An Approach to Folksonomy-based Ontology Maintenance for Learning Environments

Dragan Gašević, Amal Zouaq, Carlo Torniai, Jelena Jovanović, Marek Hatala

Abstract—Recent research in learning technologies has demonstrated many promising contributions from the use of ontologies and Semantic Web technologies for the development of advanced learning environments. In spite of those benefits, ontology development and maintenance remain the key research challenges to be solved before ontology-enhanced learning environments are widely used. In this paper we present an approach to ontology maintenance based on the use of collaborative tags contributed by learners while using learning environments. Our contribution is twofold: i) a visualization and user interaction interface supporting the tasks of enriching ontologies with selected collaborative tags; and ii) ontology-enhanced metrics that are used for measuring semantic relatedness between collaborative tags and ontology concepts and for recommending tags which are relevant to a given ontological concept. We developed a software architecture as a proof of concept and a tool for the evaluation of our proposal. This tool is used to conduct the evaluation of the usability and effectiveness of the proposed method.

Index Terms—Computer uses in education, ontology design, collaborative learning, Applications and expert knowledge-intensive systems

—————————— ◆ ——————————

1. INTRODUCTION

The emergence of the Semantic Web technologies in computer-based education has enabled the development of next-generation semantic-rich eLearning environments and has already provided some interesting results [22][24]. However, one of the main challenges for the next-generation learning technology-based environments is the development of a more effective and efficient paradigm for the integration and interaction between information, learners and experts, based on the learning context and semantics [32].

The Semantic Web vision relies on ontologies as its main knowledge structure. However, ontologies are difficult to build and maintain. This is the main hurdle preventing a broader adoption of ontology-based e-learning environments [47][56]. In the learning technology community, ontologies are used to model various aspects of the educational process including (but not limited to) domain knowledge, knowledge artifacts, pedagogical models, user behavior and characteristics, and social interactions. Furthermore, learning is a dynamic process, which requires that domain ontologies describing the learning process need to evolve to reflect changes. In fact, the need for the constant rebuilding and maintenance of

domain ontologies is one of the main challenges that face current semantic-rich learning environments [18].

Recent efforts to increase the availability and reusability of ontologies have focused on the development of online ontology libraries [16] (e.g., Swoogle) or (semi-) automatic ontology development tools; however, the usage of these libraries and tools still require a high level of technical knowledge. In general, educators lack the knowledge required to effectively use such tools, as they reveal the complex details behind semantic web technologies and are more tailored to knowledge engineers. Moreover, the available ontologies may not adequately describe course content thus creating a semantic mismatch between the content and the ontology. Hence, educators are in need of interfaces that provide support in building and maintaining ontologies. Such interfaces should enable them to focus on their domain of expertise [17][18]: course content (development) and its effective conveyance to learners.

Another problem that arises from the current ontology-based learning environments is that they generally implement a very traditional instructional approach, in which learners are recipients of course content. Additionally, they have strong groundings in the traditional ideas of intelligent tutoring systems [31], which assume that a learner is *tutored* and where creation opportunities for learners are limited. Here, we look at learning environments in the broadest possible sense, where such learning environments need to support learning at all the levels of the Bloom's taxonomy [3]. For this paper, of special interest are Bloom's levels related to evaluation and creation, both of which require high degrees of self-regulated and -directed learning; such learning aims at stimulating and rewarding creativity of learners. This is also a first precondition for supporting social constructivism principles. In our research, another equally important precondition

- *Dragan Gašević and Amal Zouaq are with the School of Computing and Information Systems, Athabasca University, 1 University Dr. Athabasca, AB T9S 3A3, Canada. E-mail: {dragang, amal.zouaq}@athabascau.ca*
- *Carlo Torniai is with the Oregon Health & Science University Library, 3181 SW Sam Jackson Park Rd. – LIB, Portland, OR 97239-3098. Email: torniai@ohsu.edu*
- *Jelena Jovanović is with the Faculty of Organizational Sciences, University of Belgrade, Jove Ilica 154, 11000 Belgrade, Serbia. E-mail: jeljov@gmail.com.*
- *Marek Hatala is with School of Interactive Arts and Technologies, Simon Fraser University, 250-13450 102 Avenue, Surrey, BC V3T 5X3, Canada. Email: mhatala@sfu.ca*

for enabling social constructivism in learning environments is the provision of opportunities for a learning group to create a small culture of shared artifacts with shared meaning. This sharing should not be limited to the common understanding of only one instance of, say, a university course with never-changing course ontologies. Creating and sharing should happen within a community (or even among communities [5]) with a much longer time span in which different learners and educators participate in different and not-necessarily overlapping periods of time. Such a concept, intrinsically embedded in social constructivism, certainly requires a different approach to domain conceptualizations. In our case, this different approach would be based on the inclusion of both folksonomies and ontologies as artifacts of shared knowledge. In section 4, we show empirical results obtained through a user study that evaluated the perceived value by the target users (i.e., educators) of this approach.

The above-stated need for knowledge sharing (consistent with the social constructivist principles) is further intensified by the novel forms of online interaction brought by the social-media era. It calls upon the environments that are better aligned with the social constructivist principles, where learners' views can be represented and shared [52]. Students' perception of the course content, often reflected in the tags that they use to annotate content, may differ from the course conceptualization encoded in the domain ontology and may prove very useful to identify knowledge gaps. Actually, a recent study showed some promising opportunities for improved knowledge acquisition through collaborative tagging [26]. In general, with the motto "ontologies are us", the Semantic Web research showed a high relevance of social networking principles for knowledge acquisition and ontology development [39]. The above arguments clearly justify why more participatory (often referred to as "Web 2.0") approaches, such as collaborative tagging, have so far gained a lot of attention in technology-enhanced learning [8][34][52]. However, folksonomies, as structures of collaborative tags that are created by a community, suffer from problems related to the ambiguity of tag semantics, including ambiguous tag meaning and the lack of a coherent categorization scheme; not to mention the amount of time and the size of the community required for their emergence and stabilization [40]. For these reasons folksonomies, unlike ontologies, might increase ambiguity and lead to less precise results in the automated data analysis process.

In this paper, we propose an approach to leveraging folksonomies for the maintenance of domain ontologies used in learning systems. Through the collaborative tagging, learners create a folksonomy, which reflects the learners' perception of the domain under study. Educators, who maintain domain ontologies, may leverage these folksonomies as a useful source of (community) maintenance knowledge. To design such an ontology maintenance approach, in this paper, we focus on the two research challenges: i) creating an intuitive and highly-usable environment for maintenance, which hides the complexity typically attributed to ontology engineering

tools [24]; and ii) providing effective recommendations based on the computed relevance of folksonomy tags to domain ontology concepts. Therefore, the main contributions (Section 2) of this paper are as follows:

- an **ontology-folksonomy visualization** and **interaction** which offers an intuitive interface for the maintenance and manipulation of a domain ontology and a tag cloud;
- an efficient and automatic method to compute relations among tags and domain concepts using **measures of semantic relatedness**;
- an **ontology-based enhancement** of semantic relatedness; this enhancement relies on ontology subsumption relationships to contextualize values of the measures of semantic relatedness.

To be able to make use of the three contributions in developing concrete solutions, we developed a software architecture for folksonomy-based ontology maintenance in learning environments (Sec. 3). Our claimed contributions are implemented as an extension to LOCO-Analyst, a tool for educational feedback provisioning [29]. This implementation provided us with a suitable setting for answering some relevant research questions such as the perceived usability of the proposed visualization and interaction interfaces for ontology maintenance in learning environments (Sec. 4); and the effectiveness of an ontology-based enhancement of semantic relatedness measures, which are used for recommending collaborative tags with respect to ontology concepts (Sec. 5). In the paper, we also discuss the limitations of our experiments (Sec. 6) and compare our work with the state of the art (Sec. 7).

2. PROPOSED METHOD

As indicated in the introduction, in a social constructivist learning environment, the process of knowledge creation and evolution is constant. The process happens throughout the different dimensions of interaction (e.g., six dimensions of interactive learning environments as per [4]) and creation of shared knowledge artifacts with a commonly shared meaning [54]. In this paper, knowledge artifacts are ontologies. Such ontologies are not to be merely used as course ontologies, but they are rather to establish shared meaning within and across communities of learners[1]. Due to the well-known knowledge acquisition bottleneck, for a specific learner community (e.g., a community of learners who have taken a specific course in the period of several years at a specific university) one can reuse a general domain ontology. While this is not an ideal case, our empirical experience shows that even such ontologies can produce rather useful results. For example, we experimented with LOCO-Analyst at Athabasca University on two master's level courses for learning analytics with the feedback types outlined in [29]. In this example, the course ontology was defined on top of the ACM Computing Classification System (CCS), which is used in the experiments described later in this paper. Our exper-

---

[1] It is important to stress that domain ontologies, in learning environments, can be connected with course ontologies, ontologies for competences, learning flows and designs, and other dimensions important for a learning process (e.g., LOCO ontology framework [29])

iment showed that the produced effect of this ontology for generating learning analytic feedback types of LOCO-Analyst is highly positively valued by educators. To better capture the overall semantics of the knowledge of a specific community (e.g., further improve the granularity level of those analytics feedback types [29]), we also want to update such a general domain ontology with different sources of knowledge, which are produced by the participants in the learning process of the community (i.e., students and educators). Another important example with which we empirically experimented is the case when a community of learners wants to search external sources of learning content by using terminology they are familiar with. In [21], we also showed how ACM CCS can be adapted with course specific terminology and later leveraged for cross-community search tasks.

