# Semantic Web supporting Adaptive E-Learning to build and represent Learner Model

***Khaled M. Fouad***
Computer Science Dep. , Community College , Taif Univ., Kingdom of Saudi Arabia (KSA).

kmohammed@tu.edu.sa

***Hany M. Harb***
Computers and Systems Engineering Dept., Faculty of Eng., Azhar Univ., Egypt.

harbhany@yahoo.com

***Nagdy M. Nagdy***
Engineering Applications and Computer Systems, Al-Baha Private College of Science, Kingdom of Saudi Arabia (KSA).

prof_nagdy@yahoo.com

**Paper for :**
The Second International Conference of E-learning and Distance Education (eLi 2011) – Riyadh 2011

# الويب الدلالية لدعم التعلم الإلكتروني التكيفي لبناء وتمثيل نموذج المتعلم

**خالد محمد فؤاد ابراهيم**
علوم الحاسب – كلية المجتمع – جامعة الطائف – المملكة العربية السعودية

kmohammed@tu.edu.sa

**هاني محمد حرب**
هندسة النظم والحاسبات - كلية الهندسة – جامعة الأزهر - مصر

harbhany@yahoo.com

**نجدي محمود نجدي**
هندسة الحاسبات – كلية الباحة للعلوم – المملكة العربية السعودية

prof_nagdy@yahoo.com

**بحث مقدم للمؤتمر الدولي الثاني للتعلم الإلكتروني والتعليم عن بعد**

**الرياض ١٤٣٢ هـ**

**المحور التابع له البحث :**

**" المحور الثالث: التقنيات والبرمجيات الحديثة – أنظمة التعلم الإلكتروني الذكية"**

# ABSTRACT

In order to improve the level of the personalized service and enhance the accuracy of search, it is necessary to build learner (user) model in e-learning system. Constructing an accurate and comprehensive learner model is required in E-learning. Learner model can describe explicitly the learning goals and information needs for the learner.

Learner model contains an information about the learners. Learner model can be extracted from the personal factors like learner interest. Learner model plays an important role in providing personalized service in e-learning by representing learner's identity information and interests.

The Semantic Web is a new generation of the Web in which data is structured in such a way that it can be machine readable and exhibited in a learner-friendly way. Semantic Web is provided tools and methodologies to design adaptive and intelligent learning systems.

This paper proposes an approach to analyze learners' browsing data recorded in the web log to identify learner interests. The learner interests inside learner model is acquired using domain ontology and the space vector model. We describe the learner interests by using semantic relation tree to be built in the learning system. We use the model to realize personalized learning service. We apply knowledge engineering approach to build domain ontology and use Hozo ontology tool to build our domain ontology.



**<u>Keywords :</u>**

*E-Learning; Semantic Web; Ontology; Learner Model; Semantic Relation Tree.*

## الملخص

من أجل تحسين مستوى الخدمة الشخصية وتعزيز دقة البحث ، فمن الضروري بناء نموذج المستخدم في نظام التعلم الإلكتروني. بناء نموذج المستخدم الدقيق والشامل مطلوب في نظام التعلم الإلكتروني. يصف نموذج المتعلم أهداف التعلم والمعلومات المطلوبة عن المتعلم بطريقة واضحة.

يحتوي نماذج المتعلم على معلومات عن المتعلمين. يمكن استخلاص نماذج المتعلم على أساس العوامل الشخصية مثل اهتمام المتعلم. يلعب نموذج المتعلم دورا هاما في تقديم الخدمة الشخصية في التعلم الإلكتروني من خلال تمثيل معلومات هوية واهتمام المتعلم .

الويب الدلالية هى الجيل الجديد من شبكة الإنترنت التي تمثل البيانات بطريقة يمكن أن تكون مقروءة آليا ومعروضة في وسيلة سهلة الاستخدام. قد وفرت الويب الدلالية على الشبكة العالمية أدوات وطرق لتصميم أنظمة التعلم الذكية القادرة على التكيف.

داخل ورقة البحث يتم اقتراح حل لتحليل بيانات تصفح المتعلمين المسجلة في ملف سجل الويب لتحديد اهتمامات المتعلم. يتم الحصول على اهتمامات المتعلم في نموذج المتعلم باستخدام "تمثيل المعرفة" (**Ontology**) بالمجال ونموذج المتجه للفضاء. نصف اهتمامات المتعلم باستخدام شجرة العلاقة الدلالية لاستخدامه في نظام التعلم ولتحقيق خدمات التعلم الفردية. لقد تم تطبيق حل هندسة المعرفة لبناء تمثيل المعرفة الخاصة بالمجال. نستخدم أداة تمثيل المعرفة التي تسمى (**Hozo**) لبناء تمثيل المعرفة الخاصة بنا.

**الكلمات المفتاحية :**

التعلم الإلكتروني ، الويب الدلالية ، تمثيل المعرفة (Ontology)، نموذج المتعلم ، شجرة العلاقة الدلالية.

# 1. Introduction

As the amount of information on the web increases rapidly, retrieving information effectively is getting a big problem. Search engines allow learners to enter keywords to retrieve documents based on these keywords. When the same keywords are submitted by different learners, a typical search engine returns back the same result, regardless of who submitted the query. But it is highly unlikely that the millions of learners with access to the Internet are so similar in their interests. In fact, it has been reported that approximately one half of all retrieved documents are irrelevant [6].

Electronic learning or E-Learning [27] is an interactive learning in which the learning content is available online, and it provides automatic feedback to the learner. In fact, it is much like computer-based training (CBT), and computer-aided instruction (CAT), but the point is that it requires Internet to access to the learning material and to monitor the learner activities.

In the context of e-learning, adaptive systems adapt learning content and the presentation of this content [26]. According to [27], an adaptive system focuses on how the knowledge is learned by the student and pays attention to learning activities; cognitive structures and the context of the learning material. The system intervenes at three stages during the process of adaptation. It controls the process of collecting data about the learner, the process of building up the learner model (learner modeling) and the adaptation process. Basic limitation of adaptive systems and in fact the current state of web is lack of standardized machine readable semantic information in the web content. As solution an idea of Semantic web [15] has been proposed by the chief of World Wide Web (W3) Consortium Tim Berners-Lee. Besides, a claim for effective adaptation methods as integral part of Semantic web has been declared in this vision.

