Investigating eLearning Research Trends in Iran via Automatic Semantic Network Generation

Maedeh Mosharraf, Fattaneh Taghiyareh & Sara Alaee


To link to this article: http://dx.doi.org/10.1080/1097198X.2017.1321355

Published online: 01 Jun 2017.

Submit your article to this journal

Article views: 13

View related articles

View Crossmark data
Investigating eLearning Research Trends in Iran via Automatic Semantic Network Generation

Maedeh Mosharraf, Fattaneh Taghiyareh, and Sara Alaee

Electrical and Computer Engineering Department, Faculty of Engineering, University of Tehran, Tehran, Iran

ABSTRACT
The purpose of this study is to investigate Iran’s eLearning research status in comparison with the world. We propose a method based on a text mining approach for extracting knowledge from Iranian published articles and generating the corresponding semantic network automatically. eLearning concepts are extracted from papers published in 6 years’ proceedings of ICeLeT, an International Conference on eLearning and eTeaching, in Iran. After extracting the domain-specific concepts, each pair of concepts get the possibility to be linked together based on co-occurrence in the articles. A weight is assigned to each edge according to the pointwise mutual information value of the pair of concepts. To identify gaps between the latest local and global research, the obtained semantic network is compared with another semantic network extracted from 6 years’ proceedings of ICALT, an International Conference on Advanced Learning Technologies. By applying a hybrid clustering algorithm on two networks based on the combination of label propagation and Markov clustering, and identifying the differences between node memberships and hubs, strengths and weaknesses of each network are demonstrated.

KEYWORDS
International Conference on Advanced Learning Technologies (ICALT); International Conference on eLearning and eTeaching (ICeLeT); Iran; knowledge extraction; research trend; semantic network

Introduction
Twenty-first-century skills such as critical and creative thinking, communication and collaboration, information literacy, and professional developments are demands of new ways of learning. Learning these new skills cannot just be added to existing learning goals. It means that they ask for a fundamental rethinking of learning trajectories. For this purpose, not only existing learning contents and objectives should be improved, but also learning methods should be evolved. Technology plays a significant supportive role in learning upheaval. Unlike learning, there is fairly clear agreement among academic members on the definition of technology. Technology is understood as the process by which humans modify nature to meet their needs and wants (Selwyn, 2011).

Integrating technology into education leads to introduction of technology-enhanced learning (TEL). TEL methods can change the deployment of the most important resources in the educational systems (i.e., teachers’ and learners’ time). Although TEL is not restricted to high technology (Richey, 2008), electronic educational technology has become an important part of today society. eLearning is a clear example of TEL.

With the expansion of global research in the field of eLearning, Iranian researchers and industrialists—like their counterparts in other countries—began to address this new paradigm in their studies. Establishment of the Iranian Association of eLearning was an Iranian endeavor to employ the country’s professional capacity and success in eLearning practices. It is a scientific, non-
governmental, non-profit, and non-political community with the following objectives (Iranian eLearning Association, 2015):

- To improve teaching–learning process;
- To reinforce fledgling nature of eLearning in Iran;
- To promote knowledge, insight, and scientific culture in the field of eLearning;
- To fix flaws in eLearning plans and implementations;
- To assess policies and administrative programs of eLearning systems at the national level; and
- To supervise relevant eLearning departments to provide adequate legislations and guidelines.

Another example is the International Conference on eLearning and eTeaching (ICeLeT) held annually since 2009 (ICeLeT, 2015). This conference aims to provide an opportunity for presenting the latest research and technological advances in the field of eLearning with the cooperation of universities, research centers, scholars, and engineers. Although this conference receives papers from the entire world, most of them are from Iran. Same as other international conferences, ICeLeT is enthusiast to bridge some professional eLearning gaps. While Iranian publications in international professional societies (i.e., worldwide journals and conferences) are not negligible, their propensity to ICeLeT is more considerable. ICeLeT can be regarded as Iran’s research performance indicator in the field of eLearning. Therefore, papers presented in this conference are expected to show Iran’s progress in eLearning studies. Comparing Iran’s progress with proceedings of the entire world can warn this country’s politicians and researchers. It is hoped that recognition of backwardness motivates them to advance promptly. Therefore, recognizing the strengths and weaknesses of Iran’s state in eLearning research can help decision makers develop better policies.

