Fuzziness and Semantic Web Technologies in Personalized eLearning

Marek Reformat¹ and Patricia Boechler²

¹Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6G 2V4
²Educational Psychology, University of Alberta, Edmonton, Canada T6G 2V4

Abstract

The rapid development of information technologies has created suitable environment for construction of eLearning systems. The pivotal point of those systems lies in web technologies. Due to those technologies eLearning leads to a number of improvements in: delivery, access, distributed authority, and personalization, when compared with traditional training. Additionally, integration of fuzziness with processes of customization and selection of adequate material for the user creates a chance to built truly personalized and adaptive systems.

The paper presents an overview of an architecture of eLearning and its basic component – eLerning Engine – taking a full advantage of ontology, tagging, and users’ feedback represented with linguistic descriptors and quantifiers.

Keywords: eLearning, CI, Semantic Web

1. Introduction

The concept of Semantic Web proposed by Berners-Lee [1] introduces an idea of machine-processable information – ontology [7]. This form of representation enables better and more semantic-oriented processing of information, as well as reasoning about it. Its application to eLearning creates opportunities to built systems capable of analyzing learners’ needs and behaviors, and more accurate selection of learning materials. However, in spite of those abilities, there is a need to deal with missing or inaccurate data [30]. Learners themselves bring uncertainty – they are imprecise in expressing their needs and opinions. Their decisions regarding selection of most suitable alternatives heavily depend on current circumstances, their understanding of situations, and their needs and requirements – things that are “equipped” with ambiguity.

In many situations, identification of the most suitable alternative is influenced by opinions provided by others. In such circumstances, a selection process is a multi-criteria decision-making process where each alternative is evaluated based on multiple measures originated from multiple sources. Learners are interested in other’s experiences – social software becomes an important vehicle for communication and exchange of information among individuals. Its integration with eLearning seems natural and highly beneficial.

The paper is a description of a framework, and its development phases, for constructing human-centric eLearning systems – systems with capabilities to recognize learners’ learning styles and adapt to learners’ needs and preferences. Such systems combine 1) technologies of the Semantic Web – ontology and forms of its representation, 2) aspects of social software – blogs and tagging, and 3) techniques of Computational Intelligence (CI) – fuzziness and multi-criteria decision-making. Additionally, such systems provide instructors with abilities to enter their suggestions and recommendations, and observe learners’ learning activities and comment on them. It is apparent that a proper representation of data and adequate processing of information are necessary steps leading towards knowledge-oriented systems.

The ability to find and represent different types of relations between pieces of information is a necessary condition for creating semantically conscious applications. This semantic awareness allows for more accurate identification of relevant information. At the same time building any type of system that interacts with a human requires ability to handle imprecision and ambiguity. From this point of view the application of fuzziness and approximate reasoning creates a promising avenue of introducing human aspects to software systems, and development of more human-aware and human-like systems.

2. Related work

Development of eLearning systems and supporting them technologies represents an ongoing challenge of fundamental interest and practical relevance. Existing approaches in this area are quite diversified enjoying the reliance on various methodologies and effective algorithmic developments. A substantial number of them adhere to the fundamentals of general schemes of web technologies.

The domain of eLearning is progressing rapidly. The Semantic Web proposed knowledge representations – RDF (Resource Description Framework) [32] and ontology – quickly finds their way in eLearning applications. Such aspects as formal taxonomies expressed with web ontology languages RDFS and OWL [33], and rules represented using the web rule
language RuleML [34] play a key role in enabling the representation and the dynamic construction of shared and re-usable learning content [25], [23], [12]. Some of the highly visible trends are worth briefly recalling here. An idea of ontology-enabled annotation becomes an important element of eLearning systems. A number of ontologies representing different aspects of knowledge are used to support personalized annotation, real-time discussion, and semantic content retrieval [28]. Specialized annotation tools are developed [4]. Successful e-Learning systems have to address how individuals may want to learn. So, e-Learning designs need to support personalization [15]. Personalized hypertext relations powered by reasoning mechanisms over distributed RDF annotations have been proposed in [9]. Several ontologies have been used which correspond to the components of an adaptive system: a domain, a user ontology, and an observation ontology. In [5] a student model has been developed and integrated with an ontology, enabling the personalization system to guide the student’s learning process. Ontology-based solutions for dynamic assembly and personalization of learning content parts have been also proposed in [11]. An integrated learning environment called TANGRAM has been developed. It relies on two ontologies for representing learning object content structure and content type (i.e. pedagogical role). A user model ontology is defined to represent relevant information about TANGRAM’s users. Applications of Web Intelligence and Artificial Intelligence in eLearning systems enhance adaptivity and enhanced learner comfort. They enable course sequencing and material presentation, automatic discovery, invocation, and composition of educational Web services [2].

Social networks make a great basis for developing new generation of education portals. They provide a structure for exchange of information, and communication between authors, teachers, and educational institutions [18] [31]. Such networks allow for combining educational portals, ontologies, and search agents with functions such as Web mining, and knowledge management to create, discover, analyze, and manage the knowledge of different domains presented in educational material.

Multiple aspects of eLearning has been addressed by techniques involving fuzziness. Fuzzy-based methods are used for user profiling, determining learners profiles, evaluating quality of eLearning systems, as well as enhancing their capabilities. In [16], the authors use fuzzy terms to describe pedagogical resources as well as users’ profiles. They propose application of fuzzy operators to identify links between available resources and users’ preferences. It leads to flexibility and customization of learning processes. In [10], the author applies fuzzy clustering to determine categories of learners’ profiles. He performs mining of Web log files to categorize learners’ behavioral patterns. Another interesting application of fuzziness can be found in processes of evaluating quality of eLearning systems [29]. The method relies on analysis of questionnaires completed by learners. Fuzzy relations and weighted factor are used to determine learners’ perceptions of quality of learning experience.

3. Concept

The ultimate objective of this work is to develop a comprehensive framework for development of eLearning Engines (eLEngine) required for building human-centric eLearning Systems. Human-centric aspects will be accomplished via utilization of fuzziness and approximate reasoning that are able to express and process ambiguity and imprecision – two very characteristic features of human selection and decision-making activities. These techniques combined with RDF and ontology-based representation of knowledge and elements of social networks – blogging and tagging – will lead to a new way of designing eLearning Systems. Development of such a system will allow us to draw conclusions of immediate practical relevance to creation of these systems.

In contrast to the research results reported commonly in the literature, we are concerned with such aspects of human nature as imprecision, incomplete information, and approximate reasoning in order to ensure engaging and comfortable, yet practical and efficient learning environment. The proposed framework evolves around two observations: 1) omnipresent ambiguity and imprecision in many forms of information provided and used by humans; 2) multiple sources of information that influence the content and the form of course materials. Figure 1:

- personal preferences and feedback – in the form of learner’s goals and his/her learning style, as well as his/her involvement in content annotation, tagging, and contributions to blogs;
- required material, suggestions, and constraints provided by instructors;
- feedback, content annotations, notes, and evaluations contributed by other learners.

These sources of information – quite often equipped with ambiguity – determine relevant components of domain knowledge and influence selection of the teaching material that is most suitable to the learner. This process can be performed in a variety of ways with different levels of precision and importance assigned to each of these sources. For example, for courses that are fundamental and need a rigorous approach, the suggestions and constraints imposed on the content by instructors will have higher priority, influence, and precision than preferences provided by the learner, as well as the feedback provided by other learners. Such an approach offers flexibility in the process of constructing course materials.

Another important aspect is the ability to create an environment in which the learner feels that
he/she participates at the same time in a one-to-one session with an instructor and has access to the feedback from his/her peers at the same time. Such ability brings a different and very important dimension to the eLearning experience.

The principal challenge is to combine fuzziness (capable of addressing impressions and approximate reasoning) with the SW technologies (for knowledge representation) and with mechanisms of social networking. All those elements are needed to achieve the ultimate goal – to build an environment for human-centric learning experience where each learner feels comfortable and is keen to learn. The proposed framework is multifaceted and is based on the most essential features, such as:

- new ways for multi-domain semantic annotation of course material and evaluating the learning experience that fully exploits social software techniques;
- novel methods of extracting and representing knowledge based on the integration of CI technologies with new information representation forms, i.e., ontology and RDF triples;
- application of fuzziness and approximate reasoning to represent, analyze and process learners’ preferences and course evaluations; to identify relevant – to learners’ needs – subjects, courses and materials; and to mimic learners’ selection and decision-making processes.

