AN APPROACH FOR CONTENT AND KNOWLEDGE PROVISION IN EXTENDED ORGANIZATIONS

Sonja D. Radenkovic, Vladan Devedzic
University of Belgrade, Serbia E-mail: sonjafon@gmail.com, devedzic@gmail.com

Abstract- This paper describes the Content and/or Knowledge Provision service (CKP service, for short), a service that is developed as a part of research related to content management and semantic search, within IntelLEO, an FP7 project in the area of technology-enhanced learning. The project aims at supporting workplace learning situations where employees are often faced with the need to efficiently find, browse, share, and otherwise use electronic learning resources, but are usually constrained by limited time and other factors when they perform all these activities on their own. CKP service is one of the core services that the IntelLEO project develops. By using the CKP service, employees can access semantically annotated heterogeneous resources.

Keywords: knowledge management, Semantic Web technologies, workplace learning

I. INTRODUCTION

In today's world, where increasingly rapid changes are taking place, it is imperative that a higher level of knowledge is attained to keep pace with new developments. Some individuals who were considered knowledgeable in the past may prove incapable of dealing with present or future demands. In the world of changes the workplace learning is defined as a multibillion-dollar enterprise in which employees learn new skills designed to help them keep their organizations competitive in an increasingly global economic environment [1]. The creation of a climate of learning and innovation within an organization requires the facilitation of new experiences by workers, creating a blurring of traditional task and job boundaries [2].

In order for workplace knowledge and experience to increase and evolve over time, collaboration and extension of learning activities across organizational borders may be essential. Intelligent Learning Extended Organization (IntelLEO) paradigm represents a learning community emerging as a temporal integration of two or more different business and educational communities and organizational cultures (industrial, research, and educational) [3]. For example, an enterprise can team up with an academic research group to create such a temporary IntelLEO, in which both the employees from the enterprise and the researchers can pursue some of their interests more efficiently. For instance, the employees may need some guidance from researchers when it comes to selection and use of new technologies, whereas the researchers may want to try out their results in a real-world environment.

The IntelLEO paradigm is the subject of exploration within the IntelLEO FP7 EU research project. The main purpose of IntelLEO is to enhance cross-organizational Learning and Knowledge Building (LKB for short) practices at the workplace. The IntelLEO paradigm is supported by a set of services, based on the common ontology framework, facilitates and further promotes LKB activities. One of those services is Content and Knowledge Provision service (CKP service for short) that enables employees to upload different kinds of learning resources into a knowledge repository, annotate them, and (re-)discover relevant learning resources by performing semantic search over the knowledge repository. Hence, this service effectively serves as a content management system and semantic search engine within an IntelLEO.

This paper presents the concept and the implementation of the CKP service and shows how it allows employees to effectively use knowledge and content from all over the Web, as well as from the extended organization's knowledge/content repositories, without wandering and wasting their time on irrelevant resources.

The paper is organized as follows: the next section presents some application scenarios for the CKP service in extended organizations, thereby also indirectly illustrating the current implementation of the CKP service. Section three shows the main concepts of IntelLEO Service Oriented Architecture. Section four highlights the conceptual framework of the CKP service, whereas section five shows related work. The last section gives the conclusions and indicates directions for future work.

II APPLICATION SCENARIOS

Let us suppose there is an IntelLEO comprising a research institution and an SME in the area of e-Engineering and e-Manufacturing. The SME wants to continually improve its products and remain the global leader in the field; thus, its employees need to stay up-to-

214
date with the latest technological developments relevant to the company. On the other hand, the research institution needs a real world setting to assess the feasibility of its research work and get feedback from end users. This implies that both members of this IntelLEO (i.e., both the employees and the researchers) frequently need access to learning resources relevant for the accomplishment of their learning goals. The learning goal in this context is defined as the harmonization of personal and organizational goals [4], and is often about the acquisition of one or more competences [5]. In addition, they need a seamless way of managing and exchanging knowledge assets, since it comes as a natural part of their learning and work-related practices. As explained in the previous sections, these activities can be facilitated and improved through the use of the CKP service supported by the other IntelLEO services (see Section 3). In what follows, we provide more insight into the support offered by this service.