In this article, we consider Iranian articles published in the ICeLeT proceedings as our dataset. For our baseline, we chose articles published in the proceedings of ICALT (the International Conference on Advanced Learning Technologies). We generate a semantic network for each group and compare their structures.

Semantic network is formalism for knowledge representation that allows demonstrating semantic relations between concepts as a graph. It expresses concepts logically and in the form that is machine-understandable and tractable (Steyvers & Tenenbaum, 2005). Semantic network can be an intermediate for translating natural language to machine-understandable language. In such network, which is a labeled directed graph, nodes represent concepts and edges stand for binary relations between nodes. In the present work, we model the entire ICeLeT articles, as well as the ICALT articles, as a semantic network. In the network generation, each of the eLearning concepts extracted from the papers corresponds to a unique node. Using pointwise mutual information (PMI) metric, the weight of the edge between each pair of nodes is calculated. The generated semantic network is utilized for sub-domain clustering of eLearning concepts and speeds up investigations.

The generated semantic network as an integrated knowledge-base extracted from different papers can be used to model education domain and to build, organize, and update specific learning resources. This network specifies a shared understanding of the domain, since it contains a set of domain concepts, together with their synonyms and relationships. The implemented method for explicit representation of eLearning concepts and their relations in the form of a network can lead to the straightforward generation of domain-specific ontology. In eLearning, like many other fields of research, ontology can easily manage the domain knowledge and allow a more detailed organization of the concepts. Some other benefits of deploying the ontology or even the semantic network in eLearning systems are as follows:

- Generating the domain-specific lexicon, which explicitly represents the knowledge of many documents related to the domain. This lexicon can have many appliances in the future of eLearning;
eLearning document clustering, which enhances categorizing data archives, retrieval efficiency, and accessibility in adaptive content management systems (CMS);

- Richer annotation of learning materials, which is a way to deal with some new learning demands, such as sharing, reusing, and remixing of learning objects;
- Semantic querying and navigating through learning materials;
- Agent-based architecture of eLearning environments, which brings high-level supports with automatic, flexible, and interoperable agents;
- Web services usage, which reduces eLearning systems loads, and brings them versatility, code re-using, and cost saving; and
- Modeling learning activities and materials, which will improve personalization process and recommender modules in learning systems.

This network tries to reduce or eliminate conceptual and terminological confusion among the members of a particular community who need to share different kinds of documents and information. We believe that integrating the generated network with a CMS or a learning management system (LMS) will improve its services. Using the CMS repository, we can incrementally refine and update the generated network and consequently better annotate the archives. We hope applying the same methodology to create the semantic network for each domain helps improve organized information systems in that domain. Collecting, organizing, filtering, storing, communicating, processing, and distributing data can be facilitated in information systems using the semantic network. Organizing concepts and their predefined relations are some beneficial steps toward the ontology generation, machine-readable processes, and semantic web.

The rest of the article is organized as follows: In the next section, a review of the related works on knowledge extraction from text documents is presented. Our proposed methodology for automatic knowledge extraction in the form of semantic network is then explained. Next, research findings are described. Following that, some further discussions are imparted, and finally, the article is concluded.

**Related Work**

Increase in the number of unstructured contents brings new challenges to knowledge extraction, classification, and retrieval. Specific text mining techniques are used to extract knowledge from unstructured textual data (Feldman & Sanger, 2007). These techniques apply sets of linguistic, statistical, and machine learning methods to model and structure the information contents. In addition to these methods, some of the textual analysis techniques use external knowledge-bases to better interpret the meaning of contents. Wikipedia and its link structure (Wang & Domeniconi, 2008), WordNet taxonomies (Wei, Lu, Chang, Zhou, & Bao, 2015), and ontologies and their conceptual relations (Rajpathak, 2013) are examples of these knowledge sources.

