thematic fragments
    Action 4. Create a Topic $Map_i$ associated to Di by Extracting a list of topics
              and associations from $D_i$ and its segments.
    Action 5. Enrich Topic $Map_i$ with new ontological and structural links
              extracted from the domain thesaurus.
    Action 6. Enrich Topic $Map_i$ using the set of potential questions associated to
              the document.
    Action 7. Validate Topic $Map_i$ by domain experts.
    Action 8. Merge Topic $Map_i$ associated to $D_i$ with the global Topic Map.
End
```

The Topic Map validation step consists in defining (or fixing) the semantics of some links, adding or deleting some topics and/or associations. This phase is realized thanks to the collaboration of domain experts.

In this paper, we will focus only on the phases 3, 4, 5 and 6 of our approach. Sections 3.1, 3.2, 3.3 and 3.4 describe these phases.

3.1 Thematic segmentation of documents

Automatic text segmentation identifies the most important thematic breaks in a document in order to cut it into homogeneous parts. In [6], these parts, called “document units”, are defined as parts of text with strong intrinsic relationships, disconnected from other adjacent parts. More precisely, the segmentation task consists in partitioning a text into contiguous areas, by determining boundaries between them. These areas of text are called thematic segments.

Many segmentation methods have been proposed. Most of them rely on statistical approaches such as TextTiling [7], C99 [8], DotPlotting [9] or Segmenter [10]. They are based on the distribution of the words in the text, in order to determine the thematic changes. They are mostly applied to given types of corpus.

To demonstrate the feasibility of our approach, we used a large corpus of English and French documents. These documents are in different formats (pdf, foc, html, txt, etc). Most of them are voluminous.

To perform the segmentation, we choose to use TextTiling algorithm since it takes into account these types of documents and it can be applied not only to English documents but also to French ones. Moreover, based on experimental results in the literature [11], TextTiling is very performing when applied to voluminous documents and gives good results, in time and cost, in comparison with other segmentation algorithms like C99 or Segmenter.

3.2 Extraction of Topics and Associations

The goal of this phase is to extract from a document D_i a set of topics and associations between them. As mentioned in section 1, these associations are defined as ontological links (“is a”, “part of”, etc) and semantic links related to the Topic Map application domain. Topics could be themes or terms.

For thematic topics, we take advantage of what we did in the previous phase. In fact, through the segmentation process, we have decomposed documents into thematic segments. So to each defined segment will be created a thematic topic to which it will be linked.

For the extraction of topics representing terms and associations, we propose to use existing Natural Language Processing techniques and tools.

This issue is addressed by various approaches. Most of them are developed for ontology learning from textual documents. Most of them are implemented. We can classify them according to the techniques used: statistical, syntactic and text mining based approaches. Statistical methods are only applied for the extraction of terms. To select candidate terms, they use techniques based on measures. Among the most popular measures, we can cite: term frequency measure used to assign a weight to each term according to its frequency in the corpus [12], [13], [14], tf-idf measure, T-test measure [15], [16], [17].

Syntactic methods are applied for the extraction of terms and associations. However, they require a manual naming phase for the extracted associations. They are based on the analysis of grammatical dependencies between words or group of words in a phrase. Some of them exploit the hypothesis that grammatical dependencies between terms can be used to define semantic relationships [18], [19]. Other research works propose to use syntactic patterns [20], [21], [22] to detect relations between terms.

The last category of methods exploits text mining techniques [23], [17], [12], [14], [24], etc. Approaches, such as [23] and [17], use classification techniques and an ontology to realize the matching between candidate concepts (those present in the documents) and those of the ontology. [12] and [14] propose to apply clustering techniques in order to group terms according to their co-occurrences in the corpus. The gathering of concepts within a same cluster allows the deduction of possible relationships between these concepts. The approach proposed in [24] extracts, from the corpus, association rules between terms. Each association rule identifies a relationship between two concepts. A manual labeling process is then performed for naming the produced relations.

The literature also supplies tools to extract concepts and associations from textual documents: Nomino [25], Lexter [26], Fastr [27], Mantex [28], Likes [29], Acabit

[30], Syntex [31], OntoGen [32] and Text2Onto [33], etc. Syntex is a text analysis tool based on identifying syntactic dependencies between concepts. Text2Onto is an ontology learning tool from textual data. Text2Onto combines machine learning approaches with basic linguistic processing such as tokenization, lemmatizing and shallow parsing in order to identify concepts, relations between them and concept instances. It is based on the GATE framework [33] for processing texts.

Finally, we note that many text mining frameworks have been developed for linguistic processing of textual documents [35], [36] such as GATE (General Architecture for text Engineering). The latter could be used for semantic annotation and information extraction from text corpora.

To extract topics representing terms and associations, we choose the GATE platform which has the advantage to propose a generic solution for the linguistic processing of textual documents through a set of configurable modules. These modules can be combined, enriched and adapted to our needs. Moreover, GATE offers a module, called "gazetteer", for the recognition of named entities from pre-defined dictionaries. These dictionaries can be enriched with the terms of the application domain. In addition, GATE allows the integration of external resources, such as domain thesaurus, to build the hierarchy of topics and to add other links in the Topic Map.

3.3 Topic Map enrichment with new ontological links

This phase aims at organizing topics extracted from sources documents by adding new ontological and structural links ("is-a", "part-of", etc). For that, we propose to explore relations between the terms of the thesaurus. The ISO 2788 and ANSI Z39 standards have proposed the guiding principles for building a thesaurus. A thesaurus is a terminological resource in which terms are organized according to restricted relations: equivalence, hierarchical relations and non-taxonomic relations (associative links).

To the best of our knowledge, most of the existing Topic Map construction approaches do not propose to use a thesaurus to build a Topic Map. However, many ontology building approaches are based on existing thesaurus as a starting point to create an ontology, such as Hernandez's research work [37] who proposes to re-use a thesaurus to create and maintain a domain ontology. The authors define a method to extract conceptual schema elements of an ontology from a domain thesaurus and textual documents. The process is based on a set of transformation rules to re-use thesaurus relations. These rules explore "is more specific than" (IST), "is more generic than" (IGT), "use term instead" (USE) and "used for" (UF) relationships to generate ontology concepts, labels associated to each concept and hierarchies of concepts.

Our Topic Map enrichment approach with new ontological links is inspired from Hernandez' approach since we propose to reuse thesaurus relations to identify ontological and structural links. In fact, topics are organized in a hierarchical structure by means of "is-a" relationships. These relations are directly identified from "is more specific than" and "is more generic than" explicit relationships of the thesaurus. We use also "use term instead" and "used for" relations to add new names to a topic or to group two or more topics in one topic. The method that we have been proposing is defined as an algorithm executed in two steps. The first step concerns the use of (USE) and (UF) relations to group topics. The second step refers to topics organisation in hierarchical structures.

Let $SYN(Term_i)$ be the list of terms composed of Term_i and all the terms related to Term_i with USE and UF relations in the thesaurus. Let $/SYN(Term_i)/$ be the preferred term in SYN($Term_i$). Let $CHILD(/SYN(Term_i)/)$ be the list of terms identified from the thesaurus when parsing all paths starting from $/SYN(Term_i)/$. These paths contains only “is more generic than” links. In the following, we present the algorithm to group topics and affect multiple names to a topic:

Algorithm 2. Algorithm for grouping and labelling topics

```
For each topic $T_i$ do
            If $T_i$ is in the thesaurus
            Then compute $SYN(T_i)$, $/SYN(T_i)/$ and CHILD ( $/SYN(T_i)/$)
                  T will have as base name $/SYN(T_i)/$ and the other names
                  found in $SYN(T_i)$
End For
For each couple of topics $T_1$ and $T_2$ do
   If $SYN(T_1) = SYN(T_2)$
   Then group T1 and T2 in T3 such as the base name of T3 is $/SYN(T_1)/$ and
   The other names are those included in $SYN(T_1)$.