The semantic web [14] is a space understandable and navigable by both human and software agents. It adds structured meaning and organization to the navigational data of the current web, based on formalized ontologies and controlled vocabularies with semantic links to each other. The semantic web-based educational systems need to interoperate, collaborate and exchange content or re-use functionality. A key to enabling the interoperability is to capitalize on (1) semantic conceptualization and ontologies, (2) common standardized communication syntax, and (3) large-scale service-based integration of educational content and functionality provision and usage. The vision of the semantic web-based E-Learning is founded on the following major premises:

- Machine-understandable educational content
- Shareable educational ontologies, including
    - Subject matter ontologies
    - Authoring ontologies (modeling authors' activities)
- Educational semantic web services, for supporting
    - Learning, e.g., information retrieval, summarization, interpretation (sense-making), structure-visualization, argumentation, etc.
    - Assessment, e.g., tests and performance tracking
    - Collaboration, e.g., group formation, peer help, etc.
- Semantic interoperability

Ontology [22] comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic. Ontologies applied to the Web are creating the Semantic Web. Ontologies facilitate knowledge sharing and reuse, i.e. a common understanding of various contents that reaches across people and applications.

The behavior of an adaptive system [24] varies according to the data from the learner model and the learner profile. Without knowing anything about the learner, a system would perform in exactly the same way for all learners. It was described the application of learner models as follows:

"Users are different: they have different background, different knowledge about a subject, different preferences, goals and interests. To individualize, personalize or customize actions a user model is needed that allows for selection of individualized responses to the user."[26].

It was summarized [21] the role of personalization in learning environments as follows:

- personalized learning environments enable one-to-one or many-to-one learning paradigms (one teacher - one learner, and many teachers – one learner), contrary to traditional learning environments that always adopt one-to-many learning paradigm (one teacher, many students);
- personalized learning environments impose no constraints in terms of learning time, location, etc., whereas traditional ones are fairly restricted by the learning setting;
- personalized learning environments recognize the huge variety in the learner's characteristics and preferences in terms of the learning style, media, interests.
- personalized learning environments tailor instruction to suit the learner's requirements (self-directed learning); in traditional learning environments, in which the curriculum, learning units, and the selection and sequencing of learning material are determined by the tutor.

Learner interests [4,13] can be collected by two ways: explicit and implicit collection. Explicit collection is predefined or feedback by learner's ratings through an interface. There are roughly two kinds of automatic way to capture a learner's interest implicitly: behavior-based and history-based. The behavior-based research proves that the time spent on a page, the amount of scrolling on a page and the combination of them has a strong positive relationship with learner interests. Browsing histories capture the relationship between learner's interests and his click history in which sufficient contextual information is already hidden in the web log.

In this paper, we propose and represent the learner model based on learner interest, which is acquired based on domain ontology and the space vector model [17]. We also describe the learner interests based on semantic relation tree to be built in the learning system, and use the model to realize personalized service. We apply knowledge engineering approach to build domain ontology [7], and use Hozo ontology tool [30] to build our domain ontology.

The paper is organized as follows: Section 2 discusses related works. Section 3 describes the proposed architecture. We describe Suggestions and recommendations in section 4. The conclusion is described in section 5.

## 2. Related Works

An accurate [16] representation of a learners interests, generally stored in some form of learner model, is crucial to the performance of personalized search or browsing agents. Learner model is often represented by keyword/concept vectors or concept hierarchy. The acquired model can then be used for analyzing and predicting the future learner access behavior. Learner model may be built explicitly, by asking learners questions, or implicitly, by observing their activity.

In [8], authors investigated the techniques to create a user profile automatically using the ontological approach. They used a framework to gather the user information from different search space where user's details could be found. The details include user's general information to specific preferences. They used Meta search in user's blog, personal/organization web page, and any other cites to collect information about user. This information is assigned to a pre- structure hierarchy or in reference ontology to create an initial user profile. More clearly, initial profile is learned by the concept/document collected from user's details. In traditional user profiling system feature extraction from document is done by vector space model or considering term frequency, *tf-idf* methods only. In this research, authors considered WordNet and Lexico-Syntactic pattern for hyponyms to extract feature from document. This profile further improved by taking collaborative user methods. Where, they found a group of users with similar interest by taking similarity score among them. After that an ontology matching approach is applied to learn the profile with other similar user which is called improved profile.

In [9], authors introduced a method for learning and updating a user profile automatically. The proposed method belongs to implicit techniques. It processes and analyzes behavioral patterns of user activities on the web, and modifies a user profile based on extracted information from user's web-logs. The method relies on analysis of web-logs for discovering concepts and items representing user's current and new interests. Those found concepts and items are compared with items from a user profile, and the most relevant ones are added to this profile. The mechanism used for identifying relevant items is built based on a newly introduced concept of ontology-based semantic similarity.

There are many different methods to Construct Learner Models [26]: Machine Learning Methods, Bayesian Methods, Overlay methods, Stereotype methods, Plan Recognition. Update of Learner Models methods are: Analysis of Learner Responses, Analysis of the Process of Problem Solution, Analysis of Learner Actions, Discounting Old Data.

In [4], student model mainly included the cognitive model and the interest model. Cognitive model mainly pay attention to learner background knowledge, study style and cognition level, through synthesizing the domestic and foreign research practice authors proposed to use Solomon Study Style Measure Meter as preceding measure to test learner's study style. Regarding the student's cognition level's estimate they took the thought of fuzzy set.

It was combined the ontology and concept space [10], indicated the feature items of user profile with semantic concepts, calculates learner's interest-level to the topic through establishing the word frequency and utilize the suitable calculation methods, mining the concepts within the user's feedback files and the relationship between concepts, combines user's short-term interests and long-term interests to create user profiles model with semantic concept hierarchy tree and embody the drifting of user profile and improves and completes the user profiles model consistently on the related feedback mechanism.

