Automated Domain Model Creation for Adaptive Social Educational Environments

Marián Šimko*
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 3, 842 16 Bratislava, Slovakia
simko@fiit.stuba.sk

Abstract
Adaptive educational systems tailor learning material to user goals, needs and characteristics. While supporting more effective learning, they require semantic descriptions enabling adaptation engines to make at least basic reasoning. However, creating such descriptions manually is extremely demanding task. The situation is even more complicated when considering adaptation in social collaborative environment with content generated dynamically by learners on daily basis. In this paper we present a method for automated domain model creation based on processing of resources authored by teachers and social annotations created by learners.

Categories and Subject Descriptors
I.2.6 [**Learning**]: Knowledge acquisition, Concept learning; I.2.7 [**Natural Language Processing**]: Text analysis; K.3.1 [**Computer Uses in Education**]: Collaborative learning, Computer-assisted instruction (CAI)

Keywords
adaptive learning, social learning, domain modelling, text mining, user generated content, usage mining

1. Introduction
Massive spread of Web 2.0 technologies changed the face of web-based learning since they improved user experience resulting in more interaction, communication and collaboration during learning. A learner is no longer a passive consumer of information. He is not isolated by reading static pages, now he benefits from learning and collaboration in a social learning environment. Even more, he himself contributes to the content by various forms of annotations: he tags, rates, comments and gives a feedback.

The paradigm shift affects adaptive learning as well. In adaptive learning, the learning experience differs among different learners. Delivered learning material is tailored according to user goals, needs and characteristics [2] resulting into more efficient learning. The magic behind adaptivity is domain and user modelling enabling to track user knowledge of individual domain concepts and to apply predefined adaptation rules. Adaptive learning system needs semantic descriptions of learning resources varying from conceptual maps (typically consisting of concept hierarchies) to complex domain ontologies [7].

The inconvenience with adaptive learning systems is caused by the complexity of semantic description creation. The vast number of concepts and relationships makes it almost impossible to create an adaptive course metadata during reasonable period of time manually. The ability to provide learning resources with necessary descriptions is reduced even more when considering adaptive learning in dynamic environments with user generated content being created on daily basis.

One of the current challenges in adaptive learning is extending adaptation beyond static content created by a teacher towards content generated by students. The obvious question arises: To what extent are we able to assign user generated content appropriate descriptions automatically? In order to make it feasible we proposed an adaptive learning framework based on *lightweight domain modelling*, which considers pitfalls of collaborative social learning [14]. Besides lightweight domain modelling, the framework also addresses problems with social-based user modelling and personalization [1]. We believe that lightweight domain modelling, although slightly simplifying domain resources' semantic descriptions, will result into real adaptive collaborative social learning environment offering more attractive and efficient learning.

In this paper we present a method for domain model automated creation by leveraging both teacher authored content and social annotations. The rest of the paper is structured as follows. In section 2 we discuss related work. Section 3 gives an overview of the proposed method for automated domain model creation and describes its steps. In section 4 we sum up and conclude our work.

*Doctoral degree study programme in field Software Engineering. Supervisor: Professor Mária Bieliková, Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies, STU in Bratislava. Work described in this paper was presented at the 7th Student Research Conference in Informatics and Information Technologies IIT.SRC 2011 and awarded by ACM Slovakia Chapter Award.

Šimko, M.: Automated Domain Model Creation for Adaptive Social Educational Environments. Information Sciences and Technologies Bulletin of the ACM Slovakia, Special Section on Student Research in Informatics and Information Technologies, Vol. 3, No. 2 (2011) 119-121

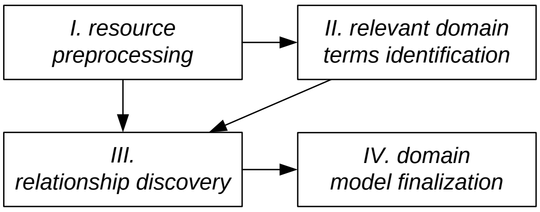

Figure 1: A scheme of the method for automated domain model creation.

2. Related Work

The number of approaches related to automated semantic acquisition for adaptive web-based learning systems is small. The only relevant evidence of (semi-)automatic domain model generation we are aware of is presented by in [5], where relationships between domain entities are acquired based on the comparison of their domain attributes. There are aslo attempts to create other models of adaptive systems – a goal model by generating course prerequisites [12] or an adaptation model by approximating adaptation rules [8].

There are many approaches to term extraction known as automatic term recognition (ATR) algorithms [9]. Although they are domain independent, those yielding best results are supported with background corpora, which are targeted on resources in English. Concept relationship discovery is a task from ontology learning field [4]. The relations being created typically have taxonomic character (is-a). Considering text mining, related approaches mainly utilize natural language processing (NLP) techniques. Relations are induced based on linguistic analysis [6]. Better results achieve approaches relying on preceding text annotation [3], i.e., they need a human assistant.

