Towards a Knowledge-Based Model for Instructional Design

Frosina Koceva

Università degli Studi di Genova, Italy
frosina.koceva@edu.unige.it

Abstract. This thesis will discuss a knowledge-based model for the design and development of units of learning and teaching aids. The idea behind this work originates from previous theoretical work on Educational Concept Maps - a logical and abstract annotation system derived from the theories of instructional design. Our work is motivated by the open issues in designing instructional authoring system and from the lack of a well-defined process able to merge pedagogical strategies with systems for the knowledge organization of the domain.

Keywords: Knowledge representation, Instructional Design, Semantic technologies, Topic Maps

1 Introduction

Teaching and learning have undergone profound changes in recent years, partly a consequence of the evolution of learning theories, in part dependent on the development and evolution of network technologies. The emergence of constructivist theories of learning models [1] was accompanied by the evolution of the management of learning processes that have facilitated the dynamics of sharing and co-construction of knowledge[17]. The evolution of this scenario prepared the ground to new challenges to research on issues such as interoperability and reusability of learning materials, accessibility, personalization, the definition of standards, quality, etc.

The impetus for PhD thesis starts from this awareness. The final goal is the definition of a knowledge-based model and the implementation of a software tool for instructional design with specific focus on the educational content designe, to be used in e-learning environments. Our model takes into account the perspectives of development that appear to promise the web today, and is grounded on pedagogical reflection and scientific knowledge we have today.

This thesis addresses the problem of instructional authoring system from different points of view trying to integrate them into a same model. From the pedagogical point of view, the framework of reference is that presented by Stelzer and Kingsley in "Theory of Subject Matter Structure"[8] and later revised in [9]. From the point of view of the representation of the subject matter the reference model is that of subject centric networks with specific focus on the Topic Maps model [10]. From the point of view of technology related works are that carried out by projects and research consortia working on *Topic Maps (TM)* [3]. The real difficulty is the integration between pedagogical and technological aspects in a common tool that is easy to use for teachers and students.

The paper is organized as follows. After the next preliminary section introducing the problem, in Sect. 3 the details about our approach and model will be illuminated, in Sect. 4 we pass from the model to its implementation. Sect. 5 shows the relevancy. The Hypotheses and the preliminary results are found in Sect. 6. Finally the evaluation plan and reflection are given in Sect. 7 and Sect. 8.

2 Problem Statement

The approach proposed in this thesis finds its foundation in the work of those who have addressed the problems underlying the processes of learning and knowledge representation, with particular attention to the semantic web research area.

The idea originates from the analysis of the open issues in instructional authoring systems, and from the lack of a well-defined process able to merge pedagogical strategies with systems for the knowledge organization of the domain. In particular, the plan is to ground the work on the *ECM (Educational Concept Map)* model: a logical and abstract annotation system, derived from the theories of instructional design, developed with the aim of guaranteeing the reusability of both teaching materials and knowledge structures [2]. By means of ECMs, it will be possible to design lessons and/or learning paths from an ontological structure characterized by the integration of hierarchical and associative relationships between the educational objectives. The specific problem I address is a knowledge-based model for the design and development of *units of learning (UoL)* and teaching aids that can find a “suitable” teaching and learning path through an ECM. The learning path is a sequence of concepts characterizing the subject matter under definition (a lesson or an entire course).

The ECM is implemented by means of semantic web standards and technologies [3], and has a two level structure: the level of concept (the domain knowledge) and the level of resources (the information domain). The level of concepts gives an representation of the subject matter where the topics are semantically associated between them and each topic can be associated (level of resources) with one or more resources describing the topic itself (documents, pictures, movies, ...).