In order to improve the informal knowledge during the working process an employee or a researcher runs the CKP service accessible from his/her Web browser. Currently, the service is integrated in Mozilla’s Firefox Web browser as a toolbar (Fig. 1). Its integration with Chrome browser is underway. When a Web page is opened in the browser and the CKP service is activated, it invokes the KIM server [6] to automatically annotate the resource with the concepts from a domain-specific ontology.

Clicking the IntelLEO button opens the “Bookmark and Share in IntelLEO” dialog (Fig. 2), extracting at the same time the URL and the title of the opened Web page [32]. This action initializes the data about the user who performs the annotation, the annotation itself, as well as the learning resource to be bookmarked/tagged/annotated. If the current Web page has already been bookmarked by some other member of extended organization, his/her name will be shown in the Also Tagged by part (Fig. 2B). By clicking the name/figure of user shown in this part, a Human Resource Discovery (HRD) service is called (see Section 4 for details) to retrieve and show the user’s public profile. There are three types of annotations that the CKP service makes use of to annotate different kinds of learning resources: metadata corresponding to commonly used vocabularies (Title, Author(s), Subject, and Description in Fig. 2A) concepts from a domain-specific ontology (Related Domain Concepts in Fig. 2D); these concepts are automatically extracted, user-defined tags stored in the repository (User’s Tags, Fig. 2E).

The user bookmarks the learning resource (i.e., the current Web page in the given scenario) by choosing the specific annotations from Related Domain Concepts and User’s Tags, as well as by manually adding annotations in the Tags field (Fig. 2). Having selected the IntelLEO tab (Fig. 2C), the user is presented with a list of learning goals associated with the learning resource (i.e., the Web page) being bookmarked, Fig. 3. The Source column shows the user or the organization that has originally defined a learning goal, whereas the Learning goal column shows the name of the corresponding learning goal. The Relevance column uses the 0 to 1 scale to indicate the relevancy of the given learning resource to the specific learning goal.

The Action column includes two links: Details and Add. By clicking Details, the user can reveal details about the learning goal, including the set of competencies that are to be achieved and the learning activities that should be done to reach the goal. These come directly from the Learning Path Creator service (see Section 4 for details), where the corresponding learning goal is seen “in context”, i.e. along with the other learning goals set by the organization or by a specific user. By clicking Add, the user annotates the current learning resource with the chosen learning goal, that way indicating the goal’s relevancy for the resource being bookmarked.

The bookmarking process is completed by clicking the Save button; this saves all the data (annotations) relevant for the bookmarked learning resource in the repository. Clicking Upload button (Fig. 1) allows the user to physically upload some learning resource into the repository. By "uploading" we mean physical storing a learning resource into a designated repository and annotating it (tagging, annotating with concepts from a domain-specific ontology and with descriptive metadata such as Dublin Core Terms vocabulary), so that other users can access it and use it. The user interface of the Upload functionality is very similar to the one shown on Fig. 2. The only difference is the file upload part, Fig. 4. Here the user can click Browse button to select and upload a document from her/his local folders.
The Resources button (Fig. 1) allows users to browse the repository of learning resource and view individual resources, to update their bookmarks, and delete them, as well as to perform semantic search of the repository (Fig. 5). Updating the bookmarks often includes additional manual annotation and it is performed by clicking the button Update (Fig. 5). The same dialog as the one shown in Fig. 2 will appear, this time for an
already bookmarked resource, allowing the user to make the necessary changes (e.g., add tags) and save them.

Semantic search in the CKP service (Fig. 5) is based on measuring the semantic relatedness using the Cosine similarity metrics [25] between two vectors – the vector of user’s profile and the vector of specific resource.

The vector of user’s profile is calculated as the cumulative vector consisting of the term frequency vector of user’s learning goals and the competencies within them, as well of the vector of user’s favourite concepts. The vector of specific resource is the term frequency vector of concepts obtained by the process of semantic annotation.

