Authoring Educational Topic Maps: Can We Make It Easier?

Darina Dicheva and Christo Dichev
Winston-Salem State University, Winston-Salem, N.C. USA
{dichevad, dichevc}@wssu.ed

Abstract

In this paper we examine some problems related to capturing the structure and the topic name space of learning content in the context of Topic Map authoring. We demonstrate that the use of traditional course taxonomies as ontological resources is problematic. Based on these findings and on the results of a locally conducted study, we propose an empirically justified minimal ontology for Topic Maps-based e-Learning. The proposed minimal set contains five relation types and is implemented in the TM4L Editor to support authors that experience difficulties in articulating and naming relationships.

1. Introduction

Current Web-based educational practices indicate that courseware authors' ability to gather and generate information exceeds their ability to organize, manage, and effectively use it. Ontologies are a key technology emerging to facilitate Web information processing by supporting semantic structuring, annotation, indexing, and search. Ontologies allow organization of learning material around components of semantically annotated topics. This enables ontology-based educational systems to do efficient semantic querying and navigation through the learning content. We have developed an authoring tool, the TM4L (Topic Maps for Learning) Editor [1][2], which enables the development of ontology-aware courseware based on the new Semantic Web technology Topic Maps [3].

In the Topic Map (TM) paradigm an ontology is an accurate description of the essential entities and relations which are found in the modeled domain, and can be represented as a set of *topics* linked by *associations*. Therefore the Topic Maps technology is well suited for structuring learning material around subject ontologies.

We are currently experimenting with the TM4L Editor by creating educational Topic Maps for different university courses. In this paper we discuss the main problems authors face when creating ontology-based courseware and propose strategies for overcoming some of the TM authoring problems.

2. Difficulties in authoring educational TM

The authors of educational Topic Maps are typically instructors who create maps for the courses they teach. In order to find out what are the major difficulties they face we conducted a study in which seven TM were created with the TM4L Editor by different authors. The study showed that authors generally didn't have problems in selecting appropriate learning content and resources. However, they had difficulties in structuring the content and defining and naming relationships between concepts (topics). In the next sections we discuss these problems.

2.1. Classifying learning content

In a typical courseware structure, learning content is laid out in a tree-like structure of course units (lectures, sections, subsections etc), an approach adopted from the traditional textbooks organization. In general, the names of the course units have some relationships based on the author's notion of classification. It is appealing to think that the course units can be organized based on a hierarchy of concepts (terms), like a taxonomy. However, this is not true in general. Frequently the concept structure used to organize the learning content is not a proper hierarchy, and the concepts naming the sub-units of the learning material do not represent more specialized content of their "parents". For example, in the Prolog book of Sterling and Shapiro the topic "Lists" is under "Recursive programming" while in Bratko's book it is under "Lists, Operators, Arithmetic". This shows that the order in learning content classifications is often subjective and arbitrary and could easily be reversed. However, if the order could be altered and still make sense then that topic structure doesn't represent a true hierarchy. Apparently this subjective approach encourages ad hoc concepts organizations. Moreover, authors' intuition about where to place a unit sometimes is inconsistent with the broadly adopted structure. This may result in putting a unit in an unexpected for learners place. When several authors are involved, it is even harder to keep a consistent organization of the instructional material. An additional complication is that both the names and organization of concepts are subject of change over time. Even the ACM Computing Classification System has been changed several times since its first publication.

2.2. Identifying topics

The structure of the learning content usually reflects the author's concept of systematization. The titles (topic names) and their relations depend on authors' knowledge and goals at the time of creation. Being subjective they are of variable quality and with uncontrolled terms. There are no rules to limit the authors to use specific information for describing content. Uncontrolled vocabularies make it easy to record information but shift the load of interpretation to the users.

Among the principal problems with identification and naming topics are:

- The titles (topic names) are not necessarily unique.
- Generally there are no agreed terms for all topics.
- The titles may not be informative enough.
- The titles fail to group related materials together in a more or less standard way.

Further practical questions include: What is the relationship of title subject and the actual content? How does one phrase come to present a multitude of subjects? These questions reflect problems related to browsing and searching for relevant resources. An example in this context is the use of different titles to represent the same topic. In such cases, a search for "”Pattern Matching” will not pick up items indexed with the term “Unification”.

On the other hand, the title is an identifying label by which we refer to the subject. The extension of titles' original use in the new context of e-learning requires a stable, durable topic identification system.

2.3. Articulating and naming relations

As we already mentioned, TM authors have difficulties in deciding what type of relationships to use and how to name them. Authors generally try to follow the content structure of the used textbook, which might not be taxonomy: any given topic may represent a subclass, an instance, a property, a “See also” relationship, etc. Typical collections of learning content are a somewhat incoherent combination of taxonomies, partonomies, and other (possibly unnamed) schemes. The authors shared that often they would encounter a relation between two concepts of the kind “more general - less general” but could not determine its exact type, not to speak about a name. When the relation between the topics was ‘hierarchical’, they would place the new topic as a child of an earlier created (parent) topic in the topics hierarchy (in TM4L representing “class-subclass” relations). The author of the Topic Map “Basic Counting Principles”, for example, used only the default hierarchical type beside the relation “created by” for expressing all relations between the defined topics.