Knowledge extraction from text is the process of transition from a non-formalized text written in natural language to a formalized actionable language. Over the past few decades, considerable works have been published concerning the development of methods to harvest knowledge automatically. The proposed methods are grouped in three categories: supervised, unsupervised, and logic-based.

The supervised methods require a set of training reviews annotated according to predefined features. In Jahiruddin, Abulaish, and Dey (2010), information extraction process is based on a biomedical named entity recognizer, which tags genes, proteins, and other entity names in the text. Using natural language processing and latent semantic analysis (LSA), key concepts are identified. Ciravegna (2001) proposed an adaptive algorithm that utilizes examples from user-defined tags to learn symbolic rules that insert tags into texts. New tags assigned to texts are representatives of knowledge hidden in the texts. TEXTRUNNER is a tool developed in Banko, Cafarella, Soderland, Broadhead, and Etzioni (2008) for information extraction from web documents, which is based on three phases. In the self-supervised learning phase, a classifier is generated that labels candidate
extractions. In the single-pass extracting phase, tuples from all possible relations are extracted from entire corpus. Each retained tuple is assigned a probability, which determines the likelihood that it is a correct instance of the relation in the redundancy-based assessor phase (third phase). Mooney and Nahm (2005) used a learned information extraction system to transform text into more structured data, which is then mined for interesting relationships. Wu and Weld (2010) presented a self-supervised learning method that uses bootstrapping from entries in Wikipedia info-boxes to learn extraction patterns in dependency parsers.

In the unsupervised methods, hidden knowledge is discovered from unlabeled data with no predefined features. In Shinyama and Sekine (2006), a simple procedure is applied to extract knowledge from a collection of documents. At first, the entire documents are clustered, so that all the documents of a group discuss similar topics. Then, named-entity recognition and co-reference resolution are performed within each cluster, and deep linguistic parse structures are obtained to identify set of entities and their relations. Opine (Popescu, 2007) is an unsupervised information extraction system, which extracts fine-grained features and associates opinions to customer comments. It takes advantage of the unique characteristics of web texts and leverages existent search engine technology. Knowledge extraction can be accomplished using feature extraction, which is particularly important in some text analytics and applications. Carennini, Ng, and Zwart (2005) introduced a method for feature extraction that draws on the unsupervised method presented in Hu and Liu (2004) as a text mining approach. In order to represent these extracted features, they organized them in a hierarchical fashion. In order to support decision making about resource planning of an emergency call center, Barrientos and Sainz (2012) used a fuzzy version of an unsupervised decision tree, which merges decision trees and clustering.

In the third group of these methods—which are based on logic—machine learning and rule discovery approaches are applied. Blaschke and Valencia (2002) proposed an approach to develop information extraction rules by encoding regular expressions. Its focus is on biomedical texts to identify the desired entities or relations of interacting proteins. Banko and Etzioni (2007) introduced an agent whose goal is to automatically discover a collection of concepts, facts, and generalizations, which describe a particular topic of interest directly from a large volume of web texts. This agent is claimed to be a lifelong learning agent, since it iteratively adds general domain knowledge to its theory. Pujara, Miao, Getoor, and Cohen (2013) proposed a method based on probabilistic soft logic to reason about candidate facts and their associated extraction confidences. The paper combines the tasks of entity resolution, collective classification, and link prediction mediated by rules based on ontological information. In order to combine diverse data from hundreds of millions of web pages and construct a knowledge base, Niu, Zhang, Re’, and Shavlik (2012) employed machine learning and statistical inference techniques. These techniques are applied on relational features, which are extracted from web pages.

