Social Semantic Web-Based Analytics Approach to Support Lifelong Learning

Khaled Halimi, Hassina Seridi-Bouchelaghem

Abstract—The purpose of this paper is to describe how learning analytics approaches based on social semantic web techniques can be applied to enhance the lifelong learning experiences in a connectivist perspective. For this reason, a prototype of a system called SoLearn (Social Learning Environment) that supports this approach. We observed and studied literature related to lifelong learning systems, social semantic web and ontologies, connectivism theory, learning analytics approaches and reviewed implemented systems based on these fields to extract and draw conclusions about necessary features for enhancing the lifelong learning process. The semantic analytics of learning can be used for viewing, studying and analysing the massive data generated by learners, which helps them to understand through recommendations, charts and figures their learning and behaviour, and to detect where they have weaknesses or limitations. This paper emphasises that implementing a learning analytics approach based on social semantic web representations can enhance the learning process. From one hand, the analysis process leverages the meaning expressed by semantics presented in the ontology (relationships between concepts). From the other hand, the analysis process exploits the discovery of new knowledge by means of inferring mechanism of the semantic web.

Keywords—Connectivism, data visualization, informal learning, learning analytics, semantic web, social web.

I. INTRODUCTION

BEYOND formal learning activities designed by specialists of official educational institutions, a great need to take into consideration is the everyday learning experiences which we encounter everywhere that define “lifelong learning”. We learn throughout our life from our first words to our oldest age, we make new experiences, acquire new knowledge and new skills. Certainly, we also learn at school, in business, at university, and in training institutions; but even in these instituted places of training and learning, what we learn to be really important often has nothing to do with official programs. We experience situations, acquire skills, test our emotions and feelings in the most effective ‘school’: the ‘University of Life’ [1]. So, we learn and train ourselves in conversations with friends, watching television and reading books, flipping through catalogues or surfing the Internet, as well as when we think and make plans. We will still always be “lifelong” learners.

After the extensive use of Web 2.0 tools, the learning process occurs through users’ connections which enable knowledge construction within networks [2]. In this way, a new learning theory has emerged: it is the “connectivism” which aims to understand learning processes by asking how people use and develop their networks of social relations for their learning and professional development [3].

Since lifelong learning is a strongly learner-centred and learner-controlled process and with large data generated by users, learning analytics approaches (LA) combining with use of social semantic web technologies represent recent trends in the development of learning systems that support lifelong learning. These techniques can understand how the learning occurs and can be enhanced in networks, how learners create meaning and construct knowledge when connecting with others, and how the learning takes place, etc. [4].

The major aim of this work is to describe how the lifelong learning process can take advantage of the learning analytics approaches, semantic web and the connectivism theory in order to enhance and support the learning operation. For this reason, a system called SoLearn (Social Learning Environment) that supports this approach is presented. Therefore, all system’s knowledge is modelled by Semantic Web, then a deep analysis of learner’s behaviour and learning content using Semantic Social Web techniques is conducted; results of these analytics will be provided to learners as recommendations and data visualisation in order to facilitate, support and enhance informal learning. We call a “learner” any person looking for any kind of knowledge in any learning situation: formal, non-formal or informal learning.

The rest of this paper is organised as follows. In next section, a brief overview about related concepts of lifelong learning, connectivism and learning analytics is introduced. In Section III, the model of the semantic learning analytics approach is presented. In Section IV, the approach is evaluated through experimentation. Finally, we conclude in Section V.

II. RELATED WORK AND THEORIES

Lifelong learning is the main mode of people’s learning outside schools. It defines learning that happens in everyday activities related to work, family or entertainment. It is unorganised according to objectives, time, and learning services. It can be a tacit learning where the learner is unconscious of and not intentional; an incidental learning, which was not previously planned but it is recognised when the learner becomes conscious of it as learning or self-directed learning when learner initiates learning and is aware of it as learning [5].

Recently, research focused on social software to design systems that support the lifelong learning process. For instance, [6] indicate that people can learn in multiple ways using Web 2.0 tools. Reference [7] proposed to use social software informally to help students in formal school learning.
Reference [8] proposed a Web 2.0-based learning tool to manage learning resources through collaborative analysis of content. Reference [9] claims that social networking tools have a big capacity to enhance learning, since these tools include all the features of interactivity, collaboration, active participation and resource sharing.

