

# The Knowledge Puzzle: An integrated Approach of Intelligent Tutoring Systems and Knowledge Management

Amal Zouaq<sup>1</sup>, Roger Nkambou<sup>2</sup>, Claude Frasson<sup>1</sup>

<sup>1</sup>University of Montreal, <sup>2</sup>UQAM

{zouaq,frasson}@iro.umontreal.ca, nkambou.roger@uqam.ca

## Abstract

*In this paper, we present The Knowledge Puzzle, an ontology-based platform designed to facilitate domain knowledge acquisition for knowledge-based systems and especially for intelligent tutoring systems. We present a new content model, the Knowledge Puzzle Content Model, that aims to create Learning Knowledge Objects (LKO) from annotated content. Annotations are performed semi-automatically using natural language processing algorithms. These LKOs are then aggregated in an Organizational memory (OM) which serves as a knowledge base for an Intelligent Tutoring System (ITS).*

## 1. Introduction

Knowledge acquisition has always been the major bottleneck for knowledge-based systems (KBS). This is especially true for the domain of Intelligent Tutoring Systems (ITS) [16]. To make them more widespread in academic and industrial settings, easy tools must be elaborated for knowledge acquisition and authoring. Moreover, one other problem of KBS is that knowledge must be created from scratch. However, crucial information resides in existing documents within a community or an organization. We believe that the ability to reuse document content could represent a great opportunity to capture tacit and explicit domain knowledge and a key issue in competence development. However, few research works paid attention to this important question in e-learning or AIED communities [7, 13, 27]. In general, current approaches to this issue do not provide an integrated vision of the real knowledge need and the delivered training. The context in which the learning takes place is not modeled. We think that this dimension is crucial for the delivery of just-in time, just-enough learning. Moreover, these approaches assume that training resources are already available. The idea of extracting domain expertise and learning materials from any kind of documents within a community (a university, a company) is not exploited and is not used to enhance domain ontologies especially for an intelligent tutoring system.

In this paper, we present an ontology-based approach for the creation of **Learning Knowledge Objects (LKO)** from annotated documents. Natural language processing tools are used to annotate documents and to discover new domain concepts, which are then added to the domain ontology. Semantic links are also created between LKOs. The resulting LKOs and links are integrated into an Organizational Memory (OM), which gathers knowledge, expertise, solutions and experiences. Our platform, named The Knowledge Puzzle, relies on Ontologies and Semantic Web Languages to create and organize the OM content. This paper is organized as follows:

First, we present a new content model to describe learning knowledge objects (LKO) and their metadata and to provide resources compliant with standard learning content models (SCORM, LOM).

Second, we show how we create these LKOs from ontology-based annotations of documents. This enables us to formalize and re-purpose key knowledge and store it in an Organizational Memory (OM).

Third, we discuss how OM can support learning and job-aid tasks and how it serves as a knowledge prosthesis for an intelligent tutoring system.

## 2. Related Work

Our work has its roots in several disciplines ranging from ITS and ITS Authoring Tools (Educational side) to Organizational Memories and Semantic Annotations (Knowledge Management side).

### ITS Authoring Tools: Acquiring Domain Knowledge in ITS

As Murray stated [16], building ITS domain knowledge usually consists in equipping the author with tools (such as CREAM-Tools [17], EON [16], etc.) that enable him to define domain concept maps (domain model) and problem solving rules (expert model). Even with the use of these tools, domain knowledge extraction or creation remains a very difficult task, as the expert must work hard to determine the concepts that should be

considered and the relationships between them. No automatic or semi-automatic knowledge extraction tools are provided to the expert to help him elicit his knowledge. Moreover, once a domain model is elaborated and stored in a knowledge base, it is difficult to maintain it and to make it evolve without resorting another time to experts. This is a very costly approach especially in organizational contexts. In fact, domain knowledge evolution is an intrinsic part of organizations' lives thus (semi) automatic knowledge extraction tools must be made available.

### Organizational Memories

Knowledge is the key asset of the modern knowledge intensive organization. According to Conklin, "Organizational memory extends and amplifies this asset by capturing, organizing, disseminating, and reusing the knowledge created by its employees." [6]. We think that an OM can be set-up within any community with shared knowledge. Researches in the field of OM are more oriented towards enterprise communities such KnowMore [1] and CoMMA [10]. Few studies tried to merge eLearning with OM [2] or Knowledge Management (LIP [21]) and we intend to present an integrated vision of both worlds (OM and ITS).