The goal of this thesis is to develop a system that assists the teacher for the design of a course curricula by proposing a pliable model of domain knowledge on the base of a course with the aim of guaranteeing the reusability of both the teaching aids and knowledge structure of a single disciplines. The knowledge structure could be: the general one, representing the knowledge domain i.e. the ECM, or a *CCM (Course Concept Map)*, representing the specific user vision of the domain, e.g., specific teacher vision of a subset of the course domain. As to reusability, the ECMs are designed to maintain the concept layer separate from the resources, making it possible to provide courses with the same CCM from the ECM but with different resources. Since we need a representation of the domain that can be seen from different points of view, each view showing a different structure, different set of parts, differently related [15]. It seemed to us that TM are an appropriate abstraction for designing UoL. Furthermore, for the implementation of efficient information search, metadata will be a central component and a pedagogical ontology describing the characteristics of the didactic resource will be defined. In TM metadata can be isolated and stored separately from the object, but still closely connected to the object. Once an educational objective is define the system will assist the design of the course by automatically identifying the "prerequisites", in other words the concept that a student must know before attending a given UoL and the learning outcomes, on base of the relations (see approach). In order to propose to the teacher a possible sequence of topics where each topic can appear only once and cannot be preceded by any of its successors, the system implements a topological order modified algorithm that provides all the possible sequence of topological sorting (see approach). This is possible since between the UoL and between the topics there could be a propedeutic relation (is-requirement-of) which is a unidirectional relation that imposes a precedence relationship that makes the unit of learning an acyclic graph.

3 Approach

ECMs are a formal representation of the subject matter structure in the context of learning environments, and a formal definition of the model is available in [2]. To understand the work of this thesis it is necessary, however, to describe here some concepts. An ECM is a logical and abstract annotation model created with the aim of guaranteeing the reusability of teaching materials, as well as of knowledge structures, and designed taking into account the pedagogical requirements defined by Educational Modeling Language research group [14]. It has been developed by means of an ontological structure characterized by the integration of hierarchical and associative relationships. Firstly, it asks teachers and instructional designers to focus their attention on learners' profile (in particular educational background, learning and cognitive styles) and objectives. Taking into account these elements, the model suggests how to identify, within the discipline's subject matter, the key concepts and their relationships so as to identify effective strategies of contents presentation and to support the activation of meaningful learning processes.

According to that model, a profiled learner has a goal identified by an objective (or a composition of objectives) that is achieved by a UoL, or by a composition of UoLs. The *Course Unit (CU)* is the indivisible union of an objective with its unit of learning and can be composed by creating the tree structure of the course (learning units, sub-learning units, etc.). The course units may be connected each other by means of the *Educational Associations (EA)* that may represent a link or a propaedeutic relationship the units have (see Fig. 1.). In particular, four types of EA have been identified:

- *is-requirement-of*: denoted as $R_{req,}$ identifying a transitive and propaedeutic association between two or more topics, e.g. it may be used with the aim of specifying the logical order of contents;

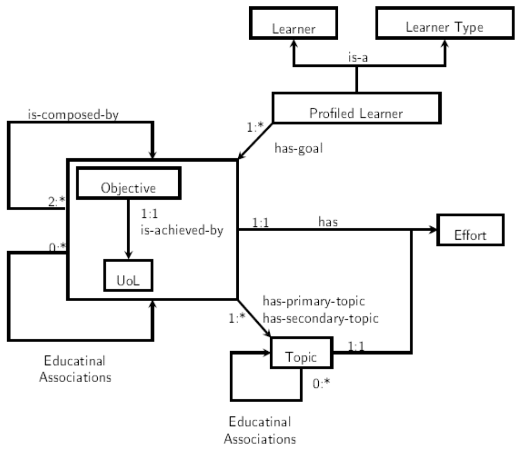

Fig. 1. ENCODE model

- *is-item-of*: denoted as $R_{it,}$ identifying a hierarchical asymmetric association among two topics in order to denote a specific relationship of an individual topic and its more general topic, or a membership relationship;
- *is-related-to*: denoted as $R_{rel,}$ identifying a symmetric association among closely related topics (e.g., it may be used with the aim of creating learning paths without precedence constraints);
- *is-suggested-link-of*: denoted as $R_{sug,}$ identifying an asymmetric association among a topic and it's in-depth examination, e.g. this relationship type may be used in order to suggest in-depth resources.

These relation types have been defined with the aim of allowing teachers to create different learning paths (with or without precedence constraints among topics). The same types of relationship can be found between topics. The latter are the smaller granularity of the ECM model. They represent the concepts of the domain: any subjects a teacher may want to talk about. Moreover, the UoL are connected to the topics through two relationships:

- *has-primary-topic*: where a *primary topic PT* identifies the "prerequisites" of a CCM, in other words the concept that a student must know before attending a given unit of learning;
- *has-secondary-topic*: where *secondary topic ST* identifies the concepts that will be explained in the present unit of learning (this kind of topics will have specific learning materials associated).