In this paper, the source of knowledge of primary interest is collaborative tags created by a community of learners and then incorporated by educators. However, ontology evolution in learning environments should not be limited to this source of knowledge. The other sources could be also specific artifacts collected and/or created by students (e.g., Web articles) or even textual content of the course itself. While in our other research we covered those other sources and leveraging of automated techniques for learning and evolving ontologies from text [24][56], in this paper we specifically focus on collaborative tags for ontology maintenance and evolution. In the rest of this section, we present the main elements of our method for ontology maintenance.

2.1. Ontology Maintenance Operations

Ontology maintenance [27], as an area of ontology engineering, is strongly grounded in a much more established discipline of software maintenance and evolution [38]. Two key issues in software (as well ontology) maintenance) are sources of change [10] and maintenance operators. Due to the nature of ontologies, it is hard to limit sources of change only to explicitly defined new requirements and defects, as it is common in software maintenance. Rather, ontology maintenance requirements emerge from the activities of the community. In the case of this paper, collaborative tags are investigated as a source of knowledge evolution and maintenance. According to [1], software comprehension is the foremost technical issue which determines "how quickly a software engineer can understand where to make a change or a correction". In fact, this is also defined in the well-known ISO 9126 standard, as one of the main characteristics of maintainability – understandability. In this paper, we only focus on this characteristic of maintainability. In our case, in order to provide for better comprehension of the relatedness between ontology concepts and collaborative tags, we introduced a visualization and interaction method described in Sect. 2.2, which is empowered with the semantic relatedness measures presented in Sect. 2.4. For better inspection and navigation through the ontology, the proposed interaction and visualization is implemented with various features for searching, drag-and-drop, and zooming operations as explained in Sect. 3.

Maintenance operators typically include: adding new elements, updating, refining, merging, and removing existing ones. All these are supported in the proposed ontology interaction (Sect. 2.2) and its implementation (Sect. 3). The measures of semantic similarity (Sect. 2.4) are however primarily designed to empower comprehension of ontology maintainers for the adding and refining operators. The adding operator assumes adding new classes into the ontology under maintenance. Those classes can be then associated with other already existing classes. The refining operator assumes changing class names, adding new class aliases, and creating equivalent classes.

2.2. Domain Ontology-Folksonomy Visualization and Interaction

One of the main drawbacks of current ontology editors is their lack of an intuitive interface that helps users in the ontology maintenance and management tasks. This is especially true in the domain of technology enhanced learning, where clear and straight-forward interactions targeted at users with limited knowledge of the semantic web-related technologies is an important part for the success of any tool.

Our primary goal is to provide educators, who are our target ontology maintainers, with an environment in which they can comprehend and intuitively interact with a domain ontology under maintenance and a folksonomy of a community of interest (e.g., a study group). Thus, our method for user interaction in ontology maintenance is supposed to support the common tasks for ontology maintenance by leveraging well-known and intuitive user interaction operations. This is the reason, why we decided to uses tag clouds[2] and ontology graphs (Fig. 1). The learning content that is tagged can take any level of granularity level for content units, as per the ALOCoM ontology [30]. The ontology visualization supports this task with features that allow users to search for ontology concepts based on keywords. The ontology visualization tool also has several facilities for zooming.

The next important task for ontology maintenance is to identify relations between tags and ontology concepts. In our approach, we decided to make use of *tag coloring as an indicator of relatedness* of tags to a given concept. That is, the tag cloud leverages the size and color of the presented tags to convey to educators information about the tags' popularity and relevancy, respectively. The size of a tag reflects its popularity, which is calculated by the number of times that tag was used to annotate a particular piece of learning content. The saturation of a tag's color reflects its relatedness to the selected concept of the domain ontology (see Fig. 1) – *a darker color denotes a more related tag*. This relatedness is computed by using measures introduced in Sect. 2.3 and 2.4. This is **the point of connection between our user interaction interfaces** (introduced here and evaluated in Sect. 4) **with measures of semantic relatedness** (introduced in Sect. 2.3-4 and evaluated in Sec. 5).

---

[2] As we had several options for representing tag clouds, we conducted a small scale pilot study showing participants different kinds of tag clouds, among which was a tag cloud that used different colors. The results of this pilot showed that the "standard" one (typical for the majority of apps that male use of tag cloud) was perceived by the participants as the most intuitive, and thus the reason for its selection in our approach.
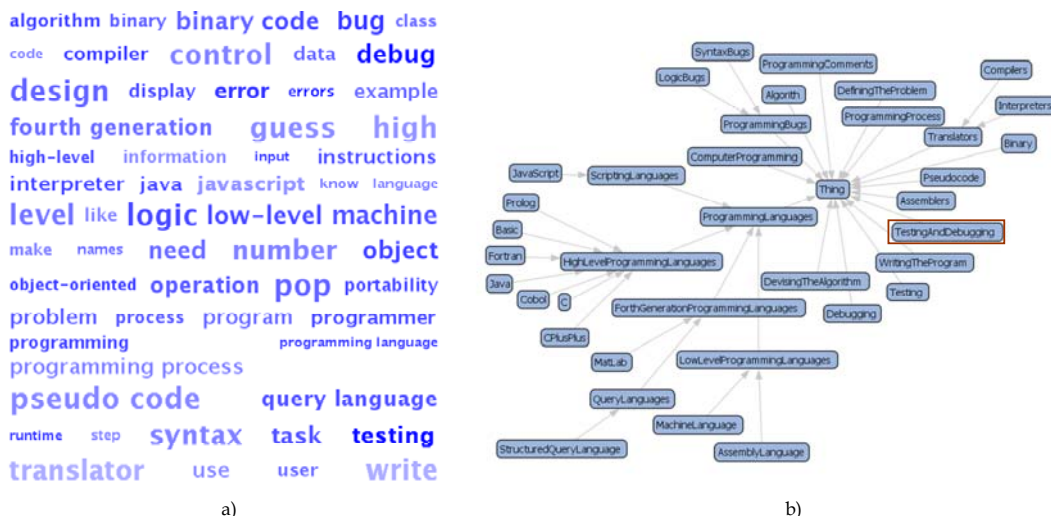
Fig. 1. Interaction interface between tag clouds and ontology graphs: a) Tag cloud contains tags whose popularity determines their size, while saturation of their color is determined by the similarity to the selected ontology concept; and b) Interactive ontology visualization. In the example from the figure, concept Testing and Debugging in the ontology graph (b) is selected, while the saturation of the color of the tags in the tag cloud (a) is based on their level of semantic relatedness with the selected concept.

The final task for ontology maintenance analyzed in this paper, is editing of ontologies based on the tags from folksonomies. Given that the source of ontology updates are collaborative tags, our proposed interaction method should facilitate an intuitive way for editing ontologies based on the tags. In our implementation, we supported this with the ontology visualization that has features for editing by using drag-and-drop interactions from the tag cloud. Once a tag is dropped on an ontology concept, the user interface associated with the ontology visualization allows for connecting the dropped tag with the selected concept through different relation types – subclass, synonym, or some custom relations. It also has various highlighting features for related concepts and concepts of a particular learning object which facilitates browsing and comprehension of the ontology.

2.3. Measures of Semantic Relatedness (MSRs)

As indicated in the tag-concept visualization and interaction component of our research, the color saturation of tags reflects the relatedness between a given concept and a set of tags from a tag cloud. We foresee several methods to relate tags to a domain ontology including using algorithms for determining semantic relatedness, eliciting expert ratings, or calculating co-occurrence of tags and ontology concepts from past experience. Our approach to establishing and weighting the relations between collaborative tags and ontology concepts is to use *measures of semantic relatedness*, which have been used successfully in the field of natural language processing to assess the similarity between terms based on some corpus [23][43].

MSRs can be defined as computational means for assessing the relative meaning of terms [53], and assigning values that describe the degree to which two terms are related. Here, these terms are represented by pairs of concepts and tags. Some learning content can be indexed by a set of ontological concepts and can be annotated by a set of learners' tags. For this reason, we propose computing the semantic relatedness among each element of these two sets. The most related concept-tag pairs are then proposed to the educator to update the ontology.

For experiments and implementation of our approach, we use the MSR Server [53], which implements various MSRs including Normalized Search Similarity (NSS), Point-Wise Mutual Information (PMI), WordNet-based measures, and Latent Semantic Indexing. We chose two widely used metrics among those implemented by the MSR Server: PMI and NSS. PMI is a well-established and proven measure for approximating human semantics [50]. PMI is based on the probability of finding two terms of interest (t1 and t2) within the same window of text versus the probabilities of finding each of those terms separately. NSS [13] measures the similarity between two terms by using probabilities of co-occurrences extracted using the Google corpus (i.e., the Web). The use of the Web as a corpus has gained more and more importance [23] as the large amount of data permits to discover interesting associations and guarantees maximum coverage.

