ETM Toolkit: A Development Tool Based On Extended Topic Map

Lu Jiang, Jun Liu, Zhaohui Wu, Qinghua Zheng, Yanan Qian

School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, P.R. China
roadjiang@yahoo.com, liukeen@mail.xjtu.edu.cn, wzh@stu.xjtu.edu.cn, qzheng@mail.xjt.edu.cn, yaya213@gmail.com

Abstract

By research on Topic Map standard, the Extended Topic Map (ETM) is proposed as a novel model for better organization and management of the massive knowledge resources in E-learning. Based on the model, an Extended Topic Map Toolkit is designed and implemented, which supports exploration, search, consistency check and etc. The ETM Toolkit not only provides learners with visual navigation and search on massive E-learning resources, but also offers a way for instructors to collaboratively build the shareable and reusable domain knowledge efficiently. By ETM Toolkit, an extended topic map with a certain scale on Computer Networks has been built and is currently available for students in our university.

Keywords: Knowledge Element, Extended Topic Map, EXT M, ETM Toolkit, E-learning.

1. Introduction

E-learning can be defined as training or instruction that is delivered electronically. Typically it uses the Internet to bridge distances and enables people to learn no matter where they are. A crucial issue in e-learning is how to organize the learning content so that learners and instructors can find what they need indeed. To this end, the resource management based on topic map is proposed. Topic maps provide a bridge between the domains of knowledge representation and information management [1]. The basic structure - TAO (Topic Association Occurrence) structure of the topic map is designed to indicate the semantic structure of topics within knowledge resources.

Knowledge elements as the smallest and integral knowledge units in a special discipline [2] widely exist in E-learning resources. Without proper organization of overwhelming knowledge elements, one will easily suffer disorientation and cognitive overload. According to the Constructivism theoretical point of view, the structural knowledge organization based on the knowledge elements and their associations is an effective mean to reduce the cognitive load of learners [3]. Unfortunately, the TAO structure of topic map may not suitable to reflect the associations between knowledge elements while maintaining the relations of concepts because both concepts and knowledge elements are treated as topics indiscriminately in TAO, which may lead to deficiencies in meeting learners’ multi-level demand for knowledge acquisition. To solve this problem, a novel logical organization model of domain knowledge ETM (Extended Topic Map) is proposed in this paper, which is compatible with existing ISO standard Topic Map and allows for efficient management of both topics and knowledge elements.

Based on the ETM model, we implemented a prototype of ETM Toolkit. The software will be of great benefit to both instructors and learners in E-learning. The editor provides a way to help instructors to collaboratively build domain ETM conveniently and correctly. And the navigator facilitates learners’ burden in their studies and assists them finding what they want efficiently.

The rest of this paper is organized as follows: Section 2 introduces the related work. Section 3 defines the formal and XML representation of ETM. Section 3 presents the framework of ETM Toolkit and discusses tree design issues in it. Section 4 shows a prototype of ETM Toolkit. Section 5 concludes the results and discusses the future work.

2. Related work

Resource management techniques can be divided into physical organization technique and logical organization technique. In aspect of the logical organization, which is the key technique to resource management, meta-data based and topic map based management are proposed presently. Meta-data based management such as LRM (Learning Resource Metadata)[4] and LOM(Learning Object Metadata)[5] etc. organize the knowledge resources merely focusing on the external feature such as subject, title and authors, ignoring semantic features residing in resources. Therefore they can hardly provide uses with the semantic knowledge of resources. The concept of Topic Map was first put forward by W3C, and has been
adopted as the ISO standard [6]. A topic map annotates and provides organising principles for large sets of information resources. It builds a structured semantic link network above the resources.

The logical representation of knowledge usually stored in the structured or semi-structured data, which is easily understood by machines rather than human being. Besides, real topic maps usually are very large and may contain thouds of topics and associations. Due to the lack of expression of XML representation, the Graph User Interface (GUI) is widely studied these years. Several GUIs on editing and navigating have been proposed such as TM4J (Topic Map for Java)[7], TM4L (Topic Map for e-Learning)[8][9][10] and UNIVIT (Universal Interactive Visualization Tool)[11] etc. TM4J is an open source toolkit written entirely in Java for creating, manipulating and publishing topic maps. TM4L is Ontology based navigating and editing environment, which is implemented in Java and uses TM4J Topic Map engine. UNIVIT is a 3D Topic Map viewer using virtual reality techniques. These tools are all based on Topic Map model thus still can’t support knowledge elements navigating and editing.

3. Knowledge representation in ETM

In this section, formal representation of the ETM is presented. Based on it, XML representation is brought out. Principles of conversion from formal representation to the easy-to-machine XML document are also studied.

