Proceedings of the 1st Doctoral Consortium in Technology Enhanced Learning

Crete, Greece, October 2, 2006

1st PROLEARN Doctoral Consortium in Technology Enhanced Learning Crete, Greece, October 2, 2006

In today's knowledge society companies are aware of the crucial role that education and training, life-long learning, and knowledge management play in shaping a workforce able to strengthen its competitive position. The ever increasing proliferation of advanced information and communications technologies offers a fertile terrain for cultivating novel pedagogical scenarios for learning at the workplace. To effectively design, implement, and manage technology to enhance professional learning requires an interdisciplinary understanding of a diversity of scientific fields within the broad categories of educational sciences, computer science, networks, electronics, and marketing. Such is the challenge of the PROLEARN Network of Excellence, funded by the European Commission's Information Society Technology programme. PROLEARN's core mission is to bring together the most important research groups, professionals, as well as other key organisations and industrial partners working in the area of Technology Enhanced Learning. By fostering joint research, education and training, and virtual professional communities, PROLEARN is bridging the currently existing gap in cross-domain research, and training a new generation of researchers and experts to be able to manage complex technology enhanced learning solutions for the workplace.

The PROLEARN Academy (www.prolearn-academy.org) is more especially implicated in developing an institutional culture, bonding distributed researchers across Europe into one community. To achieve this goal we organise three main activities:

- To continuously educate, train and network world-class graduate and postgraduate researchers or junior faculty to maintain leadership at the PEOPLE level;
- To offer a forum for publication, debate and envisionment in order to maintain European leadership in topics across the full range of research issues in Technology Enhanced Learning covered by PROLEARN working groups;
- To implement a technology infrastructure to support a pan-European virtual community of researchers and training measures in the area of Technology Enhanced Learning.

Operational since 2004, the PROLEARN Academy has successfully inaugurated an annual summer school for early stage doctoral students. We have held two summer schools. The first PROLEARN Summer School took place in Sile, Turkey, September 5th -9th, 2005, with about 40 doctoral students working in the diverse areas of technology enhanced learning. The second PROLEARN Summer School took place in Bled, Slovenia, June 5th - 9th, 2006, which attracted about 45 doctoral students again.

As a result of our experience with the summer schools involving mostly early-stage PhD students, we wanted to create such a forum for doctoral students which are more advanced in their thesis work and who would welcome deeper and more personalized feedback about their research from distinguished researchers in the area of technology enhanced learning. A major conference like the EC-TEL is an excellent event for such a forum because many of these famous researchers will be congregating there. Therefore, we decided to initiate our first Doctoral Consortium at EC-TEL 2006. We were very proud and pleased to have received such a large number of proposals of very good quality. Unfortunately, we were only able to have a limited one-day time slot at the conference. Because we wanted to give detailed and constructive feedback to the students participating at the event, and be able to discuss doctoral work in technology enhanced learning, we made acceptance decisions not only based on the quality of the proposal but also on the relevance to technology enhanced learning and the maturity of conducted research. Thus we had a 58% rate of acceptance.

These are the proceedings of the First PROLEARN Academy Doctoral Consortium 2006 which will be held at the First European Conference on Technology Enhanced Learning in Crete, Greece, October 2, 2006. They include ten high quality papers from eleven authors, revised by at least two reviewers from the programme committee of the EC-TEL 2006 conference. The feedback was used by the doctoral students to improve their papers. Even more feedback will be given at the doctoral consortium itself.

We would like to thank first of all the doctoral students for their implication in advancing the state-of-the-art in the research field of technology enhanced learning. We would like to thank all of the reviewers whose precious time has been so instrumental in giving professional guidance to the doctoral students. We would like to thank the organisers of the EC-TEL for giving us the opportunity to host our Doctoral Consortium at their event. And of course, we would like to thank the PROLEARN and the PRO-LC cluster in disseminating and promoting this event, thus contributing to the institutionalisation of our virtual PROLEARN Academy.

Katherine Maillet and Ralf Klamma
Doctoral Consortium Chairs
Aachen and Evry, September 22, 2006

Table of Contents

Author Index

Aspects of Personalization in Language Learning Process and Greek Literacy in a New Technology Based Environment

Olga Fragou

Educational Technology Laboratory (ETL),
National and Kapodistrian University of Athens (NKUA), School of Philosophy,
Department of Philosophy, Pedagogy and Psychology,
University Campus, Ilissia, Athens, Greece
ofragou@ppp.uoa.gr

Abstract. An output of up to 60.000.000 prototype words of the whole Greek Literature corpus extending from 8BC to 16^{th} century AD, shows the multiplicity of structure schemata and the plethora of conjugation types of ancient greek language. Focus of this paper is New Technologies' implementation for structuring *a flexible, web based learning environment in a Greek, University students' undergraduate course*, under a critical analysis scope of trends in modern language learning methodologies.

Strategies of language acquisition/ competence, *motivational factors*, aspects of the so called "traditional method" of ancient greek teaching shall be further examined to promote an elevated sense of personal control and examine parameters of skills linking *instructional strategies*, *motivational processes* and *learning outcomes*. As current trends in learning theory evoke changes in forming education the issue of classics' modernisation and preservation of its aura comes forward by enriching it with a "modern persona" for youngsters .Learners' active engagement in *"decoding text"* is intended to become a core element for setting the design principles of a rich in stimuli learning environment in an effort to understand the underlying process of ancient greek learning and creating a *process syllabus* for the subject.

Keywords: *language learning methodologies, classics education, web based learning environments, Greek Higher education*

1 The research field: emerging issues

Focus on educational research has been on the development of new pedagogical models for e-Learning, mobile learning, learning on demand and professional life long learning. The achievement of individualized learning at a massive scale using traditional approaches seems a utopia: diversity of target population participating in learning activities, diversity in learner' s access media and modalities, the anticipated proliferation of free educational content set the scene for a new "learning" reality. Under this scope, *personalization* [1] procedures benefit from learner models which enclose observations from different sources (user interaction, learner self-reflection, peer reflections) while important parameters are: interoperable learner model artifacts, techniques for describing such artifacts, methods for extracting relevant learner model parts for particular learning situations or services, techniques for exchanging and communicating such artifacts. An important issue to realize the potential of personalized and adaptive learning is to integrate the variety of perspectives combined for these two tenets: Psychological Backgrounds, Diagnostics, Models of Competence, Personality, Knowledge, Social Psychology; Cognitive Science, Models of Cognition, Linguistic Approaches, Artificial Intelligence, Computer Science Approaches, User Modelling, User Tracking, Profiling, Machine Learning Methods, Business Applications of Personalization Tools, Human Resource Management, Employee Self Services, Business Processes underlying Personalization. This integration can build on a sound body of work in these areas [2], [3], [4], [5].

There are several systems that employ adaptive techniques to enable and facilitate different aspects of learning [6]. A dichotomy appears between typically commercial, standard e-learning systems and research prototype of Adaptive Learning Environment. The choice of Web becoming a standard as a development platform actually has granted long life to a few Web-based adaptive educational hypermedia systems such as ELM-ART [6],[7], InterBook [8] and 2L670 [9], Hy-SOM [10], AHM [11], CHEOPS [12].

Models' expansion in language learning has actually occurred with the advent of *Intelligent Language Tutoring Systems (ILTS)* emphasizing on learners' language input and feedback provision. *Computer Assisted Language Learning (CALL),* major trend in modern languages as a highly eclectic field integrates principles from Computer Assisted Learning and linguistics based on *tools*, (text processors, voice processing programs etc) , *instructional programs* (drills and practice, tutorials, etc), databases (information sources, databases, text corpuses, hypermedia,) and testing programs (computer language testing programs). Raising awareness as well as empowering learners to be able to deal with stereotypical viewpoints are important aims of modern language curriculum, as human language is used for comprehension and generation of optimal learners and optimal comprehenders.

Contextualizing' lesson presentations has become a widely accepted rule of good language teaching adherent to current English as a Foreign Language (EFL) pedagogy, which expounds that more responsibility for learning be placed upon the learner, taking into consideration learning styles and preferences. *An interesting categorization of cognitive styles has been made by Sternberg and Grigorenko* [13] *while r*esearch on learning theories and refinement of the constructivist learning model proceeds by user requirements' analysis and their feeding into the appropriate standards corpus.

A major issue in modern language methodologies is the distinction between the subconscious process of *acquisition* and *learning*, the conscious process of "*knowledge about language"* [14]. Smith [15] presupposes two conditions for language acquisition: learners' supposition they are successful learners of target language (TL) and their consideration as members of a language users' community.

Theories of *Comprehensible input* as the essential ingredient in language acquisition and the only way of stimulating the Language Acquisition Device (LAD) [16] as well as The *Affective Filter Hypothesis* [17] have prevailed in language learning tenets [18]. Language classes have been effectively supplemented in recent years in academic situations by the *sheltered subject* class matter [19].

2. The Existing Approaches

A massive expansion in the application of information theoretic and machine learning (ML) methods to *natural language processing (NLP)* for the past 15 years has produced impressive results in accuracy and coverage for engineering systems, addressing a wide variety of tasks in areas like speech recognition, morphological analysis, parsing, semantic interpretation, and dialogue management. Natural Language Processing [20] is often integrated in CALL, as the domain of language learning is the first "candidate" for the application of computational linguistics tools.

Different language technologies are applied in "programs designed to help people learn foreign languages": morphology and lemmatisation, syntax, corpus-based language acquisition, speech processing, etc. Chomsky [21], [22], [23] argues for a richly articulated language acquisition device to account for the speed and efficiency with which humans acquire natural language on the basis of "sparse" evidence. Chomsky's *Universal Grammar (UG)* and Ferguson's definition of *diglossia ,* encourage an individualized, learner-centered methodology in which different learners are motivated to set their own goals and assess their own progress. The *Common Underlying Proficiency Model* by Cummins (CUP) [24] emphasizes the importance of first language development in achieving high levels of competence in their second language especially in literacy related academic abilities.

Various methodologies have been developed in language learning field deriving from the need of adjusting to current educational trends and practices but also from the impetus for taking into consideration personalization processes. The *Grammar Translation Method*, and *Direct Method* have been widely used in classics domain [25], [26] . Attempts of modernisation have been mainly focusing on the structural aspects of ancient greek language : the *Greek Course* (*GC*)(early 70s) by the *JACT* scientific team based on structuralism highlighting cultural components and having a communicative orientation, as well as the digital Library of *Perseus* [27] (http://www.tufts.edu) , an 8 year research enterprise of the Tufts University in USA , allowing territory for personal learning action, structured by links and approaching texts through images.

3. The Proposed Approach

Learning literacies in a changing world [28] sets new educational needs, demanding reflection and data for learning process and language acquisition. A critical analysis of methodological approaches in *effective language learning* [29] provide basic features for designing a flexible, web based learning environment for ancient greek. Based on principles of comparative teaching methodologies in modern language learning field (*Grammar Translation method, TPR, CLL, Communicative* *method*), an eclectic approach is intended to be adopted and didactics' exploration of a classical subject in an effort to "merge" functional principles of both tenets. Basic aim is to personalize knowledge designing a learner directed process- syllabus for a classics subject, encouraging learners' discovery and acquaintance with basic principles of ancient greek language structural system to increase contextual familiarity [30] and self relevance [31]. Consideration of Technology Based Language Learning trends and theories (eg., *CALL, TELL*) is intended, re-enforced also by principles of *language competence* [32].

As text is a non changeable educational material and as learners/ educators change their approach depended on current educational needs, there is an impetus for *deep language processing,* aimed at better structure/ genre comprehension [33] and choices in interpretation encouraging learner's "personal critical reading". Means for achieving critical thinking skills in LL is autonomy, an important issue in self directed language learning**:** a goal of developmental learning control of one's learning [34] which allows for learners' choice based on effective *strategy use* [35]. Under this scope basic aim is to provide heuristics for "constructing" the autonomous language learner through use of prior learning experience and foster awareness on meta cognitive strategies and reflection, supported by the revelation of learners' epistemologies. Language awareness and meta cognition concepts are approached through personal organization of learning experience aided by the orchestration of an array of technological tools, adherent to learners' personal needs.

To achieve autonomy, focus on monitoring learners' target setting is intended, supported through partnership with others beyond the formal learning environment in an effort to critically examine prerequisites of self regulation and language competence: this, provided careful attention to individual learning styles and motivation factors. Standardized methods and classroom approaches will no longer be sufficient for individual needs: exploration of theories, dominant practices, the *repertoire* concept prove significant in a "learning society", based on "reflection on action" brought forward by Schon's theory [36].

"Drill and practice" software use in language learning is not a panacea: technological determinism seems to lose ground whenever careful consideration of educational settings and learner's needs is excluded. Previous research on software design identified different approaches of handling complex design and used bipolar descriptions, such as "top down" versus "bottom up" [37] or "planning" versus "bricolage" [38].Under this scope, use of *exploratory software* [39] is intended in an effort to uncover learners' actual planning decisions as a means of organizing language and literacy activities. Learners do act as designers when confronting a complex task building thus on incremental learning [40] and become active architects of their own knowledge. "Microworlds" [41] highlighting the "thinking through process" through playfulness and experimentation and computer's use as *a tool-kit* lead to perceiving learning as a process of "continuous adjustment" [42] : these are tools for revealing adults' language learning process by reflecting their conceptions and attitudes on the specific subject's methodology. Multiplicity in dynamic manipulation, visualization, information processing, , symbolic and written expression, learning mechanisms' stabilization achieve flexibility in applied educational practice: *programmability and interactivity* [43]. Under this scope, principles of *active learning* [44] and *self-regulation* [45], [46] are also intended to be highlighted so as to achieve a sense of relatedness and control or competence, crucial task aspects [47] Being adjusted to learners' goal systems, exploratory software encourages intrinsic motivation linked to deep level learning, self regulation [48] promoting further *ownership* [49].

. Models for Learning Processes and especially for collaborative processes are a focus of attention from different research areas recently: instructional designers and practitioners, look for a notation to represent their learning designs both for communication among each other and that can be implemented in computer based scenarios [50]. In this context, *Learning Management Systems (LMS)* [51], [52] as tools for designing, managing and delivering online collaborative learning activities are based on principles of Learning Design in a systematic process of translating general principles of learning and instruction into plans for instructional material of study.

4. The methodology scheme

Researcher's aim is to explore New Technologies' use in language learning *trying to identify design principles of a dynamic, pedagogically flexible environment for language concepts' and literacy ideas' exploration* based on Greek Higher Education context, standards of classical language pedagogy and findings of modern language learning methodologies.

A *learner-centered technology- based approach is intended highlighting self-pacing:* student judgment in selection of objectives, management of materials and resources used, self-direction in reaching objectives and self-evaluation so as to develop a flexible pedagogic and technological architecture that enables structuring of learning objects and their components . *An effort to set a taxonomy of language behavior in ancient and modern greek* embedded within setting learning objectives by identifying criteria for *skill levels*, within the concept of language skills is crucial; a detailed taxonomy of *notional-functional categories* for *syllabus design* driven by meaningful categories of communication.

i. Research design

A *design research* scope [53] is intended to be adapted enabling intertwining of designing learning environments' central goals and developing theories, with continuous cycles of design [54] to reveal broadly the nature of learning in a complex system, refine generative or predictive learning theories and models of successful innovation [55] rather than artifacts. Supported by Laurrilard's *"conversational framework"* [56], [57] and Engestrom's expanded framework [58] an effort to refine locally innovation's nature and develop globally usable knowledge of the field is intended. Ethnography practices (observation, video taping) are expected to be used as a further research tool to complement for the technological data collection tools (artifacts) and ascertain validity of these. Researcher's contribution is expected in the scope of participant observation [59].

ii. Participants- Sampling Method

Re –evaluation and modification of methodological tools is expected (the thesis is in an embryonic stage) leading to amendment, customization of research tools and software standardization. Undergraduate students of Educational Technology Laboratory at NKUA, having already formed their personal language culture and become adequately acquainted with language principles and strategies in Greek secondary education, as experienced language learners, consist our sample.

iii. Data collection instruments, procedures, analysis, interpretation

Researcher's intention is to use a combined set of research tools and methods of contemporary educational practice as data resources:

- learners' prototypes-artifacts identified by extensibility and openness to their intervention based on *Ariadne*, an E-Slate microworld [60] designed by the researcher
- educational *formats (activity plans)* based on technology's creative implementation in language, literacy concepts and cognitive functions involved (http://noe-kaleidoscope.org)
- web based tools (e-portfolio/ LMS) as collection of realia and digital resource to monitor learner's language learning process, an effort for post-task self-ratings of comprehension, testing performance as a function of choice and personalization
- students' questionnaires and semi structured interviews

Description of data collection procedure is expected to take place in real time settings.

iv. Expected results

.

- data for language learning process/ acquisition providing integrated methodology heuristics for developing software environment and effective language tasks in ancient greek learning , based on communication, ability of self monitoring and group monitoring in process of activities
- Web based data come reflecting learners' beliefs and preconceptions on the specific subject, their perceptions of learning organization so as to determine core functional aspects of language learning methodology
- learners' flexibility and autonomy re-enforcement (macro and critical skills development) and exploration of conditions that sustain intrinsic motivation in a technology based ancient greek language learning environment
- New Technologies' use to enhance learners' independence and awareness of in-use strategies providing cognitive and hermeneutic tools for reading, speaking, writing ancient greek : monitor and develop text processing procedures, goal setting in genre, morphology, syntax, semantics, pragmatics
- production of educational formats - intended to cater for exploration of language concepts and written expression, retrieval, actual processing and organization, analysis and presentation of information

v. The original value of research

- explore and highlight important issues in learning ancient greek by approaching a classics subject through language learning methodologies' principles and New Technologies' use
- examine principles for a what so called *deep approach in classics' learning* , revealing emerging language learning models, promoting personalization in learning process , based on language perception, usage and characteristics
- explore didactical framework for effective text comprehension/ interpretation and critical reading/writing systematization and develop language learners' intelligence as adaptation : a comparative and co-relational method of learning modern and ancient greek
- determine and examine interconnections between self efficacy in language learning and New Technology use, exploring the nature of technology -based didactical interventions on the specific subject

DISCUSSION

People's learning as consistent sequence implies acting differently as a result of experience. Construction and reconstruction reveals self determination and autonomy in learning : *planning (goals, strategies), decision on further knowledge or resources, monitoring progress, evaluating goals, and terminating when the goals have been met.* Identification of functional components that capture major aspects of human teaching strategies such as *Theorization, Representation, Recognition, Execution, and Evaluation* seem important. Changing action theories, attitudes, and behavioral patterns, takes more than information: there is an impetus for a theoretical system that experience tests out, as well as reflection on the meaning of experience. An important way of integrating technology into powerful learning environments, putting the tools of creation into learners' ' hands, designing instructional units that channel student creativity into effective language learning activities are at the research scope, along with the principles of: *extension of learning boundaries, multiple intelligences' development, expanding curriculum choice and structure of Connected Learning Communities.*

Resolving tensions between a universal system and personal needs presupposes exploration of students' engagement and learning through a combination of modalities: *sharpening our perception of what is required of a syllabus which develops language competence , promoting ways of developing awareness, seems a starting point not a destination.* Based on the idea of educator Antoni Boleslaw Dobrowolski for *"universitas rediviva",* i.e., "renewing, living" university, learners need to concentrate on and develop their "learning skills" and "learning strategies". Performed thus in a global environment, language acquisition and learning (even in classics) is *a long-term investment* but not a skill or a body of knowledge that is likely to become outdated in a few years' time.

REFERENCES

[1] Weber, G. & Brusilovsky, P. : ELM-ART (2001): An adaptive versatile system for Web-based instruction. International Journal of Artificial Intelligence in Education 12 (4), Special Issue on Adaptive and Intelligent Web-based Educational Systems,pp. 351-384.

[2] Naeve, A., Nilsson, M. & Palmér, M. (2001) **:** *The Conceptual Web - our research vision*, Proceedings of the first Semantic Web Working Symposium, Stanford,www.semanticweb.org/SWWS/program/position/soi-nilsson.pdf

[3] Allert, H. & Richter, C. (2002) **:**Re-Designing an Educational Setting - Trails of Competency in an Open Learning Repository. E-Learn 2002: World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn 2002). Oct. 15-19, Montreal, Canada.

[4] Di Nitto, E. & Sassaroli, G. & Zuccalà, M (2003). : "Adaptation of Web Contents and Services to Terminals Capabilities: the @Terminals Approach", In Proceedings of First IEEE International Conference on Pervasive Computing and Communication (PerCom03), Fort Worth, USA

[5] Matera, M. & Maurino, A. & Ceri, S. & Fraternali, P (2003). : Model-driven Design of Collaborative Web Applications. Software Practice and Experience

[6] Brusilovsky, P.: Methods and techniques of adaptive hypermedia (1998). In: P. Brusilovsky, A. Kobsa and J. Vassileva (eds.): Adaptive Hypertext and Hypermedia. Dordrecht: Kluwer Academic Publishers, pp. 1-43

[7] Brusilovsky, P., Schwarz E. and Weber G.,(1996a): ELM-ART: An intelligent tutoring system on World Wide Web. In: C. Frasson, G. Gauthier and A. Lesgold (eds.): Proceedings of Third International Conference on Intelligent Tutoring Systems, ITS-96. Lecture Notes in Computer Science, Vol. 1086, Berlin: Springer Verlag, pp. 261-269.

[8] Brusilovsky, P., Schwarz, E. and Weber G.(1996b): A tool for developing adaptive electronic textbooks on WWW. Proceedings of WebNet'96, World Conference of the Web Society, San Francisco, CA, pp. 64-69, Available online at http://www.contrib.andrew.cmu.edu/_plb/WebNet96.html

[9] De Bra, P. M. E.,(1996), Teaching Hypertext and Hypermedia through the Web. Journal of Universal Computer Science 2(12), 797-804.

[10] Kayama, M. and Okamoto, T., (1998), A mechanism for knowledge-navigation in hyperspace with neural networks to support exploring activities. Proceedings of Workshop `Current Trends and Applications of Artificial Intelligence in Education' at the 4th World Congress on Expert Systems, Mexico City, Mexico, pp. 41-48.

[11] Pilar da Silva, D., Durm, R. V., Duval, E. and Olivie¨ , H.,(1998), Concepts and documents for adaptive educational hypermedia: a model and a prototype. Proceedings of Second Adapt- ive Hypertext and Hypermedia Workshop at the Ninth ACM International Hypertext Con-ference Hypertext'98, Pittsburgh, PA. Computing Science Reports 98/12, Eindhoven University of Technology, pp. 35-43.

[12] Negro, A., Scarano, V. and Simari, R., (1998), User adaptivity on WWW through CHEOPS Proceedings of Second Adaptive Hypertext and Hypermedia Workshop at the Ninth ACM International Hypertext Conference Hypertext'98, Pittsburgh, PA. Computing Science 108 PETER BRUSILOVSKY Reports 98/12, Eindhoven University of Technology, pp. 57-62.

[13] Sternberg, R., J (1988), Handbook of Human Intelligence, New York, Cambridge University Press

[14] Krashen S., (1993) The power of reading, Libraries Unlimited INC, Englewood Colorado

[15] Smith , E., (1982), Writing and the Writer, NY, Holt , Rinehart and Winston

[16] Krashen S., (1989), Language Acquisition and Language Education: extensions and applications, NY, Prentice Hall

[17] Dulay H., and Burt , M (1977)**,** Remarks on creativity in language acquisition. In Burt , M., Delay, H., and Finocchiarro, M., (Eds). Viewpoints on English as a Second Language. NY. Regents, pp.95-126

[18] Krashen S., (1981b), Second Language Acquisition and Second Language Learning. Oxford. Pergamon Press

[19] Krashen S., (1982), Principles and Practice in Second Language Acquisition. Oxford. Pergamon Press.

[20] Nerbonne, J. 2002. *Computer-Assisted Language Learning and Natural Language Processing.* In: R. Mitkov (Ed.) Handbook of Computational Linguistics, Oxford Univ. Press, pp. 670-698.

[21] Chomsky, N. (1986), *Knowledge of Language: Its Nature, Origin, and Use*, Praeger, New York.

[22] Chomsky, N.(1995), *The Minimalist Program*, MIT Press, Cambridge, MA.

[23] Chomsky, N.(2000), *New Horizons in the Study of Language and Mind*, Cambridge University Press, Cambridge.

[24] Cummins J., (1980) The cross- lingual dimensions of language proficiency: Implications for bilingual education, *TESOL Quarterly* (14), pp. 175-187

[25] Stray., C., (1998), Classics transformed: schools, universities and society in England, 1830-1960, Oxford, Clarendon Press

[26] Forrest, M., (1996), Modernising the classics: a study in Curriculum development, Exeter, University of Exeter Press

[27]Crane G., (1998) The Perseus Project and beyond: how building a digital library challenges the humanity and technology. Retrieved on March 2004 from http://www.dlib.org/dlib/january98/01crane.html

[28] Cope, B., Kalantzis, M., (2000) (Eds), Multiliteracies. Literacy learning and the design of social futures, Routledge , London and New York

[29] Lewis, T., Rouxeville, A., (2001), Technology and the advanced language learner; London: Association of French Language studies in association with CILT Publications

[30] Dalgarno B. (1998) Choosing learner activities for specific learning outcomes: a tool for constructivist computer assisted learning design. In C.McBeath, C., & Atkinson, R.,(Eds), Planning for Progress, Partnership and Profit. Proceedings EdTech'98 . Australian Society for Educational Technology, Perth.