Our approach novelty lies in unsupervised automatic processing of both learning objects (created by teachers) and social annotations (created by learners). To our best knowledge, we are not aware of similar approaches in the field of adaptive social learning.

3. Automated Domain Model Parts Creation

A lightweight domain model's metadata layer consists of relevant domain terms (RDTs) interconnected by various forms of relationships, associated to resources (explanations, user-generated content) [14]. The goal of the proposed method is automated creation of a domain model, i.e., identifying relevant domain terms for underlying resources and creating relationships between them. The method builds on the research conducted on "static" adaptive course automated domain model creation [13].It consists of the following steps (see Figure 1): I. resource preprocessing, II. relevant domain term identification, III. relationship discovery, and IV. domain model finalization.

In the first step, a resource representation for further processing is prepared. The learning objects are analyzed and extended vector representation is composed. In the second step we select the most relevant terms in each learning object representation and create resource-term associations. Relationship discovery step is crucial for the whole method. We perform graph, linguistic and statistical analysis to discover two types of relationships: relatedness and is-a relationships. The fourth step of the method is domain model finalization, where a teacher (optionally) modifies a created course according to his needs.

3.1 Resource Preprocessing

In this step we create a vector representation of resources contained within a course. According to the originator of resource, we differentiate between learning objects authored by a teacher and social annotations (as a common term for all student generated content). Learning objects preprocessing consists of text analysis, formatting analysis and vector representation composition [13]. In the fist step we perform lexical analysis, identify lexical units, remove stop words and retrieve word's lemmas. We compute lemmas weights (we employ tf-idf measure) in order to compose a vector representation. If available, the obtained tf-idf weights are consequentially modified according to formatting rules in a source learning object [10].

In addition to learning objects, we preprocess annotations created by students by applying the following steps:

1. annotation content processing – a particular content processing technique differs among annotation type being processed,
2. annotation filtering – we filter non-descriptive and low rated (user rating, popularity) annotations,
3. learning object vector representation adjustment – we extend a vector representation of each learning object based on its associated social annotations.

3.2 Relevant Domain Term Identification

The goal of relevant domain terms identification is to select domain terms, which are relevant as domain descriptors. This part consists of the following steps:

1. relevant domain term selection,
2. relevant domain term weight normalisation,
3. relevant domain term to learning object relationship creation.

In the first step, we select top-k% associated terms, referred to as relevant domain terms, for each learning object. In the second step we recompute and normalise weights of each RDT for each learning object it is associated with. Finally, we create weighted *has-rdt* relationships between learning objects and RDTs. We created a bipartite graph consisting of two node types (learning objects, RDTs) and one edge type (*has-rdt* relationship), which is utilized in the next step of the method. Additionally, we create *has-rdt* relationships also between annotations and relevant domain terms.

3.3 Relationship Discovery

In this step we discover relationships between relevant domain terms. We consider two relationship types: *relatedness* relationship and *is-a* relationship. Relatedness relationship represents a basic, unspecified, level of relatedness between RDTs. In order to discover relatedness relationships, we utilize several knowledge discovery techniques based mainly on existing graph analysis [13]. In our current work, we extend relationship discovery with creating hierarchical is-a relationships. In order to identify is-a relationship, we apply two independent techniques, each resulting into own relationship set:

- lexico-syntactical patterns identification,
- resource-RDT (*has-rdt*) relationship processing.

In the first technique we perform syntactical analysis of preprocessed resources. We aim at lexico-syntactical patterns identification [6]. We currently assemble such patterns for Slovak language. As the state-of-the-art methods' precision of pattern matching is not optimal, we also consider existing relatedness relationship between RDTs – based on the assumption that two RDTs are more likely to have *is-a* relationship, if there already exist relatedness relationship in between.

The second technique is based on resource-concept relationships. We employ set theoretical principle often used when constructing taxonomies [13]. It builds on an assumption that "in an ideal situation, Entity A is a supersumed by Entity B if the set of entities classified under B is a subset of the entities under A". For each RDT we find a set of subsumed RDTs (i.e., hyponyms) by traversing all *has-rdt* relationships. In contrast to approaches to taxonomy composition, where concept graph has to form a component, when creating relationships between RDTs for adaptive web-based domain model such a restriction does not exist. Discovered *is-a* relationships are rather an enrichment of a domain model, which are aimed to improve adaptivity of a system; they are not mandatory to basic adaptive web-based system functionality.