### Semantic Annotations

Many uses of annotations have been described [15] and many tools for annotating content have been implemented (such as CREAM [11], KIM [18]). Annotation tools dedicated to training materials are particularly interesting in our case. These annotation tools, such as TANGRAM [13] or AMG [5], aim to create metadata for learning objects or parts of them. The use of ontologies as the backbone of the annotation process is more and more adopted [3] in eLearning communities. However, few annotation platforms for training materials integrated, like we do, natural language processing (NLP) in the annotation process.

## 3. Our Solution: The Knowledge Puzzle

The Knowledge Puzzle is an integrated platform of knowledge management and intelligent tutoring systems. It aims to provide easy tools to annotate documents with the objective of creating an organizational memory, which gathers information, and knowledge that flows into the organization. The content of this OM is in turn used by an ITS to provide just-in time, just-enough learning.

The just-in time learning objective is realized by the definition of competencies (denoting training needs) and their link with roles and tasks. Whenever a human agent (an employee in a company, a student in a university...)

must be trained, a set of competencies is defined according to the task he or she must fulfill.

The just-enough learning objective is reached thanks to the fine-grained decomposition of documents into knowledge objects, assets and asset categories. The use of ontologies to annotate knowledge objects and their components aims at facilitating their reusability.

### Knowledge Puzzle Content Model

We developed a learning content model called The Knowledge Puzzle Content Model. It enables to describe a learning object and its components at a very fine-grained level. A number of learning content models already exist; such as the SCORM content aggregation model [22], the learning object taxonomy [8] or the ALOCOM Model [26, 27].

In [26], Verbert and Duval studied six content models and showed that they could map on their abstract model (ALOCOM). So we compare our Knowledge Puzzle Content Model to the ALOCOM one, and we refer the reader to [26] for a comparative analysis and a survey of these models. Briefly speaking, ALOCOM distinguishes between content fragments (CF), content objects (CO) and learning objects (LO). Content fragments represent basic learning resources such as text fragments, images, audio, etc. Content objects aggregate content fragments and can be composed of sub-content objects. Learning objects aggregate content objects, can be composed of sub-learning objects and finally identify a learning objective for the learning object.

We suggest that content fragments should not only be represented by their format (text, fragment, image, etc.) but also by their intended uses or meanings (which we call asset categories) in educational settings. We also think that the three-level decomposition (CF, CO, LO) is not necessary to obtain a learning object with a learning objective.

So, our model is based on the following statements:

First, we introduce the concept of categories for content fragments. We name content fragments as **assets** (see figure 1). Asset Categories, as indicated in figure 2, could evolve and are not limited to the proposed list. We took as a starting point the work of Ullrich [25] to elaborate these categories.

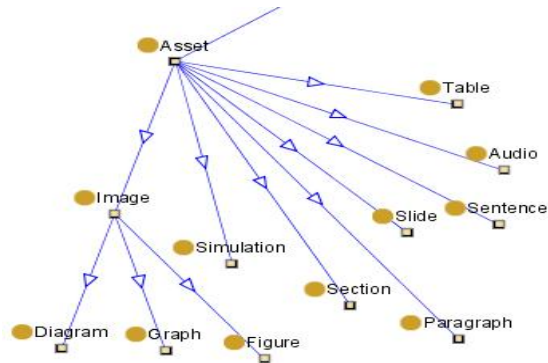


Figure 1: Assets

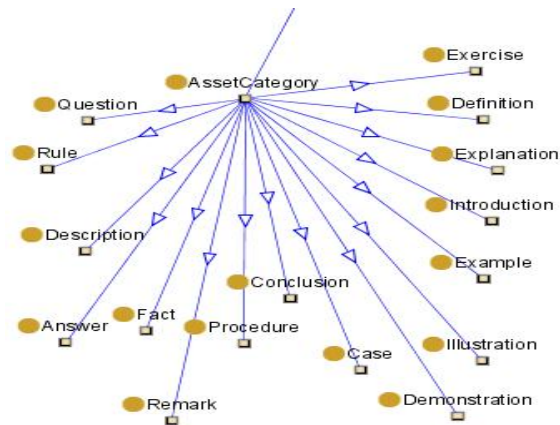


Figure 2: Asset Categories

Second, a document is represented by what we call a **Knowledge Object**, which is linked to annotation objects. An annotation object could in turn be specialized into a Structural Annotation and an Instructional Annotation.