   As a result to this merging, all the other characteristics of $T_1$ and $T_2$
   (association roles and occurrences) are also merged.
End For
```

To organize topics, we propose to use two existing techniques among those that we have already proposed in [38] for building and maintaining ontologies. The first one, called “translation”, allows to transform a concepts hierarchy into constraints between these concepts. The second technique, called “normalization”, aims at building concepts’ hierarchy starting from a set of constraints between concepts. In the Topic Map enrichment process, we apply the first technique on the thesaurus in order to extract constraints between terms represented as topics in our Topic Map. After identifying these constraints, we apply the normalisation technique on the Topic Map.

In the particular context of Topic Map enrichment, we consider two types of constraints: semantic exclusion constraint and semantic inclusion constraint. A semantic exclusion constraint, denoted $\not\leftrightarrow$, defined between two concepts T_1 and T_2, expresses the fact that T_1 and T_2 don’t have the same semantics. A semantic inclusion constraint, denoted “$\mapsto$”, defined from a concept T_1 to a concept T_2 ($T_1 \mapsto T_2$) means that the semantic of T_1 contains the semantic of T_2 (the inverse is not necessarily true).

To extract constraints between terms represented as topics in our Topic Map, we apply the translation technique on the thesaurus. We use for “is more generic than” (IGT) relations already present in the thesaurus. This technique is based on three translation rules which are:

- *Rule 1*: if, in the thesaurus, a concept T_2 is more generic than a concept T_1, then ($T_1 \mapsto T_2$).
- *Rule 2:* if, in the thesaurus, there is no concept T_3 as a first node of two paths, one path goes to T_1 and the other goes to T_2 , these paths are only made of “is more generic than” links then $T_1 \nleftrightarrow T_2$.
- *Rule 3*: if, in the thesaurus, two concepts T_1 and T_2 are both linked to a concept T_3 by “is more generic than” link , then T_1, $T_2 \mapsto T_3$.

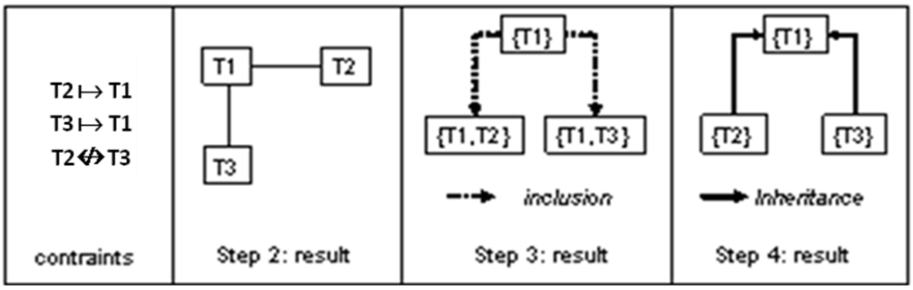
Let CHILD (T) be the set of terms of the thesaurus collected during the browsing of all paths consisting only of EPG links and having as starting node T. The extraction of constraints is done as follow:

Algorithm 3. Algorithm for extraction of constraints

```
For each topic $T_1$ do
    For each $T \in$ CHILD($T_1$)
        $T \mapsto T_1$
    End For
End For
For each couple $T_1$ and $T_2$ present in the Topic Map do
    If CHILD($T_1$)$\neq$ CHILD($T_2$) then $T_1 \nleftrightarrow T_2$
    Else if CHILD($T_1$)$\cap$ CHILD($T_2$) = $T_3$ then $T_1$, $T_2 \mapsto T_3$
End For
```

Once semantic constraints derived, we apply on the Topic Map the normalization algorithm in order to organize, in a hierarchical form, all its concepts. The latter encompasses four steps. First, it builds a complete non oriented graph with all the topics as nodes. Second, it eliminate any link between two nodes T_1 and T_2 if $T_1 \nleftrightarrow T_2$. Then, it deduces all possible cliques* that are compatible with the set of semantic inclusion constraints and organizing these cliques into an inclusion graph. Finally, by reversing inclusion links and eliminating redundancies, we obtain the hierarchy of topics.

Figure 5 shows an example of applying the normalization algorithm:



[*] A clique is a complete sub graph

Fig. 5. Example of applying the normalization algorithm

3.4 Topic Map enrichment with queries

The main interest of the Topic Map is to assist the user in his information retrieval. It should allow him to browse topics in order to find relevant documents. Another kind of guidance would be to introduce knowledge about questions frequently asked related to the source documents. To implement this type of guidance, we have proposed to represent in the Topic Map potential questions extracted from questioning sources such as FAQ, user or expert queries related to the source documents, phone discussions and consultations with people working in the domain. This extraction (Action 2 of Algorithm 1) is done, at the present time, manually. For that purpose, we have defined a new type of association, called "usage links" (see section 2.3). The latter allow linking questions represented as topics to associated answers represented also as topics. To offer the user the possibility to retrieve documents through submitting queries, we have proposed to link a topic question to all the keywords that compose it using "is composed of" link. Keywords are also topics but are not displayed (see the example given in Figure 6).

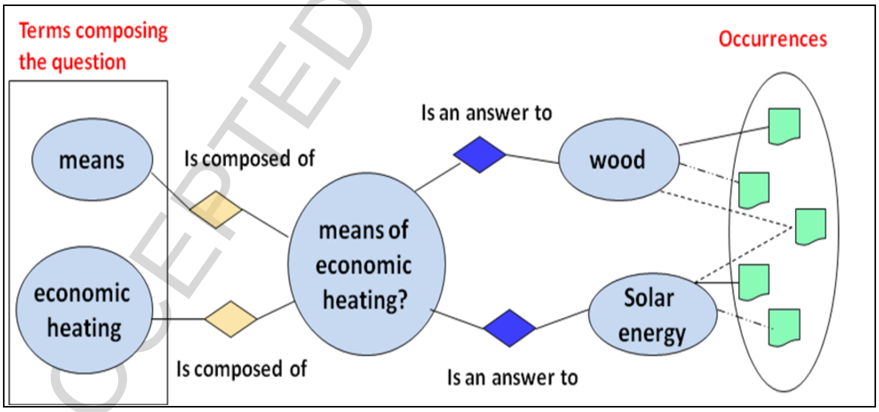


Fig. 6. Topic Map enrichment with usage links

Adding keywords to the Topic Map will allow automatic retrieval of documents and/or document segments containing answers to user queries. Indeed, if we compute the Salton vector [42] associated to a given user query (the set of keywords composing this query), we can automatically retrieve answers by comparing this vector to those stored in the Topic Map. This comparison is done by calculating distances (cosinus of the angle made by two vectors) between user's Salton vector and those stored in the Topic Map.

The enrichment process consists on integrating questions, keywords, answers and occurrences of answers into the Topic Map. The integration of the two last concepts requires computing answers from questions using linguistic processing techniques like [15], [16], [19].

To insert a topic question, already identified by the enrichment process, in the Topic Map, three cases can occur: (1) the topic already exists in the Topic Map with the same name, (2) the topic already exists in the Topic Map but labeled with a different name and finally (3) the topic does not exists in the Topic Map. For the second situation we propose to add a name to the existent topic. For the last situation, we proceed to the insertion of a new topic. The matching between a new topic and those in the Topic Map is used using domain ontology.