3.1. Formal representation of ETM

**Definition 1:** A domain Extended Topic Map can be represented as a septuple \((C, KE, RC, RKE, \alpha, \beta, \theta)\) where

- \(C = \{c_1, c_2, ..., c_k\}\) \((k > 0)\), refers to a finite nonempty set of concepts or topics within a specific domain. Each concept in set \(C\), in its most generic sense, can be any “thing” whatsoever regardless of whether it exists 0.
- \(KE = \{ke_1, ke_2, ..., ke_n\}\) \((n > 0)\), refers to a finite nonempty set of knowledge elements within a specific domain. Each knowledge element in \(KE\) is the smallest and integral knowledge unit in a special discipline, for example, a definition of computer network, the IP address formats and a routing algorithm [2].
- \(RC = \{part\ of, \ kind\ of, \ instance\ of, \ attribute\ of\}\), is a predefined set of concept association types between topics, including part of, kind of, instance of and attribute of four types of concept association.
- \(RKE = \{reason, \ precondition, \ case, \ reference\}\), is a predefined set of association types among knowledge elements, including reason, precondition, case and reference four types of knowledge element association.

\(\alpha\) is the mapping function from \(C \times C\) to \(RC\), i.e. \(\alpha \subseteq (C \times C \rightarrow RC)\), referring to the associations among concept sets. \(\beta\) is the mapping function from \(KE \times KE\) to \(RKE\), i.e. \(\beta \subseteq (KE \times KE \rightarrow RKE)\), referring to the associations among knowledge elements set \(KE\).

\(\theta\) is the mapping function from \(C \times KE\) to \\{0, 1\}, i.e. \(\theta \subseteq (C \times KE \rightarrow \{0, 1\})\), referring to the associations between concept \(C\) and knowledge element set \(KE\). If knowledge element \(ke_i\) element is on concept \(c_i\), the value of \(\theta(c_i, ke_i)\) is 1, 0 otherwise.

The ETM model extends the representation by adding the knowledge element layer, shown in the figure 1. Compared to TM model, the ETM model is more in line with the learners’ cognitive model and more expressive in managing knowledge elements.

![Figure 1. Comparison between the TM and ETM](image)

3.2. Formal representation of ETM

The formal definition of ETM is accuracy, but can’t be processed directly by machine. Therefore easy-to-machine representation is required. A XML based Topic Map representation XTM (Xml Topic Map) DTD proposed by Newcomb and Biezunski[6], was adopted by the international standard ISO/IEC3250 in 2002. The XTM makes computer qualify to recognize the structure and links in TM model. XTM consists of 19 basic labels including \(<\text{topic}>\), \(<\text{basename}>\), \(<\text{basenameString}>\), \(<\text{association}>\), \(<\text{instanceOf}>\), \(<\text{occurrence}>\) etc.

Designing the XML representation standard of ETM-EXTM (Extended Xml Topic Map) should follow two principles. First, the EXTM should be compatible with existing ISO standard XTM. In the light of this feature, a system based on EXTM is capable for accessing, manipulating and sharing XML files with a system based on XTM. Data interactivity and collaborative work will be carried out smoothly between the system based on XTM and EXTM. Second, the design considerations such as universality, readability and accuracy should be also taken into account in EXTM design. Based on these principles,
four XML expansion labels: <knowledgeElement>, <knowledgeElementRef>, <elementAssoc>, <topicElementAssoc> are introduced in EXTM, representing respectively knowledge element, knowledge element reference, association among knowledge elements and association between a topic and a knowledge element. Compared to the XTM standard, the EXTM makes expansion in the following five points: (The explanation of other labels are defined in [6])

1) Extend triple of topic map to sextuple topicMap (topic|knowledgeElement|association|elementAssoc|topicElementAssoc|mergeMap).
2) Add knowledgeElement label to identify $k_e$. Each knowledge element has the following description labels: zero or one instanceOf, zero or one subjectIdentify, zero or more baseName or Occurrence.
3) Add elementAssoc label to identify $\beta_i$. Each knowledge element association has the following description labels: zero or one instanceOf, zero or one scope, one or more member, and one or more member reference label knowledgeElementRef.
4) Add length attribute to identify the length of $\alpha_i$ and $\beta_i$.
5) Add topicElementAssoc to identify $\theta \in \theta$. Each topic element association has the following description labels: zero or one instanceOf, zero or one scope, one or more member, and one or more member reference label knowledgeElementRef or topicRef.

4. ETM Toolkit

While topic maps offer a powerful and promising technology for intelligent organization and access of information in general, creating topic maps for e-learning is not a clear and simple task at present [11]. Based on the ETM logical representation of knowledge resource, we developed a suite of software, named ETM Toolkit, which is a free software coded in Java, to assist learners and instructors manipulating, sharing, navigating the domain knowledge (ETM Toolkit uses Tough Graph as the graphic engine). In this section we first present the framework of ETM Toolkit and then discuss three design issues in the framework.