Depending on the purpose of knowledge extraction from text, each of the described methods can be applied. Extracting entities, terms, and concepts, finding their relations, and creating either semantic network or ontology are some purposes of knowledge extraction. Named entity recognition involves identifying references to particular kinds of objects such as names of people, companies, and locations (Bikel, Schwartz, & Weischedel, 1999). Terms are linguistic realizations of domain-specific concepts. The literature provides many examples of term extraction techniques—most of them are based on information retrieval methods for term indexing, which are combined to more or less advanced levels of linguistic processing (Buitelaar & Magnini, 2005). It is not clear what exactly constitutes a concept. So, extraction of concepts from text is a challenging task. Relation extraction consists of extracting hierarchical relations (is-a, has-a) and non-hierarchical (all kinds) relations. Even though the aims of knowledge acquisition and semantic network generation tend to overlap, there are a number of innovative aspects for semantic network generation, which sets it apart from much of the previous works in knowledge acquisition. Concept recognition, synonym detection, and taxonomy construction are some of these aspects. Hourali and Montazer (2011) used neural network clustering for concept detection. Yildiz, Yildirim, and Diri (2014) applied lexico-syntactic patterns to
a big corpus to determine synonyms. de Knijf, Frasincar, and Hogenboom (2013) used the extended subsumption method to generate semantic taxonomies.

**Semantic Network Generation**

Educational improvements do not necessarily have to be driven by digital technologies. Nevertheless, the interest of research on using these technologies to improve teaching and learning is ever-increasing. This article proposes a method to analyze the progresses in technology-enhanced education, particularly eLearning. By looking at eLearning research trends from 2009 to 2014 in Iran and other countries, we can identify the near-future research in Iran, decide where to focus, and design the educational research plan.

This work aims to introduce a method to extract Iran’s progress in eLearning domain and compare it with the global state. In this respect, the semantic network of eLearning concepts is extracted from the published papers. Generation of the semantic network is accomplished in three phases according to Figure 1.

As we can see in Figure 1, all the articles from the input corpus will eventually be assimilated into a graph. In the pre-process phase, all the candidate words from each paper are extracted, stemmed, and unified. Setting the co-occurrence window as the size of the title and abstract, each two words in the same window are linked together with an edge. The edge weight specifies the normalized number of co-occurrence between nodes. Finally, a partitioning method is applied on the graph. The following sections explain these processes in more detail.

**Pre-Processing Step**

Each paper has a theme that can be described through a set of words. We need to extract these sets of main words from each article. Finding these words is the same problem as assigning keywords to

![Figure 1. The process of semantic network generation.](image)
each paper. The author-specified keywords are not sufficient, since they do not describe the article comprehensively. The specified keywords are biased based on authors’ opinions. In addition, these keywords are chosen so that the paper can get higher rank by search engines. Therefore, we cannot assign a set of words to each paper based on the list of keywords. In our proposed method, a window with the size of title and abstract is defined for each paper. The words within this window are considered as candidates for the whole paper. The following steps are performed to find the describing paper nomenclatures.

**Candidate Words Extraction**
This step consists of extracting all the words in each paper that constitute our corpus and removing general words from them. It begins with sentence detection in the text content and determining the grammatical roles of the words. Stop words are not represented in the domain concepts set. They include verbs such as “is,” “am,” and many other verbs such as “provide,” “introduce,” etc. (which are identified using the bag-of-words approach), identifiers such as “the,” “a,” “an,” etc., and propositions such as “at,” “in,” “on.” In order to identify the parts of speech of the words based on their functions in sentence (noun, verb, etc.), we use the Stanford part of speech (POS) tagger (The Stanford Natural Language Processing Group, 2015). POS tagger is a tool for identifying the parts of speech corresponding to each word in a given text. It is a preliminary and syntactic processing stage in language processing.

**Compound Words Solidification**
Compound words are written differently. For example, “e-Learning” and “eLearning” are different representations of one word. Yet with a hyphen, an automatic tool might recognize them as different words. In addition, compound words might be written with whitespace. In automatic text processing, whitespaces are typically not processed. For example, a machine processor assumes “learning management system” as three separate words, while it actually illustrates one concept. In general, different forms of writing compound words, such as summarizing either all parts of compound words or some parts (“eLearning” vs. “electronic learning”), or adding or removing some parts from the compound words (“educational data mining” vs. “educational mining”), are some examples of existing problems with automatically identifying compound words. Several techniques such as association rule mining have been introduced to help with this process, but they all need huge corpora. To avoid these problems in our work, we decided to solidify compound words by removing hyphens, spaces, and other delimiters between each part of a compound word, manually (e.g., by expert judgments).