4 Topic Map pruning process

The Topic Map pruning process is a big issue to be addressed in our work. Indeed, a Topic Map is essentially used to organize document content and to help users finding relevant information in this content. Therefore, it is required to maintain and enrich the Topic Map along the time in order to satisfy users' queries and handle changes in relation with the repository evolution and in relation with the usage of the Topic Map.

We are interested in this research work by the size of the Topic Map that can grow very quickly if the content size grows. This problem can affect the quality of the Topic Map from a usage point of view. To overcome this problem, we need a pruning process. To decide about its execution, we need some indicators. At a present time, we focus on topics pruning. For this purpose, we defined a score indicator that reflects the importance of a topic. It is calculated from usage-based meta-properties which are metadata assigned to topics (see section 2.4). It corresponds to the weight average of values associated with usage-based meta-properties. Let (1) *T* be a given topic, (2) *DN* be the value corresponding to the number of documents indexed by *T*, (3) *FN* be the number of FAQs referring to *T*, (4) *CN* be the number of consultation of *T*, the formula allowing the computation of the score *S* of *T* is as follow:

$$S = (\alpha *\mathrm{DN} + \beta *\mathrm{FN} + \gamma * \mathrm{CN})/ \alpha + \beta + \gamma).$$

α, β and γ are weights assigned to each meta-property. They are configurable by the Topic Map creator. However we suggest setting γ greater than α and β in order to better reflect the usage of the Topic.

This score indicator contributes not only to the evolution of a Topic Map but also can compel temporarily the display of some topics. It is obvious that this score is not the only indicator that can contribute to the management of the evolution of a Topic Map. So we can consider this work as a starting point of further research.

5 Validation

To demonstrate the feasibility of our approach, we have developed a prototype and we have applied our approach to the sustainable construction domain. These two points are presented in the next paragraphs.

5.1 Prototype presentation

CITOM have been implemented as a JAVA application composed of two main functional packages: a Topic Map construction package and a Topic Map editing package.

The Topic Map construction package is a set of modules: Configuration module, Storage module and Topic Map generation module. The configuration module is used to set the execution of the application to a given repository. It allows a Topic Map creator to choose:

- The Topic Map storage format : XTM, RDF or OWL format (the XTM format is a default setting);
- The language in witch the Topic Map will be visualized for browsing.

The Topic Map generation module takes as input a repository composed of multilingual textual document and thesaurus paths and an ontology. It applies thematic segmentation to the source documents using TextTiling program. Then, it extracts topics and associations. The extraction is performed via a connection to the Gate platform. The generated topic is stored using the storage module into an XTM file.

The Topic Map editing package contains two consultation modules. The first one allows the user to navigate through the Topic Map. The second one is dedicated to querying the Topic Map using Tolog as a Topic Map query language. To query a Topic Map the user enter his request, the querying module transform it into a Tolog query and returns a list of document and/or segments. To visualize a Topic Map we have chosen after testing many visualization tools like TM4J tool [44], Ontopia Vizdesktop [45] and TMNav [46], the Treebolic tool [47]. This tool is not dedicated to the visualization of Topic Map. It allows to visualize a Topic Map as an hyperbolic tree. In our application, we have performed some adaptations that have required more effort. However, we have obtained good results in terms of Topic Map visualization compared with Topic Map tools.

5.2 Application for the sustainable construction domain

We experiment our approach CITOM on a real corpus from the sustainable construction domain. This corpus contains 105 documents with different formats (pdf, html, doc, txt). The whole size of our corpus is 14 Méga bytes. We have downloaded them from various bilingual web sites (English/French) specialized in the domain and especially those concerned with solutions for energy economy. We quote below some of these sites:

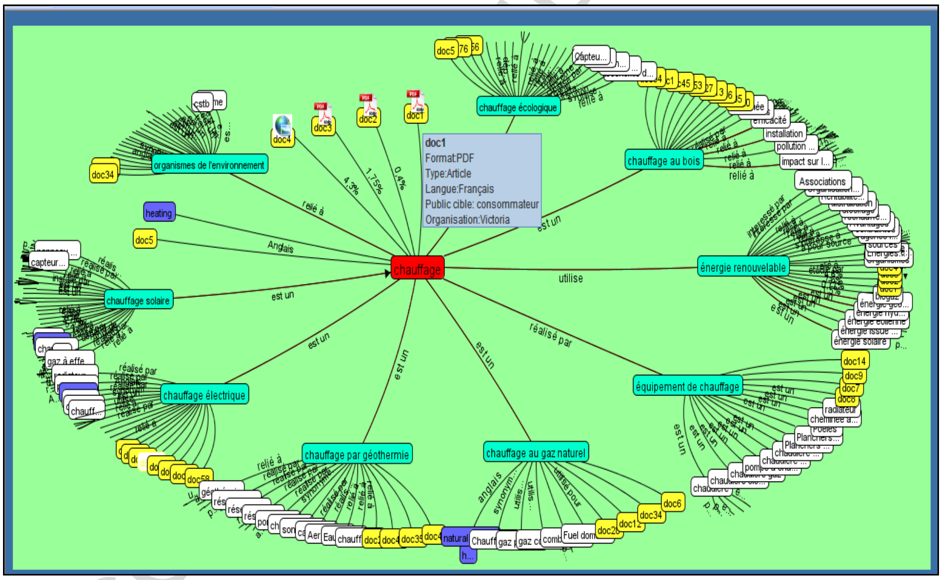
- http://www.ademe.fr, web site of Agency of Environment and Energy Management;
- http://www.cstb.fr, web site of scientific and technique center of building;
- http://www.rncan.gc.ca, web site on natural resources in Canada;
- http://www.ec.gc.ca, web site on environnement in canada;
- http://www.avenir-energie.com, web site on solutions for ecologic and economic heating.

We used as input a bilingual thesaurus (French/English) called CTCS (Canadian Thesaurus of Construction Science and Technology) from the domain of construction science [43]. This thesaurus contains 15331 terms organized in hierarchy of 10 levels. Each term is described in html file. Relations between terms are external cross links contained in this file.

The segmentation allows us to identify nine themes (or subjects) in the corpus: heating, economic heating, wood heating, solar heating, natural gaz heating, geothermal heating, fuel heating, electrical heating and ecological heating.

We also selected questions from a list of FAQ extracted from the mentioned websites and integrate them in the Topic Map.

Figure 7 presents the Topic Map generated from the corpus.



doc1
Format:PDF
Type:Article
Langue:Français
Public cible: consommateur
Organisation:Victoria

Fig. 7. Topic Map visualization

The user can get metadata values associated with a document or a segment of document like document type, documents format, document language, organization, etc. (see the gray rectangle in Figure 7). He can also access to the whole document or segment of document when he selects the document or segment node. Figure 9, is an example of document provided by the application.

Our application allows us also to focus on a topic by highlighting only the topics related to it (see Figure 8).

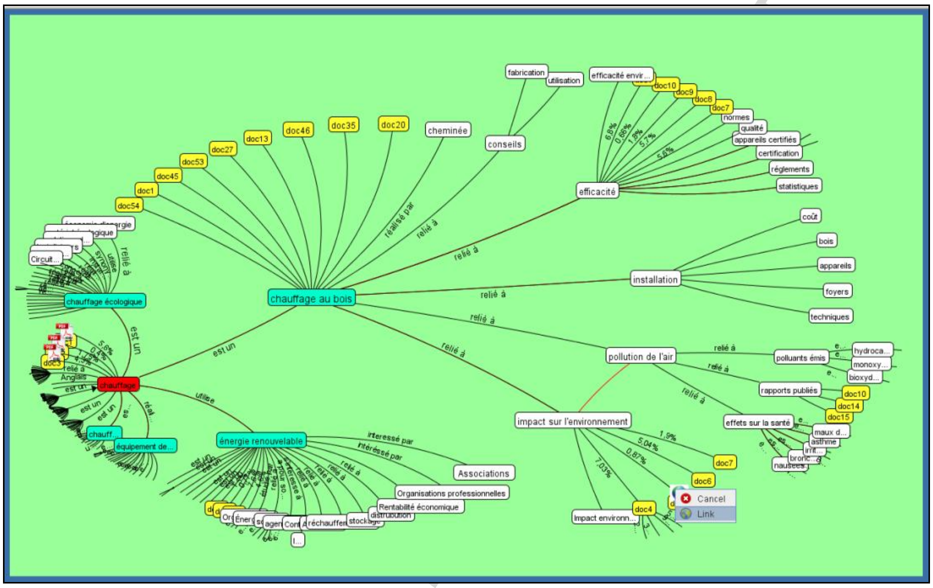
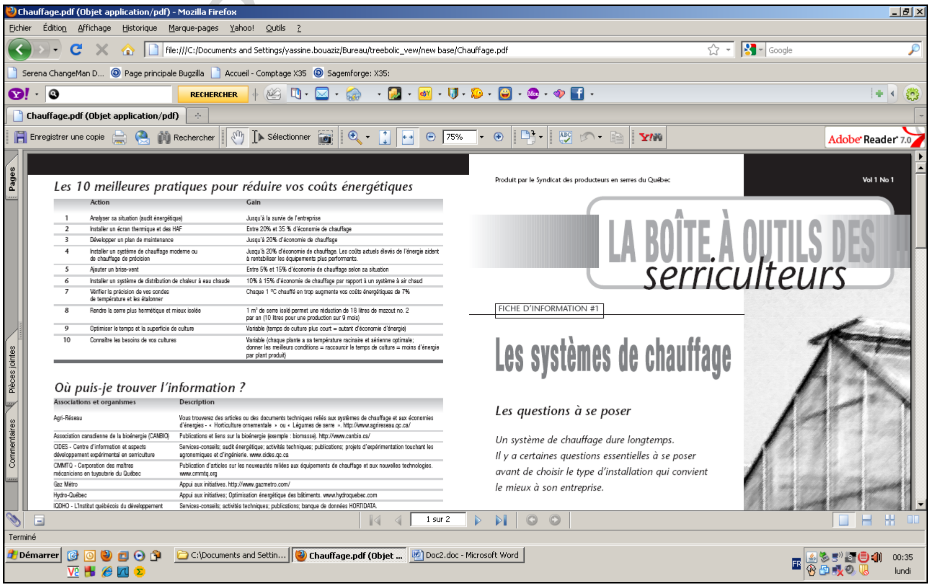


Fig. 8. Example of choosing “wood heating” as focus topic

As we stated before, when he navigates, the user can access to a document through the Topic Map, For example, as shown in the figure 9, the user choose to visualize a document related to heating systems:



file:///C:/Documents and Settings/yassine.bouaziz/Bureau/treebolic_vew/new base/Chauffage.pdf

Les 10 meilleures pratiques pour réduire vos coûts énergétiques

	Action	Gain
1	Analyser sa situation (audit énergétique)	Jusqu'à la survie de l'entreprise
2	Installer un écran thermique et des HAF	Entre 20% et 35 % d'économie de chauffage
3	Développer un plan de maintenance	Jusqu'à 20% d'économie de chauffage
4	Installer un système de chauffage moderne ou de chauffage de précision	Jusqu'à 20% d'économie de chauffage. Les coûts actuels élevés de l'énergie aident à rentabiliser les équipements plus performants.
5	Ajouter un brise-vent	Entre 5% et 15% d'économie de chauffage selon sa situation
6	Installer un système de distribution de chaleur à eau chaude	10% à 15% d'économie de chauffage par rapport à un système à air chaud
7	Vérifier la précision de vos sondes de température et les étalonner	Chaque 1 °C chauffé en trop augmente vos coûts énergétiques de 7%
8	Rendre la serre plus hermétique et mieux isolée	1 m^{3} de serre isolé permet une réduction de 18 litres de mazout no. 2 par an (10 litres pour une production sur 9 mois)
9	Optimiser le temps et la superficie de culture	Variable (temps de culture plus court = autant d'économie d'énergie)
10	Connaître les besoins de vos cultures	Variable (chaque plante a sa température racinaire et aérienne optimale; donner les meilleurs conditions = raccourcir le temps de culture = moins d'énergie par plant produit)

Où puis-je trouver l'information ?

Associations et organismes	Description
Agri-Réseau	Vous trouverez des articles ou des documents techniques reliés aux systèmes de chauffage et aux économies d'énergies - « Horticulture ornementale » ou « Légumes de serre ». http://www.agrireseau.qc.ca/
Association canadienne de la bioénergie (CANBIO)	Publications et liens sur la bioénergie (exemple : biomasse). http://www.canbio.ca/
CIDES - Centre d'information et aspects développement expérimental en serriculture	Services-conseils; audit énergétique; activités techniques; publications; projets d'expérimentation touchant les agronomiques et d'ingénierie. www.cides.qc.ca
CMMTQ - Corporation des maîtres mécaniciens en tuyauterie du Québec	Publication d'articles sur les nouveautés reliées aux équipements de chauffage et aux nouvelles technologies. www.cmmtq.org
Gaz Métro	Appui aux initiatives. http://www.gazmetro.com/
Hydro-Québec	Appui aux initiatives; Optimisation énergétique des bâtiments. www.hydroquebec.com
IQDHO - L'Institut québécois du développement	Services-conseils; activités techniques; publications; banque de données HORTIDATA.

Produit par le Syndicat des producteurs en serres du Québec

Vol 1 No 1

LA BOÎTE À OUTILS DES serriculteurs

FICHE D'INFORMATION #1

Les systèmes de chauffage

Les questions à se poser

Un système de chauffage dure longtemps.
Il y a certaines questions essentielles à se poser avant de choisir le type d'installation qui convient le mieux à son entreprise.

Fig. 9. Visualizing a document through the Topic Map

6 Related works

Research on Topic Map engineering has lead at various proposals. Most proposals have focused on the construction of Topic Map. Few of them have concerned the extension of the Topic Map model. These extensions have all been used for the construction of specific Topic Map. Further studies are more oriented to the Topic Map merging issues. Other ones have addressed issues related to the quality or to the visualization of Topic Maps.

6.1 Extended Topic Map models

The standard Topic Map model (ISO 13250) has been used by most Topic Map construction approaches. However, some propositions have suggested an extension of this model ([48], [49], [50] and [51]). Ueberall and al. introduces concepts by like entity, view, facet classes, restriction values and some metrics measuring the navigational quality of facets [48]. These concepts have been added in order to facilitate the exploratory search of content. The HyperTopic Model of [49] was designed to represent knowledge about collections like products, projects, books and to give multiple points of view of this knowledge. It incorporates the concept of “entity”, “point of view” and “standard attribute”. SocioTM model also extends the Topic Map paradigm by adding relevancies metadata to each Topic Map elements (topic, occurrence, association, etc.) in order to generate, on the fly, a dynamic Topic Map associated to the user profile [50]. The authors of [51] present an extended topic model where a Topic Map is structured on three levels: a clustering level, a knowledge element level and a topic level. They also define pre-order and post-order associations between knowledge elements. Let us note that some Topic Map models have been used in distributed knowledge context. For example, Khortaus et al. have built a Topic Grid infrastructure based on the Topic Map standard model [52]. Lu et al. propose a distributed knowledge system based on their extended model [51].