4.1. Framework of ETM Toolkit

As shown in figure 2, the ETM Toolkit is designed in layered manner aiming at reducing complexity and improving extensibility. Based on the knowledge representation model discussed above, the logical representation of knowledge which is built upon knowledge resources stored in one or more EXTM documents. From down to top there are 3 layers:

In the function layer, modules directly interact with logic representation of knowledge (EXTM files). Each module in the layer is responsible for providing one basic function. For example, the Local Access module is responsible for saving and opening EXTM files on users’ machine. In the logical layer, modules are responsible for handling the logic of user’s actions. For example, before adding a new topic in ETM, the syntactic and semantic consistency will be checked in this layer. A module in the layer communicates with a module in function layer through function calls. In the user interface layer, modules are served as Graphic Interface which listens to user’s actions and notifies the proper module in logic layer. A user can develop his own user interface in this layer if he follows the interface criterions defined in the layer.

Based on the proposed general framework, the ETM Toolkit includes two typical applications: ETM Editor and ETM Navigator. ETM Editor allows instructors to build domain knowledge based on extended topic map. Products of the Editor are fully compatible with the XTM standard and thus interchangeable and interoperable with any tools based on XTM and EXTM standard. This brings flexibility and expediency to reuse or extension of existing domain knowledge. ETM Navigator facilitates learners’ burden in their study and assists them finding what they want efficiently. Study using the navigator may significantly reduce cognition load of learners.

Generally, there are three critical issues should be considered in designing the ETM Toolkit. First, because an extended topic map may contain millions of concepts and knowledge elements, reducing learners’ information overload through navigation is essential. Second, since instructors may unconsciously make mistakes in editing, each edit operation must be checked for consistency. Third, in order to achieve the goal of collaboratively building the ETM, concurrent management should be performed to maintain the
consistency among versions. In the rest part of the section, we will discuss these issues.

4.2. Information overload mitigation

Intuitively, an excellent navigation interface may significantly reduce cognition load of learners. In the ETM Navigator, the EXTM document is visually displayed as the double layered networks, as shown in figure 3.

![Figure 3. The User interface of ETM Editor](image)

Topics and topic associations are represented in the above topic layer in which each node is regarded as a topic and each edge is regarded as an association of topics, while knowledge element and their associations are in the knowledge element layer. The navigator hires dynamical layout policy \[12\] to display the extended topic map structure.

When a learner selects a knowledge element by clicking the node in the knowledge element layer, it is highlighted in the center and other knowledge elements which are immediately related to the currently selected are displayed. Meanwhile, in topic layer, the topic of selected knowledge element is also highlighted. The transformation carried out on navigator guarantees that only the most relevant information concerned by learner is provided, hence may sharply reduces the information overload. Learners can move the graph as well as re-organize its topological structure. Moreover, search is recommended in the environment. By inputting a topic name, navigator makes an abstract click then the given topic and its related topics and knowledge elements are located on the screen. Once a learner finds what he wants, he can choose one or more useful learning resources related to the topic or knowledge element from resources list.

4.3. Consistency check

In order to maintain the consistency in the ETM, consistency check is preformed at the end of each operation. There are two types of check: the syntax check and the semantic check. The executed order of them is exhibit in figure 4.

![Figure 4. The workflow of edit operations](image)

Syntax check is mainly focus on the correctness and completeness of input data, including:

1) The name of topic or element is not allowed to contain &, <, > and other special symbols.
2) The name of topic or element is not allowed to contain reserved word.
3) The name of topic or element must not be NULL.
4) The association length must be in the range of [10,999].
5) The two roles of any association must be present without ambiguity. 

On the other hand, semantic check mainly focuses on the semantics in knowledge element associations. As mentioned above, there are four types of knowledge element association and only two of them (sequence and case) need to be checked.

**Definition 2:** the sequence set of knowledge element associations \(\beta_{seq}\) \((\beta_{seq} \subset \beta)\) is semantically consistent if there doesn’t exist a sequence of relations \(\beta_i, \beta_{i+1}, \ldots, \beta_n (\forall i \in [1, n], \beta_i \in \beta_{seq})\) such that each pair \(\beta_i, \beta_{i+1}\) has at least one same member knowledge element and \(\beta_1 = \beta_n\).