The framework will support and quantify the methodological investigations and ensure correctness of developed methods and approaches.

4. Technologies and methodology

4.1. Fuzziness and Semantic Web

Fuzziness provides a unique approach to deal with a very human concept of imprecision. Abilities to use such imprecise terms as much, so-so and linguistic quantifiers like most, at least, any make fuzzy-based methods most suitable for dealing with human evaluation of different items and their description. Another important aspect of fuzziness is its ability to express levels of membership of terms to specific concepts. Using fuzzy-based mechanisms for processing and reasoning, we can deduce new facts and their levels of belonging to specific categories, as well as precision levels of their descriptions.

The technologies of the SW provide a new and highly comprehensive approach to knowledge representation. The basic framework used to represent concepts, their definitions and instances, as well as relationships existing between concepts is ontology. The ontology defined in the context of the SW uses the concept of Resource Description Framework (RDF) triple as the foundation of knowledge representation. An RDF triple is a triple subject-predicate-object, where: subject identifies what object the triple is describing; predicate (attribute) defines the piece of data in the object a value is given to; and object is the actual value of the attribute; for example, the triple “John likes books” has John as subject, likes as predicate and books as object. A collection of RDF triples can be treated as multi-graph, and as such, an RDF-based data model is a very attractive knowledge representation form.

The increasing involvement of users in providing and evaluating contents of the web is expressed by popularity of systems that rely on principles of social software. One of the most important aspects that leads to such popularity is an easy way of providing and augmenting web content and expressing individuals’ opinions about it. The two methods that are directly involved in those activities are blogs and tagging. A blog allows its users to post commentary, descriptions of events, or any other material related to a specific topic or topics. Those descriptions could be of any form – text, graphics, or even video. The process of tagging is nothing else but labeling – annotating – resources [17] [8]. It is performed by users who use tags to annotate resources easily and freely without knowing any taxonomies or ontologies. Tags represent any strings that users consider appropriate as descriptions of resources. Resources, on the other hand, could be any items that have been posted and are accessible by users. This can lead to an interesting way of describing resources [19] [27]. Those technologies are applied to design and develop elements and features of eLearning Engines.

4.2. Categorization of learners

Identification of categories of learners is another important aspect of eLearning. Each learner is different, and each learns in a different way. Educators recognize this, and express these differences as learning styles, cognitive styles, multiple intelligences or cognitive traits. With adaptive learning systems, two approaches have been used: 1) assessing cognitive/learning styles at the outset and presenting the system to match the learner’s profile, or
2) having no preset initialization of the system and allowing adaptations to occur based on the learners’ use of the system only. In order to understand the advantages and disadvantages of both approaches, both approaches need to be tested.

In regards to the former approach, the exact method for characterizing learner profiles is highly debated. Some learner assessment approaches rely on cognitive style measures arising from psychological theory [24], that is, measures of general cognitive tendencies or approaches that endure across numerous types of stimuli. Others focus on learning styles, categorizations of learners preferences in educational contexts and finally, some methods employ the measurement of basic cognitive traits (e.g., working memory capacity) as a means to predict what material and style of presentation is desirable for a particular learner [6]. From a theoretical standpoint, it appears none of these approaches have been unchallenged in regards to their validity and reliability [24].

