RSS-generated contents through personalizing e-learning Agents

C. De Maio, G. Fenza, V. Loia, S. Senatore
Dipartimento di Matematica e Informatica
Universita' degli Studi di Salerno
via Ponte don Melillo
84084 Fisciano (SA), Italy

M. Gaeta, F. Orciuoli
Dipartimento di Ingegneria dell'Informazione e Matematica Applicata
Universita' degli Studi di Salerno
via Ponte don Melillo
84084 Fisciano (SA), Italy

Abstract

Nowadays, the emphasis on Web 2.0 is specially focused on user generated content, data sharing and collaboration activities. Protocols like RSS (Really Simple Syndication) allow users to get structured web information in a simple way, display changes in summary form and stay updated about news headlines of interest. In the e-Learning domain, RSS feeds meet demand for didactic activities from learners and teachers viewpoints, enabling them to become aware of new blog posts in educational blogging scenarios, to keep track of new shared media, etc.

This paper presents an approach to enrich personalized e-learning experiences with user-generated content, through the RSS-feeds fruition. The synergic exploitation of Knowledge Modeling and Formal Concept Analysis techniques enables the definition and design of a system for supporting learners in the didactic activities. An agent-based layer supervises the extraction and filtering of RSS feeds whose topics are specific of a given educational domain. Then, during the execution of a specific learning path, the agents suggest the most appropriate feeds with respect to the subjects in which the students are currently engaged in.

1. Introduction

E-learning is defined as a wide set of applications and processes, which use available electronic media and tools to deliver vocational education and training. More recently, an alternative definition of e-learning defines it as a solitary, individual activity, or a collaborative group activity both synchronous and asynchronous communication modes may be employed. In this context, the diversity of students characteristics and background is one of the most important issues. Enrolled students usually come from many different places and they may have different linguistic, cultural, and academic background, hence the conventional e-course materials cannot always meet different students needs. Thus, we can affirm that 'one curriculum for all' is no longer suitable for the e-learning environments. The aforementioned statement suggests that a great expectation for personalized e-learning is raising [3].

Nowadays, the Semantic Web technologies are considered the most promising solutions to effectively organize and manage available e-learning resources according to peculiar necessities of both teachers and students. Brase and Nejdl in [4] have showed the increasing importance, in the e-Learning field, of knowledge modeling through metadata definition standards. However, these standards introduce the problem of incompatibility between disparate and heterogeneous metadata descriptions across domains, which might be avoided by using ontologies as a conceptual backbone in an e-learning scenar [5]. Several systems have been developed to handle learning resources by means of Semantic Web technologies. Another representative system that uses Semantic Web techniques in an e-Learning environment is the Courseware WatchDog [7]. The Semantic Web approach for the personalization of e-learning processes is tightly coupled with the availability of great volumes of reusable educational content. This is clear considering that having more available learning objects means that there are more opportunities to better satisfy the learners' preferences. The raising of numerous IEEE LOM-compliant

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.124

Learning Object Repositories (e.g. MERLOT[1] with 10607 public objects stored, eRIB[2] with 49761, EdNA Online[3] with 30300, etc.) can significantly improve the quality of the proposed vision. On the other hand, the exigency of an "intelligent" support is unquestioned due to decentralized infrastructure in e-learning systems [15]. Multi-agent systems for e-learning and skill management are scattering in many applications [13], [14], synergically integrated in semantic technologies [10]. Furthermore, in the last years the impulse carried out by the so called Web 2.0 [8] has involved also the e-learning field where new trends are rising. Tools like Blogs (used to share ideas), Wikis (used as a way to construct knowledge in a collaborative way), Podcast (used to distribute multimedia files over the Internet) and other Web Sharing Applications (e.g. Flickr, YouTube, del.icio.us, etc.) are exploited by Internet communities in order to work and make business but also to teach and learn. The coherent utilization of the aforementioned tools in e-learning processes is called e-learning 2.0.