MSR measures are trained on various corpora such as the entire Web, the New York Times or Wikipedia. Depending on the selected corpus, the performance of the measure may differ. This is why we chose to test various combinations of measure-corpus pairs to identify the best performing one(s). Performance is measured by comparing MSR answers with human answers taken as a gold standard. This will be further explained in Section 5.

It is important to indicate that corpora such as Wikipedia or the entire Web are not complete and fine grained models of many domains. Still, they can be considered rather representative models of the world. As such, they might be a solid source for calculating probability of concurrence or semantic relatedness along with all the limitations that can be introduced in this calculation process. The value can especially be important for semi-automatic systems, where computer measures are only used to recommend, while human users make the final decisions (i.e., our approach). In an ideal world, we would have a community created and standardized ontologies. However, the history of ontology engineering showed that it is hard to expect that ontologies will be created as standard-

ized ontologies; at least not in the foreseeable future. The majority of the Semantic Web community abandoned that approach; social technologies and lightweight ontologies are looked at as much more promising research venues, and Wikipedia is the best known example.

We also wanted to use WordNet-based measures (as WordNet provides a curated set of links among the terms it contains). Those were mostly unsuccessful in identifying similarity values between concept-tag pairs in our experiments. Corpora such as the entire Web and Wikipedia on the other hand have grown at a quick pace and contain many more terms (including domain specific); the reason why results based on these corpora are more successful.

2.4. Ontology-based Weighting of Semantic Relatedness

In addition to the aforementioned metrics, we were interested in exploring how the relatedness weight might be affected by the taxonomical structure of the ontology under maintenance. In fact, many traditional methods for computing semantic relatedness rely on hierarchical links and explore path lengths among nodes in taxonomies [45] to identify concept similarity. Therefore, instead of depending only on a given concept, semantic relatedness in this work also relies on the context of this concept (parents and children) to find the most accurate links between that concept and the folksonomy tags. In other words, we wanted to emphasize the context in terms of the domain ontology under maintenance, so that MSR values can be contextualized in that sense as well. That is, contextualization also considers the relatedness of the surrounding of a given concept with a collaborative tag, where the surrounding is represented with the concepts that are related to the given concept through hierarchical relationships. In what follows, we describe two methods to compute a context-based relatedness measure.

We can define the *context* of a concept C in an ontology α as being C_α, therefore:

$C_\alpha = \{Sup_1(C), Sup_2(C), ..., Sup_m(C), Sub_1(C), Sub_2(C), ..., Sub_k(C)\}$

where $Sup_i(C)$ (i=1..m) are the concepts to which C is related with super-class relations and $Sub_j(C)$ (j=1..k) are the concepts related to C through sub-class relations.

In order to take into account the context of a concept when computing its relatedness with a tag, we define the *Weighted Measure of Semantic Relatedness* (WMSR) between a concept c_i and the tag t_j as follows:

$$WMSR(c_i, t_j)_n = MSR(c_i, t_j) +$$

$$+ \frac{1}{|Sub(c_i, n)|} \sum_{k=1}^{|Sub(c_i, n)|} \frac{MSR(Sub(c_i, n)_k, t_j)}{Dist(c_i, c_k) + 1} +$$

$$+ \frac{1}{|Sup(c_i, n)|} \sum_{k=1}^{|Sup(c_i, n)|} \frac{MSR(Sup(c_i, n)_k, t_j)}{Dist(c_i, c_k) + 1} \quad (1)$$

where $MSR(c_i, t_j)$ is the measure of semantic relatedness between concept c_i and tag t_j. MSR can be any of the measures mentioned in Sect. 2.3, and in our evaluation in Sect. 5, we experimented with different MSRs; $Sub(c_k, n)$ is the predicate that returns all the sub-concepts of concept c_k where the sub-concepts are up to n sub-concepts rela-

tionships distant from c_k; and $Sup(c_k, n)$ is the predicate that returns all the super-concepts of concept c_k where the super-concepts are up to n super-concepts relationships distant from c_k. It is important to emphasize that both these predicates return all the sub-/super-concepts of c_k, not just immediate ones. Each sub-concept and super-concept MSR is weighted by its distance (Dist) from the given concept c_i. However, given the use of MSR measures, it is important to indicate that they already consider the relatedness of terms in a semantic space and measure their semantic distances. Thus, if we introduced weights, one could anticipate that that the gap between terms will further be modified.

Thus, we introduce the second metric called ***non-Weighted Measure of Semantic Relatedness (nWMSR)***, which is the same as (1) except that it does not assign any weight to each sub-concept and super-concept MSR based on its distance from the given concept c_i. It is calculated as follows:

$$nWMSR(c_i, t_j)_n = MSR(c_i, t_j) +$$

$$+ \frac{1}{|Sub(c_i, n)|} \sum_{k=1}^{|Sub(c_i, n)|} MSR(Sub(c_i, n)_k, t_j) +$$

$$+ \frac{1}{|Sup(c_i, n)|} \sum_{k=1}^{|Sup(c_i, n)|} MSR(Sup(c_i, n)_k, t_j) \quad (2)$$

The formulation in (2) states that the non-weighted measure of semantic relatedness between a concept c_i and a tag t_j is computed as the average of the measures of semantic relatedness of concepts related to c_i (i.e., the concepts included in its *context*).
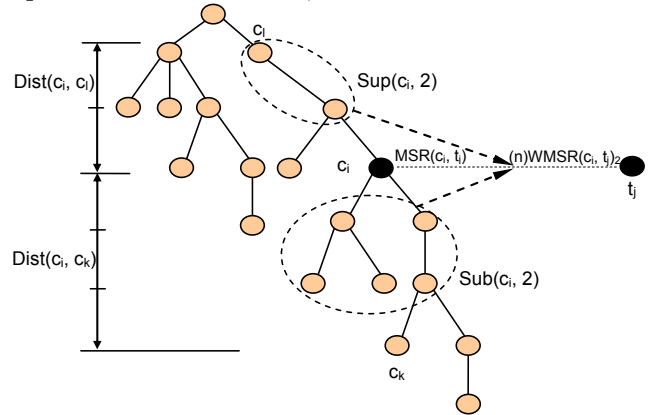

Fig. 2. A graph depicting the hierarchy structure of a sample ontology. The figure explains the computation of the *Dist*, *Sub*, and *Sup* predicates. Dist returns the distance between the two ontology concepts connected through the sub-/super-classing relations (e.g., *Dist(c_i, c_k)* = 3 and *Dist(c_i, c_l)* = 2). The *Sub* (resp. *Sup*) predicate returns a set of subconcepts (resp. superconcepts) for a given distance *n*. The figure also demonstrates the impact of super- and subclasses on computation of (n)WMSRs.

Figure 2 shows how the two metrics use the domain ontology hierarchical structures to return sub-concepts, super-concepts and their distance from the given concept.

3. ARCHITECTURE AND IMPLEMENTATION OF A SEMANTIC-RICH E-LEARNING ENVIRONMENT

Aiming to provide an appropriate context for deploying

our proposed method, we designed the required software architecture which is illustrated on Fig. 3. The main challenge here is to be able to integrate learners' tags and ontologies developed by educators as two different viewpoints that can enrich each other. To combine these two perspectives, the metrics defined in Sections 2.3-2.4 are used to determine the relatedness between the domain concepts originating from the ontology and students' tags. The framework is then able to suggest the most appropriate tags for a given concept based on the returned semantic relatedness value. The measures used for this relatedness could be selected based on the results reported in Section 5. In our work, we implemented our architecture in order to have a proof of concept and a tool for the evaluation of our proposal. The implementation is an extension of our ontology-based tool LOCO-Analyst. LO-CO-Analyst was designed to provide educators with information about the use of their content and pedagogies. Thus, the integration of our facilities for ontology maintenance naturally complemented an existing tool supporting the evolution of learning content [47]. Of course, this implementation is just one possible (context of) implementation and developers might find some other tool more suitable for the implementation of these ideas.
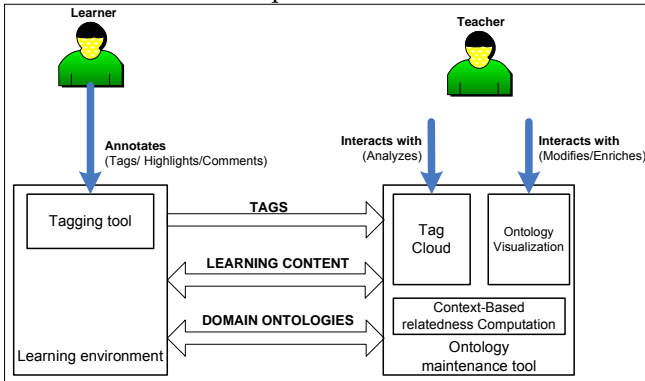

Fig. 3. A framework for a folksonomy-driven ontology maintenance

In our experiments, the Open Annotation and Tagging System (OATS) was used for collecting tags. OATS allows learners to create and share knowledge by allowing students to add highlights, tags and notes in Web-based content [6]. The tool has so far been integrated into the iHelp Moodle Courses Learning Content Management Systems (LCMSs) and allows learners to tag learning objects. These learning objects are also annotated by the educator using a previously-developed domain ontology. Results of students' tagging activities are accessed by the LOCO-Analyst tool which in turn makes them accessible to educators. LOCO-Analyst also provides an interactive visualization of the course domain ontology aiming to facilitate the process of ontology maintenance. This interactive visualization includes all the features described in Sec. 2.2, including ontology graph display, easy drags and drops from the tag cloud to the ontology graph, zoom-in and zoom-out capabilities, etc. Computation of (non-)weighted semantic relatedness measures among tags and concepts (Sec. 2.4) is used to assist educators in ontology maintenance by suggesting visually the most relevant tags for a particular concept (Sec. 2.2).