In [23], the authors proposed a new approach to User Model Acquisition (UMA) which has two important features. It doesn't assume that users always have a well-defined idea of what they are looking for, and it is ontology-based, i.e., it was dealt with concepts instead of keywords to formulate queries. The first problem is that most approaches assume users to have a well-defined idea of what they are looking for, which is not always the case. They solved this problem by letting fuzzy user models evolve on the basis of a rating induced by user behavior. The second problem concerns the use of keywords, not concepts, to formulate queries. Considering words and not the concepts behind them often leads to a loss in terms of the quantity and quality of information retrieved. They solved this problem by adopting an ontology-based approach.

In [20], the student interactions with the system are monitored and stored in log files. The recorded data are then cleaned and preprocessed (e.g. compute the relative frequency of learner actions, the amount of time spent on a specific action type, the order of navigation etc). Subsequently, these behavioral indicators are analyzed and based on them the system can

infer different learning preferences of the student. Finally, the identified learner model is used by the decision-making component to select the most appropriate adaptation actions, in order to provide the student with the educational resources that suit her/his specific needs.
The users' requirements [3] can be represented as a case in the defined structure which can be reasoned to enable the requirements analysis. In this model, a number of techniques perform the analysis of: (1) articulating users needs and capturing the user requirements in a profile; (2) mapping the user requirements onto information dimensions in an information space through a reasoning process; (3) configuring the dimensions of the information space into information provision specifications; and (4) organizing information objects in a repository to enable information provision specifications, discovering and retrieving suitable content. The method defines constraints through a norm construct which systemically controls the information provision process.

## 3. Proposed System Architecture

The design idea of adaptive learning system based on creating an ideal learning environment for the learners so that system can provide the adaptive learning support according to the learner's individual differences, and to promote learners to study initiatively, and to achieve the knowledge construction. There are some objectives in the design of adaptive learning system. Firstly, system can provide the adaptive learning content based on the learner's interest. Secondly, the system can support the self-directed learning and collaborative Learning. Thirdly, the system can help teachers to understand the learning process of learners, and adjust the pedagogical activities, and support the learning evaluation. Lastly, the system will support the courses development for staff. Based on these considerations, a new architecture of adaptive learning system is proposed in current paper. It is illustrated in figure 1. According to proposed architecture, the learning system is mainly composed of four processes: Learner's Web Log Analysis, Learner Interest Acquiring, Learner Interests Representing, and Domain Ontology Developing.

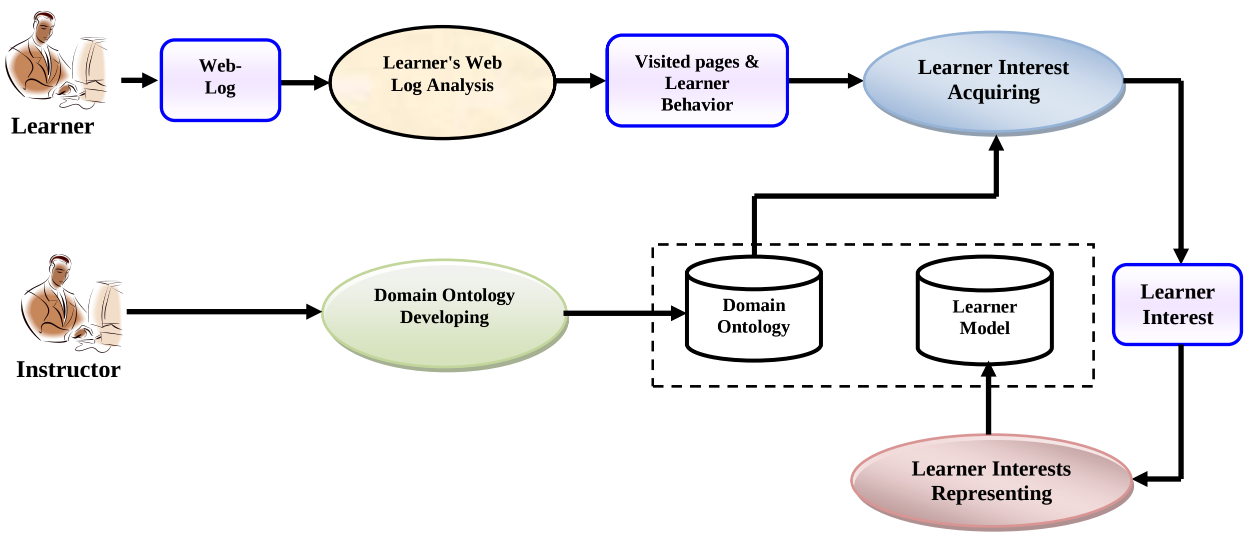



Figure 1: Proposed System Architecture to build learner model

## 3.1 Web Log Analysis

Web usage mining [5], the process of discovering patterns from web data using data mining methods, strives to find learner preferences based on the web-logs that reside on servers. Web log [32] records each transaction, which was executed by the browser at each web access. Each line in the log represents a record with the IP address, time and date of the visit, accessed object and referenced object. In such data we follow sequences in visiting individual pages by the learner, who is, under certain condition, identified by the IP address. In sequences we can look for learners behavior patterns.

The data from Web logs, in its raw form, is not suitable for the application of usage mining algorithms. The data needs to be cleaned and preprocessed. To perform log data analysis the data pre-processing process must be accomplished. The data pre-processing is the process of cleaning and transforming raw data sets into a form suitable for web mining. The task of the data pre-processing module is therefore to obtain usable datasets from raw web log files, which, in most cases, contain a considerable amount of incomplete and irrelevant information.

The overall data preparation process [25] is briefly described in figure 2.

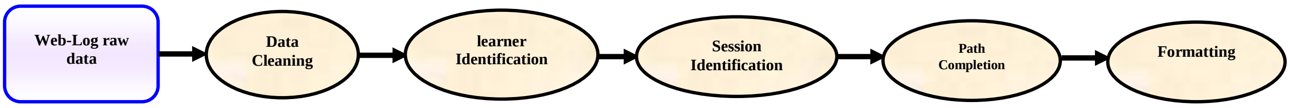


Figure 2: data preparation process for web log

***Data Cleaning:*** to remove accesses to irrelevant items (such as button images), accesses by Web crawlers (i.e. non-human accesses), and failed requests.

***Learner Identification:*** Because web logs are recorded in a sequential manner as they arrive, therefore, records for a specific learner are not necessary recorded in consecutive order rather they could be separated by records from other learners.