The parent-child topic classification reflects the title-subtitle tradition established by the conventional textbook organization and the created TM-based learning material mirrors the context in which authors used to see the composing items. The problem though is that an organized collection of learning items often represents a kind of contextually related topics which is difficult to translate into conventional hierarchical structures. Another problem comes from the fact that in a TM browser only one hierarchical relation type is usually displayed as a tree. Therefore only the topics connected with relationships of that type would be displayed in the tree. All other topics will be seen not linked to it. Thus, the authors cannot actually see all created topics arranged nicely in a tree-like structure. An attempt to “fix” this problem brought the author of the “Number Systems” TM to define two different relationships instead of one between many topics – “instance-of” which she felt is the proper one and “class-subclass” just to display the topics in the domain term (ontology) hierarchy.

When a relationship between two topics is not ‘hierarchical’, the problem is even worse since the author has to decide also how to name it. Unlike concepts, which are generally named by terms from the subject domain, there are no established/agreed names for relations. The latter are usually named by common language words and there is a variety of words that can express their meaning. The choice of different words by different authors poses though a serious problem for the reusability and exchangeability of created courseware.

3. Support for authoring educational TM

Educational Topic Maps authors are generally untrained in information classification and work in lack of controlled vocabularies and support from ontology analysts. We decided to support the TM4L users in two ways: (1) Help them reuse existing established classifications (combined with controlled vocabularies); (2) Make available some predefined relations to them.

Here we propose an empirically justified minimal ontology for expressing contextual relationships in educational Topic Maps. By predefining the minimal set of relation types we enable authors that experience difficulties in articulating and naming relationships to use the predefined general relationships. Our model contains two layers: *standard* and *author's* layer. The standard layer contains five predefined generic relationships: the classic “superclass-subclass” and “class-instance”, one general “hierarchical” and two “horizontal” relationships. The author’s layer might contain specialization of the general relations defined by the author.

3.1. Modeling general hierarchical relations

Our goal was to define a *generic* relation that captures the current practice of organizing learning material in a structure of topics similar to a *table of content*. Following the work on SKOS [4], we suggest extending the Topic Map model for the needs of the educational TM with a new relation, which we call “super-sub”**.** This relation carries weak semantics used to express the fact that one topic *is more general* than another. This implies that the related topics can be arranged into a hierarchy, without being too strict about the exact semantics of the relationship. The proposed relation is a generalization of “part-whole” and can be used to characterize associations with *asymmetric* roles assigned to two role players, such as X *is-part-of* Y, X *is-based-on* Y, etc. “Super-sub” is not a version of “superclass-subclass” or “class-instance” relationships: it is *less restrictive* than these two. The most informative properties of our “super-sub” relation derive from *transitivity* used also in type hierarchies such as “instance-of” and “class-subclass”. However it is more general and defined as an *asymmetric, transitive relation* satisfying the condition *if (a R b) then not (b R a).* The insight was to provide a generic, extensible hierarchical relation, intended to serve as a type of a family of relationships with more specific meaning where the semantics of the new relations is inherited from the “super-sub” relation. The “super-sub” relationship captures in generic sense hierarchical relationships within the modeled domain, such as “part-whole”, “section-subsection”, “folder-subfolder”, “based-on”, “supported-by”, etc. Thus, instead of classifying concepts such as “facts”, “rules” and “queries” as instances or subclasses of “Prolog”, authors can use the predefined relation “super-sub”.

3.1. Modeling “horizontal” relations

If semantically similar resources are scattered among distinct topics, a simple hierarchical browsing will not find them, while the TM technology allows finding references that are *horizontally* related and unexpected. We propose two *horizontal* relationships between topics: “relevant to” and “mentioned by”. The first one is introduced to capture relations with *symmetric* roles assigned to two role players, such as “co-refers”, “is similar to”, “is synonym of”, “is of same level of complexity”, etc. It can be used to represent mapping between *equivalent* topics. The relation satisfies the condition *if (a R b) then (b R a).*

The second relation “mentioned by” represents the family of *asymmetric, not necessary transitive* relations. It is used to express the fact that two topics are related, but the relationship can not be used to build hierarchies. The relation “mentioned by” is intended to capture in a generic sense asymmetric relations of the type X *mentions* Y, X *refers-to* Y, X *discusses* Y, X *is used by Y*, X *is created by* Y etc. The key relations in the pragmatic sense are the binary relations. The basic intuition is that the five proposed relations (“superclass-subclass”, “class-instance”, “super-sub”**,** “relevant to” and “mentioned by”) represent a sufficient basis of generic relations for creating educational Topic Maps. They can be used as a generic grouping of concepts and resources that might be difficult to articulate relieving the authors from creating dummy or incorrect relation types.

4. Conclusion

We have created an authoring tool, the TM4L Editor that supports the development of standards-based ontology-aware online learning materials. In this paper we discuss some problems authors face when creating ontology-based courseware with TM4L and a strategy for overcoming one of them. We propose an empirically justified minimal relation set for TM based e-learning in an attempt to support authors that experience difficulties in articulating and naming ontology relationships.

5. Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. DUE-0333069 “NSDL: Towards Reusable and Shareable Courseware: Topic Maps-Based Digital Libraries.”

6. References

[1] Dicheva, D., Dichev, C. “A Framework for Concept-Based Digital Course Libraries”, *J. of Interactive Learning Research*, 15(4), 2004, pp. 347-364.

[2] Dicheva D., Dichev C., Sun, Y., Nao, S. “Authoring Topic Maps-based Digital Course Libraries”, *SW-EL@AH 2004*, The Netherlands, 2004, 331-337.

[3] Biezunski, M., Bryan, M., Newcomb, S.: ISO/IEC 13250:2000 Topic Maps: Information Technology, www.y12.doe.gov/sgml/sc34/document/0129.pdf.

[4] SKOS Core Vocabulary Specification, http://www.w3.org/2004/02/skos/core.