Fig. 4 presents the user interface of LOCO-Analyst that enables educators to refine domain ontologies (Fig. 4, item C) based on students' tagging activities (Fig. 4, item B) captured in OATS. An educator's interaction with the LOCO-Analyst's features for ontology maintenance can be described as follows: as the educator selects a lesson (or a complete learning module) from the tree-like representation of the course structure (Fig. 4, item A):

- The visual representation of the ontology (Fig. 4, item C) changes to emphasize the concepts relevant for the selection being made. More precisely, ontological concepts referenced in the content of the selected lesson change color to become visually distinctive.
- The tag cloud (Fig. 4, item B) is populated with tags related to the selected lesson.
- The educator selects (in the visual representation of the ontology, Fig. 4, item C) a concept that (s)he wants to inspect. As soon as the concept is selected, the tag cloud changes, displaying the tag color saturation according to the computed relatedness to the selected concept. The educator is then free to choose a tag (from the tag cloud) that (s)he finds the most relevant for the selected concept and drag-and-drop it over the concept. Once this is done, a popup menu appears offering different kinds of relationships for establishing a connection between the selected concept-tag pair. As soon as the selection is made, the ontology is updated allowing the educator to see his(her) changes in real-time. The educator can also postpone a decision for later in which case this potential relation is automatically added to the user's notes for later reflection.

In the next two sections, we report the results of our experiments that aimed to evaluate:

i)   The perceived value of the tag-concept visualization and user interaction for ontology maintenance in learning environments (Sect. 4), and
ii) The effectiveness of MSR, WMSR, and nWMSR measures for ontology maintenance using folksonomies based on our proposed method (Sect. 5).

4. USABILITY EVALUATION

In our usability evaluation, we wanted to investigate the following research questions:

RQ1 – What is the perceived intuitiveness and usability of the proposed method for ontology maintenance?
RQ2 – Is there any relation of the perceived intuitiveness of the ontology maintenance process with the used ontology visualization and interaction interfaces?
RQ3 – Is there any difference in the perceived value of the proposed ontology maintenance method between different groups of participants – instructors, teaching assistants, and research students/practitioners?
RQ4 – What are the most and least valued characteristics of the proposed ontology maintenance method?

4.1. Methods

4.1.1. Design

To investigate the perceived usefulness of the proposed

ontology maintenance method, we wanted to study users' impressions after a session with the tool supporting the proposed method. Users' observations were obtained through a questionnaire, which was used after the session

with the tool. Once data were collected, we used quantitative and qualitative (coding and content analysis) methods for data analysis.
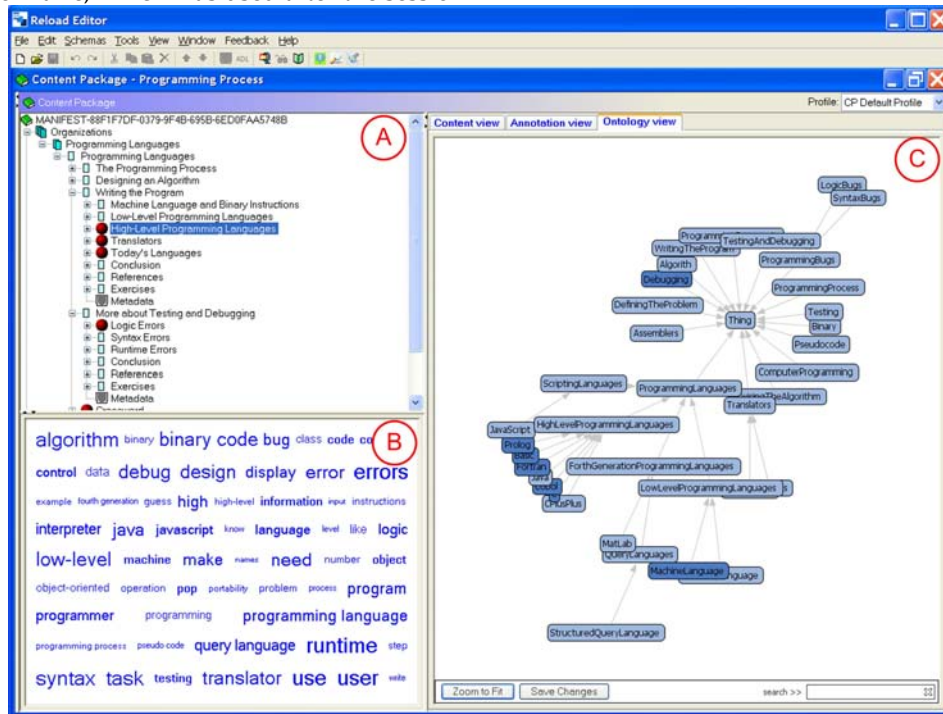

Fig. 4. LOCO-Analyst's user interface for the ontology maintenance: A) Tree-like representation of the course structure; B) Tag cloud; C) Visual representation of the ontology. In the LOCO-Analyst implementation, the concepts of the ontology (B) which are related to the selected lesson (A) are darker colored. The color saturation principles illustrated in Fig. 1 for relations between concepts and tags remain the same.

4.1.2. Participants

For our experiment, participants were recruited in October 2009 from Simon Fraser University, Athabasca University, University of Belgrade and a private Canada-based company developing and offering technology and content for professional training. Overall, 22 persons (17 men and 5 women) responded to our invitation and all of them successfully completed all the steps of the experiment. The participants were also asked to express their role in online education. We distinguished between the following three roles:

- *instructors* – persons who had independently instructed at least one entire course. There were six participants in this group and they had on average 10.67 years of experience (Standard Deviation, SD=7.09);
- *teaching assistants* - persons who had had previously only teaching assistant experience. There were eight participants in this group and they had on average 3 years of experience (SD = 1.06); and
- *research students/practitioners* – persons who had done research related to online education, or practiced online education in industry through software and content development and delivery. There were eight participants in this group and they had on average 6.75 years of experience (SD=5.23).

4.1.3. Materials

The LOCO-Analyst tool with its features for ontology maintenance was presented to the participants. To

demonstrate implemented features of the ontology maintenance process in the LOCO-Analyst tool, we created video clips describing each individual feature in detail. The clips also served as a guide on how to use the implemented functionality and made sure that its interpretation was clearly carried to the participants of the study. These videos are available on the website of LOCO-Analyst[3]. The participants were provided with a complete and correct domain ontology (i.e., ACM CCS) and a set of collaborative tags; the set is described later in Sect. 5.1.2.

The evaluation of the ontology maintenance method was done together with a general evaluation of all the other features of the LOCO-Analyst tool using a questionnaire. While the general questionnaire consisted of 21 questions, three questions specifically addressed the ontology maintenance method. The three questions had the statements as shown in Table I and answers to them had two parts: i) a five level Likert scale answers where each level had an associated code on the 1-5 scale expressing the level of agreement with the statement (i.e., from Strongly Disagree – 1 to Strongly Agree – 5); ii) an open-ended part allowing participants to further reflect on the asked question in a free text form. The latter part was optional. Each question in the questionnaire had an URL of the specific video clip to which the question was related.

4.1.4. Procedures

The participants were presented with guidelines that ex-

[3] http://www.jelenajovanovic.net/LOCO-Analyst/videos.html

plained the purpose of the evaluation and outlined the steps they should take. In a nutshell, the participants were asked to watch the demo videos explaining the functionality of the tool. They were then asked to download the tool and try the presented functionality. They were also encouraged to send any further clarification questions to the evaluation team. In the guidelines, we asked them to perform the implemented functionalities of the method for comprehension and maintenance operators outlined in Sect. 2.1. Together with the guidelines, we also supplied the questionnaire. Once finished, the participants were asked to send the completed evaluation questionnaire back within a week from the time of their initial acceptance to participate in the study. Finally, after receiving the answers from all the participants, we entered answers into an Excel spreadsheet for further analysis.