***Session Identification:*** To divide pages accessed by each learner into individual sessions. A session is a sequence of pages visited by a learner. We also call it as a usage sequence.

***Path Completion:*** To determine if there are important accesses which are not recorded in the access log due to caching on several levels.

***Formatting:*** Format the data to be readable by data mining systems.

Once web logs are preprocessed, useful web usage patterns may be generated by applying data mining techniques.

Table 1 shows a sample of web log data after preprocessing process.

Table 1: Sample of Web Log Data

| Visit Time | UserId | URL | …. |
|---|---|---|---|
| 20090405202122 | 10 | http://www.cs.bu.edu/teaching/ | …. |
| 20090405203225 | 19 | http://www.cs.bu.edu/teaching/unix/intro/ | … |
| 20090406081905 | 10 | http://www.cs.bu.edu/teaching/cs113/spring-2000/object/ | ….. |
| 20090407091215 | 11 | http://www.aw-bc.com/brookshear/ | ….. |
| 20090407082621 | 19 | http://chortle.ccsu.edu/java5/Notes/chap21/ch21_1.html | …. |

The outputs of this step are the web based learning materials; that the learner explored and preferred it, and the behavior pattern of the learner. The learner behavior is used to acquiring knowledge requirement for learners based on course ontology.

## 3.2 Domain ontology developing based knowledge engineering approach

Ontology engineering is a subfield of knowledge engineering that studies the methods and methodologies for building ontologies. It researches the ontology development process, the ontology life cycle, the methods and methodologies for building ontologies, and the tools suite and languages that support them. Knowledge Engineering field usually uses the IEEE 1074-2006 standard [31] as a reference criteria. The IEEE 1074-2006 is a standard for developing a software project life cycle processes. It describes the software development process, the activities to be carried out, and the techniques that can be used for developing software.

It was proposed [7] a knowledge engineering approach to build domain ontology. Figure 3 shows main steps of the ontology development process.

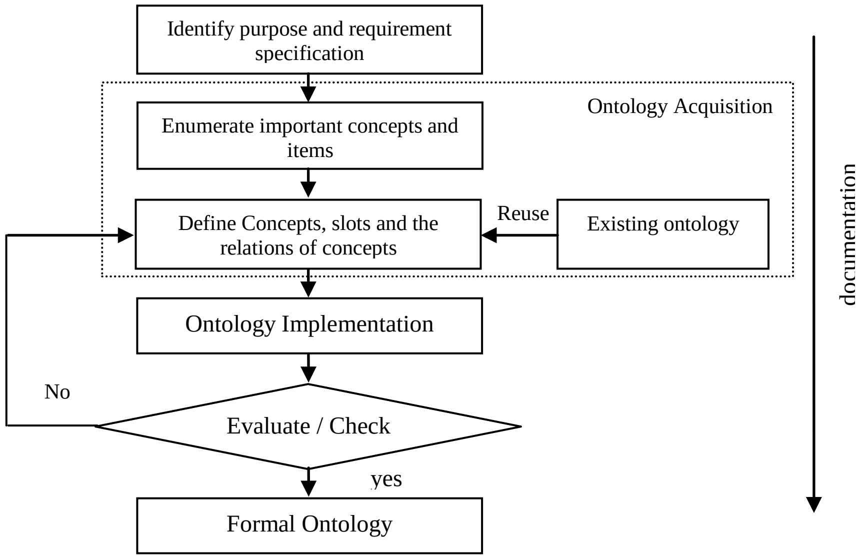
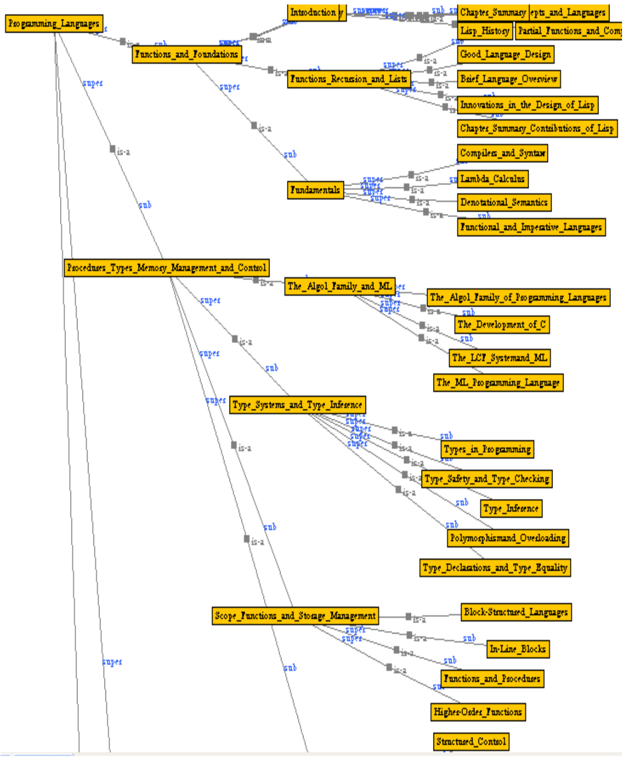



Figure 3 Main steps of the ontology development process

Identify purpose and requirement specification concerns to clearly identify the ontology purpose, scope and its intended use, that is the competence of the ontology. Ontology acquisition is to capture the domain concepts based on the ontology competence. The relevant domain entities (e.g. concepts, relations, slots, and role) should be identified and organized into hierarchy structure. This phase involves three steps as follows: first, enumerate important concepts and terms in this domain; second, define concepts, properties and relations of concepts, and organize them into hierarchy structure; third, consider reusing existing ontology. Ontology implementation aims to explicitly represent the conceptualization captured in a formal language. Evaluation/Check means that the ontology must be evaluated to check whether it satisfies the specification requirements. Documentation means that all the ontology development must be documented, including purposes, requirements, textual descriptions of the conceptualization, and the formal ontology.

