Ontology-based document-driven memory for e-Learning

Emil M. Popa, Daniel Hunyadi, Iulian Pah
University “Lucian Blaga” of Sibiu
University “Babes Bolyai” of Cluj Napoca
5-7 Dr. Ioan Ratiu Street, Sibiu
ROMANIA

Abstract: - E-learning leads to evolutions in the way of designing a course. Diffused through the web, the course content cannot be the direct transcription of a face to face course content. A course can be seen as an organization in which different actors are involved. These actors produce documents, information and knowledge that they often share. We present in this paper an ontology-based document-driven memory which is particularly adapted to an e-learning situation. The utility of a shared memory is reinforced in this kind of situation, because the interactions do not usually occur in the same place and in the same time. First we precise our conception of e-learning and we analyze actors needs. Then we present the main features of our learning organizational memory and we focus on the ontologies on which it is based. We consider two kinds of ontologies: the first one is generic and concerns the domain of training; the second one is related to the application domain and is specific to a particular training program. We present our approach for building these ontologies and we show how they can be merged.

Key-Words: - E-learning, Ontology, Organizational memory, Topic maps

1 Introduction

Information Technology has already transformed the way people work and has an increasing impact on the long life learning. New approaches mainly based on the utilization of web technologies are proposed. They often refer to the concept of “e-learning”. Unfortunately, the term e-learning is used to designate various types of situations such as administrative course management, web-based learning, or videoconferences. Usual e-learning definitions put the emphasis more on network utilization and pedagogical content than on distribution of courses. Numerous documents resources may be used during e-learning. Some are internal and made by several actors implied in the e-learning. Others are available on the web: on-line courses, course supports, slides, bibliographies, frequently asked questions, lecture notes, etc. The increasing number of available resources is a real problem in content management systems.

Research work on the Semantic Web aims at addressing this kind of problem. The Semantic Web is an extension of the current web in which information is given well-defined meaning, enabling computers and people to better work in co-operation [4]. The idea is to represent web data, to define and link them so that they can be used for more effective discovery, automation, integration, and reuse across various applications. The Web can reach its full potential if it becomes a place where data can be shared and processed by automated tools as well as people. Sharing data is one of the basic principles the Semantic Web will operate on. In order to be able to exchange the semantics of information, one first needs to agree on how to explicitly model it. Ontologies are a way of representing such formal and shared information. They can be used to index data indicating their meaning, thereby making their semantics explicit and machine-accessible.

In the DANTE project, which relates to this research field, we propose to consider an e-learning training as an organization and to manage the resources of this organization by the means of an ontology-based “learning organizational memory”[1]. This memory allows, on one hand, to capitalize the learning knowledge, and on the other hand, to better index resources, taking into account the learning context.

2 e-Learning

Our conception of e-learning

The term ‘e-learning’ is currently very used and refers to various notions such as logistic (administrative management), resources (course broadcasting) or technology (virtual conference tools). Numerous definitions of e-learning have been proposed. They usually put the emphasis on network utilization (explaining the « e » in e-
learning) and on Information Technology. E-learning must not be reduced to the use of new technologies to serve old learning modes. It is supposed to lead to new learning forms. This implies some consequences. For example, e-learning needs at least:

- A reflection on the content: goals, concepts to study, competences to acquire, etc.
- A reflection on the content organization: relations between learning concepts,
- A construction of new resources taking into account possibilities offered by Information Technology: direct digitalization of old resources is not sufficient,
- A redefinition of actors (teachers, learners) roles.

Within the DANTE project, we are interested in the building of a pedagogical content under a granular form represented by ontology of concepts. Users must have free access to this ontology. Indeed, we consider that the learner must have an active role in his learning. Available documents are not simply transcription of classical courses. They consist in a set of resources that intend to be easy-to-access because of their indexation by the ontology of learning domain concepts. The courses we deal with are scientific courses taught at university.

Use scenario

In our conception of e-learning, knowledge and information structuring is central as well for learners as for teachers. The ontology-based organizational memory we propose aims at helping them to structure and manage knowledge related to a given course or training unit. It relies on an organization model of this course unit and takes into account teachers and learners viewpoints.