4.1.5.   Content Analysis

To analyze the observations in open–ended questions, we followed the approach introduced in [35]. Initially, we developed a coding scheme based on the participants' answers. The coding scheme consisted of three general categories: i) *Positive comments* – expressing positive opinions without any concerns; ii) *Positive comments with some observations* – expressing positive opinions, but the participants either had some observations that questioned some decisions or suggested some improvements; and iii) *Negative comments* - expressing either negative observations or some concerns questioning the decisions made in the design. Each of these three categories were further subcategorized into three new subcategories, namely: i) *Feedback features* – observations about specific feedback mechanisms supported by the user interface of LOCO-Analyst (not applicable to the ontology maintenance features of interest for this paper); ii) *Intuitiveness* – observations about the intuitiveness of the user interface; and iii) *General comments* - conceptual comments, applicable to different features of LOCO-analyst (not necessarily to ontology maintenance).

The early version of the coding scheme was first tested by two raters. To perform the testing, they applied the scheme to five randomly selected answers to each of the three questions (Table I). Consequently, they fine-tuned the scheme and revised the usage guidelines. In the next step, the two raters applied the fined-tuned scheme independently to rate all the answers. This was followed by a meeting of the two raters where all the differences in the assigned codes to each individual answer were reconciled. Finally, to evaluate the reliability of the inter-rater agreement, we used Cohen's kappa. The result of 0.88 of Cohen's kappa can be interpreted as an *almost perfect agreement* according to the conventional interpretation [4].

4.2. Results

4.2.1.   Quantitative Analysis

Before discussing the specific results, we report the internal reliability of the collected Likert scale data. For this, we used the standard Cronbach's α coefficient. We obtained $\alpha = 0.90$ which is higher than 0.80, the value typically used as a minimal threshold for reliability.

To evaluate the perceived level of intuitiveness and usability of the proposed method for ontology maintenance (i.e. RQ1), we used the descriptive statistics (Table I). The presented values are based on the participants' responses to the questions using the five-level Likert scale.

Table I. Descriptive statistics of the participants' answers to the three Likert scale questions: M – mean, SD – standard deviation, N – number of answers.

Code	Question Content	M, SD, N
Q1	I found the process of extending the domain ontology intuitive and easy to accomplish	4.64, 0.79, 22
Q2	I found the visual representation of the ontology and the interaction with it helpful for the ontology maintenance process	4.86, 0.47, 22
Q3	I found students collaborative tags as a relevant source for ontology maintenance	4.23, 0.87, 22

It is apparent that almost all participants strongly appreciated the ontology visualization and interaction proposed in our ontology maintenance method (Q2). That is, 20 out of 22 participants strongly agreed that the process is intuitive and easy to accomplish. Just slightly lower, but still very highly recognized is the intuitiveness of the ontology maintenance process (Q1). For the question about the suitability of the use of student generated collaborative tags (Q3), the descriptive statistics reveal a high approval by participants. Overall, the participants expressed very positive attitude about the intuitiveness and usability of the proposed method. Still, some salient comments emerged in the open-ended answers, which are reported in the results of the qualitative analysis.

To determine if there is any relation between the perceived intuitiveness of the ontology maintenance process and the ontology visualization and interaction interfaces (i.e., RQ2), we calculated Pearson's bivariate correlation (two-tailed) between observations stated in the answers to Q1 and Q2. The results reveal that there is a *significant association* of the proposed *ontology visualization and interaction* with the *intuitiveness and ease-of-use of the proposed maintenance method* ($r = .633$, $p < 0.01$). These results corroborate our previous experimentation results where educators also indicated that a graph-based visualization of ontologies is rather intuitive for the ontology representation [24]. Yet, that experiment [24] also revealed that an ontology visualization is not enough and can even be confusing if there is no effective interface for the interaction of users with the visualization.

To address RQ3, we used one-way ANOVA to test if there is any difference in the perceived value of the proposed ontology maintenance method among the three groups of the participants. For each of the three questions, our results showed no significant difference between the three groups (i.e., Q1 – $F_{(2, 19)} = 1.084$, $p = .358$; Q2 – $F_{(2, 19)} = .565$, $p = .578$; and Q3 – $F_{(2, 19)} = 1.076$, $p = .361$).

4.2.2.   Qualitative Analysis

The goal of the qualitative analysis was to investigate the most and least valued characteristics of the proposed ontology maintenance method (i.e., to address RQ4). Table II presents the percentage of the total number of answers as per their categorization obtained by applying the coding

scheme in our content analysis. Please note that not every participant provided answers to all the open-ended parts of the questions, as they were optional (i.e., 72.74% participants provided open-ended answers to Q1, 68.19% to Q2, and 54.44% to Q3). The responses of the participants are predominantly grouped in the first two categories – positive comments and positive comments with some observations. This directly addresses our RQ1 and further corroborates the results of the Likert scale responses, confirming an overall positive perception of the intuitiveness and ease-of-use of the proposed method and its tooling.

To address RQ4, we provide here the specific qualitative observations of the participants. We start with the observations related to Q1. A large majority of positive comments stressed **the importance of the visualization** of the ontology in the process and that it was reportedly a missing feature in the other related tools (the participants already experienced some of these tools such as Protégé). The participants mentioned some specific features related to the ontology maintenance and leveraging collaborative tags. In particular, a few participants appreciated the use of "drag-and-drop". The participants also appreciated the supported navigation through ontologies/folksonomies and ontology editing such as "… *the simplified method of adding new topics as subclasses or related topics.*" Finally, the participants appreciated the implemented functionality to search ontologies with keywords, as important for large scale (real-world) ontologies.

Table II. Frequencies of the participants' observations according to the codes assigned during the content analysis. Formulations of questions Q1-3 are given in Table I.

Category	Sub-category	Q1	Q2	Q3
Positive Comments	Intuitiveness	45.45%	54.55%	-
	General comments	13.64%	13.64%	50.00%
Positive Comments with some observations	Intuitiveness	4.55%	-	-
	General comments	4.55%	-	4.55%
Negative Comments	Intuitiveness	4.55%	-	-
	General comments	-	-	-

On the other hand, some participants, in spite of appreciating the given visualization, expressed some concerns on the lack of enough guidance: "*Visualization is good. However the interaction with drag-and-dropping the words from tag cloud to the ontology concepts could be made more evident in the interface (display some tips that it can be done, etc.)*". In fact, this type of observations are in accordance with our other experiment in the area of ontology engineering [24] where participants also expressed **a need for better guidance** in the ontology development process. This is certainly an important topic to be investigated in future research and to be carefully addressed in the development of similar types of tools. This also indicates that a more explicit user interface intervention is needed in addition to the supported tag coloring. Also, the participants raised another important concern – how to effectively support ontology comprehension when there are so many crossing links representing properties among concepts in the ontology visualization. Indeed, this has re-

cently been recognized as an important research challenge in the semantic technologies research community [33]. Only some preliminary work has been done proposing a more comprehensive visualization based on different coupling metrics [20].

An observation of another participant is even more critical in this regard, since it points out that the current approach lacks any indication if a tag has already been included into the ontology: "*This can be a problem if [an] ontology has many concepts and it's hard to visually see if a tag doesn't appear in the ontology. In this case, [a] teacher must first search for the tag using [the] search field. A solution to this can be to color differently tags already included in the ontology, or filter just the tags which are not in the ontology (using, for instance, checkbox).*" This can certainly be a valuable input for improving the intuitiveness of the support tool. This is in line with the HCI research which indicates that differences in color are detected faster than any other visual variables [55]. Although this is to some extent leveraged in our research, there are certainly many other aspects that should be investigated.

From the above comments on the process, it is very clear that the majority of the participants fully equated the maintenance process with its actual tooling support, i.e., visualization and interaction interfaces. This corroborates the earlier reported association in the quantitative results. In addition to the already mentioned observations, the participants, in response to Q2 from Table I about the proposed visualization and interaction interfaces, also indicated the appreciation of the use of different colors, effective use of the small screen space for complex visualizations, and that the tool uses "*no excessive and useless options, no[t] trying flashy effects.*" Also, they indicated that the tool had a better visualization comparing to the other ontology tools they knew of, such as Protégé.

When asked about the usefulness of the collaborative tags for ontology maintenance (in Q3 from Table I), some participants wondered if collaborative tagging is useful at all since students (users of ontology-based learning systems) do not see the ontology most of the time. We concur that ontologies should not be visible to the end-users, as most of the software artifacts are not anyhow. Yet, collaborative tags reflect, at least to a certain extent, the community's shared conceptualization of a given domain. As such they have a rather similar purpose to ontologies in terms of knowledge sharing. Indeed, our motivation for the use of collaborative tags was consistent with the opinion of other participants "*… because students can be considered as people who are, at least partially, familiar with the area the topic is coming from, and because of the quantity of tags which help to make better tag cloud.*" Another participant stated that "*most of the times instructors/content authors are not sure what concepts they should include within their domain ontology. These tags come from a real context of usage and interaction and can perfectly reflect the concepts of the domain ontology.*" While some participants indicated a need for more automation of the process and a possible automatic inclusion of tags into the ontology, we intentionally did not push this functionality, as our previous study [24] indicated a strong preference of educators to be in the con-

trol of the ontology engineering process. Thus, our ontology-folksonomy visualization and interaction only indicates (color saturation) the relevant tags. Based on that educators can make a decision on which tags are to be integrated into the ontology under maintenance.

