Detecting the learning troubles through a semantic analysis of the posted messages

Samia Ait Adda  
Department of Computer Sciences,  
Mouloud Mammeri University of Tizi-Ouzou  
Algeria  
s_ait_adda@esi.dz

Balla Amar  
National High School of Computer,  
Ouad Smar, Algiers, Algeria  
ESI  
m_balla@esi.dz

Abstract— Communication tools on distance learning platforms are ways that allow learners to exchange messages between them or with their teachers. It is also a way to interpret their social behavior patterns and their learning styles. In this paper, we are interested in the semantic analysis of the content's messages published by learners by use of domain ontology. The purpose of this analysis is to identify the domain concepts that are most published and shared by learners and to keep them into the learner’s model as concepts not well mastered. We hypothesize that all concepts edited and exchanged over email, chat and especially in discussion forums can be considered as knowledge poorly or badly acquired by learners and deserve thus more attention and consideration both by the tutor for the pedagogical monitoring of learners on these concepts and from designer of course, to restructure and more enrich the educational contents which articulate these concepts identified beforehand in this analysis.

Keywords— e-Learning, communications tools, domain ontology, conceptual indexing, learner model.

I. INTRODUCTION

Our work is placed in the field of the Interactive Educational System; we set us therefore in a context of an online learning. Hence, learners work remotely on platforms which allow them to attend course, make tests and exercises or discuss by means of communication tools that are always integrated on these platforms.

The progressive integration of these communication tools offers opportunities for learners geographically dispersed to exchange without any time constraint and allows considering a new forms of social interaction between learners and between learners and teachers. [1] With this simplified information sharing, new forms of interaction have emerged, and new skills have been developed both in social, cognitive, or metacognitive, [2].

These communication tools are generally places of meetings and discussing for learners who are often in difficulties on some concepts of the taught domain, which may be weakly assimilated. If we take the example of a learner who publishes a question on a forum, this fact can interprets an obstacle on concepts contained on the content of this question. As well as the messages exchanged between learners by email or chat.

We aim in this article to analyze the content of messages discussed by the learners to identify the most exchanged concepts and be able to help learners in their needs. We introduce among other things, to do this, ontology of taught domain to correlate information between the different contents of messages published by the learner and those of the studied course. The goal that we want to achieve through this study is essentially enriching and updating the learner profile, by marking the domain concepts that pose a problem for him and that are detected due to its posted messages. Recognize the problematic concepts of the learners participating in the learning process can help the designer of the course to review the course content on these concepts and the tutor to assist and support learners who are already published these concepts.

The paper will be structured as follows, in the first section we present some research that are carried on the analysis of communication tools and their users, then we are interested in some semantic web tools, such as ontology and the possibility, thanks to these tools to exploit the contents of messages to detect concepts badly acquired by learner, further we detail and explain our approach, we finish with a conclusion and some perspectives.

II. RELATED WORK

In several studies, the ability of message to structure the written exchanges may offer an interesting educational trail for helping learners to succeed. Bruer [3] highlights the need to translate his thoughts in writing forces the user of message to a rigorous structure and requires a time for reflection. This advanced online writing will encourage a more deep communication and learning. So for Mangenot [4], we may recognize in forum the benefit to mobilize ideas, through the gradual structuring of discussion and to allow in this way the gradual emergence of themes.

From a constructivist point of view, the recourse to messages stimulated students to make connections between the many messages in which they are confronted and give them the opportunity to help building their own comprehension.

Another important feature of this written output relates to the permanence of messages. In [5], the archiving of messages provides indeed the possibility for learners to consult them thereafter at any time. Asynchronous communication’s space also guarantees a real capitalization of knowledge. In [6], the forum can be considered as a free complement and practice to
deepen in concepts discussed in the context of online learning. Some authors [7] analyze the external factors of messages, these analyzes attempt to account for what is played on forums from readily observable indicators such as the number of posts by each user or group, the number of responses, the length of discussions, the number of learners participating in discussion the average duration of a session on the forum … etc.