Given the inconclusive nature of research in these areas, we adopt a method that creates coherence between the initially derived learner profiles and the mechanisms for updating learner profiles that exist within this particular adaptive system. We choose a cognitive style measure that reflects the specific mechanisms in our adaptive system and that we believe we can modify it to limit the demands of assessment on the learner in the initial phase of using the system. For the assessment of the structure of content, we rely on the Wholistic/analytic dimension on the Cognitive Styles Analysis (CSA) [22]. This provides stimulus to adaptive mechanisms of the system. For the assessment of media preference, we rely on the Verbal/imagery dimension on the CSA. This will be linked to the course material representation mechanisms and the personalization mechanisms of the system. For social preferences, we will again turn to the Verbal/imagery dimension of the CSA as research supports its identification of preference for learning contexts that include social interaction. This measure will become an input to the individual and collaboration-based selection mechanisms. The CI technologies and knowledge representation as proposed by SW, and SS creates a unique opportunity to design and develop eLearning systems that are able to “analyze” those measures, and adapt to needs and preferences of each learner.

4.3. Representation of course material

Applications of digital technologies in education means that learners experience new ways of interaction with their learning environment. The aspect of effectiveness of digital media in a teaching process is of major importance.

The task aims at the exploration of how digital representation of material selectively extends but also constrains what a learner sees, experiences and has access to, and how it enhances but also shapes instructors’ representations and presentations of their knowledge in an eLearning system.

Such activities as taking notes, marking important and/or difficult parts of presented materials, and writing feedback comments (i.e., confirmation, corrective, explanatory, diagnostic, and elaborative information) are vital components of a learning process at any level of education. The proposed framework will be equipped with a number of techniques and methods, such as content annotation, blogs, and tagging, to allow users for labelling the teaching material and providing their opinions about its content. A particular emphasis will be put on non-intrusive ways of inputting information – for example, via voice.

Information provided by users is stored using the knowledge representation schema based on ontology and RDF triples. Algorithms are used to process users’ inputs and to annotate course materials with terms and keywords reflecting users’ opinions and notes.

4.4. Personalization and context dependence

One of the essential challenges of eLearning systems is to satisfy learner’s needs and preferences. It is of critical importance to be able to properly elicit and store information about learners, their likes and dislikes, and what kind of methods, techniques and media they enjoy during learning activities. The techniques should ensure utilization of two types of information:

- learner’s needs, what he/she already knows, his/her goals, things already done, things left to do, timing, ability to learn, ways of learning, most suitable media (slides, notes, short lectures);
- current context – time of a day, an amount of time a learner can spend, place, ability to listen/watch.

Special mechanisms and techniques are used to keep track of things that work for the learner and that he/she likes, his/her comments, and information about his/her favorite instructors.

All information about an individual learner is stored in a specialized ontology. Such ontology is created and maintained for each learner. The mechanisms that support storing imprecise (fuzzy-based) information are developed. The learner is able to provide her priorities regarding needs and preferences in a human-like form. Special mechanisms for estimating relevance of that information are proposed and validated. We address temporal aspects of information. Such concepts as “fading out” and “old” are defined and their importance and usefulness are investigated.
4.5. Instructors’ input mechanism

The involvement of instructors in the education process is irreplaceable. Therefore, the proposed framework provides a number of ways that an instructor can monitor students’ activities and act accordingly. Instructors are able to query the system for items related to available materials and learners taking their courses, as well as read comments provided by learners and related to the material they prepared. The system also allows instructors to enter answers to learners’ questions, suggest alternative material to learners, and correct them.

4.6. Individual and collaboration-based material selection

The process of selecting most suitable lecture materials – choice between multiple versions, multiple sections, multiple media, is the most critical part of any eLEngine-based system. Multiple sources of information about lecture materials have to be evaluated and ranked based on:

- multiple criteria provided by the learner – learner’s goals and profile (preferences, likes/dislikes), as well as comments and notes given to similar course materials;
- instructors’ suggestions, recommendations, and constrains – instructors’ notes, observations and expertise will be used to important selection criteria;
- multiple evaluations of possible material – the process of collecting learners’ feedback via social software methods will lead to an extensive annotation of material, a schema for integrating annotations and extracting most common opinions is required; those opinions play the role of criteria during a selection process.

The proposed selection methods mimic human-like aggregation processes from learners’, other users’, and instructors’ points of view. The levels of importance are taken into consideration. The methods have to deal with imprecision information (evaluations, criteria, annotation), different priorities, and constraints. We examine the following aggregation approaches: fuzzy-based methods; evidence theory; different aggregation techniques including linguistic based aggregation. Overall, the proposed selection mechanisms perform decision-making tasks taking into account:

- what the instructor thinks is important;
- what is important for the learner;
- what peers think is important.