[31] Cordova D. & Lepper M. (1996) Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, pp.715–730.

[32] Richards, J., Rodgers, T., (2001), Approaches and methods in language teaching; Cambridge: Cambridge University Press

[33] Schiefele, U .,(1991), Interest, Learning and Motivation, *Educational Psychologist, 26*, pp.229-323

[34] Benson., P., (2001), Teaching and researching autonomy in language learning; Harlow: Longman

[35] O Malley, J., M., Chamot, A. U., (1990), Learning strategies in second language acquisition; Cambridge: Cambridge University Press

[36] Schon, D. A., (1984), The reflective practitioner: how professionals think in practice. New York, NY: Basic Books

[37] Jeffries, R., Turner, A. A., Polson, P.G., Atwood, M.E. (1981), The processes involved in designing software . In J.R. Anderson, (Ed) , Cognitive skills and their acquisition Hillsdale NJ:Lawrence Erlbaum Associates

[38] Turkle, S., Papert,S., (1991), Epistemological pluralism: styles and voices within the computer culture. In I. Harel & S., Papert (Eds), Constructionism. Norwood, NJ:Ablex

[39] Papert, S., (1980). Mindstorms: Children, computers and powerful ideas. New York: Basic Books

[40] Gangarian, G., (1996), The art of design. In Y. Kafai & M. Resnick (Eds), Constructionism in practice: designing, thinking and learning in a digital world, NJ: Lawrence Erlbaum Associates

[41] Kynigos, C.,(1999), Software development, Innovative practice , Research and The School Context: is synergistic progress possible? Proceedings of the 7th European Logo Conference, Sofia, Bulgaria, Nikolov R., Sendova E., Nikolova I., Derzhanski I., (Eds) pp.62-79

[42] Crook, (1994), Computers and the collaborative experience of learning; London and New York: Routledge

[43] diSessa, A., (2000). Changing minds: Computers, Learning and Literacy, MIT Press

[44] Meyer C., & Jones, T. B. (1993). Promoting active learning: Strategies for the college classroom. San Francisco: Jossey-Bass.

[45] Bastiaens, T., Martens,R.,(2000), Conditions for web based learning with real events. In *Abbey, B (Ed), Instructional and cognitive impacts of web- based education*. London: Idea Group Publishing.
[46] Herrington J., Oliver, R., (2000), An instructional design framework for authentic learning environments, *Educational Technology Research and Development, 48(3),* pp.23-48

[47] Ryan , R. M., Deci, E. L., (2000), Self determination theory and the facilitation of intrinsic motivation, social development and well being, *American Psychologist, 55(1)*, pp.68-78

[48] Marton E., Saljo, R., (1984), Approaches to learning**.** In *E. Marton, D. Hounsell and N. Entwistle (Eds), The experience of learning.* Edinburgh. Scottish Academic Press

[49] Watson, D. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies*, 6 (4), 251-266.

[50] Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed). Three worlds of CSCL. Can we support CSCL (p. 61-91). Heerlen, Open Universiteit Nederland.

[51] Hernandez, D., Asensio, J.I., Dimitriadis, Y. (2004). IMS Learning Design Support for the Formalization of Collaborative Learning Patterns. Proceedings of the 4th International Conference on Advanced Learning Technologies (Aug.30 - Sep. 1, 2004, Joensuu, Finland), IEEE Press, p.350-354.

[52] Kollar, I., Fischer, F., & Hesse, F. (2003). Cooperation Scripts for Computer-Supported Collaborative Learning. In: Wasson, B., Baggetun, R., Hoppe, U., Ludvigsen, S.: CSCL 2003 Community Events, Bergen

[53] Design –Based Research: An emerging paradigm for educational inquiry, (2003), The Design Based Research Collective , in *Educational Researcher*, January / February 2003, 32(1), pp.5-8

[54] Cobb, P., (2001), Supporting the improvement of learning and teaching in social and institutional context. In *S., Carver and D., Klahr, (Eds), Cognition and Instruction: Twenty five years of progress.* Mahwah, NJ: Lawrence Erlbaum Associates

[55] Campione J.C., Brown, A.L., (1990**),** Guided learning and transfer: implication for approaches to assessment. In *N., Frederiksen, R., Glaser, A., Leigold , M., Shafto (Eds): Diagnostic motoring of skill and knowledge acquisition,* Hillsdale, NJ., Lawrence Erlbaum Associates ,pp.141-172

[56] Laurillard, D. (2002). *Design Tools for E-learning*. Keynote presentation for ASCILITE2002. http://www.unitec.co.nz/ascilite/proceedings/papers/key_laurillard.pdf [30th July 2003]

[57] Hedberg, J. (2003). Ensuring quality e-learning: creating engaging tasks. *Educational Media International,* 40 (3/4).

[58] Engestrom ,Y., (1987), Learning by expanding: an activity theoretical approach to developmental research. Helsinki:Orienta-Konsultit. [on line]. Available at **http://communication.ucsd.edu/MCA/Paper/** Engestrom/ expanding/toc.htm

[59] Robson, C., (1993), Real World Research; Blackwell: Oxford UK

[60] Kynigos, C., (1995), Programming as a means of expressing and exploring ideas: three case studies situated in a directive educational system. In *diSessa A., Hoyles, C., (Eds), The design of computational media to support exploratory learning*, ASI Series, Springer-Verlag

Research and Development of a Positioning Service for Learning Networks for Lifelong Learning[1]

Marco Kalz
Open University of the Netherlands
marco.kalz@ou.nl

Abstract. Positioning in learning networks is a process that assists learners in finding a starting point and an efficient route through the network that will foster competence building. This contribution presents the rationale for the positioning project and provides an overview about methodological questions for the research and development of a positioning service for lifelong learning and a short outlook on objectives and expected results.

Recognition of Prior Learning for Learning Networks

Lifelong technology enhanced learning should support flexible ways to build competencies across institutions and different contexts. While most of the former elearning models are built from an institutional perspective now the individual should stand in the centre of every effort in lifelong learning. The concept of learning networks addresses this problem and offers a framework to bridge the different contexts of current technology enhanced lifelong learning [1]. A learning network connects actors, humans as well as agents, institutions and learning resources. Information and communication technologies are used in such a way that the network self-organizes. The actors in the learning network share one common goal: furthering the development of competence by learners. Competence is defined here as effective performance in a domain at different levels of proficiency. Competencies include skills and they can be divided in 5 main competencies (cognitive, functional, personal, ethical and meta-competences) [2].

A traditional approach to overcome the limitations of institutional dependencies is the concept of Accreditation or Recognition of Prior Learning (APL/RPL) [3]. APL offers methods and techniques to identify prior learning experiences from formal and informal education. This procedure is especially important if a person crosses the boundaries between work and learning or between academic disciplines. Current practice in APL implements procedural solutions to the APL problem. These solutions stipulate the steps in a procedure in which the learner can present material to substantiate claims on particular exemptions. These claims are then evaluated on equivalence: has the learner substantiated the claim that s/he has acquired learning outcomes equivalent to those produced by the courses for which exemption is sought. Most of these APL methods rely on experts who study the learner profiles and decide

[1] *This project is carried out in the 6th EU Framework Programme for Research and Technological Development in the Integrated Project TENCOMPETENCE and it is funded through the European Union.*

which learning activities leading to a competence should be exempted and which ones are best suited as starting position for the student. But this way of positioning a learner is a very time-consuming and expensive approach. This project concentrates on computational approaches to address the positioning problem for lifelong learning and delivers techniques and tools that may help to clarify claims on equivalence thus bringing content-based analysis to the procedures of APL.

Research Question

The project concentrates on the following research question: Taking into account the goals and the history of a learner what is the best place for the learner to start and which activities should be exempted? To address this problem this project researches models to address the positioning problem and develop a prototypical solution useful for lifelong learners. The project will focus on the process of mapping competence information of learners to the competence information in learning networks. Since there are already (technical) solutions to present the individual situation of a learner like electronic portfolios and there are already many learning resources to support formal and informal learning there is a lack of research how these distributed parts of technology-enhanced lifelong learning can be connected. A positioning service helps learners to find the best entry position in learning networks according to his competence development goals. Conceptually it is important to stress that the positioning service should be seen as a recommender system for starting positions/exemptions but not as an automated system without human decisions. The service could be used either by the learner or by an expert who has to decide about exemptions. Since both sources of data (learner data & network data) can change very fast in an environment for lifelong learning a dynamical positioning service would have to redefine the learners' position every time a new competence related information has been added either to the learner profile or the learning network. The quality and success of a positioning service should be controlled through the criteria of reliability (the same situation leads to the same recommendation) as well as validity (the recommendation matches that of experts). The project will focus on the comparison of similar data for positioning. Three cases will be explored and addressed during the project: A content-based approach, an approach that additionally uses metadata for positioning and a third approach combining content, metadata and the use of competence ontologies and semantic-web technology.

State of the Art and existing solutions

From an educational perspective it is important for successful technology-enhanced learning to offer learners individualized learning experiences and educational resources that fit to the needs and individual situation of the learners. The positioning service should reduce the time required to reach learning objectives based on an automated portfolio assessment. Previous work from user and learner modeling for the generation of possible routes and the selection of learning resources for tailored and individualized instruction also takes into account prior knowledge and goals. But there are no consistent models and techniques to assess the prior knowledge of learners.

The project will research the development of a web service to position learners in learning networks according to the competence development goals they want to achieve. Web services provide a standard means of interoperating between different software applications, running on a variety of platforms and/or frameworks. Different situations can be given for the comparison of the competence related information of the learner with competence related information in the learning network. To focus the research project it will be limited to three different cases to address the positioning problem. The project will cover only situations where similar data can be compared. The following figure shows the focus of the project.

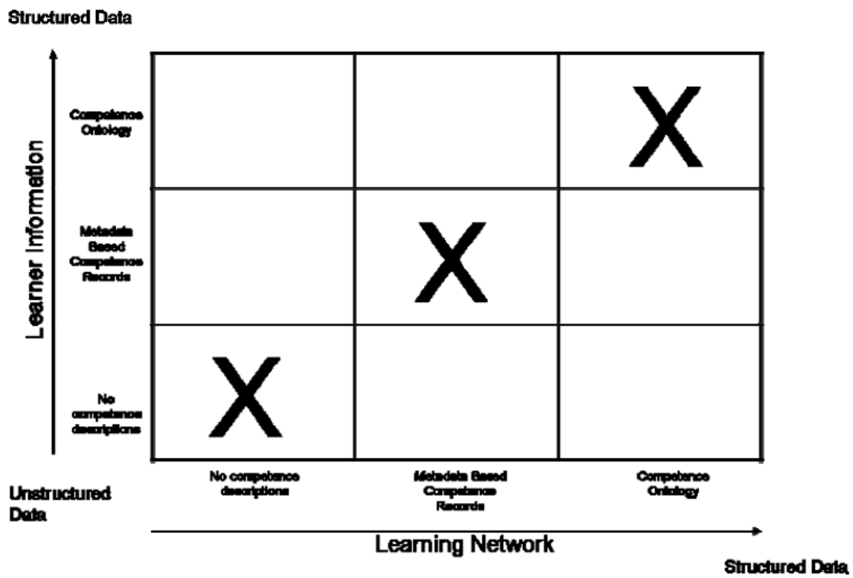

Fig. 1. Positioning Situations Matrix

In case one a positioning service has to handle a situation without any competence information in the learner profile and the learning network. This situation could occur when a learner does not already have an electronic portfolio and his learning network does not contain explicit competence related information but only learning activities and content. To approach this situation we make use of content analysis to compare the learner profiles and the learning activities in the learning network. We assume that there is a similarity of the contents of (learning) materials studied or produced by the student (source material) and the material contained in the learning activities in the learning network (target). If a positioning service determines that the content of source and target materials overlap substantially, the target activity is exempted. In the content-based positioning service document similarity is computed using latent semantic analysis (LSA).

The second case deals with metadata-based competence related information in the learner profile and the learning network. There are different standardization efforts in the field of competence related metadata. Five standards and specifications deal with the competencies of the learner: The IMS Reusable Definition of Competency or Educational Objective (RDCEO) specification aims at a standard description of competencies and educational objectives for online and distributed learning. RDCEO is expected to promote common understanding of competencies that can be used in competency development (learning and career development) or in specifying learning pre-requisites or learning outcomes [4]. The RDCEO offers a unique identifier to assign an unstructured competency description to an object for example in a Unit-of-Learning (UoL). Based on the RDCEO a draft standard for Reusable Competency Definitions (RCD) is being defined in the IEEE. Although RCD does not intent to offer a solution to the aggregation of competencies from sub-competencies the data-model allows the integration of relational information or competence ontologies through embedding additional metadata [5]. For portfolios two specifications are of interest. The IMS Learner Information Package Specification (LIP) is designed to package learner information for the exchange of data [6].

The IMS ePortfolio specification builds on the LIP specification to ensure portability and exchange of ePortfolio records for learners [7]. The specification is addressing different usage possibilities (assessment, planning of learning) and it can store produced artifacts form the learner and formal achievement records like references. A slightly different approach comes from the HR-XML Consortium. The consortium develops a standard suite of XML-specifications to allow the exchange of Human-Resource-related data, such as a competency schema for a variety of business contexts that is applicable in recruitment processes [8]. The model allows the evaluation, rating and ranking of competences which are an important issue in recruiting processes. On the resource side it is important to mention that the IMS Learning Object Metadata (LOM) has no element to store competence related information at the moment [9]. They could be stored in the educational segment of the metadata as proposed in [10] but this does not seem to be a widely adopted solution to the problem. Nonetheless in case two of the positioning problem competence descriptions in these metadata could be used for positioning by mapping them and finding similarities between the descriptions. The research needs for this part of the positioning problem are the methods and techniques that can be applied to compare competence related metadata and use similarities in metadata for the positioning service.

The third case is the comparison of competence ontologies in the learner profile and the learning network. Ontologies are metadata schemas providing a controlled vocabulary of concepts and they can be useful to share common understanding in a domain in a machine-readable way. For competence development ontologies or taxonomies can be used to define competences related to competence development programs in the learning network. Competence ontologies could be either added to the learner profiles [11], learning objects [9] or the competence development programs [12]. But the design and implementation of competence ontologies is still a very complex and time consuming task [13]. For competence similarity the number of different ontologies and the difference between vocabularies used is important. There are several models and techniques for measuring similarity of ontologies [14, 15]. As we may find competence ontologies in formal education it is very unlikely to find them in non-formal education. In an ideal situation every learning network could share a common understanding of the competences needed for successful running through the program based on ontologies. The ontologies in the programs could be added to the learner profiles step-by-step after they have successful finished the connected learning activities. In this case positioning could happen through the mapping of competence ontology inside the learner profile with the competence ontology in the learning network. The project will concentrate in this phase on the measurement of similarities for competence ontologies the use of this information for positioning.

Methodological Questions

Because of feasibility reasons we will concentrate on a situation from formal education where we will have predefined activities in the learning network. All experiments will be done in an introductory psychology learning network of the Open University of the Netherlands. In the beginning we will collect student data about their educational history and prior learning experiences. Besides plain information on their former courses we need also material or hints to material they have produced or studied in their former education. We will start to compare the content of the learner portfolios with the course content for case one. In the content-based approach document similarity is computed using latent semantic analysis (LSA).

LSA is a technique from natural language processing that was originated in the field of information retrieval [16]. LSA is based on word (co)-occurrences in documents, thus all order (syntax) of words or semantics in the original documents is ignored. All analyses are performed on a Term-by-Document matrix with word frequencies in the cells. The dimensions of this matrix are computed and the largest dimensions found (the semantic factors) are retained to reproduce the original matrix [17]. In the reproduced matrix each document is represented as a vector. The smaller the angle between two document vectors the higher they are correlated, that is, they are expected to contain materials that have substantial overlap. After the data collection we will build a corpus for each student and compare them through LSA to the content from the activities in the learning network.

The result will be a correlation between the data of the students and the activities in his chosen competence development program. To integrate also a self-estimation by the students we will ask them before and after every activity about their prior learning experiences according to the competences connected to the activities. By this means we can compare at the end (data interpretation) the results from the LSA engine to the results from the experts and the self-estimation by the students. The result of the experiment should lead to a model for content-based positioning in learning networks. To evaluate the results from the experiment we have to answer the following questions:

1. Is the recommendation valid?
   A valid positioning service should deliver recommendations that can be compared to human judgments. To control the validity of the recommendations from the positioning service the results will be compared to the results from two domain experts who should compare the learner profiles with the introductory psychology learning network. If they give a very similar recommendation after comparing the

learner profiles and the activities in the learning network the positioning service delivers valid results.

2. Is the recommendation reliable?
   The reliability of the recommendation should be controlled through the repetition of the positioning with very similar situations. If the same situation delivers always the same results the positioning service and the delivered recommendation is valid.
3. Is the recommendation efficient?
   The efficiency of the service should be compared to the time and cost of domain experts. For this issue we need to know how long a domain expert would need to analyze the learner profile with the learning network. Another efficiency issue is the need of computing power for a positioning service. The efficiency of the positioning service can at a later stage of the project be compared in relation to given data. In the second experiment we will compare the efficiency of metadatabased positioning to the content-based approach. As a result we should know after the project about the data requirements for positioning.

4. Are the basic assumptions sound?
   For the evaluation of the basic assumptions for positioning we will compare the results of the experiments. For example the comparison of competence ontologies or competence maps should give a direct comparison of competences while we assume in our first experiment that the comparison of content can also give sufficient results for the accreditation of prior learning experiences in learning networks.

The second model will concentrate on the comparison of a combination of different metadata for positioning. The hypothesis of the experiment is that positioning recommendations can be based on metadata in the learner profiles and the content of a learning network and that these recommendations are comparable to those of experts. The aim of the model is to develop a solution for a positioning service if metadata are available in the learning network and the learner profiles. As a first step we will add metadata to the learner profiles and the activities from the introductory psychology learning network. For this part of the experiment we have to be aware that the quality of the metadata-based positioning depends on the quality of the provided metadata. Since we will use in parts the same learner profiles as in the first experiment we can compare the results of both experiments to control the added-value of integrating metadata in a positioning service. The approach for the research and development of apositioning service is planned incrementally so we will also test the combination of a content-based and a metadata-based approach.

The third case will concentrate on the comparison of competence ontologies. In the third experiment we can compare the added-value of positioning of all three experiments and we can also model a combination. Assuming we will have a rich set of data (content, metadata and competence ontologies) how can a positioning service give a valid recommendation? Since the research project is divided into three phases defined by the above described cases, the result of every phase should add a model and techniques to the development of a prototypical positioning service.

Objectives and Expected Results

The first objective is the research and development of a model for a positioning service. The model should cover three cases for the comparison of similar competence related data and result in a recommendation for a starting position or an exemption decision inside a learning network. The second primary objective is the development of a prototype of a positioning service. This web service should support lifelong learners and help them to be positioned in learning networks. The computational support of the accreditation of prior learning is previously not well researched in the field of educational technology. Additionally the combination of the three cases for a lifelong learning perspective is unique and can provide a solution to bridge the previously unconnected contexts of lifelong learning. As a secondary result the project should stimulate the discussion of a positioning service for portfolio assessment and formulate requirements for electronic portfolios. Since the project will be carried out in the EU funded integrated project TENCOMPETENCE all results of the project will be published under an open source or open content license.

References

1. Koper, R., Rusman, E. & Sloep, P. (2005). Learning Network connecting people, organisations, software agents and learning resources to establish the emergence of effective lifelong learning, LLine: Lifelong Learning in Europe, vol. 9, no. 1, 18-27.
2. Cheetham, G. & Chivers, G. (2005). Professions, Competence and Informal Learning, Cheltenham:Edward Elgar Publishing.
3. Merrifield, J., McIntyre, D & Osaigbovo, R. (2000) Mapping APEL: Accreditation of Prior Experiential Learning in English Higher Education. London.
4. IMS Global Learning Consortium (2002). IMS Reusable Definition of Competency or Educational Objective Specification. Retrieved 1st of July, 2006 from http://www.imsglobal.org/competencies/
5. IEEE Learning Technology Standards Committee (2006). Reusable Competency Definitions (P1484.20.1, Draft 3). Retrieved 1st of July, 2006 from http://www.ieeeltsc.org/wg20Comp/wg20rcdfolder/IEEE_1484.20.1.D3.pdf
6. IMS Global Learning Consortium (2001), IMS Learner Information Packaging Specification. Retrieved 1st of July, 2006 from http://www.imsglobal.org/profiles/
7. IMS Global Learning Consortium (2005). IMS ePortfolios Specification. Retrieved 1st of July, 2006 from http://www.imsglobal.org/ep/index.html
   HR XML Consortium (2004). Competencies (Measurable Characteristics). Retrieved 1st of July, 2006 from http://ns.hr-xml.org/2_3/HR-XML-2_3/CPO/Competencies.html
9. Ng, A., Hatala, M. & Dragen Gasevic (2006). Ontology-Based Approach to Learning Object Formalization, In: Sicilia, M.-A. (Ed.), Competencies in Organizational ELearning: Concepts and Tools, Idea Publishing 2006.
10. Sanchez-Aloso, S . & Sicilia, M.-A. (2005). Normative Specifications of Learning Objects and Learning Processes: Towards Higher Levels of Automation in Standardized e-Learning, International Journal of Instructional Technology and Distance Learning. Vol 2. No. 3., ISSN 1550-6908.
11. Dolog, P., Schaefer, M. (2005). A Framework for Browsing and Manipulating and Maintaining Interoperable Learner Profiles, Proc. of UM2005 - 10th International Conference on User Modeling, July, 2005, Edinburgh, UK. Springer Verlag.
12. Woelk, D. (2002). e-Learning, Semantic Web Services and Competency Ontologies,

Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002

13. Posea, V. & Harzallah, M. (2004). Building an Ontology of Competencies. Proceeding of Workshop on Ontology and Enterprise Modelling: Ingredients for Interoperability. In Conjunction with 5th International Conference on Practical Aspects of Knowledge Management. December 2, 2004. Vienna, Austria.
14. Ehrig, M., Haase, P., Hefke, M., Stojannovic, M. (2004). Similarity for ontologies – a comprehensive framework. Proceedings of the Workshop on Ontology and Enterprise Modelling: Ingredients for Interoperability, Vienna 2004. 13 – 23.
15. Maedche, A. & Staab, S. (2002). Measuring Similarity between Ontologies. Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW). Springer.
16. Deerwester, S., Dumais, S.T., Furnas, G. W., Landauer, T. and Harshman, R. (1990).Indexing by latent semantic analysis. Journal of the American Society for Information Science, vol. 41, 391-407.
17. Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, vol. 104, no. 2, 211-240.
18. St. Pierre, M. & LaPlant, W.P. (1998). Issues in Crosswalking Content Metadata Standards, National Information Standards Organization (NISO), Retrieved 1st of July, 2006 from http://www.niso.org/press/whitepapers/crsswalk.html

Publication of Distributed Linked Content

Paul Libbrecht and Christopher Brooks
paul@(activemath.org|dfki.de), cab938@mail.usask.ca

DFKI GmbH, University of Saskatchewan

Abstract. The goal of this work is to investigate how content for e-learning can be authored and published in a distributed fashion: this includes the management of links between content pieces and their meta-data items as well as the evolution of both of these items. We investigate the notion of *derivative works*, which appears to be an alternative to verbatim or copy-and-paste reusability by overlaying local content over remote content. Our investigations take place under the light of the authoring experience gained with the ActiveMath [3] and iHelp Courses [4] learning environments, and we propose solutions to work with these systems.

Reusability is not measurable in our system as it is based on the copy and paste method that can be applied on various levels – learning unit, material, content block and index. [1]

1 Introduction

The current practice in the authoring of standards-based e-learning material relies on publishing content inside of a learning content management system (LCMS) [2]This content is generally created with third party tools (e.g. Macromedia Dreamweaver, Microsoft PowerPoint, etc.) and aggregated together into IMS content packages [5], or LCMS-specific structures. Publication thus typically relies on the fusion of the content-storage and learning-environment roles. For an instructor intending to publish learning material to his students, all of the resources of a planned learning content need to be assembled and uploaded in the LCMS. This practice is problematic since LCMSs see learning content collections as isolated pieces of content whereas the instructor has likely assembled it by reusing other resources. This in turn leads to unmanaged duplication, a form of reuse which does little to enhance the quality of development of the original resources.

Provided that a reference can be made within the content, ActiveMath and OMDoc [7], and newer versions of IMS content packaging allow transparently the re-use of learning content items by the inclusion of a reference in table-of-contents. This is a first form of reuse which we call *verbatim reuse*.