Structural Annotation identifies document structure. It decomposes the document into paragraphs, sentences, and holds metadata such as key concepts, title, author, format and others.

Instructional Annotation identifies the list of asset categories that exist in a document. These kinds of annotations are created by a human or software annotator and thus could represent different points of view over the same document.

Third, a Learning Knowledge Object is linked to an instructional objective and aggregate one or more assets chosen according to a pedagogical scenario. This pedagogical scenario uses asset categories to fulfill the instructional objective.

The Knowledge Puzzle Content Model is shown in figure 3.

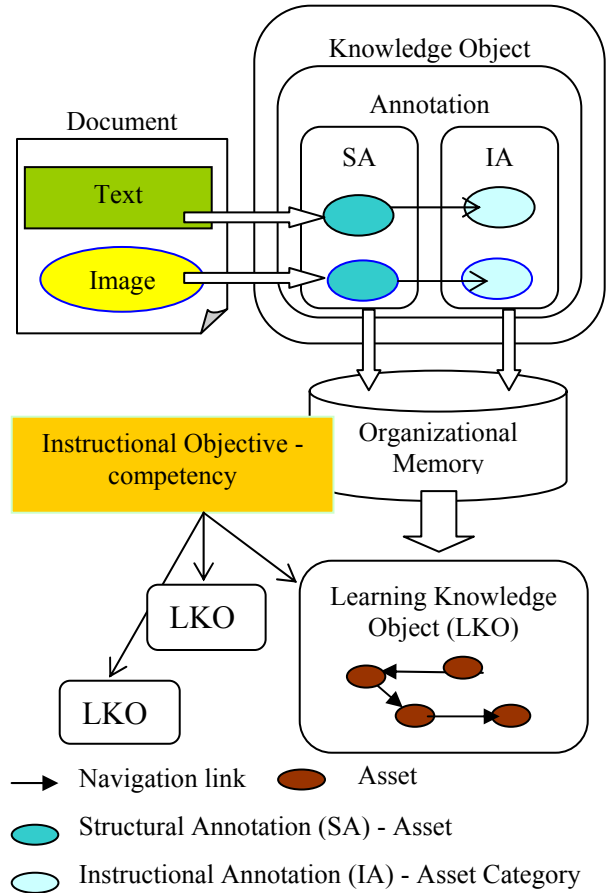


Figure 3: Knowledge Puzzle Content Model

Moreover, the Knowledge Puzzle Content Model is based on an ontological model, which we introduce in the following section.

### Knowledge Puzzle Ontological Foundation

The use of ontologies in the specification of a content model and its resources eases its reusability and its common comprehension within the targeted community. Stojanović et al. [23] describe three dimensions to document comprehension and usage in eLearning: structure, content and context. Our ontologies describe the same dimensions.

- **Document Structure Ontology**

This ontology decomposes a document into relevant knowledge parts called assets and asset categories (figure 1 and 2). It represents the annotation process described in the Knowledge Puzzle Content Model. The instantiation of this ontology transforms a document into a Knowledge Object linked to annotations.

- **Domain Ontology**

Domain Ontology is organized around the notion of concept. A concept can be linked to another one by a number of relationships such as hierarchical links, composition links, prerequisites links, etc. This is represented through the Relation Class in the ontology. A concept is defined by one or more terms in the Term class.

Knowledge objects or assets are related to domain concepts either in their content or in their metadata (description, key concepts, etc.). In fact, all the ontology's classes are described by one or more domain concepts.

- **Organization Ontology**

The organization ontology describes the targeted community and its structure in term of actors, tasks and processes. In the case of a company for example, it describes its divisions, its employees, their roles and other entities (such as human and software agents, places, meetings).

Each class of the organization ontology is also described by a number of concepts of the domain ontology. Moreover, Knowledge Objects and Assets can be linked to organizational annotations such as named entities annotations (people, places, meetings, etc.).

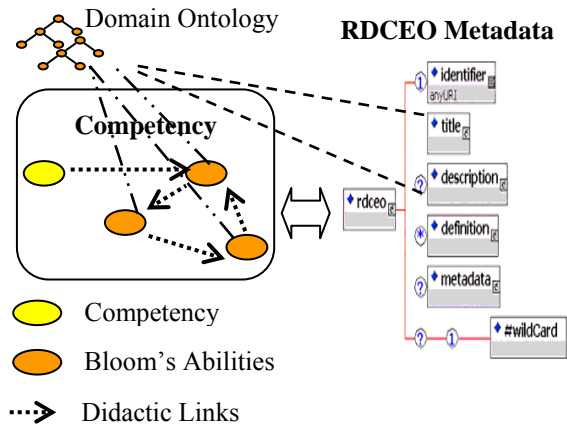
- **Competency Ontology**

Many definitions of the notion of competency exist in the literature. Among these definitions is the IMS Reusable Definition of Competency or Educational Objective Specification (RDCEO) [20]. This specification, elaborated by the IMS Global Learning Consortium, enables to create common understandings of competencies.

Based on the RDCEO, we define a competency as an educational, instructional or a learning objective. According to Nkambou *et al.* [17], a learning objective is a description of a set of behaviors (or performances) a learner should be able to demonstrate after a learning session. It can also describe the set of abilities or skills to be mastered by a student after a pedagogical activity. Nkambou *et al.* [17] created a pedagogical objectives model (CREAM-O) in which pedagogical objectives are represented and connected by didactic links.

Following Nkambou *et al.* [17], we conceptualize a competency as a set of abilities an agent must master in various contexts (educational context, working context, etc). Ability is defined according to domain concepts. We also represent different levels of competencies by using Bloom's taxonomy [4]. In fact, Bloom's taxonomy uses action verbs to qualify the ability involved in a competency. It is largely used in education in general [17] and its integration to our ontological model enables the

definition of competencies at a very detailed level, which again foster knowledge reusability. Our competencies are associated with RDCEO compliant metadata, which also refer to domain ontology concepts. Figure 4 illustrates our competency's definition.



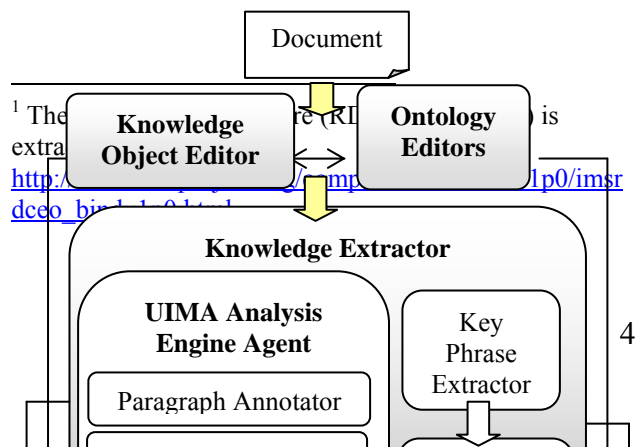
**Figure 4: Competency Components and Metadata<sup>1</sup>**

An example of a competency would be: “Learn Remote Method Invocation (RMI) usage”. The set of abilities and concepts associated to this competency would be: “**define** RMI”, “**identify** RMI goals”, and “**indicate** an example of RMI usage”. The abilities in the example are indicated in bold.

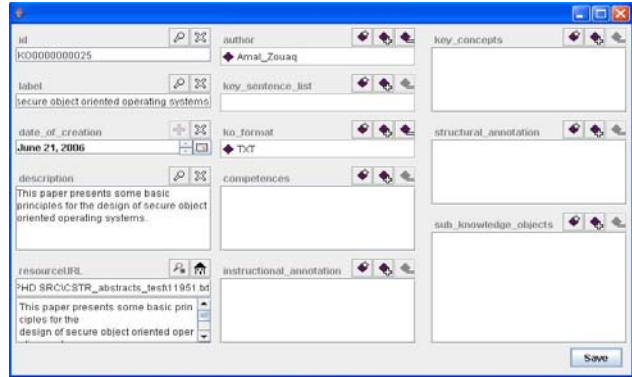
To summarize and to map our ontologies to Stojanović *et al.*'s dimensions [23], we could say that document structure ontology describes the structural dimension; domain ontology depicts the content dimension whereas organization and competency ontologies represent the context dimension.

### Knowledge Puzzle's Architecture

Knowledge Puzzle's architecture can be decomposed into its knowledge management component and its knowledge exploitation component. It is a fully ontology-based environment. Ontological schemas were developed using Protégé Ontology Editor [19] and we use an open-source Java library, the Protégé OWL API [19], to create and update our ontological knowledge base. Figure 5 shows the Knowledge Management Component of the Knowledge Puzzle's Architecture.