4 From the model to its implementation

The ECM model is the theoretical framework for the design of our so called *ENCODE – ENvironment for COntent Design and Editing* application, currently in the implementation phase, with some innovative features described in the following:

1. The possibility to generate a linearized path, useful, for example, for a teacher to produce a lesson or a document about a given subject matter. In this latter case, a *Suggested Paths Strategy* is necessary, to be expressed by means of is-requirement-of relationships.
2. The possibility to publish a CCM on the Web and the relationships suggest the different navigation strategies of the underlying subject matter.

To explain the strategy behind the Suggested Paths Strategy, let us also consider the idea of preparing a lesson on a given subject matter, using the previous ECM model.

The R_{req} (is-requirement-of) relationships order the topics T of the lesson according to the propaedeutics rules, therefore in the graph $G=(T, E)$ there cannot be loops, thus obtaining a Direct Acyclic Graph (DAG), where T are nodes and E arcs, with: $(t_i,t_j) \in E \leftrightarrow R_{req}(t_i, t_j)$. In this context, a *Topological Order* on a CCM is a sequence $S = \{s_1, s_2, \ldots s_{|T|}\}$ where each element T appears only once and cannot be preceded by any of its successors; given pair of nodes (t_i,t_j) in S if there exists an arc from t_i to t_j of type is-requirement-of, it follows that the node t_i is before the node t_j in the list: $\forall(t_i,t_j) \in S$: $(t_i,t_j) \in E \rightarrow i < j$.

The algorithm implementing the Topological Order is derived by Topological sorting algorithm [7] with a main modification in order to get all the possible sequences of topological sorting. Therefore we let the teacher to choose which of this sequences better answers the accomplishment of the didactic objectives. The result of topological ordering is a XML structure that can be imported in a text editor for further adaptations.

Furthermore, for better presentation of the knowledge structure and effective navigation a cluster with a name N_c is defined, by grouping all the topics t_i that are in a is-item-of relationship, i.e. $R_{it}(t_i, t_j)$ with a common topic t_j. More formally we define a cluster $C=\{N_c, T_c\}$ as a non-empty finite set of topics T_c, where $\forall t_i \in T_c$ and $\exists t_j \notin T_c$ where $R_{it}(t_i, t_j)$.

For as much as the topics are topologically ordered this doesn't take into account the distance factor in between the topics, thus a signaling (denoted as *Topic Aider - TA*) is introduced in the sequence S before the distant topic to remind its subject. The TA is a suggestion for the teacher to introduce an exercise, an example, a text or a valuation test. This TA is also reported in the final sequence in order to highlight not only to the teacher, but also to the student the place where s/he should evoke a determinate argument. The choice to have not a single path but a list of paths to suggest to the author leaving the final choice to the author him/herself, is also to answer to the non-equifinality problem posed in [15]. The "suggested" order lists is on the basis of the principle of reducing as much as possible the distance between two topics of the list that are contiguous on the graph.

In order to implement such a model, TM has been chosen. TM is an ISO multi-part standard [3] designed for encoding knowledge and connecting this encoded knowledge to relevant information resources. The standard defines a data model for representing knowledge structures and a specific XML-based interchange syntax, called *XML Topic Maps (XTM)* [4]. The main elements in the TM paradigm are: *topic* (a symbol used to represent one, and only one, subject), *association* (a relationship between two or more topics) and *occurrence* (a relationship between a subject and an information resource). Therefore, two layers can be identified into the TMs paradigm:

- the *knowledge layer* representing topics and their relationships, allowing to construct the ECM model;
- the *information layer* describing information resources, to be attached to the ECM topics.

Each topic can be featured by any number of *names* (and *variants* for each name); by any number of *occurrences*, and by its *association role*, that is a representation of the involvement of a subject in a relationship represented by an association. All these features are statements and they have a *scope* representing the context a statement is valid in. Using scopes it is possible to avoid ambiguity about topics; to provide different points of view on the same topic (for example, based on users' profile) and/or to modify each statement depending on users' language, etc. Therefore, to solve ambiguity issues, each subject, represented by a topic, is identified by a *subject identifier*. This unambiguous identification of subjects is also used in TMs to merge topics that, through these identifiers, are known to have the same subject (two topics with the same subject are replaced by a new topic that has the union of the characteristics of the two originals).