**Definition 3:** the reference set of knowledge element associations \(\beta_{ref}\) is semantically consistent

\[
\forall \beta_{h} = (k_{e_{j}}, k_{e_{j}}) \in \beta_{ref}, \exists \beta_{h}' = (k_{e_{j}}, k_{e_{j}}) \in \beta_{ref} .
\]

According to the definition 2 and 3, checking sequence associations is to detect the semantic loop in the ETM graph and checking reference associations is to make sure all associations in graph are symmetrical. Reference associations check is quite straight, here we only gives the algorithm (Figure 5) to perform the sequence check in ETM.
Initialization: pre and post are two int array with initial value -1; cnt = 0; cntP = 0; acyclic = true;

Recursion: hasCycle(β)
β = (ke₁, ke₂); pre[ke₂] = cnt++;
For each keᵢ adjacent to the node ke₂
β’ = (ke₂, keᵢ)
IF pre[keᵢ] == -1 hasCycle(β’);{  
Else IF (post[keᵢ] == -1) acyclic = false;
Return; }
post[ke₂] = cntP++;

Sequence Consistency Check: checkSeq(ETM)
For each association βᵢ = (keᵢ, keᵢ) ∈ β_seq  
IF (pre[keᵢ] == -1)  hasCycle(βᵢ); return acyclic;

Figure 5. The algorithm of semantic loop detection

4.4. Concurrent update management

The ETM Editor accepts new versions of the EXTM and reproduces old versions on request. Concurrent update management of EXTM involves two main tasks: version recording and merging. Version recording records the generation tree of versions and merging automatically integrates information from diverse sources into a coherent new topic map. Because ETM benefits from the Topic Maps’ fundamental feature of supporting effective merge of existing knowledge while maintaining their meaningful structure, we only discuss version recording here. (More details on topic map merging can be found in [13])

Definition 4: Each operation provided by ETM Editor can be represented as opᵢ = (id, opName, objᵢ), where id is the owner identification and opName is the operation name and objᵢ is the target object. The modification of EXTM can be represented as Δ = (op₁, . . ., opᵢ).

Version recording records the submitted EXTM documents and applies positive deltas to obtain the newer version. Here we assume the consistency check in editor makes sure that there is no conflict in Δ. When two participants each obtain a copy of the TM₀ and one edits his copy into version TM₁ while the other edits his version into TM₂. According to definition 4 we have TM₁ = Δ₁(TM₀), TM₂ = Δ₂(TM₀). Because TM₁ is not submitted, the server records the Δ₁ rather than TM₁ in generation tree and uses Δ₁ to generate newer TM₁ version on request. Soon, a user may obtain a copy of the TM₁ and edit into version TM₃ and submit the TM₃. This time, the server will put TM₃ rather than Δ₁ in generation tree as in Figure 6.

Figure 6. The generation tree of versions

Based on the discussion above, the workflow of collaborative work in building ETM can be given, shown in Figure 7.

Figure 7. The sequence of collaborative work in ETM

As illustrated above, before a user makes any modification, he must acquire a copy of the newest version of EXTM from the server. After the end of editing process, the user may want his EXTM to be brought up to date. Then the server calls a specific merger to merge the EXTM with the version from which it derive and add the EXTM in the generation tree. After that, the server notifies the user with conflicts (if any) which are to be corrected by hand. This correction may require consultation with the other participants, since some conflicts may be misunderstood among the other members. The submitting operation is allowed only if the EXTM document to be submitted derives from the newest version. Concurrent update management of ETMs is not an easy task. The process may involve different strategies to generate a global ETM. Due to the lack of space we can’t present the detail here.

5. Prototype of ETM Toolkit

Using our ETM Toolkit, we have created an extended topic map on Computer Networks with 310 concepts, 786 knowledge elements and 699 concepts associations and 1823 knowledge element associations. We also have deployed the ETM Toolkit on Internet so that users can learn through their Web Browsers.
Figure 8 is a screenshot from our Internet based ETM Navigator. The left part of the screen is the ETM Navigator and the right part is description of current selected concept and available E-learning resources list on it.

5. Conclusion and future work

In this paper, we present a novel logical organization model of domain knowledge ETM (Extended Topic Map) aiming at managing both topics and knowledge elements effectively. Based on that, we implement a suit of software ETM Toolkit to assist learners and instructors involving in E-learning. Three key technologies in designing are discussed. Finally we deploy the prototype of ETM Toolkit on Internet which is currently available for students in our university.

We only focus on the individual learning in which learners are isolated with each others. Some useful information may be lost in this way. Now, based on our prototype, we are collecting the personalized information and studying a personalized navigator using such personalized information.

Acknowledgement

The research was supported by the National High-Tech R&D Program of China under Grant No.2008AA01Z131, the National Science Foundation of China under Grant Nos.60825202, 60803079, the National Key Technologies R&D Program of China under Grant No.2006BAJ07B06.

References