Some participants also pointed out some possible threats of the use of collaborative tags: "*Unless the students are familiar with the domain than the use of the collaborative tags for the ontology extension is not a reliable solution*" and that "*the collaborative tags may not [be] correct and relevant to the ontology at all.*" That is, students might not always tag things in terms relevant for the ontology or we further say that they might not tag relevant content for the ontology at all. The purpose of our MSR, WMSR and nWMSR (from Sect. 2.2-3) is exactly to compute semantic relatedness of tags with a selected concept in the ontology visualization. The values of those metrics are in the range 0-1. The color of strongly related tags (closer to 1) will be darker and of weakly related ones (closer to 0) lighter. The saturation of the color can go to the point to become invisible in the case when there is no semantic relatedness. Based on these relatedness measures and their reflections through tag colors, educators (i.e., ontology maintainer) can make informed decisions.

Finally, the participants from industry, although having positive comments about the tool, were reserved about the applicability of the approach for their target population of learners – workplace training where learners typically want to go through the content in a minimal time and are not interested in additional interaction (even tagging). Thus, collaborative tags would be hard to produce in that context. The observation is valid for some domains, but the adoption of social technologies in the corporate sector call for similar studies in that context [8].

5. RELATEDNESS MEASURES EVALUATION

In the interaction and visualization interfaces for ontology maintenance introduced in Sec. 2.2, we integrated a tag coloring method, as a way to recommend relevance of a tag for a given concept. Underneath those interfaces, the key component for recommending the relatedness between collaborative tags and ontology concepts are the measures introduced in Sec. 2.3-4. Values computed by those measures determine the level of color saturation of collaborative tags (i.e., darker colored tags are more relevant). In this section, we briefly summarize results of our evaluation of the proposed (WMSR and nWMSR) measures over the existing MSRs.

For our experiments, we used a sample ontology derived from the ACM CCS[4], represented in OWL as developed in [21]. The sample was related to the Software category (i.e., category D). We decided to use concepts *Programming Languages* (D.3) and *Software Engineering* (D.2) along with their sub-concepts. We ended up with an ontology consisting of 33 classes with a maximal depth of 5. Total of 58 tags were used for experiments from the annotation performed by three human experts and enriched by a set of tags generated by an automatic keyword extractor

for the content of the course "Introduction to Computer Science" deployed in the iHelp Courses LCMS at the University of Saskatchewan[5]. We involved 21 participant out of the 22 participants from Sect. 4 (one did not have a background in the area of the ontology domain) to create a gold standard for a selected group of tag-concept pairs.

Our results showed that the **best performing metric for all the gold standard baselines is nWMSR PMI-Gwikipedia**. Generally, PMI-based metrics have provided ratings that are more similar to human judgments and this finding is confirmed by other relevant experiments [46]. Our experiments also showed that the **nWMSR metrics outperform the WMSR metrics**. Our results that there is no need to go farther than two levels of depth, while computing an (n)WMSR metric.

Considering the obtained results, for the computation of recommendations of relevant tags (i.e., computation of color saturation of tags in the visualization and interaction interfaces introduced in Sect. 2.2), the most suitable metric is nWMSR based on PMI-Gwikipedia and with the depth level either 1 or 2. This measure for depth 1 is thus included in our final implementation of the ontology maintenance method in the LOCO-Analyst tool. We opted for one level of depth due to less computation steps needed than for depth 2 (as per equation (2) from Sect 2.4). In our future work, we plan to make further use of the findings of the experiments with the semantic relatedness measures. In particular, we plan to test if there is a significant difference in using the best and the worst performing metrics in the quality of maintained ontologies. We will also test if such measures create a significant difference in time to complete maintenance tasks related to understandability, and a possible impact on changeability of the proposed maintenance method.

6. THREATS TO VALIDITY

In this section, we discuss some potential threats to our experimentation results.

With respect to *internal validity* of the usability evaluation, we consider if some confounding factors would make a difference in the analyses [11]. In our experiment, the following confounding factors can be found: *different roles the participants played in education, experience, and motivation.* As reported in Sect. 4, we did not find any significant differences in the responses of the three groups based on their roles in education. We exclude the motivation as a confounding factor because the participation in the study was on a voluntary basis, and none of our participants left the experiments, while a great majority responded to the optional open-ended questions.

External validity of the results is the extent to which reported results can be generalized [11]. Here, we can first start from the population involved in the experiments, where there was a smaller number of experienced instructors. Still, as already indicated, our analyses (ANOVA) over the participants' responses grouped in different roles did not reveal any differences. A replicated experiment should further investigate the validity of this analysis.

---

[4] http://www.acm.org/about/class/1998

[5] http://pami.uwaterloo.ca/projects/lornet/software/ame.php

Another external factor of validity is the population, which was predominately composed of computer scientists (21 out of 22 participants). Our previous experience in evaluating the same usability factors in ontology learning [24] showed that computer scientists had a significantly higher level of expectations. Thus, computer scientist had significantly more negative observations than the participants with non-computer science background in evaluations of existing ontology tools for developing educational ontologies. A replicated experiment with non-computer scientists is needed to confirm this hypothesis.

Our results are applicable to a specific set of possible learning applications – university education. As our participants from the industry indicated, it is hard to believe that the same results would be applicable to some types of the corporate training. Still, this does not mean that there is no relevance of the lessons learned to contexts which are not universities [8].

For external validity of our findings related to the measures, it is most likely that other specialized domains might lead to rather similar conclusions, but that is a task for future replicated experiments in other domains. Also, the way of creating our gold standard might be a possible threat to the validity of the experiment. We used three different gold standard baselines and concluded that the minimal one is the optimal. However, we think that this issue needs to be carefully researched in the future, as some gold standard baselines might be better suited for some metrics and/or purposes. This is why we reported results for the three gold standard baselines.

7. RELATED WORK

There have been numerous proposals for leveraging ontologies in e-learning systems in general [19]. The evaluation of our Learning Object Context Ontologies (LOCO) framework for capturing learning contexts has proven that educators strongly appreciate the qualitative benefits which stem from the use of ontologies in providing educational feedback [29]. There have also been recent proposals to leverage folksonomies for ontology evolution in the context of learning technologies [41][42], mainly in the LT4eL project. However, their approach has some significant differences: first, it does not rely on the same measures as proposed here; and second, it does not take into consideration the interaction and visualization aspects even though these HCI aspects are of tremendous importance from a user perspective. This is why our approach also emphasizes the usability evaluation of the tool. Moreover, our approach is independent from any external semantic resource contrary to what is proposed in [41], where DBpedia is used. In our previous work, we also showed that collaborative tagging might be leveraged for different tasks related to learning content maintenance besides for ontology maintenance. Thus, that work, which is also integrated in the LOCO-Analyst tool, nicely complements the approach proposed in this paper and equips further educators with a comprehensive tool for course-related knowledge management.

Currently, there are two main kinds of approaches to linking folksonomies and ontologies. The first kind of approaches relies on altering the collaborative tagging process, so that it creates "semantic tags". Semantic tags are disambiguated by a user (i.e. tags are mapped to concepts in an upper-level ontology) [6] or tag relationships are defined by the community [36]. Neither method has proven to be overly successful. We attribute this to the fact that the additional effort required by typical taggers in creating the semantic tags, outweighs the perceived benefits. The second kind of approaches has even a more ambitious goal of automatically or semi-automatically linking collaborative tags with ontologies. While these kinds of approaches have had some promising results they have not yet revealed a general purpose and reliable solution [2]. With respect to all the aforementioned approaches, our contribution goes in the direction of leveraging folksonomies for ontologies evolution through a user-driven interaction based on interactive visualizations and system recommendations relying on context-based relatedness measures. In fact, measures of semantic relatedness (MSRs) are widely used for natural language processing tasks such as word-sense disambiguation [43] and analysis of the structure of texts. These measures rely on various knowledge sources including lexicons, thesauri, Wikipedia and the Web. One interesting aspect in using a resource such as MSR Server [53] is that it is easy to experiment with various metrics even for non-specialists, which is of great interest for the educational community.

In general, the interesting feature of MSRs is that they provide an automatic way of linking pairs of terms and are therefore perfectly applicable to pairs of concepts and tags. Methods for measuring semantic relatedness between concepts within and across ontologies are explored in [15]. Similar to our idea of evaluating a concept in its context, measures of semantic relatedness between ontological concepts are proposed by considering each concept as a set of its descendent leaf concepts. However, to our knowledge, this has not been done before in the learning technology community. In fact, our approach does not aim at defining new measures for computing semantic similarity between concepts from the same or different ontologies. Rather, we focus on suggesting relevant folksonomy's tags for concepts of a given ontology. We propose a simple way to take into account relationships for a particular concept to provide a "contextualized" usage of the already available measures.