Very often, however, verbatim reuse like this, even of fragments, is not realistic. Often authors of content wish to either make very small changes to content to adapt it to their needs. For this reason, copy-and-paste has been considered as the sole possible mechanism for reuse within many systems (e.g. [1]). Relying on this paradigm can, however, only lead to proliferation and duplication whereas it could be possible to repurpose while keeping a reference to the original version. One could draw a parallel to the issues associated with the pass-by-value and pass-by-reference methods for procedure calls as well as to the ability to override variables and methods of subclasses of object-oriented programming..

Reuse can also be made possible through *external linking*. External linking in web-based educational systems (as well as many other domains) has a number of detriments, the most profound being that the content at the end of external links are isolated and outside control of the link author, so changes to it cannot be easily reviewed before it is made available to learners within the learning environment. Issues of reliability and availability of content are also important.

To get around all the problems of links, it is common to informally include someone else's content, for example using copy-and-paste. Instead of copying, we prefer a slightly more general action of reuse which has been coined by Erik Gevers[1] as *appropriation*: the action of taking another piece of content and incorporating parts of it while still making modifications on it. An equivalent formulation would be the notion of *derived work* where the content that is appropriated is changed, filtered, and repurposed.

The wish to keep a link to the original work upon which a derivation is built allows for better management since it may enable the author of the derived work to incorporate further enhancements that the original author is doing or the author of the original work to track usage of his content. It also allows the author to situate the derived work as an extension or patched version of the original work. This is very appropriate to enable smaller, quality-oriented, adaptations to a piece of content or to its metadata which can be published and shared. We further note that modification to a learning object can be broken down into two kinds of changes: changes in the physical content of the learning object (e.g. words, sentences, graphics), and changes to the metadata which describes that learning object (e.g. where it is useful, what it is used for, etc.).

The goal of our paper is to initiate the study of appropriations as part of an authoring workflow, to demonstrate its importance for realistic reuse, and explore possible implementations within our learning environment.

1.1 Outline

Our paper is structured as follows: we first describe our current experiences in content publication within the ActiveMath and iHelp Courses environments. We then state approaches that provide partial answers to these wishes. Finally, we describe the steps we are taking to solve this within each learning environment and indicate open research questions.

[1] Personal communication, 2004.

2 Authoring in Learning Environments

2.1 Publishing in ActiveMath

The ActiveMath learning environment [3] is an intelligent web-based learning environment which serves mathematical content to learners and supports the learning experience by user-modelling, pedagogical advice and interactive exercises, both based on the knowledge structure of the content, an extension of the OMDoc language [7].

ActiveMath delivers its content based a serving-interface called MBaseRef which allows individual fragments (the content of a definition or an exercise for example) to be listed, to be extracted, and relations to be queried in both ways: e.g. the relation that an exercise trains a given concept can be queried from the concept or from the exercise. The requirement of ActiveMath to support a classroom delivery implies a requirement of high performance for this content-serving-interface, estimated at about 300 queries per second.

The content in ActiveMath is organized into content collections where each collection contain all OMDoc files and their supporting resources (e.g. pictures) and define the behavior of ActiveMath on them: for example it provides the pointers to the table of contents of books to be presented. Content collections are somewhat similar to an IMS content package [5] but are not manipulated as a set of resources to be directly served to the learner's browser. Instead, the content is processed through a presentation layer with, for example, consistent mathematical notations. Finally, metadata describing the elements of the content form the basis of knowledge for learner modeling.

Content authoring is done in a text editor using a readable syntax, see [6]. Loading the content in the ActiveMath content-store is done using a publish procedure in the authoring tool which also validates all references. This validation procedure has turned out to be a precious help to support the authors in their task which include the input of many relations: the inclusion of an item within a book, the mathematical relationships between such items as an exercise and the conceptual items it trains, the hyperlinks, and the usage of mathematical symbols. Currently, however, the system can only validate references to content that is loaded in the local content-store.

The first authors' PhD project is the design, implementation, provision and experimentation of authoring and storage tools for the ActiveMath learning environment. It addresses the study of authoring semantic mathematical documents (e.g. with OMDoc [7]), the investigation on how to support authors best by error-reporting feedback and suggestions, and the publication mechanisms to promote the collaboration between authors as well as the reusability promised by the semantic nature of the content. The first facet is reported about in [6], the last facet is the focus of our article.

The approach proposed below intends to make the content-serving-interface of Activemath able to serve efficiently content from remote sources, possibly with modifications typical of an appropriation.

2.2 Publishing in iHelp Courses

iHelp Courses [4] is an open source web-based learning object content management system deployed within the Department of Computer Science at the University of Saskatchewan. It is used to deliver some of the introductory courses in computing concepts for non-majors and is available for learners in both a blended environment and a purely online environment.

This system aims to be as standard compliant as is possible in a research environment, and delivers content packaged using the IMS Content Packaging format [5] and the IMS Question Test Interoperability (QTI) framework [8]. Similar to other LCMSs, content can be in any format renderable by a browser (including Flash, Java Applets, etc.). The flexibility of the format allows an instructor to take existing learning materials and readily integrate them into a new course.

Authoring for this system consists of three parts. First, the author collects the materials they would like deliver together, and modifies them to fit their intended audience. This step can include taking resources off of the web, out of a learning object repository, or creating them from scratch, and is typically done with an (X)HTML editor. Once collected and edited, the resources are sequenced into activity hierarchies using a SCORM or IMS Content Packaging compliant editor such as RELOAD[2]. This is saved as a single archived file and uploaded to a particular course within iHelp Courses. The final step in the authoring process is the setting of rules based upon the IMS Simple Sequencing [9] which restrict the learning workflow for the student based on the roles they may have (e.g. advanced student, tutorial assistant, regular student, etc.).

The second authors' recently-finished MSc project was the investigation of how version control can be managed in a decentralized authoring environment like this, where authoring is a step in a process instead of a feature in a content management system. Particular care was taken in trying to preserve the history of content evolution while taking into consideration the state-of-the art in e-learning standardization. The result of this project was a prototype tool for maintaining a history of syntactical changes to learning material, and a simple visualization environment for seeing the interdependencies between versioned learning objects.

3 Related Works

We draw inspiration for distributed authoring of learning objects from the area of software configuration management (SCM), where versioning issues amongst software developers are a well studied problem. We note that there are some differences in the workflow and goals of educational content authors when compared to software programmers. Whereas programmers tend to work in highly coupled groups with one end goal, educational content authors are often widely distributed, work asynchronously, and may have individual teaching goals that

[2] See `http://www.reload.ac.uk/`

are not common. Further, learning content tends to be stored in a disparate array of repositories while software tends to be managed centrally.

A quick survey of current versioning tools show that very little is done with them in order to decentralize the repositories (e.g. CVS[3] and Subversion[4]. More elaborate approaches exist such as GNU Arch, Bazaar-NG,[5] and SVK[6] which allow cascaded repositories whereby a repository can be derived of another repository still follow a logically and administratively centralized model and changes committed to one physical repository propagate to others in the system. The appropriation approach described in section 1 aims at reuse and can be realized without an underlying versioning system by considering only the relation and differences between the original and derived works. Of course, an elaborate implementation of this approach could make use of the versioning systems as remote content storage.

Current learning object repositories have, at best, limited versioning capabilities. The majority of these repositories (e.g. Merlot[7]) serve as metadata catalogues only, with links to content stored on author websites. Changes to content can result in invalidation of metadata, and metadata is thus only used for high level description of the content that is being referenced. The repositories that store the content as well as metadata such as EducaNext[8] tend to work on a "most recent commit" model, where the only version available is the latest one uploaded. Thus each version of a learning object is seen as a revision of the previous version, and there is no value given to maintaining separate versions of a given work whereas this answers precisely a need of content editors.

3.1 Vocabulary for Versioning

Core to the notion of distributed authoring is the issue of versioning. Versioning, as studied for software artifacts, is governed by three principles:

Sameness principle Versioning requires a method by which two artifacts can be examined to see if they are of the same version [10]. Being syntactically the same is often not enough to identify that two artifacts are of the same version; sometimes the ancestry of the artifact is important in determining the context that it will be used in. In addition to this, syntactic comparison is often unachievable because it requires knowledge of the syntax being used.

---

3 Concurrent Version System (CVS), see `http://cvshome.org/`

4 Subversion, see `http://subversion.tigris.org/` only support single centralized repositories)

5 Bazaar NG is a descendant of GNU Arch, for both see `http://bazaar-vcs.org/RcsComparisons`

6 SVK is an extension of Subversion to support decentralized operations, see `http://svk.elixus.org/`

7 Merlot is one of the large learning objects repositories. See `http://merlot.org/`.

8 EducaNext is one of the large learning object repositories. See `http://educanext.org/`.

Immutability of artifacts Artifacts in a versioning model are considered immutable, in the way that changes to an artifact force the creation of new version of that artifact. This ensures that the history and traceability of the artifact is never lost [11].

Unique identification Given the sameness and immutability principles, there must be some way of identifying each artifact, and each version of each artifact, uniquely. This is usually implemented by giving each software artifact a unique object identifier, and then giving each version of that artifact a unique version identifier [10].

Versions that are meant to replace previous versions are called *revisions*, while versions that are meant to coexist with previous versions are called *variants* [10]. The set of all revisions and variants of an artifact is known as a *revision group*. The reason for changes in an artifact is domain dependant; Conradi and Westfechtel cite bug fixes, extending functionality, and changes in dependencies as common reasons for creating new versions when dealing with software artifacts [10], while Noy and Klein identify changes in a domain, conceptualization of a domain, or specification of a domain as reasons for change in ontology based knowledge representation systems [12]. Our anecdotal observations of learning object authors suggest that teaching style and curriculum changes tend to drive development, though external influences, such as tools being used to experiment with a particular subject, can also lead to the need for revisions.

3.2 E-Learning Standards for Versioning

Current metadata standards have little or no support for the versioning of learning objects. The Dublin Core contains only two useful elements, the source element and the relation element. The source element identifies the resource that the new learning object was derived from, using some uniquely identifying index (usually in the form of a Uniformed Resource Identifier (URI)). There are no options to further annotate this element to capture how the current learning object differs from the previous one. The relation element allows for a more general form of expressing the relationships between two resources. It allows for a resource to indicate that the current learning object is either a variant of another learning object, or a revision of another learning object by using the modifiers *IsVersionOf* and Replaces respectively. Two other modifiers, *HasVersion* and *IsReplacedBy* allow original learning objects to identify the same relationship with respect to their derivative works.

The creation of the LOM was influenced by a number of different metadata specifications, and thus it contains many mechanisms that are similar in name and function to those in the Dublin Core. One of these mechanisms is the relation element that allows authors to describe relations using a simple vocabulary. This vocabulary corresponds closely to that of the Dublin Core, and allows for using the *hasversion* and *isversionof* keywords. The relation element also allows for identifying a derivative object similar to the source element, by using the *isbasedon* and *isbasisfor* modifiers keywords.

Also included in the LOM is the lifecycle category, which contains a number of data elements that describe the history of a learning object. The version element describes the edition of the object using human readable plain text, while the status element describes the completion status of the object using a small, predefined vocabulary (draft, final, revised, or unavailable). The contribute element identifies what entities have influenced the history of this learning object (such as other authors, organizations, etc).

While both the Dublin Core and the LOM provide solid foundations for describing a simple revision tree, they fail to offer full support of a version model. Indeed, in an analysis of the LOM done by CanCore, it is stated as only generally addressing the concerns of versioning, and even then it does "not do so in a way that is sufficient for the requirements of many projects" [21]. Instructional designers or software components who are interested in determining the significance of a given change to a learning object are forced to obtain versions of both learning objects, and attempt to reason about the metadata. Comparing metadata records is not a trivial task – the majority of elements in both the Dublin Core and the LOM are meant for human consumption and would be difficult or impossible for a software agent to compare with faithful semantics. Vocabularies are often left open ended, and force implementers to either come up with their own terminology or adopt a given application profile. This leads to data that is difficult even for intelligent software components to understand when trying to dynamically obtain and compose learning objects into larger educational units.

3.3 Decentralized Networks

The requirement for content to be distributed through several network nodes is an obvious scalability requirement. Such scalability issues are already faced currently in the LeActiveMath project by the current content store used for ActiveMath. Another objective for *decentralized* network is to allow each authoring project to use one's own content and other's content without the load or visibility of a single server (such as in the Wikipedia project).

An area exploring decentralized collaboration is the area of peer-to-peer networks. In [13], a scenario for a personalized learning environment with content distributed through peers is proposed based on RDF-queries for content extraction. Similarly, the POOL, POND, SPLASH environment [14] implements a hierarchal peer network of learning object repositories, where desktop repositories are connected to one another through institutional repositories, which are connected to one another through regional repositories.

4 Managing Learning Object Authoring

We outline here two different approaches to handling version control issues in distributed learning object authoring environments. The first approach is to be used by ActiveMath. It requires the content storage interface to be converted partially to a proxy to include external content repositories as source. The second was developed to be compatible with standards-based content repositories, and is being considered as an approach for iHelp Courses. While it this time it has only been prototyped, it allows for loosely coupled and highly decentralized development of content and makes use of the growing number of learning standards.

4.1 Proposed Approach for ActiveMath Authoring

We have indicated that the ActiveMath learning environment *reads* authored content through a content-delivery-interface which serves fragments and relations. We propose to provide implementation of this interface that allow remote content stored on a simple web-server to serve as source and an implementation of this interface which realizes a *content filter* by giving precedence to local content to answer queries before it queries the remote content. This is pictured in figure 1 where the learning environment is a consumer of three content collections, two of which are *filtered* so as to be adapted for an author preparing next year's course based on shared content and the course material of last year. This picture is an architectural vision for a single author (or author team) and could be reproduced for each learning environment, keeping shared content repositories centralized.

Link validation is to be coordinated by the local content store which will be configured appropriately so that the relations from items of the derived content of the shared repository can be published and do not produce reference errors once there. Between published repositories, links notification methods should be used (e.g. as is done by the TrackBack protocol [15]) which do work between trusted repositories. Using the knowledge of these links allows an author to measure the impact of a change before uploading to a shared repository, for example by testing it using his own learner. Because changes in a shared repository can happen anytime and because the remote connections introduce a breakage possibility, caches need to be provided with management methods to allow complete cache fill-ups and control updates, existing tools and practice for such caching is common since HTTP caching is well defined and cared for in such recommendations as the *Architecture of the Web* [16].

Error reporting can go beyond simple link validation. When several authors work on a content collection, they need to be able to watch the overall evolution. Assertions on the content structures can be checked and reported about similarly to errors: for example, it is natural to assert that each concept in a given set, should have enough exercises for each targeted educational context so that the adaptive content selection can work, or to assert the absence of cycles in the *prerequisites* relationships.

The fact that the fulfillment of assertions or the fix for an error may be the task of another author in collaborative content authoring adds another facet to the error-reporting: errors should be *talked-about* which means they need to be exchanged and read by others, just as a Web-document would.

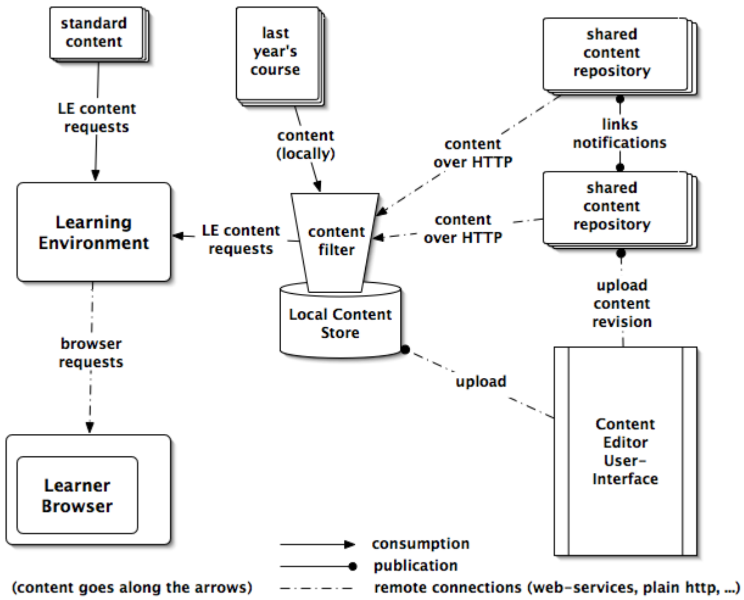

Fig. 1. Architecture overview of the learning environment as client of several repositories combined by the content filter.

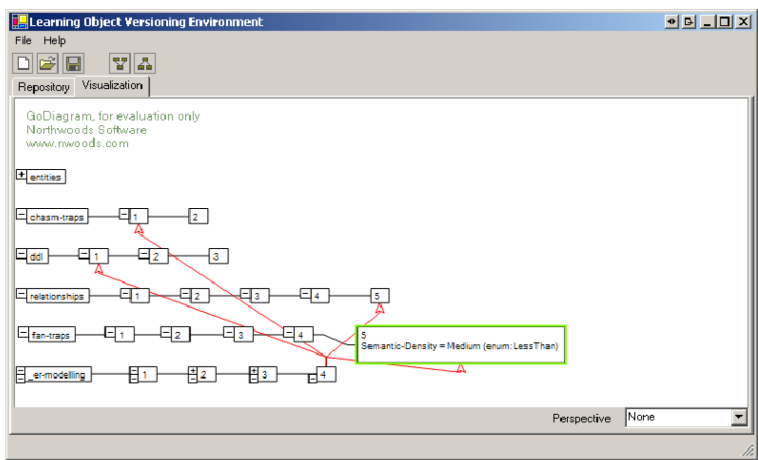

Fig. 2. A visualization of a learning object repository. The top five timelines indicate simple learning objects (e.g. content packages without sequencing) where each node is a revision of the previous node. The fifth timeline has been currently expanded by the author to show available metadata, while the sixth timeline represents a sequence of content packages (the arrows refer to specific version offsets in the referenced content packages).

4.2 Proposed Approach for iHelp Courses Authoring

Versioning for iHelp Courses has been prototyped at a very low level by the creation of a change-aware authoring environment. This tool, the Learning Object Versioning Environment (LOVE) [17], was developed as an integrated authoring environment where an instructional designer can build, sequence, and visualize learning materials. Version changes are coupled directly with the learning object they refer to, and are stored inside of the IMS Content Packaging manifest that describes that content. Changes to sequences of content are realized as tree manipulations on the IMS Simple Sequencing activity tree, and are also stored within a content package that describes the whole course. The coupling of content changes with the content package makes creating derivative works as simple as copying the original package and editing it. The environment also provides methods to control IEEE LOM-based metadata, as well as a means of showing the coupling between sequences of learning objects (figure 2).

While this method of versioning allows an author to retain the history of a learning object, it loses the link to the instance of that object which is currently deployed. This has two effects. Firstly, changes to the deployed learning object will not be propagated to the derivative work. This provides stability, but makes upgrading content (e.g. bug fixing) a tedious task. A notification framework similar to RSS with author-assisted merging is being considered to mitigate this. Second, new metadata about the deployed learning object is not propagated to copies of that object. One of the key beliefs of the second author is the notion that learning object metadata is a *process*, not just a data structure, and that the collection of usage-based metadata is imperative for creating intelligence systems. This belief is based upon McCalla's ecological approach [18], and greatly complicates the issue of version control. When metadata is automatically collected around a particular version of a work it is unclear what the implications are for earlier or later versions of that work.

5 Conclusions and Open Issues

As noted in [19], legal aspects play an important role to enable an appropriator to act. The work done in the Creative Commons project is leading the way by providing a small set of licenses with readable summaries (both by human and machine readable) which let an appropriator decide in one glance on his allowances to act. The right to *derive* is explicitly mentioned as a right that can be included in the original work's license or not.

Validation of content becomes much more important when it comes to distributed content. We have indicated defensive methods to use a cache to protect oneself against unwanted changes of the original work of an appropriation. A much more evolutionary attitude can be taken if trust to the original author and its repository can be granted by enabling stronger validation checks based on an authors' assertions about certain global qualities of the learning content. These qualities include requiring a consistency of changes (e.g. compatibility of differences), the verification that *non-contradictory* definitions occur for concepts, or that metadata do not contradict one another (within a given scope and for a given purpose). The experience reported about in [6] indicated that simple link validation is fundamental, and this supposes that authors may benefit from stronger validation methods and management of their results.

It is the firm belief of the authors that visualizations of content interconnections, both within a learning object repository as well as within a course delivery tool, will greatly aid in the usability and quality of learning content produced. Nonetheless, the production of these visualizations has been largely ignored by the educational technology community – where as visual differencing programs for plain texts or word-processing are common, both viewing of differences of metadata as well as viewing semantic differences as in [20], have not been explored.

We have presented approaches to publish content in an evolutive and distributed fashion where the initiative of a change lies in any individual that may edit content for it to be delivered to his intended recipients. It provides a realistic answer to the wishes of reusability provided the rights to derive are cleared and opens the door to the evolution of content collections by the shared contribution of multiple parties involved in their realization and use.

We hope to implement and investigate our implementation, based on existing authoring tools such as [6] or [17]. A common implementation may arise as a filter which brings together parts of several remote repositories with local modifications. The support for external linking available in newer versions of the IMS content packaging [5] will be relevant. Practices of actual reuse will be gathered following these implementations and reported about.

References

1. Milos Kravcik, Marcus Specht, and Reinhard Oppermann, *Evaluation of WINDS Authoring Environment*, Proceedings of Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2004, Eindhoven, LNCS 3137, 2004.
2. Mark Power, *Interoperability in Action*, a video introduction of the RELOAD project, Technical Report, CETIS Group, 2004, see `http://www.cetis.ac.uk/content2/20041102022838`
3. Erica Melis, Eric Andrès, J. Büdenbender, Adrian Frischauf, George Goguadze, Paul Libbrecht, Martin Pollet, and Carsten Ullrich. *ActiveMath: A Generic and Adaptive Web-Based Learning Environment.* International Journal of Artificial Intelligence in Education, 12(4):385–407, 2001.
4. C. Brooks, L. Kettel and C. Hansen. (2005) *Building a Learning Object Content Management System*, World Conference on E-Learning in Corporate, Healthcare, & Higher Education (E-Learn 2005)
5. *IMS Content Packaging Information Model*, Version 1.1.4, IMS Global Learning Consortium Inc., 2003.
6. P. Libbrecht and C. Gross, *Experience Report Writing LEActiveMath Calculus*, Proceedings of Mathematical Knowledge Management 2006, LNCS 4061.
7. Michael Kohlhase, *OMDoc: Towards an OpenMath representation of mathematical documents*, version 1.1, Technical report, Universität des Saarlandes, 2000. Available at `http://www.mathweb.org/omdoc/`.

8. IMS Global Learning Consortium Inc., *IMS Question and Test Interoperability*, version 1.2 , 2003.
9. IMS Global Learning Consortium Inc., *IMS Simple Sequencing Information and Behavior Model*, Version 1.2, , 2003.
10. Bernhard Westfechtel and Reidar Conradi. *Software configuration management and engineering data management: Differences and similarities*. In Boris Magnusson, editor, Software Configuration Management – ECOOP'98 SCM8 Symposium, pages 95–106, Brussels, Belgium, 20-21 July 1998. Springer Verlag LNCS 1439.
11. L. Bendix, Antonina Dattolo, and Fabio Vitali. *Software configurationmanagement in hypermedia engineering*. Handbook of Software Engineering and Knowledge Engineering, 1, 2001.
12. N. F. Noy and M. Klein, *Ontology Evolution: Not the Same as Schema Evolution*. Knowledge and Information Systems, 6(4), pp. 428-440.
13. Peter Dolog, Nicola Henze, Wolfgang Nejdl, and Michael Sintek: *Personalization in Distributed eLearning Environments*. In Proc. of WWW2004 - The Thirteen International World Wide Web Conference, New York, USA, 2004. ACM.
14. Richards, G. and Hatala, M. (2002). *POOL, POND and SPLASH - A Peer to Peer Architecture for Learning Object Repositories. Internet 2 Conference*, Workshop on Collaborative Computing in Higher Education: Peer-to-Peer and Beyond.
15. Six Apart SA, *TrackBack Technical Specification, Revision 1.2*, accessible at `http://www.sixapart.com/pronet/docs/trackback_spec`.
16. Ian Jacobs and Norman Walsh, *Architecture of the World Wide Web, Volume One*, W3C recommendation, December 2004. Available at `http://www.w3.org/TR/webarch/`.
17. C. Brooks: LOVE: C. Brooks. *Supporting Learning Object Versioning*. Master's thesis, Department of Computer Science, University of Saskatchewan. 2005.
18. McCalla, G, *The ecological approach to the design of e-learning environments: purpose-based capture and use of information about learners*, In T. Anderson and D. Whitelock (Guest Eds), Journal of Interactive Media in Education (JIME), Special Issue on The Educational Semantic Web, Vol. 1, p.18, 2004, Available at: `http://www-jime.open.ac.uk/2004/1`.
19. Reusable Learning Project, (Robby Robson, Geoff Collier, Brandon Muramatsu), *Reusability Framework*, see `http://www.reusablelearning.org/`.
20. Svetlana Radzevich, *Semantic-based Diff, Patch and Merge for XML-Documents*, Masters Thesis, Univeristy of Saarland, 2006.
21. CanCore Initiative. CanCore Learning Object Metadata: Metadata Guidelines, Version 1.1, 2002.