<sup>1</sup> The extra information is available at [http://www.imsc.org/imsr/dceo\\_briefing.html](http://www.imsc.org/imsr/dceo_briefing.html)



**Figure 6: Knowledge Object Editor**

- **Knowledge Extractor**

Knowledge Puzzle's Knowledge Extractor (figure 7) extracts semantic annotations (metadata) both at the knowledge object level and asset level. This tool uses natural language processing technologies for metadata extraction.

First, we developed an **Analysis Engine Agent** that uses annotators (Paragraph Annotator, Sentence Annotator and Named Entity Annotator) to extract relevant knowledge from document. This analysis engine agent is based on IBM's Unstructured Information Management Architecture (UIMA) [24]. UIMA is an architecture and software framework that helps to build a bridge between unstructured information sources and structured knowledge. In our case, the agent is used to generate document's structure (mainly paragraphs and sentences) and to detect named entities in text (people, names, places, etc.) of particular interest to the community. The analysis results are stored in the Document Structure Ontology.

Second, we used a key phrase extractor algorithm called Kea-3.0 [9] to find document's key concepts (a key phrase or concept can be composed of more than one word). Briefly speaking, Kea-3.0 identifies candidate phrases in a document using lexical processing, computes features (TF \* IDF and position of first occurrence) for each candidate, and finally generates a classifier using machine learning. This classifier determines which candidates could be chosen as key phrases. More details could be found in [9, 12]. Key concepts are stored at the Knowledge Object level (Document Structure Ontology) and are also added to the Domain Ontology if they do not already exist. This process can be fully automatic or semi-automatic. Indeed, a human annotator can control key phrases' pertinence and update key concepts accordingly by simple drag and drops from document content to key concepts list.

**Figure 5: Knowledge Puzzle's Knowledge Management Component**

### **Knowledge Management in Knowledge Puzzle**

The knowledge management component is composed basically of three kinds of tools: Ontology Editors, a Knowledge Extractor and a Knowledge Annotator.

- **Ontology Editors**

Knowledge Puzzle's Ontology Editors are used to create or update ontologies content. The Knowledge Object Editor (figure 6) is of special interest because it is used each time a document is processed. This editor works at the knowledge object layer level and allows creating, automatically or manually, basic knowledge object metadata such as title, date of creation, description, creator, format, etc.

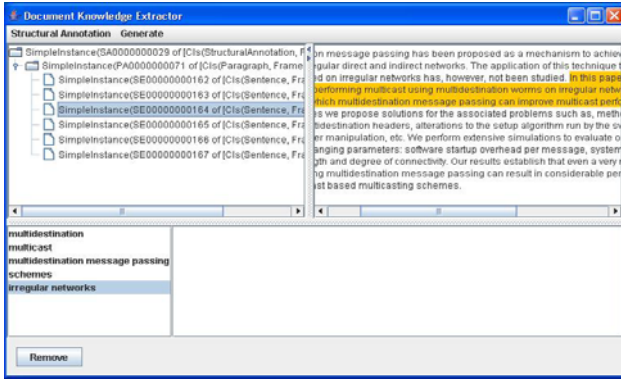


Figure 7: Knowledge Extractor

Third, sentences containing key concepts are considered as **key sentences**. Again, a human annotator can evaluate and update key sentences list. Then these sentences are transformed into **concept maps**. We used the Stanford Parser [14] to obtain a lexical concept map for each key sentence and we developed a semantic concept map extractor that transforms the lexical concept map into a semantic concept map. This semantic map is used to discover new concepts and links and to update domain ontology. Figure 8 shows a lexical concept map for the sentence “An intelligent tutoring system provides individualized training” and figure 9 illustrates how this map is transformed into a semantic one. Here, domain ontology is enriched by two concepts: “intelligent tutoring system” and “individualized training”. These two concepts are related to each other by the new link “provides”.