Our Topic Map model takes into account the key concepts of the Topic Map standard model. However it differs from other proposed models since it incorporates other concepts allowing to access to document segments and to proceed to the pruning of the Topic Map. Moreover, through the usage association concept, our contribution anticipates the possible user's queries and then we allow the user to retrieve rapidly the knowledge he needs. Finally, by defining two types of topics (theme topics and term topics) we can expect, in the future, two levels of details in the display of topics according to user's needs.

6.2 Topic Map construction approaches

Many Topic Map construction approaches are proposed in the literature. They mainly differ (a) by the sources required as input (XML documents, Web resources, RDF metadata, knowledge bases, thesauri, ontologies, text documents, etc.), (b) by the techniques used during the process (fusion, linguistic analysis, learning, classification, etc.), (c) by the underling Topic Map model (standard or extended model), (d) by the type of the generated Topic Map (Centralized or distributed Topic Map) and (e) by the degree of collaboration for its construction (individual or collaborative construction) and finally by the availability or not of a supported tool. Table 1 summarizes this comparison for some recent proposals. It also positions our approach according to the above criteria. A detailed survey of existing approaches is presented in [2].

Comparison criteria **Approaches**	**Required Inputs**		**OutPut**	**Technique used**	**Underling model**	**Collaborative construction**	**Availability of Construction Tool**
	Main source	**Secondary source**	**Centralized /Distributed Topic Map (TM)**				
Reynolds et Kimber (2002) [53]	XML documents	Domain ontology	Centralized TM	-Extraction technique -Merging technique	Standard	No	No
Folch et Habert (2002) [54]	Textual documents		Centralized TM	-NLP techniques -Clustering technique	Standard	No	Yes
Böhm et al. (2002) [55]	Textual documents		Centralized TM	-Text mining techniques -Linguistic analysis techniques	Standard		
Köhler et al. (2004) [56]	textual documents		Centralized TM	-indexing process (lexical analysis, stopword removal, stemming) Term weights computation technique NLP techniques	Standard	No	Yes
Mas et al. (2006) [57]	Web sites User Web histories		Centralized TM	-Structured-Based Hierarchical clustering -Rule-based mapping technique	Standard	No	No
Kasler et al. (2006) [58]	Bilingual textual documents (English and Hungarian)	-Domain ontology -Taxonomies -Dictionaries	Centralized TM	-machine learning techniques information retrieval techniques -pattern matching techniques	Standard	No	Yes

Zaher et al. (2006) [59] **Zaher et al. (2007)[60][61]**	Web ressources	-Co-designer group -User advocate actors (regulators)	Centralized TM	-"SeeMe" modelling method -conflict-based co-construction method	HypertTopic	Yes	Yes
Pepper (2007) [62][63]	Dublin Core metadata		Centralized TM	Rule-based mapping technique	Standard	No	No
Roberson et Dicheva (2007) [64]	Web Pages		Centralized TM	-Crawlng websites -Parsing HTML codes -Rule-based mapping technique	Standard	No	Yes
Librelotto et al. (2008)[65]	Heterogeneous information systems (databases, xml documents, etc.)	Domain ontology	Centralized TM	Dataset extraction techniques	Standard	No	
Neidhart et al. (2009) [66]	Relational databases (SQLite)		Centralized TM	Rule-based mapping technique	Standard	No	Yes
Zheng et al. (2009) [67]	Textual documents		-Distributed TM -Global TM	Merging Learning techniques	Extended	No	Yes
Weber et al. (2010) [68]	Relational database	Domain ontology	Centralized TM	Mapping rules	Standard	No	Yes
Garshol et Fischer (2010)[69]	Liferay's CMS	Ontopoly Topic Maps editor of Ontopia	Centralized	User interaction	Standard	Yes	Yes
Eslami et al. (2011)[70]	Relational databases (Microsoft SQL Server)		Centralized TM	Rule-based mapping technique	Standard		
Dharavath et al. (2011) [71]	Hidden Web		Centralized TM	-Crawling technique -Pattern matching techniques	Standard	No	Yes
CITOM	**Multilingual textual documents**	**Domain Thesaurus WordNet**	**Centralized TM**	**NLP techniques Merging**	**CITOM TM**	**No**	**Yes**

Table 1. Comparison of most recent Topic Map construction approaches

As the table 1 shows, CITOM is an approach dedicated to the organization of textual documents contents. A careful analysis of these approaches leads us to the three following comments:

1) No existing approach handles multilingualism, except Kasler's approach [58] which takes into account only two languages: English and Hungarian language;

2) Adding occurrences to topics is too labor-intensive. None of the existing approaches is incremental. As a consequence, adding a document requires the reconstruction of the entire Topic Map;

3) No approach leads to facilitate the knowledge search within a document. So the user has to cross the entire document to find the searched knowledge;

Our Approach CITOM overcomes these three inconvenient. It is an incremental approach. It also handles multilingual documents. It allows the association between document fragments and topics. Finally, it can be seen as usage oriented method since it allows the anticipation of user requirements by storing sample queries.

Let us mention that [72] proposes an automatic approach for labelling topics extracted from English textual documents. This approach uses English Wikipedia and is based on Information retrieval techniques, NLP techniques and some techniques to learn the association of a label candidate with the topic terms like lexical association measures. This approach, although may contribute to the construction of Topic Map, it remains dedicated to monolingual Topic Maps.

6.3 Topic Map merging

The Topic Maps model defines a generic merging function called "MergeMap" based on merging rules that use the equivalence principles to determine whether two or more Topic Map elements (topics, associations, etc.) can be merged. This function doesn't allow us to merge similar Topic Map elements. To address this issue, some research studies such as that of Maicher et al. [73] and Chung et al. [74] focused on defining similarity measures between Topic Map elements. Other ones goes beyond by defining merging approaches of local Topic Maps (based on a standard or extended model) into global ones [75], [76], [52], [67], [77], [78], [79] and [80]. Some of them operate at the syntactic level. Those that operate at a semantic level use domain ontologies or common dictionary like WordNet.

This topic is also the object of many research works in the closed domain of ontologies merging [81], [82], [83], [84], [85], [86] and [87].

In CITOM, the integration process builds on the work of our team on merging ontologies [38] and conceptual schemas [88]. It takes into account the multilingual nature of our Topic Map.

6.4 Topic Map tools

Topic Map technology is becoming more and more popular. The need for support tools is inevitable. However, most of them allows manual creation and browsing of Topic Maps: Mondeca [89], TM4J [44], Topic Map Designer [90], Ontopia Navigator Framework [45], TM4L ([91], [92]), TMshare [93], tools of Godehardt et al. [94], TopiMaker [95], TMEd [96], Agorae tool [97], TROPICS [98] etc. Some of them are domain specific. For some of these tools, navigation is realized via indexes. Each index corresponds to a topic. Selecting a topic allows us to view information about it. For other tools, navigation is graphical, usually in the form of hyperbolic trees, sometimes with different levels of navigation (class topics, topics, resources). Most of these tools don't allow the visualization of the whole Topic Map.

Let us note here the specificity of the tool TROPICS that allows Intra-Topic Map navigation, Inter-Topic Maps navigation and Merged navigation. It also offers the possibility of querying the resources using a query string.