Finally, MSRs are recognized as being of great interest to next generation Semantic Web applications [23] including ontology maintenance and matching. The problem of ontology maintenance has also been recognized in a related, but broader field of knowledge management and some solutions have been offered. For example, del.icio.us Brainlet [49] is a plugin for the DBin Semantic Web platform for personal knowledge and information management that enables a user to import tags from his/her del.icio.us account into a local RDF store, transform them into ontology classes and insert them in the class hierarchy. User interaction with a domain ontology and tags aimed at ontology enrichment is what makes DBin del.icio.us Brainlet similar to our work. However, the interaction in our approach is more promising for two reasons: 1) we calculate the relatedness between ontology

concepts and tags and use it to measure the relevancy of a tag for a given concept in the context of the given ontology; 2) tags are not presented in the form of a flat list (e.g., del.icio.us Brainlet), but in the form of tag cloud, so that the user can spot the popularity of each tag, its relevancy to the selected domain concept, and how it compares to other tags used to describe concept-related content.

Hepp et al [28] suggest Wikis' infrastructure and culture as an environment for constructing and maintaining consensual vocabularies for knowledge management and using the Wikipedia URIs as unique identifiers for concepts for annotating knowledge assets. This seems to be an appealing solution from the perspective of knowledge engineers as it would provide them with an easy-to-use working environment. However, this solution produces an "informal ontology", that is, a collection of named conceptual entities with a natural language definition, and such an ontology can not address specific requirements of e-learning environments.

8. CONCLUSIONS AND FUTURE WORK

With respect to the first component of our research contribution – visualization and interactive interface for ontology maintenance – our analysis revealed a very high perceived value by the educators involved in our experiments. It also showed that the participants identified the usability and interactivity of tools for ontology maintenance with the maintenance process itself. This indicates that visualization and interactive user interfaces are firstclass citizens when any tools for ontology maintenance are to be developed for learning technologies. While our usability experiment provides useful data about different usability aspects of the tooling proposed for ontology maintenance, all our findings are based on subjective variables; as such they are suitable for our research question – evaluate the perceived usability of the tool. In future research, we will work on an experimentation setting that will allow us to collect data about more objective variables. In particular, we envision setting up experiments in which educators will be asked to complete a set of tasks for ontology maintenance (e.g., for a given concept, select relevant tags and connect them with the concept). Such experiments will then provide us with objective means to measure usefulness and effectiveness of particular aspects of the ontology maintenance user interface (e.g., number of selected tags per concept, types of used relations, time for finding relevant tags for different concepts). In terms of the ISO 9126 standard, we focused on the understandability characteristic of maintainability. In our future work, we will investigate the analyzability and changeability characteristics as well. These experiments might also be useful to evaluate the effectiveness of the proposed recommendation process quantitatively (i.e., the saturation of the color of more related tags is darker). A more ambitious and longer-term goal would be to have a standard experimental setup for evaluating future solutions in this area of research, similar to those adopted in software maintenance [9].

Although our participants appreciated the use of on-tology visualization and interaction interfaces, it was well-observed that this type of solution might not scale for large ontologies. Depending on the size of a learning domain covered by a specific ontology (e.g., course ontology might not be that big, but an entire study program might be rather large), development of effective visualizations/interaction interfaces might be more or less challenging. One promising research direction for future work is to investigate the combined use of different ontology coupling [20] with tag popularity and relatedness metrics for visualizing and interacting with large scale ontologies and tag clouds. Equally important to this will be to investigate user interfaces that will guide educators more effectively in ontology maintenance and development. Similar to the findings of the experimentation presented in this paper, our other related experiment [24] confirmed that users require better guidance than currently offered while completing different ontology editing tasks. That is, besides our tag coloring strategy, we will need to investigate some more "intrusive" guidance in the user interfaces.

Our experiments with the different relatedness metrics showed that the best performing metric is nWMSR PMI-Gwikipedia. This best performance of a Wikipedia-based metric was already hypothesized in our preliminary work [47]. Even though our experiments were related to the domain ontologies for computer science education, we can hypothesize similar results for other domains. Of course, that hypothesis is to be confirmed in future studies. In addition to the consideration of sub-class relations, our future work should also consider other types of relations such as synonymy, polysemy, or custom relations.

For the already mentioned challenge of a better guidance for educators in ontology maintenance tasks, our future research needs to further investigate recommendations of relations which might be established between collaborative tags and ontology concepts. Our current implementation for ontology maintenance offers a fixed set of possible relations to choose from. However, even from that set of offered relations there are no recommendations on the most suitable ones, or the ability to discover some potentially new ones. For this purpose, we see as a promising direction Hearst patterns [25], following approaches like the one introduced in [44] or external sources of collectively accumulated knowledge such as DBpedia [7].

REFERENCES

[1]   Abran, A., Moore, J.W., Bourque, P., and Dupuis, R. (Eds.), "Guide to the Software Engineering Body of Knowledge (SWE-BOK)," IEEE Computer Society, Washington DC, (2004).

[2]   Al-Khalifa, H.S. & Davis, H.C. "Exploring The Value of Folk-

sonomies For Creating Semantic Metadata," *Int'l J. on Sem. Web and Inf. Sys.* 3(1), (2007), 13-39.

[3] Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*, Longman, UK, (2001).

[4] Anderson, T.D., Garrison, D.R., "Learning in a networked world: New roles and responsibilities," *In C.C. Gibson (Ed). Distance learning in higher education: Institutional responses for quality outcomes*. Madison, WI: Atwood Publishing, (1998) 65-76.

[5] Aroyo, L, Dolog, P., Houben, G.J., Kravcik, M., Naeve, A., Nilsson, M., Wild, F., "Interoperability in Personalized Adaptive Learning," *Edu. Tech. & Soc.*, 9(2), (2006), 4-18.

[6] Bateman, S., Brooks, C., and McCalla, G., "Collaborative Tagging Approaches for Ontological Metadata in Adaptive E-Learning Systems," *In Proc. of the Int'l WSh on Applications of Semantic Web technologies for E-Learning*, (2006), 3-12.

[7] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S., "DBpedia – A Crystallization Point for the Web of Data," *J. of Web Semantics*, 7(3), (2009), 154–165.

[8] Braun, S., Kunzmann, C., Schmidt, A., "People Tagging & Ontology Maturing: Towards Collaborative Competence Management," *In D. Randall and P Salembier (eds.) From CSCW to Web2.0: European Developments in Collaborative Design Selected Papers from COOP08*, Springer, (2010), 133-154.

[9] Briand, L., Bunse, C., Daly, J., "A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs," *IEEE Tran. Soft. Eng.*, 27(6), (2001), 513-530.

[10] Buckley, J., Mens, T., Zenger, M., Rashid, A., Kniesel, G., "Towards a taxonomy of software change," *Journal of Software Maintenance*, 17(5), (2005), 309-332.

[11] Chin, D. N. "Empirical Evaluation of User Models and User-Adapted Systems," *User Modeling and User-Adapted Interaction* 11 (1-2), (2001), 181-194.

[12] Cilibrasi, R. & Vitanyi, P.M.B. "Similarity of objects and the meaning of words," *In Proc. of the 3rd Conf. Theory and Applications of Models of Computation*, (2006), 21-45.

[13] Cilibrasi, R., & Vitanyi, P. M. B., "The Google similarity distance," *IEEE Trans. Know. & Data Eng.*, 19(3), (2007), 370-383.

[14] Cohen, J., "A coefficient of agreement of nominal data," *Educational and Psychological Measurements*, 20(1), (1960), 37-46

[15] Cross, V., "Semantic Relatedness Measures in Ontologies Using Information Content and Fuzzy Set Theory," *In Proc. of the 14th IEEE Int'l Conf. on Fuzzy Systems*, (2005), 114–119.

[16] d'Aquin, M., Noy, N.F., "Where to Publish and Find Ontologies? A Survey of Ontology Libraries," *TR BMIR-2010-1413*, SCBIR Stanford University, 2010, http://bit.ly/hGDUbr

[17] Denaux, R., Dimitrova, V., Cohn, A.G., Dolbear, C., Hart, G., "ROO: Involving Domain Experts in Authoring OWL Ontologies," *In Proc. 9th Int'l Semantic Web Conf.*, (2008).

[18] Dimitrova, V., Denaux, R., Hart, G., Dolbear, C., Holt, I., Cohn, A.G., "Involving Domain Experts in Authoring OWL Ontologies," *In Proc. 9th Int'l Semantic Web Conference*, (2008), 1-16.

[19] Dzbor, M., Stutt, A., Motta, E., Collins, T., "Representations for semantic learning webs: Semantic Web technology in learning support, *J. of Computer Assisted Learning*, 23(1), (2007), 69–82.

[20] Garcia, J., Garcia, F., & Therón, R., "Visual Modeling of the Hierarchy, Metrics and Coupling among classes in an OWL ontology," *In Proc. CAiSE 2010 Workshop on Ontology, Conceptualization and Epistemology for Information Systems, Software Engineering and Service Sciences*, (2010).

[21] Gašević, D., Hatala, M., "Ontology mappings to improve learning resource search," *British Journal of Educational Technology*, 37(3), (2006), 375-389.