It is essential to perceive that social technologies have changed meaningfully the learning concept. As a result of this perception, Siemens presented the “connectivism theory”, which we believe to be the best learning theory that supports lifelong learning in a social context. The connectivism with its eight principles was introduced by George Siemens and Stephen Downs and was richly described by Siemens in [10]. Connectivism is defined by Downs as “the thesis that knowledge is distributed across a network of connections” [3]. He assumes that being a member of a network, communicating with others and being able to filter information and ideas will lead to knowledge creation and progress of learning. Connectivism supports the active participation of persons in communication, rather than transferring knowledge from educator to learner. In addition, it promotes a learning organisation, where knowledge is spread over the network and users’ participation is the learning itself. We believe that connectivism may seem attractive in the sense that the process of the lifelong learning would rely on the students’ ability to build their knowledge based on connection with other sources.

Since the lifelong learning has no official beginning, end or well-defined learning objectives, it is challenging to assist learners during their learning. Thereby, the use of learning analytics (LA) approaches, based on artificial intelligence and data mining tools, could be very necessary to support and enhance the learning process. LA aims to develop tools making learning activities ready for analysis [11]. LA used to collect and analyse information of learners’ activities, such as: communication, posting on forums and online social interactions, etc.

According to the definition of the Society for Learning Analytics Research [12]:

“Learning analytics is the collection, measurement, analysis and reporting of data about learners and their contexts”.

Reference [13] defined LA as:

“...an emerging field in which sophisticated data analysis tools are used to improve learning and education”.

Reference [14] defined LA as:

“The application of analytic techniques to analyse educational data, including data about learners’ activities, to identify patterns of behaviour and provide conclusions to improve learning”.

Referring to study in [15], LA would be based on: users’ skills and competencies, users’ characteristics (educational history, learning objectives, skills, emotional status and knowledge), user-generated data (shared resources, social interactions and practices, etc.) and external data (online searches, social networking, communication, video-conferencing, etc.). LA could guide learners to reflect on their actions and outcomes and if necessary, provide appropriate recommendations for improvement.

Semantic learning analytics which represents the main core of the presented approach in this work, is one of the most recent developments in the LA field, the fact that the analysis process can greatly benefit from the possibilities of web semantic, especially the meaning expressed by semantics presented in the ontology and the discovery and the inferring of new knowledge [16].

Some approaches that apply different concepts mentioned above will be presented in the following.

Reference [17] analysed the social network of a connectivist MOOC through the use of the centrality, proximity and intermediacy measurement. Reference [18] identified types of behaviour by measuring the correlation between these behaviours and students’ skills of the 21st century. He used the factor analysis and the regression models. The latter takes the students’ activities as independent variables (send a message, share information, become a friend, join a group, etc.) and the students’ skills as dependent variables (negotiation, networking, critical thinking, gambling, multitasking, appropriation, transmedia navigation, etc.). Reference [19] proposed a method to measure students’ engagement from their digital portfolios and to show how these new indicators can improve the quality of prediction. They used the Classification Algorithms (Naïve Bayes, decision tree, logistic regression, etc.). Reference [20] evaluated LARAe, a dashboard for teachers that provides a graphical representation of traces, badges and student activities, etc. MeLOD [21] is a system that supports analytics of learners’ activities in a mobile learning setting based on Semantic Web. PBL3.0 (Problem-Based Learning 3.0) [22] is a system that integrates learning analytics and semantics in problem-based learning.

In order to develop our approach, we conducted a comparative study between the systems mentioned above. We took into consideration the different services they provide. Throughout this comparison and on the basis of concepts and theories mentioned above, we were able to extract necessary features to support the lifelong learning process. We describe in detail in next section our approach implemented into SoLearn, a semantic social-based environment that supports lifelong learning.

III. PROPOSED APPROACH

This work represents an approach aiming to organise the data generated by students through a semantic representation and to extract the necessary knowledge in order to find out patterns for future analysis.

A. Conceptual Model

In the following, we present in detail the conceptual model of the proposed approach as presented in Fig. 1. Firstly, all the knowledge used is modelled by the Semantic Web. Then, a deep analysis of learner’s activities and content using the MeaningCloud text analysis is performed. Results of the analytics process will be provided to learners as
recommendations and data visualisation in order to facilitate and support their learning experiences.