**Words Unification**
To avoid excessive repetitions of words in text documents, it is good to use their synonyms. For example, “student” and “learner” are often used interchangeably. These synonyms can be merged to create a super node in the semantic graph. Due to lack of special-purpose eLearning lexicon and high error rate in using general-purpose dictionaries, we performed this task manually (i.e., with a combination of help from both available lexicon and domain experts). In other words, we identified the set of words that have same meanings and are applied interchangeably by investigating all the candidate words extracted from the eLearning corpus and using the knowledge of eLearning and language experts to unify them.

**Words Standardization**
Natural language concepts may be expressed in a multitude of forms. In order to detect different forms of concepts, the similarity of multi-word expressions should be defined. Case folding and stemming are the two techniques for this purpose. In the case folding process, all the characters of extracted words in the corpus as well as those in the lexicon are converted to either lower case or upper case. Here, we transformed them to lowercase form. In addition, the stemming is performed on each extracted word. Stemming is the process for reducing words to their root forms. This is necessary because words with
the same roots should be considered as one concept in our method. We used Porter’s Stemmer Algorithm (Porter, 1997), a well-known stemmer for English words, for this purpose.

**Network Generation**

A graph-based representation of structural information combined with a clustering technique has been shown to be successful in knowledge discovery (Aggarwal & Zhai, 2012). For this reason, we propose a new method to represent the entire available articles in the corpus as a form of a network. The following steps are performed for creating the extracted graph.

**Low-Score Words Elimination**

In a large domain-specific corpus, a low frequency word can be regarded as an outlier. The word’s score is defined as the number of occurrences of that word in the abstract. By the defined threshold, only words with higher scores are preserved and all the lower score words are eliminated. Removing useless data leads to clean corpus, which is a factor in increasing accuracy and the speed of clustering process. Synonyms are also replaced with super nodes. As mentioned earlier, using synonyms of a word in the text is similar to repeating that word. Therefore, the score of super node is equivalent to sum of the scores of all its sub-nodes. Elimination process is not applied on the words from titles, meaning that all the extracted keywords from titles are considered concepts in eLearning domain, and hence, nodes of the graph. This is because each title is a compact version of its abstract and expresses the essence of its corresponding paper.

**Edge Weight Calculation**

Semantic network generation is based on word co-occurrence. If we assume a semantic network as a graph of words, each concept would be represented as a node, the corresponding relations with other words would be the edges, and PMI provides the edges weights. The PMI value between two nodes is calculated according to Equation 1:

\[
p(M_{ij}) = \log \frac{p(x_i, x_j)}{p(x_i)p(x_j)} = \log \frac{f(x_i, x_j)}{f(x_i) + f(x_j)} = \log \frac{N * f(x_i, x_j)}{f(x_i)p(x_j)}
\]

In this equation, \(x_i\) is a node in the semantic network, \(f(x_i)\) is the frequency of occurrences of \(x_i\) in the predefined window, and \(f(x_i, x_j)\) is the frequency of occurrences of \(x_i\) and \(x_j\) together in the entire window.

**Graph Creation**

Based on Li (2002), Qiu, Wang, Wang, Dong, and Zhang (2011), and Yamaguchi (2001), each pair of terms with high frequency on co-occurrence statistics can have conceptual relationships. In the previous sub-step, an edge was drawn between every pair of words that occurred together in the predefined window (the weight equals their PMI is assigned to the edge). However, a small value of PMI cannot be representative of a conceptual relation. Defining a threshold for PMI values, all edges whose weights are lower than this threshold will be removed.