[22] Gašević, D., Jovanović, J., Devedžić, V., "Ontology-based Annotation of Learning Object Content," *Interactive Learning Environments*, 15(1), (2007), 1-26.

[23] Gracia, J. & Mena, E., "Web-Based Measure of Semantic Relatedness" *In Proc. of the 9th international Conference on Web information Systems Engineering*, (2008), 136–150.

[24] Hatala, M., Gašević, D., Siadaty, M., Jovanović, J., Torniai, C. "Can Educators Develop Ontologies Using Ontology Extraction Tools: an End User Study," *In Proc. of the 4th European Conference on Technology-enhanced Learning*, (2009) 140-153.

[25] Hearst, M.A. "Automatic acquisition of hyponyms from large text corpora," *In Proc. 14th Int'l Conf. on Computational Linguistics*, (1992), 539-545.

[26] Held, C., Cress, U., "Learning by Foraging: The Impact of Social Tags on Knowledge Acquisition," *In Proc. of the 4th European Conference on Technology Enhanced Learning*, (2009), 254-266.

[27] Hepp, M., De Leenheer, P., de Moor, A., and Sure, Y. (Eds.) *Ontology Management for the Semantic Web, Semantic Web Services, and Business Applications*, Springer, Berlin, Germany, 2007.

[28] Hepp, M., Siorpaes, K., & Bachlechner, D., "Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management," *IEEE Int. Comp.*, 11(5), (2007), 54-65.

[29] Jovanović, J., Gašević, D., Brooks, C., Devedžić, V., Hatala, M., Eap, T., Richards, G., "Using Semantic Web Technologies for the Analysis of Learning Content," *IEEE Internet Computing*, 11(5), (2007), 45-53.

[30] Jovanović, J., Gašević, D., Devedžić, V., "Ontology-based Automatic Annotation of Learning Content," *Int'l Journal on Semantic Web and Information Systems*, 2(2), (2006), 91-119.

[31] Jovanović, J., Gašević, D., Torniai, C., Bateman, S., Hatala, M., "The Social Semantic Web in Intelligent Learning Environments-State of the Art and Future Challenges-," *Interactive Learning Environments*, 17(4), (2009), 273–308.

[32] Kamtsiou, V., Naeve, A., Stergioulas, L.K. & Koskinen, T., "Roadmapping as a Knowledge Creation Process: The PROLEARN Roadmap," *J. of Universal Knowledge Management*, 1(3), (2006), 163-173.

[33] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E. "Ontology visualization methods—a survey," *ACM Comput. Surv.* 39(4), (2007), 10.

[34] Klamma, R., Chatti, M.A., Duval, E.; Hummel, H., Hvannberg, E.T., Kravcik, M., et al. "Social Software for Life-Long Learning," *Edu. Tech. & Soc.*, 10(3), (2007) 72-83.

[35] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology (2nd edition)*. Thousand Oaks, CA: Sage Pub., 2004.

[36] Lachica, R., "Towards holistic knowledge creations and interchange Part 1: Socio-semantic collaborative tagging." *Talk at the TMRA Int'l Conf.*, (2007)

[37] Landis, J.R., Koch, G.G., "The measurement of observer agreement for categorical data," *Biometrics*, 33(1), (1977), 159-174

[38] Mens, T., Demeyer, S., *Software Evolution*, Springer, Berlin-Heidelberg, Germany, 2008.

[39] Mika, P., "Ontologies are us: A unified model of social networks and semantics," *J. of Web Semantics*, 5(1), (2007), 5-15.

[40] Mikroyannidis, A., "Toward a Social Semantic Web," *Computer*, 40(11), (2007), 113-115.

[41] Monachesi, P, Markus, T., "Using Social Media for Ontology Enrichment," *In Proc. of the 7th Extended Semantic Web Conference*, (2010), 166-180.

[42] Monachesi, P., Markus, T., Mossel, E., "Ontology Enrichment with Social Tags for eLearning," *In Proc. of the 4th European Conf. on Technology Enhanced Learning*, (2009), 385-390.

[43] Patwardhan, S., Banerjee, S., Pedersen T., "Using Measures of Semantic Relatedness for Word Sense Disambiguation," *In Proc. 4th Int'l Conf. Computational Linguistics and Intelligent Text Processing*, (2003), 241-257.

[44] Plangprasopchok, A., Lerman, K., "Constructing folksonomies from user-specified relations on flickr," *In Proceeding of the 18th Int'l WWW Conference*, (2009), 781-790.

[45] Rada, R., Mili, H., Bickell, E., & Blettner, B., "Development and application of a metric on semantic nets," *IEEE Transactions on Systems, Man and Cybernetics*, 19(1), (1989), 17–30.

[46] Recchia, G.L., & Jones, M.N., "More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis," *Behavior Research Methods*, 41(3), (2009), pp, 657-663.

[47] Torniai, C., Jovanović, J., Bateman, S., Gašević, D., Hatala, M., "Leveraging Folksonomies for Ontology Evolution in E-learning Environments," *In Proc. 2nd IEEE Int'l Conf. Semantic Computing*, (2008), 206-215.

[48] Torniai, C., Jovanović, J., Gašević, D., Batemen, S., Hatala, M., "E-learning meets the Social Semantic Web," *In Proc. of the 8th IEEE Int'l Conf. on Advanced Learning Tech.*, (2008), 389-393.

[49] Tummarello, G. & Morbidoni, C., "Collaboratively building structured knowledge with DBin: from del.icio.us tags to an 'RDFS Folksonomy'," *In Proc. of the WWW 2007 WSh on Social and Collaborative Construction of Structured Knowledge*, (2007).

[50] Turney, P.D., "Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. *In Proc. 12th European Conference on Machine Learning*, (2001), 491-502.

[51] van Hage, W., Katrenko, S., & Schreiber, G., "A Method to Combine Linguistic Ontology-Mapping Techniques," *In Proc. of the 4th Int'l Sem. Web Conf.*, (2005), 732-744.

[52] Vassileva, J. (2009) Towards Social Learning Environments, *IEEE Transactions on Learning Technologies*, 1(4), 199-214.

[53] Veksler, V. D., Grintsvayg, A., Lindsey, R., & Gray, W. D., "A proxy for all your semantic needs," 29th Annual Meeting of the Cognitive Science Society, (2007).

[54] Vygotsky, L., *Mind in Society*. London: Harvard U. Press, (1978).

[55] Winn, W.D., "Encoding and Retrieval of Information in Maps and Diagrams," *IEEE Trans. Prof. Comm.*, 33(3), (1990), 103–107.

[56] Zouaq, A., & Nkambou, R., "Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project," *IEEE Trans. Knowledge and Data Engineering*, 21(11), (2009), 1559-1572.

Dragan Gašević is a Canada Research Chair in Semantic Technology and an Associate Professor at Athabasca University. He is also an Adjunct Professor at Simon Fraser University. He received his Dipl.Ing., MSc, and PhD degrees in computer science from the University of Belgrade. His research interests are in semantic technologies, software language engineering, technology-enhanced learning, learning analytics, and services computing. Being a (co-author) of numerous publications, often keynote speaker and event organizers in the areas of his research interest, he has been a co-founder of the International Conferences on Software Language Engineering (SLE) and Learning Analytics & Knowledge (LAK). More details about Dragan can be found at http:/dgasevic.athabascau.ca

Amal Zouaq (http://azouaq.athabascau.ca) is an Assistant Professor at the Royal Military College of Canada. Her research interests include natural language processing, Semantic Web, ontology engineering, knowledge extraction and technology-enhanced learning. She serves as a member of the program committee of various conferences in technology-enhanced learning and Semantic Web and she is part of the editorial review board of the Interdisciplinary Journal of E-learning and Learning Objects and a member of the editorial board of the Journal of Emerging Technologies in Web Intelligence. She also serves as a reviewer in many conferences and journals in knowledge and data engineering, natural language processing, eLearning and the Semantic Web.

Jelena Jovanovic is an Assistant Professor of Computer Science with the Department of Software Engineering, FOS - School of Business Administration, University of Belgrade, Belgrade, Serbia. She received her B.S., M.Sc. and PhD degrees in informatics and software engineering from University of Belgrade in 2003, 2005 and 2007, respectively. Her research interests are in the areas of semantic technologies, Web technologies, technology-enhanced learning, and knowledge management. She is a member of the GOOD OLD AI research network. She can be reached at http://jelenajovanovic.net.

Carlo Torniai is currently an Assistant Professor at Oregon Health & Science University in Portland, Oregon. Throughout his work experience he has been developing and using Semantic Web technologies in several domains. In particular his research interests include best practices of ontology development for multimedia and biomedical resource annotations and exploring interactions between Social and Semantic Web in learning environments.

Marek Hatala is an Associate Professor and a Graduate Chair at the School Interactive Arts and Technology at Simon Fraser University. He received his PhD degree in cybernetics and artificial intelligence from the Technical University of Kosice. His research interests are in the areas of knowledge management, artificial intelligence, distributed systems, user modeling, interoperability, security and trust policies, e-learning and collaborative systems. More details about Marek can be found at http://www.sfu.ca/~mhatala