Our domain focuses "programming languages" course. We use Hozo [29] as our ontology editor. Since Hozo is based on an ontological theory of a role-concept, it can distinguish concepts dependent on particular contexts from so-called basic concepts and contribute to building reusable ontologies. A role-concept [7] represents a role which an object plays in a specific context and it is defined with other concepts. On the other hand, a basic-concept does not need other concepts for being defined. An entity of the basic concept that plays a role-concept is called a role-holder. Figure 4 shows part of our domain ontology and the extracted OWL [33].




```
- <owl:Class rdf:ID="Programming_Languages">
  <rdfs:label>Programming_Languages</rdfs:label>
  </owl:Class>
- <owl:Class rdf:ID="Functions_and_Foundations">
  <rdfs:label>Functions_and_Foundations</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Programming_Languages" />
  </owl:Class>
- <owl:Class rdf:ID="Introduction">
  <rdfs:label>Introduction</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Functions_and_Foundations" />
  </owl:Class>
- <owl:Class rdf:ID="Programming_Language">
  <rdfs:label>Programming_Language</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
  </owl:Class>
- <owl:Class rdf:ID="Goals">
  <rdfs:label>Goals</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
  </owl:Class>
- <owl:Class rdf:ID="Programming_Language_History">
  <rdfs:label>Programming_Language_History</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
  </owl:Class>
- <owl:Class rdf:ID="Organization_Concepts_and_Languages">
  <rdfs:label>Organization_Concepts_and_Languages</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Introduction" />
  </owl:Class>
- <owl:Class rdf:ID="Computability">
  <rdfs:label>Computability</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Functions_and_Foundations" />
  </owl:Class>
- <owl:Class rdf:ID="Partial_Functions_and_Computability">
  <rdfs:label>Partial_Functions_and_Computability</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Computability" />
  </owl:Class>
- <owl:Class rdf:ID="Chapter_Summary">
  <rdfs:label>Chapter_Summary</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Computability" />
  </owl:Class>
```


Figure 4: part of the domain ontology of "Programming Languages" course

## 3.3 Learner's interest acquiring

In the proposed system [1], learner interest model's knowledge expression uses the thought, which is based on the space vector model's expression method and the domain ontology. This method acquires learner's interest was shown in [1]. Figure 5 shows certain steps to acquire learner interest.

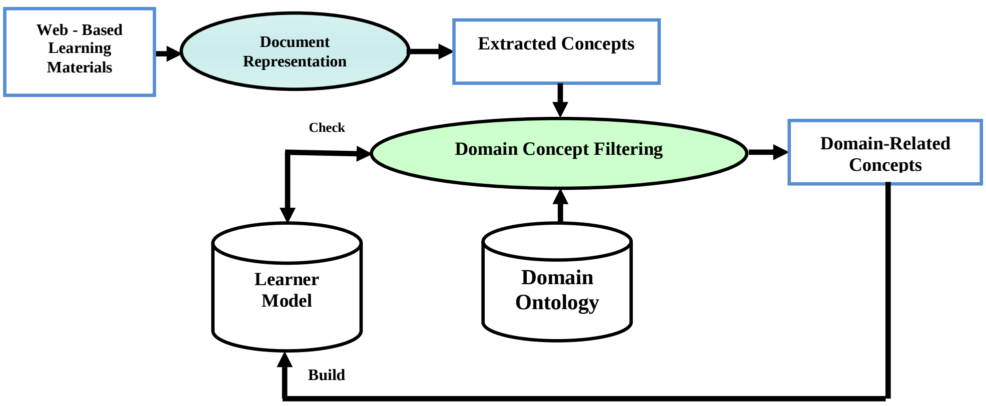



Figure 5: steps to acquire learner interest

### 3.3.1 Document Representation

The Vector Space Model [11, 18, 19] is adapted in our proposed system to achieve effective representations of documents. Each document is identified by n-dimensional feature vector where each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight. The weight is a function of the *term frequency*, *collection frequency* and *normalization factors*. Different weighting approaches may be applied by varying this function. Hence, a document j is represented by the document vector $d_j$:

$d_j = (w_{1j}, w_{2j}, ..., w_{nj})$ Where, $\boldsymbol{w_{kj}}$ is the weight of the $k^{th}$ term in the document j.

The term frequency reflects the importance of term k within a particular document j. The weighting factor may be global or local. The global weighting factor clarify the importance of a term k within the entire collection of documents, whereas a local weighting factor considers the given document only.

The document keywords were extracted by using a term-frequency-inverse-document-frequency (*tf-idf*) calculation [18, 19], which is a well-established technique in information retrieval. The weight of term *k* in document *j* is represented as:

$$w_{kj} = tf_{kj} \times (\log_2^n - \log_2^{df_k} + 1)$$

Where: $tf_{kj}$ = the term *k* frequency in document *j*, *dfk* = number of documents in which term *k* occurs, *n* = total number of documents in collection. Table 3 shows the term frequency in different documents.

Table 3: Sample of the Documents with their representation

| DOC/items | Computer science | AI | Programming | Software eng. | Networks | LAN | WAN | Computer Arch. | Processors | Parallel processing |
|---|---|---|---|---|---|---|---|---|---|---|
| Doc1 | 20 | 25 | 20 | 15 | 10 | 0 | 0 | 5 | 0 | 0 |
| Doc 2 | 15 | 15 | 25 | 15 | 5 | 0 | 0 | 0 | 0 | 0 |
| Doc 3 | 0 | 5 | 25 | 10 | 25 | 10 | 15 | 5 | 0 | 0 |
| Doc 4 | 0 | 5 | 25 | 10 | 20 | 5 | 15 | 5 | 0 | 0 |
| Doc 5 | 0 | 0 | 5 | 0 | 10 | 0 | 0 | 5 | 20 | 20 |
| Doc 6 | 0 | 0 | 0 | 5 | 20 | 0 | 0 | 0 | 25 | 30 |
| Doc 7 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 5 | 20 | 10 |
| Doc 8 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 10 | 25 | 5 |
| Doc 9 | 15 | 5 | 30 | 30 | 0 | 0 | 0 | 5 | 0 | 0 |
| Doc 10 | 5 | 0 | 25 | 25 | 0 | 0 | 0 | 0 | 0 | 0 |
| Doc 11 | 10 | 0 | 0 | 0 | 10 | 0 | 0 | 25 | 10 | 30 |
| Doc 12 | 10 | 0 | 0 | 0 | 10 | 0 | 0 | 30 | 10 | 25 |
| Doc 13 | 20 | 25 | 5 | 0 | 10 | 0 | 0 | 5 | 0 | 0 |
| Doc 14 | 15 | 15 | 5 | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| Doc 15 | 0 | 30 | 20 | 15 | 25 | 0 | 5 | 0 | 0 | 0 |
| Doc 16 | 0 | 25 | 25 | 15 | 20 | 0 | 0 | 0 | 0 | 0 |

The main purpose of this step is to extract interested items in the web page, then get term frequency that reflects the importance of term. Finally get the weight of terms in the selected page. The output of this step is the weight of terms in selected page that can be used to build learner interest profile. Table 4 shows a sample of the weighted terms in the documents; that found in table 3.