In an e-learning situation, learners are often geographically distant. It is thus necessary for them to have an easy access to documents and more generally to resources they need. But because of the distance, they often need to get in contact and to dialogue with teachers and with other learners. Furthermore, certain types of activities (such as practical work) explicitly require cooperation between students.

During training, learners are often led to ask questions regarding the content of a course. For example: What are the goals of this lesson? What are the notions to be learnt? What are the prerequisites? Is there any order in these notions? Are there any documents to consult (slides, books, etc.)? What is it possible to do in order to improve a lesson? Is there any web site, newsgroup dealing with this lesson?

During training, students have often to produce documents that are sent to teachers for evaluation or that are kept. In this last case, documents can be for example work or synthesis documents or annotations. The students can decide (or propose) later to make these documents available for other users. It is therefore useful to allow the attribution of different grants to documents.

The definition of a shared vocabulary is a key point in order to facilitate access to documents, dialogue with teachers and collaboration with other learners.

Learning organizational memory

A course unit is based on knowledge and competencies it should provide, on actors (learners, instructors, trainers, course designers, administrators, etc.) and on resources of different types (definitions, exercises with or without solution, case studies, etc.), and different forms (reports, books, web sites, etc.). In this sense, a course is an organization.

A common approach to tackle the knowledge management problem in an organization consists in designing an organizational memory. Such a memory can be seen as “an explicit and persistent representation of knowledge and information in an organization, in order to facilitate their access and reuse by members of the organization for their tasks” [9].

An organizational memory allows capitalizing not only pedagogical resources related to the contents of the course but also information on actors themselves (specificities, background, profile, etc.). It allows administrative management (registration, notes, etc.) of the course too.

In order to share information in an organization, actors have to use a common terminology, especially when they are geographically distant. A given word or expression must have the same meaning for everyone. It is one of the reasons why organizational memories are often based on ontologies.

Organizational Memories and Learning Organizational Memories

A learning organizational memory is different from an organizational memory because of its goal, which is to provide users with content and more precisely pedagogical content. This pedagogical content is composed of the notions to acquire, the links between these notions and the resources they
index.

Notions are not only chosen because they are related to the course unit, they are also the result of a reflection on the course itself. A pedagogical work has to be done. For example, with the first object, why and how to make a link between the “loop” and “array” notions?

Resources have to be selected relying on pedagogical goals. The choice of their indexation terms is related to this goal too. It is not an automatic indexation. The course manager (with the help of an editorial committee if needed) is responsible for the pertinence of the links. It is not because a document treats of a notion to acquire that it will be necessary indexed by this notion. The choice is explicit, that is to say that the document must have been evaluated as sufficiently adapted to the learning of this notion.

These choices are part of the pedagogical scenario the course manager wants to implement. In a classical organizational memory, there is no pedagogical scenario because the objective of this kind of memory is not training.

The learning organizational memory we propose aims at facilitating knowledge organization and management for a given course or training, and at clarifying competencies it allows to acquire.

**Notion to learn**

The design of an e-learning application implies to focus on the learner, giving him/her the means to be active, to make him/her understand the resources that are at his/her disposal and to teach him/her how to search and to use them. Articulating a course starting from knowledge grains offers more individualization possibilities. For some authors[5], it consists in dividing the course content in fine grains, using a semantic mark-up.

On the contrary, we do not use the expression ‘notion to learn’ to refer to a course unit part, but to a notion to acquire. Consequently, there is no need to cut off existing documents or to produce new documents corresponding to these notions. Authors remain free regarding the making of their documents. They do not have to follow graphical or contents guidelines. Moreover they can reuse existing documents.

Notions to learn are used as indexes to access documents related to them. A notion to learn can refer to several documents (giving several means to acquire it) and a document can be referred to by several notions (giving several means to retrieve it)[2].

**Pedagogical resources**

Pedagogical resources are generally documents: course texts, course notes, slides, e-books, reports, books presentations, links to web sites … Among the represented documents, some (digital documents) are stored in the memory and others are references to physical documents.

Resources can be accessed according to different rights. They can be private. In this case, users only store them in the memory and do not want to give other users access to them. They can be annotations, work in progress, downloaded and not yet analyzed documents. Resources can also be semi-public or public, that is to say shared by part or all of the users. For example, an annotation of a reader giving his/her motivated impression on a document can help memory users to choose appropriate documents. Moreover, several annotations written by different authors or relying on different notions can be attached to a same document.