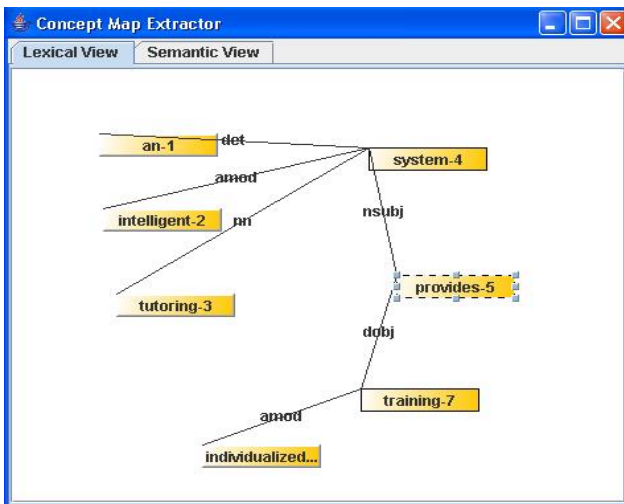


Figure 8: Lexical Concept Map Example

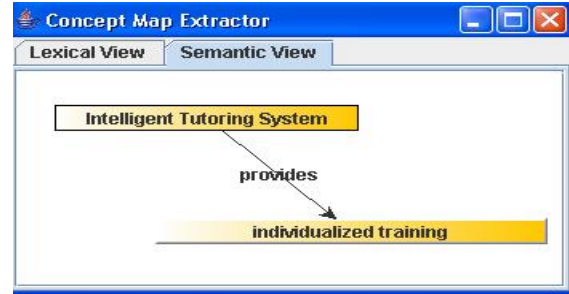


Figure 9: Semantic view of figure 8's concept map

- **Knowledge Annotator**

The knowledge Annotator, illustrated in figure 10, is a tool to manually annotate documents. By simple drag-and-drops, the annotator can define structural annotations (sections, images, key sentences, etc.) as well as instructional annotations based on the document structure ontology. These annotations are then linked to the Knowledge Object that represents the document.

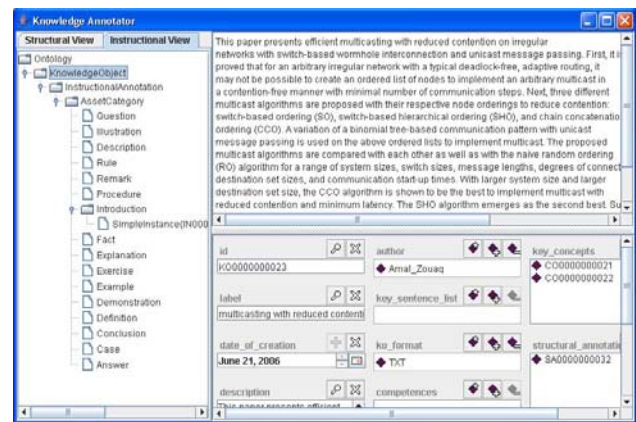


Figure 10: Knowledge annotator

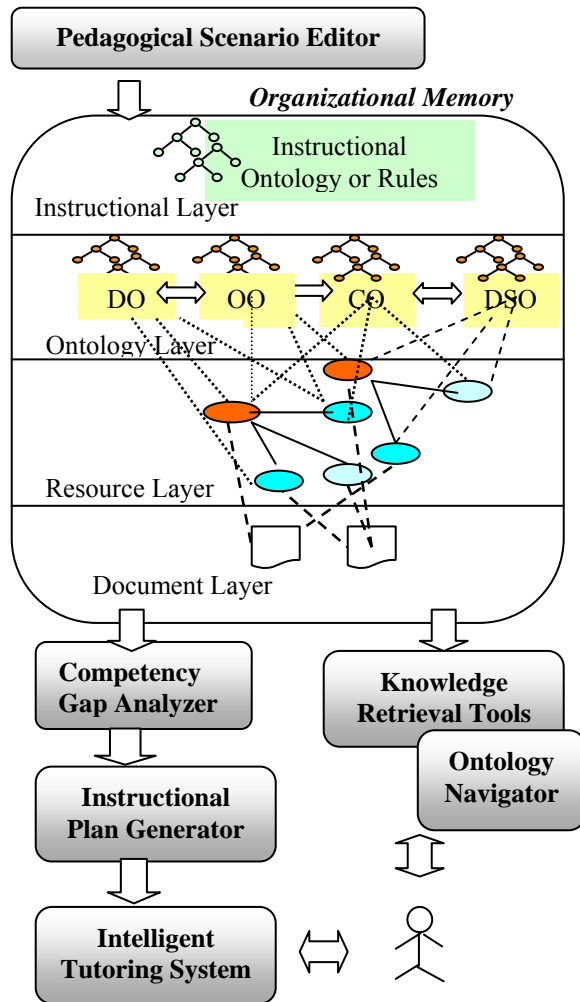
The Knowledge Annotator also uses ontologies to index current documents content. More precisely, domain ontology is traversed to map any existing concept to knowledge content as well as competence ontology.