Semantically-enabled Model Driven Course Composition

Mark Melia and Claus Pahl

School of Computing, Dublin City University,
Dublin 9, Ireland
mmelia@computing.dcu.ie

Abstract. Modeling is an essential tool in the development of complex software. In recent times models have gained importance becoming actual development artifacts, replacing traditional code. Here, we discuss the use of models in the development of courseware. Course authoring is a time consuming and costly activity. A model based approach to course development may prove to alleviate some of its' associated costs and improve the speed at which courses are developed. The use of semantics embedded in models is also investigated. Ontologies allow for the formal expression of semantic information. This information can be used to validate models and even provide the course developer with fixes for invalid models.

1 Introduction

The "reuse rather than create" paradigm has been successfully applied in the software engineering sector. The advent of new software engineering frameworks such as component-based software and service-orientated architecture cuts the time and cost involved in creating complex software systems through reuse. By providing ready-made solutions to common business problems, developers can integrate services or components to provide a solution for larger, more complex business problems.

Learning technology systems have recently seen a surge in the use of this paradigm. Course developers are looking to reuse and re-purpose learning material created by others to cut course development time and costs. With the advent of Learning Object Repositories (LORs) this style of course development is becoming ever more popular.

We propose using a type of Model Driven Development (MDD) to encourage the course creator to consider aspects of pedagogy, such as how to sequence learning content when creating a course. This is done by creating a visual environment for setting course sequencing and navigational strategies. In software engineering, MDD raises the level of abstraction software developers develop at, by using precise, platform-independent models as software development artifacts [2]. Semantic Web technologies [3] combined with MDD technologies are also used to allow for automated selection/recommendation of learning objects, and validation of course models, reducing the course creator's workload.

In this paper, we propose an intuitive authoring framework based on modeling, which allows the course creator to compose courses and specify sequencing behaviour. Our framework also aids the course creator in locating, selecting and validating LOs. Our approach is complimentary to other Learning Technology movements, such as LORs, Learning Management Systems (LMSs) and the various learning interoperability standards.

2 Project Objectives

This project consists of four main objectives, these are:

- Define an intuitive modeling language to set out a course conceptual structure which provides the course creator with a simple way to embed learning objects into the course. This modeling language is also used to specify sequencing of LOs and course components.
- Map models to learning interoperability standards using MDD technologies.
- We assume the uptake of semantically enabled LORs, containing LOs, which are semantically annotated. These semantics enable the validation of our models, ensuring that all LOs are suitable within the context they are placed. Semantic Web technologies, particularly ontology technology, enable this functionality. We also investigate parallels between LO composition and software component/service composition using Semantic Web technologies.
- Semantic knowledge contained within a course model graph is used to locate suitable content in semantically enabled LORs.

3 Rational, Benefits and Innovation

Learning Object Repositories are becoming the norm in course development. We see this with the growing popularity of repositories such as MERLOT [13], which at the time of writing has over 36,512 members, with nearly 1,000 new members added in the previous month. In terms of reusable material available to users MERLOT has 14,791 individual LOs at the time of writing with over 150 new ones being added in the previous month.

With Learning Object Repositories becoming proliferate, we believe course developers will reuse and re-purpose learning material available to them through repositories much more and only create original course content when no suitable existing content is available. This will cut the time and cost involved in creating courseware. This new approach to course development means the course developer will now have to invest time in locating course content, evaluating it and integrating it with the course. In this paper, we investigate a model-based infrastructure for these tasks.

Our project sets out to alleviate the problems associated with reusing LOs by providing an intuitive model-based environment to create courses. Our course development approach also uses semantics in LO annotation to provide for more sophisticated LO composition, discovery and validation.

Courses can be modeled both conceptually (in a hierarchical diagram) and sequentially (in a flow diagram). The conceptual diagram outlines the course structure, while the sequence diagram allows the course creator to specify LO sequencing behaviour.

Completed courses can be validated using an ontology describing the learning process in a similar way software compositions can be validated using software process ontologies [4]. Validation ensures that each learning object have been placed in the course in a pedagogically sound manner.

Should a model be deemed invalid, the semantics contained within the models provide sufficient information to allow for the automated discovery of LOs, which when added to the model will validate the model. For example, if a knowledge gap is found in the sequencing, where the learner will not have the necessary pre-requisite knowledge for a desirable LO, the semantics within the models can be used to search a LOR for a LO which will provide the learner with the necessary knowledge in order to fulfill the pre-requisite requirements of the desired LO.

4 Research Methodology

The ultimate goal of this project is the development of an intuitive model-based course engineering framework allowing for the development, validation and correction of courses.

We firstly investigate similar frameworks, such as those found in the software engineering domain. We investigate what aspects of the approaches found in similar frameworks can be applied to course development.

We then investigate modeling notation, which allows the course creator to intuitively manipulate a course and can be translated into a learning interoperability standard, such as SCORM [5]. There are three possible ways for doing this:

- Develop a UML profile which is intuitive enough that it will gain acceptance from course developers. This would be the simplest solution as an XML binding for UML is available (XMI) and can be mapped to learning technology standards. UML has been used to model course creation methodologies before, but only at a superficial level [6].
- Look at how future standards, which are not based on software engineering principles, such as the Business Process Modeling Notation (BPMN) [7] could be used to model course composition, and investigate possible XML bindings.
- Investigate the creation of a new modeling notation for course composition, although this would be the most difficult it may be necessary in order to gain acceptance from course developers.

As a proof of concept, we will extend an existing tool, such as Reload [8] with the new model-based course composition framework, and test it with course developers. Semantic Web technologies that originate from the software engineering domain will be adapted to validate the course models. This will allow for the validation of each LO in the course and ensure it is in harmony with the course as a whole. One possible method for doing this is to adapt existing infrastructure used for the automated composition of software services using process modeling ontologies [4]. Our framework will also recommend solutions for an invalid course using semantic information captured within the modeling notation to automatically generate LOR queries. We hope this will minimise manual work carried out by the course creator.

5 Related Work

5.1 The Reload Editor

The reload project is a JISC funded project, managed by the University of Bolton [8]. The project aims are as follows:

- to facilitate the course creator in sharing learning objects
- to allow the course creator to encapsulate pedagogical information in lesson plans.

The Reload editor allows the course creator to create a SCORM package at a slightly higher level of abstraction, abstracting such things as the XML implementation details of a SCORM package. Despite this, course developers still require an in-depth knowledge of the principles of the SCORM standard when using Reload.

We see our framework as the perfect companion to Reload. Reload allows course developers to create courses, and to add sequencing information, but in a less intuitive way than specified here. Adding our framework as a plug-in to Reload would empower the less technical course creator to develop courses that make use of explicit LO sequencing.

5.2 My Online Teacher (MOT)

The "My Online Teacher" (MOT) system [9], developed at the Technical University of Eindhoven, allows course developers to create adaptive courses using the LAOS system of layered models [10].

Similarities between our approach and that of LAOS can be noted. LAOS contains a domain map, which is also implicit in our approach as LOs pre-requisites and postrequisites reference a domain model in order to share a common semantic space. On top of the domain map in LAOS a goal and constraints map contains pedagogical information. In our approach the author is in effect setting a goal and constraint map when he or she set out sequencing paths and sequencing strategies using models.

5.3 Sequencing Objects

Su et. al. recognised the steep learning curve faced by the course creator in SCORM 2004 compliant courses, and suggests using an Object Orientated Methodology (OOM), to simplify the creation of SCORM compliant courses [11]. Using the OOM, complicated sequencing rules are encapsulated into objects, known as a sequencing object.

The course creator then adds learning content to sequencing objects, and joins the sequencing objects together to create a course. This course is then translated into a SCORM 2004 package using the "Course Sequencing to Content Package" (CS2CP) algorithm.

Sequencing objects are excellent examples of how models can be used to elevate some of the complexity involved in creating SCORM sequencing. Sequencing objects use a self-defined notation, and are not standardised, and therefore do not take advantage of the automated transformation processes developed for standardized modeling languages such as the UML.

5.4 Visual Online Authoring Tool (VOAT)

Yang et. al. identifies difficulties with reusing and repurposing LOs in [12], and has created a suite of tools which aim to aid the course creator in the creation of courseware.

There are three components of Yang's course creation tool suite, Ontology-Based Outline Authoring Tool (OBOAT), the Visual Online Course Authoring Tool (VOCAT) and the Visualized Online Simple Sequencing Authoring Tool (VOSSAT). The course creator uses OBOAT to create a conceptual map of the content of the course to be created. The construction of the conceptual map is guided by a domain ontology.

Once a conceptual map has been created it is loaded into the VOCAT. VOCAT is used to add and assemble Learning Objects and Content Packages to the course.

The final tool in the suite is the VOSSAT, which allows the course creator to specify LO sequencing information. Sequencing information is specified through a form based user interface. Once sequenced, courses are exported using SCORM and can then be used by learners in a SCORM-compliant LMS.

Yang et. al.'s approach is heavily reliant on the ADL SCORM [5] standard and cannot be used with other standards. The tool also assumes theoretical knowledge of SCORM introducing an additional barrier to course developer who wishes to use the tool.

6 Conclusion

The main goal of our work is to aid the course creator in the course creation process. We do this by providing the course creator with an intuitive environment in which to allocate and sequence LOs in courseware. We also provide a way for the course creator to check that a course adheres to basic pedagogical principles. Courses are validated based on these principles. If a course has been deemed invalid, the framework can recommend a remedy to validate the course. Once the course is valid it can be exported using one of the learning technology interoperability standards, and delivered on a standards-compliant LMS.

References

1. Wiley, D.A.: The Instructional use of Learning Object. Chapter, Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor and taxonomy. Association for Educational Communication and Technology, 2001
2. Frankel, D.S.: Model Driven Architecture. Wiley Publications, Indianapolis, Indiana (2003)

3. Daconta, M.C.: Obrst, L.J., Smith, K.T.: The Semantic Web: A guide to the future of XML, Web Services and Knowledge Managements. Wiley Publications, Indianapolis Indiana, (2003)
4. Pahl, C.: Ontology Transformation and Reasoning for Model-Driven Architecture. In: Proceedings of the International Conference on Ontologies, Databases and Applications of Semantics (ODBASE05), Springer-Verlag LNCS Series (2005)
1170-1187
5. Advanced Distributed Learning: SCORM 2004 Overview (2004) Available from: http://www.adlnet.gov/scorm/index.cfm
6. Hu, S.C.: Applications of the UML in modeling SCORM-conformant Contents. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT05), IEEE Computer Society (2003).
7. White, S.A.: Introduction to BPMN: Technical Report (2004)
8. The RELOAD Project: The RELOAD Metadata and Content Packaging Editor: Available from http://www.reload.ac.uk/editor.html
9. Cristea, A.I., Smits, D., deBra, P.: Writing MOT, Reading AHA! – converting between an authoring and a delivery system for adaptive educational hypermedia. In proceedings of the Third International Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia at AIED05. (2005)
10. Cristea, A.I., de Mooij, A.: LAOS : Layered WWW AHS Authoring Model and their corresponding Algebraic Operators : In proceedings of the Twelfth International World Wide Web Conference (WWW03), Alternate track on Education, ACM, (2003)
11. Su, J.M., Tseng, S.S., Weng, J.F., Chen, K.T., Liu, Y.L., Tsai, Y.T.: An Object Based Authoring Tool for Creating SCORM Compliant Course: In proceeding of the Nineteenth International Conference on Advanced Information Networking and Applications. Vol. 1. IEEE (2005) 209-214
12. Yang, J.T.D., Chen, W.C., Tsai, C.Y., Chao, M.S.: An Ontological Model for SCORM-Compliant Authoring Tools. Journal of Information Science and Engineering **21**(5). (2005) 891-909.
13. The MERLOT Project: http://www.merlot.org

Modularization of Existing Learning Resources for Repurposing

Marek Meyer

SAP Research CEC Darmstadt
Bleichstr. 8, 64283 Darmstadt, Germany
marek.meyer@sap.com

Ph.D. Advisor: Prof. Dr.-Ing. Ralf Steinmetz

Abstract. Reusability of Learning Resources often fails because the existing Learning Resources are not available at the proper level of granularity. Hence, an ex post modularization is necessary to nevertheless re-use parts of these Learning Resources. The major research question, which I want to address in my Ph.D. work, is how modularization of existing Learning Resources for different re-use purposes can be performed efficiently and in a user-friendly way.

1 Background & Motivation

E-Learning has gained an important role in education, and especially in professional learning. Web-Based Trainings (WBT) allow learning at home or at ones desk. But although E-Learning has been used for some years and is a focus of many research projects, repurposing of E-Learning courses is still a challenge. Today, re-use is mainly thought of as re-use “as is”, which means that a course is used again for another learner or group of learners, but in its original, unchanged form. However, reusing complete courses unchanged is often insufficient. If a learner wants to learn only a part of a whole course, he does not want to seek within a course those chapters which are relevant. Also, an author sometimes wants to re-use parts of one course for creating another course – either with another topical focus or for another target group and hence using different didactics. In these cases, having modular content available at a finer level of granularity would significantly improve the situation. Production of fine-grained, modular content is often demanded. But in practice authors do not obey for different reasons: planning and authoring of modular content takes more time and thus produces higher costs; and secondly deadlines for content delivery are usually short. As a result, creation of modular content is a nice concept in theory, but does not work satisfactory in practice in many cases.
If content is though supposed to be re-used in the described manner, it has to be modularized ex post. Monolithic Learning Resources have to be transformed into smaller, reusable modules. The process of transforming a Learning Resource into one

or more smaller Learning Resources for re-use is called *modularization*. In my Ph.D. work, I focus on methods for modularization of existing Learning Resources for subsequent repurposing.

2 Identification of Problems in the Field of Research

In order to support modularization of Learning Resources, several problems have to be solved. The most important challenges identified so far are:

- Granularity of Learning Resource fragments
- Identifying reusable Learning Resource fragments
- Requirements on modules for subsequent aggregation
- Transformation of suitable fragments into reusable modules
- Retrieval of suitable Learning Resources

The first challenge is to identify the proper granularity of Learning Resource fragments that are suitable for repurposing. A fragment is every part of a Learning Resource, which can be clearly separated from the rest of the Learning Resource; Learning Resource fragments are also called *content fragments* or simply *fragments* in this paper. Therefore everything from a single sentence over chapters up to the whole course can be regarded as a fragment. The question, which range of fragment sizes is suited best for repurposing, is still open. But as authoring of larger courses by aggregation is a possible usage of the resulting modules, the module granularity may be smaller than that of a traditional Learning Object.

The second challenge is to identify those fragments within a Learning Resource, which are suitable for a particular purpose. What is suitable for re-use depends on the user’s purpose. Reusable fragments may be determined by different criteria such as topic, didactical functions or involved media types.

The modularization process shall result in modular Learning Resources, which can be re-used immediately or for aggregation into larger courses. Therefore, it is required to analyze and specify the requirements on modular Learning Resources. These requirements comprise technical aspects (formats), appearance (layout and design), didactic aspects, and possibly legal aspects as well.

When reusable fragments have been determined and the requirements on modular content are known, the fragments have to be transformed into modular Learning Resources which are suitable for aggregation. The challenges here are not only the technical decomposition but also include producing clean modules: There should not remain void references to unavailable parts of the original Learning Resource, neither technical (e.g. links in HTML), nor textual references. Also, depending on the chosen granularity, some texts like an introduction, summary or bridging texts have to be adapted.

And finally, retrieval of suitable Learning Resources remains a challenge. Sometimes, already modularized content might be available in a learning object repository. But more often the user has to find Learning Resources, from which he may re-use parts by modularizing it. Thus, it is necessary to find Learning Resources not only by their overall properties, but also by properties of individual reusable parts.

3 Overall Approach

For solving the research question, the whole modularization is divided into smaller issues. The modularization of a Learning Resource can be modeled as a process, which consists of several consecutive process steps. By splitting the process into smaller steps, the requirements on the whole modularization process can be assigned to the different process steps.
The basic approach is to model a modularization process which makes use of supportive functionality by a repurposing framework. The framework shall be used as an abstraction layer to the Learning Resource contents: it provides facilitated access to the Learning Resources regarding structure, contents and semantics. The modularization process is arranged between retrieval and re-use (see Fig.1). A retrieval component is used for finding Learning Resources that contain suitable contents for re-use. After the modularization process, different re-use components may process the resulting modules. Possible re-use scenarios could be aggregation, rearrangement or adaptation of modules.

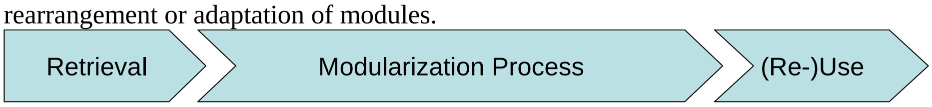

Fig. 1. Modularization process environment.

There are many different tasks which have to be performed for modularizing a Learning Resource. These tasks may also differ for different Learning Resource formats and re-use purposes. However, all relevant tasks shall be clustered into a smaller number of process steps for facilitating the planning, implementation and discussion of modularization methods. The process steps should be ordered, so that the results of one step are only required by successors, but not by predecessors. I have chosen to organize my modularization process in six process steps (see Fig. 2): they are consecutive; each of these steps has its own challenges; but each process step is defined precise enough to address and solve the corresponding problems. The six process steps are:

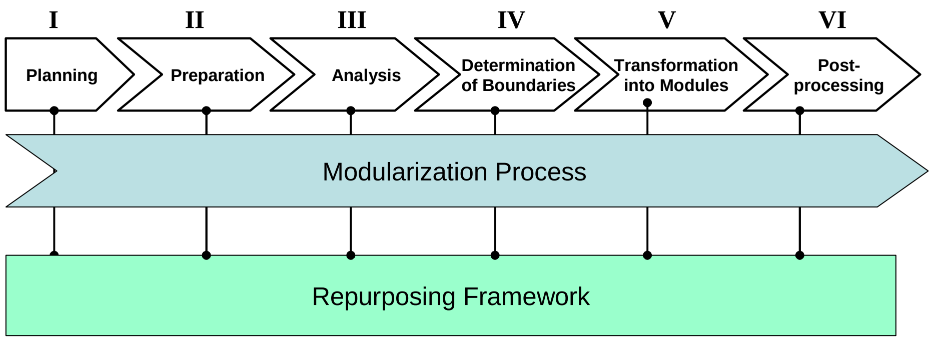

Fig. 2. Steps of the modularization process.

I) Planning. In the planning phase, the modularization goal has to be determined and methods and criteria to apply have to be identified. This includes the intended re-use

purpose for the resulting modules and also the question which criteria to apply (contents, didactic and media criteria) and which module granularity to chose.
II) Preparation. Before modularization can take place, a preparation may be required. If a Learning Resource exists in an arbitrary format, it might be desired to transform it into a format that is better suited for re-use. If SCORM content is modularized, it is recommended to break large Shareable Content Objects (SCOs) into smaller SCOs or assets first, and also remove internal navigation elements of the SCOs.
III) Analysis. In a further step, the Learning Resource fragments are analyzed regarding different criteria, which may have impact on the determination of reusable fragments. The analysis comprises properties of the contents of fragments, their didactic functions and media properties. Content properties, for example, may be covered topics, similarity of fragments or references between fragments.
IV) Determination of module boundaries. Based on the information, which has been collected in the previous process steps, the fragments that are to be re-used are chosen. Usually, resulting modules are supposed to be consistent and self-contained; but depending on the re-use purpose, a deviation from this principle might be superior. Two basic kinds of modularization are to be distinguished: Segmentation and selection; though these two kinds are supposed to appear combined in practice. Segmentation is a partitioning of a Learning Resource into several modules, whereas selection decides which fragments are re-used and which are not.
V) Transformation into modules (decomposition). The transformation step realizes the determined module boundaries by decomposing the Learning Resource. The resulting modules are made compatible to a predefined modular Learning Resource format, which is suitable for the intended use. If, for example, the result is supposed to be integrated into a larger course, the modules have to be aggregateable.
VI) Post processing. The results of decomposition are very often not yet suitable for re-use. There are still references to Learning Resource fragments, which no longer exist in the target modules, contents have become inconsistent and old metadata records do no longer fit the new modules. These shortcomings have to be eliminated in a post processing step. Not all of these tasks can be automated – some require a manual intervention of the user.
These six process steps cover all necessary tasks for modularization of existing Learning Resources. The level of automation and how much a user is involved in manual tasks may differ between implementations. For my proof of concept, I assume that especially the determination of module boundaries requires involving the user as decision maker.

4 Recent Contributions

Some contributions have already been made for enabling modularization. A repurposing framework has been developed for facilitating the design and implementation of repurposing applications [4]. It is suitable for a modularization application, but also allows adapting Learning Resources to different learning or teaching contexts. Additional components may be connected to the framework for realizing the content analysis methods, which are required for the analysis step of the modularization process.

The current version of the Shareable Content Object Reference Model (SCORM) does not support aggregation of packages to larger units. I have developed an extension to SCORM, which enables to aggregate Learning Resources by reference [5]. This extension of SCORM will be the primary target format for modularized content for my proof of concept implementation.

5 Current Solution Approaches and Work in Progress

5.1 Preparation of Learning Resources for analysis and decomposition

There are many existing courses which are labeled as SCORM compliant but use SCORM only as a wrapper for arbitrary content. In the worst case, the package contains only one single SCO, which provides an internal navigation for the whole course. Modularization can be facilitated by decomposing SCOs first into smaller SCOs or assets. The decomposition comprised determining the SCO’s internal structure and making it explicit; and also to identify and remove navigation elements of the SCOs.

5.2 Statistical analysis of content fragments

Methods for statistical analysis of Learning Resource fragments are currently evaluated, and are already showing promising results. There are two approaches for unveiling topics of and correlations between fragments. The first approach transfers text segmentation methods to Learning Resources. Calculating pair-wise statistical similarity of fragments allows detecting topic shifts within a Learning Resource. The second approach is based on the hypothesis that Learning Resources and fragments of Learning Resources are similar to Wikipedia articles that are concerned with the same or similar topics. I currently work on methods for predicting covered topics and especially topic shifts by the comparison of content fragments at different levels of granularity to Wikipedia articles.

5.3 Interactive GUI for Module Boundary Determination

The third issue which is currently worked on is an interactive graphical user interface (GUI) for the determination of module boundaries. A preliminary implementation visualizes the structure of a Learning Resource and lets the user assign module boundaries. The goal of the GUI is to support the user’s decision making by providing him all necessary information. Hence an important research issue is: Which information is necessary for the determination of module boundaries and how should it be presented to the user? This question is strongly related to content analysis.

6 State of Art and Discussion

The need for modularization of existing Learning Resources has already been expressed in other scientific works. Duval and Hodgins mention in their LOM Research Agenda *decomposition* as an open research issue [2].

A guideline for the transformation of existing course materials into reusable learning objects is provided by Doorten et al. [1]. That work describes a manual decomposition process for domain experts, which is focused mainly on didactic properties of course materials and how to achieve self-contained learning objects. There exists a lot of authoring tools for Learning Resources which provide decomposition into smaller units. However, this functionality is often limited to exporting a chapter within the course structure, which has to be manually selected by the user. Examples for such tools are the Reload Editor, which may export a part of a SCORM package as a new package, and the Phoenix editor, which can be used for creating documents with pedagogic markup. Phoenix enables the decomposition of pedagogic units which have been created as reusable units before [3]. Another project which targets at decomposition of Learning Resources is ALOCoM. ALOCoM decomposes slide presentations and enables the re-use of individual slides for new presentations [6].

My solution differs from the existing approaches for different reasons. First of all, there is no approach known to me that supports the whole modularization process by a technically-integrated solution. Existing solutions are either guidelines for manual modularization or do not support the whole process from retrieval to the final modular contents. Also, existing approaches do not take into account that Learning Resources have to be modularized differently for different re-use purposes. And finally, available approaches do not sufficiently regard the role of the user as the real decision maker, who has to be supported by providing adequate information.

References

[1] Doorten, M., Giesbers, B., Janssen, J., Daniels, J., Koper, R.: Transforming existing content into reusable learning objects. In R. McGreal, Online Education using Learning Objects, London, 2004, S. 116-127.

[2] Duval, E.; Hodgins, W.: A LOM Research Agenda. In *Proceedings of the twelfth international conference on World Wide Web*, 2003; S. 1-9.

[3] Fernandes, E., et al.: Phoenix Tool: A Support to Semantic Learning Model, Proceedings *of the 5th International Conference on Advanced Learning Technologies (ICALT '05)*, 2005, S. 948-949.

[4] Meyer, M., Hildebrandt, T., Rensing, C., Steinmetz, R.: Requirements and an Architecture for a Multimedia Content Re-Purposing Framework. To appear in the *Proceedings of the First European Conference on Technology Enhanced Learning* (EC-TEL 2006).

[5] Meyer, M., Rensing, C., Steinmetz, R.: Supporting Modularization and Aggregation of Learning Resources in a SCORM Compliance Mode. To appear in the *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies* (ICALT '06).