2 Motivations

The personalized e-learning experiences could really improve the learning processes realized by means the Information and Communication Technologies [1] and could become more effective and efficient if there is the availability of a great number of educational content to be dynamically filtered and assembled with respect to learners' preferences and cognitive states. Our idea consists of exploiting the Web as a prominent source of educational content for supporting and improving personalized e-learning processes. In order to reach this goal, important issues will be faced in the next sections: (a) managing the several types of educational content the Web systems offer (*Interoperability*), (b) extracting from the Web and providing to students only educational content that are relevant with respect to the subjects currently treated within their e-learning experiences (*Contextualization*) and (c) driving the Web search activities (for relevant educational content) in order to identify promising Web zones in which to find educational valued content about a domain of interest (*Vertical Search*). In our approach we propose:

- The adoption of a standard publication language, like **RSS (Really Simple Syndication)** [8], as a "lingua franca" used to simplify the information management (extraction, filtering, classification, delivery, etc.) and to solve the *Interoperability* problem. This choice simplifies our architectural design assuming to (a) search only for Web Repositories and Sites that publish their content using RSS feeds, (b) use RSS fields values in order to perform filtering, classification, etc. and (c) exploit RSS feeds readers and aggregators [8] to delivery educational content to students.
- The introduction of a mathematical model based on the **Formal Concept Analysis** (briefly FCA) [16] which allows the *Contextualisation* of the educational content, customized to personal e-learning skills. Through the building of the relative lattice contexts, FCA enables the representation of the relationships between feeds and topics of learning objects. Thus, the system exhibits only content which matches the user experiences.

Finally, the distributed nature of the Web emphasizes the use of *Agent Technologies* as a backbone for the system interactions. In fact, the main activities listed above require communications between distributed components, monitoring of the local computations and sensing of the environment. Agents paradigm suits to carry out these goals, enabling the achievement of a more dynamic and proactive e-learning platform.

The paper is organized in two main parts. The first part introduces a personalized e-learning solution based on the explicit knowledge representation (through the use of ontologies) of educational domains provided by an e-Learning Platform called Intelligent Web Teacher (IWT) [9]. Furthermore, an interaction scenario is presented, from the students' learning experiences point of view. In the second part we go deeply into the details of the defined techniques and algorithms. Conclusions and future works close the paper.

3. A user scenario in e-learning domain

In order to provide a whole overview of the proposed approach, an interaction scenario is depicted through the main steps the learner accomplishes during the fruition of the application interface.

Let us suppose, Mark, an undergraduate student is attempting the course of Mathematics. He is registered to the on-line site of his degree course. Mark usually accesses to the web site (see *1. Log In* in Fig. 1) and navigates through the educational content organized by means the *learning path* relative to his own subjects (*2. Course Fruition*). Mark is studying the differential equation, as evidenced in the learning path schema by the relative node. Let us note previously the teacher loaded educational content relative to his own teaching course; in particular, the teacher defined learning paths for the courses. Each learning path establishes an ordered sequence of subjects (i.e. learning objects) the learner has to study for passing the exams. The e-learning system allows the user to benefit from the educational content (exercises, documents, tests, etc.), as well