Table 4: Sample of the weighted terms in the documents

| DOC/items | computer Science | AI | programming | Software Engineering | Network | LAN | WAN | Computer Architecture | Processors | Parallel Processing |
|---|---|---|---|---|---|---|---|---|---|---|
| Doc1 | 0.40 | 0.46 | 0.28 | 0.25 | 0.14 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 |
| Doc2 | 0.30 | 0.27 | 0.35 | 0.25 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Doc3 | 0.00 | 0.09 | 0.35 | 0.17 | 0.35 | 0.40 | 0.51 | 0.09 | 0.00 | 0.00 |
| Doc4 | 0.00 | 0.09 | 0.35 | 0.16 | 0.28 | 0.20 | 0.51 | 0.00 | 0.00 | 0.00 |
| Doc5 | 0.00 | 0.00 | 0.07 | 0.00 | 0.14 | 0.00 | 0.00 | 0.09 | 0.48 | 0.48 |
| Doc6 | 0.00 | 0.00 | 0.00 | 0.08 | 0.28 | 0.00 | 0.00 | 0.00 | 0.60 | 0.72 |
| Doc7 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.48 | 0.24 |
| Doc8 | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.18 | 0.60 | 0.12 |
| Doc9 | 0.20 | 0.09 | 0.42 | 0.50 | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 |
| Doc10 | 0.10 | 0.00 | 0.35 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Doc11 | 0.20 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.45 | 0.24 | 0.72 |
| Doc12 | 0.20 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.54 | 0.24 | 0.60 |
| Doc13 | 0.40 | 0.45 | 0.07 | 0.00 | 0.14 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 |
| Doc14 | 0.30 | 0.27 | 0.07 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Doc15 | 0.00 | 0.54 | 0.28 | 0.25 | 0.35 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 |
| Doc16 | 0.00 | 0.45 | 0.35 | 0.25 | 0.28 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

### 3.3.2 Domain Concept Filtering

This process discovers concepts which represent the learner's interests. These concepts and items are compared to the domain ontology to check the relevant items to the learner profile. The most relevant ones update the learner profile. The items relevance is based on ontology-based semantic similarity where browsed items by a learner on the web are compared to the items from a domain ontology and learner profile. The importance is combined with the semantic similarity to obtain a level of relevance. The page items are processed to identify domain-related words to be added to the learner profile. A bag of browsed items is obtained via a simple word indexing of the page visited by the learner. We filter out irrelevant words using the list of items extracted from domain ontology. Once domain-related items are identified, we evaluate their relevance to learner's interests.

The selected method was used in [12,28] to compute semantic similarity function (S) based on a domain ontology. The similarity is estimated for each pair of items where one item is taken from a learner profile, while the other one from a set of browsed items.

The functions *Sw* is the similarity between synonym sets, *Su* is the similarity between features, and *Sn* is the similarity between semantic neighborhoods between entity classes *a* of ontology *p* and *b* of ontology *q*, and $w_w$, $w_u$, and $w_n$ are the respective weights of the similarity of each specification component.

$$S(a^p, b^q) = w_w \times S_w(a^p, b^q) + w_u \times S_u(a^p, b^q) + w_n \times S_n(a^p, b^q) \quad ; \text{For } w_w;\ w_u;\ w_n \geq 0:$$

Weights assigned to Sw, Su, and Sn depends on the characteristics of the ontologies.

The similarity measures are defined in terms of a matching process [12,28]:

$$\mathbf{S}(a, b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a,b)|A / B| + (1 - \alpha(a,b))|B / A|}$$

where *A* and *B* are description sets of classes *a* and *b*, i.e., synonym sets, sets of distinguishing features and a set of classes in semantic neighborhood; *(A∩B)* and *(A/B)* represent intersection and difference respectively, | | is the cardinality of a set; and α is a function that defines relative importance of non-common characteristics. A set of browsed items that are similar to items from the learner profile is considered as a set of items that can be added to this profile.

## 3.4 Learner interest represenatation

The learner model representation adopted in this paper is based on keywords expression that is based on semantic relation tree [2]. This expression shows the relationship between the learners' interests and express series of the learners' interests. Setting the relationship value between the nodes can not only better reflect the learner's interest in features, but avoid a separate list arising from the use of keywords disconnect problem. The structure of semantic relation tree is shown in figure 6.

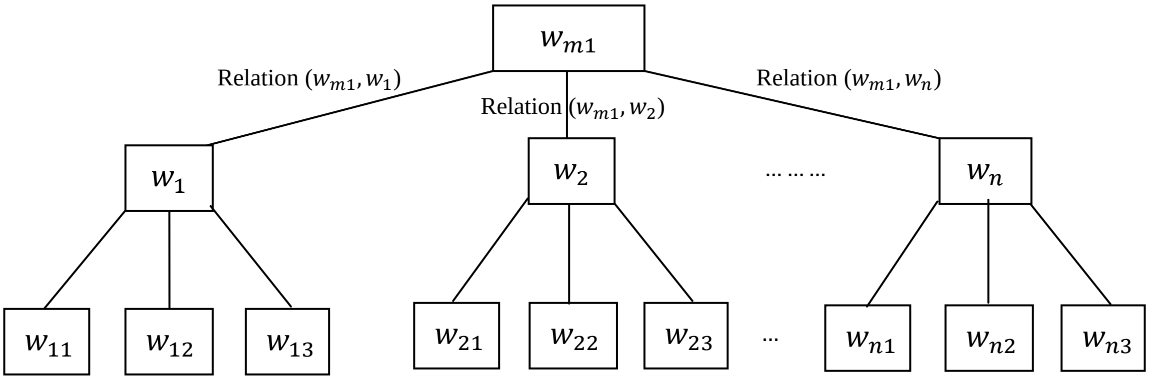



Figure 6: Semantic Relation Tree

In the tree, node={*w*,*update_time*}, where (*w*) represents a word and its weight and *update_time* represent the last time that word was updated. Vector ( $w_{m1}, w_{m2}, \ldots w_{mn}$) that associated semantic tree root, comes from the weights of each keywords in the documents that is shown in table 4. The relation ($w_{m1}$, $w_n$) indicates the relationship value between the nodes $w_{m1}$ and $w_n$ .This relation is calculated as found in the equation :