Resources can also have different status. They can be terminated and validated documents, or on the contrary, working documents written by one or more users and therefore shared by them during the time of their realization.

**3 Ontologies**

**Ontologies for e-learning**

For navigating through the memory, the end-users (learners, teachers, etc.) need a shared vocabulary. That is why we decided to model the memory with ontologies. From the different ontology types defined by Van Heijst[10], generic ontologies, domain ontologies, application ontologies and meta-ontologies, we only use the second and third categories. We have to consider two aspects for modeling the memory and building ontologies [6]. First the domain of training has its own characteristics. Secondly, it must be linked to the application domain of a particular training program.

The first ontology (domain ontology) we have to specify, describes the concepts of the « training » domain. They can be users types (tutor, secretary), documents types (book, slides for oral presentation, web page, site, etc.), media types (text, image, audio, video). They can also be pedagogical characteristics (activity type) and they can refer to point of view (annotation). It is difficult to directly reuse part or a whole of existing ontologies because they mainly depend on objectives and choices for specific needs,
but we must consider the help they can bring.

The second ontology (application ontology) specifies the organization of theoretical notions that are studied during training session. In the example of an initiation to algorithmic, some notions like data structure or control structure are explained. It is possible, but not mandatory, to consider “tree” and “array” as sub-concepts of the concept “data structure” and to define the relation “uses” between the concepts “data structure” and “iterative structures” (in this case they are the domain and range value of this relation).

These ontologies are not independent; the second one is necessarily attached to the first one. For example, to express that a document is an introduction to data structures we join the two concepts “introduction” and “data structures” that do not belong to the same ontology. Pedagogical relations like “prerequisite” or “uses” that occur between concepts of the application ontology are defined in the domain ontology.

Integration of ontologies

In the DANTE project the domain is the training itself. Its corresponding ontology has to be linked to application ontologies. Figure 1 shows this integration. The root of the project ontology is danteObject. First, this concept must be the root of all the concepts belonging to application ontologies. The sub-concept knowledgeBeanObject allows the integration of application ontologies. Their root concept must extend it.

Secondly, the danteObject concept must also be the root of all the concepts that belongs to the training ontology. Its root is called here trainingOntologyObject. The project defines a special concept called knowledgeBean whose elements are the concepts of application ontologies. They are the notions that learners have to study in the training. This concept extends the specific trainingClass containing all the concepts of the domain ontology.

Elements of the domain ontology

Figure 1 shows the upper elements of the domain ontology. We give in this section more details about it (see Figure 2). Actors of the training program are instances of the concept Person and we consider four categories listed in the figure. A person can also play a role in a relation: author, responsible, or tutor for example. Documents are organized according to their form, more or less structured. We present in Figure 2 the main categories. Each document is associated with a support (ResourceAccess in Figure 2), digital or not.

Figure 1. Integration of ontologies

Secondly, the danteObject concept must also be the root of all the concepts that belongs to the training ontology. Its root is called here trainingOntologyObject. The project defines a special concept called knowledgeBean whose elements are the concepts of application ontologies. They are the notions that learners have to study in the training. This concept extends the specific trainingClass containing all the concepts of the domain ontology.

TrainingProperty in Figure 1 is the class of relations occurring between concepts. Some are more pedagogical as prerequisite for example. Other are more general as writtenBy that allows to link a document and a person. Binary relations have a domain and a range for constraining instances of the relation, but more generally we can include a relation inside a Cartesian product of generic concepts.

When writing application ontology compliant with that domain ontology only few constraints appear:

- The root of the ontology must extend the concept knowledgeBeanObject as firstObject and secondObject in Figure 1.
- Each concept of the application ontology must be an instance of the concept knowledgeBean.
- It is possible to use relations defined in the
domain ontology.

It is possible to create relation between concepts of both ontologies.

