

## Intelligent machine for ontological representation of massive pedagogical knowledge based on neural networks

Abdelladim Hadioui, Yassine Benjelloun Touimi, Nour-eddine El Faddouli, Samir Bennani

RIME TEAM-Networking, Modeling and e-Learning-LRIE Laboratory-Research in Computer Science and Education Laboratory, Mohammadia School Engineers (EMI)-Mohammed V University in Rabat, Morocco

---

### Article Info

#### Article history:

Received Feb 24, 2020

Revised Aug 7, 2020

Accepted Nov 12, 2020

---

#### Keywords:

Artificial neural networks

Big data

E-learning

Linked data

Ontology engineering

---

### ABSTRACT

Higher education is increasingly integrating free learning management systems (LMS). The main objective underlying such systems integration is the automatization of online educational processes for the benefit of all the involved actors who use these systems. The said processes are developed through the integration and implementation of learning scenarios similar to traditional learning systems. LMS produce big data traces emerging from actors' interactions in online learning. However, we note the absence of instruments adequate for representing knowledge extracted from big traces. In this context, the research at hand is aimed at transforming the big data produced via interactions into big knowledge that can be used in MOOCs by actors falling within a given learning level within a given learning domain, be it formal or informal. In order to achieve such an objective, ontological approaches are taken, namely: mapping, learning and enrichment, in addition to artificial intelligence-based approaches which are relevant in our research context. In this paper, we propose three interconnected algorithms for a better ontological representation of learning actors' knowledge, while premising heavily on artificial intelligence approaches throughout the stages of this work. For verifying the validity of our contribution, we will implement an experiment about knowledge sources example.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Abdelladim Hadioui

RIME TEAM-Networking, Modeling and e-Learning

LRIE Laboratory-Research in Computer Science and Education Laboratory

Mohammadia School Engineers (EMI)-Mohammed V University in Rabat

Agdal AV. Ibn Sina Agdal Rabat BP. 765 Morocco

Email: ahadioui@gmail.com

---

## 1. INTRODUCTION

Cloud computing is an immense infrastructure based on different models for providing various services for software and hardware. Cloud computing paradigm has resulted from integration of online systems. Observable is that tertiary education is currently incorporating technological solutions (LMS) into education, that is to say, computerization of educational processes for the benefit of the current learners and the other coming learning actors alike. Recent relevant research [1] has shown the vital importance of LMS in many educational establishments. We can incorporate many learning activities in the form of learning scenarios in MOOC systems, which can generate massive traces through actors' usage. That is, in the course of learning, such systems produce big data building on the learning actors' interactions (including their effects and productions). These big data were dealt with in depth in our big data preprocessing work [2]. Based on the massive data produced via interactions, we proposed, as a first step, a machine learning system,

while respecting the rules and objectives described in [2]. The second step concerns ontological construction based on the massive data preprocessed by the operational massive data layer.

The aim is to produce massive knowledge using ontologies-grounded approaches as well as the concepts of artificial intelligence (artificial neural networks). The large amount of data produced by MOOCs users invites researchers in this field to study the best methods for extracting and representing knowledge in the fields of education. Following our analysis of traces produced by online learning systems (MOOCs) [3], we notice that, currently, one of the main obstacles hampering the evolution of these systems, at the level of teaching entities, is the absence of capitalization and representation of the knowledge resulting from actors' systemic interactions. To clarify the point, the extraction process of knowledge sources is highly complex because of the said absence; therefore, the learning traces produced are in a raw format that cannot be used. Accordingly, we introduce ontology-based system. Online learning systems are characteristic of limits and shortcomings notwithstanding the fact that these systems offer platforms, which are rich with educational functions. In such platforms we noted the existence of some limits linked to the lack of means of extracting and representing the knowledge produced by learning actors.

In this sense, through our analysis of the work [4-6], we notice that it proposes some solutions for extracting and representing the knowledge produced via MOOCs. However, on the one hand, it has been noted that these systems do not offer solutions for capitalizing the knowledge that circulates in these systems. On the other hand, knowledge extraction and representation systems proposed by the said research work have some processing limits due to the absence of a generic intelligent approach in their operations. However, the knowledge produced by actors is generally made up of massive traces (log files, data files, etc.), which take in most cases an unstructured format. Therefore, they do not have the semantic aspect; that is, they are in a noisy format, which cannot be used, and which makes no sense for the involved actors. In addition, in MOOC systems, there is no instrument for representing the actors' knowledge. Hence, comes the necessity of proposing a solution for the representation of learning actors' knowledge.

In this article, we study the intelligent ontological representation of learning actors' knowledge based on the massive data preprocessed in [2], while making use of the ontologically based semantic web approaches as well as artificial intelligence algorithms. Indeed, this stage represents the continuity of the realization of our framework whose architecture for extracting knowledge is based on the results of interactions of MOOCs' learning actors [7]. By analyzing knowledge representation methods, our proposed system reduces the limits characterizing knowledge extracted at the end of a learning session.

The main objective in this work is the construction and the representation of learning actors' knowledge resulting from their interactions in an online learning platform based on ontologies and concepts of artificial intelligence (neural networks). In this article, we present an intelligent machine for representing such knowledge. Accounting for such an objective, the article is made up of 6 sections: in the first section we lay out the general context of the study. In the second section, we shed light on the existing research relevant for the subject matter dealt with in this work. The third section points out the theoretical foundation of the taken approach. The fourth one presents our system of representing learning actors' knowledge. In the fifth section, we tackle the knowledge representation algorithms proposed in this work based on artificial neural networks. The final section is devoted to an evaluation of the system and, then, draws a conclusion and perspectives.

The existing knowledge representation systems cover many issues [8-10]; the most common ones are as follows:

- Correspondence between database attributes and ontological attributes for establishing connection between the elements of a database and those of ontologies
- Semantic similarity of acquired knowledge: The calculation of semantic similarity through identifying knowledge sources which are equal weight
- Calculation of the distances between knowledge sources for a better extraction
- Classification of collection measures
- Identification of a new knowledge source.

Considering the progress of approaches relating to knowledge ontological representation and the problems solved by researchers' proposals, we will start this work with a literature review of the current problems relating to the context of the present paper. Our study deals with knowledge representation systems specifically adapted to e-learning system, based on massive data already pre-processed by our work [2], which covers many constraints and problems.

## 2. STATE OF ART

Our research focuses on knowledge extraction from massive data pre-processed in [2] and their ontological representations. Therefore, emphasis is to be placed on the big data grounded ontology (NOSQL). It is noteworthy that knowledge representation remains a domain which is still knowing expansion by the concerned researchers. Indeed, our study reveals that numerous research projects have emerged to provide solutions to problems linked to the development of the relating fields [11]. That is, it has been used in medicine for complex processes such as the identification of DNA, in the satellite field for identification of knowledge [12], in addition to our field, namely, education in which numerous research projects have been conducted [4]. These include projects on the representation of actors' knowledge and the use of ontologies for the indexing of numeric resources [13]. In accounting for knowledge representation, we have reviewed the relevant literature, focusing on the works recently proposed by researchers in the field in order to arrive at the best solutions to the problems dealt with in our research. In this vein, focus is lent to many methods related to our research context.

### 2.1. Ontological engineering

Knowledge mapping represents an important stage in actors' knowledge representation. Given this fact, we have made an in-depth analysis of the best methods used by re-searchers in ontological mapping. The main objective of this stage is:

- To make correspondence between the database elements and ontological elements, on the one hand, and between ontology sources and ontological elements, on the other hand.
- To make correspondence between two or more ontology associated with our knowledge representation system. In fact, some researchers use linked or open data for the representation of know-ledge based on the SPARQL language [14], whereas others use the ontological representation [15] as a solution to problems relating to knowledge mapping.

Knowledge identification is achieved via the use of two types of methods, namely: non-supervised [14] and supervised Web methods processed by experts in the field [16]. In the second case, there are technical teams working on the extraction of knowledge produced by actors in the field of education. Currently, using systems which are based on web related technologies, the Web domain produces massive data resulting from users' interactions. Numerous researchers have worked for years on the integrated web' methods to better manipulate those data [17]. Building on their analysis, we have come to grips with linked and open data methods facilitating data access. As a case in point, the researcher in [18] has used such methods to facilitate access to cloud data. Linked and open data are based on a standard SPARQL [19], which offers a uniform approach for accessing massive data. To put it another way, web-based methods, including linked and open data, integrate digital resources or massive data for any area. Several works have been proposed in ontological mapping; among these works [20], we find articles which propose solutions related to ontological mapping between Dataset and ontology, on the one hand, and mapping between two or more ontologies on the other, as mentioned above. Accordingly, the main objective of mapping function is to make correspondence/matching between a given operational data layer and ontological knowledge representation for any given field (i.e., medicine, education, and so on). Thus, the above-mentioned correspondence is used to create an ontology common to knowledge extraction systems for the benefit of actors involved in online learning [7]. However, ontological mapping still encounters problems relating, for instance, to similarity between knowledges, as introduced in our work [2], which represents a preprocessing machine for big data produced by interactive actors. The preprocessed data output of our machine are represented in XML and RDF formats. Having reviewed some studies, we have noted that ontological mapping in the case of correspondences between ontologies is performed by a given type of methods and tools, whereas in the case of mapping between dataset and ontology, it is functioned by another type. Given this distinction, we have investigated the best methods for mapping whatever type of input knowledge sources. The Table 1 shows in detail the correspondence between knowledge sources (Dataset and ontologies) and ontologies output:

Table 1. Ontological mapping algorithms

Tools	Dataset and textural	RDF/OWL	SPARQL	Limits
ONTOP	Dataset	√	√	Specific to structured data.
DataMaster	Dataset	√	√	Process all types of texts
XApi	Both	√	√	
Word2Vec	Both	√	√	

Ontological learning: This process aims to fill the proposed output ontology [7] with knowledge extracted from the big data layer [2]. Ontological learning is implemented by some researchers who have

created algorithms in their research works [21], whereas other researchers use existing algorithms proposed by the scientific community. In the Table 2, we show the efficiency of some proposed algorithms by researchers who dealt with ontological learning in their scientific proposals:

Table 2. Learning ontology algorithms

Algorithms and Methods	input	output	Limits
Onto 4 KDD	Onto	KDD	The limits of these approaches appear either on the processing time or by the flexibility with the sources of knowledge.
KDD 4 Onto	KDD	Onto	
OntoKDD 4 Onto	OntoKDD	Onto	
Onto 4 OntoKDD	Onto	OntoKDD	
Onto 4 Onto	Onto	Onto	

In this stage, researchers [18, 21] proposed some ontological learning solutions for integrating the input knowledge sources (i.e., a dataset and an ontology) and the output ontology. For instance, in his work, the author undertook an in-depth study on ontological matching for knowledge extraction. During the ontological matching stages, he carried out the ontological alignment between one or more knowledge sources. Having analyzed these works, we noted that they are characteristic of shortcomings in terms of the used approaches. Our analysis of the results of the recently proposed works [3, 22] revealed that the best methods of ontological learning are built on artificial intelligence (AI) approaches due to their contribution with regard to Knowledge extraction.

**Ontological enrichment:** As the literature demonstrates, many methods of ontological enrichment have been introduced by researchers. The main objective of their works is to identify novel knowledge sources for a given field, including the knowledge relevant for education (the context of our study). Other studies as in [23] proposed methods for tackling correspondence between the initially created ontologies and the one identified by our machine in order to enrich our ontological knowledge representation system for educational systems. The field of education produces knowledge through e-learning systems that fall into different knowledge sources categories and found in different places. Researchers are working on two phases of such knowledge identification: the first concerns localization, while the second suggests the integration of intelligent agents for a better identification of delocalized knowledge [24]. Other researchers have offered web services to identify new distant knowledge in the field.

## 2.2. Knowledge extraction based on neural networks

**Artificial intelligence (AI) approaches:** such approaches have been introduced into many areas of scientific research. It should be noted that sciences subsumed within AI represent the trends that will dominate the scene in the coming years due to their effectiveness in the development of many sectors. More recently, researchers in the field have proposed methods for accentuated handling in the stages of carrying out the research results in various fields. In this context, researchers have proposed methods for accentuated investigation and processing in the phases of carrying out their work in various fields. Among these fields we may refer to the field of robotics. For example, the author in [25, 26] has made a comparative study of works relating to this vibrant field and its outcomes and implications for industry. Another example is that researchers have undertaken comparative studies to make agent-based models. Majority of the pinpointed ideas and concepts have an automatic aspect in that they tackle methods which have been proposed for dealing with massive data integration in a given area. Among these concepts, we may refer to artificial neural networks (ANN) and the relating approaches, which can be incorporated in various fields, especially the field of education, where they can be used in parallel processing of big data as an input, and classification of knowledge as an output. Among these works we cite [27].

**WORD2VEC:** It represents the embedding of lexical words families (i.e., word embedding), which allows vectorial representation based on knowledge sources. We can use word2vec in two different ways: Continuous bag of words model (CBOW) and Skip-gram (Gram jump). On the one hand, the realization of the words vectorial representation with CBOW is obtained by predicting their content from knowledge of the neighboring words. In this case, the notion of "words bag" does not impact the order of words. On the other hand, the second method, Skip-gram, is based on an approach of predicting words based on the context of their occurrence. In this case, the words with heavy weight are neighboring ones. The calculation of the word centers is made by artificial neural networks which simplify the data processing. Drawing on our analysis of the works which integrate Word2Vec approaches, we noted that Word2Vec has an implication for the cases of knowledge extraction from textual documents [13, 28]. In other works, it was used for knowledge extraction from structured documents. However, we equally noticed that such works have some shortcomings in terms of their knowledge sources nature and types.

### 3. THE METHODOLOGY

In order to achieve our main objective, which has already been laid out above, and which revolves around the representation of massive knowledge produced by learning actors, the present work adopts the methodology presented in Figure 1.

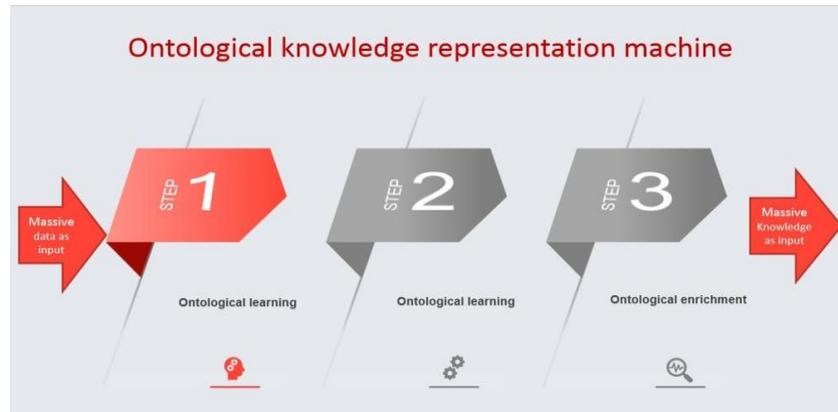


Figure 1. The methodology adopted for ontological knowledge representation in MOOCs

Based on this methodology, we adopt three algorithms: the first for ontological mapping, the second for ontological learning, while the third for ontological enrichment. Throughout the stages of the realization of the three stated algorithms, we use artificial intelligence approaches for the extraction/parallel processing of knowledge received from massive data as an input. The three elements are introduced as follow:

- The first algorithm: In this algorithm, the system receives massive data, which are preprocessed by generic interfaces and adapted to the educational field, from the operational massive data layer [2]. Worth noting is that the system involves two types of interfaces: database interface and the ontological one. In this stage, our algorithm makes the matching between database attributes and ontological attributes, which have already been proposed in [7], that is, the conceptual framework used for knowledge extraction from learning actors.
- The second algorithm: In this algorithm, the system incorporates ontological learning approaches during knowledge extraction from the database layer into the ontological layer. Hence, the system synchronizes the big data induced by the operational big data layer (SQL and NOSQL) and the ontological layer which is on the other interface. In this stage, the algorithm proposes an approach based on artificial neural networks techniques to better identify new sources of local or distant knowledge, on the national or international scale; then, it feeds the extracted knowledge into our representation system. In parallel, it verifies the quality of the acquired knowledge as well as its impact on the actors' profiles in the educational field.
- The third algorithm: In this algorithm, the system makes use of some methods for identifying knowledge in local MOOCs/or other Moroccan universities MOOCs. Then, it adapts them to our ontological knowledge representation system [2]. The algorithm is based on Web intelligent methods configured for the identification of new actors' knowledge sources; then, it adapts them to the general context of our knowledge extraction system.

The objective of the abovementioned algorithm is to create an intelligent knowledge representation system, while incorporating the best methods of massive knowledge identification and extraction depending on the results induced by the learning' actors' interactions in a MOOC system.

### 4. MASSIVE DATA PREPROCESSING

In this stage, we are proposed a MapReduce machine learning based on bigdata produced by MOOCs [2] to perform the parallel pre-processing of massive data falling within various structures. The pre-processing of knowledge sources is achieved by virtue of analyzing the results induced by the learning actors' interactions in MOOCs. This system [2] processes SQL and NoSQL massive data by an algorithm based on the HADOOP echo system associated approaches. It offers massive semi-structured data in XML and RDF format for the ontological knowledge representation system, the subject matter of the paper.

**5. PROPOSED MACHINE**

In this diagram, we propose a global architecture of our ontological knowledge representation system for learning actors based on big data resulting from their interactions in online learning. It should be noted that our system is open to all sorts of knowledge sources at the national level; it is enough to link the sources to the system. Figure 2 proposes the elements prerequisite to make the semantic representation of the actors' knowledge.

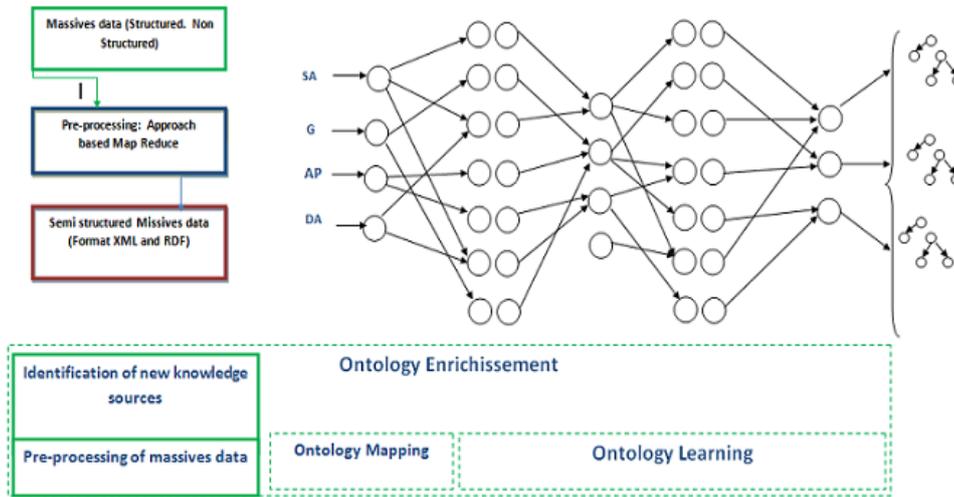


Figure 2. Actors' knowledge representation system based on MOOCs

**5.1. Ontological machine for knowledge representation in MOOC**

This system makes the ontological representation of the big data pre-processed in [2]. The representation begins with the ontological mapping algorithm (algorithm 1) between the input DataSet and/or ontologies knowledge sources and the output ontologies. Then, the ontological learning algorithm (algorithm 2) feeds knowledge sources extracted from online learning through into the output ontologies. This process is realized through transforming massive data into massive ontological knowledge, while making use of the artificial neural networks throughout the stages of ontological knowledge construction.

Finally, in an important later stage, the ontological enrichment algorithm (algorithm 3) detects new knowledge sources in MOOCs. This algorithm represents one stage among others proposed in our Framework [7]: In this stage, we focus on some definitions of our system variables:

Definition 1: For each given ontology as input of the system we have:

$$O_{ini} = \{ O_{in1}, O_{in2}, \dots, O_{inn} \}$$

$$\text{In that: } O_{ini} = \{ C_i, R_i, V_i \} \text{ And } (C_i = \{ C_1, C_2, \dots, C_n \}, R_i = \{ R_1, R_2, \dots, R_n \},$$

$$V_i = \{ V_1, V_2, \dots, V_n \} ) \text{ in wich } R_i = \{ is\_part, part\_of \};$$

Definition 2: For each given ontology:  $DS_{ini} = \{ DS_{in1}, DS_{in2}, \dots, DS_{inn} \}$  in that:

$$DS_{ini} = \{ A_i, T_i, V_i \} \text{ and } (A_i = \{ A_1, A_2, \dots, A_n \}, T_i = \{ T_1, T_2, \dots, T_n \},$$

$$V_i = \{ V_1, V_2, \dots, V_n \} ).$$

Definition 3: Calculation of the similarity [1] between concepts  $Sim(C_i, C_{i+1})$  in that:

– Extraction function based word

$$sim(w_1, w_2) = \frac{\vec{v}(w_1) \times \vec{v}(w_2)}{|\vec{v}(w_1)| \cdot |\vec{v}(w_2)|} \tag{1}$$

- Extraction function based concepts

$$Sim(c1, c2) = \frac{2 \times N3}{N1 + N2 + 2 \times N3} \quad (2)$$

Such as: N represents the level of the node;  $w_i$  represents the entered words;  $C_i$  represents the concept in the ontologies.

Definition 4: For each given ontology  $Oout_i = \{Oin\_pert, Oin\_nonpert\}$  in that:

$$\begin{aligned} Oout\_pert &= \{C_i, R_i, V_i\} \text{ and } (C_i = \{C1, C2, \dots, Cn\}, R_i = \{R1, R2, \dots, Rn\}, V_i = \\ &\{V1, V2, \dots, Vn\}) \\ \text{and } Oout_{nonpert} &= \{C_i, R_i, V_i\} \text{ and } (C_i = \{C1, C2, \dots, Cn\}, R_i = \{R1, R2, \dots, Rn\}, V_i = \\ &\{V1, V2, \dots, Vn\}), \text{ For each given ontology } \rightarrow Orefsuch \text{ Oref} = Oout\_pert. \end{aligned}$$

Definition 5: artificial neural networks are defined by learning function ontology:

$$f(x) = \sum_{i=1}^n w_i x_i + \beta \quad (3)$$

Such as:  $w_i$  and  $x_i$  the input of the ANN machine,  $\beta$  the activation threshold of node

Definition 6: The two mathematical models for extracting the words are CBOW and Skip-gram:

- CBOW

$$f(x) = \frac{1}{T} \sum_{t=1}^n \sum_{-c \leq j \leq c, j \neq 0} \log p(w_t / w_{t+j}) \quad (4)$$

- Skip-gram

$$f(x) = \frac{1}{T} \sum_{t=1}^n \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} / w_t) \quad (5)$$

Such as: Let are the central target word  $w_t$  and the context word  $w_j$  respectively indexed to  $t$  and  $j$  in the dictionary. The conditional probability of generating the word context for the given central target word can be obtained by performing the system on the internal product of the vector.

## 5.2. The contribution of this machine

### 5.2.1. Ontological mapping layer

In this stage, our system uses an approach which is more adequate for making correspondences between database attributes (SQL AND NOSQL) and ontological attributes (OWL, RDF). Our investigation of the relating literature reveals that researchers have proposed many approaches and methods dealing with massive data integration [29]. Indeed, the relevance of our approach is apparent at the level of incorporating the added values of big data-based linked and open data through web knowledge integration. Hence, this algorithm offers generic solutions for the mapping between the elements of RDBMS and those of RDF and OWL. This linkage is introduced by our conceptual modelling proposed in our ontology-based conceptual framework [7], by so doing, we determined a structured and/or semi structured dataset in online learning systems through:

- DS (Attribute, type, value)

And OWL ontological representation through the following attributes:

- ON (class, R, Value)

The objective in this stage lies in the transformation of dataset into ontology. Online learning systems, in this stage, produce big data in accordance with several structures that are pre-processed by our matching learning; this work performs the pre-processing of the massive data produced by actors [2]. This work takes big data from any structure as input; it follows that it applies many methods to analyze and pre-process them. Our aim here is to make an automatic mapping between database attributes and the ontological ones, respecting the standards of ontological mapping [18]. In a given stage, ontological mapping uses linked and open data for incorporating knowledge from various sources based on SPARQL language, which is a query language for web semantic data access.

The mapping algorithm: The mapping functions dealt with in various works are of different sorts. Drawing on such works, we determine mapping functions between knowledge sources (a Dataset and/or ontology input) and ontology output in our knowledge extraction system [2]. The following algorithm shows, in detail, all the steps to follow:

### Algorithm

#### Input:

$\Omega = DS \{A, T, V\}$  or  $O_{in}\{C_{in}, R_{in}, V_{in}\}$  // Knowledge sources and knowledge destination are not the same concepts (DataSet or owl to OWL case).

#### Output:

$O_{out}\{C_{out}, R_{out}, V_{out}\}$  // knowledge collections of learning actors according to actor categories.

#### Start

IF (DS  $\neq$  0) THEN

$\Omega \leftarrow DS$

**Step 1:** // Source and destination are not the same concepts (the case DataSet to ontology)

$$W_{i,j} = \frac{f_{i,j}}{\max_z f_{z,i}} \times \log\left(\frac{D}{d_i}\right) // \text{Extraction of key concepts } w_i, j$$

$$\vec{v} = \frac{1}{T} \sum_{t=1}^T \sum_{-c < j < c, j \neq 0} \log p(W_{t+j} | W_t) // \text{matching between sources attributes (dataset) and ontologies destination (knowledge)}$$

$$srel(w_1, w_2) = \frac{\vec{v}(w_1) \times \vec{v}(w_2)}{|\vec{v}(w_1)| \cdot |\vec{v}(w_2)|} // \text{Calculation of the semantic similarity of knowledge}$$

[relatedness, our ref].

IF (srel = 1) THEN

    w1 et w2 are synonymous

ELSE

    IF (srel = 0) THEN

        w1 and w2 are synonymous

    ELSE

        IF (srel = 0) THEN

            w1 and w2 are not similar otherwise // we will look for the type of relationship, we propose an evaluation rule (metonymy, antonym, hyperonym, hyponym, ...).

            SVM on calculating offset

            Ri = SVM (wi, wj) // call to our semantic relation extraction algorithm

$O_{out} \leftarrow O_{out} + \{W_i, R_i, V_i\}$

        END IF

    END IF

END IF

ELSE

$\Omega \leftarrow OWL$ .

**Step 2:** // OWL Alignment of source ontology and global ontology objects (C, P, V) the ontology

    O1 (C1, P1, V1) the ontology O2 (C2, P2, V2) the ontology

    AL = {en1, en2, coresp} en1 an entity of O1, en2 an entity of O2, coresp

    correspondence between w1 and en2 //Correspondence between the source

    attributes (dataset) and destination (knowledge)

$$Sim(c1, c2) = \frac{2 \times N3}{N1 + N2 + 2 \times N3} // \text{Calculate the similarity between knowledge}$$

[relatedness, note ref].

FOR each w1 in Entity1 then

    Sim\_temp = null;

FOR each w2 de Entity2

    Sim = {sim (name1, name2), sim (label1, label2), sim (content1, content2)}

    IF sim > sim\_temp

        sim\_temp = sim

        element = w2

    end if

END

    Al  $\leftarrow$  Al + {w2, element, sim\_temp}

end

$O_{out} \leftarrow$  Al

END IF

    return  $O_{out}$

**End**

### 5.2.2. Ontological learning layer

In this stage, our machine parameterizes the knowledge representation system [7], which is based on the resulting ontology of all knowledge sources (NOSQL or SQL) [2] emanating from the operational big data layer. Our objective here is to propose an optimal method for automatic feeding of our ontological knowledge base through detecting the new knowledge produced in our basic system and by measuring the similarity of the received knowledge. This objective is achieved resting on the already conducted studies on the best methods of ontological learning through the calculation of the similarity of the acquired knowledge. In order to have an environment capable of systematic and rapid learning, we have made in-deep analysis on the existing learning methods in the literature [30]. Among the methods, we have identified supervised and unsupervised learning methods [31]. In this algorithm, there is a range of processes to be applied on our system. Following an analysis of the works published in scientific journals relating to our research area, we observed that all of the taken approaches were based on one or all of the following steps:

- Identification of structured data received from the operational massive data layer.
- Definition of the transmission channel between the massive data and massive knowledge interfaces.
- Definition of the transmission means from pre-processed massive data to ontological layer
- Implementation of ontological learning algorithm.
- Integration of artificial neural networks-based approaches for parallel processing of different knowledge sources.

**Loading algorithm:** In this algorithm, we modify the data identified by our knowledge extraction system. In this stage, the goal is to make our learning system optimal and fast, and, thus, the most integrated at the level of our target knowledge representation system. Hence, this algorithm is defined as follows:

#### Algorithm

##### Input:

Mega\_data: the preprocessed missives data;  
Global\_Onto: the output ontologies;  
C: communicational channel.

##### Output:

Space of knowledge relevant for learning actors: SPAQL or NSPARCL

##### Start:

Adding the stream received from Meta\_bigdata in O\_collection while respecting the structure defined at the start(wi , xi )

$$f(x) = \sum_{i=1}^n w_i \times x_i + \beta \quad // \text{ Application of the formula for calculating different}$$

possible cases

IF (f(x) > β ) then The activation of neurons if they pass the threshold for the acquired knowledge.

Y = SVM (act (activities), note, TR (quarter), mm (memory), st (style), kg, peda (pedagogies), inter (interpretation), mode\_e (mode of education), us\_it (IT usage), inno (Innovation))// we define two classes pert and not pert in such a way that the actor with the note higher than 10 belongs to the relevant class c

IF(y <1) then

O<sub>out nonpert</sub> = O<sub>out nonpert</sub> + {Ci, Ri}

ELSE

O<sub>out pert</sub> = O<sub>out pert</sub> + {Ci, Ri}

END IF

END IF

End

### 5.2.3. Ontological enrichment layer

This step is proposed in order to give our knowledge representation system more scalability in case of acquisition of other knowledge sources (SQL and/or NOSQL) and, thus, enrich it by each received ontological knowledge. The system in question provides a rich environment for educational entities with regard to learning actor's knowledge representation. For the enrichment of such representation, as our studies show, Moroccan government, like other governments, has set the objective of creating online learning systems as a reference (i.e., MOOCs); our system makes use of actors' interactions in these systems as knowledge sources. The integration of MOOC systems is being generalized at the level of all Moroccan universities [22, 32]. By analyzing these systems, we find that they yield massive knowledge that can be integrated into our knowledge representation system. It follows that it is prerequisite that we think of the means of ontological enrichment that make discovery of the best new knowledge when connected with some MOOCs in universities.

This objective is achieved through the integration of artificial intelligence-based agents to the discovery of knowledge at the level of the Moroccan MOOC system. In the literature, we have identified numerous researches works that deal with ontological enrichment projects, the basic principle around which revolves the ontological representation of educational resources. In addition to these methods, this algorithm aims to identify new sources of knowledge, and then adapt such sources to our proposed system [7] which provides a framework for extracting knowledge from learning actors.

Enrichment Algorithm: Ontological enrichment: As presented in the foregoing, this algorithm, in principle, proposes a method of identifying distant knowledge sources, be it ontology ( $O_i$ ) or massive data (DSi). The algorithm provides in detail the steps to follow for a better ontological enrichment:

#### Algorithm

##### Input:

Ip, structure, Area, country,  
Massive data  $D = \{d_i\}$   
An ontology:  $O = \{o_i\}$  as input

##### Output:

Collection of knowledge based on domain ontology.

##### Start:

```
Adding new knowledge to an external knowledge source location
  IF Dsi ≠ 0
    Ω = Dsi
    • Call to big data preprocessing approaches;
    • Call to the ontological mapping algorithm;
    • Call to the Step 1 function of the mapping algorithm;
    • Call to the ontological loading algorithm.
  ELSE
    • Call to the ontological mapping algorithm;
    • Call to the Step 2 function of the mapping algorithm.
  END IF
```

End

## 6. EXPERIMENT

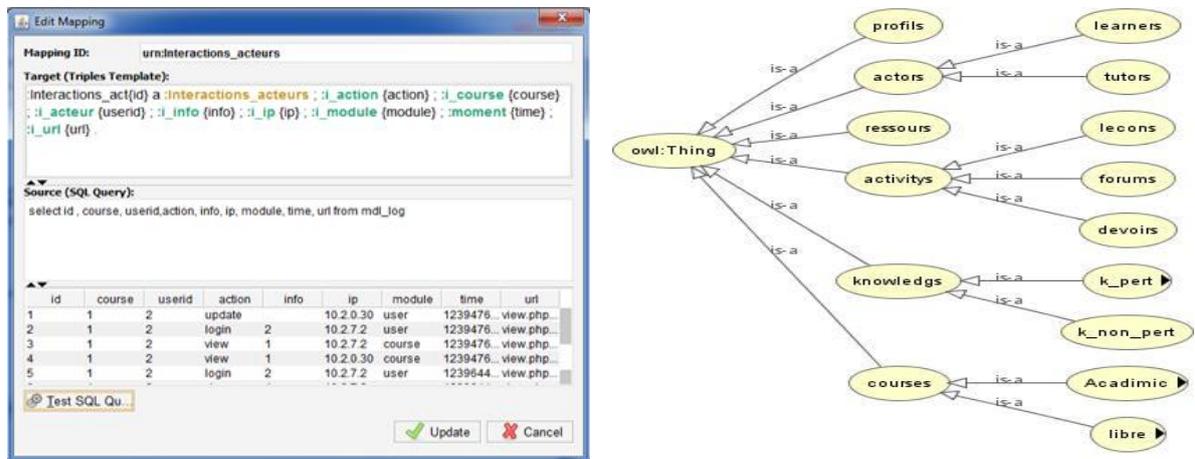
In this experiment, we emphasize the importance of the ontological mapping stage in the experimental evaluation of the functioning of our machine. The results of this experiment give us an idea about the reel experience of this machine. In this experimental phase, our system takes as input massive data of learning activity traces in XML format (massive data pre-processed by our learning machine [2], for example dataset (229 022 row)).

Massive data traces were extracted from the online learning system of the Mohamadia School of Engineering (EMI). To do the ontological mapping, we will use the ontological representation editor Protégé, which represents an ontological mapping tool that integrates plug-in for transforming knowledge sources (Dataset, XML) to ontology (OWL). This tool performs ontological mapping for the semi structured data (XML). In this mapping stage, the Protege ontological representation tool integrates plug-in for mapping, among these tools we have in the Ontop mapping literature which represents an ontological mapping tool. The correspondence is made either by coding or by graphic assistant. Figure 3 shows in detail the correspondence made in our case.

In this step, we have as output an ontological knowledge representation of learning actors. The Figure 3 shows the representation class, resulting from the application of the ontological mapping function provided by the Ontop plug-in. Following the analysis of the knowledge produced by our system, we observe that this latter sorts out knowledge, which is available for use as part of both formal and informal online learning sessions, depending on actors' profiles. Accordingly, it can create an environment for developing such profiles. We noted that knowledge falls into two categories:

- a. Relevant knowledge, which is part of the massive knowledge, depends on the flow of the massive data extracted at the level of the operational data layer. Indeed, the extracted data are of different sorts. We may refer for example to the following:
  - Knowledge related to effects: this type represents the knowledge reflecting the effects of the actors, such as copying the file, consulting the documents, and reading documents, etc.
  - Knowledge yielded due to productions during a learning session: it is the result of actors' interactions (answering questions, test evaluations, and exams, etc.)
  - Knowledge produced by virtue of access to files: it is the type of knowledge relating to the summaries produced from the files consulted during a learning session etc.

- b. Irrelevant knowledge, which is part of massive knowledge, is knowledge produced by actors who are ill-placed in this type of system.



ontop query editor:

Interactions_act	I_action	I_course	I_url	I_ip	I_module	I_info	I_userid
Interactions_act13	"view"	1	"view.php?id=1"	"10.2.7.2"	"course"	"1"	2
Interactions_act14	"view"	1	"view.php?id=1"	"10.2.7.2"	"course"	"1"	2
Interactions_act15	"login"	1	"view.php?id=0&course..."	"10.2.7.3"	"user"	"1"	1
Interactions_act16	"view"	1	"view.php?id=1"	"10.2.7.3"	"course"	"1"	1
Interactions_act17	"error"	0	"index.php"	"10.2.7.3"	"login"	"admin"	1
Interactions_act18	"error"	0	"index.php"	"10.2.7.3"	"login"	"admin"	1
Interactions_act19	"view"	1	"view.php?id=1"	"10.2.7.3"	"course"	"1"	1
Interactions_act20	"login"	1	"view.php?id=0&course..."	"10.2.0.30"	"user"	"2"	2
Interactions_act21	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act22	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act23	"update"	1	"view.php?id=-1&course..."	"10.2.0.30"	"user"	""	2
Interactions_act24	"assign"	1	"admin/roles/assign.ph..."	"10.2.0.30"	"role"	"Course creator"	2
Interactions_act25	"assign"	1	"admin/roles/assign.ph..."	"10.2.0.30"	"role"	"Course creator"	2
Interactions_act26	"assign"	1	"admin/roles/assign.ph..."	"10.2.0.30"	"role"	"Course creator"	2
Interactions_act27	"logout"	1	"view.php?id=2&course..."	"10.2.0.30"	"user"	"2"	2
Interactions_act28	"login"	1	"view.php?id=0&course..."	"10.2.0.30"	"user"	"3"	3
Interactions_act29	"change password"	1	"view.php?id=3&course..."	"10.2.0.30"	"user"	"3"	3
Interactions_act30	"view"	1	"view.php?id=3&course..."	"10.2.0.30"	"user"	"3"	3
Interactions_act31	"logout"	1	"view.php?id=3&course..."	"10.2.0.30"	"user"	"3"	3
Interactions_act32	"login"	1	"view.php?id=0&course..."	"10.2.0.30"	"user"	"2"	2
Interactions_act33	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act34	"logout"	1	"view.php?id=2&course..."	"10.2.0.30"	"user"	"2"	2
Interactions_act35	"login"	1	"view.php?id=0&course..."	"10.2.0.30"	"user"	"2"	2
Interactions_act36	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act37	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act38	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2
Interactions_act39	"view"	1	"view.php?id=1"	"10.2.0.30"	"course"	"1"	2

Figure 3. The show of the ontological machine proposed in this work

In a nutshell, it is clear-cut that this system offers a rich environment for relevant and irrelevant knowledge. Concerning the relevant knowledge, it is within the reach of learning actors for the follow-up of the courses in an online learning system, while irrelevant knowledge is available to study the actors' problems in online learning sessions. In the e-learning system we have two types of knowledge, such as relevant knowledge and irrelevant knowledge. To evaluate this machine, we have measured the probabilities for knowledge representation via two approaches (the first is based on our machine and the second on prediction). The Table 3 shows the result for some types of knowledge representation based on different sources circulated in the e-learning EMI.

By analyzing the predicted results and the machine proposed in this paper, we found that the knowledge produced by our machine is relevant compared to the predicted knowledge extracted from an e-learning system even if we have diverse knowledge sources. Indeed, our knowledge representation machine is more performant than the probability analysis based on knowledge representation. A second experiment on these knowledge sources was carried out; we kept that knowledge whose similarity between concepts varies between 5 and 10, having a great value of similarity between concepts. We applied our machine to these concepts and we observed that the number of acquaintances remains invariable. It should be noted that we retain only the pertinent knowledge that circulates in the platform. The Tables 3 and 4 give more details on this second experiment.

Table 3. Knowledge classification experiment 1

		DataSet SQL and NoSQL	Semi structured Data	Document Textual
Knowledge Pertinent	The Machine proposed in this Paper	150 022	140 000	2200
	Predicted	30 000	2000	400
Knowledge Non Pertinent	The Machine proposed in this Paper	44 000	60 000	500
	Predicted	5000	7600	300

Table 4. Knowledge classification experiment 2

		DataSet SQL and NoSQL	Semi structured Data	Textual Document
Knowledge Pertinent (10)	The Machine proposed in this Paper	164 022	175 000	2300
	Predicted	44 000	7000	600
Knowledge Non Pertinent (5)	The Machine proposed in this Paper	30 000	25 000	150
	Predicted	1 500	2 600	100

The results displayed in the Tables 3 and 4 were very encouraging, and the number of knowledge produced by our machine is very high compared to the probabilistic approach, exactly as in the first experiment. These results show the effectiveness of our approach to representing knowledge of learning actors. Having analyzed the obtained results, we can say that our system provides learning actors with a rich environment for developing their profiles. This system offers different kinds of knowledge because it can integrate delocalized knowledge. A case in point is the integration of knowledge produced by external educational entities. It is characterized by:

- Activation of the actor's motivation in an online learning session
- Giving autonomy to the actors
- Proposing efficiency measures to actors
- Providing an environment to recommend knowledge for actors.

To sum up, the system offers a knowledge-rich environment which can be used by learning actors in online learning, either for the development of their profiles, or for conducting studies on problems relating to the use of learning systems.

## 7. CONCLUSION AND PERSPECTIVE

In this work, we carried out research on the most recent works dealing with projects of knowledge extraction and representation for all domains, placing heavy importance on the projects of knowledge representation in the domain of education. We tried to handle the issue of actors' knowledge extraction, which is based on their interactions in an on line learning session. For such purpose, we identified the most appropriate knowledge representation methods in this area. For achieving our main objective; we made such an extraction and representation from various sources, based on three basic algorithms: First, mapping algorithm which performs an ontology mapping between a knowledge source (Dataset and/or ontology) and an output ontology, which represents the ontological representation of the actors' knowledge; this is done through several transformations covering the entire scope of our objective linked to this stage; second, learning algorithm deals with the ontological feeding stage of our output ontology. This process is carried out via the determination of the transmission channel of the massive pre-processed data towards massive ontological knowledge, that is, the output of our knowledge representation system based on e-learning systems; Third, enrichment algorithm is intelligent in that it identifies the new knowledge around which our global system of knowledge representation revolves; then it adapts it to the global context of our actors' knowledge representation system.

The overall results of this ontology-based system of shows the existence of two sorts of knowledge (relevant and irrelevant), which can be used by learning actors in an online learning system like MOOC. We believe that this work will lead to a better implementation of our representation system in a real production environment. Thus, it leads to a better assessment of such actors' effectiveness, the system's impact on their profiles, and finally the integration of a generic approach to actors' knowledge recommendation in an online learning session.

## REFERENCES

- [1] Carolina Costa, L. Teixeira, and H. Alvelos, "Exploring the Usage of MOOCs in Higher Education Institutions: Characterization of the Most Used Platforms," *International journal of information and communication technology education IJICTE*, vol. 14, no. 4, pp. 1-17, 2018.
- [2] A. Hadioui, Y. B. Touimi, and S. Bennani, "Machine Learning Based on Big Data Extraction of Massive Educational Knowledge," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 11, pp. 151-167, 2017.
- [3] Y. Gutiérrez, D. Tomás, and I. Moreno, "Developing an ontology schema for enriching and linking digital media assets," *Future Generation Computer Systems*, vol. 101, pp. 381-397, 2019.
- [4] P. Hallinger and W. Hammad, "Knowledge production on educational leadership and management in Arab societies: A systematic review of research," *Educational Management Administration & Leadership*, vol. 47, no. 1, pp. 20-36, 2019.
- [5] H. Li, J.-X. You, H.-C. Liu, and G. Tian, "Acquiring and sharing tacit knowledge based on interval 2-tuple linguistic assessments and extended fuzzy Petri nets," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 26, no. 01, pp. 43-65, 2018.
- [6] M. L. Zeng, "Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article," *El profesional de la información*, vol. 28, no. 1, pp. 1-35, 2019.
- [7] A. Hadioui, et al., "An Ontological Extraction Framework of the Actors' pedagogical Knowledge," *Journal of Theoretical and Applied Information Technology*, vol. 93, no. 1, pp. 69-80, 2016.
- [8] S. W. Tu, H. Eriksson, J. H. Gennari, Y. Shahar, and M. A. Musen, "Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: application of PROTÉGÉ-II to protocol-based decision support," *Artificial Intelligence in Medicine*, vol. 7, no. 3, pp. 257-289, 1995.
- [9] H. Li, T. Yuan, H. Wu, Y. Xue, and X. Hu, "Granular computing-based multi-level interactive attention networks for targeted sentiment analysis," *Granular Computing*, vol. 5, pp. 387-395, 2020.
- [10] G. Zhu and C. A. Iglesias, "Computing semantic similarity of concepts in knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 72-85, 2016.
- [11] W. You, Z. Yang, G. Guo, X.-F. Wan, and G. Ji, "Prediction of DNA-binding proteins by interaction fusion feature representation and selective ensemble," *Knowledge-Based Systems*, vol. 163, pp. 598-610, 2019.
- [12] G. Giuliani, et al., "Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on Land degradation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, pp. 1-12, 2020.
- [13] Z. Gao, et al., "edge2vec: Representation learning using edge semantics for biomedical knowledge discovery," *BMC Bioinformatics*, vol. 20, no. 1, p. 306, 2019.
- [14] L. Yang, K. Cormican, and M. Yu, "Ontology-based systems engineering: A state-of-the-art review," *Computers in Industry*, vol. 111, pp. 148-171, 2019.
- [15] Y. Blanco-Fernández, A. Gil-Solla, J. J. Pazos-Arias, M. Ramos-Cabrer, A. Daif, and M. López-Nores, "Distracting users as per their knowledge: Combining linked open data and word embeddings to enhance history learning," *Expert Systems with Applications*, vol. 143, p. 113051, 2020.
- [16] H. Xia and Y. Yang, "Fine-grained Aspect Extraction for Online Reviews of E-commerce Products Based on Semi-supervised Learning," *WHICEB 2018 Proceedings*, vol. 36, 2018.
- [17] I. Harrow et al., "Ontology mapping for semantically enabled applications," *Drug Discovery Today*, vol. 24, no. 10, pp. 2068-2075, 2019.
- [18] A. Konys, "Knowledge systematization for ontology learning methods," *Procedia Computer Science*, vol. 126, pp. 2194-2207, 2018.
- [19] G. Gómez-Jiménez et al., "Chapter Four-The OECD Principles for (Q)SAR Models in the Context of Knowledge Discovery in Databases (KDD)," in *Advances in Protein Chemistry and Structural Biology*, vol. 113, pp. 85-117, 2018.
- [20] G. Petrucci, M. Rospocher, and C. Ghidini, "Expressive ontology learning as neural machine translation," *Journal of Web Semantics*, vol. 52-53, pp. 66-82, 2018.
- [21] M. Rani, A. K. Dhar, and O. P. Vyas, "Semi-automatic terminology ontology learning based on topic modeling," *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 108-125, 2017.
- [22] M. Ali, S. Fathalla, S. Ibrahim, M. Kholief, and Y. Hassan, "Cross-Lingual Ontology Enrichment Based on Multi-Agent Architecture," *Procedia Computer Science*, vol. 137, pp. 127-138, 2018.
- [23] Y. Deng, "Construction of higher education knowledge map in university libraries based on MOOC," *The Electronic Library*, vol. 37, no. 5, pp. 811-829, 2019.
- [24] A. A. Kornienko, A. V. Kornienko, O. B. Fofanov, and M. P. Chubik, "Knowledge in Artificial Intelligence Systems: Searching the Strategies for Application," *Procedia-Social and Behavioral Sciences*, vol. 166, pp. 589-594, 2015.
- [25] B. Pandey and R. B. Mishra, "Knowledge and intelligent computing system in medicine," *Computers in Biology and Medicine*, vol. 39, no. 3, pp. 215-230, 2009.
- [26] S. Leo Kumar, "Knowledge-based expert system in manufacturing planning: state-of-the-art review," *International Journal of Production Research*, vol. 57, no. 15-16, pp. 4766-4790, 2019.
- [27] P. Pham and P. Do, "W-MetaPath2Vec: The topic-driven meta-path-based model for large-scaled content-based heterogeneous information network representation learning," *Expert Systems with Applications*, vol. 123, pp. 328-344, 2019.
- [28] P. Ristoski and H. Paulheim, "Semantic Web in data mining and knowledge discovery: A comprehensive survey," *Journal of Web Semantics*, vol. 36, pp. 1-22, 2016.

- [29] J. Liu, Y. Li, X. Tian, A. K. Sangaiah, and J. Wang, "Towards Semantic Sensor Data: An Ontology Approach," *Sensors*, vol. 19, no. 5, p. 1193, 2019.
- [30] W. Gao, J. L.G. Guirao, B. Basavanagoud, and J. Wu, "Partial multi-dividing ontology learning algorithm," *Information Sciences*, vol. 467, pp. 35-58, 2018.
- [31] C. Miller, Z. Nagy, and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1365-1377, 2018.
- [32] I. Pembeci, "Using word embeddings for ontology enrichment," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. 3, pp. 49-56, 2016.

## BIOGRAPHIES OF AUTHORS



**Abdelladim Hadioui** RIME TEAM-Networking, Modeling and e-Learning- LRIE Laboratory-Research in Computer Science and Education Laboratory Mohammadia School Engineers (EMI)-Mohammed V University in Rabat Agdal AV. Ibn Sina Agdal Rabat BP. 765 Morocco.  
Email: ahadioui@gmail.com



**Yassine Benjelloun Touimi** RIME TEAM-Networking, Modeling and e-Learning- LRIE Laboratory-Research in Computer Science and Education Laboratory Mohammadia School Engineers (EMI)-Mohammed V University in Rabat Agdal AV. Ibn Sina Agdal Rabat BP. 765 Morocco  
Email: yassine.benjelloun.touimi@gmail.com



**Nour-eddine El Faddouli** RIME TEAM-Networking, Modeling and e-Learning- LRIE Laboratory-Research in Computer Science and Education Laboratory Mohammadia School Engineers (EMI)- Mohammed V University in Rabat Agdal AV. Ibn Sina Agdal Rabat BP. 765 Morocco  
Email: faddouli@emi.ac.ma



**Samir Bennani** Deputy Director of students and academic affairs. RIME TEAM-Networking, Modeling and e-Learning-LRIE Laboratory-Research in Computer Science and Education Laboratory Mohammadia School Engineers (EMI)-Mohammed V University in Rabat Agdal AV. Ibn Sina Agdal Rabat BP. 765 Morocco  
Email: sbennani@emi.ac.ma