[6] Verbert, K., & Duval, E. Towards a Global Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, S. 202-208.

Metrics for Learning Object Metadata

Xavier Ochoa[1] and Erik Duval[2]

[1]Escuela Superior Politcnica del Litoral (ESPOL), Via Perimetral Km. 30.5, Guayaquil, Ecuador
xavier@cti.espol.edu.ec
[2]Computerwetenschappen Dept., Katholieke Universiteit Leuven, Celestijnenlaan 200 A, B-3001 Leuven, Belgium
erik.duval@cs.kuleuven.be

Abstract. Previous research has set the foundation of Learning Object Technologies; unfortunately just the foundation is not enough to convince instructors and learners to use the technology. Mature tools are needed in order to breach the early-adopter - mainstream gap. Following what has been done with other related technologies, this work presents research to create automated measurements (metrics) that could enable the creation of a new generation of smarter and friendlier learning object applications. The methodology for the proposal and test of the metrics is discussed, along with early encouraging results in the area of metadata quality metrics. Finally we present possible applications of the current and future results of this research.

1. Research Context and Justification

During almost 15 years of research, the foundation of Learning Objects Technologies has been set. There are standards that define the metadata that describe a learning object [1] and how to sequence it [2]. Thanks to these standards, Learning Management Systems (LMS) are able to import and export learning objects of different granularity. There are several repositories worldwide where the instructors can publish the learning objects that they create and search for learning objects published by peers [3]. Thanks also to standardization [4], these repositories can query each other and present the user with a considerable amount of results.

Despise all the development in this necessary foundation, the tools that the end user access to index, search, integrate and re-use learning objects are still immature if we compare them to similar tools used in related fields. For example, web pages creators or papers publishers do not need to manually index their work [5] [6]. The basic text-matching or field-matching techniques of current learning object search tools are not enough to sort the huge amount of results returned by federated queries, a problem that has been solved in the World Wide Web thanks to PageRank-like algorithms [7]. There is not readily available equivalent in the learning object community for the Amazon's book recommending feature [8]. This lack of maturity is reflected in the low level of adoption of learning object technologies among instructors and learners [9] [10].

In order to improve the adoption of learning object technologies, smarter and friendlier end user tools must be developed. These tools should capitalize the vast amount of information that is present in the learning object metadata and other sources as context and usage. To be exploitable, that information should be automatically measured and processed to extract deep knowledge of the characteristics, relations, usefulness, behavior and recommended usage of individual learning objects, as well as, complete learning object repositories. This research will create and test automatic quantitative measurements (metrics) for learning object metadata. The objective of the metrics will be to find useful calculations that use intrinsic and extrinsic information to improve the performance and usability of the current tools. For example, if the number of times an object is reused (metric) is a good predictor of the relevance of a learning object (learning object characteristic), it could be used inside the sorting algorithm of federated search application (tool). We have called this initiative "Metrics for Learning Object Metadata".

The idea of using metrics to automate or improve tools or procedures has been borrowed from other fields: Software Engineering uses metrics to semi-automatically determine the cost and duration of software projects [11]. Scientometrics creates metrics to automatically predict the "impact" of a journal or gain insight about the research environment at a given moment and field [12]. Webometrics creates metrics to determine the relevance of a web page [13] (for example Google's PageRank metric [7]). This new small research area, Metrics for Learning Object Metadata (Learnometrics), will create metrics based on the information present in the metadata record(s) of a learning object and the contexts where it is used to enable the creation of a new generation of Learning Object end user tools.

2. Research Focus

During this research work, we will create and test metrics for three main areas:

Quality of Metadata:
The quality of the metadata record that describes a learning object affects directly the chances of the object to be found, reviewed or reused [14]. To deal with the exponential grow of metadata records available and at the same time to be able to retain some sort of quality assurance for the information contained in the metadata record, we propose automating the quality assessment of learning object metadata. This automated evaluator will assess intrinsic characteristics of the metadata itself, measured through the use of one or more synthetic metrics.

Similarity between Learning Objects:
We propose to measure the semantic distance between two or more learning objects. This metric (or set of metrics) will enable the creation of a service that can automatically cluster two or more learning objects together based on their characteristics. All kind of tools, from Automatic Metadata Generation to Recommendation Systems, can use this service to improve their performance.

Relevance of Learning Objects:
This group of metrics should give insight on the relevance of learning objects in a specific situation based on the available information from the learning object metadata, usage and context information, as well as, users' annotations. These relevance metrics could be the seed to create a LearnRank [15] algorithm.

3. Methodology

Finding useful metrics is an experimental endeavor. A lot of try-and-error is needed to find good predictors. First, (1) based on theoretical or empirical considerations, a metric is proposed. Then, (2) an experiment is set-up to obtain a real value for the characteristic we want to predict with the metric. Finally, (3) the real values are compared with the values obtained from the metric. If the values correlate for different learning objects or users, then the proposed metric (or metrics) is (are) a good predictor for the selected characteristic. This metric could then be use to provide services to improve a tool (4).

As an example of this procedure: (1) The set similarity theory will suggest that counting the number of fields where two learning objects have the same value could be used to predict how similar they are. (2) The experiment will be to present a set of different learning objects to a group of human examiners. They will score the grade of similarity between them. (3) The scores will then be correlated with the values calculated with the similarity metric for the same objects. As result we will know if the similarity metric is a good predictor of the similarity as perceived by human reviewers, and (4) we can use it inside a learning object clustering service.

To facilitate this experimentation, a framework is being built. This framework, called M4M, enable easy prototyping of metrics and tests against real data. While M4M has been developed for learning object metadata it is based on a XML database, in this way, it could be easily extended to support any other kind of metadata.

4. Current Status

At the time of writing, the most developed research field is Metadata Quality Metrics. We have developed a group of interesting metrics to assess the quality learning object metadata records, loosely based on a metadata quality framework proposed in [16]. In an accepted paper [17] we describe these metrics and their rationale. The formula of this metrics can be seen in Table 1. These metrics has been tested against the quality value obtained from an online experiment were human reviewers grade the learning object metadata using the same quality framework in which the metrics where based. The results obtained indicated that one of the metrics, the Textual Information Content provide a good predictor (correlation factor = 0,842) of the evaluation of average score given by human reviewers to the evaluated objects.

Table 1. Formulas of the Quality Metrics

Metric Name	Metric Formula
Simple Completeness	$$\frac{\sum_{i=1}^{N} P(i)}{N}$$ Where P(i) is 1 if the ith field has a non-null value, 0 otherwise. N is the number of fields.
Weighted Completeness	$$\frac{\sum_{i=1}^{N} \alpha_i * P(i)}{\sum_{i=1}^{N} \alpha_i}$$ Where α_I is the relative importance of the i^{th} field.
Nominal Information Content	$$\sum_{i=1}^{K} -\log(P(value_i))$$ Where K is the number of nominal fields. $P(value_i)$ is the probability of a value of the ith nominal field.
Textual Information Content	$$\log\left(\sum_{i=1}^{N} tf(word_i) * \log\left(\frac{1}{df(word_i)}\right)\right)$$ Where $tf(word_i)$ is the term frequency of the ith word, $df(word_i)$ is the document frequency of the ith word.
Readability	$$\frac{Flesch(title \,\&\, description\_text)}{100}$$ Flesch(description_text) is the value of the Flesch index for the text present in the title and description of the record

This metric (Textual Information Context) has been used in a tree-map visualization tool where the quality of the metadata present ARIADNE repository can be easily assessed. Figure 1 present a snapshot of this application: red represents low quality metadata records, yellow medium quality metadata records and green high quality records. The size of the box represents the number of metadata records that has been created by a given author in a given local repository. This tool can be accessed online at [18].

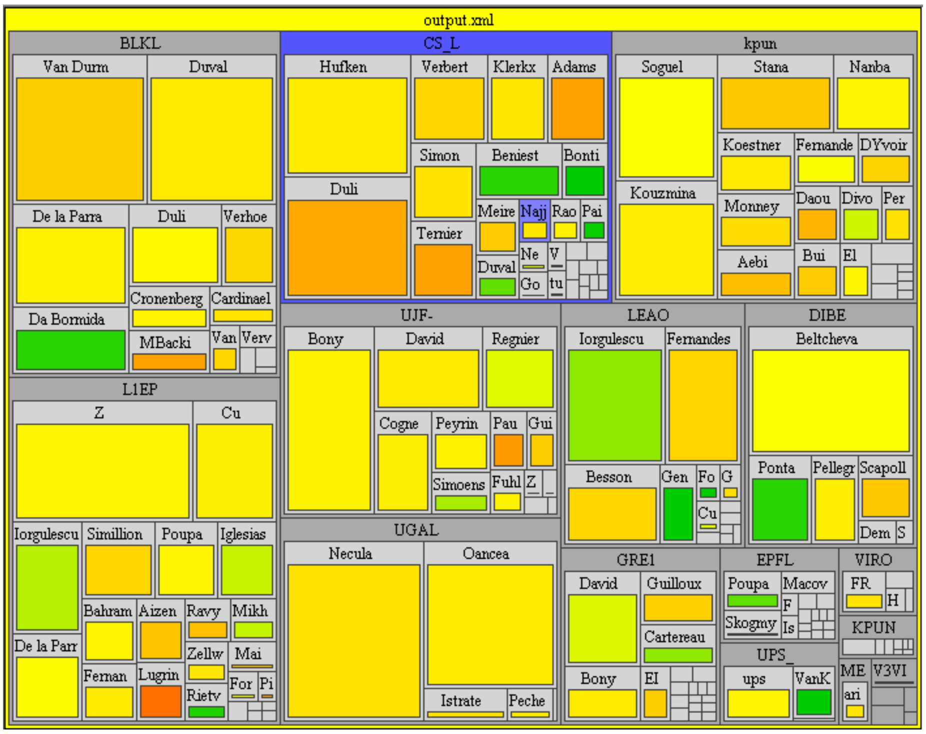

Figure 1. Visualization of the Metadata Quality of a Learning Object Repository

Currently, we are using the Textual Information Content metric to compare the quality of several learning object repositories members of the GLOBE consortium [19]. In the field of Similarity Metrics, just early experiments have been performed. A clustering search application can be accessed at [20]. It clusters learning objects based on the text that is contained in the title and description fields of the metadata record.

4. Expected Results and Application

As a result of this research work, it is expected that several useful metrics has been found and can be used inside learning object tools in order to provide a smarter interaction with the final user. The indirect desired effect is that learning object tools become mainstream among instructors and learners worldwide.

As for the possible applications that we envision this metrics will enable, we can list a few:

- **Automated Evaluation of Quality:** To establish if the information available in the metadata record makes the object useful inside a certain application.
- **Ranking of Objects:** To assign a comparative value to an object in the result list of a search tools reflecting the relevance of the object.

- **Replacement or Updating of Objects:** To find similar, more recent/available, objects based in the characteristics or use of the original one.
- **Recommendation:** To establish the relevance of an object to an user based on usage patterns and user profiles (for example the information extracted from Social Software or Learning Management Systems)
- **Measuring the impact of Learning Objects:** Calculating how useful a learning object has been to a certain community.
- **Interoperability:** To find semantic-corrective calculations that will enable the exchange of information between two or more collections with different cultures.

4. Conclusions

As well as Webometric research lead to the development of smarter Web Search Engines, if we want to develop better learning object tools we must research about the basic characteristics of the learning object information (metadata, usage, context) and provide ways in which this characteristics could be measured and used. A new field, Learnometrics, should be researched. This work is a step in that direction.

As with the development of metrics in other fields, this task is mainly based on extensive experimentation. While this work focus on the finding of metrics in three main fields (Quality, Similarity and Relevance), it also set a methodology to explore new areas where metrics could be used to extract useful information to be used in novel tools.

This work is complementary to recent learning object research. It will consume information from Attention Logging [21], Federated Searches [4] and Social Networks and will provide services that can be used by Adaptive Learning, Social Recommending [22], Automatic Generation of Metadata [23] or Visualization [24] applications.

References

1. IEEE (2002). IEEE Standard for Learning Object Metadata. http://ltsc.ieee.org/doc/wg12/
2. ADL (2004). SCORM (Sharable Courseware Object Reference Model) http://www.adlnet.gov/scorm
3. Neven, F., Duval, E. (2002) Reusable Learning Objects: a survey of LOM-based Repositories, Proceedings of the 10th ACM international Conference on Multimedia, pp. 291-294, Juan-les-Pins, France.
4. Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S. et al. (2006). A Simple Query Interface for Interoperable Learning Repositories. In B. Simon, D. Olmedilla, & N. Saito (Eds.), (pp. 11-18).
5. Bollacker, K., Lawrence, S., and Giles, G. (1998) CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications. In Agents '98, http://citeseer.ist.psu.edu/
6. Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In Proceedings of the Seventh International World Wide Web Conference (WWW7), Apr. 1998.

7. Page, L., Brin, S., Motwani, R. & Winograd, T. (1998), The pagerank citation ranking: Bringing order to the web, Technical report, Stanford Digital Library Technologies Project.
8. Linden, G., Smith, B., York, J. (2003) Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, Volume 4, number 1
9. Koppi, T., Lavitt, N., (2003) Institutional Use of Learning Objects Three Years on: Lessons Learned and Future Directions, Learning Objects 2003 Symposium: Lessons Learned, Questions Asked, Honolulu, Hawaii, USA.
10. Mohan, P., Greer, J., (2003) Reusable Learning Objects: Current Status and Future Directions, ED-Media Conference, Honolulu, Hawaii, USA.
11. Leung, H., Fan, Z. (2002) Software Cost Estimation. Handbook of Software Engineering and Knowledge Engineering, (eds) Chang S. K. Vol. II, World Scientific Publishing Co., ISBN 981-02-4974-8
12. Garfield, E. (1995) Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. Science, 122(3159), p.108-11, No: 25
13. Almind, T., Ingwersen, P., (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'. Journal of Documentation 53 (4): 404-426.
14. Currier, S., Barton, J., O'Beirne, R., & Ryan, B. (2004). Quality assurance for digital learning object repositories: issues for the metadata creation process. ALT-J, Research in Learning Technology, 12, 5-20.
15. Duval, E. (2005). LearnRank: the real quality measure for learning materials. In A.McCluskey (Ed.), Policy and Innovation in Education - Quality Criteria (pp. 26-29). European Schoolnet.
16. Bruce, T. & Hillman, D. (2004). The continuum of metadata quality: defining, ex-pressing, exploiting. In D.Hillman & L. Westbrooks (Eds.), Metadata in Practice (Chicago): American Library Association.
17. Ochoa, X. & Duval, E. (2006). Quality Metrics for Learning Object Metadata. Accepted for Ed-Media Conference 2006.
18. Ochoa, X (2006), Quality Visualization Tool. http://ariadne.cti.espol.edu.ec/M4M/tremap/applet.htm
19. GLOBE (2006). Global Learning Objects Brokered Exchange. http://taste.merlot.org/initiatives/globe.htm
20. Ochoa, X (2006). Learning Object Clustered Search Demo. http://ariadne.cti.espol.edu.ec/clustered-search
21. Najjar, J., Meire, M., & Duval, E. (2005). Attention Metadata Management: Tracking the Use of Learning Objects through Attention.XML. In Proceedings of ED-MEDIA 2005.
22. Recker, M. M., Walker, A., & Wiley, D. A. (2000). "Collaboratively filtering learning objects." In D. A. Wiley (Ed.), The Instructional Use of Learning Objects: Online Version.
23. Cardinels, K., Meire, M., & Duval, E. (2005). Automating metadata generation: the simple indexing interface. In Proceedings of the 14th international conference on World Wide Web (pp. 548-556). New York, NY, USA: ACM Press.
24. Klerkx, J., Duval, E., Meire, M. (2004) Using Information Visualization for Accessing Learning Object Repositories. IV 2004: 465-470

MD2 Method:

The Didactic Materials Development from a Model perspective

Carmen L. Padrón Nápoles

DEI Laboratory
Computer Science Department
Universidad Carlos III de Madrid, clpadron@inf.uc3m.es

Abstract. The didactic material development is a complex, intensive, time-consuming process that needs an effective support. This is the goal of our PhD research and the MD2 method presented in this paper. We propose the MD2 model to describe general didactic materials requirements like its content, pedagogical, technical and quality features. MD2 method uses the information provided by the MD2 model to guide the different development stages and they are the foundation of a generative authoring tool for didactic materials. The main idea is that users will specify material features at a high level of abstraction and the generative tool will be capable infer and fill the information for lower level of design, allowing to assemble material components in delivery time or runtime and to generate accurate semantic annotations. That tool will also control the quality and usability of material based on the coherence and completeness of contents and their capability to effectively support the achievement of stated learning objectives.

Keywords: Authoring Tools, Metadata, Learning Objects, Modeling, Development Methods.

1 Introduction

E-Learning provides, among other advantages, the means to integrate teaching and learning into every facet of each person's life, promoting the life-long education and increasing the globalisation of education [6]. Thanks to e-learning there are accessible a wide range of new opportunities for instructional reusability and personalised learning needed for long-life education. But those opportunities can not be efficiently deployed if there are not available enough didactic materials able to support those features and authoring tools to create such types of materials.

The didactic material creation is a complex, intensive, time-consuming process. This creation process is composed by a set of phases: analysis of requirements, design, development or implementation and evaluation [16]. Each of those phases needs special support. Instructional Design initiatives like the Instructional Engineering Method (MISA) [22] focuses on proving solutions to first two phases. But our research interests are more related to the development or implementation phase. The recent results of research related to e-learning standards and specifications have a decisive role in this area. Since they provide common means and prescribe guides to make materials interoperable among heterogeneous systems (e.g. IMS CPIM [7]; SCORM [27]); accessible based on suitable dealings with intellectual property rights (e.g. IMS LRMDI [10]); capable for management and personalization (e.g. IMS LIP [9]) and, flexible enough to perform composition and integration into new materials (e.g. IMSLD [8], IMS QTI [11]).

The need of authoring tools neither has been ignored and there is available an important number of those tools in recently years. Among them are worthy to mention: RELOAD [24], CopperAuthor [6], Aloha [1] and OLAT QTI Editor [19]. They have in common the successful technical implementation of learning technology standards and specifications such as IMS LD, IMS LOM, IMS QTI or ADL SCORM. In that way, their use represent a step ahead assisting the development of didactic materials and ensuring some important features like material's interoperability, accessibility and personalization. But the design of such tools and the solutions they provide are suitable for a small segment of all the possible intended users: those practitioners with a deep technical understanding of the standards. By other side, important issues arise during the development process like how to find most suitable contents, where can be they retrieved from, which are appropriated criteria to select contents, when and how to control the coherence and accuracy of contents? Or how to ensure the reusability of the material based on the reusable nature of its components? Unfortunately those tools lack of support to solve all those issues.

Those are some unsolved problems we had found in our analysis of current support for the development of didactic materials in e-learning. Thus, our research is focused on defining development framework to systematically and rationally solve such problems.

Once we had introduced the motivation of our PhD research work and identified some significant problems on this research area, the rest of this paper is structured as follows: An outline of the current knowledge of the problem domain as well as the state of existent solutions is presented in Section 2. Our proposal approach with its preliminary ideas, some details about the applied research methodology and the achieved outcomes are presented in Section 3. Finally in Section 4 a description about our approach's contribution to the problem solution and a brief discussion of how the proposed solution is different to existing approaches to the problem are outlined.

2 Related work

Didactic materials are herein considered as the combination of certain contents with a pedagogical strategy defined by an instructional design. Our understanding of didactic materials concept is based on two known definitions: reusable learning objects and unit of learning. According to Polsani in [23] a learning object is an independent and self-standing unit of learning content that is predisposed to reuse in multiple instructional contexts. Meanwhile, units of learning are pedagogical strategy representations defined in IMS LD [8] as a diverse group of prescribed activities based on a given set of contents (in the form of learning objects) that allows learners to obtain certain learning objectives (i.e. acquisition of knowledge, skills, competences and/or attitudes). A unit of learning also includes assessments, services, and support facilities provided by teachers, trainers and other staff members.

On the other hand, authoring tools can be described, according to Murray in [18], in term of two aspects: an underlying representational framework (conceptual model), and the user interface (the authoring modules) that reify this framework for the user, allowing her to create, visualize and modify elements of the didactic materials. Thus, the power and effectiveness of authoring tools strongly depends on the power and fidelity of the underlying conceptual model and the usability of its interface.

Such conceptual models must describe material main elements and also the stages or phases of its development process. Reusable didactic materials can be considered as knowledge based systems thus, it contents element should be described as part of certain knowledge domain and the instructional design element, as a teaching strategy belonging to the pedagogical domain. By other hand, the development process can be divided in three stages or phases in order to provide answers to the problems stated in previous section. Those stages are: selection (i.e. election of contents); composition (i.e. assembling of contents and their integration with the teaching strategy), and evaluation (i.e. preview of created material and the assessment of how it does satisfy predefined material requirements).

Many research initiatives has developed interesting approaches to solve problems related to the development of didactic materials from model based perspective in recent years. Next we present a short analysis of some of those approaches according to the following criteria:

- Which element of didactic materials has been modelled in the approach?
- Is the approach related to e-learning standards or specifications? What kind of relation has it?
- Which development stages are supported?

First we analyze content-centered modelling approaches and later, those efforts focused on instructional design and pedagogical strategies.

TeachML has been developed as part of the project Targeteam (TArgeted Reuse and GEneration of TEAching Material) [28]. This approach concentrates on the contents and their reusability. It is a language to mark-up media independent learning resources and to create course structures in XML. TeachML language has been designed for representing modular teaching materials as reusable units or modules.

In the Targeteam approach contents integration is done on the abstract content-structure level, not on the layout or presentation level. The module structure is used to produce a comprehensive XML document from a module and all its sub-modules and atoms. Thus, materials are integrated into a uniform TeachML document which can then be processed for different presentation media. In this approach, contents are modelled as documents and the reusability of material is viewed as their capability for being represented in diverse delivery formats. It is an interesting solution for content assembling, but it does not address problems related to the selection of those contents. Although Targeteam approach is open source solution that uses XML documents as standard mean for representation contents, the resulting contents are not compliant to any e-learning standard.

Other interesting approach is the Knowledge Based Tutor Meta-model. It has resulted as conclusion from the analysis of the two dozen of authoring tools for advanced-technology learning environments [18]. The KBTMM is a systematic synthesis and abstraction of important elements found in that analysis. This meta-model is concerned with the representation of contents based on its pedagogical knowledge (i.e. declarative information specific to a knowledge domain which is relevant for teaching about that domain). KBTMM proposed a layered curriculum object framework, called Decision Architecture, which illustrates common pedagogical components of the tutor systems and allows their composition.

The pedagogical knowledge information provided by KBTMM can help solving issues related to selection of the appropriated contents for material development. Nevertheless it does not address the representation of instructional design or teaching strategy, the composition of a system described by the Decision architecture can be also useful for content assembling and their integration with a teaching strategy, described through an instructional design model.

The effectiveness of KBTMM has been proved in the development of more than 10 different tutoring systems using Eon Authoring Tools [18]. But unfortunately none of those tutoring systems are compliant to e-learning standards.

The IBM Dynamic assembly of learning objects (IBM DALO) is a solution designed and tested by the Learning Objects Framework Research Team at IBM [5]. It is a proprietary solution, which models material based on the SCORM specification. This solution uses a fixed and predefined instructional design of materials to create explanatory courses (how-to types) based on IBM products. The knowledge enclosed in learning objects (content) is modelled as topic graphs represented through RFD schemas. Learning objects are stored in repositories and they are properly annotated with a subset of IMS LOM tags (General, Educational, and Classification) to ensure their localization.

In IBM DALO approach, the selection stage is based on user queries about topic contents, some learning constrains and user preferences. And contents are retrieved by means of a crawling mechanism. The composition stage uses semantic connections of the topic graph to control coherency of the retrieved set of contents and to provide a logical sequencing of those contents according to the pedagogical strategy.

This approach has been experimentally validated by its implementation in the Dynamic Learning Experience System. The results had confirmed that such kind of course assembly provides high quality design courses and it is suitable for effective personalization. But this solution left open some issues mainly related to reusability of created materials and generalization of its composition mechanism. The first is that material reusability is partially ensured since it is a content-centered modelling approach and just one pedagogical strategy is used, hence obtained courses can be only reused in learning situations with the same pedagogical strategy. The proposed composition mechanism is based on sequencing rules according to that fixed pedagogical strategy. Therefore the second open issue is about how this mechanism can be extended to develop materials with different strategies. Those issues can be solved with means to generally describe and support different pedagogical strategies, and that is one of main goals of the research efforts about educational modelling languages

There is available a notable number of researches about educational modelling languages, as representational frameworks to describe learning contents and processes, at a pedagogical level of abstraction. Among them are Palo language [25] and the OUEML [13].