The MeaningCloud API [23] enables users to embed text analytics and semantic processing in any application. It combines the use of a semantic web with a cloud-based framework. This is what prompted us to choose this tool, the fact that our approach implements a semantic representation of learners’ activities and learning resources. Steps of the analysis approach are summarized in this algorithm:
1. Data collection and storage.
2. Data processing and filtering.
3. Apply analytics using MeaningCloud.
4. Send recommendations based on the analysed data.
5. Show visualisation based on the analysed data.
6. Save history of analysis to the knowledge repository.

1) Knowledge Representation

In order to realise our objective, all the knowledge managed in the system is described with the use of an ontology, to formalise the knowledge of the learning domain. The use of Semantic Web technologies allows us to enhance the semantic representation with standardised tools and to associate formal descriptions to learning resources, to make formal reasoning (resources retrieval, resources compositions, etc.), to search pedagogical resources tailored to the learner, to compose new resources from existing resources and to adapt the interaction between the system and the learner according to her/his preferences [16].

a) The Proposed Ontology

We developed an ontology that describes three types of knowledge: concepts, object properties and individuals. For each property, a subject and object concept is defined. The use of Semantic Web technologies allows us to enhance the semantic representation with standardised tools and to associate formal descriptions to learning resources, to make formal reasoning (resources retrieval, resources compositions, etc.), to search pedagogical resources tailored to the learner, to compose new resources from existing resources and to adapt the interaction between the system and the learner according to her/his preferences [16].

The learners’ model stores all the knowledge about their actions and activities in RDF files. We have modelled different types of knowledge: personal information, knowledge level, learning style, learning interests, friendship networks, questions, messages, tags, emotions and actions, etc. As shown in Fig. 3, the User class is related to “LearningDomain” class and “LearningStyle” class through the “WantToLearn” relationship.
The emotions of learners are presented using the property “hasEmotion”.

2) Learning Content Model

To facilitate the learning analytics process, the learning objects’ model was designed as low granulated pedagogical entities with a semantic representation in order to support sharing, reusability and flexibility. The model is based on dividing the course into a set of learning entities (definition, summary, illustration, example, etc.) and each entity could have different formats (for example: definition is a text file while illustration is a video), while the semantics relationships between these entities are also defined, for instance <definition, isPrerequisiteOf, illustration>.

![Image](https://via.placeholder.com/150)

**Fig. 3 The learner’s model**

C. Data Pre-Processing and Filtering

Data pre-processing is an important step in the field of learning analytics. The fact is that we live in the world of social Web and informal learning, where any user can create the knowledge and share it, and therefore, data can be incomplete, noisy, and inconsistent and it does not make any sense to analyse in this form. The study employed the inference mechanism of the Semantic Web in order to provide a formal representation of learners’ activities. In general, the inference is a process of reasoning which is based on knowledge acquisition and allows obtaining new information and discovering of new relationships between instances. For instance, by returning to the ontology, the system can easily realize that the concepts: “Object-Oriented Programming”, “OOP”, or “Object-based Programming” have the same meaning, the fact that these concepts are related by the “sameAs” relation. Without using the ontology, a user interested in “Object-Oriented Programming” and another one interested in “OOP” will not have same interests, since the two terms are syntactically different despite the fact that both belong to the same learning domain.

All knowledge about content and users’ actions generated in all previous phases will be stored in the knowledge base as RDF files, in order to use them in different learning analytics modules which will be presented in the following section.

D. Learning Analytics

In this section, we describe how the analytics approaches can be applied to support the lifelong learning process, and how they can get the benefit of the Semantic Web models. In fact, tracking learners’ activities by means of a semantic model, supports the implementation of analytics processes that can take advantage of the meaning expressed by semantics contained in the relationships between concepts. The key element of the analytics approach presented in this work is based on the use of MeaningCloud API. The latter performs various operations, namely: text classification, sentiment analysis, language identification, topic extraction and text clustering.

1) Application of the Analytics Process

In the following, we will see how both teachers and students get benefit of the analysis process through the following scenario presented in Fig. 4. Suppose that a teacher following the pedagogical model presented before adds “course1” that consists of a set of pedagogical entities: “introduction”, “example1”, “example2”, “definition”, “summary1”, “exercise2” and “exercise3. As we can see, “example1” has the format Image, defined semantically with the triplet: <lo:Example1 lo:hasFormat lo:Image> and “Exercise2” has the format: Text.