$$Relation\ (w_{m1}, w_n) = 0.7 \times w_{ci} + 0.3 \times w_{ti} \quad .$$

$w_{ci}$ represents the weight of the child node, that was shown in (section 4.3.1) ; $w_{ti}$ represents the child node viewing time, $w_{ti}$ is calculated as :

$w_{ti} = t_i \,/\, t_{total}$ .

where $t_i$ represents the time viewed by learner; $t_{total}$ represents the sum of learner browsing time, that can be calculated by the difference between the time learner entering and exiting of the system. This relation equation was derived and proved in [2]. The next example shows the method to structure the semantic relation tree by applying the relation equation. Table 5 shows the weight values ($w_{ci}$) , time viewed ($w_{ti}$) for different terms and the result relation of the interests of learner. Figure 7 shows the semantic relation tree of the example.

Table 5: Weight values, time viewed for different terms and relation of the interests

| | computer | computer Science | AI | programming | Software Engineering | Network | LAN | WAN | Computer Architecture | Processors | Parallel Processing |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $w_{ci}$ | 0.14 | 0.40 | 0.46 | 0.28 | 0.25 | 0.14 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 |
| $w_{ti}$ | 3/5 | 3/14 | 2/11 | 2/9 | 1/7 | 1/5 | 1/4 | 1/3 | 1/5 | 1/3 | 1/4 |

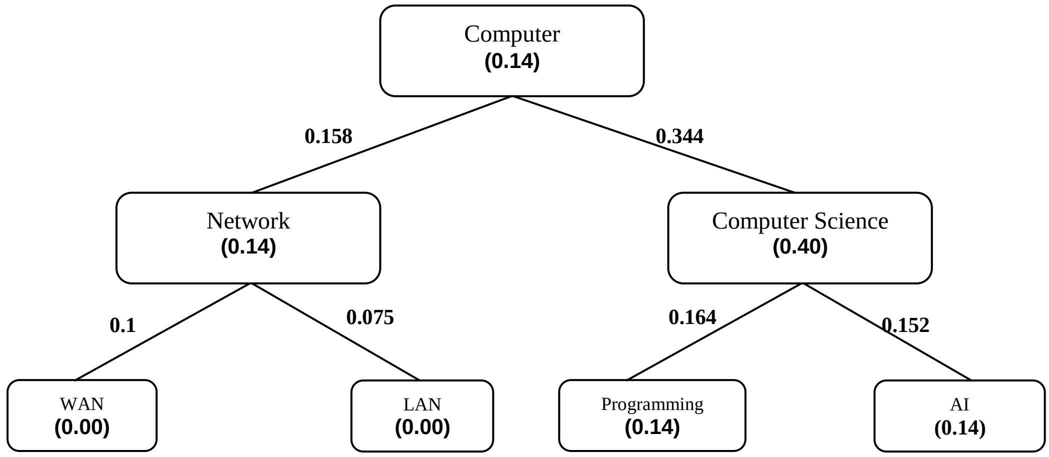



Figure 7 Semantic Relation Tree of the example in table 5

# 4. Suggestions and Recommendations

In order to achieve personalization and improve the precision and recall of learning resources retrieval, we need a good similarity assessment algorithm for the use of semantic web technology in semantic search to fetch the relevant learning content to the learner.

We suggest to use our results of learner model and semantic web to personalize the semantic search of learning content in e-learning systems to improve the precision and recall of learning content retrieval for the learner.

We recommend to use fuzzy clustering technique that allows an entity to belong to more than one cluster with different degrees of accuracy. While hard clustering assigns each entity exactly to one of the clusters. Fuzzy clustering is suitable for constructing the learner profiles representations such as (learner ontology).Such representation of learner profiles is useful because some information is not forced to fully belong to any one of the learner profiles. Fuzzy clustering methods may allow some information to belong to several learner profiles simultaneously with different degrees of accuracy.

# 5. Conclusion

The current paper is aimed to describe the learner model based on semantic web to be built in the learning system, and use the model to realize personalized e-learning. The advantages of the system were: first the system can provide the adaptive learning content based on the learner's interest. This system can support the self-directed learning and collaborative Learning, and help the teachers to understand the learning process of learners.

This method uses the learner behavior in web-log and ontology-based semantic similarity to compare items browsed by a learner on the web with the items from a domain ontology and learner profile to acquire learner interest. Domain ontology was built based on a knowledge engineering approach and we used Hozo as our ontology editor.

We used semantic relation tree to represent learner model based on keywords expression. This expression shows the relationship between the learners' interests and express series of the learners' interests.

Current work showed that the semantic web technologies can be used to achieve the personalization and improve the precision and recall of learning resources retrieval in learning systems.