**Graph Refinement**

Although connected nodes are related statistically, not all of them are associated semantically. In order to combine semantics and statistics in the network relations, all the edges from the previous sub-step should be refined with the following rules. If none of these rules is applicable to an edge, that edge will be removed.
• An edge represents the inclusion or inheritance relation of two nodes, and thus forms a concept hierarchy.
• From two nodes, which are linked using an edge, one of them is a tool for doing or promoting another.
• One of the nodes involved in an edge is an action in learning or eLearning processes. Verbs such as “assess,” “assign,” “learn,” “teach,” “game,” “study,” and “collaborate” are examples of these nodes.

Network Partitioning

A large semantic network makes analysis process difficult. Graph partitioning can facilitate the process of analyzing the structural and functional properties of this complex network. Graph partitioning should be done semantically, meaning that the nodes placed in a group should be semantically related. In this respect, sets of nodes should be determined so that weights of relations between the nodes inside the sets are semantically higher than the relations weights of nodes outside the sets. This definition means modularity maximization (Clauset, Newman, & Moore, 2004). To this end, we applied a combination of label propagation (Wang & Zhang, 2006) and Markov clustering (Dongen, 2000) algorithms.

In label propagation, which is run iteratively, each node of the network is initially given a unique label. 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 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 is no overlap between clusters obtained from two algorithms, then the cluster with the maximum modularity will be the final cluster. Modularity is defined by Equation 2 (Clauset et al., 2004):

\[
Q = \frac{1}{2m} \sum_{v,w} A_{vw} \delta(C_v, C_w) \tag{2}
\]

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 is no edge between nodes \( v \) and \( w \), and \( A_{vw} = 1 \) means there is an edge between the two nodes. If we suppose the nodes are divided into clusters such that node \( v \) belongs to group \( C_v \), \( (C_v, C_w) \) 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.
**Research Findings**

**Datasets and Semantic Networks Features**

This study aims to build two distinct eLearning semantic networks based on the articles from two conference repositories, namely ICALT and ICeLeT, from 2009 to 2014.

ICALT and TEL are organized by the IEEE Computer Society and the IEEE Technical Committee on Learning Technology. This conference accepts English papers reporting original academic or industrial research in the world. ICALT aims to bring together the world people who are working on the design, development, use, and evaluation of technologies that will be the foundation of the next eLearning systems generation. Despite the conference slogan, which is focused on a special theme each year, ICALT received papers in various tracks of eLearning. Numbers of accepted papers from the countries that participated in this conference—diverse tracks of the conference, various topics of accepted papers, desirable average citation count, and the strict process of reviewing received papers—motivated us to choose the proceedings of this conference as the latest research in the eLearning around the world. Table 1 represents the detail of ICALT proceeding in each year.

ICeLeT is the other international conference, which accepts both domestic and foreign papers in both Persian and English languages. As mentioned earlier, the reason for choosing the ICeLeT papers as the experimental input corpus was to focus on research status in Iran. Therefore, the foreign papers were removed from the ICeLeT proceedings as outliers. The Persian-written articles were also removed, since our methodology was only focused on English language.

Two datasets used in this research—namely the ICALT and ICeLeT papers—have the size 1,270 and 91, respectively. The big difference between the number of articles in ICALT and ICeLeT makes their comparison difficult. Providing similar conditions in generating semantic networks for both datasets—for example, choosing proportional thresholds in the generation phase—could be one of the solutions. Through this approach, which is like a porous chunking of the node scores in each dataset, the scores in the lowest decile are chosen to be removed. Considering the edge weight interval, the one-third lowest weights are also chosen to be removed.

Some nodes in our semantic networks seem to be redundant. These are nodes connected to most or all of the other nodes, and therefore contain no specific information. Network analysis and representation becomes clearer when such nodes are removed from the graph. These include “eLearning” and “education,” which can be considered as the hubs of the semantic networks. In further refinements, these nodes and related edges are removed from both networks. Table 2 illustrates the graph characteristics corresponding to these two networks after these refinements. Despite the different number of ICALT and ICeLeT articles, their graph characteristics are not significantly different. This is because of the frequently repeated domain-specific concepts in most articles.