III. INTELLEO SOA-BASED ARCHITECTURE

CKP is one of the core IntelLEO services, Fig. 6. Technically, the IntelLEO project team applies the principles of Service-oriented Architecture (SOA) [7] in the development and integration of services. In that way, the deployed architecture can provide a loosely-integrated suite of IntelLEO services that can be used within multiple business domains, which is the ultimate goal of the IntelLEO project. The IntelLEO ontology framework (Fig. 6) includes a set of IntelLEO ontologies and a set of services that allows for retrieval and management (store, update, and delete operations) of ontology instance data stored in (RDF) data repositories. The IntelLEO software architecture includes a set of core services (CS) that can be divided into two main groups of services. The first group – Learning and Knowledge Building Services – contains the following services: 1) Services for Collaborative Learning, in particular Human Resource Discovery (HRD) CS, Working Group Composition (WGC) CS; 2) User Monitoring (UM CS); 3) Content/Knowledge Provision (CKP) CS. The second group – Harmonization Services – includes 1) Learning Path Creator (LPC) CS; and 2) Organizational Policy Service (OP) CS.

These services are often orchestrated from applications to interact with the CKP service. However, a detailed description of the other IntelLEO services is beyond the scope of this paper.

The IntelLEO CSs rely on an interlinked set of IntelLEO ontologies as their common underlying data model. This ontology set has been developed through a combined top-down (review of existing work in the field) and bottom-up (requirements elicitation from IntelLEO application cases) approach. It includes a number of ontologies designed with modularity and flexibility in mind so that they can be easily reused and extended. In accordance with the recommended practices in ontology engineering [8] and publishing Linked Data on the Web [9], when developing the IntelLEO ontologies we relied on and linked to the ontologies already available and in use. Specifically, we leveraged widely accepted and used ontologies for modeling people and online communities (Friend Of A Friend [26], and communications and content created and exchanged within those communities (Semantically Interlinked Online Communities [27]; ontologies for content annotation (Dublin Core [28] and CommonTag [29]); as well as some of the ontologies of the LOCO framework [30] for modeling characteristics of learning situations. The specifications of all the ontologies are available on the project’s web site.
For the CKP service, particularly relevant are domain ontologies that define concepts and instances describing specific domains and are typically used for semantic annotation of different kinds of resources. For the purpose of testing and initially applying the CKP service, we have developed an ontology that is actually a SKOS [31]-based version of the ACM CCS taxonomy of the computing domain. Specifically, it defines a number of concepts (as instances of the skos:Concept class) from the area of computing. The ontology relies upon skos:narrower and skos:broader properties for structuring the concepts in appropriate concept hierarchies. This ontology and any other domain-specific ontology, can be used in the context of an IntelLEO for annotation of the knowledge being shared, undertaken or planned learning activities, content that was used or produced, etc. The advantage of semantic annotation over the popular tag-based annotations is that ontology concepts have unambiguously defined semantics accessible to both humans and machines. It also allows for semantic interlinking (i.e., connecting based on meaning) of diverse kinds of components of a learning process (e.g., activities, content, knowledge, people).

IV. CKP SERVICE CONCEPTUAL MODEL

The time that employees can allocate to learning activities at the workplace is very limited. It means that they won’t allow themselves the comfort of browsing endlessly through unstructured documents and knowledge sources in order to perform the workplace learning. Another option for them is to ask someone else about (new) learning resources that are interesting for their learning. Unfortunately, the other persons that are usually experts in the area are not available at any time they want. It causes that either the work performance drops, or motivation to learn drops, or both. CKP service is in charge of locating, retrieving, and making appropriate learning and knowledge objects accessible to either members of an IntelLEO (i.e., end-users) or other IntelLEO services, based on the requirements of a given learning context[32]. The users of this service are all the extended organization members that collaborate and learn with support of the IntelLEO framework. All of them access the CKP service through the browser toolbar, or through other IntelLEO services, such as Learning Path Creator, Organizational Policy and Collaborative Learning Services.

One of the main advantages that CKP service offers to employees is related to the type of content that they can use. Here the term “content” refers to digital entities that contain and represent various forms of implicit organizational knowledge such as reports, documents and notes related to a project, forum/blog posts, micro-blogging posts (e.g. tweets), discussions/messages, Wiki entries and “traditional” learning objects related to the domain knowledge of the task at hand. All of these heterogeneous pieces of content are automatically annotated with concepts from specific domain ontology (as indicated in Section 2). At the same time, they are filtered according to the user (employee) profile.

The main components of the CKP service are shown in Fig. 7.

4.1. Unstructured Data Sources and Client Application

CKP service can be used with sources of structured content, but it is more likely that it will be used with unstructured content. In an extended organization, unstructured content sources will typically be various documents and the organization's forum/blog posts, discussions/messages, and Wiki entries. While the content and its formats vary, the topics discussed/written about are not specific to a single source. Therefore, using a semantics-based approach to annotating content allows for interlinking of (topically) related content irrespective of the source the content originates from. In addition, it provides greater flexibility when incorporating new content sources and when distributing queries among the various sources of annotations.