### How Knowledge Puzzle is used for training purposes?

The main objective of the Knowledge Puzzle Architecture is to use organizational memory content for efficient training. Figure 11 illustrates Knowledge Puzzle's Knowledge Exploitation Component. As indicated in the Knowledge Puzzle Content Model (figure 3), aggregating a number of assets creates a Learning Knowledge Object (LKO). This aggregation is guided by a pedagogical scenario, which uses asset categories (definition, example, case, etc.) and formal instructional

theories (Gagné, Merrill ...). This scenario is created through a **Pedagogical Scenario Editor (PSE)**.

A **competency gap analyzer (CGA)** compares the user profile (stored in the Organization Ontology) with the competency definition to detect training needs for that specific user. User's learning objectives are then indicated to the **Instructional Plan Generator (IPG)**. Then according to a pedagogical scenario, the planner searches the OM to gather relevant assets and generates **Learning Knowledge Objects**. Finally, the intelligent tutoring system deploys the learning session in conformance with the generated plan.



**Figure 11: Knowledge Puzzle's Knowledge Exploitation Component**

Moreover, the ITS offers a number of tools to the user enabling him to search the organizational memory to find experts, knowledge objects, documents, problem solving methods, etc. (**Knowledge Retrieval Tools and Ontology Navigators**). These last tools are already implemented but we are still in the process of creating the learning part of the architecture (PSE, CGA, IPG, ITS).

Knowledge Puzzle uses two algorithms that have already been evaluated (Kea 3.0 and Stanford Parser). We refer the reader to [12] for Kea 3.0's evaluation and to [14] for Stanford Parser's evaluation. We would like to underline that first experiments have shown that the good performance of these two algorithms are preserved in our context.

#### 4. Conclusion and Further Work

In this paper, we presented an ontology-based approach to automatically and manually annotate document content. We also introduced a new content model, the Knowledge Puzzle Content Model that allows decomposing a document from a structural and an instructional point of view. This decomposition makes possible to retrieve and uses document components (assets and assets categories) and to automatically create pertinent learning knowledge objects (LKO). These LKO are stored into an Organizational Memory, which is then used as a knowledge base for an intelligent tutoring system.

Our content model maps to a number of existing content models. Thus our annotations can be used to generate standard metadata such as SCORM and LOM and standard learning objects (SCORM). The next goal in this research will be to generate these standard objects and metadata as well as to explore the use of natural language processing to automatically extract other kinds of metadata such as asset categories (which are manually annotated for the moment).

#### 5. References

[1] Abecker, A., Bernardi, A., Sintek, M, "Proactive Knowledge Delivery for Enterprise Knowledge Management", In Proceedings of the 11th International Conference on Software Engineering and Knowledge Engineering, Learning Software Organizations, Methodology and Applications, Lecture Notes In Computer Science, , 1756, 1999, pp.103–117.  
 [2] Abel, M.-H., Benayache, A., Lenne, D., Moulin, C., Barry, C., Chaput, B., "Ontology-based Organizational Memory for e-learning", Educational Technology & Society, 7 (4), 2004, pp. 98-111.  
 [3] Aroyo L, Dicheva D, "The New Challenges for E-learning: The Educational Semantic Web", Educational Technology & Society, 4(7) , 2004, pp. 59-69.  
 [4] Bloom B. S. and Krathwohl D. R., "Taxonomy of Educational Objectives: The Classification of Educational Goals", by a committee of college and university examiners. Handbook I: Cognitive Domain. New York, Longmans, Green, 1956.