These studied works aim to show the cognitive, social and constructive advantage of published messages. Our approach is inscribed in the same perspective. On the other hand, we propose a method which gives a semantic meaning to learners messages with an ultimate goal of detecting concepts of the studied domain that are poorly learned by the learner. To do this, the concepts of the studied domain are modeled through ontology of the taught course, as we describe it below.

III. DOMAIN ONTOLOGY

The Semantic Web [8] is an understandable and navigable space by both human and software agents. It introduces an additional meaning to the navigational data of the classical web, based on a formal ontology and controlled vocabularies through semantic links. In standpoint of e-learning, it can help learners to locate, access, querying, processing and evaluating learning resources across distributed heterogeneous network, or assist teachers in creating, using, locating, or the sharing and exchanging learning objects. The semantic web technologies [9] have also been used as an alternative to allow adaptability and interoperability of e-learning systems. This is to overcome the limitations of the current web to promote interoperability, enabling different applications to share, interact, collaborate and exchange resources. It also allows the development of intelligent agents [10], capable of ensuring adaptability, using inference and search engines, implemented through domain ontology [11].

Ontology [12] includes a set of terms, knowledge, including vocabulary, semantic relations, and a number of logic-inference rules for some particular domain. The ontology applied to Web creates thus the Semantic Web [13]. Ontologies [14] facilitate the sharing and reuse of knowledge, i.e. a common understanding of diverse content by persons and machines.

The use of ontology in learning environments aims to provide mechanisms to improve the search process and semantic discovery of learning resources. It also offers the capacity to organize and display information such as the viewing of relationships between concepts.

In our case, the role of domain ontology lies in the conceptual indexing of the edited messages to facilitate their identification and semantic search of such content by the learner since they become knowledge’s basis available for consulting [5] [6]. On top of that indexing, the most edited domain concepts will thus detected. This ontology also represents the structure of the learner’s model, since it is part of the domain model, i.e. the domain ontology in our case.

The model of domain ontology that we propose is shown in Figure 1. The considered educational resources are described by a set of metadata (LOM) [15].

![Figure 1. Model of domain ontology](image)

IV. THE ARCHITECTURE OF PROPOSED APPROACH

To detect knowledge of domain supposed poorly assimilated by learners, we propose an approach that consists in constructing a text document relative to the messages published by the learners in their email, chat and discussion forums.

These documents are taken from the database of the platform and then indexed according to the concepts of the taught course through the domain ontology (conceptual indexing). The most edited concepts by a learner or group of learners are then highlighted.

Our goal is threefold. On the one hand we try to index the messages to facilitate their research and consultation, on the other hand to detect the most edited concepts of domain by learners, and finally we want to identify learners who have used these concepts via the communication tools to appropriate more knowledge about some domain concepts.

The approach that we propose is divided into three basic processes: (1) Building of the messages corpus, (2) Conceptual Indexing and (3) the management and processing of results, this is what will be detailed in the following of this paper.

A. Building of the edited messages corpus

To perform indexing, we need textual content messages. It just consists to access on the data base of the platform to get massages exchanged by all learners from the corresponding tables. The result will be in the form of text documents, thus closing the content of message, paths to the attachments if exists, information about the sender and receivers of the message and the time of dispatching. These documents will be saved in a repository to constitute a local corpus of documents.

B. Documents representation

To assess the exchanged domain concepts and learners who have used it, which is the objective of this study, we represent each message by its keywords. The vector model is adapted in the proposed approach for an effective representation of documents. Each document is identified by a vector of n dimensions, where each dimension corresponds to different words. Each term in a document vector is associated with a weight. The weight is a function of term frequency, collection and normalization factor, \( tf \times idf \) (term frequency-inversed
Document Frequency). Different approaches for calculating frequency may be applied by the variation of the function. However a document i is represented by a document vector \( d_i \):

\[
D_i = (W_{1i}, W_{2i}, \ldots, W_{ni})
\]