Students during their learning perform several actions on course1, they can comment, tag, share, like, dislike any of its entities. Analysing these actions through the learning analytics module will give conclusions for both teachers and students as follows:
a) Sentiment Analysis

We use the sentiment analysis module to analyse students’ comments; for instance, in order to determine if they express a positive, negative or neutral sentiment. To do this, the local polarity of the different sentences in comments is identified and the relationship between them is evaluated, resulting in a global polarity value for the whole text. Additionally, sentiment analysis detects if the text processed is subjective or objective and if it contains irony marks; the latter gives the teacher additional information about the seriousness of learners in dealing with the course. To proceed, we use the RDF models of learners as presented in Fig. 5.

```
1: lo:student
2: lo:putTag lo:PPP
3: lo:putComment lo:comment1
4: lo:dislike lo:dislike
5: lo:putTag lo:SNV
6: lo:wantToLearn lo:AJAX
```

Fig. 5 Extract of the learner’s RDF file

Then, we apply a SPARQL query to extract learners’ comments as shown in Fig. 6.

```
1. SELECT DISTINCT ?comment
2. WHERE { ?x lo:putComment ?comment
```

Fig. 6 SPARQL query to extract learners’ comments

Submitting learners’ comments to the sentiment analysis module shows that they created a controversy about the difficulty or the clarity of Exercise2. Most students who have taken this exercise showed a negative reaction, contrariwise they show a positive reaction towards Exercise3. The system then, detects that Exercise2 is probably inappropriate and suggests to the teacher to redesign it with what suits the students’ cognitive levels.

b) Social Network Analysis

Social network analysis (SNA) tools could be used for topics analysis to identify potential communities in the network which may share similar learning attitudes and behaviours. As we can see in Fig. 4, we can group students that have a positive reaction towards such a pedagogical entity in the same clusters, and then observe how students contribute and how they influence the evolution of these communities. Participation can be measured via centrality analysis which is mostly used in SNA research.

c) Text Classification

Using the text classification module, we can extract concepts contained in any text shared on the system (documents, posts, answers, etc.); then, we project extracted concepts to the ontology, and through the use of semantic relations, inference rules and SPARQL queries, we can give learners all the necessary knowledge to understand any content, namely: the domain and objectives of content; semantic relations with other concepts; users who deal with the content; etc.

d) Text Summarisation

Using the text summarisation module, learners can easily extract a summary for a given document, select the most relevant sentences and get a whole overview of what it is about. For instance, quickly, they have an overview of the teachers’ courses, so that any user can understand the content of the course whatever his/her background, his/her level or his/her age.

E. Recommendation and Visualisation

1) Recommendation

The recommendation module, which is based on the results of the analytics approaches presented above combining with use of inferring and reasoning mechanisms of the Semantic Web, gives us a complete overview about learners’ behaviour and activities and of the teachers’ resources. Thus, the system will be able to determine what type of actions, content or users must be proposed to learners in order to help them achieve their learning goals. On the one hand, the system recommends for authors to redesign courses or parts of the course, so as to
assist learners who encounter problems, communicate with learners who show negative sentiment, and to encourage positive learners to continue working, etc. On the other hand, the system recommends for students to communicate with learners who have the same interest and learning objectives and proposes for them a set of learning objects with different formats, students who show a clear expertise in a certain learning domain and students who show a strong desire to be social users, etc. Here is an example of implicit recommendations:

\[
<\text{lo:Student1 hasNegativeEmotionOn lo:Exercise3}> \text{ and } <\text{lo:Exercise3 required lo:Summary8}> \text{ } \Rightarrow \text{ } <\text{lo:Student1 shouldTake lo:Summary8}>.
\]

Therefore, the system recommends to student1 to take the summary again, because, s/he did not understand it properly.

2) Visualisation

The system provides a visualisation tool to help teachers monitor learners’ activities such as their social activities, learning outcome, learning progress, etc. On the other hand, the system gives learners, using a simple graphical representation, the possibility to understand their behaviour, their activities and to extract interesting conclusions that can be used to enhance their future learning experiences. For instance, a teacher can use a chart showing the number of comments mentioning a particular topic to confirm whether it was discussed according to the syllabus of his course. A bar chart represents learner’s social actions: Commenting, messaging, tagging and friendship request. A gauge chart shows the learner’s social state at a period of time. Consequently, a teacher can identify which learners are not participating actively in learning sessions and can understand the reasons behind this lack of active participation.

IV. TEST AND RESULTS

An experimental study is in progress using a prototype of SoLearn (a system under development). Actually, the advantages of implementing a learning environment based on the Social Semantic Web technologies have been widely discussed and detailed in [16]. The purpose of the current experimentation is to demonstrate the effectiveness of the proposed concept that implements the use of learning analytics approaches combined with the use of Semantic Web representations.