# References


[1] H. Hany, F. Khaled. (2010). “Semantic web based Approach to learn and update Learner Profile in Adaptive E-Learning ”, Al-Azhar Engineering Eleventh International Conference, December 23-26, pp., 2010.
[2] Y. Li, J. Wu. (2010). Personalized User's Interest Model Study Based on Semantic Relation Tree, 2nd International Conference on Computer Engineering and Technology, 978-1-4244-6349-7/10, IEEE.
[3] S. Lily, J. Cleopa. (2010). Case-based analysis in user requirements modeling for knowledge construction, Information and Software Technology 52 (2010) 770–777, doi:10.1016/j.infsof.2010.03.010, 0950-5849, Elsevier.
[4] B. Qiu, W. Zhao. (2009). Student Model in Adaptive Learning System based on Semantic Web, In First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466.
[5] G. PAOLO, F. SILVIA. (2009). Applied Data Mining for Business and Industry Second Edition, Chapter 6: Describing website visitors, ISBN: 978-0-470-05887-9 (Pbk), John Wiley.
[6] W. Shuqing, S. Li, L. Zhen, F. Yan. (2009). Algorithm Research on User Interests Extracting via Web Log Data, 2009 International Conference on Web Information Systems and Mining, 978-0-7695-3817-4/09, IEEE.
[7] H. YUN, J. XU, M. J. XIONG. (2009). Development of Domain Ontology for E-learning Course, 978-1-4244-3930-0/09, IEEE.
[8] D. Trong, N Mohammed, L. Delong , and J. Geun. (2009). A Collaborative Ontology-Based User Profiles System, N.T. Nguyen, R. Kowalczyk, and S.-M. Chen (Eds.): ICCCI 2009, LNAI 5796, pp. 540–552, Springer-Verlag Berlin Heidelberg.
[9] R. Marek and K. Sayed. (2009). Updating User Profile using Ontology-based Semantic Similarity, FUZZ_IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.
[10] W. Cuncun, H. Chongben, T. Hengsong. (2009). A Personalized Model for Ontology-driven User Profiles Mining, 2009 International Symposium on Intelligent Ubiquitous Computing and Education, 978-0-7695-3619-4/09, IEEE.
[11] B. Qiu, W. Zhao. (2009). Student Model in Adaptive Learning System based on Semantic Web, 2009 First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466.
[12] R. Marek and K. Sayed. (2009). Updating User Profile using Ontology-based Semantic Similarity, FUZZ_IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.
[13] L. Fang, L. Yihong,W. Yanchen, Z. Kai,L. Feng,W. Xinguang. (2008). Discovery of a User Interests on the Internet, 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, DOI 10.1109/WIIAT.2008.18, 978-0-7695-3496-1/08, IEEE.
[14] D. Dicheva. (2008). Ontologies and Semantic Web for E-Learning, in book "Handbook on Information Technologies for Education and Training", 978-3-540-74155-8 (Online), Springer Berlin Heidelberg.
[15] K. Breitman, M.Casanova and W. Truszkowski. (2007). Semantic Web: Concepts, Technologies and Applications, e-ISBN: 978-1-84628-710-7, Springer-Verlag London Limited.
[16] Z. Hui, S. Yu, and S. Han-tao. (2007). Construction of Ontology-Based User Model for Web Personalization, C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 67–76, Springer-Verlag Berlin Heidelberg.
[17] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng, and W. Daming. (2007). Ontology Based User Profiling in Personalized Information Service Agent, Seventh International Conference on Computer and Information Technology, 0-7695-2983-6/07, IEEE.
[18] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng, and W. Daming. (2007). Ontology Based User Profiling in Personalized Information Service Agent, Seventh International Conference on Computer and Information Technology, 0-7695-2983-6/07, IEEE.

[19] P. Jianguo, Z. Bofeng, W. Shufeng, W. Gengfeng. (2007). A Personalized Semantic Search Method for Intelligent e-Learning, 2007 International Conference on Intelligent Pervasive Computing, 0-7695-3006-0/07, IEEE , DOI 10.1109/IPC.2007.48.
[20] P. Elvira, T. Philippe, B. Costin. (2007). Adaptive Educational Hypermedia Systems: A Focus on Learning Styles, EUROCON 2007 The International Conference on "Computer as a Tool" Warsaw, September 9-12, 1-4244-0813-X/07, IEEE.
[21] D. Vladan. (2006). SEMANTIC WEB AND EDUCATION, ISBN: 0-387-35417-4, Springer Science+Business Media, LLC.
[22] G. Fayed, D. Sameh, H. Ahmad, M. Jihad., A. Samir, and S. Hosam. (2006). E-Learning Model Based On Semantic Web Technology, International Journal of Computing &Information Sciences Vol. 4, No. 2, August 2006, On-Line. Pages 63 – 71.
[23] P. C´elia and T. Andrea G.B. (2006). An Evolutionary Approach to Ontology-Based User Model Acquisition, V. Di Ges´u, F. Masulli, and A. Petrosino (Eds.): WILF 2003, LNAI 2955, pp. 25–32, c_Springer-Verlag Berlin Heidelberg.
[24] F. Christoph. (2005). User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis at Graz University of Technology, Copyright 2005 by Christoph Fr¨oschl.
[25] K. Natheer, C. Chien-Chung. (2005). Web Usage Mining Using Rough Sets, NAFIPS 2005 - 2005 Annual Meeting of the North American Fuzzy Information Processing Society, 0-7803-9187-X/05, IEEE.
[26] F. Christoph. (2005). User Modeling and User Profiling in Adaptive E-learning Systems, Master's Thesis At Graz University of Technology.
[27] A. YACINE, B. RACHID, B. JAWAD. (2004). Learning Objects Based Framework for Self-Adaptive Learning, Education and Information Technologies 8:4, 345–368, 2003, 2004 Kluwer Academic Publishers. Manufactured in the Netherlands.
[28] R. M. Andrea, E. Max J. (2003). Determining Semantic Similarity among Entity Classes from Different Ontologies, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 15, NO. 2, MARCH/APRIL, 1041-4347/03, IEEE.
[29] K. Kozaki, Y. Kitamura, M. Ikeda, and R. Mizoguchi. (2002). Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of Role" and "Relationship", Proc. of the 13th International Conference Knowledge Engineering and Knowledge Management (EKAW2002), Sigüenza, Spain, October 1-4, 2002, pp.213-218.
[30] K. Kozaki, Y. Kitamura, M. Ikeda, and R. Mizoguchi. (2002). Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of Role" and "Relationship", Proc. of the 13th International Conference Knowledge Engineering and Knowledge Management (EKAW2002), Sigüenza, Spain, October 1-4, 2002, pp.213-218.
[31] IEEE Standard for Developing Software Life Cycle Processes, IEEE Computer Society, New York (USA), April 26, 1997.
[32] http://www.w3.org/TR/WD-logfile.html.
[33] http://www.w3.org/TR/owl-ref/.