Where, \( W_{kj} \) is the weight of the \( k \)th term in the document \( j \) and \( n \) as a total number of documents in corpus

The frequency of the term reflects the importance of the term \( j \) in document \( i \). The weight factor can be local or global. The overall weight factor takes into account the importance of the term \( j \) in the collection of documents, while the local weight factor considers the term only in the relevant document [16] [17].

\[
Tf_{ij} \times idf_j
\]

where \( tf \), \( j \) is the frequency of term \( j \) in documents \( i \), \( N \) is the total number of documents, \( f_i \) is the number of documents in the corpus containing the term \( j \).

A. The conceptual indexing

Once we have the keywords of each document of the corpus (the term vector), we proceed to their conceptual indexing. To perform conceptual indexing, we need an external resource that offers an organization by concepts and structure [18] [19].

We use in our case of study domain ontology formalized with SKOS\(^1\) format (Simple Knowledge Organization System) [20] [21]. We also use a mapping pattern in the level of concepts (ontology), a tool for selecting terms (from the document) able to solve the semantic ambiguities, and finally a conceptual weighting system.

1) Concept Identification

The purpose of this step is to identify ontology concepts that correspond to document words. Concept identification is based on the overlap of the local context of the analyzed word with every corresponding ontology entry. Concepts are referred to in text documents with simple or compound words. These terms have been extracted during the preceding phase as a vector component. It remains to make the correspondence between the terms and concepts (the mapping on the domain ontology).

This ontology contains a set of concepts \( C \) and a set of relationships between concepts \( R \) (\( c_1, c_2 \)). As shown in the ontology extract of the PHP course over the Figure 2, each concept is associated with one or more terms with SKOS properties "prefLabel" (preferred label) or "altLabel" (alternative label), as well as a semantic relationship. Hence, concept C2 is a sub concept of C1 via the hierarchy link "Broader" of SKOS.

\[
\text{idf}_i = \log(N/f_i) + 1
\]

Where \( tf \), \( j \) is the frequency of term \( j \) in documents \( i \), \( N \) is the total number of documents, \( f_i \) is the number of documents in the corpus containing the term \( j \).

In the ontology, a set of terms is used for labeling concepts and relationships between concepts. That set forms the vocabulary of the ontology. Thus, the method that we proposed for extracting concepts from the vector document is to assign to each term of document's vector a concept associated to an entry of the domain ontology. To respond nevertheless in case if the processed term is ambiguous, a disambiguation step is so necessary.

2) Term Disambiguation

Each term \( t_i \) in document may be associated to a number of related possible ontology concepts. Thus we distinguish the situation of semantic or polysemy ambiguity. For example, the term "table" has a three meaning in PHP ontology: (1) table of data structure, (2) table in database, and (3) table in the html structure. It can refer to three different concepts. In this case we proceed as follows, for an ambiguous term \( t_i \) in the document, we seek a label of a concept \( C_k \) linked in the ontology with a concept \( C_i \) which is indicated by the ambiguous term \( t_i \). If \( C_k \) exists, \( C_i \) is taken as the concept designated by the term \( t_i \).

3) Concept Weighting

The extracted concepts are weighted according to a method more general than \( tf \) *idf named \( C_{fc} \) * idf (concept-frequency-inversed document frequency). In this method each extracted term represents necessarily a concept of the ontology since we used ontology to identify them. For a concept \( C \) its frequency in a document depends on the frequency of the word itself [22]. It is calculated as follows:

\[
C_{fc} = \sum_{t \in \text{term}(c)} tf_{tm}
\]

Where: \( t \) (\( c \)) is the set of terms corresponding to different concept \( C \).

The weight of each concept in a document \( d \) is so calculated as follows:

\[
C_{f \text{idf}} = C_{fc} \times \text{idfc}
\]

\( \text{idfc} \) is the inverse frequency of the concept \( C \) by counting the number of documents in which appears the concept \( C \), analogously to the weight of each term in a document (see formula 2).