Few of them are dedicated to automatic or semi-automatic creation or merging of Topic Map (ETM Toolkit [99], Tools of Roberson and Dicheva [64], Kasler et al. [58] and Korthaus et al. [52], Metamorphosis [65], ITM Tool [51], etc.). CITOM tool allows the generation of Topic Maps, its visualization and browsing.

6.5 Topic Map quality

In our opinion, very few research works have tackled the Topic Map quality issue [94], [100], [101], [102], [103]. Studies addressing this problem can be classified into two classes. The first one gathers research works focusing on the evaluation of the Topic Map quality from the visualization point of view [94], [100]. The second one concerns research works treating this aspect from the Information Retrieval point of view [101], [102], [103]. To our knowledge, none of the existing works have addressed the Topic Map quality issue from the user's point of view or from the Topic Map evolution point of view. We think that our proposition of meta-properties for the pruning of Topic Maps, although incomplete and requires further research, will contribute to this last issue.

7 Conclusion and future work

In this paper, we have presented CITOM, an incremental approach to build multilingual Topic Map. CITOM was implemented and tested on a corpus of documents concerning sustainable construction.

Unlike existing methods, CITOM has the advantage to take into account, through the definition of usage links between potential questions extracted from all the possible questioning sources and the associated answers, the future usage of a Topic Map. Moreover, any potential question (expressed in natural language) represented by a topic is also connected to each of its constituent keywords via the "is composed of" link. Thus, information retrieval could be realized not only through navigation but also through automatic search of "similar questions".

CITOM takes into consideration multilingual resources. Thus, a user may, when browsing topics, access to documents that are not in its native language. The great advantage of this approach, compared to simple translations of answers, is to provide, the user, documents corresponding to concepts that not exist necessarily in its language or culture. This is, from our point of view, a cultural enrichment.

Moreover, the underling Topic Map model of CITOM provides usage-based meta-properties that could be used not only for the evaluation of the quality of a Topic Map but also for a dynamic pruning of topics during the display of the Topic Map or for the management of its evolution. The Model provides also structural-based meta-properties that can contribute to the choice of topics to display (only question, only themes, etc.).

In our view, there are still several areas that need further study and exploration. Indeed, the CITOM validation step is until now performed by the creator of the Topic Map. In the near future, we will focus on this step to propose a collaborative validation approach involving different domain experts.

Another enrichment will be to extend and improve the Topic Map pruning process by integrating new criteria (meta-properties) to evaluate Topic relevance and by defining meta-properties for the other elements of the Topic Map like associations.

Acknowledgements

The authors would like to thank Pr. Jacky Akoka and the reviewers for their useful comments.

References

1. ISO/IEC WD 13250-2, Topic Maps - Data Model (TMDM), 2008-06-03, International Organization for Standardization, Geneva, Switzerland. http,//www.isotopicmaps.org/sam/sam-model/2008-06-03/.
2. N. Ellouze, E. Métais, M. Ben Ahmed, State of the Art on Topic Maps Building Approaches, In R.-D. Kutsche and N. Milanovic (Eds.), MBSDI 2008, Model Based Software and Integration Systems, CCIS 8, pp. 102–112, © Springer-Verlag Berlin Heidelberg, 2008.
3. Ontology Definition Metamodel (ODM) Version 1.0, http://www.omg.org/spec/ODM/1.0/
4. S. Pepper, Article for the Encyclopedia of Library and Information Sciences, http://www.ontopedia.net/pepper/papers/ELIS-TopicMaps.pdf, 2008.
5. V. C. Storey, S. Purao, Understanding Relationships: Classifying Verb Phrase Semantics, Conceptual Modeling – ER 2004, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Volume 3288/2004, ISSN 0302-9743 (Print) 1611-3349 (Online), pp. 336-347, 2004.
6. G. Salton, A. Singhal, C. Buckley, M. Mitra, Automatic text decomposition using text segments and text themes; Proc. Hypertext '96, The Seventh ACM Conference on Hypertext Washington DC, pp. 53-65. ACM. 1996.
7. M. Hearst, Texttiling: segmenting text into multi-paragraph subtopic passages; Computational Linguistics, 23(1), pp. 33-64. 1997.
8. F. Choi, Advances in domain independent linear text segmentation; Proc. of the first conference on North American chapter of the Association for Computational Linguistics (2000), San Francisco, CA, USA, 26-33. 2000.
9. M. Kan, J. Klavans, K. McKeown, Linear segmentation and segment significance; Proc. 6th Workshop on Very Large Corpora, ACL SIG-DAT, 197-205. 1998
10. J. Reynar, Topic Segmentation: Algorithms and applications. PhD thesis, University of Pennsylvania, Seattle, WA. 2000.
11. O. Ferret, Approches endogène et exogène pour améliorer la segmentation thématique de documents, 148 TAL. Volume 47 – n° 2/2006, pp. 111-135, 2006.
12. E. Agirre, O. Ansa, E. Hovy, D. Martinez, Enriching very large ontologies using the WWW, In ECAI 2000 workshop on Ontology Learning, Berlin, Germany, 2000.

13. A. Faatz, R. Steinmetz, Ontology enrichment with texts from the WWW, In the Semantic Web Mining Conference WS'02, 2002.
14. V. Parekh, J-P. Gwo, T. Finin, Mining Domain Specific Texts and Glossaries to Evaluate and Enrich Domain Ontologies , In International Conference of Information and Knowledge Engineering, 2004.
15. P. Velardi, M. Missikof, P. Fabriani, Using text processing techniques to automatically enrich a domain ontology , In Proceedings of ACM- FOIS, 2001.
16. F. Xu, D. Kurz, J. Piskorski, S. Schmeier, A domain adaptive approach to automatic acquisition of domain relevant terms and their relations with bootstrapping , In the 3rd international conference on language resources and evaluation, 2002.
17. K. Neshatian, M. R. Hejazi, Text categorization and classification in terms of multi-attribute concepts for enriching existing ontologies, In 2ndWorkshop on Information Technology and its Disciplines, p. 43–48, 2004.
18. R. Bendaoud, M. Rouane Hacene, Toussaint Y., Delecroix B., Napoli A., Construction d'une ontologie à partir d'un corpus de textes avec l'ACF , IC 2007.
19. C. Roux, D. Proux, F. Rechermann, L. Julliard, An ontology enrichment method for a pragmatic information extraction system gathering data on genetic interactions , In Proceedings of the ECAI2000 Workshop on Ontology Learning, OL 2000.
20. M.A. Hearst, Automatic acquisition of hyponyms from large text corpora, Rapport technique S2K-92-09, 1992.
21. A. Maedche, S. Staab, Mining ontologies from text, Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. volume 1937, Springer-Verlag, 2000.
22. G. Stumme, A. Hotho, B. Berendt, Semantic web mining: State of the art and future directions. Web Semantics: Science, Services and Agents on the World Wide Web, 4(2):124–143, June 2006.
23. E-H. Han, G. Karypis, Centroid based document classification: Analysis and experimental results , In The 4th European Conference of Principles of Data Mining and Knowledge Discovery, pp 424–431, 2000.
24. R. Agrawal, R. Srikant Mining generalized association rules, Future Generation Computer Systems, 13(2–3): 161–180, 1997.
25. L. Dumas, A. Plante, P. Plante, ALN: Analyseur Linguistique de ALN, vers.1.0 . ATO, UQAM, 1997.
26. D. Bourigault, LEXTER, a Natural Language Processing tool for terminology extraction, Proceedings of the 7th EURALEX International Congress, Goteborg, 1996.
27. C. Jacquemin, D. Bourigault, Term Extraction and Automatic Indexing , in Mitkov R. (ed), The Oxford Handbook of Computational Linguistics, Oxford University Press, pp. 599-615, 2003.