- [5] Cardinaels K., Meire M., Duval E., "Automating Metadata Generation: the Simple Indexing Interface", Proceedings of the 14th international conference on World Wide Web, Chiba, Japan, 2005, pp.548 – 556.
- [6] Conklin, E.J. "Designing Organizational Memory: Preserving Intellectual Assets in a Knowledge Economy", white paper, Group Decision Support Systems, Washington D.C. Available at : <http://cognexus.org/dom.pdf>.
- [7] De Hoog, R., Kabel, S., Barnard, Y., Boy, G., DeLuca, P., Desmoulins, C., Riemersma, J., Verstegen, D., "Re-using technical manuals for instruction: creating instructional material with the tools of the IMAT project", In: Barnard, Y. (Ed.), ITS (Intelligent Tutoring Systems) 2002, Workshop proceedings Integrating technical and training documentation, San Sebastián, Spain, 2002, pp. 28-39.
- [8] Duval, E., Hodgins, W. A., "LOM Research Agenda", WWW2003 Conference, 2003, Budapest, Hungary.
- [9] Frank E., Paynter, G.W., Witten, I.H., Gutwin, C., Nevill-Manning, C.G., "Domain-specific keyphrase extraction" Proc. Of the 16th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers, San Francisco, CA, 1999, pp. 668-673.
- [10] Gandon, F., "A Multi-Agent Architecture For Distributed Corporate Memories", Third International Symposium, From Agent Theory to Agent Implementation, at the 16th European Meeting on Cybernetics and Systems Research, Vienna, Austria, 2002, pp. 623-628.
- [11] Handschuh S., Staab S., "CREAM - Creating Metadata for the Semantic Web", Computer Networks, 42, 2003, pp. 579-598.
- [12] Jones, S., Paynter, G. W., "Human evaluation of Kea, an automatic keyphrasing system", ACM/IEEE Joint Conference on Digital Libraries, 2001, pp.148-156.
- [13] Jovanovic, J., Gasevic, D., Devdzic, V., "Ontology-based Automatic Annotation of Learning Content", International Journal on Semantic Web and Information Systems, 2(2), 2006, pp. 91-119.
- [14] Klein, D., Manning, C. D., "Fast Exact Inference with a Factored Model for Natural Language Parsing", In Advances in Neural Information Processing Systems, 15 (NIPS 2002), 2002.
- [15] Marshall, C., "Toward an ecology of hypertext annotation", in ACM Hypertext. Pittsburgh, PA, 1998, pp. 40-49.
- [16] Murray, T., "An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art", In Murray, T., Blessing, S. & Ainsworth, S. (Eds.), Authoring Tools for Advanced Technology Learning Environments Toward cost-effective adaptive, interactive, and intelligent educational software., Kluwer Publishers, 2003, pp. 491-544.
- [17] Nkambou R., Frasson, C., Gauthier G., "CREAM-Tools : An Authoring Environment for Knowledge Engineering in Intelligent Tutoring Systems", In : Murray, T., Blessing, S. and Ainsworth, S. (Eds) : Authoring Tools for Advanced Technology Learning Environments : Toward cost-effective adaptive, interactive, and intelligent educational software., Kluwer Publishers, 2003, pp. 93-138.
- [18] Popov B., Kiryakov A., Ognyanoff D., Manov D., Kirilov A., "KIM - A Semantic Platform For Information Extraction and Retrieval", Journal of Natural Language Engineering, 10(3-4), Cambridge University Press, 2004, pp. 375-392.
- [19] Protégé ontology builder and Protégé OWL API. Available at: <http://protege.stanford.edu/>.
- [20] RDCEO. Available at : <http://www.imsproject.org/competencies/index.html>
- [21] Schmidt, A., "Bridging the Gap Between E-Learning and Knowledge Management with Context-Aware Corporate Learning (Extended Version)" In: Professional Knowledge Management (WM 2005) Post Proceedings, Springer, 2005.
- [22] SCORM CAM. Available at : <http://www.adlnet.gov/downloads/70.cfm>
- [23] Stojanović, L. J., Staab, S., Studer, R., "E-learning based on the Semantic Web", In Proceedings of the 6th World Conference on the WWW and the Internet, Orlando, FL, 2001.
- [24] UIMA. Available at : <http://www.research.ibm.com/UIMA/>
- [25] Ullrich, C., "The learning-resource-type is dead, long live the learning-resource-type!", Learning Objects and Learning Designs, 1(1), 2005, pp.7-15.
- [26] Verbert, K., Duval, E., "Towards a Global Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models", In P. Kommers & G. Richards (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2004 Chesapeake, VA: AACE., pp. 202-208.
- [27] Verbert, K., Klerkx, J., Meire, M., Najjar, J., Duval, E., "Towards a global component architecture for learning objects: An ontology based approach", In Proceedings of the OTM 2004 Workshop on Ontologies, Semantics and E-learning, Agia Napa, Cyprus, 2004, pp. 713-722.