5. Description of eLearning system

The architecture of the eLEngine-based system is presented in Figure 2. In a nutshell, the system knows what the learner wants (learner’s profile) and likes (annotations, blogs, tags), knows what instructors suggest and recommend (modifications, suggestions, annotations, blogs, tags), and knows what other learners say about available material (annotations, blogs, tags). Based on that knowledge and syllabuses, the system provides the learner with most suitable alternatives regarding sets of education material. The system allows learners to make notes and record opinions. At the same time instructors have the ability to monitor the learner and provide modifications and additions to the material the learner is currently using. In order to accomplish that, all the tasks performed by the system are divided into three categories:

- multi-domain annotation of course material stored in a repository combined with techniques and methods of extracting important options based on tag clouds, blogs and learners’ notes; instructors’ suggestions and constrains are also used to annotate available material;
- personalization that leads to creation and updating of learner’s profile that contains information about learner’s preferences, needs, likes/dislikes;
- prioritization-based multi-criteria selection that performs selection of most suitable material based on learner’s profile, other learners’ opinions, and instructors’ inputs.

Figure 2: Architecture of eEngine-based System.

The annotation activities are presented in Figure 3. All annotation is performed on course material stored in the repository of the system, and it reflects three domains (dimensions): learners’, instructors’, and knowledge relevant to course topics. The learners-wise annotation is supported by such processes as identification of concept and keywords in annotations and blogs, as well as analysis of tag-clouds. The instructor-wise annotation takes into account all requirements and recommendations provided by instructor. The domain-wise annotation leads to annotation of all repository materials with terms originated from specific knowledge domains. In the end, all materials is annotated with three types of terms originated from three sources of annotations.
The personalization activity is also shown in Figure 3. It uses results of the same process as annotation: concept and keyword identification and analysis of tag-clouds to extract information that is related to a single learner. This information is used for updating of the learner’s profile. Both annotation and personalization process are continuously performed so annotations and profile is up-to-date.

The multi-criteria selection mechanisms are presented in Figure 4. The first step in the process selects a few sets of alternative materials. The selection uses the annotated course material and is based on learner’s needs and goals, and provided syllabuses. An important element is extraction of evaluations of those materials done by other learners. The last and most important step of selection is identification of the most suitable material. This step relies on a number of different decision-making methods that are able to deal with multiple criteria with multiple levels of priorities [26], imprecise information and human-like aggregation of evaluations [21] [20]. These processes can use different sources of information, for example RSS-feeds, and different analysis techniques including formal ones, for example FCA [3].

6. System development

6.1. Overview

The framework objectives require a structured work plan comprised of a number of parallel and complementary activity lines. Progress within those lines will be interactive, incremental, and, when necessary, iterative. The conceptual activity line dominates as the eLearning Engine methodology evolves. It requires a very thorough and extensive exploration of the developed concepts that leads to transition into algorithmic development and validation – the algorithmic activity. This is followed by integration and experimentation using existing software tools – the experimentation activity. For instance, validation of software implementing a given algorithm provides insights into conceptual foundations, leading to revision of the algorithm and its new implementation. Similarly, improvements in implementation impact an integration process.

Conceptual activity line: within this activity the focus is put on the fundamentals and conceptual underpinnings, in particular the following aspects are investigated: annotation mechanisms, personalization techniques and a variety of selection schemes, as well as diversity of possibilities in eliciting opinions of users – learners and instructors. The mechanisms of human-centricity, which becomes present when forming a feedback relevance loop, could be conceptually diversified leading to a number of general pursuits.