28. P. Frath, R. Oueslati, F. Rousselot, Identification de relations sémantiques par repérage et analyse de cooccurrences de signes linguistiques, in Ingénierie des connaissances. Évolutions récentes et nouveaux défis. Eds. Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault, Eyrolles, Paris, pp 291-304, 2000.
29. F. Rousselot, P. Frath, R. Oueslati, Extracting concepts and relations from Corpora . In Proceedings of the Workshop on Corpus-oriented Semantic Analysis, European Conference on Artificial Intelligence, ECAI 96, Budapest, 1996.
30. B. Daille, Identification des adjectifs relationnels en corpus, in Actes de la Conférence de Traitement Automatique du Langage Naturel (TALN'99), Cargèse, 1999.
31. D. Bourigault, C. Fabre, C. Frérot, M.-P. Jacques, S. Ozdowska, Syntex, analyseur syntaxique de corpus , in Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles, Dourdan, France, 2005.

32. B. Fortuna, M. Grobelnik, D. Mladenic, Semi-automatic data driven ontology construction system . In Proceedings of the 9th International multiconference Information Society IS-2006, Ljubljana, Slovenia, 2006.
33. P. Cimiano, J. Volker, Text2onto - a framework for ontology learning and data-driven change discovery . In A. MONTOYO, R. MUNOZ & E. METAIS, Eds., Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB), volume 3513 of Lecture Notes in Computer Science, p. 227–238, Alicante, Spain, Springer, 2005.
34. The GATE platform: http://gate.ac.uk/
35. D. Ferruci, A. Lally, UIMA: an architecture approach to unstructured information processing in a corporate research environment. Natural Language Engineering, 10(3-4), p. 327–348, 2004.
36. H.-M. Muller, E. E. Kenny, P. W. Sternberg, Textpresso: an ontology based information retrieval and extraction system for biological literature , PLoS Biology, 2(11), 1984–1998, 2004.
37. N. Hernandez, J. Mothe, D'un thesaurus vers une ontologie de domaine pour l'exploration d'un corpus , In Actes de la conférence Veille Stratégique Scientifique & Technologique VSST, 2006.
38. N. Lammari, E. Métais, Building and Maintaining Ontologies: a Set of Algorithms , Data and Knowledge Engineering, 48(2): 155-176, 2004.
39. D. Calvanese, G. D. Giacomo, M. Lenzerini, A framework for ontology integration, In Proc. of the First Semantic Web Working Symposium, 2001.
40. N. F. Noy, M. A. Musen, Prompt: Algorithm and tool for automated ontology merging and alignment, in Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, (AAAI Press / The MIT Press, 2000.
41. P. Buneman, S. B. Davidson, A. Kosky, Theoretical aspects of schema merging, in EDBT '92: Proceedings of the 3rd International Conference on Extending Database Technology, (Springer-Verlag, London, UK, 1992.
42. G. Salton, C. Buckley, Term-weighing approaches in automatic text retrieval. In Information Processing & Management, 24(5): 513-523, 1988.
43. Canadian Thesaurus of Construction Science and Technology: http://irc.nrc-cnrc.gc.ca/thesaurus/.
44. TM4J website: http://tm4j.org/
45. Ontopia Solution – Navigator Framework: http://www.ontopia.net/
46. http://tm4j.org/tmnav.html
47. http ://treebolic.sourceforge.net/
48. M. Ueberall, O. Drobnik, Facet-based Exploratory Search in Topic Maps .Maicher, L.; Garshol, L. M. (eds.): Subject-centric computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, Germany, October 16-17, 2008, Revised Selected Papers. (Leipziger Beiträge zur Informatik: XII) - ISBN 978-3-941152-05-2.
49. M. Zacklad, J. Caussanel, J.P. Cahier: Un méta-modèle basé sur les Topic Maps pour la structuration et la recherche d'information, 2003. http://enssibal.enssib.fr/autres-sites/RTP/websemantique/octobre/octobre4/zacklad.pdf
50. S. Ruda, S. Rudan. SocioTM – Relevancies, Collaboration, and Socioknowledge in Topic Maps. Maicher, L.; Garshol, L. M. (eds.): Subject-centric computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, Germany, October 16-17, 2008, Revised Selected Papers. (Leipziger Beiträge zur Informatik: XII) - ISBN 978-3-941152-05-2.
51. H. Lu, Feng B., An intelligent topic map-based multi-resource knowledge service system. Information & Computational Science Journal, 7(3), p. 657-665, 2010.

52. A. Korthaus, A. Markus, H. Stefan, A distributed knowledge management infrastructure based on a topic map grid, Int. Journal of High Performance Computing and Networking, 6(1), p. 66-80, 2009.
53. J. Reynolds, W.E. Kimber, Topic Map Authoring With Reusable Ontologies and Automated Knowledge Mining. XML 2002 Proceedings by deepX, 2002.
54. H. Folch, H. Habert, Articulating conceptual spaces using the Topic Map standard. Proceedings XML'2002, Baltimore, december (2002), 8-13.
55. K. Böhm, G. Heyer, U. Quasthoff, Ch. Wolff, Topic Map generation using text mining. Journal of Universal Computer Science, 8(6), p. 623-633, 2002.
56. A. Korthaus, C. Köhler, M. Schader, Semi-automatic topic map generation from a conventional document index. Int. Conf. Knowledge Sharing and Collaborative Engineering (KSCE’04), US Virgin Islands, pp.101-108, 2004.
57. M. Mase, S. Yamada, Extracting Topic Maps fromWeb histories by clustering with Web structure and contents. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06).
58. L. Kasler, Z. Venczel, L.Z. Varga, Framework for Semi Automatically Generating Topic Maps. TIR-06. Proceedings of the 3rd international workshop on text-based information retrieval. Riva del Grada, 24-30, 2006.
59. L’H. Zaher, J.-P. Cahier, W.A. Turner, M. Zacklad, A conflictual co-building method with Agoræ. In: Proc. of Workshop on Knowledge Sharing in Organizations, COOP 2006, Carry le Rouet, France 2006.
60. L’H. Zaher, J.-P. Cahier, C. Guittard, Cooperative Building of Multiple Points-of-View Topic Maps with Hypertopic. TMRA 2007: 154-159, 2007.
61. L’H. Zaher, J.-P. Cahier, M. Zacklad, Cooperative Building of a Multi-Points of view Topic Maps with Hypertopic and SeeMe. RCIS 2007: 361-366, 2007.
62. S. Pepper, Methods for the Automatic Construction of Topic Maps, 2002: http://www.ontopia.net/topicmaps/materials/autogen-pres.pdf
63. S. Pepper, Expressing Dublin Core in Topic Maps. TMRA’07, p. 186-197, 2007.
64. S. Roberson, D. Dicheva, Semi-automatic ontology extraction to create draft topic maps, The 45th ACM Southeast Conference, p. 23-24, 2007, March 23-24, 2007, Winston-Salem, North Carolina, USA.
65. G. R. Librelotto, J. C. Ramalho, P. R. Henriques, A framework to specify, extract and manage topic maps driven by ontology. SIGDOC’08, p. 155-162, 2008.
66. T. Neidhart, R. Pinchuk, B. Valentin, Semantic Integration of Relational Data Sources. With Topic Maps. Maicher, L.; Garshol, L. M. (Eds.): Linked Topic Maps. Fifth International Conference on Topic Maps Research and Applications, TMRA 2009 Leipzig, Germany, November 12–13, 2009 Revised Selected Papers. Leipziger Beiträge zur Informatik. ISBN 978-3-941608-06-1.