In order to further improve the precision of both techniques, we also consider common learning objects book-like structure. The basic idea lies in the assumption that if concepts are related via *is-a* relationship, they are more likely to be assigned to resources arranged in hierarchical way (e.g., chapter and subchapter). This way we fine-tune both sets of relationship candidates. When creating *is-a* relationships, each relationship candidate is assigned a reliability weight indicating a certainty level of correctly identified relationship type. It is derived from *has-rdt* relationships and it also considers a relevancy of a particular technique. When fine-tuning relationship candidates, we automatically filter out only relationship candidates with reliability weight not exceeding a defined threshold.

The last step of the method is domain model finalization. A teacher can verify appropriateness of the generated domain model. He eventually modifies generated RDTs and relationships. Although the step is optional, the method probably will never produce a "perfect" output. Knowledge discovered by the method can be supplemented with the teacher's real-world experiences and skills.

4. Conclusions and Future Work

In this paper we tackled the potential bottleneck of adaptive social web-based learning – complex domain modelling emphasized by frequent content changes typical for dynamic social learning environments. We advocated *lightweigt* domain modelling, which represents potential benefit for social learning: social learning benefits from lightweight domain modelling as it facilitates user generated content processing; lightweight domain modelling profits from social learning as social annotations can be automatically used as domain resource descriptions.

We proposed a hybrid method for automated domain model creation based on heterogeneous sources (content provided by a teacher and social annotation coming from learners) processing. The novelty of the model lies in unsupervised automatic processing of both resources types. Currently, we work on multi-layered evaluation of the proposed method. First results yielded 70.6 % precision of generated *is-a* relationship constituting a promising basis for further experiments and potential enhancements of the method.

Acknowledgements.
This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010, and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

References

[1] M. Barla. Towards Social-based User Modeling and Personalization. In: *Information Sciences and Technologies Bulletin of the ACM Slovakia*, Vol. 3, No. 1, 2011, pp. 52–60.

[2] I. Beaumont, P. Brusilovsky. Adaptive educational hypermedia: From ideas to real systems. In Maurer, H., ed.: *Proc. of ED-MEDIA'95 – World conf. on educational multimedia and hypermedia*. Graz, Austria, 1995, pp. 93–98.

[3] P. Buitelaar, D. Olejnik, M. Sintek. A protégé plug-in for ontology extraction from text based on linguistic analysis. In: *Proc. of the 1st European Semantic Web Symposium*, 2004.

[4] P. Cimiano. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer, 2006. 347p.

[5] A. I. Cristea, A. de Mooij. Designer Adaptation in Adaptive Hypermedia Authoring. In: *Proc. of the Int. Conf. on Information Technology: Computers and Communications ITCC'03*. Las Vegas, US, IEEE, 2003, pp. 444–448.

[6] M. Hearst. Automatic acquisition of hyponyms from large text corpora. In: *Proc. of the 14th International Conference on Computational Linguistics*, COLING, 1992, pp. 539–545.

[7] N. Henze, W. A. Nejdl. Logical Characterization of Adaptive Educational Hypermedia. In: *New Review of Hypermedia and Multimedia, NRHM*, 10(1), 2004, pp. 77–113.

[8] P. Karampiperis, D. Sampson. Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. In: *Educational Technology & Society*, 8 (4), 2005, pp. 128–147.

[9] P. Knoth, M. Schmidt, P. Smrž, Z. Zdráhal. Towards a Framework for Comparing Automatic Term Recognition Methods. In: *Znalosti 2009*, SUT, Bratislava, 2009, pp. 83–94.

[10] M. Lučanský, M. Šimko, M. Bieliková. Enhancing Automatic Term Recognition Algorithms with HTML Tags Processing. In: *Proc. of Int. Conf. on Computer Systems and Technologies*, ComSysTech'11. Accepted, 2011.

[11] P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In: *LNCS 3729, The Semantic Web, Proc. of 4th Int. Semantic Web Conference*, 2005, pp. 522–536.

[12] S. Sosnovsky, P. Brusilovsky, M. Yudelson. Supporting Adaptive Hypermedia Authors with Automated Content Indexing. In: *2nd Int. Workshop on Authoring Adaptive and Adaptable Educational Hypermedia*, AH'04, 2004, pp. 23–26.

[13] M. Šimko, M. Bieliková. Automated Educational Course Metadata Generation Based on Semantics Discovery. In: *LNCS 5794, Proc. of European Conf. on Technology Enhanced Learning*, EC TEL'09, Nice, France. Springer, 2009, pp. 99–105.

[14] M. Šimko, M. Barla, M. Bieliková. ALEF: A Framework for Adaptive Web-based Learning 2.0. In Reynolds, N., Turcsányi-Szabó, M., eds.: *KCKS 2010, IFIP Advances in Information and Communication Technology*, Vol. 324. Springer, 2010, pp. 367–378.