ENCODE application is designed in a layered manner aiming at reducing complexity and improving extensibility. The implementation of the application is based on extending and reusing already available modules, i.e. from the wandora engine [6] for the topic management and of an external open-source text editor for the Semantic Structure Text Editing. In Fig.2 is shown the conceptual architecture of ENCODE.

ENCODE user interface allows a user to build/import/export/visualize domain/course knowledge. The main end user is the Teacher that uses the tool as a Map Designer; Resource Creator or Corse Consumer. The Corse Consumer benefits from the tool by using the linear path of the CCM imported in the Semantically Structured TextEditor as an educational material to use it at school; by using the html format of the CCM to publicate the course materials; importing the XTM format of the structure of the course in an LCMS (i.e., moodle). Other end users are the authors of books or research papers, and in the end the students for explorative studying and/or for the design of concept maps. Thus, ENCODE User Interface layer includes: Editor that import ECM/CCM in XTM format; exports ECM/CCM in a XTM format; generates ECM/CCM web pages for a web view navigation of the knowledge i.e. ECM/CCM Navigation; uploads ECM/CCM to the ECM/CCM web Repository; imports the linearized ECM/CCM in an semantically structured text editor for further editing. Having a user-friendly interface and a graph representation of the subject matter is the main requisite for this layer.

The Logical Layer handles the user's actions and in order to maintain the consistency in the ECM/CCM, performs consistency checks and reasoning for the *TAO (Topic Association Occurrence)* operation. The consistency checks can be of a syntactic and semantic nature. A schema is needed that contains all the information necessary for the validation process, i.e. the ECM is validated against a template ECM

Schema. The ECM Schema is defined by *TMCL (Topic Maps Constraint Language)* [16], which does not have any syntax of its own, since it is defined simply as a Topic Maps vocabulary.

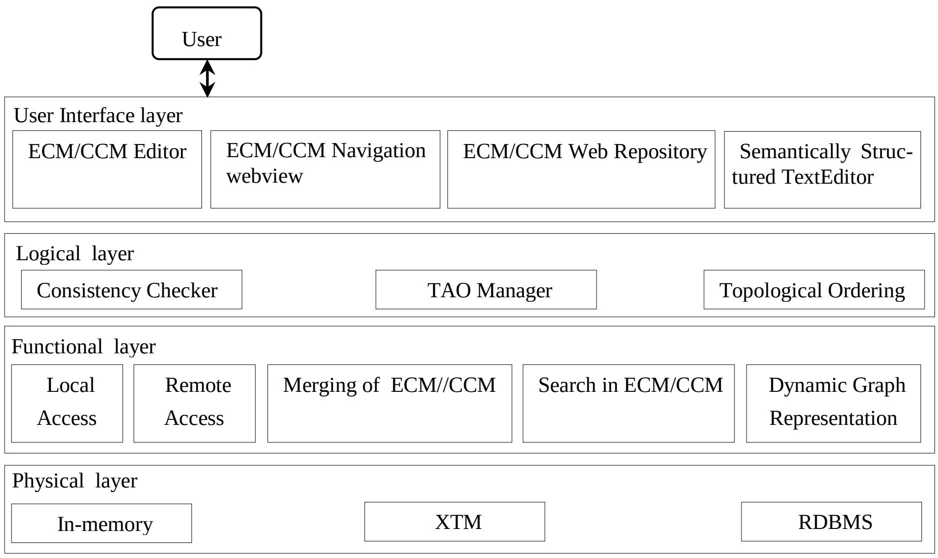

Fig. 2**.** ENCODE application

However, a number of CTM templates are defined in this International Standard in order to facilitate authoring of TMCL schemas using *CTM(Compact Topic Maps syntax)*. Thus a ECM.ctm was defined. The ECM.ctm restricts the Topic Map model to education topic type; four education association types specifying the relation properties (symmetric, transitive, asymmetric) and the education role types involved. TMCL is firmly meant for validation, and not for reasoning. Consistency check are made for:

- detecting unreachable topic $t_i \in$ CCM: i.e. t_i is unreachable if $t_i \notin$ PT and for $\forall R_{req}(*,t_i)$ $\nexists path(t_{pt},*)$ where $t_{pt} \in$ PT. Formally a $path(t_i,t_j)$ among two topics t_i and t_j is defined by a direct relation $R_{req}(t_i,t_j)$ in case of adiacent topics (in this case the "length" of a path is 0), or by the recursive definition $R_{req}(t_i,t_u) \Lambda\ path(t_u,t_v) \Lambda\ R_{req}(t_v,t_j)$ in case of not adjacent topics (in this case the "length" of a path is define by the number of topics between t_i, and t_j). It's worth nothing that between two generic topics t_i and t_j more than one path can exist.

- On adding a is_requirement_of association $R_{req}(t_i,t_j)$ where $t_i,t_j \in$ ECM or $t_i,t_j \in$ CCM, loop verification is needed starting from topic t_i, where a loop is formally defined as $path(t_i, t_i)$ with "length" > 0;

- The deletion of a topic t_i demands deletion of all the associations in which the topic t_i is involved. After deletion a check on possibly unreachable topics is needed. If there is a path of "length" 1 , i.e., $R_{req}(t_u,t_v)\ \Lambda\ R_{req}(t_v,t_k)$ than on deletion of the t_v a $R_{req}(t_u,t_k)$ is created.

- The topic t_j that is an in-depth examination of some other topic it can't have exiting arcs, i.e. for $\forall R_{sug}(t_i,t_j)$ $\nexists R(t_j, *)$ where $R \in \{R_{sug}, R_{in}, R_{req}, R_{rel}\}$.

- There can be only one relation between two topics, i.e. for every $t_i,t_j \in$ ETM or $t_i,t_j \in$ CCM $\exists! R(t_i, t_j)$ where $R \in \{R_{sug}, R_{in}, R_{req}, R_{rel}\}$.

Reasoning is adopted in order to keep the CCM consistent from the propaedeutic point of view. During the CCM creation, after user chooses the desired topics, ENCODE fills up the CCM recursively with the missing preparatory topics starting from the added topics, i.e. if $\exists R_{req}(t_i,t_j)$ where t_i, $t_j \in$ ETM and $t_j \in$ CCM, than ENCODE adds t_i to the CCM.

The Functional Layer modules directly interact with the logical representation of the knowledge in order to save the project locally; upload it to the server repository; merge of ECM/CCM; visualize and adds associations to the concept map graph.

The Physical Layer represents the backend of ENCODE, where the knowledge base is saved: during editing in-memory; permanently as XTM files locally or remotely; or in a non thread-safe RDBMS.

5 Relevancy

The problems faced in this thesis are still open issues in instructional authoring systems, since there is a lack of a well-defined process able to merge pedagogical strategies with systems for the knowledge organization of the domain. By means of the logical and abstract annotation model of ECMs, it will be possible to design lessons and/or learning paths (see previous section). Once an ECM for a subject matter is defined by a teacher, the design of a lesson (for the teacher) and the navigating through a learning path for a student become a problem of topological sorting (on a graph) [7]. The ability to adaptively sort the ECM becomes a powerful tool both for teachers, during the instructional design phase, and for students, during the learning phase.

Indeed, once an ECM is defined, the teacher can design a lesson adapting it on the previous background of its class, and a student can personalize the learning path depending on its specific knowledge and skills.

6 Hypotheses and preliminary results

It's hypothesized that the availability of “sound” knowledge-based tools will increase the productivity of teachers (time and quality) in the daily process of instructional design.

This being said the system is in the initial stage of implementation. The decision on the implementation framework to use for the development was conditioned on the usage of an open source framework that implements the TM standard possibly with active community and an open source text editor. At first we focused on testing two open-source tools, Ontopia [5] and Wandora [6]; then we opt for building the system on top of Wandora being a well-established topic maps creation tool in continuous evolution, with good reputation and with a powerful and flexible graphical presentation tool. The system besides the Wandora engine for the TAO operation, has TM validator, TM repository. Also further functionalities are in process of implementation for the topological ordering and the assistance in the building of the CCM.