[1] www.merlot.org
[2] www.edusource.ca
[3] edna.edu.au

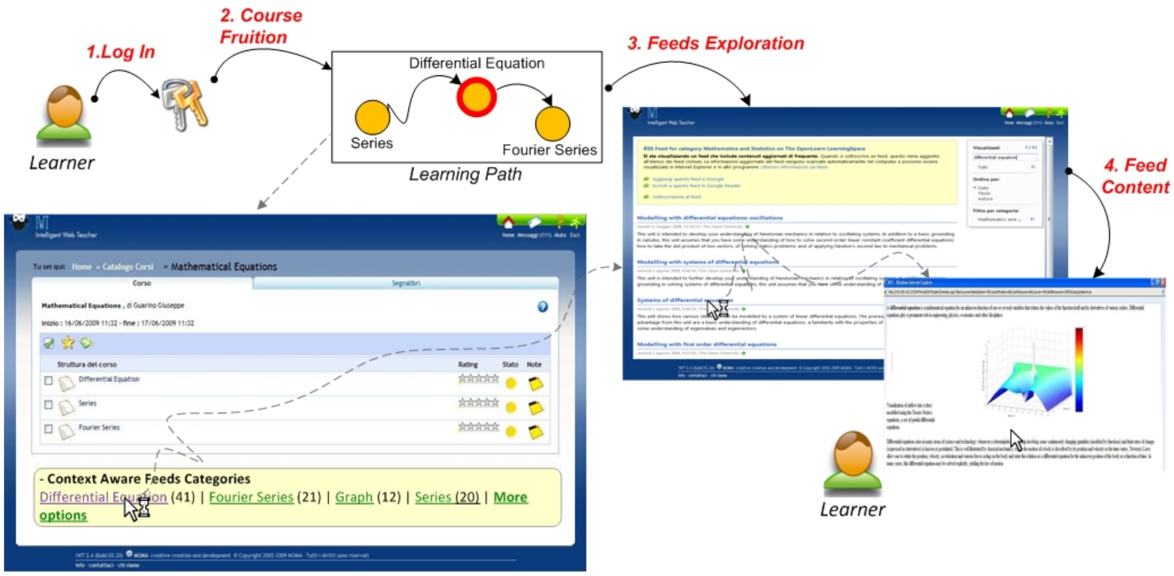


Figure 1. An overview of the user scenario in the e-learning domain

as automatically acquire further learning resources. Mark notes additional resources, provided in RSS-Feed format, connected to his current study subject. In fact, as shown in Fig. 1 the principal web page (on the left of the figure) displays a bottom frame named "Context Aware Feeds Categories" which shows a list of Rss-feeds categories, related to the learning subject. Mark is interested in inspect the proposed feeds and explores the category (*3. Feeds exploration*) which considers more interesting for his own target subject. Then, he opens the web page associated to the relative link in the category list (*Differential Equations*) and envisions the relative feed content (*4. Feed content*).

4. The Proposed Solution

In this section we are going to illustrate the architectural sketch (Fig.2) of our proposed solution decomposing the whole system into software components and describing each component from the adopted technique and the algorithmic point of view. The integrated system is composed by two main modules. The first one is called Intelligent Web Teacher (IWT) that is an e-learning platform enabling the definition and the execution of personalized learning experiences. The second one is the Content Collector which maintains the acquired and inferred knowledge. It is composed of two software modules. The first module (**Training**) is used to store the processed knowledge (in form of *formal lattice*, discussed in the next sections), extracted by the RSS-feeds, collected for each domain of interest, retrieving from the Web. The second module (**Q/A**) enables the querying of the lattice in order to retrieve all the RSS Feeds which are relevant for the context in which the query has been performed. The architecture lays on the agent-based platform Jade[4] which controls the system interactions. The modules interaction indeed, is managed and monitored by the agents which guarantee the communication activities. The agents interaction are accomplished through ACL message exchange. Mainly, the data flow and the interaction between these modules are in charge of the following types of agents:

- Learner Manager: maintains the information of each learner. It starts up whenever a learner logs into the system and loads the appropriate context, according to the target subjects the learner has to study.

- Teacher Manager: similarly to the Learner Manager, it actives itself when a teacher accesses to the IWT system and loads the personal teaching context; furthermore, it acquires the feeds indicated by the teacher for lattices processing and updating.

- Content Manager: creates and manages the lattices (serialized into XML documents), related to the learning contexts. It is in charge of updating the lattices, when new feeds are suggested by the teachers.

- Personal RSS Reader: proposes ad-hoc feed content to the learner, with regard to his educational background. On one hand, it requires to the Content Manager specific content, stored in some concepts of the lattice; on the other hand, by interacting with the Learner Manager, gets information related to the educational subjects of the current learner. This way, it can filter ap-

[4]Java Agent DEvelopment Framework - http://jade.tilab.com/

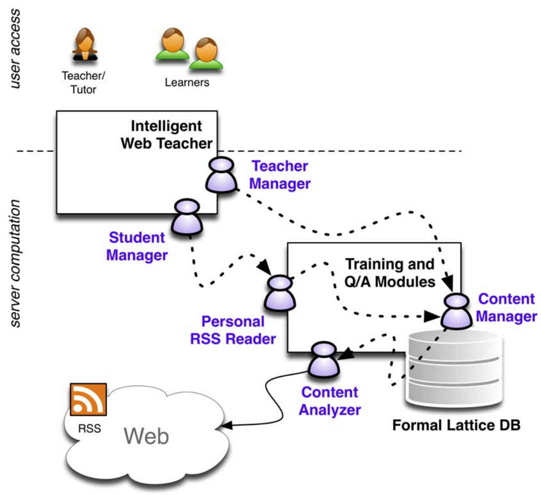


Figure 2. The Architectural Sketch.

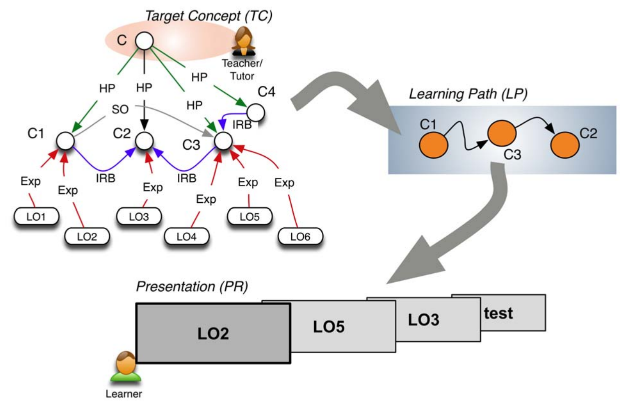


Figure 3. The IWT e-Learning Experiences Definition Process.

propriate feeds to present to the learner, according to his targeted subjects and learning paths.