A. Methodology

To realise the experimental study, we used a simple random sampling where we have chosen randomly 25 students. The topic of the study was general concepts on computing. We have focused only on the verification of the question: “Is there any relation between the difficulty or the design of a pedagogical entity and the learners’ sentiments expressed in comments about that entity?”

To answer this question, students were asked to download an exercise proposed by a teacher on the system. For the exercise, which is about “Java classes” and is considered a bit difficult compared to their levels, the students were asked to add comments on the system evaluating the difficulty of the exercise. At the same time, they were asked to answer the following question on a printed questionnaire: How do you evaluate the difficulty of the exercise? Where the students’ answers were rated vary from very easy, easy, neither easy nor difficult to very difficult, with corresponding values (-2, -1, 0, +1 and +2). Results of the sentiment analysis process and the questionnaire are presented in Fig. 7.

B. Results and Discussion

As we can see in Table I, results of the proposed approach show that, to some extent, there is a big convergence between results of the sentiment analysis and the learners’ answers on the questionnaire.

We note that the rate of learners who evaluated the exercise as “difficult” on paper and at the same time showed a negative sentiment towards it in the system is about 60%. Therefore, we can say that the students’ sentiments towards a pedagogical entity clearly determine its level of difficulty or its inappropriate design. The learning analytics module, was able to give the system means to detect where learning content is difficult without the need for human intervention. This will be very beneficial, especially in the context of informal lifelong learning, where the content is not structured and presented by anyone. In this way, the system will be able to automatically classify the content according to the discussion that is going on around it, and it can suggest to the content’s authors to
revise it and reorganise it.

<table>
<thead>
<tr>
<th>TABLE I RESULTS &amp; INTERPRETATION OF THE TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>N= 25 Learner Results</td>
</tr>
<tr>
<td>Sentiment analysis module</td>
</tr>
<tr>
<td>On the questionnaire:</td>
</tr>
<tr>
<td>Easy</td>
</tr>
<tr>
<td>Neither easy not difficult</td>
</tr>
<tr>
<td>Difficult</td>
</tr>
<tr>
<td>Very difficult</td>
</tr>
<tr>
<td>Sentiment analysis module</td>
</tr>
<tr>
<td>Positive sentiment</td>
</tr>
<tr>
<td>Neutral sentiment</td>
</tr>
<tr>
<td>Negative sentiment</td>
</tr>
<tr>
<td>Strong negative</td>
</tr>
<tr>
<td>Interpretation</td>
</tr>
<tr>
<td>General matching (15 learners)</td>
</tr>
<tr>
<td>Negative &amp; Difficult</td>
</tr>
<tr>
<td>Precise matching (08 learners)</td>
</tr>
<tr>
<td>Strong Negative &amp; Very Difficult</td>
</tr>
<tr>
<td>Negative &amp; Difficult</td>
</tr>
<tr>
<td>Wrong matching (02 learners)</td>
</tr>
<tr>
<td>Difficult &amp; Positive</td>
</tr>
<tr>
<td>In general, 15 learners have evaluated the exercise as difficult on the questionnaire and they showed a negative sentiment about the exercise on the system.</td>
</tr>
<tr>
<td>Three learners have evaluated the exercise as very difficult on the questionnaire and they showed a strong negative sentiment about the exercise on the system.</td>
</tr>
<tr>
<td>Five learners have evaluated the exercise as very difficult on the questionnaire and they showed a strong negative sentiment about the exercise on the system.</td>
</tr>
<tr>
<td>Two learners have evaluated the exercise as difficult on the questionnaire but they showed a positive sentiment on the system.</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

In this paper, we presented a conceptual model for using diverse concepts to study the efficiency of semantic social networks in supporting informal learning analytics, based on connectivism theory that takes into consideration knowledge creation and sharing through Web 2.0 tools. The analysis process enables the optimal use of data generated by students. In this context, LA can offer answers and proper tools to enhance informal learning settings. A system called SoLearn is built on the basis of a social semantic infrastructure with the final aim of supporting informal learning analytics. A preliminary evaluation shows that the implemented analytics approach is effective in enhancing the connectivist informal learning process by providing learners the freedom to choose what, when, where, with who and in what way they will learn. There are no conditions and barriers to learning.

REFERENCES