The work on Educational Modelling Languages like OUEML and the subsequent integration of a subset its elements in the IMS Learning Design Specification (IMS LD), is the most important initiative to date, to integrate Instructional Design concerns into the international standards movement. The IMS LD generally models pedagogical strategies and their delivery modes in such way that material technical interoperability and reusability are ensured. It describes a formal way to represent the structure of unit of learning and the concept of pedagogical method specifying roles and activities that learners and academic staff can “play” (carry out) using learning objects and services in the environment. It can also describe methods as workflow processes, that can provide alternative plays adapted to different target populations and delivery methods. IMS LD can serve to support different instructional or pedagogical strategies like collaborative learning, problem solving, project-based learning, communities of practices, and multi-facilitators support as found in distance education universities. Unfortunately, the specification does not specify proper guidelines for the development of materials but it provides useful information to create material structure templates for different pedagogical strategies that can help in the material composition stage.

But conceptual models are not enough to create effective authoring tools, mechanisms to process the information they provide are also needed. Research efforts related to Ontologies can help on those tasks because these afford means to represent a portion of our mental model about a specific domain [3] in a computer-usable and machine-understandable way (e.g. software agents, sophisticated search engines or web services) that facilitates the automated processing of elements from that domain. The research on educational ontologies is not scarce. Outstanding examples are Murray’s proposal [17], the Mizoguchi’s approach [15] and more recently the Leidig proposal [14]. Educational ontologies can provide the means to create templates, wizards and consistency-checking tools that help authors during the development of didactic materials [14]. They can help in the selection stage since they support appropriate semantic interpretations for search engines in order to localize and retrieve didactic materials from distributed repositories. They also facilitate the automated composition and configuration of learning processes as long as the conceptual model contains relationships between instructional design and the contents.

All those approaches give partially solutions to some of the issues we had presented as motivation of our work. They offer valuable solutions for modeling the main elements of didactic materials and provide means to address some of the issues emerging during different development stages like content selection and assembly or their integration with a pedagogical strategy. Thus, the analysis of their contributions let us to think that we can design a general development framework that rationally and systematically solve the problems related to didactic material development.

3 The MD2 APPROACH

Our approach is named MD2 after the Spanish acronym for Development Method of Didactic Materials (and their automation) [20]. Some of the most important rationales behind the MD2 model and method to support the development of didactic materials are presented in this section. And, we also outline the applied research methodology and achieved results at the time of this writing.

Our PhD work is part of the MD2 research project. It is concerned with one of the MD2 project main endeavors: the definition development method for didactic materials as basis for a set of tools that guides and assist creators during the entire development process, the MD2 method.

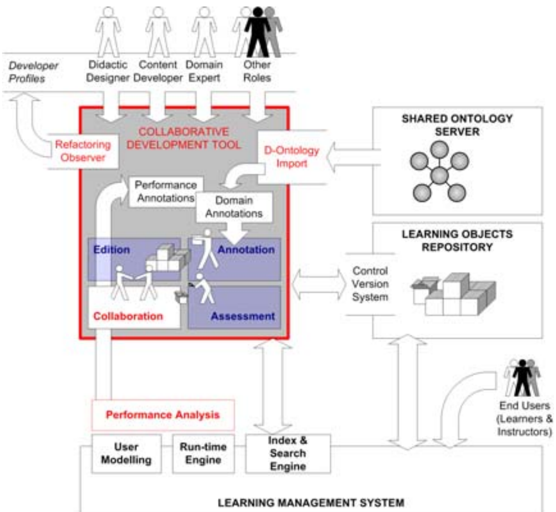

Fig 1. General architecture of MD2 platform

Figure 1 depicts the architecture of a distributed MD2 platform. Its main component is a collaborative development tool, composed by different functional modules. The core of this tool is a set of modules related to edition and annotation of didactic materials which, implements the selection and composition procedures defined by MD2 method. Meanwhile a part of the assessment module is also responsible for the quality assessment of the material during the evaluation stage of the development process according to MD2 method. The collaborative support is provided by another module that is based on a distributed collaborative server. The rest of the components presented in the figure are external subsystems (e.g. learning management systems, a shared ontology server and repositories of learning objects, assessment and activities) which interact with the main tool through a web services based architecture.

Once we had explained the relation of our PhD research with its homonym project, next we summarize the steps of the research methodology applied in our work. We had used as reference the methodology proposed in [26]. It prescripts the following steps:

1. Formulation of the research question
    a. Definition of the research question context and the motivation to work finding a solution.
    b. Outline the current knowledge of the problem domain, as well as the state of existing solutions.
    c. Clear formulation of the research question.
2. Formulation of a work hypothesis
    a. Definition of the main work objectives.
    b. Design of a solution based on the hypothesis, capable to be qualitatively and empirically evaluated.
3. Evaluation of work hypothesis
    a. Selection of case studies to validate the hypothesis.
    b. Definition of the evaluation criteria.
    c. Definition of the evaluation procedures and tasks scheduling to evaluate the hypothesis solution using the selected case studies.
4. Analysis of evaluation results
    a. Elicitation of conclusions based on the obtained evaluation results and their comparison with the defined work objectives (2a).
    b. Generalization of results and analysis of hypothesis applicability to other knowledge domains.

We have been working on tasks prescribed by the two first steps of this methodology until this moment. The Introduction and Related work sections of this paper contain a concise summary about results obtained from prescribed tasks of the Step 1.

As we explained in previous section, the development of didactic materials can be effectively supported if we count with means to describe material main features: contents and pedagogical strategy and, a method as guide for the different development stages: selection, composition and evaluation. Thus, our hypothesis work is the following:

- If we have a model that generally describes pedagogical, technical, knowledge and quality requirements for contents, then those descriptors can be used as criteria to select the most appropriated contents for the material and to check contents quality and accuracy.
- If that model can also describe pedagogical strategies based on e-learning standards, it will provide information about what kind of material structure for its implementation is needed and how contents can be connected or plugged to such structure.
- Then, a method for didactic materials development can use the information provided by the model to solve issues related to each one of the development stages. And those model and method can be the foundation of a generative authoring tool for didactic materials.

The main idea is that users will specify material features at a high level of abstraction and the generative tool will be capable infer and fill the information for lower level of design, allowing to assemble material components in delivery time or runtime and to generate accurate semantic annotations. The tool will also control the quality and usability of material based on the coherence and completeness of contents and their capability to effectively support the achievement of stated learning objectives.

We had design a solution based on that hypothesis. The MD2 model generally describes didactic materials features like its content, pedagogical, technical and quality requirements by means of four view elements or descriptors. Those views are Knowledge Domain (KD), Pedagogical (P), Support (S) and Quality-Usability (U-Q). The rationale behind the model is to use those descriptors to provide mappings form high level technical descriptions of learning technology standards and specifications to simpler and closer descriptions to practitioners about material requirements. The mappings from some of the KD, P and S view elements allow guiding the selection and composition stages of the didactic material development process. Meanwhile a set of Q-U view elements are used for checking and to control the usability, pedagogical value and quality of created material during the development evaluation stage.

The MD2 method is composed by a set of 13 steps based on simple questions to developers about material requirements. The method steps follow a logical order, first are going questions (1) related to the KD view since they are foundation of the educational process. In second place are questions (2-5) about P view elements; they will determine the material pedagogical requirements. Next, the questions (6-12) are related to S view to define technical support requirements. Developer or creator answers to those questions are input data for inference mechanisms based on MD2 model, which helps finding or creating the most appropriated resources during the selection stage. The inference mechanisms also uses another set of those answers to obtain mappings from the stated requirements to the proper information model elements from a delivery and publishing standard, thus the material presentation structure can be defined and completed through the composition stage of development. Final method step (13) is concerned with the evaluation of material completeness, coherence and accuracy of its contents and its capability to effectively support an educational process. The results of such evaluations will provide developers with some confidence information about the quality and effectiveness of the created material. Also all values assigned to MD2 model elements will be automatically stored as extended semantic annotations of the created material once the evaluation stage has finished. By that way, material future localization and retrieval for reutilization proposes is ensured. All details about the MD2 model and MD2 method can be found in the proceedings of this conference [21].

This development framework must be implemented in order to validate our working hypothesis according to Step 2 of the research methodology. Thus, we had defined a set of tasks to carry out such implementation and we explain their accomplishment state at the time of this writing.

Once the MD2 model had been designed are necessary:

- A design and implementation of Ontology for MD2 model in order to conduct automatic classification with a reasoner that evaluates model consistency and its effectiveness to classify different types of didactic material examples. This task has been finished.
- A definition of an extension for the IMS LOM to include information about the material design rationales used in the material development. This information will be used to check if their use allows easier and proper resources retrieving from repositories for composition stage in such

development situations when other practitioners need to create didactic materials with similar requirements. This task still in progress.

After the definition of MD2 method was concluded, we started the design and implementation of a prototype for MD2 generative tool to evaluate the method technical feasibility. Thus, the following tasks were mandatory:

- A design of materials repository and its implementation. It will be used as local storage for the different kind of materials described by MD2 model: assessments, activities and learning objects. Also a selection of external repositories and their access rules and retrieval mechanisms will be needed. The first part of this task is finished and the work on second part is in progress.
- A definition and implementation of fuzzy aggregation mechanisms for the material development evaluation stage. We had implemented such mechanism and we are working on its refinement.
- A design and implementation of material structure patterns represented by means of the adequate e-learning standards: IMS LD, IMS QTI and ADL SCORM. The selection of standard depends on the type of pedagogical strategy. The material structure pattern for assessment materials based on IMS QTI is done; we are working to define structure patterns based on IMS Learning Design UOL structures for other pedagogical strategies.
- Design and implementation of MD2tool prototype modules responsible for semantic annotations and each development stage: selection, composition, evaluation. Their functionalities are implemented according to MD2 method guidelines for each development stage. We had developed a prototype and we are conducting some tests to check the method effectiveness to support the development of assessment materials.

We have also started to work in the tasks related to the evaluation of the work hypothesis (Step 3). The evaluation will have two-fold objective validation: the defined model and the proposed development method by the hypothesis. To achieve those endeavours we are defining some case studies in order to check accuracy, description capability and generality of the designed model and the effectiveness of the development process guided by the method. Such study cases are development scenarios where didactic materials based on different pedagogical strategies must be created using the proposed generative authoring tool.

4 Discussion

In this section we briefly present MD2 contributions and analyze how it is different from the approaches presented in the section 2.

The MD2 contributions can be summarized in the following: the MD2 model which generally describes requirements for the didactic material development by means of four views. The MD2 method uses simple MD2 model descriptors of content, pedagogical, technical requirements to provide mappings to high level technical descriptions of learning technology standards. Such mappings are the foundation of MD2 method that guides selection and composition stages of development. The interoperability of created material is ensured thanks to those mappings as well. The MD2 model quality descriptors are used to check if features of obtained material satisfy usability and quality requirements. Thus, creators or developers will have some confidence about the achievement of those properties at the end of evaluation phase. MD2 method also defines that all those data exchanged with creators during the whole development process will be stored as an extended set of material semantic annotations in a Repository. Such kind of annotations contains material descriptions and its design rationales. The availability of such information will ensure material accessibility, reusability and its personalization capability.

We had use a set of supported features to compare how MD2 approach is different from the solutions provided by other approaches. Those criteria are:

1. Existence of any representation or description of the material main elements: contents and pedagogical strategy.
2. Presence of descriptions about the pedagogical knowledge information.
3. E-learning standards compliance of the obtained material (product) and its development process.
4. Support for the different development process stages and what mechanisms are used to achieve it.
5. Use of any design rationales support.
6. Semantic annotation support.
7. Reusability of the obtained material. Which of the main element of didactic materials is used to ensure it?

The Table 1 summarizes each approach contributions to the development of didactic materials. As can be seen from the table, their more relevant solutions to the different comparison criteria are highlighted and almost all approaches are focused on the composition stage of the development process. Most those approaches use a representation or description of just one of the material main elements. Therefore whether reusability is ensured, it is based on that described element. By contrast the MD2 approach tries to generalize some of their most important contributions, and it provides support for the three development stages by means of the MD2 four-view descriptions of both material main elements: contents and pedagogical strategy. Among those descriptors is also included the pedagogical knowledge information, which is used for the election of appropriated contents during the selection stage. The MD2 is the unique approach that has e-learning standard compliant development process and product. It also provides means to ensure that resulted didactic materials will have proper semantic annotations including its design rationales. Also, the integration mechanism, conceived for the MD2 composition stage (i.e. plugging selected contents into the implementation structure of the pedagogical strategy), ensures the reusability of the obtained material based on its both main elements.

At the time of this writing we are conducting evaluations to check the effectiveness of a MD2 tool prototype to support the development of assessment materials. Once we will finish all tasks related to Steps 3 and 4 from the applied research methodology, we will be in proper conditions to value the generality of our working hypothesis.

Acknowledgments. This work is part of the MD2 project (TIC2003-03654), funded by the Ministry of Science and Technology, Spain.

References

1. Aloha project website Retrieved December 16, 2005 from http://aloha2.netera.ca/
2. CopperAuthor project website. Retrieved February 3, 2006 from http://www.copperauthor.org/
3. Daconta, M., Obrst, L., Smith, K. (2003). The Semantic Web. A Guide to the Future of XML, Web Services, and Knowledge Management. Wiley Publishing, Inc., Indianapolis, Indiana.
4. Dick, W., Carey, L. & Carey, J. O. (2001). The systematic design of instruction, 5th Ed. New York: Longman.)- Retrieved form http://www.umich.edu/~ed626/Dick_Carey/dc.html
5. Farrell, R., Liburd, S.D., Thomas, J. (2004) "Dynamic Assembly of Learning Objects" in PGL Workshop On E-learning Objects and Systems.. Retrieved from http://pgl.ufl.edu/events/pgl2/Farrell/dynamic_assembly_pres.pdf
6. Hummel, H., Manderveld, J., Tattersall, C. and Koper, R. (2004) Educational modelling language and learning design: new opportunities for instructional reusability and personalized learning, *Int. J. Learning Technology*, Vol. 1, No. 1, pp.111–126, 2004.
7. IMS-CPIM. IMS Content Packaging Information Model. Version 1.1.2 final specification, 2001. Retrieved from http://www.imsproject.org/content/packaging/cpv1p1p2/imscp_infov1p1p2.html
8. IMS-LD. IMS Learning Design. Version 1.0 - final specification. Retrieved April, 2005 from http://www.imsglobal.org/learningdesign/index.html
9. IMS-LIPS. Learner Information Package Specification. Version 1.1-final specification, 2001. Retrieved from http://www.imsglobal.org/profiles/lipinfo01.html
10. IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata Version 1. Retrieved April, 2005 from http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html
11. IMS Question and Test Interoperability: Information Model. Retrieved April, 2005 from http://www.imsglobal.org/question/index.html
12. Jochems, W., van Merrienboer, J., Koper, R. (Eds) Integrated e-Learning: implications for pedagogy, technology and organization. Routledge Farmer, London, 2004.
13. Koper (2001) Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind EML. Retrieved from eml.ou.nl/introduction/docs/ped-metamodel.pdf
14. Leidig, T. (2001) L3—towards an open learning environment, Journal on Educational Resources in Computing (JERIC) Volume 1, Issue 1es (March 2001), 7 Retrieved in July 2005 from http://doi.acm.org/10.1145/376697.376702
15. Mitzoguichi, R., Sinitsa, K. E Ikeda, M (1996) Task Ontology Design for Intelligent Educational/Training Systems. Workshop on Architectures for Designing Cost-Effective and Reusable ITSs, Montreal, Canada p 1-21
16. Molenda, Michael, Pershing, James A., and Reigeluth, Charles M. (1996). Designing Instructional Systems. In Robert L. Craig (Ed.), The ASTD Training and Development Handbook 4th ed. (pp. 266-293). New York: McGraw-Hill.

17. Murray T. (1996) "Special Purpose Ontologies and the Representation of Pedagogical Knowledge". International Conference for the Learning Sciences (ICLS-96). Evaston, Charlottesville(USA)
18. Murray T. (2003) Principles for Knowledge based Tutor Authoring Systems in Authoring tools for Advanced Technology Learning environments: Towards cost-effective adaptive, interactive and intelligent educational software. T. Murray, S. Blessing and S. Ainsworth (Eds) pp 439-467. Kluwer Academic Publishers, 2003
19. OLAT web-based Open Source Learning Management System (LMS) project website. Retrieved February 3, 2006 from http://www.olat.org/public/index.html
20. Padrón, C. L., Dodero, J.M., Aedo, I. and Díaz, P. (2005). The collaborative development of didactic materials. *Computer Science and Information Systems Journal* Volume 02, Issue 02 (December), 2005.
21. Padrón C.L. (2006). MD2 method: the Didactic Material Creation from a Model Based Perspective. In Proceedings of ECTEL06 Conference LNCS 4227, Crete, Greece, October 2006.
22. Paquette, G. (2003). Educational Modeling Languages, From an Instructional Engineering Perspective. http://www.licef.teluq.uquebec.ca/gp/fr/publications/documents/ArticleEML-MISA.doc
23. Polsani, P. (2003) "Use and abuse of reusable learning objects." *Journal of Digital Information*, Volume 3, Issue 4. (2003). Retrieved January, 2006 from http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/
24. RELOAD Reusable E-Learning Object Authoring and Delivery project Retrieved January 5, 2006 from http://www.reload.ac.uk/
25. Rodríguez-Artacho, M., & Verdejo Maíllo, M. F. (2004). Modeling Educational Content: The Cognitive Approach of the PALO Language. In Journal of Educational Technology & Society, 7 (3), 124-137
26. R. Sierra. Tesis doctorales y trabajos de investigación científica. Thomson editores Spain Paraninfo, S.A., 1986
27. SCORM. Sharable Content Object Reference Model (SCORM). 2004. Retrieved form http://www.adlnet.org/scorm/history/2004/documents.cfm
28. Teege, G. Targeteam: TArgeted Reuse and GEneration of TEAching Materials, Retrieved form http://www11.in.tum.de/forschung/projekte/targeteam/

Table 1. Comparison of MD2 contributions with different approaches to the development of didactic materials

Supported features/ Approach	**TeachML**	**IBM DALO**	**OUEML/ IMS LD**	**KBTMM**	**MD2**
Content representation	reusable unit or module/sub-module/atoms	Learning objects	Learning objects/QTI tests	Instructional units/Topics	Learning objects/QTI tests/ LD activities
Pedagogical Knowledge domain model	Pedagogical information not included	Pedagogical information not included	Not included	Knowledge types: fact, concept, principle	KD view elements with Pedagogical information: Learning outcomes to promote, Estimated time Topics
	Topics hierarchy/	Topics graphs/ Generalization and specialization relationships		Topics hierarchy/	Knowledge view: Topics hierarchy/
	Generalization and specialization relationships			Generalization and prerequisite relationships	Generalization and prerequisite relationships
Pedagogical strategy representation	Not supported	An explanatory strategy using SCORM	Common description of structure and behaviour for different pedagogical strategies: Exploratory: CL, PBL, Explanatory	Not supported.	Pedagogical and Support Views: Different pedagogical strategies using structure patterns based on IMS LD, QTI
Based on e-learning standards	e-learning standards are not supported	SCORM, IMS LOM	IMS LOM, IMS SS, IMS QTI/IMS LD	e-learning standards are not supported	IMS LOM, IMS SS, IMS QTI, IMS LD
Product Standard compliance	TeachML documents/ XML	SCORM, IMS LOM	IMS LD		IMS LOM, IMS QTI, IMS LD
Selection stage	Not supported	Topic, learning constrains, user preferences.	Not supported	Not supported	Using information provided by Knowledge and Pedagogical view elements
		Crawling mechanism			Filters definition
Supported	**TeachML**	**IBM DALO**	**OUEML/ IMS LD**	**KBTMM**	**MD2**

features/ Approach					
Composition stage/ Integration	Set of languages: Content structure Module integration Cross referencing Content	Using topics graph information	Using the structure provided by UOL	Decision Architecture	Using composition and integration patterns obtained from Pedagogical and Support Views
		Content sequencing rules based on the pedagogical strategy		Lessons/Topics- Topic levels/Presentation contents/Events	
Evaluation stage	Not supported	Contents Coherency using Topic graph and pedagogical strategy	Not supported	Not supported	Using Quality and Usability view elements and fuzzy aggregation mechanisms
Semantic annotations	Each module has annotations about integration and adaptation	Not supported	Considered as optional	Not supported	Values of MD2 view elements as extensions of IMS LOM
Design rationales	Not supported	Not supported	Not supported	Not supported	A log of all information used by MD2 method
Content Reusability	Same content represented in different formats	Different contents for the same pedagogical strategy	Same contents used for different pedagogical strategies	Not supported	Same contents used for different pedagogical strategies
Pedagogical strategy reusability			Same pedagogical strategy with different contents		Same pedagogical strategy with different contents

Technology Enhanced Life Long eLearning for All

Olga C. Santos[1]
[1] aDeNu Research Group, Artificial Intelligence Dept., Computer Science School, UNED, C/Juan del Rosal, 16, 28040 Madrid, Spain
ocsantos@dia.uned.es – http://adenu.ia.uned.es

Abstract. This paper presents the approach of my Ph.D work, which focuses on providing personalized learning taking into account not only the learners' background, interests and needs, but also accessibility requirements along life long eLearning. In particular, accessibility requirements should be addressed in the full life cycle of eLearning, but an initial research of the state-of-the-art has shown that currently there is no sufficient technological support. The approach of my research work implies the design, implementation and validation of methodologies and developments to show how to cope with these requirements and improve the effectiveness (in the sense of achievement of learning goals) of learning for all along the full life cycle of life long eLearning.

Keywords: Educational standards, accessibility requirements, learning systems, authoring tools, life long eLearning, personalization.

1 Research question and existing problems in the field

ICTs (Information and Communication Technology) claim to support the Life Long Learning paradigm integrating education and work in a continuous learning process. To successfully lead this process, [1] proposes a user-centred approach in which individual learning needs are taken into account. Moreover, to avoid building new social, physical or cognitive barriers, this process should also attend the functional diversity of learners. For this reason, since technology is playing an increasing role in mediating the learning,it has to deal with the special needs of the actors involved in the learning process to provide an effective life long learning for all (in the sense of achievement of the learning goals proposed).

International standards (IMS, SCORM and WAI) have been specified to facilitate reusability of learning designs, adaptation to learning needs and accessibility for all. Moreover, collaboration and communication facilities are demanded to support and motivate learners during the learning process. However, limitations in current authoring tools and learning environments to cope with all these requirements together make that this approach has not yet become a reality.

Therefore, taking into consideration the problems mentioned above, the following questions are to be worked in this research: 1) What special needs issues are involved in eLearning?; 2) How can personalized eLearning address these special needs?; 3) What can be done to support life long learning in this context?; 4) Which requirements should be met by eLearning technology to support these needs?

2 Problem domain and state of the art of existing solutions

To address LLL for all, the technology has to support authors, tutors and learners in their learning process over time. First, it has to provide tools for authors to help them specify the learning design they have in mind addressing the particular individual needs. This specification has to be reusable among different courses and platform independent. Second, learning environments are needed that can read these specifications and deliver the appropriate material to the learner, promote collaboration and involve the tutor in the learner support. Third, dynamic support should also be provided taking into account past and current interactions of the learners in the course. And all these functionalities have to be provided considering the special needs authors, learners or tutors may have.

Educational standards (SCORM, IMS) have been developed to respond the requirements of reusability and adaptability demanded by authors. The IMS Global Learning Consortium develops and promotes the adoption of open technical specifications for interoperable learning technology. The results of these specifications are XML files (imsmanifiest.xml) that serve as a vehicle between the author specifications and the learning environment configuration. Thus, no constraint or requirements on the technology are given. Different specifications have been defined, such as Learning Object Metadata (LOM), Content Packaging (CP), Learner Information Packaging Accessibility (LIP-Accessibility), Learning Design (LD), Question and Test Interoperability (QTI) and Enterprise. In turn, Sharable Content Object Reference Model (SCORM) is a collection of standards and specifications adapted from multiple sources to provide a comprehensive suite of e-learning capabilities that enable interoperability and reusability of web-based learning content. Unlike IMS specifications, SCORM describes the learning management system (LMS) requirements for managing the run-time environment (RTE) (i.e., content launch process, standardized communication between content and LMSs, and standardized data model elements) used for passing information relevant to the learner's experience with the content and even provides a working example of the run-time environment.

On the other hand, from the functional diversity standpoint, the W3C Web Accessibility Initiative provides guidelines for web content (WCAG), authoring tools (ATAG) and user agent (UAAG) accessibility. More specifically, WCAG explain how to make web content (e.g. web pages or web applications) accessible to people with disabilities; ATAG describes how to make authoring tools accessible to people with disabilities and how tools can help web developers produce web content that conforms to WCAG; UAAG explain how to make user agents (browsers, media players, assistive technologies) accessible to people with disabilities, particularly to increase accessibility to web content. There are also accessibility requirements specified by national governments, such as the American Section 508 of the Rehabilitation Act, the English SENDA or the German BITV.

The required compliance with educational standards, accessibility guidelines, collaboration facilities and dynamic support entails the necessity of having authoring tools and learning environments that provide them. However current developments do not fully support these requirements, as the following summarized review of the state-of-the-art shows.