- Content Analyzer: performs the keywords extraction through the analysis of the RSS-Feeds contents. Preprocessing activities such as normalization, POS tagging, lemmatize and stop-word removal processing have been applied [19].

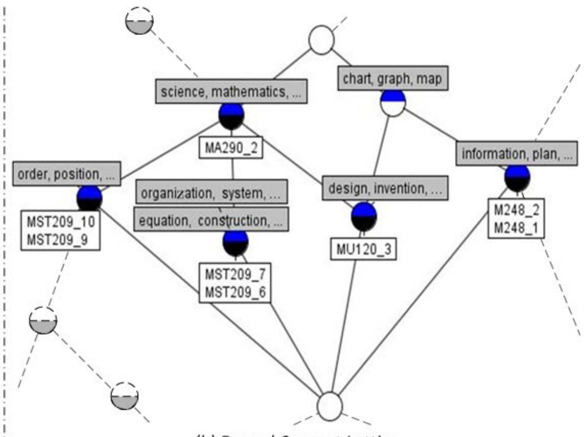
4.1 Intelligent Web Teacher

The Intelligent Web Teacher (IWT) is an e-Learning Web-based Platform whose distinctive features are the construction and delivery of personalised e-learning experiences through the execution of specific algorithms [11]. Using these algorithms it is possible to generate courses tailored to a class, to a specific group and even to single learners. The foundation of IWT is the *Learning Model* described in [9]. The *Learning Model* allows to automatically generate a Unit of Learning (i.e. a course, a module or a lesson structured as a sequence of Learning Activities represented by Learning Objects and/or Learning Services) and to dynamically adapt it during the learning process according to the learner's preferences and cognitive state (personalization process). A Unit of Learning (UoL), during its execution, represents what we have previously named e-learning experience. In IWT, the piece of the educational domain that is relevant for the e-learning experience we want to define, concretize and broadcast, is formalized in a machine-understandable way. The used mechanism is named ontology, i.e. an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words.

In the IWT approach, the vocabularies are composed by terms representing subjects (or concepts within an e-learning ontology) that are relevant for the frame of the educational domain we want to model. Subjects are associated to other subjects through a set of three conceptual relations: *HasPart* (in brief *HP*) that is a part-of relation, *IsRequiredBy* (in brief *IRB*) that is an order relation and *SuggestedOrder* (in brief *SO*) that is a "weak" order relation. The ontologies constructed following the few aforementioned informal rules are named e-learning ontologies [12]. Furthermore, Learning Objects are associated to subjects within a specific e-learning ontology by means the relation *Explain* (in brief *Exp*). E-learning experiences are defined as: (*i*) a set of Target Concepts (TC), i.e. the set of high-level concepts to be transmitted to the learner; (*ii*) A Learning Path (LP), i.e. an ordered sequence of atomic concepts (subjects) that is necessary to explain to a learner in order to let him/her learn TC. Given the personalization on a particular learner, the sequence does not contain subjects already "learnt" (i.e. known with a grade greater than the fixed threshold) by that learner (these information are managed by means the *Learner Model*); (*iii*) A Presentation (PR), i.e. an ordered list of learning objects that the learner has to use in order to acquire knowledge about subjects included in LP. Fig.3 shows an overview of the personalized e-learning experience definition process foreseeing the selection of the TC (performed by the teacher), the automatic extraction of LP (performed by the system algorithms) and the automatic binding between Learning Objects and LP concepts (performed by the system optimization algorithms). The exclusion of the subject C_4 and the selection, for instance, of the learning object LO_2 in the place of LO_1 is due to the personalization process that takes into account the cognitive state and the learning preferences of the current involved learner.

	science, mathematics, ...	organization, system, ...	design, invention, ...	order, position, ...	equation, construction, ...	chart, graph, map	information, plan, ...
MA290_2	1.00		0.35	0.55		0.55	058
MST209_6	0.6	0.99			1.00		0.4
MST209_7	0.6	0.96		0.57	0.97		0.56
MST209_9	0.72			1.00			
MST209_10	0.93			1.00			
M248_1	0.5					0.70	1.00
M248_2	0.48			0.52		0.61	0.61
MU120_3	0.6		0.65	0.58		0.78	