---

\(^1\)<http://www.w3.org/2006/07/SWD/SKOS/reference/20090315/implementatio n.html>
V. SEMANTIC REPRESENTATION OF DOMAIN KNOWLEDGE

Each indexed document is represented by a vector of weighted key concepts. For this purpose, all documents that constitute the corpus C will be represented by an occurrence matrix of document and concept. We will distinguish two kinds of corpus, a first corpus Clearner consists of each messages published by a learner, and a second CGRP represents the sum of all messages edited by all learners who participated in the learning process.

Thanks to the first corpus, we can specify the knowledge which pose problem to the learner. Since the occurrence matrix, we naturally recognize the most published concepts by the learner through adding instances of the same matrix row; the result is a vector Vlearner containing the concepts occurrences in the corpus Clearner.

\[ V_{\text{learner}} = (W_{e1}, W_{e2}, \ldots, W_{en}) \]

Concepts with a high weight by estimating a threshold \( a \), that we will fixed by experimentation, will be reviewed as problematic domain concepts for a learner. Therefore, the tutor may intervene to help the learner on these concepts.

For the general corpus CGRP, we proceed with the same process as a learner, so the result will be a general vector of concepts occurrence in the corpus CGRP. Similarly, the concepts greater than a threshold \( \gamma \), which we also determine by evaluation, will be considered as wrong developed concepts in the course. To this end, the designer of course can review the content of resources that explain these identified concepts, and further enrich its course on these concepts.

VI. BUILDING AND ACQUIRING OF THE MODEL KNOWLEDGE MODEL

The learner’s models are cognitive models which allow to provide relevant information for a learning system in order to adapt learning to the knowledge, competences, features, preferences and objectives of apprenticeship to learner in particular domain (these which is taken in the learning environment) [23].

There are five popular and useful features for an individual learner’s representation, these are: the learner’s knowledge, interests, goals, background, and individual traits [24]. The user’s knowledge of the subject being taught or the domain represented in hyperspace appears to be the most important user feature. For existing learning adaptive system, the knowledge is frequently the only user feature being well-modeled [25], [26].

We distinguish several ways to represent the learner’s knowledge; the largest used method is the Overlay model [27]. This model represents the learner’s knowledge as subset of the domain model, which reflects the expert-level knowledge of the subject. The domain model is presented in our case of study as domain ontology.

For each domain concept, the Overlay model stores estimation for the level of user’s knowledge. This estimation, may be Boolean (known or not known), qualitative (good-average-poor), or a quantitative measure, such as the probability that the user knows the concept. The power of this model is tied in fact that is extensible, flexible and can measure independently the user's knowledge for different domain concepts. One extensions of the overlay knowledge model is a layered overlay model. A layered model stores several values to represent the state of user’s knowledge of each concept. These values can be updated independently and without crushing.

Hence, in our case of study, we propose an overlay knowledge model with two layers (2 levels). The first layer concerns assessment and contains the mark obtained by the learner in the test on the concept. As for the second layer, it stores the weight of the edited concept on the communication tools using (4).

Technically, overlay knowledge model of user is represented as triples of « concept-aspect-value », in which « aspect » interprets the layer. The details and the structure of the domain model (domain ontology) are stored once in the system and are typically not replicated for individual learners. As a result, the amount of information stored for each individual learner is very small, whatever the large of the domain ontology.

The state of the learner’s knowledge may be changed from session to session and even several times within the same session. This means that an adaptive system relying on learner knowledge has to recognize the changes in the learner's knowledge state and update the learner model accordingly. Thus, for a better representation of the learner's knowledge state in each learning session, we suggest adding the identifier of this session. From that moment on, the learner’s mastered level for a concept is represented like quadruple of « concept-session-aspect-value ».

Below figures an example of the learner’s knowledge state presentation of the concept C1 during 2 sessions and for two sources of observation, these are Assessment and communication.