67. Q. Zheng, Z. Wu, L. Jiang, J. Liu, A collaborative knowledge construction system design for massive knowledge resources. CSCWD’09, pp. 137-142, 2009.
68. G. E. Weber, R. Eilbracht, S. Kesberg, Topic Maps for Improved Access to and Use of Content in Relational Databases – a Case Study on the Descriptive Variety Lists of Germany’s Bundessortenamt, Sixth International Conference on Topic Maps Research and Applications, TMRA 2010, Leipzig, Germany, September 29 – October 01, 2010.
69. L. M. Garshol, M. Fischer, Extending Content Management with Topic Maps – Ontopia/Liferay Integration, Sixth International Conference on Topic Maps Research and Applications, TMRA 2010, Leipzig, Germany, September 29 – October 01, 2010.
70. S. Eslami, E. Nazami, An automatic approach for Topic Maps development using relational databases ICCRD ‘2011, Chine, pp 304-308, 2011.
71. K. Dharavath, S.K. Saritha, Organizing Extracted Data: Using Topic Maps. 11 Eighth International Conference on Information Technology: New Generations. Las Vegas, Nevada

USA, April 11-April 13, pp 1048-1049, ISBN: 978-0-7695-4367-3 By. Issue Date:April 2011. pp. 1048-1049, 2011.

72. J. H. Lau, K. Grieser, D. Newman, T. Baldwin, Automatic Labelling of Topic Models. ACL 2011: 1536-1545, 2011.
73. L. Maicher, H. F. Witschel, Merging of Distributed Topic Maps based on the Subject Identity Measure (SIM) Approach. Proceedings of LIT'04. pp. 301-307, 2004.
74. H.S. Chung, J.M. Kim, Conflict Detection and Resolution in Merging of Topic Maps. International Conference on Convergence Information Technology, pp. 907 – 912, 2007.
75. X. F. Wu, L. Zhou, L. Zhang, Q.L. Ding, TOM algorithm in distributed topic maps merging. Engineering Journal of WuHan University, 39(5), p. 131-136, 2006.
76. J. M. Kim, H. Shin, H.J. Kim, Schema and constraints-based matching and merging of topic maps, Information Processing and Management, 43(4), p. 930-945, 2007.
77. M. Ouziri, Semantic integration of Web-based learning resources: A Topic Maps-based approach. In: Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT 2006), 0-7695-2632-2/06 $20.00 ©, IEEE, Los Alamitos, 2006.
78. H. Lu, B. Feng, Y. Zhao, Q. Zheng, J. Liu, Distributed knowledge management based on extended topic maps. International Conference on Computer Science and Software Engineering, Volume 01, p. 649-652, 2008.
79. H. Lu, B. Feng, An intelligent topic map-based approach to detecting and resolving conflicts for multi-resource knowledge fusion, Information Technology Journal, 8(8), p. 1242-1248, 2009.
80. Y. Xue, W. Liu, B. Feng, W. Cao, Merging of Topic Maps Based on Corpus. International Conference on Electrical and Control Engineering (ICECE), pp 2840 – 2843, 2010.
81. S. Raunich, E. Rahm, ATOM: Automatic target-driven ontology merging, IEEE 27th International Conference on Data Engineering, ICDE 2011, pp. 1276-1279, 2011.
82. F. Hamdi, B. Safar, N. Niraula, C. Reynaud, TaxoMap alignment and refinement modules: results for OAEI 2010, Ontology Alignment Evaluation Initiative (OAEI) 2010 Campaign, ISWC Ontology Matching Workshop, Shanghai International Convention Center, Shanghai, China, Nov. 7, 2010.
83. A. Mechouche, N. Abadie, S. Mustière, Alignment based measure of the distance between potentially common parts of lightweight ontologies, Fifth International Workshop on Ontology Matching (OM 2010), Shanghai, Nov, 2010.
84. S. Sorrentino, S. Bergamaschi, M. Gawinecki, L. Po, Schema label normalization for improving schema matching, Data & Knowledge Engineering 69 (2010) 1254–1273, 2010.
85. J. Kim, P. Kim, Hyunsook Chung, Ontology construction using online ontologies based on selection, mapping and merging, International Journal of Web and Grid Services 2011 - Vol. 7, No.2 pp. 170 - 189, 2011.
86. P. Krbálek, M. Vacek, Collaborative knowledge mapping, September 2011, i-KNOW '11: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, Publisher: ACM Request Permissions, 2011.
87. C.K.Sarumathy, J.Gokulraj, P.Selvaperumal, Integration of Online Ontologies Using Combined Ranking Algorithm, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-6, January, 2012.
88. N. Lammari, S. Besbes Essanaa, Rétro-Conception de Sites Web : Extraction du Contenu Informatif, In Vers l'Ingénierie des Evolutions, Revue des Sciences et Technologies de l'Information série Ingénierie des Systèmes d'Information (RTSI série ISI), 0(0), 2009.
89. http://www.mondeca.com/
90. http://www.topicmap-design.com/
91. D. Dicheva, C. Dichev, D. Wang, Visualizing Topic Maps for e-Learning. Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)
92. D. Dicheva, C. Dichev, TM4L: creating and browsing educational topic maps, British Journal of Educational Technology (BJET), 37(3), p. 391-404, 2006.

93. K. Ahmed, TMShare - Topic Map fragment exchange in a peer-to-peer application. XML Europe conference, London, UK, 2003.
94. E. Godehardt, N. Bhatti, Using topic maps for visually exploring various data sources in a Web-based environment, Third International Conference on Topic Maps Research and Applications (TMRA'07), p. 51-56, 2008.
95. D. De Weerdt, R. Pinchuk, R. Aked, J.J. de Orus, B. Fontaine, Topimaker - an implementation of a novel topic maps visualization. In Proceedings of International Conferences on Topic Maps Research and Applications, Leipzig, 2006.
96. A. Hatzigaidas, A. Papastergiou, G. Tryfon, A Topic Map Editor and Navigation Tool, IADIS International Conference WWW/Internet, 2004.
97. L.H. Zaher, J. P. Cahier, M. Zacklad, The Agoræ / Hypertopic approach, International Workshop IKHS - Indexing and Knowledge in Human Sciences, 2006.
98. D. Damen, M. Patriksson, Putting Topic Maps to REST, Sixth International Conference on Topic Maps Research and Applications, TMRA 2010, Leipzig, Germany, September 29 – October 01, pp. 9-17, 2010.
99. L. Jiang, J. Liu, Z. Wu, Q. Zheng, Y. Qian, ETM toolkit: a development tool based on extended topic map. The 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD'09), p. 528-533, 2009.
100. B. Legrand, S. Michel, Visualisation exploratoire, généricité, exhaustivité et facteur d'échelle. In Numéro spécial de la revue RNTI Visualisation et extraction des connaissances, mars, 2006.
101. D. Dicks, V. Venkatesh, S. Shaw, G. Lowerison, D. Zhang, An Empirical Evaluation of Topic Map Search Capabilities in an Educational Context. In L. Cantoni et C. McLoughlin (Eds.): Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, pp. 1031-1038, 2004.
102. O. S. Gyun, O.N. Park, Design and Users' Evaluation of a Topic Map-Based Korean Folk Music Retrieval System. In L. Maicher, A. Sigel, and L.M. Garshol (Eds.): Proceedings of TMRA 2006, LNAI 4438, pp. 74–89, 2007.
103. P. Haluzová, Evaluation of Instances Asset in a Topic Maps-Based Ontology, Sixth International Conference on Topic Maps Research and Applications, TMRA 2010, Leipzig, Germany, September 29 – October 01, 2010.