(a) Fuzzy Formal Context



(b) Formal Concept Lattice

Figure 4. Portion of fuzzy formal context (a) and the relative concept lattice with threshold T = 0.6 (b)

4.2 RSS Trainer and Q/A Systems

The theoretical model behind to the module is the Formal Concept Analysis [16]. Specifically, Formal Concept Analysis is a technique of data analysis, which exploits the ordered lattice theory. Recently, FCA and fuzzy techniques are integrated in order to deal with uncertain and vague information. In particular, this approach exploits a fuzzy extension of FCA. Fuzzy Formal Concept Analysis (FFCA) [17] combines fuzzy logic into FCA representing the uncertainty through membership values in the range [0, 1]. Through formal contexts, FFCA enables the representation of the relationships between objects and attributes in a given domain. Some definitions about Formal Concept Analysis and its fuzzy extension are given.

Definition 1: *A* **Fuzzy Formal Context** *is a triple K = (G,M, I = φ(G M)), where G is a set of objects, M is a set of attributes, and I is a fuzzy set on domain G M. Each relation (g, m) $\in$ I has a membership value μ(g, m) in [0, 1].*

Definition 2: Fuzzy Representation of Object. *Each object O in a fuzzy formal context K can be represented by a fuzzy set Φ(O) as $\Phi(O)=A_1(\mu_1)$, $A_2(\mu_2)$,..., $A_m(\mu_m)$, where A_1, A_2,..., A_m is the set of attributes in K and μ_i is the membership of O with attribute A_i in K. Φ(O) is called the fuzzy representation of O.*

This formal models describe RSS-feeds in the e-learning domain. A straight mapping translates the set of attributes M (linguistic terms extracted from feeds content) into the ontological concepts (additional details about the feeds parsing and the terms extraction are given in [19]) and the set G of objects represents the RSS-feeds collection. Thus, in the e-learning context, (g, m) $\in$ I means that the specific RSS-feed *g* is described by the concept *m* (composed of one or more terms extracted by the feed content). A fuzzy formal context is often represented as a cross-table as shown in Fig. 4(a). Let us note, each cell of the table contains a membership value in [0, 1]. According to fuzzy theory, the definition of Fuzzy Formal Concept is given as follows [18].

Definition 3: Fuzzy Formal Concept. *Given a fuzzy formal context K=(G, M, I) and a confidence threshold T, we define $A^*=\{m \in M \mid \forall\, g \in A: \mu(g, m) \geq T\}$ for $A \subseteq G$ and $B^*=\{g \in G \mid \forall\, m \in B: (g, m) \geq T\}$ for $B \subseteq M$. A fuzzy formal concept (or fuzzy concept) of a fuzzy formal context K with a confidence threshold T is a pair ($A_f = \varphi(A)$, B), where $A \subseteq G$, $B \subseteq M$, A*=B and B*=A. Each object $g \in \varphi(A)$ has a membership μ_g defined as*

$$\mu_g = \min_{m \in B} \mu(g, m)$$

where $\mu(g, m)$ is the membership value between object g and attribute m, which is defined in I. Note that if B={ } then μ_g = 1 for every g. A and B are the extent and intent of the formal concept (φ(A),B) respectively.

Definition 4: *Let (A_1, B_1) and (A_2, B_2) be two fuzzy concepts of a fuzzy formal context (G, M, I). ($\varphi(A_1)$, B_1) is the* ***Subconcept*** *of ($\varphi(A_2)$,B_2), denoted as ($\varphi(A_1)$, B_1) $\leq$ ($\varphi(A_2)$,B_2), if and only if $\varphi(A_1)$ $\varphi(A_2)$ ($B_2 \subseteq B_1$). Equivalently, (A_2, B_2) is the* ***Superconcept*** *of (A_1, B_1).*

Definition 5: ***A Fuzzy Concept Lattice*** *of a fuzzy formal context K with a confidence threshold T is a set F(K) of all fuzzy concepts of K with the partial order $\leq$ with the confidence threshold T .*

Fig. 4(b) shows an example of lattice coming from the related table, with threshold T = 0.6. Let us note each node

(i.e. a formal concept) includes the feeds that are better described by a set of terms. The lattice structure emphasizes a taxonomic arrangement of concepts and the subsumption relationships.
The *Training* and *Q/A* modules, through the agent's interaction, manage the lattices building, loading and updating; furthermore they enable the user querying, eliciting personalized feed content.