Fig. 7. The CKP service in an IntelLEO solution
The client application (e.g., a Web browser) provides a user-friendly interface to the CKP service (e.g., Fig. 1), and search and retrieval options, so the learner can request specific kinds of content or specific concepts associated with the content, regardless of their source.

4.2. The Role of Ontology Tool

The ontology tool enables employees to access domain-specific ontologies that can be used to create and query annotations. It also enables selected users (i.e., domain experts) to incrementally build and extend domain ontologies as needed.

The domain ontologies to be used to annotate unstructured data sources provide the employee (through the client application) with a set of relevant domain-specific concepts and relationships. The second aspect of the ontology management is the ontology maintenance and evolution. In the CKP service, it is based on a variant of the process of ontologization [10] – extracting a set of tags that are semantically related to the given domain ontology and using them to extend the ontology.

The process of extracting a set of semantically relevant tags happens in three phases [11]:

1) lexical processing – based on semantic similarity measures, selecting those tags that are meaningful and should be taken into account as a basis of the semantic enrichment

2) disambiguation and semantic expansion – disambiguation is related to eliminating tag polysemy (i.e. the possibility that the same tag has different meanings in different contexts); semantic expansion refers to expanding the tag with synonyms and lexical variations, in order to increase the possibility of identifying semantic entities (concepts in the domain ontology) most likely corresponding to the tag (see phase 3)

3) semantic enrichment – identifying semantic entities (concepts) from the domain ontology that are relevant for each tag by leveraging the results of phase 1 and phase 2.

The semantics extracted this way can be used in the process of evolving/maintaining the domain ontology, but the set of semantically relevant tags is not a complete description of the domain [12]. Building that set may turn out to be a very long process. Also, that set cannot be used "alone". In the context of the CKP service, it is envisioned to help domain experts in the process of ontology evolution and maintenance.

The process of updating the ontology with a new concept involves placing that concept correctly in the hierarchy and retrieving appropriate parents. A number of categorization techniques have been used in order to augment ontology with a new concept: the k-nearest neighbor method (kNN), the category-based method and the centroid-based method [13]. These methods use vector-based features for representing concepts based on co-occurrence and word windows. The new concept can then be placed in the hierarchy by measuring the semantic relatedness with existing concepts in the ontology. [14] gives a good review about these methods.

The CKP service makes semantic enrichment of tags and ontology evolution continuous processes, Fig. 8. The experts create initial domain ontology in order to enable semantic annotation as well as information retrieval from the repository. During their everyday learning and knowledge building activities at the workplace, IntelLEO users annotate relevant resources thus creating a folksonomy [15]. This folksonomy evolves over time, and is used periodically to perform ontologization and thus enrich the domain ontology and improve semantic search. In other words, using the tags' acceptance (popularity) and contextualized measures of relatedness between the tags and ontology concepts [23], the CKP service provides domain experts with recommendations on how to enrich their domain ontologies.

It is important to note that all these activities related to ontology management pertain only to domain ontologies, not to all ontologies from the IntelLEO ontology framework.

4.3. Annotation Tool

The annotation tool is in charge of semantically annotating new learning resources to be stored in the shared repository, and for providing access to existing semantic annotations. This is the component that allows the client application performing semantic search to query the full set of annotations for those that meet a set of criteria/parameters. In order to query annotations efficiently, this component incorporates specialized indices that provide quick access to annotations that have specific characteristics. That way, this component enables efficient discovery, linking, and sharing of various document data across the relevant application, enterprise, and community boundaries.
The annotation tool of the CKP service is broken down into three pieces: the annotator, the persistence mechanism and the access API. The annotator creates annotations of knowledge objects stored in the IntelLEO knowledge base. The annotations take the form of RDF triplets based on the Annotation ontology [24] and appropriate domain ontologies (in our initial testing, it is the Computing ontology (based on the ACM CCS taxonomy)). Annotation instances have some general information associated with them, including temporal descriptions, as well as the information necessary to index back into the unstructured source to which the annotation refers. The annotations are stored in RDF triple stores. The access API is a client layer that provides the client application with a structured API for creating and retrieving annotations.