2.1 Authoring Tools for course development

There exist already a great variety of editors (commercial and opensource) that produce SCORM and IMS compliant educational material, such as Reload, eXe, Respondus, CopperAuthor, Ask LDT, AltEd, CoSMoS, Mot Plus or aLFanet LD & QTI Authoring Tools. However, none of them complies with accessibility requirements. Search done so far has revealed that only two (commercial) products claim to produce accessible content for standards based educational material, CourseGenie and YAWC (Yet Another Word Converter) Online eLearning Edition.

Moreover, there exist web accessibility verifiers that ensure that the web pages generated by authoring tools are accessible to people using assistive technologies. For instance, A-Prompt examines web pages for barriers to accessibility, performs automatic repairs when possible, and assists the author in manual repairs when necessary; Imergo supports quality assurance methods by validating entire sites against standards and configurable accessibility rule-sets. More specific tools such as MAGpie, to create closed captions and audio/video descriptions are available as well.

2.2 Learning systems

The accessible and standard based specifications for the course material produced by the authoring tools have to be managed by the learning environments to provide all learners the author's design. Again, to not exclude people with special needs, the learning environments should comply with the accessibility requirements so learners can plug in the needed assistive technologies. The table below summarizes the educational standards and accessibility requirements the major learning platforms conform to, as well as the technology upon they are built.

Table 1. State of the art of major learning systems.

Platform	Tech.	SCORM	IMS	Accessibility
ATutor	PAM	1.2	-	WCAG 1.0 AA+
ANGEL	Windows	1.2	Ent.1.1	WCAG 1.0 A, S.508
BlackBoard	Win/Unix	1.2	MD,CP1.1.2,QTI1.2, Ent.1.01	WCAG 1.0 A, S.508
Design2Learn	Windows	1.2, 2004	Ent1.1,CP,QTI1.2,LOM	Section 508
dotLRN	OpenACS	1.2	MD1.2.1,CP1.1.2,QTI1.2.1Ent.1.1LD	WCAG 1.0 A
Docebo	PAM	1.2	-	WCAG 1.0 A, S.508
Moodle	PAM	1.2	-	Text based browser
webCT	Win/Unix	1.2	CP1.1.2,QTI1.1.2,Ent1.0.1	WCAG 1.0 A, S.508

2.3 Discussion

The review of the state of the art has shown that there is not a real support in terms of tools to address special needs during the learning process, and far less a technological framework that guarantees the global process from authoring to learners performing the course activities. There is a wide offer in authoring tools for SCORM and IMS (CP, QTI, LD). However, when analysed learning environments, SCORM support is more highly available than IMS since SCORM provides the requirements for building run time environment whereas IMS does not. Nevertheless, IMS specifications have a greater potential than SCORM in attending special needs issues. SCORM can be assimilated to IMS-CP. What makes really the difference is IMS-LD, which allows to specify different learning routes (level A) that are to be selected according to the values that defined properties (level B) have at runtime. The first IMS-LD engine that covered A, B and C levels was CopperCore, implemented for aLFanet project (see later) and disseminated in UNFOLD (IST-2003-507835). In relation with aLFanet and Coppercore, dotLRN is the only learning collaborative environment that has already implemented support for IMS-LD (thanks to E-LANE project) .

Furthermore, the tools analysed above were selected because they claimed both to support educational standards and accessibility requirements. However, there is a tricky issue that is usually not considered. If courses are assimilated to web content and we restrict to WCAG 1.0 (the current stable version) checkpoint 6.3 says that "Ensure that pages are usable when scripts, applets, or other programmatic objects are turned off or not supported. If this is not possible, provide equivalent information on an alternative accessible page." But SCORM RTE is based on a script language (javascript). To provide SCORM functionality (e.g. user tracking) interactive support is given, and it cannot be provided without the use of scripts or any programmatic object (and the same would happen with IMS). So, SCORM/IMS courses cannot be accessible? That is really a big problem if we want to deliver LLL for all.

There have been long discussions differentiating between web content and web application, defining user agents, etc. which have ended up in the reformulation of this requirement in the WCAG 2.0 (currently still draft version) which says that web content still conforms using user agents that only support the technologies that are in the baseline, where the baseline is defined as the set of technologies assumed to be supported by, and enabled in, user agents in order for web content to conform to these guidelines. This change is justified because at the time WCAG 1.0 was written assistive technologies did not support scripts, but today they do.

Then, how can current authoring tools and learning management systems say that they support both SCORM/IMS and WAI WCAG? Authoring tools justify themselves saying that they support both, but not simultaneously. Learning environments usually validate the accessibility on the content on the web page, but do not analyse the accessibility of the functionality they provide (forums, assessments, course delivery, etc.). As a consequence, contents may be accessible, but not the interaction with the platform services. Powerful development technologies (such as the one upon which OpenACS/dotLRN is built) should be used to guarantee that not only web pages but the functionality provided meet accessibility requirements.

3. Research works undertaken and results obtained

To provide life long eLearning for all, two research lines are being followed. Firstly, the whole learning process has to be supported by the technology and, secondly, special needs have to be addressed during this process. The first objective was dealt with in aLFanet project (IST-2001-33288), in which I was directly involved as technical coordinator. There, we developed a prototype that supports adaptive course delivery in the full life cycle of eLearning (Design, Publication, Use and Feedback) by means of a pervasive use of (IMS) standards [2]. We provided personalised eLearning thanks to the combination of learning design (defined by the author via IMS-LD templates) and run time adaptations (by applying machine learning and collaborative filtering techniques to the learners' interaction data) [3]. Although the effectiveness of this LMS was positively evaluated, we do not intend to move this prototype into a production software because the architecture is too complex [4], the adaptation module (which supports the adaptive dynamic support) needs to be further worked (within my PhD) and it did not consider learners' special needs. Moreover, currently an open source development (i.e. dotLRN) is working on this direction and we are benefiting from (as well as contributing to) this community of developers.

Now our research efforts are led towards integrating accessibility support into the full life cycle of eLearning. Therefore, we need to provide support in the four phases mentioned above. Specially, for the Design phase we need authoring tools that produce standard based educational material that can be understood by LMS and managed by the existing assistive technologies. And for the Use phase, we want to provide an adapted response to learners, which not only adapt the course flow to the learners' background, interest and learning needs (as aLFanet approach), but the user interface and course flow to their special needs. Standardization of user profiles are needed to provide adaptation in web services which take into account the functional diversity [5]. Currently I am involved in three research projects where learners' special needs are addressed to support their learning over time. Their objectives are diverse but complementary. FAA (funded by Galician government) focuses on defining the methodology and selecting the appropriate tools to supporting the life cycle of the learning process considering special needs. The other two are European projects funded by the EC to start next fall. ALPE (eTen) will market validate the approach with real users with functional diversity that lack some basic skills. EU4ALL (IP) aims to improve the effectiveness of accessible lifelong learning by developing an open service standards based architecture for all.

4 Contributions to the problem solution

My PhD work intends to contribute to the problem of how to provide life long eLearning taking into account the special needs learners may have from a combination of a pervasive use of standards and the application of user modelling and machine learning techniques.I am tackling this issue in my PhD work (in line with the research projects I am involved) as follows:

- Analyse what has to be considered at design time in order the assistive technologies can manage learning contents. As part of my PhD work, the LFanet methodology [6] is being extended to design reusable, platform independent, objective based and adapted courses for all. This approach goes beyond being conformant with accessibility guidelines (which has been proved not to be fully appropriate) and intends to assure real accessible support taking into account current requirements for assistive technologies.

- At an initial stage, we are focusing on the development of courses which comply with SCORM and IMS-QTI standards since there are available authoring tools and learning environments that support them. Considering IMS-LD from the beginning will lead us to conceptual and practical problems, as those found in aLFanet, because of the not yet maturity of the related technology.
- Dynamic support to address individual learners' special needs has to be provided. In this respect, I am extending the user modelling in the multi-agent architecture of the adaptation module with adaptation tasks that deal with the interface configuration of the LMS, and working on the user model co it can be dynamically built from the learners' interactions by applying machine learning techniques.
- Selection of the appropriate LMS. We have chosen dotLRN for the reasons revealed in the state-of-the-art research (built upon OpenACS framework, support collaboration and educational standards (SCORM, IMS) including an IMS-LD package, broadly used).
- Complement LMS functionality to support the life long eLearning for all. ePorfolios to review students' learning for educational or employment reasons are caching on within this paradigm. However, although it may seem just the opposite, addressing accessibility requirements in ePortfolios is not easy at all, for many reasons, such as the greater level of learning authoring.

Acknowledgments. I would like to thank my advisor Jesús G. Boticario for offering me the opportunity to research in these leading and challenging issues and very specially for sharing with me all the knowledge he has gathered.

References

1. eLearningPR. Adopting a multi-annual programme (2004-2006) for the effective integration of ICT in education and training systems in Europe (eLearning Programme).
2. Rosmalen, P. van, Boticario, J.G., Santos, O.C. "The Full Life Cycle of Adaptation in aLFanet eLearning Environment".Learning Technology newsletter. Vol:4.Oct (2004) 59-61
3. Boticario, J.G. and Santos, O.C. Issues in developing adaptive learning management systems for higher education institutions. ADALE Workshop. AH'06. Dublin (Ireland), June 2006.
4. Santos, O.C., Boticario, J.G., Barrera, C., aLFanet: An adaptive and standard-based learning environment built upon dotLRN and other open source developments. Foro hispano de .LRN y software libre educativo, Madrid, (Spain), May 2005.
5. Velasco, C. A., Mohamad, Y., Gilman, A. S., Viorres, N., Vlachogiannis, E., Arnellos, A., Darzentas, J.S. Universal Access to Information Services - the Need for User Information and its Relationship to Device Profiles. In: Gulliksen J, Harker S, Vanderheiden G (eds), Special issue on guidelines, methods and processes for software accessibility. Universal Access in the Information Society, 3 (1), pp. 88—95, 2004.
6. Santos, O.C. and Boticario, J.G. Meaningful pedagogy via covering the entire life cycle of adaptive eLearning in terms of a pervasive use of educational standards: the aLFanet experience. EC-TEL'06. Crete (Greece), October 2006.

Associative retrieval of resources for work-integrated learning: Integrating domain knowledge with content-based similarities

Peter Scheir[1]

[1] Know-Center,
Inffeldgasse 21a, 8010 Graz, Austria
pscheir@know-center.at

Abstract. The dissertation project presented here aims at identifying resources for work-integrated learning that are relevant to the current context of a knowledge worker. For finding suitable resources, methods of associative information retrieval are employed. The unified approach to search for useful material is based on a formal model of the domain the knowledge worker operates in and on statistical information gained from the content of the (text) documents in the document base of the knowledge worker. Further on the chosen approach to search allows for incorporating feedback of the knowledge worker regarding the usefulness of the presented material, which will be used for future search efforts.

Keywords: Work-integrated-learning, semantic web technologies, associative information retrieval

1 Formulation of the research question

The dissertation project presented here originates in the context of work-integrated learning: To enhance the productivity of knowledge workers, they have to be enabled to directly learn at their workplaces. In work-integrated learning the knowledge workers' actual tasks, personal competency disposition and work domain form the context of the learner, which is relevant for deriving his or hers current learning needs. The environment the knowledge worker operates in can be represented by formal domain models in the form of ontologies. These forms of knowledge representation conceptualize those entities of the worker's domain, which are relevant for work-integrated learning and capture relations between the domain's concepts by defining semantic dependencies between them. For example formal domain models in the domain of requirements engineering for building airplanes, would define which methods exist for requirements engineering (the learning domain) and which entities are involved in the process of manufacturing an airplane (the application domain). Formal models define the entities of a domain and the relations between them by using formalism, e.g. a special language.

The framework presented here is designed to identify resources for work-integrated learning that are related the learner's current context. To reach this goal, methods of associative retrieval [15] are employed for finding resources to be presented to the learner. Besides content-based similarities of (textual) resources also models of the domain the knowledge worker operates in are utilized, with the goal to further enhance the identification of material appropriate to the current situation of the learner. The research project presented here does not focus on the aspect of determining the user context and only implicitly addresses the creation of formal model(s)[1]. Instead the presented work starts from a given context of the learner and aims at finding the best matching resources to this context with the help of associative retrieval methods on domain models and the content of the resources.

Associative retrieval is a variation of *information retrieval*[2] [6], which tries to find relevant information by retrieving information that is by some means associated with information that is already known to be relevant. Information items which are associated can be (parts of) documents, extracted terms but also concepts originating form knowledge representations. Various information retrieval tasks can be modelled as associative retrieval problem, as e.g. document retrieval [8], fact retrieval [14] or recommender systems [22]. Earlier associative retrieval systems, for documents retrieval, created associations based on the similarity of the content of documents. Associative retrieval systems used for fact retrieval operated solely on semantic networks. Associative retrieval approaches combining both, content based similarities of (text-) document and knowledge representations rarely exist.

One of the biggest achievements of the Semantic Web [10] community so far is the development of sound mechanisms for creating knowledge representations in the form of ontologies (see [19] or [20] for exact definitions of the concept ontology). This knowledge representation formalism will be used in the presented research project, allowing for the encapsulation of domain knowledge in a separate structure. Separating domain knowledge (i.e. the domain models) form operational knowledge leads to higher adaptability of the system [28].

Both, knowledge representations as well as associative document retrieval systems, work by relating items. The central research hypotheses of the project presented here is that combing the two approaches into one model will enhance the performance of a search infrastructure for work-integrated learning resources.

2 Current knowledge of the problem domain

Early systems for associative document retrieval date back to the 1960s. [35] gives an overview on the first associative retrieval systems. [34] describes a system for *linear associative retrieval*, which is based on the vector space model and associates text-documents based in similarity of content. Later systems [8] [32] employ the *spreading activation* model for similarity search. The spreading activation model

[1] This does not mean that these two aspects are less important, they are addressed in separated research projects.

[2] Within this work information retrieval is seen as in [30], i.e. information retrieval can be document retrieval as well as fact retrieval.

originates from cognitive psychology (cf. [4]) where it servers as mechanism for explaining how knowledge is represented and processed in the human brain. The human mind as modelled as network of nodes, which represent concepts and are connected by edges. Starting from a set of initially activated nodes in the net, the activation spreads over the network [37].

Associative retrieval systems employing spreading activation associate information in a graph, which is often referred to as *associative network*. During search, 'energy' flows from a set of initially activated information items over the edges to their neighbours. The information items with the highest level of energy are seen to be the most similar to the set of nodes activated initially. (See [15] for a detailed introduction to spreading activation in information retrieval.) One of the first works in spreading activation in information retrieval is [29]. In [14] a set of constraints for better control over the spread of activation is introduced, coining the term *Constrained-Spreading-Activation* for this approach. Mandl [26] concludes from the good performance of the spreading activation based systems [11] and [24] at the TREC[3], that the spreading activation model is comparable to other retrieval techniques in its performance.

Besides systems that use spreading activation for finding similarities between text documents or search terms and text documents, approaches exist, which employ spreading activation for finding similar concepts in knowledge representations [2] [9] [14] [31]. For example [9] describes a tourist information system, that's underlying knowledge base is searched via spreading activation.

Currently no approaches to associative retrieval are known the author, that combine the two methods described above, into one model, i.e. that provide information based on content-based similarities and a knowledge representation. A step into this direction was I3R [17]: A network structure was introduced allowing for connecting documents, extracted terms and concepts from a semantic network. In the presented prototype only content-based similarity of text documents and similarities based on the same author and co-referenced works were used. In [1] a similar network structure is suggested. It remains unclear to the author how in both approaches content-based similarity and the semantic layer are connected.

3 Significant problems in the field of research

3.1 Relating ontology and resources

The approach presented here combines two sources for determining the similarity between learning resources. On the one side content-based similarity of resources is used, while on the other side similarities stem from a knowledge representation. To integrating the two sources, they are transformed into a network representation. The resulting architecture can be represented with three layers, a layer for (parts of) documents, a layer for terms extracted from documents and a layer for concepts

[3] Text Retrieval Conference - http://trec.nist.gov/

originating from the ontology used (see Fig. 1). After the creation of the three layers the challenge of combining the already well integrated document and term layer with the concept layer exists (symbolised by the dashed lines in Fig. 1). While the document and term layers stem from resources present in the organisational memory, the concept layer originates from the knowledge representation(s) used. Per se those two tiers are not connected, as the domain models are created by experts for a special purpose, while the resources in the organisational memory are created by the workers in the company during their daily job. The concepts from the domain models can be seen as metadata to the documents [25]. Because manually assigning this metadata to the documents in the organisational memory is a burdensome task, connecting domain knowledge with resources raises an interesting challenge. A similar problem is present when trying to establish a semantic web: A lot of resources already exist in the current form of the web, but most of them are not annotated semantically. Together with the high number of resources already present on the web (or in a company) this is one of the major obstacles in establishing a broad use of semantic technologies. Therefore this research project will precisely observe current and future developments in fields such as ontology learning [12] and the (semi-)automatic creation of semantic metadata [21].

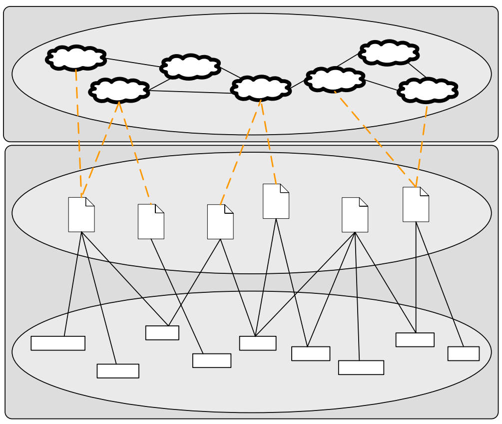

Fig. 1. The three layer architecture of the model.

3.2 Spreading activation

For searching the network representation, a graph based approach named spreading activation shall be used. While the basic spreading activation algorithm is straightforward, complex problems can arise in action. Typical problems that can occur are thinning out of activation, resulting from graphs with many edges, and over activation, which can easily happen in cyclic graphs. While variations of spreading activation are used on a regular basis, details regarding the implementation of the used algorithms are rare.

Pure spreading activation is a numerical method, which calculates a level of activation for a node depending on the activation levels of its neighbour nodes and the associations' strengths of the edges connecting the nodes. Semantic relations between concepts are only taken into account, by assigning a different numerical value to edges in the associative network, depending on the semantic type they have in the original knowledge representation. A variation of pure spreading activation is marker passing (cf. [38]). There instead of numerical values, more complex data-structures are passed between nodes, which allows for more control in the processing of the network structure. This is the way, which shall be used in the research project presented here.

3.3 Learning by relevance feedback

Associative networks relate information, which results in a weighted graph. In this property the model presented here is similar to what is commonly understood as *neural net* (cf. [7] for a differentiation between neural and semantic nets). [18] defines neural nets as a form of connectionist models, which are used for modelling information processing in the nervous system. Connectionist models are understood as a more general concept: "*they include several related information processing approaches, such as artificial neural networks, spreading activation models, associative networks, and parallel distributed processing*". Following this definition the model presented here is seen as a connectionist approach.

A characteristic of connectionist approaches in information retrieval is the possibility of influencing the results of a search through the modification of the underlying graph. By adding new edges or changing the weight of existing ones the system can 'learn'.

In the work presented here this ability to learn shall be used to include user feedback regarding the usefulness of the retrieved resources into the system, to influence future search results. User feedback in this form is known as *relevance feedback* (cf. [36]). Besides the symbolic model of the domain in form of an ontology a sub-symbolic model of the domain is created trough the feedback of the system's users (the system presented in [16] aims at a similar goal).

Resulting from the network structure of the proposed model, learning algorithms developed for connectionist systems look promising for the integration of relevance feedback into the proposed model. For example the approaches presented in [8] or [23] use a variation of *hebbian learning* (cf. [27]) for modifying the underlying network structure.

4 Preliminary ideas, proposed approach, results achieved so far

Starting with the goal to develop a system for finding resources appropriate to the current context of a learner, the first step was a literature study covering topics as associative retrieval, associative networks, spreading activation or the determination of similar concepts in an ontology. Yielding the hypothesis that retrieval performance could be improved by integrating domain knowledge with similarity by content. To be able to combine both aspects into one model and to allow for machine learning based in relevance feedback, a network structure was proposed. This conceptual phase was followed by the design and the implementation of two prototypes. Both artefacts are products of the design research method and went trough the process steps shown in Figure 2. The development of the artefacts served for concretising the research hypothesis.

The first artefact performs associative retrieval on a collection of text documents. Staring from a set of text documents, other documents are found, that are similar to the initial set. For this, terms that are part of both the initial set and the result set of documents are used. While this is a typical functionally of current text search engines, the prototype models the relations between documents and terms as associative network and searches related documents via spreading activation

The second artefact operates on a knowledge representation. It transforms an ontology into a associative network and allows for finding related concepts to an initial set of concepts. Concepts in the ontology result in nodes in the network. Relations between concepts in the ontology are transformed to edges in the associative network. While the relations in the ontology are typed, in the associative network an additional weight is added by analyzing the structure of the ontology, which represents the similarity of the two concepts participating in the relation. Also in this prototype spreading activation is used for search.

A final goal for this research project is the integration of the two artefacts described above. While the first artefact operates on the term and document layers of Figure 1, the other addresses the concept layer. To combine the two artefacts into a larger framework, the ontology layer and the documents/term layers have to be interrelated. A first approach to this challenge was the development of a third artefact for the semantic annotation of documents. In this artefact, which was realized as plug-in for the ontology editor Protégé, the ontology engineer is enabled to assign documents to a set of concepts. To increase the efficiency of the annotation process, a text-classification algorithm is included, which suggests concepts for (text-) documents to be annotated. The classification is based on previously created document to concept relations.

5 Applied research methodology

The research methodology applied in the work presented here is *design research*. It aims at the design, construction and evaluation of a human created artefact: *"Design research involves the analysis of the use and performance of designed artifacts to* *understand, explain and very frequently to improve on the behavior of aspects of Information Systems. Such artifacts include - but certainly are not limited to - algorithms (e.g. for information retrieval), human/computer interfaces and system design methodologies or languages."* [38]

In the flowing the process steps of design research (cf. Fig. 2) are explained in detail:

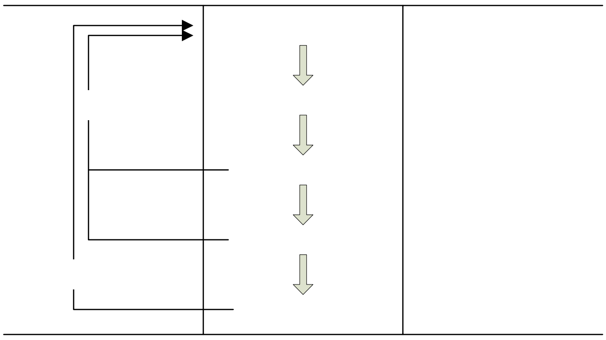

Fig. 2. The Design Research methodology.

There can be various sources for the *awareness of the problem*. Examples are a study of literature or the task of designing a concrete artefact. In the *suggestion phase* a tentative design of an artefact is created. New functionality is designed, based on the combination of new and already existing. In the *development* phase a concrete artefact is created from the tentative design of the previous step. Traditional development methods are used, as the innovation lies in the design of the artefact and not in its implementation. Depending on the criteria defined in the proposal an *evaluation* of the artefact is carried out. While in the positivistic research approach the research hypothesis is verified or falsified, in design research the results of the evaluation are used to refine the beforehand fuzzily defined research hypothesis. Often this step leads to changes in the design of the artefact. The research task ends with the *conclusion*, when the results of the previous steps are meaningful enough to assess the hypothesis. This can also be the case when there are still uncertainties over the behaviour of the artefact. The final assessment can be positive (the artefact confirms the hypothesis) or negative.

Opposite to the classical positivistic research approach in design research the hypothesis is refined during the act of research. Antonsson [5] notes, that this fact is not a real antagonism as also in the positivistic approach preliminary work is needed before the concrete formulation of the hypothesis.

6 Contribution to the problem solution

The research presented here is part of the APOSDLE[4] project, where the presented framework for associative retrieval shall be used to identify resource for work-integrated learning that fit the current work and learn context of the user. To maximize the possibilities of identifying useful material, content-based similarities are combined with a semantic model of the application domain.

The contribution of the dissertation project presented here addresses the topics described in Section 3:

- Research in the field of spreading activation will be carried out, resulting in a literature study and implementation details for different spreading activation variants.
- A methodology for associating knowledge representations with document collections will be designed and evaluated.
- A framework for associative search of resources for work-integrated learning based on document similarities and a knowledge representation will be designed and implemented. The evaluation of the created artefact will lead to new knowledge in this field.
- The developed artefact will allow for adaptation of its internal model that is used for associative retrieval, based in relevance feedback of users of the artefact.

7 How is the suggested solution different, new, or better?

The work presented here is a new approach in the context of associative retrieval and work-integrated learning. If knowledge representations where used in associative retrieval systems, this only happened implicitly (e.g. as in [8] or [17] via identical authors of a document). The **knowledge representation** was statically implemented into the system and not **easily changeable**, as it is in case when using an ontology [28].