Algorithmic activity line: Being cognizant of the numerous possibilities explored at the conceptual level, the algorithmic activity is concerned with the underlying aspects of their realization and establishing their feasibility from the computational standpoint. Likewise, the focus is put on investigation of the realization of the specific conceptual developments (e.g., through a critical assessment of required learning capabilities which are necessary to arrive at the effective algorithm). The activities focus on two main groups of tasks performed by eLearning-based systems: annotation and selection. The annotation tasks are presented in Figure 3. The course materials are constantly annotated based on three domains: users’ input (annotations, blogs and tagging); instructors’ input (suggestions, modifications, constraints); and knowledge domain (vocabularies related specific topics covered by course materials). Closely related is a process of extracting learner’s profile. It leads to updates of the profile based on learner’s comments left in blogs, tagging, and notes left during interaction with other materials. The selection tasks are shown in Figure 4.
Experimentation activity line: This constitutes an important phase of work on the proposed framework and is exclusively devoted to the construction of the software prototype of the eLearning environment which once designed will be exploited as a comprehensive experimentation platform.

6.2. Development phases

The first phase is dedicated to construction of repository for learning materials, development of ontologies, as well as designing and implementation of mechanisms for concept identification from blogs and tag clouds, and for annotation and selection. All this leads to preliminary prototypes of the system. Intermittent pilot testing occurs to develop and test the reliability of the modified CSA and to inform the design process. A measure of ICT Literacy will be developed based on [13] and [14] frameworks that include measures of general cognitive abilities as well as technical skills.

The second phase focuses on modifications and extensions of already implemented mechanisms, addition of fuzzy-based selection schemas, as well as development of procedures for identification of learners’ profiles and evaluation of course materials. Based on design changes from the pilot test results, the first version of the proposed framework is tested with a sample of undergraduate education students. At the outset of the study, students complete the modified, shortened CSA and the ICT Literacy Test. Students have the opportunity to use the system on three occasions and will be asked to complete specific tasks within the system. After using the system, students complete a questionnaire on the usability of features of the system. Focus groups with small groups of students (4-6) is conducted to provide further data on the student’s experiences of using the system (preferences, impressions etc.). The full CSA is administered to test for correlation with the modified version. Students’ patterns of use in relation to their measured learner profile are examined.

Based on the assessment performed in the second phase, the system undergoes final design changes – tuning of algorithms and procedures. Eventually, the final version of the system is tested for an extended time period (a four week module in an online education course) in order to assess both usability and learning. Students again complete the modified, shortened CSA and the ICT Literacy Test. Relationships between learner profiles, technical skills and observed user behavior patterns are examined. To ensure that no learner profile is disadvantaged in the final learning assessment phase, learning is assessed with both fact-based multiple choice questions and open-ended conceptually-based questions. Again, the full CSA is administered to test for correlation with the modified version. Focus groups are conducted to access students’ qualitative experiences of the system.

7. Conclusion

One of the strengths of the proposed framework for development of eLearning systems is its multidisciplinary nature. This work leads to interesting results in several important areas:

- ontology and RDF triples – those are new and conceptually challenging forms of knowledge representation; activities related to adaptation of these forms to eLearning systems and their integration with interactive systems should lead to improvements in system’s abilities to store, access and analyze information;
- blogs and tagging – incorporation of these activities with eLearning systems should improve agility of eLearning systems and their ability to absorb users’ feedback, additionally their integration with new forms of knowledge representation should lead to a better analysis of information embedded in blog posts and used tags;
- fuzziness and multi-criteria decision making – as the core technologies of the framework they should play an important role in the process of creating more human-centric systems; at the same time the framework should provide an evidence of necessity of application of these techniques to development of real-world system able to support users’ activities in a personalized way; the combination of fuzziness with the new forms of knowledge representation – ontology and RDF triples – should increase presence of CI technologies on the Web.

The proposed eLearning framework constitutes a very important step towards direct application of fuzziness and new forms of knowledge representation to “real world” needs of eLearning systems. The framework also addresses the challenges imposed by human-centric systems.

References

[5] P. Gomes, B. Antunes, L. Rodrigues, A. Santos, and J. Barbeira, Using Ontologies for eLearning Personalization, 3rd ELearning Con-
[32] http://www.w3.org/RDF/ (access: March 27, 2013)
[33] http://www.w3.org/TR/owl2-overview/ (access: March 27, 2013)
[34] http://ruleml.org/ (access: March 27, 2013)