![Figure 3. Example of presentation of the learner’s knowledge state on concept C1](image)

Consequently, the learner model in the proposed system is shown in an ontological form since using the domain ontology to represent the domain model.

The learning model is constructed by mapping information and shared learning on ontology concepts. Thereby converting the content exchanged by the learners in the form of a concept of ontology, and thanks to these concepts the learner’s knowledge model is thus built.
In this kind of system, different layers are maintained separately and combined only during the adaptation process. Therefore, to compute the value $V$ of concept $C$, we use the following equation:

$$V = \frac{0.7 \times (\text{mark}) - 0.3 \times (\text{edited weight})}{2} \quad (5)$$

VII. EXPERIMENTATION

A. The Test Collection

For our experiments, we have proceeded to test on group of computer science students in the second years, with the number of 27. We have proposed to them PHP course, shown in eFAD\(^2\) platform and modeled with the ontology of SKOS format. The course was presented in three sessions of one hour.

The PHP course is mainly composed of 8 top concepts and 49 sub-concepts. To consolidate our experiment, we have conceived a questioner paper which we have distributed to students, asking them to place concepts that pose them problem. At the end of the test, a written assessment was performed for all learners.

B. Evaluation of results:

The result of the experimentations consists of 27 corpus of each learner plus the general corpus. Therefore, we constituted a number of 105 textual documents of the messages extracted from the platform database. The following diagram shows the editing weights characterizing the main domain concepts for each learner:

![Figure 4. The editing weights of the main domain concepts](Image)

The challenge of this test was to find the concepts insufficiently mastered by each learner. That is to say; the threshold $\alpha$, that we set to evaluate the most concepts which posed problem for student $j$, is estimated by the median of the weights of concepts edited by this learner. This threshold is different from learner to learner; accordingly we have counted 27 values of this threshold. Indeed, we found that the concepts C4 and C6 have posed a problem for some students who are recognized by the following process.

As that is signaled, a written assessment was performed for each learner on each concept of the domain as well as a questionnaire which we have asked them to indicate the concepts not mastered. Therefore, the following diagram shows a comparison of different results obtained by the assessment, the weight of edited concepts and questionnaire responses.

![Figure 5. Synthesis of the Mark-Weight-Questionnaire results for each learner](Image)

As for the threshold $\gamma$, which we considered to estimate the concepts which are badly defined in course, it is determined by comparing the result with that obtained through the written assessment and questioner responses, the value is fixed at 0.24. As a result, we detect that 5 learners have problems on some concepts of domain (learners 2, 10, 14, 18, 22).

VIII. CONCLUSION

The educational adaptive systems provide good support for learners on their individual characteristics. It can also provide information on the needs and deficiencies of these learners, either for tutor or designer of the course, even for evaluation, monitoring and customizing the process and strategy of learning. In fact, the learner model must be developed for each student, containing information of the history of social interactions, objectives and knowledge badly acquired.

In this article, we have highlighted the need to analyze the messages edited by learners during the learning sessions and we have proposed an approach for semantic analysis that we have presented and explained which permits to detect domain concepts that were difficult for learners, by comparing the content of their messages with domain ontology of the studied course. An experiment was carried out on a group of students taking a PHP course, and has enabled us to validate the proposed approach and to set some parameters.

This result needs to be further refined by additional tests, which we are currently conducting. Moreover, we considered only the context of communication, our outlook is to take into account the context of research and visit on the Web, that is to say, detect searched and viewed concepts on the Web.

IX. REFERENCES


\(^2\)www.ufc.efad.dz
Conference, ESWC and Ontologies as Services

Information Retrieval Driven by Ontology from Query to UK, 4 June to 8 June 2005.

Conceptions of Libraries and Information Science, Glasgow, Representation”, CoLIS5: Fifth Inter

1979.

K. Sparck Jones

S. Robertson and K. Sparck Jones.

Metadata , 2006

Guide for IEEE 1484.12.1

IMS Global Learning Consortium. IMS Meta


U. Siti, A. Rohiza, A.

Heidelberg A