Advantages of the employed approach of associative search based on associative networks and spreading activation are (cf. [3]): The **independence of** the **content type** of objects present in the associative network (but different similarities measures have to be defined) and the **robustness to missing information.** The system allows for **ostensive** [13] retrieval (a form of retrieval where a user not explicitly formulates a query for a search, but selects material that currently is useful for him), as needed for the user context based approach. Additionally connectionist approaches (as the one presented here) feature the characteristics [33] of: **adaptivity**, **distributability** and **parallelizability**.

Acknowledgments. The author thanks Dr. Stefanie N. Lindstaedt and Dr. Klaus Tochtermann for their ongoing support of the doctoral work presented here.

[4] http://www.aposdle.org/ Advanced Process Oriented and Self-Directed Learning Environment, Integrated Project funded under FP6, Call 4 in the strategic objective “Technology Enhanced Learning”

This work has been partially funded under grant 027023 in the IST work programme of the European Community. The Know-Center is funded by the Austrian Competence Center program Kplus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (http://www.ffg.at), by the State of Styria and by the City of Graz.

References

1. Agosti, M., Melucci, M., Crestani, F.: Automatic authoring and construction of hypermedia for information retrieval. Multimedia Syst. 3(1) (1995) 15–24
2. Alani, H., Dasmahapatra, S., O'Hara, K., Shadbolt, N.: Identifying communities of practice through ontology network analysis. IEEE Intelligent Systems **18**(2) (2003) 18–25
3. Alves, M.A., Jorge, A.M.: Minibrain: a generic model of spreading activation in computers, and example specialisations. In: ECML/PKDD 2005 workshop Subsymbolic paradigms for learning in structured domains. (2005)
4. Anderson, J.R.: A spreading activation theory of memory. Journal of Verbal Leaning and Verbal Behaviour **22** (1983) 261–295
5. Antonsson, E.K.: Development and testing of hypotheses in engineering design research. ASME Journal of Mechanisms, Transmissions, and Automation in Design **109** (1987) 153–154
6. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1999)
7. Barnden, J.A., Lee, M.G., Viezzer, M. In: Semantic networks. MIT Press, Cambridge, MA, USA (2002) 1010–1013
8. Belew, R.K.: A connectionist approach to conceptual information retrieval. In: ICAIL '87: Proceedings of the 1st international conference on Artificial intelligence and law, New York, NY, USA, ACM Press (1987) 116–126
9. Berger, H.: Activation on the Move: Adaptive Information Retrieval via Spreading Activation. PhD thesis, Technische Universitat Wien (2003)
10. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001) http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 (29. 3. 2005).
11. Boughanem, M., Dkaki, T., Mothe, J., Soule-Dupuy, C.: Mercure at trec7. In: The Sevent Text REtrieval Conference (TREC-7). (1999)
12. Buitelaar, P., Cimiano, P., Magnini, B., eds.: Ontology Learning from Text: Methods, Evaluation and Applications. Volume 123 of Frontiers in Artificial Intelligence and Applications Series. IOS Press, Amsterdam (2005)
13. Campbell, I., van Rijsbergen, C.J.: The ostensive model of developing information needs. In: Proceedings of COLIS-96, 2nd International Conference on Conceptions of Library Science, Kobenhavn, DK (1996) 251–268
14. Cohen, P.R., Kjeldsen, R.: Information retrieval by constrained spreading activation in semantic networks. Inf. Process. Manage. 23(4) (1987) 255–268
15. Crestani, F.: Application of spreading activation techniques in information retrieval. Artif. Intell. Rev. 11(6) (1997) 453–482
16. Crestani, F., Rijsbergen, C.J.V.: A model for adaptive information retrieval. J. Intell. Inf. Syst. 8(1) (1997) 29–56
17. Croft, W.B., Lucia, T.J., Cohen, P.R.: Retrieving documents by plausible inference: A priliminary study. In: SIGIR '88: Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM Press (1988) 481–494

18.Doszkocs, T., Reggia, J., Lin, X.: Connectionist models and information retrieval. Annual Review of Information Science & Technology 25 (1990) 209–260
19.Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N., Poli, R., eds.: Formal Ontology in Conceptual Analysis and Knowledge Representation, The Netherlands, Kluwer Academic Publishers (1993)
20.Guarino, N.: Formal Ontology and Information Systems. In: International Conference On Formal Ontology In Information Systems FOIS'98, Trento, ITALY, Amsterdam, IOS Press (1998) 3–15
21.Handschuh, S., Staab, S., eds.: Annotation for the SemanticWeb. IOS Press (2003)
22.Huang, Z., Chen, H., Zeng, D.: Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Trans. Inf. Syst. 22(1) (2004) 116–142
23.Kwok, K., Grunfeld, L.: Trec-4 ad-hoc, routing retrieval and filtering experiments using pircs. In Harman, D.K., ed.: The Fourth Text REtrieval Conference (TREC-4). (1996)
24.Kwok, K.L.: Query modification and expansion in a network with adaptive architecture. In: SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM Press (1991) 192–201
25.Lux, M., Tochtermann, K., Granitzer, M.: Retrieval basierend auf Semantischen Metadaten. In Lehner, F., Noesekabel, H., Kleinschmidt, P., eds.: Multikonferenz Wirtschaftsinformatik 2006. Volume 2., Passau, Germany, Gito-Verlag (2006)
26.Mandl, T.: Tolerantes Information Retrieval. Neuronale Netze zur Erhöhung der Adaptivität und Flexibilitt bei der Informationssuche. PhD thesis, University Of Hildesheim (2001)
27.Medler, D.: A brief history of connectionism (1998)
28.Noy, N.F., McGuinness, D.L.: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 (2001) http://www.ksl.stanford.edu/people/dlm/papers/ontologytutorial-noy-mcguinness.pdf (2006-05-04).
29.Preece, S.E.: A spreading activation network model for information retrieval. PhD thesis, University of Illinois at Urbana-Champaign (1981)
30.Raphael, B.: Sir : Semantic information retrieval. In Minksy, M., ed.: Semantic Information Processing. MIT Press, Cambridge, MA (1968) 33–145
31.Rocha, C., Schwabe, D., Aragao, M.P.: A hybrid approach for searching in the semantic web. In: WWW '04: Proceedings of the 13th international conference on World Wide Web, New York, NY, USA, ACM Press (2004) 374–383
32.Rose, D.E., Belew, R.K.: Legal information retrieval a hybrid approach. In: ICAIL'89: Proceedings of the 2nd international conference on Artificial intelligence and law, New York, NY, USA, ACM Press (1989) 138–146
33.Rumelhart, D.E., McClelland, J.L., eds.: Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations. MIT Press, Cambridge, MA, USA (1986)
34.Salton, G.: Automatic Information Organization and Retrieval. McGraw Hill (1968)
35.Salton, G.: Associative document retrieval techniques using bibliographic information. J. ACM 10(4) (1963) 440–457
36.Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41 (1990) 288–297
37.Sharifian, F., Samani, R.: Hierarchical spreading of activation. In Sharifian, F., ed.: Proc. of the Conference on Language, Cognition, and Interpretation, IAU Press (1997) 1–10
38.Vaishnavi, V., Kuechler, B.: Design research in information systems (2005) http://www.isworld.org/Researchdesign/drisISWorld.htm (25.09.2005).
39.Wolverton, M.: An investigation of marker-passing algorithms for analogue retrieval. In Veloso, M., Aamodt, A., eds.: Case-Based Reasoning Research and Development, Berlin,, Springer (1995) 359–370

Community Hypermedia in Collaborative and Self-reflective E-learning Applications

Marc Spaniol

Lehrstuhl Informatik V, RWTH Aachen University, 52056 Aachen, Germany
spaniol@i5.informatik.rwth-aachen.de

Abstract. The success of hypermedia enabled community applications depends on a careful design of the digital media and the related communication/collaboration tools. The success of community hosting for professionals and scientists in engineering disciplines or the humanities depends on the community engine's capabilities to reflect the discursive hypermedia knowledge contained in cross media sets drawings, animations, pictures, digital video, text etc. However, the semantics of the multimedia contents in community communication and collaboration is hard to capture and complex to compute. The rapidly changing needs of the community being hosted, the trend for multidisciplinary work and research, as well as the challenges of sophisticated multimedia technologies demand novel approaches for flexible, evolving, adaptable, and interoperable community engines. E-learning applications therefore need to reflect the nature of the underlying community processes and their discourses.

My work describes an approach of a self-reflexive information system architecture called ATLAS (Architecture for Transcription, Localization, and Addressing Systems), which is capable to support communities by multimedia services on the basis of the multimedia content description interface MPEG-7. With the opportunities given by the combination of metadata descriptions standards like MPEG-7 and server-side content-based computations, the manageability of multimedia semantics in community engines becomes more feasible. As a proof of concept, the Lightweight Application Server (LAS) for MPEG-7 Services will be introduced. As a proof of concept I present selected e-learning applications for communities of practice on the basis of ATLAS.

Keywords: E-Learning, Community Hosting, Multimedia Management, MPEG-7, Web-Services

1 Introduction

The development of e-learning applications for communities of practice [15, 16, 20] in different application domains is a challenging issue for several reasons. Principles like legitimate peripheral participation, group knowledge, situated learning, informality and co-location have to be taken seriously in the design of the community engine. For that reason, the community engine has to reflect the social learning processes taking place, which differ from community to community. Even more, communities are usually not able to express their needs in the very beginning of e-learning application usage. Thus, the communities have to gain experiences "their own" while applying the systems in use. Additionally, the multimedia technology is rapidly developing, thus creating new requirements on hardware and network capabilities. Consequently, the research question is: How to bring system developers together in order to "jointly" develop e learning applications?

My research is an approach to create a self-reflexive information system architecture that serves as a community engine for e learning applications. In order to let communities decide about the features of their community portal while constantly monitoring the impact of these activities onto the community, ATLAS (cf. figure 1) is being applied in a community-centered development process. ATLAS therefore combines approaches from various disciplines (such as software engineering, sociology and cultural sciences) led by the idea of transcriptivity as a design principle of computer science [11, 12]. The development process builds on ideas of joining usability and sociability [17] by constantly assessing and supporting community needs. In its reflective conception the socio-technical information system developed on the basis of ATLAS are tightly interwoven with a set of media-centric self-monitoring tools for the communities served. Hence, communities can constantly measure, analyze and simulate their ongoing activities. Closing the cycle, communities can thus better access and understand their community needs leading to a tighter collaboration between systems developers and communities. Even more, the server architecture of ATLAS allows communities to add, remove and exchange services in their community portal during runtime at anytime and from anywhere based on a Lightweight Application Server (LAS).

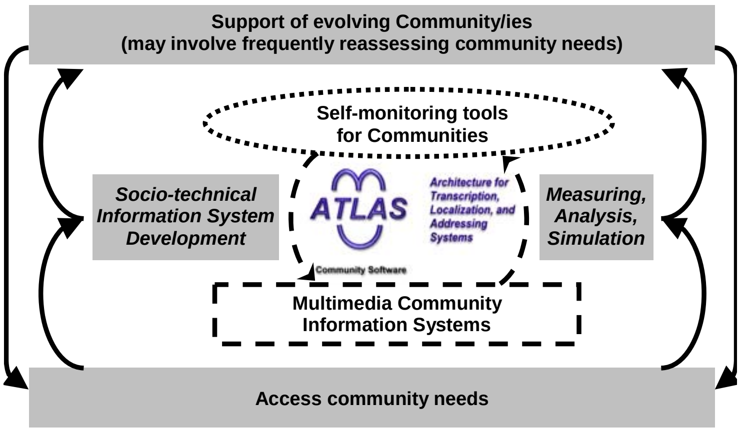

Fig. 1. Self-reflexive Information System Architecture ATLAS.

In the following, I introduce related work on community engines and compare the key concepts of LAS with these approaches. In order to proof the flexibility and adaptability of my approach, I present several communities hosted on the basis of ATLAS. My project description closes with a discussion.

2 Related Work

The main goal of a community engine for e-learning applications is to host a huge number of users in one or more learner communities. A core concept is a Lightweight Application Server for MPEG-7 Services that is capable to host a large number of communities at once and where users (agents) may have a variety of rights in several different communities. By applying interoperable multimedia metadata standards such as MPEG-7 [10] the exchangeability and interoperability of e learning contents is being improved. Furthermore, the structure of the server software should be able to run a heavyweight XML based community engine as well as a very lean hardware saving offline server on mobile devices for later synchronization in a later phase of the development. This and a clean service oriented structure keeping all business logic on server side is the main task of the LAS.

While LAS does not focus front-end generation facilities like a JSP container or a servlet engine itself, it focuses on server-side aspects mostly. Thus, LAS is much easier to configure than application servers like **Tomcat** or **ZOPE/PLONE**. In addition, it provides a very simple API to allow service developers to define access restrictions to their service methods. In addition to the described features it is possible to install, remove and replace elements (i.e. services, components and connectors) of LAS at runtime via the standard access methods. This allows server administrators to switch to the newest software versions without affecting the rest of the system. Table 1 forms a summary of the features described above by comparing LAS with the well known application servers **Tomcat** and **ZOPE/PLONE**.

Table 1. Comparison of LAS with related approaches.

	LAS	**Tomcat**	**ZOPE/Plone**
Configuration	simple, at runtime, through service	complicated, at startup	at runtime, through admin backend
Element maintenance (replacement, installation)	simple, through service	complicated, through special servlet	via admin backend
MPEG-7 compliance	fully by the LAS MPEG-7 services	not yet	not yet
Hardware needs	small to high	high	high
Flexibility	high	middle	small
User / Community management	very good	little	middle
Front-end generation	no	jsp and servlets, struts, jsf, etc.	yes

In the following, I introduce the technical key concepts of LAS in more detail. In addition, several communities hosted on the basis of ATLAS will be presented.

3 Collaborative and Self-reflective E-learning Applications based on ATLAS

The server architecture of ATLAS based on LAS, which is a platform independent Java implementation that can be flexible (re-)combined among various tools and communities. The open LAS Java API and its concepts can be used to extend the server's functionality based on a community's specific needs. Core services cover user and community management, management of access control lists, security and authorization management. A variety of databases and internet services like FTP are connected as well. For collaborative support we use the BSCW system [3] which is connected via XML-RPC. In order to support multimedia-centric community hosting, LAS provides a set of MPEG-7 services that offer methods to create, retrieve, update and delete persistent XML based MPEG-7 documents. These MPEG-7 services use the Apache XMLBeans XML Binding framework [2] for convenient navigation on MPEG-7 documents. Each service derived from this class inherits support for accessing an eXist MPEG-7 database and for using the MPEG-7 binding class library generated from the MPEG-7 XML-Schema [6]. Hence, LAS MPEG-7 services only allow valid operations to be performed on MPEG-7 documents. The MPEG-7 documents of LAS are stored in a native eXist database [8] that can be accessed by an MPEG-7 service via a built-in LAS component eXistConnector. In addition, retrieval services are provided by XQuery [4] and/or Xpath [5] expressions on the MPEG-7 database, while updates are realized by XUpdate [14] expressions. The e-learning applications show the applicability of LAS in versatile areas of applications.

3.1 MECCA & MEDINA

MECCA represents a multi-dimensional multimedia screening and classification platform. Its basic idea has been taken from the original project "Berliner sehen" developed at the MIT [9] and the Virtual Entrepreneurship Lab [13] developed by RWTH Aachen and Fraunhofer FIT. MECCA is specially designed for community hosting in multimedia-centric, interdisciplinary knowledge exchange. Thus, it serves as a media classification and monitoring system for a user community distributed among various universities in Germany. In addition, MECCA is used for web-based multimedia screenings of previously annotated media to supplement rare on-campus screenings. The scientific user community of MECCA has diverse educational backgrounds, such as film studies, history of art, graphical design and are on diverse levels of profession, i.e. full professors, research assistants, and students. Therefore, the interpretation of multimedia contents differs depending on the individual researcher's context. Figure 2 shows the front end of MECCA used for multimedia screenings based on previously annotated media. The categorization schema on the right allows metadata mediated browsing by switching between digital contents and contexts. After the start of MECCA in 2002 the community defined six main classification categories and 30 subordinate categories to classify a core set of initially 46 multimedia samples. In addition to the classification information, a wide variety of multimedia metadata are stored. Because of its multimedia-centrism and to ensure exchangeability of the contents stored in MECCA, the system is based on the LAS MPEG-7 services.

Fig. 2. User interface of MECCA applied in scholarly learning communities.

The MPEG-7 Encoding of Dublin Core Information and Naming Application (MEDINA) has been developed in order to support collaboration in communities by the exchange of multimedia contents and their low- and high-level semantic descriptions [19]. In its conceptualization, MEDINA tries to bridge the gap between “folksonomy-style” high-level semantic information about multimedia and purely technical low-level content descriptions. In this aspect, the Dublin Core (DC) metadata standard [7] has been a step forward, as it is an easy to understand and concise method for resource description on the web. However, this standard is not designed to describe temporal and media specific information connected with multimedia resources in general. In order to overcome these limitations the benefits of a loose classification scheme as in DC with the more sophisticated description elements for time based media in MPEG-7 need to be combined. Thus, MEDINA is based on an excerpt of the extensive MPEG-7 multimedia metadata standard with an integrated semi-automatic DC to MPEG-7 conversion functionality. MEDINA offers two options to annotate multimedia artifacts. First, a new multimedia annotation can be created from scratch. Second, a DC annotation might be uploaded for an automated conversion to MPEG-7. The mapping between the DC Metadata Element Set [1] and the MPEG-7 multimedia metadata standard is as described in [19]. In addition, an affiliated FTP-server is used for an automated up- and download of multimedia artifacts by the community to the common repository.

3.2 VEL 2.0

The Virtual Entrepreneurship Lab 2.0 (VEL 2.0) is the further development of its predecessor, the VEL [13]. Since entrepreneurship is a domain that has to be considered from various viewpoints and can be solved in many different ways, the problem of starting an enterprise does not involve one distinct solution. Hence, entrepreneurial knowledge is hard to formalize and best presented by success (or even failure) stories. For that purpose, a key point of the VEL 2.0 is its non-linear story-telling functionality, which is based on the Movement Oriented Design (MOD) [18] principle. MOD is a novel methodology and formalism in order to create multimedia stories by combining three facets of stories: Motivation (verbal and non-verbal knowledge), Exigency (semantic knowledge) and Structure (episodic knowledge). VEL 2.0 again contains LAS MPEG-7 services as described before. However, VEL 2.0 contains additional LAS MPEG-7 story-telling services that provide features in order to create MOD compliant non-linear multimedia stories. Thus, its user interface allows the authoring and consumption of non-linear multimedia stories, by an editor and a player (cf. figure 3). The contents themselves are media and media metadata of interviews with different entrepreneurial players from high-tech companies like Sun Microsystems, MetaCreations and ByteBurg. The interviews deal with different aspects of entrepreneurial activities like finding the right team, identifying the right opportunity, looking for start-up money etc. These issues can be temporally arranged as they depend on a distinct stage of a funding and can be associated with problems to be solved. When creating a story the author can then create paths covering different problematic aspects of entrepreneurship from the entrepreneurial interviews. Thus, the problems addressed depend on the path selected and lead consequently to different results in an entrepreneurship.

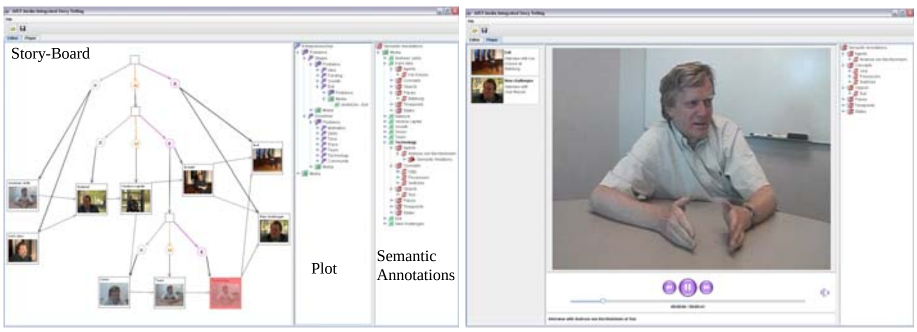

Fig. 3. Non-linear Multimedia Story (left) and Multimedia Story Player (right).

4 Discussion

Like the creation of new knowledge is a discursive and multistage process, the user requirements are rapidly changing and several new features need to be integrated into e-learning applications. In contrast to existing implementations the methodology and architecture of ATLAS is more flexible to assess the community needs over time and to integrate the community members in the development process. Even more, the multimedia services in ATLAS based on MPEG-7 provide interoperability and exchangeability of learning contents. Thus, ATLAS simplifies the community support process for the communities of practice drastically and on the same time offering more influence on the development process. However, the direct support of computer scientists and community designers is still needed. In future, graphical editing support for community web sites could leave even more responsibility on the community side.

Acknowledgments. This work was supported by German National Science Foundation (DFG) within the collaborative research centers SFB/FK 427 “Media and Cultural Communication” and by the 6th Framework IST programme of the EC through the Network of Excellence in Professional Learning (PROLEARN) IST-2003-507310. I would like to thank my colleagues for the inspiring discussions.

References

[1] ANSI/NISO Z39.85-2001: The Dublin Core Metadata Element Set. http://www.niso.org/standards/resources/Z39-85.pdf, September 10, 2001.
[2] Apache XMLBeans XML Binding Framework: http://xmlbeans.apache.org/, 2006.
[3] W. Appelt: WWW based Collaboration with the BSCW System. *Proc. of SOFSEM '99, Milovy, Czech Republic, LNCS 1725, Spinger Berlin-Heidelber*g, 1999, pp. 66 - 78.
[4] A. Berglund, S. Boag, D. Chamberlin, M. F. Fernández, M. Kay, J. Robie, J. Siméon: *XML Path Language (XPath) 2.0, W3C Candidate Recommendation,* http://www.w3.org/TR/xpath20/, 2005.
[5] S. Boag, D. Chamberlin, M. F. Fernández, D. Florescu, J. Robie, J. Siméon: XQuery 1.0: An XML Query Language. *W3C Recommendation, 3rd November 2005,* http://www.w3.org/XML/Query/, 2005.
[6] A. Brown, M. Fuchs, J. Robie, P. Wadler: XML Schema: Formal Description. *W3C Working Draft*, http://www.w3.org/TR/2001/WD-xmlschema-formal-20010320/, 2001.
[7] Dublin Core Metadata Initiative. http://dublincore.org/, 2005.
[8] Open Source Native XML Database eXist: http://www.exist-db.org/, 2006.
[9] K. Fendt.: Contextualizing content. *In: M. Knecht & K. von Hammerstein. (Eds.), Languages across the curriculum, Columbus, OH*, 2001, pp. 201-223.
[10] ISO/IEC: Information technology - Multimedia content description interface - Part 5: Multimedia description schemes. *Intl. Organization for Standardization*, 2003.
[11] M. Jarke, R. Klamma: Transkriptivität als informatisches Designprinzip. Mediale Spuren in rechnergestützten Entwicklungsprozessen. *In: G. Fehrmann, E. Linz, C. Epping-Jäger (eds.): Spuren Lektüren. Praktiken des Symbolischen, Festschrift für Ludwig Jäger zum 60. Geburtstag, München: Fink*, 2005, pp. 105-120.
[12] J. Jäger: Transkriptivität. In: Jäger, L., Stanitzek, G. (Eds.): Transkribieren - Medien/Lektüre. *In: Wilhelm Fink Verlag*, Munich, 2002 (in German).
[13] R. Klamma, E. Hollender, M. Jarke, P. Moog, V. Wulf: Vigils in a Wilderness of Knowledge: Metadata in Learning Environments. *In: Proceedings of IEEE Intl. Conf. on Adv. Learning Technologies (ICALT 2002), Kazan, Russia, September 9-12, 2002, IEEE Learning Technology Task Force*, 2002, pp. 519-524.
[14] A. Laux, L. Martin: XUpdate Language Specification. *XML:DB Working Draft,* http://xmldb-org.sourceforge.net/xupdate/xupdate-wd.html, 2000.

[15] J. Lave, E. Wenger: Situated Learning: Legimate Peripheral Participation. *Cambridge University Press, Cambridge, UK,* 1991.
[16] I. Nonaka, H. Takeuchi: The Knowledge-creating Company. *Oxford University Press, Oxford, UK,* 1995.
[17] J. Preece: Online Communities: Designing Usability, Supporting Sociability. John Wiley & Sons, 2000.
[18] N. Sharda: "Movement Oriented Design: A New Paradigm for Multimedia Design". *In: International Journal of Lateral Computing (IJLC), Vol. 1, No.1,* 2005, pp. 7-14.
[19] M. Spaniol, R. Klamma: MEDINA: A Semi-Automatic Dublin Core to MPEG-7 Converter for Collaboration and Knowledge Management in Multimedia Repositories. *In: K. Tochtermann, H. Maurer (Eds.): Proceedings of I-KNOW '05, 5th Intl. Conf. on Knowledge Management, Graz, Austria, June 29 - July 1, 2005, J.UCS (Journal of Universal Computer Science), Springer*, 2005, pp. 136-144.
[20] E. Wenger: Communities of Practice: Learning, Meaning, and Identity. *Cambridge University Press, Cambridge, UK*, 1998.