7 Evaluation plan

During the implementation of the system particular attention will be paid to the design and implementation of the user interface. There is a plan during the second half of the next year of my PhD course to experiment the prototype of the system within a selected group of teachers of the EPICT community (www.epict.it), a large community of teachers of the Italian secondary schools. The plan is to measure both the usability of the user interface and the instrument's effectiveness in terms of improving the work of the teacher. In particular, it will seek to evaluate the improvement of daily activities of instructional design carried out by the teacher in terms of both the reduction of design time, and of increased efficacy of the process of instructional design.

I will prepare questionnaire to collect quantitative data, deepen then the results with focus groups.

The experience of teachers with this system will be compared with the previous experience of the same teachers.

8 Reflections

The idea behind this thesis has been stimulated by the real needs of a community of teachers to have model and tools that facilitates some phases of instructional design. Since the concept representation is independent of its implementation, ECM lends itself for reusability of both teaching materials and knowledge structure. Thus the knowledge structure could be reused for the design of a different course according to the learner target. From student point of view, the subject-centric nature of the TM help learners to identify core concepts, while the extended TM with the learning path assists the student for proper order sequence of studying. Moreover, the underlying model, ECM, is grounded on pedagogical reflections. For these reasons we believe that this model will have a good acceptance by the community of teachers we plan to select for the testing phase.

Acknowledgments.

I would like to express my very great appreciation to my PhD Advisor Prof. Giovanni Adorni, acknowledging his valuable ideas, guidance and support during the planning and development of the ECM model and this research work. I am also particularly grateful to my paper supervisors, Prof.Dr. Judy Goldsmith and Dr. Nick Mattei, for all the useful and constructive suggestions.

References

1. Bodner G. M.: Constructivism: A theory of knowledge , J. of Chem. Education, 1986, 63: 873-878.
2. Adorni G., Brondo D., Vivanet G.: A formal instructional model based on Concept Maps, J. of E-Learning and Knowledge Society, 2009, 5(3): 33-43.
3. ISO/IEC 13250-2:2006 Information Technology -- Topic Maps -- Part 2: Data Model. Available at: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=40017
4. Garshol, L.M., Graham M.: Topic Maps - XML Syntax. Final Draft International Standard, 2006. Available at: http://www.isotopicmaps.org/sam/sam-xtm/
5. Ontopia Project, Available at: http://www.ontopia.net/
6. Wandora Project, Available at: http://wandora.org/
7. Kahn A.B.: Topological sorting of large networks. Communications of the ACM, 1962, 5(11): 558-562.
8. Stelzer J., Kingsley E.H.: An Axiomatic Theory of Subject Matter Structure. Human Resources Research Organization, Alexandria, Virginia (1974).
9. Adorni G., Di Manzo M., Frisiani A.: Evaluation of a formal approach to the structuring of subject matter. J. of Computer Based Instruction, 1981, 2: 35-42.
10. Weber G.E., Eilbracht R., Kesberg S.: Topic Maps as application data model for Subject-centric applications. In: Maicher L, Garshol L.M. (eds.), Procs. 4th. Int. Conf. on Topic Maps Research and Applications, Leipzig, Germany, 15–17 October 2008.
11. Pepper S., Vitali F., Garshol L. M., Gessa N., Presutti V.: A Survey of RDF/Topic Maps Interoperability Proposals. W3C Consortium Working Draft. Available at: http://www.w3.org/TR/rdftm-survey/
12. Garshol, L.M.: The RTM RDF to topic maps mapping: Definition and Introduction, 2003.
13. Shiladitya Munshi, Ayan Chakraborty, Debajyoti Mukhopadhyay: A Hybrid Graph based Framework for Integrating Information from RDF and Topic Map: A Proposal, 2012.
14. Koper R.: Modelling Units of Study from a Pedagogical Perspective: the pedagogical metamodel behind EML. Heerlen: Open Universiteit Nederland, 2001 (http://dspace.learningnetworks.org/handle/1820/36?mode=simple).
15. Ohlsson S.: Some principles of intelligent tutoring. In Lawler, R.W., and Masoud Yazdani, M. (eds.), Artificial Intelligence and Education: Learning Environments and Tutorial Systems v. 1, Intellect Books, 1987.
16. TMCL Final Draft International Standard, Available at: http://www.isotopicmaps.org/tmcl/tmcl.html
17. Vygotsky L S 1962 *Thought and Language*. Harvard University Press, Cambridge, MA