V. RELATED WORK AND DISCUSSION

There are a lot of research results related to the content management and information retrieval (IR). These topics have been extensively explored in the context of workplace learning, as well.

Abel et al. [16] describes the software solution, they develop within the GRAPPLE project [17], that re-use, refine and enhance previous work in the area of cross-application and generic user modeling systems. Unlike the CKP service, in this solution domain ontologies are not used. Also, this project is focused on classical technology-enhanced learning applications and does not consider the specificities of the workplace learning, which is the main area of interests in the IntelLEO.

The APOSILDE project [18] focused on the existing documents in the organization’s repository as the only resources available for supporting learning. It uses a semantic retrieval mechanism to find a set of documents related to the specific terms used in the user’s query. This mechanism relies on both, information in domain ontology and the statistical information in a collection of documents. Opposite to this approach, the CKP service uses resources available from different sources like online repositories, online communities, etc., to support learning.

In the ELENA project [19], intelligent services for learners have been developed. Personalization techniques are employed to support learners in their search for learning resources. This software offers the possibility for employees to find relevant resources by using the semi-annotated resources. However, it does not consider informal learning resources that are the major source of knowledge in workplace settings.

PROLIX project [20] consists of aligning learning with business processes in order to enable organizations to faster improve the competencies of their employees according to continuous changes of business requirements. In this approach the process of aligning learning to business is performed within a single organization. The employees do not have the opportunity to explore learning resources beyond the organization’s ‘borders’. Also, there is no collaboration between organizations and research institutions that could provide information about the newest technologies relevant for the organizations’ business processes. Finally, learning is offered in the form of formal courses that are pre-prepared and other kinds of KOs are not supported.

The CKP service differs from the other retrieval and content management services in a few ways. First, by using the CKP service within the IntelLEO software solution, employees can store, annotate and (re-)discover heterogeneous resources (e.g., documents, discussions, blog posts, wikis). The annotation is done automatically by using the concepts of appropriate domain-specific ontologies. Secondly, it allows employees to find job-specific experiences and “know-how” (in the form of, e.g., annotated wiki pages, blog posts, discussions) that are not freely available on the Web. These can originate from a member of the extended organization, a colleague from the same organization, or be a documented self-experience. Annotations of these resources with the concepts from specific domain ontologies facilitate their discovery and retrieval. The IntelLEO CKP service also aims at addressing well-known drawbacks of traditional search paradigms related to difficulties in finding relevant information [21] by improving existing search interfaces with semantic search capabilities, thus allowing the search to be based on domain topics and not only keywords. Finally, the CKP service offers a solution for leveraging folksonomies for ontology evolution through a user-driven interaction based on interactive
visualizations and recommendations that rely on semantic relatedness measures [23].

VI. CONCLUSION

The CKP service described in the paper is a Web-based learning service supporting workplace learning situations where employees are often faced with the need to efficiently find, browse, share, and otherwise use electronic learning resources.

The service is designed and implemented as a core service of IntelLEO, a new workplace learning paradigm being developed within an FP7 research project in the area of technology-enhanced learning. The CKP design is itself novel, since it uses Social Semantic Web [22] technologies to guide the process of content and knowledge provision in workplace learning.

The Web browser plug-in interface of the CKP service enables users to run it in an easy way, and from whatever Web application they use. This interface of the CKP service includes all the advantages of the bookmarklets features and also extends them for more complex operations. Evaluation of this service in real-world context is planned in the future within the three different business cases. The plan is to investigate if the synergy of harmonization and collaboration services increases learners’ motivation to get pro-actively involved in cross-organizational LKB and thus increases the responsiveness of extended organizations. In the context of the CKP service, it will be investigated if knowledge sharing process may lead to the effective collaboration of cross-organizational level.

Acknowledgments. This publication was partially supported/co-funded by the European Community/European Union/EU under the Information and Communication Technologies (ICT) theme of the 7th Framework Program for R&D (FP7). This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

REFERENCES

[17] GRAPPLE project, Online: http://www.grapple-project.org/
[18] APOSIDLE project, Online: http://www.aposdle.tue自在.at/
[19] ELENA project, Online: http://www.ele

ANNALS OF THE UNIVERSITY OF ORADEA
Fascicle of Management and Technological Engineering

221