**Populating the memory**

We give an example of annotation to show the way the memory can be populated. An annotation allows to give a suggestion about either one concept or a set of concepts. In the last case, there is no particular relation occurring in the ontology between the concepts that must be annotated. It is not an annotation of each single concept but of the reunion of all. When navigating, it is important and necessary to get the annotation document from any concept that is concerned by this annotation. The domain ontology contains the suggestion annotation relation defined by: suggestion_annotation 'Bag:about' Annotation:information. Bag is a domain concept that allows to group knowledgeBean elements. Note that we also use this facility for annotating one concept. For example, if we want to give information about the use of the concepts set and complement defined in the statistics ontology, in the population we would have:

a. bag_1, instance of Bag
b. element(Set,bag_1)
c. element(Complement,bag_1)
d. ann_1, instance of Annotation
   (Annotation is a subclass of Document)
e. suggestion_annotation(ann_1,bag_1)

For indicating the author of the annotation:

f. Pah, instance of Person
g. writtenBy(ann_1,Pah)

**A method for building ontologies**

The analysis of several research works [3] allows reaching a consensus on ontology building process. It relies on two steps: ontologization and operationalization (see Figure 3).

The ontologization step consists in building a conceptual ontology. Knowledge of a domain is elaborated in two ways:

- Human followed by machine analysis of various kinds of resources such as glossaries, books, courses, other ontologies, texts, etc., revealing terms and semantics structures.
- Expert interviews.

The operationalization step consists in coding the conceptual ontology using an operational knowledge representation language (i.e. equipped with inference mechanisms). This step can lead to loss of information.

![Figure 3. Two main steps in ontology building process](image)

Concepts are often structured using taxonomies. To build taxonomy of concepts, three approaches can be considered [7]:

- **Top-Down approach:** first top-level ontology concepts are built, and then they are specialized.
- **Bottom-Up approach:** first low-level ontology concepts are built then they are generalized.
- **Middle-out approach:** first most important concepts are built, then they are generalized and specialized.

**OntoSpec specification method**

OntoSpec is a method of semi-informal specification of ontologies [8]. It supposes that a conceptualization is made up of a set of concepts (or conceptual entities) and relations. The concepts in OntoSpec are organized in a taxonomy. Sub-concepts inherit all the properties of their super-concept. The relations make it possible to connect various concepts between them.

A conceptual entity owns a definition and denotes a set of objects having properties. The entity definition structure is based on a classification of these properties. At a first level, the properties are either Essential Properties (EP) or Incidental Properties (IP). The EPs are verified by all the objects denoted by the entity in every situation, or possible world. They are thus really definitional. Conversely, the IPs are satisfied only in a sub-range of situations. At a second level, the properties are classified according to roles they play regarding the conceptual entity. These roles can be abstract, e.g. Necessary Condition (NC), Sufficient Condition (SN), Necessary and Sufficient Condition (NSC). If the entity is defined by NSC, then its definition is complete. It is enough to characterize the entity.

An ontology is a differential set of concepts: the concepts are positioned according to their differences. In fact, the set of concepts are structured hierarchically and the properties are bound by conceptual properties. The conceptual property that structures a hierarchy of concepts is the
subsumption, which binds two concepts: the concept C1 subsumes another concept C2, (respectively the relation R1 subsumes another relation R2, if and only if all instances of C2 are necessarily instance of C1. The sub-concept is more specific than the super-concept and denotes less amount of objects (smaller extension).

Sibling concepts are organized in semantic axes according to their similarities. The set concept is specialized according to three axes: finite/infinite, countable/uncountable, subset/superset (see Figure 4).

![Figure 4. Specializations of the “set” notion](image)

OntoSpec specification method also defines a list of the more specific properties that can be associated to a concept or a relation. This will not be developed here. It is semi-formal because it requires a definition of the conceptual entities (concepts and relations) using a strongly structured language.

4 Conclusion

In this paper we addressed the problems related to pedagogical resources management for e-learning. To organize the resources in a learning organizational memory, we rely on ontologies. We consider two kinds of ontologies: the first one is generic and concerns the domain of training. The second one is related to the application domain, it is specific to a particular training program. We described our approach for building these ontologies and illustrated it with some examples of the learning memory we designed for two courses of our universities.

This research work is thus situated at the crossroad of three domains: knowledge engineering, pedagogical design and semantic web. The determination of knowledge grains and links between them relates to pedagogical design. The choices of organization, the management of resources in an ontology-based learning organizational memory concerns knowledge engineering.

Acknowledgements: The research was supported by the Grant Agency of CNMP (grant No. 73-CEEX-II-03 from 31/07/2006).

References:


