IJICTR International Journal of Information & Communication Technology Research

Volume 9 - Number 1- Winter 2017(53-61)

Pruning Concept Map to Geneate Ontology

Maedeh Mosharraf
Faculty of Engineering, Electrical and Computer Engineering Department
University of Tehran
Tehran, Iran
m.mosharraf@ut.ac.ir

Fattaneh Taghiyareh
Faculty of Engineering, Electrical and Computer Engineering Department
University of Tehran
Tehran, Iran
ftaghiyar@ut.ac.ir

Received: December 22, 2016- Accepted: February 24, 2017

Abstract—**Knowledge representation in the form of a concept map can be a good idea to categorize domain terms and their relations and help to generate ontology. Supplementing detail information to and pruning useless data from the concept map, which likes a skeleton in evolving ontology, can be semantically accomplished using the domain knowledge. In this paper, we propose a method using structural knowledge resources as well as tacit knowledge of experts to generate the ontology of eLearning domain. The concept map of eLearning is manually improved and finally verified using the group of eLearning experts. In order to enrich the ontology with merging into upcoming terms, the paper proposed an automatic method based on two external knowledge sources, Wikipedia and WordNet. The semantic similarity of concepts which is measured using the words hierarchy of WordNet combined with relations of concepts extracted from the Wikipedia graph is applied to link the new eLearning concepts to the domain ontology. The generated ontology is a dynamic knowledge source which can improve itself gradually. This integrated knowledge of eLearning domain can be used to model educational activities and to build, organize, and update specific learning resources.**

Keywords—concept map; pruning; ontology generation; ontology enrichment; elearning; graph clustering; Wikipedia; WordNet

I. INTRODUCTION

The aim of the semantic web is to enable machines to interpret and process information so that support people in doing different works on the web, especially search [1]. Several technologies that provide formal descriptions of terms, concepts, and relationships within a given knowledge base assist semantic web to its goal. Ontology is considered as one of the pillars of the semantic web technologies [2]. Although there is not a universal consensus on the precise definition of ontology, it is generally accepted that ontology is a formal specification of conceptualization [3].

Generating a worldwide ontology, which includes identifying, defining, and entering concept definitions and their relationships, is a challenging issue in the semantic web and is still far from being fully implemented. This process is so cost and time-consuming. In addition, manual process of ontology construction is limited to a special domain which requires deep understanding of that. Even in a specified domain, different opinions about concepts and their relations leads to different forms of ontologies, that none of them are sufficient certainty [4]. (Semi-)Automatic generation of ontology can overcome some of these problems.

The importance of a domain ontology is widely recognized, particularly in relation to the expected advent of the semantic web applications. The goal of a domain ontology is providing the background knowledge for any agent and function of a system and reducing the conceptual and terminological confusion among the related modules. This is achieved by the explicit representation of as more domain concepts as possible and their relationships.

ELearning as a solution of information technology to promote educational activities provides many applications, services, resources, and systems which can benefit a domain ontology to

promote their usages. The eLearning specific ontology fosters:

- Automation of many processes in eLearning applications
- Modeling and managing different modules of eLearning systems
- Communication and cooperation among different parts of a system
- Interaction between independent systems
- Sharing and reusing educational services and resources especially open educational resources
- Profiling eLearning users as well as resources
- Development of a common language for web service interactions

This paper proposes a three-phase method for semi-automatic construction of eLearning ontology and enriching it using external knowledge bases. In the first phase of ontology generation, a hybrid method of text processing and natural language processing techniques is combined with statistical analysis to extract knowledge semantically. By applying some eLearning specific rules, the process of ontology generation focuses on this domain. This simple ontology, which is actually a concept map, is generated according to a large set of papers from a famous eLearning conference as the background knowledge. In the second phase, the generated concept map is pruned and improved to the ontology. Applying tacit knowledge of domain experts, type of each node and its relations are determined to the concept map and missing relations are added. Considering comments of all the experts, the third phase of our methodology is accomplished to enrich the generated ontology with new upcoming terms in the domain and convert the ontology to a dynamic knowledge source. Wikipedia and WordNet are used to define the meaning, appliance, and relations of new terms with the other existing concepts of the created ontology.

The rest of the paper is as follows: In section 2 a review of the related works on ontology generation methods is presented. In section 3, we propose our approach to generate the eLearning specific ontology semi-automatically. Section 4 represents the experimental results, and finally in section 5 the work is concluded.

II. RELATED WORKS

Raising interests to research about semantic web, lots of methods are proposed to generate ontology. Although a manually generated ontology is much more precise and reliable, constructing ontology (semi-) automatically is the central point of recent studies. However, it could be deficient since it relies only on pure data and not on human judgments. Typically ontology can be extracted from various data types such as textual data [5], knowledge-base [6], relational schema [7], and social networks [8]. Generating or learning ontology is the process of identifying terms, concepts, taxonomic relations, non-taxonomic relations, and optionally axioms; and applying them to construct knowledge sources [9, 5].

Reviewing (semi-) automatic ontology generation techniques, [10] groups them into four main categories: 1. Conversion or translation, which transforms the representation of an existing ontology to common knowledge representations. Conversion of XML to OWL or other ontology formats is an example. For instance, [11] develops an OWL-based language that can transform XML documents to arbitrary OWL ontologies and overcomes to shortcomings of not OWL-centric methods. 2. Mining-based methods implement some mining techniques to retrieve information and produce ontology. These techniques are usually focused on processing unstructured resources like text documents or web pages through sets of linguistic, statistical, and machine learning methods [7, 3]. Linguistic-based techniques which are mainly dependent on natural language processing tools include part-of-speech tagging, sentence parsing, syntactic structure analysis, and dependency analysis [12]. Statistic-based techniques consist of information retrieval and probabilistic patterns which provide various algorithms for analyzing associations between concepts [5]. The main idea behind these techniques is that the co-occurrence of lexical units in text often provides a reliable estimate about their semantic identity. Data mining methods can also be included in machine learning based techniques which extract rules and patterns out of massive datasets in a supervised or unsupervised manner [13]. An example of the mining-based method is [14], which benefits from the combination of C-value method, artificial neural networks, Bayesian network, and fuzzy theory to construct an ontology.3. External knowledge-bases, which build or enrich an ontology using external resources like existing ontologies, search engines [15], general knowledge resources such as WordNet [16] and Wikipedia [17]. 4. Frameworks, which provide a platform with different modules to assist ontology generation. Protégé as one of the most popular frameworks is an open source platform developed by the Stanford Medical Informatics group at the University of Stanford [18].

The other view on ontology generation methods groups them in two categories as supervised and unsupervised. Supervised methods need some training data which is labeled based on predetermined features. For example, [19] implements a tool named TextRunner which operates in three phases: in the self-supervised learning phase, a classifier is generated which labels selected words. In the single-pass extraction phase, all the relation tuples are extracted from the dataset. A probability is assigned to each tuple which is evaluated in the third phase as redundancy-based assessment. However, in unsupervised methods, hidden knowledge is extracted from unlabeled data. In the unsupervised method proposed in [20], a fuzzy version of a decision tree is used. In this research done for planning an emergency center, language predictions, categories, and describing days by activities and information about the center,

the daily working cycles for each category are identified.

III. METHODOLOGY

Domain specific ontology generation needs the strong background knowledge about that domain. However, there is not a rich knowledge source to be used in automatically generating or updating eLearning ontology. Especially in the case of new terms and concepts, related background is not rich enough to show appropriate relations. So, we suggest a three-phase ontology generation method. In the first phase, using lots of domain related documents, we extract a primary concept map consisted of frequent domain terms and relations. In the second phase, all the terms and relations of the concept map are reviewed to determine the classes, instances, and type of their relations in the ontology. Clustering the generated ontology, new terms can be gradually increased to the ontology in the third phase. Fig. 1 illustrates the detail of each phase as well as the input and output of it.

A. *First phase: concept map generation*

Receiving experts' opinions person to person and without intermediaries in order to find domain concepts and their relations can be so cost and time consuming. Collection of documents generated by domain experts can be an alternative for using experts' opinions and automatically generating ontology. If this ontology is supposed to be extracted from several texts, they should be numerous enough to be sure about its comprehensiveness. Our focus is on the domain of eLearning. So, we take the proceedings of ICALT (International Conference on Advanced Learning Technologies) at six years as our input corpus.

As illustrated in Fig. 1, extracting a simple ontology is accomplished in the two steps. In the pre-process step, a collection of keywords is extracted by accomplishing candidate words extraction, compound words solidification, words unification, and words standardization procedures. The set of keywords is connected in the form of a graph in the second step. In this respect, low-score words which are considered as outliers should be removed. Afterward, each pair of words which has statically potential to be linked is connected to each other by the process of edge weight calculation. Finally, applying some rules fitting the domain of eLearning, the generated graph is refined. [21] explains these steps gradually and in full detail.

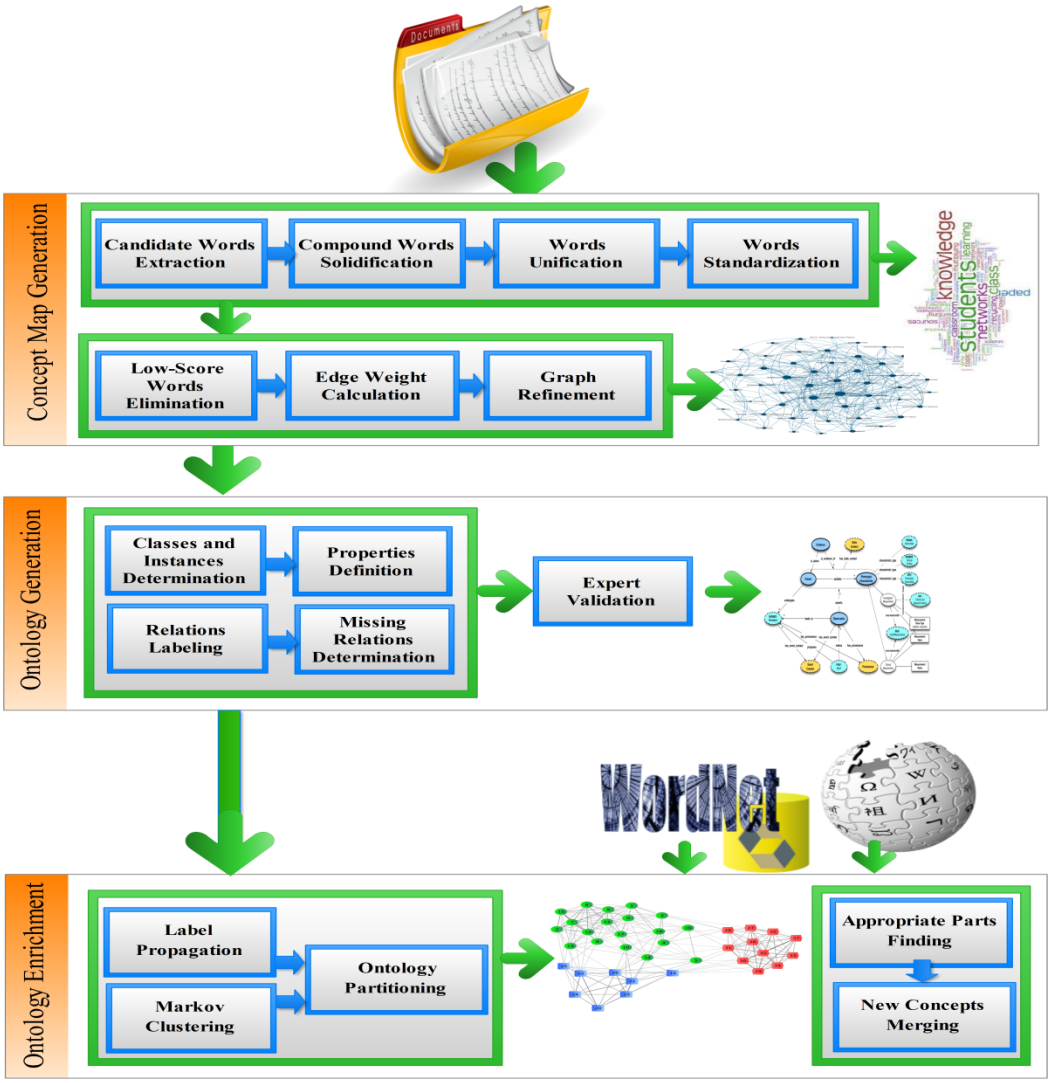

Fig. 1. Three-phase ontology generation method

However, extracting ontology from the domain related corpus leads to a simple ontology which is similar to a concept map. This concept map does not provide any information about role of the concepts and type of their relations. Therefore, the transition phase for improving it to a complete ontology is needed.

B. Second phase: concept map convertion to ontology

We described that the generated graph is a simple ontology and in the other word a concept map. This concept map shows the extracted concepts of eLearning domain and their relations. Surely, types of concepts as well as kinds of edges are not specified. We can say this concept map is a skeleton for implementing the ontology. This skeleton has some weaknesses in representing the domain knowledge.

- The initial corpus that the concept map has been extracted from is a set of research papers. In these articles, with a high probability new research findings are discussed and elementary or fundamental topics are rarely explained. Therefore, there may be some fundamental concepts not covered in this collection or removed as outliers in the first phase.
- The concept map is extracted from a set of documents, so it represents the knowledge which is embodied in them. According to [22], the type of knowledge that can be codified and represents in a text document is the explicit knowledge. In this respect, we should find a method that can complete this knowledge and enrich our ontology to the other type of knowledge which is named tacit knowledge.

We should complete the ontology using the skeleton of concept map. In the other word, we should prune some useless data from the concept map and grow some details and necessary information. The steps are taken to this end are as follows.

1) Classes and instances determination

All the terms which are included in the concept map can have different roles in the ontology such as class, instance, and even property. In order to convert the concept map to the ontology, its node should be examined from this perspective and their role should be determined. Following rules make it easier.

A term is considered as a class if:

- It has a role in eLearning systems.
- It represents a resource or tool which is available for learning.
- It introduces a learning activity.
- It plays an important role in learning processes or environments.

A term is considered as an instance if:

- It is applied as an example for a class.

A term is considered as a property if:

- It introduces a feature of a class such as an element that is used for profiling or modeling.

Linguistic rules can help to find instances and properties in a text document. Phrases “sample of”, “is kind of”, “instance of”, and “such as” are some indications of instances in a text. However, there may be an instance applied in a text document without using these phrases. Patterns of applying a property in a text are usually in the form of “class property” or “property of class”. However, there are also many violations. Benefiting experts’ knowledge, these cases can be determined. In this respect, we focus on the second phase of Nonaka and Takeuchi’s organizational knowledge creation framework - called creating concepts phase ([22]).

Terms whose roles are determined, using the specified tags are introduced to the ontology. Some examples are as follows.

```
<Declaration>
</”Class IRI=”#student>
<Declaration/>

<Declaration>
</”ObjectProperty IRI=”#age>
<Declaration/>

<Declaration>
</”NameIndividual IRI=”#MOODLE>
<Declaration/>
```

In addition, synonym terms were unified in the concept map generation phase and replaced with a super node. Now, all of the synonyms should be added to the ontology and their relation should be determined. For example:

```
<EquivalentClasses>
    <Class IRI="#student"/>
    <Class IRI="#learner"/>
</EquivalentClasses>
```

2) Properties definition

As we say, many attributes of the classes are included in the concept map and are determined in the previous sub-section. These attributes are the ones that significant number of researches being accomplished on them. User characteristics are some of these attributes used in user modeling and personalization processes. Nonetheless, many features of the ontology classes are rarely considered in researches and not included in the concept map. These features may be required in various applications and future researches, so should be defined in the ontology. Using some standards improved for the domain of learning and education, such as IEEE LOM, which is improved for modeling learning objects, and SCORM, which is improved for sharing objects, can benefit in this activity. Finally, using the knowledge of experts for completing features is the additional solution.

Assigning each property to the related class is the other activity which is done through specified format and property tags of the OWL.

```
<ObjectPropertyDomain>
    <ObjectProperty IRI="#age"/>
    <Class IRI="#student"/>
</ObjectPropertyDomain>
```

3) Relations labeling

According to [21], each edge of the concept map satisfies at least one of these rules.

- An edge represents the inclusion or inheritance relation of two concepts and thus forms a concept hierarchy.
- From two concepts which are linked using an edge, one of them is a tool for doing or promoting another.
- One of the concepts involved in an edge is an action in learning or eLearning process. Verbs such as “assess”, “assign”, “learn”, “teach”, “game”, “study”, and “collaborate” are examples of these concepts.

However, edge types in the concept map are not specified. This is done manually and by judging domain experts. Reviewing each edge of the concept map its type, which is among “sub-class”, “is done by”, “help to”, “do”, and so on, should be determined. Nevertheless, many relations in the concept map have the type “is related to”. This type can be a super type for all the other types. For example, a relation with the type “sub-class” can also be in the type “is related to”. So, we need to determine this kind of relations more accurately. If relations with the type “is related to” do not have specifically determined, it is preferred that they are pruned from the ontology. Therefore, determining the type of each relation and importing its data in the ontology, structure of the ontology can be completed.

Introducing the type of each edge to the ontology is done by calling its nodes in the format specified in OWL.

```
<SubClassOf>
    <Class IRI="#student"/>
    <Class IRI="#role"/>
</SubClassOf>
```

4) Missing relations determination

Although the ontology obtained from the previous sub-section is an acceptable ontology which contains all the concept map information and can be processed by machine, it is not necessarily complete. In the other word, this ontology should be completed using more details. Importing tacit knowledge of domain experts, the ontology concepts and their relations have been reviewed again and incomplete information is corrected and completed. Completing relations between concepts that are sometimes associated with adding new nodes to the ontology is an important task. The results of applying this step on the ontology of eLearning show that nearly 60 percent of added edges have the type “sub-class” and are completing the taxonomy of concepts.

5) Expert validation

Although a positive impact of the ontology on some applications reflects its authenticity [21], [23], we use the judgments of some experts to verify its correctness and comprehensiveness manually. In this respect, the generated ontology is sent to a group of domain experts and asked them to express their opinions about the following questions:

- Do the ontology terms cover all concepts related to eLearning?
- Does the ontology contain all relations between concepts?
- Have the type of relations been established correctly?

In this respect, we invite from seven experts of the domain to help us in this research. About all the questions, we asked the experts to express some samples which violate giving positive responses. The comments of all the experts confirm the implemented method for generating the concept map [21] and converting it to the ontology. However, considering the elimination step of node and edge outliers, some of them don’t know the generated ontology as the comprehensive knowledge base. Eliminated outliers aren’t justified statically; however the experts believe that they can be semantically corrected. The previous sub-section, which adds missing relations to the ontology, can compensate the missing information about ontology edges. But in the case of nodes, the ontology can be relatively weak. Therefore, we should find an appropriate solution for completing the ontology by outlier nodes and also upcoming new terms.

C. Third phase: ontology enrichment

Considering judgment of the eLearning experts, eliminating domain related outlier nodes from the concept map can blemish to the ontology. These nodes are removed due to their low frequency presence in the corpus documents. Therefore, we can say the background documents are not strong enough to statically support the ontology for adding low frequency terms. The dilemma of lacking adequate background knowledge increased when we want to merge the ontology with some terms which have been added to the domain concepts recently. However, each term has the specific semantic features which can be extracted from updated external knowledge bases.

The proposed approach uses Wikipedia and WordNet to specify application domain and semantic features of the input terms, which are removed as outliers in the concept map generation phase or recently added to the domain. Wikipedia as a knowledge base developed by collective intelligence distinguishes words with multiple meanings. Existence of a page related to each input term and following its input and output links can lead to determination of its domain. After defining the application domain, other characteristics of the

input term such as it synonyms, antonyms, parents, and grandparents in the hierarchy of words can be extracted from WordNet.

1) Ontology partitioning

Graph partitioning can facilitate the process of analyzing the structural and functional properties of the generated ontology, which is now a large and complex graph. Graph partitioning should be done semantically, meaning that the nodes placed in a group should be semantically related. Accordingly, the appropriate place for inserting new nodes to the ontology can be found locally. In this respect, graph partitioning methods can be applied to cluster the ontology. Therefore, sets of nodes should be determined so that the relation weights of the connections inside the sets are semantically higher than the relation weights of any connections to nodes outside the sets. This definition means modularity maximization [24]. After studding four algorithms, we applied a combination of label propagation [25] and Markov clustering [26] algorithms. Table 1 demonstrates the advantage and disadvantage of the investigated algorithms.

In label propagation, which is run iteratively, each node of the network is given a unique label initially. At each iteration, each node updates its label by choosing the label that most of its neighbors have. If multiple maximal labels exist among neighbors, the new label is chosen at random. The propagation iterations are performed until each node has a label that is the most frequent label among its neighbors.

Markov clustering partitions a graph via simulation of random walks. The idea is that random walks on a graph are likely to get stuck within dense sub-graphs rather than shuttle between dense sub-graphs via sparse connections. Utilizing this algorithm, the nodes in the graph are divided into non-overlapping clusters. Thus, nodes between dense regions will appear in a single cluster only, although they are attracted by different groups.

The fusion of the results obtained from label propagation and Markov clustering is performed as follows:

- If there is an overlap between the results of label propagation and Markov clustering, the common cluster would be the final cluster.
- If the result of clustering with one algorithm is a combination of other clusters from the other algorithm, then the largest cluster would be the final cluster. The smaller clusters might still exist in a hierarchy.
- If there's no overlap between two clusters obtained from two algorithms, then the cluster with maximum modularity will be the final cluster. Modularity is defined by equation 1 [24].

$$Q = 1/2m \sum_{vw} A_{vw}\delta(Cv, Cw) \qquad (1)$$

In this formula m is the indicative of the number of edges. Let the adjacency matrix for the network to be represented by A. $A_{vw} = 0$ means there's no edge between nodes v and w and $A_{vw} = 1$ means there is an edge between the two nodes. If we suppose the vertices are divided into clusters such that vertex v belongs to group c, (Cv, Cw) is defined to be 1 if two nodes v and w belong to the same group and zero otherwise. Q will be large for good divisions of the network, in the sense of having many within-cluster edges.

2) Appropriate part/s finding

It is likely that new concepts, adding to the existing ontology are related to each other. Therefore, we use an idea called Memory Cell. Memory Cells remember the situation of several last concepts which are added to the ontology. These cells cause in facing new concepts, the clusters of previous concepts are specially checked. Using Memory Cells is not possible for the first input concept. In addition, it is conceivable that input concepts are not related to each other. In order to increase the precision and avoid searching all the ontology for adding new concepts, we use a supplementary approach.

In the supplementary method, we calculate the semantic similarity of the input concepts with the delegate of each cluster in the ontology. The delegate node in each cluster can be the hub or a

TABLE 1. FEATURES OF THE CLUSTERING ALGORITHMS

Method	Advantage	Disadvantage
Label Propagation	Short runtime No need to information about the graph structure Propagating label of each node to its neighbors makes this method appropriate for clustering semantic networks	Failure to produce a unique answer
K-Means	Non-overlapping clusters	Need to determine the number of clusters as the algorithm input Considering the Euclidean Distance as similarity measure Unsuitable for non-spherical clusters
Markov	High speed and scalability Resistant to noise Non-overlapping clusters Considering the graph flow rather than the graph structure of the method makes it appropriate for partitioning any graph	Unsuitable for clusters with large diameter
Girvan-Newman	Focusing on edges that are most likely "between" communities	Long runtime Inappropriate for large graphs

node with the minimum total distance from the other cluster nodes. According to the size of the ontology and the number of its clusters, one/some of the clusters which has/have the closest semantic similarity to the input concepts is/are selected to search exactly. The semantic similarity measurement is done using equation 2.

Therefore, several clusters are suggested for each of the input concepts according to:

- Memory Cells
- Semantic similarity measurement.

3) New concepts merging

Adding the new concept to the ontology and linking it to the existing nodes are done based on their combinational tendency. Determining the threshold for combinational tendency is dependent to the domain of the ontology and can be accomplished based on experiments. For each of the selected clusters, the combinational tendency of the input concept and all the cluster nodes are calculated. Semantic features extracted from Wikipedia and WordNet are used to determine their combinational tendency.

At first, the synonyms of the input concept and all the concepts of selected clusters are extracted from WordNet. In the next step, corresponding pages of them on the site of Wikipedia are fetched. Existence of a direct link between concepts or a path with the length of two edges can connect each pair of concepts.

In our proposed approach, the dictionary of WordNet is applied when Wikipedia fails to link concepts. Failure of Wikipedia occurs in two circumstances:

- Lack of a dedicated page for each concept and its synonyms
- Lack of a direct link or a path with the length of two between each pair of concepts (or their synonyms)

Measuring the semantic distance of each pair of concepts using WordNet determines the possibility of their connection. If this distance is lower than the predefined threshold, two mentioned concepts are linked with the edge weighted by the inverse number of the semantic distance. Various methods of measuring similarity according to WordNet are introduced. [27] illustrates that Jiang and Conrath's measure is one of the best method as the only available information is the domain of concepts. The Jiang and Conrath's measure is given by equation 2.

$$dist_{JC}(c_1, c_2) = 2\log\left(p\left(lso\ (c_1, c_2)\right)\right) - \left(\log\left(p(c_1)\right) + \log(p(c_2))\right) \quad (2)$$

Where $\mathrm{lso}(c_1, c_2)$ is the information content of the closest common concept of c_1 and c_2. In the above formula p(c) is the probability of encountering an instance of a synset c in some specific corpus.

IV. RESULTS

As we mentioned in section 3, effectiveness of the generated concept map is evaluated through some applications [21], [23]. Table 2 represents the details of generated concept map.

TABLE 2. CONCEPT MAP CHARACTERISTICS

# nodes	# edges
108	454

Applying all the activities of second phase in order to prune useless data and improve the concept map with some details followed by completing and verifying by the group of experts, the generated ontology has the specified features (Tables 3). This ontology contains 13 different relations. Since the ontology edges are two-sided, it has 26 various types of edges.

TABLE 3. ONTOLOGY CHARACTERISTICS

# classes	#instances	#properties	#sub-class relations
171	51	86	152

The process of evaluating the third phase of ontology generation is accomplished through adding several concepts, including "conceptual model", "open source", "Kinect", "exercise", "authorship", "editor", "agent", "OER", "regular", "disable", "MOOC", and "Coursera".

Experimental results showed that the appliance of WordNet as a general purpose dictionary does not provide a good solution for eLearning domain. The main reasons are as follows:

- Various concepts in the eLearning domain are composed of multiple words and the complete form of them is not involved in the general purpose dictionary. "Open source" and "conceptual model" are some instances.
- Some domain specific words are the acronym of compound words validated only in the same domain. "OER" is an instance.
- Many words applied in the domain associate to special tools or methods of that domain. "Kinect" and "coursera" are placed in this group.

Considering the reasons, the usage of WordNet is beneficial in only six input terms. Table 4 indicates the number of edges added to the ontology graph for each concept.

TABLE 4. THE NUMBER OF EDGES ADDED TO THE ONTOLOGY USING WORDNET

Concept	#	Concept	#	Concept	#
Exercise	6	Authorship	4	Agent	1
Disable	3	Editor	2	Regular	0

The type of the new edges that connect each two concepts can be in the form of "is related to". However, in 39% of the new links, the edges are not reasonable in the domain of eLearning.

The encyclopedia of Wikipedia is an important source of information, so that each page is covering a title and its links are indicators of its semantic relations. One advantage of using Wikipedia is its

possibility to covering numerous titles. However, there are some problems in using this source.

- Many pages in the site of Wikipedia are linked to names or addresses of persons, organizations, or other proper names. These terms cannot be considered as classes. So, we can add them to the ontology in the role of instances.
- The title of many pages in Wikipedia is not a term or a concept. "List of Latin words with English derivatives", "Analysis of algorithms", "List of computer scientists", "Field-programmable gate array", "Talk: Computer architecture", and "Scientific journal" are some instances.
- Some output links of a page are more explanations or examples mentioned for justifying the page content. Many of these links don't demonstrate a semantic relation.

Table 5 demonstrates the number of links created for each of the input concepts.

TABLE 5. THE NUMBER OF EDGES ADDED TO THE ONTOLOGY USING WIKIPEDIA

Concept	#	Concept	#	Concept	#
Coursera	4	Disable	1	Agent	9
MOOC	16	Authorship	4	Regular	10
Exercise	4	Editor	2	OER	6
Open source	4	Conceptual model	5	Kinect	5

All of the created edges are from the type of "is related to". Therefore, increasing the number of input concepts transforms the ontology to a concept map again. Applying domain related rules, which combine semantic and statistic features in the concept map generation phase [21], on the new added relations can delay this conversion.

In the linking process of each concept, some new terms are added to the ontology which are counted in table 6.

TABLE 6. THE NUMBER OF TERMS ADDED TO THE ONTOLOGY USING WIKIPEDIA

Concept	#	Concept	#	Concept	#
Coursera	22	Disable	3	Agent	3
MOOC	5	Authorship	4	Regular	7
Exercise	5	Editor	2	OER	7
Open source	10	Conceptual model	4	Kinect	3

About half of the new terms added to the ontology are the name of persons, organizations, places, domains of education, and examples to complete the content. These terms have not been considered as the ontology classes; but can be added to as the instances. This type checking should be done manually. Therefore, in specified periods of time, the enriched ontology should pass the second-phase of ontology generation. This is because of determining the type of new added nodes and their relations.

V. CONCLUSION

In this paper, we proposed a methodology for extracting a semantic network from a corpus of documents. Pruning useless data and improving with additional details, we converted the semantic network to the ontology. Sciences are in progress, so we enriched our methodology to a mechanism that gradually promoted the ontology and added new terms and relations to it. We applied some rules specified for the field of eLearning in the creation of ontology, so this ontology is distinguished for this domain. However, the proposed method can be personalized for any other domain.

We believe that integrating the generated ontology with content and learning management systems (CMSs and LMSs) will improve their services. Therefore, future work would involve combining ontology with a CMS. Using the CMS repository, we can incrementally refine and update the ontology and consequently better annotate the archives. One application of the generated ontology is to cluster domain specific documents. Therefore, the other future directions include finding methods that combine different features and semantics from the ontology with more advanced techniques for clustering eLearning documents.

REFRENCES

[1] T. Berners-Lee, J. Hendler, & O. Lassila, "The Semantic Web", *Scientific American Magazine*, 2001.

[2] G. Guizzardi, & G. Wagner, "A Unified Foundational Ontology and some Applications of it in Business Modeling", *in Proceedings of the CAiSE'04 Workshops*, Riga, Latvia, 2004.

[3] T.R. Gruber, "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, Vol. 5, pp.199–220, 1993.

[4] B.E. Idrissi, S. Baina, & K. Baina, "Automatic Generation of Ontology from Data Models: A Practical Evaluation of Existing Approaches", *in Proceedings of IEEE Seventh International Conference on Research Challenges in Information Science (RCIS)*, Paris, 2013.

[5] W. Wong, W. Liu, and M. Bennamoun, "Ontology Learning from Text: A Look Back and into the Future," *ACM Computing Surveys*, Vol. 44, pp. 1-36, 2012.

[6] E. Kubias, S. Schenk, S. Staab, and J. Z. Pan, "OWL SAIQL - An OWL DL Query Language for Ontology Extraction," in *Proceedings of Third International Workshop on Experiences and Directions*, inncbruck, 2007.

[7] H. A. Santoso, S.-C. Haw, and Z. T. Abdul-Mehdi, "Ontology Extraction from Relational Database: Concept Hierarchy as Background Knowledge," *Knowledge-Based Systems*, Vol. 24, pp. 457–464, 2011.

[8] M. Hamasaki, Y. Matsuo, T. Nishimura, and H. Takeda, "Ontology Extraction by Collaborative Tagging with Social Networking," in *Proceedings of WWW*, Beijing, 2008.

[9] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, and E. Zavitsanos , "Ontology Population and Enrichment: State of the Art," in *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*. Springer Berlin Heidelberg, 2011, pp. 134–166.

[10] I. Bedini and B. Nguyen, "Automatic Ontology Generation: State of the Art," *Journal of Molecular Evolution*, Vol. 44, pp. 226-233, 2005.

[11] *M.J. O'Connor & A. Das, "Acquiring OWL ontologies from XML documents", in Proceedings of the sixth international conference on Knowledge capture*, Alberta, Canada, 2011

[12] K. Meijer, F. Frasincar, and F. Hogenboom, "A Semantic Approach for Extracting Domain Taxonomies from Text," *A Semantic Approach for Extracting Domain Taxonomies from Text*, Vol. 62, pp. 78–93, 2014.

[13] I. Serra, R. Girardi, and P. Novais, "Evaluating Techniques for Learning Non-Taxonomic Relationships of Ontologies

from Text," *Expert Systems with Applications*, Vol. 41, pp. 5201–5211, 2014.

[14] M. Hourali, A. Montazer, "A New Approach for Automating the Ontology Learning Process using Fuzzy Theory and ART Neural Network", *Journal of Convergence Information Technology(JCIT)*, Vol. 6, pp.24-32, 2011.

[15] Z. Xu, et al., "Mining temporal explicit and implicit semantic relations between entities using web search engines," *Future Generation Computer Systems*, Vol. 37, pp. 468–477, 2014.

[16] Z. Li and D. Tate, "Automatic ontology generation from patents using a pre-built library, WordNet and a class-based n-gram model," *International Journal of Product Development*, Vol. 20, pp. 142-172, 2015.

[17] D. Kucuk and Y. Arslan, "Semi-Automatic Construction of a Domain Ontology for Wind Energy Using Wikipedia Articles," *Renewable Energy*, Vol. 62, pp. 484–489, 2014.

[18] J.H. Gennari, M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubézy, H. Eriksson, N.F. Noy, S.W. Tu, "The evolution of Protégé: an environment for knowledge-based systems development", *International Journal of Human-Computer Studies*, Vol. 58, pp. 89–123, 2003.

[19] M. Banko, M. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open Information Extraction from the Web," *Communications of the ACM*, Vol. 51, pp. 68-74, 2008.

[20] F. Barrientos and G. S., "Interpretable knowledge extraction from emergency call data based on fuzzy unsupervised decision tree," *Knowledge-Based Systems*, Vol. 25, pp. 77-87, 2012.

[21] M. Mosharraf and F. Taghiyareh, "Automatic Domain Specific Ontology Generation for e-Learning Context," in Proceeding of International Conference on Virtual Learning- Virtual Reality (ICVL), Romania, 2015.

[22] I. Nonaka and H. Takeuchi, The Knowledge Creating Company. New York: Oxford Univ. Press, 1995.

[23] M.Mosharraf, F.Taghiyareh, "Investigating ELearning Research Trend in Iran via Automatic Semantic Network Generation", *Journal of Global Information Technology Management*, Vol. 20, pp.91-109, 2017.

[24] A. Clauset, M. Newman, and C. Moore, "Finding Community Structure in Very Large Networks," *Phys. Rev. E*, Vol. 70, no. 6, 2004.

[25] F. Wang, and C. Zhang,, "Label Propagation through Linear Neighborhoods," in *Proceedings of the 23 rd International Conference on Machine Learning*, Pittsburgh, 2006.

[26] V. Dongen, "Graph Clustering by Flow Simulation," PhD thesis, University of Utrech, Utrecht, 2000.

[27] A. Budanitsky and G. Hirst, "Semantic distance inWordNet:An experimental, application-oriented evaluation of five measures," in *Workshop on WordNet and other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Lingustic*, 2001.

Maedeh Mosharraf is a Ph.D. candidate at the Department of Computer Engineering in the University of Tehran. She received her B.Sc. and M.Sc. in Information Technology engineering from University of Tehran in 2010 and 2013, respectively. Her research involves semantic web technology and technology-enhanced learning.

Fattaneh Taghiyareh is an Associate Professor of Computer Engineering, Software and Information Technology at the School of Electrical and Computer Engineering, University of Tehran, where she has served since 2001. She received her Ph.D. in Computer Engineering from the Tokyo Institute of Technology in 2000, and her M.Sc. and B.Sc. from Sharif University of Technology in 1994 and 1990, respectively. Currently, she is serving both Software Engineering and Information Technology Departments as a member and she is supervising the Multi-Agent System Lab as well as eLearning Lab. Her research interests are in collaborative environments for eLearning and Multi-Agent Systems. Her current research involves the intersection of web-services, personalization, and adaptive LMSs as well as applying ontology to semantic web.

This Page intentionally left blank.