5. Conclusion

The presented work achieves an agent-based e-learning system which provides appropriate educational contents (with target subjects) through customized feeds.
Next step is to evaluate the system performance. Firstly, our idea is to exploit Information Retrieval techniques (viz. precision/recall measures) for the system assessment; then we would like to compare the learning results of different groups of students: one group will use the additional adhoc RSS-feed content, other group will exploit just the basic learning path, without the RSS-feed support. This way, we would evaluate the two learning processes by providing the same set of multiple choice tests to both the groups of students and then comparing the final results.
A conceivable future development foresees also an extension of the agent-based layer, by defining a crawler which carries out a specific "focused" spidering of the web. The crawler acts just on e-learning community sites: it is able to discriminate relevant web pages whose (learning objects) content is related to the given subjects of interest, otherwise it bypasses not relevant pages. In our vision, the *focused crawler* will provide a solution to the *Vertical Search* problem pointed in Section 2.

References

[1] Gascuena, J. M., Fernandez-Caballero, A., Gonzalez, P. Domain Ontology for Personalized E-Learning in Educational Systems. ICALT IEEE Computer Society (2006) 456-458

[2] A. Romiszowski, Hows the E-learning Baby? Factors Leading to Success or Failure of an Educational Technology Innovation, *Educational Technology*, Vol. 44, No. 1, January-February 2004, pp. 5–27.

[3] W. Chen, Y. Hayashi, L. Jin, I. Mitsuru, R. Mizoguchi, An Ontology-based Intelligent Authoring Tool, 6th Int. Conference on Computers in Education, pp. 41–49.

[4] J. Brase and W. Nejdl, Ontologies and Metadata for eLearning, Handbook on Ontologies, 2004, pp. 555–574.

[5] L. Stojanovic, S. Staab and R. Studer, eLearning based on the Semantic Web, WWW and the Internet, 2001.

[6] W. Nejdl, B. Wolf, S. Staab and J. Tane, EDUTELLA: Searching and Annotating Resources within an RDF-based P2P Network, Semantic Web Workshop, 2002.

[7] J. Tane, C. Schmitz, G. Stumme, S. Staab and R. Studer, The Courseware Watchdog: an Ontology-based tool for finding and organizing learning material, Mobiles Lernen und Forschen, 2003, pp. 93-104.

[8] Murugesan, S. Understanding Web 2.0, IT Professional, vol.9, no.4 pp. 34–41, 2007

[9] G. Albano, M. Gaeta, P. Ritrovato, IWT: an innovative solution for AGS e-Learning model, *International Journal of Knowledge and Learning*, vol. 3, no.2, pp. 209–224, 2007.

[10] J. Dunkel, R. Bruns, S. Ossowski, Semantic E-Learning Agents in *Enterprise Information Systems VI*, pp. 237–244, 2006, Springer Press

[11] G. Acampora, M. Gaeta, and V. Loia, Hierarchical optimization of personalized experiences for e-learning systems through evolutionary models, Neural Computing & Applications.

[12] M. Gaeta, P. Ritrovato, F. Orciuoli, Advanced Ontology Management System for Personalised e-Learning, Knowledge-Based Systems, 22 (2009), pp. 292-301

[13] A. Garro and L. Palopoli, F. Ricca, Exploiting agents in e-learning and skills management context, *AI Commun.*,vol. 19,no. 2, 2006, pp- 137–154, IOS Press.

[14] M. Ganzha, E-learning with Intelligent Agents, IEEE Distributed Systems Online, vol. 9, no. 2, 2008.

[15] Agarwal, R., Deo, A., and Das, S. 2004. Intelligent agents in E-learning. SIGSOFT Softw. Eng. Notes 29, 2 (Mar. 2004), 1-1.

[16] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin - Heidelberg, 1999.

[17] T.T. Quan, S.C. Hui, and T.H Cao, A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data. In Proc. of the 2004 Concept Lattices and Their Applications Workshop, pp. 1-12.

[18] B. Zhou, S. C. Hui and K. Chang, A formal concept analysis approach for web usage mining. In intelligent information Processing II, Z. Shi and Q. He, Eds. Springer-Verlag, London, 2005, pp. 437-441.

[19] C. De Maio, G. Fenza, V. Loia, S. Senatore, Towards an automatic Fuzzy Ontology generation, FUZZ-IEEE 2009, ICC Jeju, Jeju Island, Korea, 20-24 August, 2009.