As Table 2 indicates, the number of nodes and edges in the graph of ICeLeT is less than in the ICALT graph. In addition, the edge-to-node ratio of the ICeLeT graph is lower than that of the ICALT graph (3 for ICeLeT and 4.2 for ICALT). These numbers reflect the fact that articles which focus on eLearning and investigate or promote the domain concepts are inadequate in ICeLeT. The lower number of nodes in the ICeLeT graph also proves this statement. Some important concepts in the graph of ICALT that do not exist in ICeLeT include: “open learning,” “lifelong learning,”

### Table 1. Number of participating countries and paper acceptance rate of ICALT.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of participating countries</th>
<th>Acceptance rate (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>43</td>
<td>23.55</td>
<td>Aedo et al., 2009</td>
</tr>
<tr>
<td>2010</td>
<td>48</td>
<td>26.49</td>
<td>Jemni, Kinshuk, Sampson, &amp; Spector, 2010</td>
</tr>
<tr>
<td>2011</td>
<td>40</td>
<td>24.59</td>
<td>Aedo, Chen, Sampson, Spector, &amp; Kinshuk, 2011</td>
</tr>
<tr>
<td>2012</td>
<td>51</td>
<td>19</td>
<td>Aedo et al., 2012</td>
</tr>
<tr>
<td>2013</td>
<td>40</td>
<td>17.03</td>
<td>Chen, Huang, Kinshuk, Li, &amp; Sampson, 2013</td>
</tr>
<tr>
<td>2014</td>
<td>46</td>
<td>23.70</td>
<td>Sampson, Spector, Chen, Huang, &amp; Kinshuk, 2014</td>
</tr>
</tbody>
</table>
“MOOC,” “cloud,” and “disable.” They can be considered as some open topics which have not yet been investigated by Iranian researchers. Reviewing the Persian papers, which are removed from the ICeLeT corpus, also confirms the lack of these concepts in the Iranian research. The reason might be due to some gaps between the industry and academic needs. Nevertheless, scientific research can go further ahead to promote global knowledge. Therefore, it would be beneficial if Iranian researchers start working in these neglected areas after performing some feasibility studies.

As mentioned earlier, the low score nodes were removed due to noise elimination. By checking the removed nodes, we found out that some of these nodes actually contain novel concepts. Integration of these concepts into eLearning context has not yet been explored in the current studies. Table 3 shows the eLearning concepts removed in the removal step.

The same can be said about the old concepts in this domain. “Video conference,” “traditional learning,” “synchronous learning,” and “discussion board” are examples of concepts eliminated in the removal step. These concepts appeared with low frequencies in articles as examples, background terms, or even the main research subjects. By eliminating these nodes, the networks consist of only concepts that are most up-to-date and relevant in the eLearning domain.

To sustain the semantic relations, the edges with the weight lower than a threshold were also removed. Checking revealed that among the removed edges, there are few which satisfy the conditions mentioned in the section of the present article entitled “Network Generation.” Since only the small number of edges were eventually removed from the graph, the error rate in automatic removal of the related edges is relatively low. The formula used to calculate the error is given in Equation 3:

\[
err = \frac{\text{number of edges that shouldn’t be removed}}{\text{number of edges removed}}
\] (3)

The calculated error rates for ICeLeT and ICALT are illustrated in Table 4.

As shown by our results, the overall error rate for the edge removal is relatively low. Nevertheless, the error rate of ICeLeT is more than that of ICALT. The lower number of articles in ICeLeT can be one of the main reasons for its higher error rate. The small corpus reduces the probability of linking related concepts and forming a dense network.

**Concerning the ICeLeT Semantic Network**

Figure 2 illustrates the semantic network of the ICeLeT papers.

As shown in this figure, we obtained eight clusters from partitioning the ICeLeT graph. We selected the central concept of each cluster as the hub. Each hub gives the best description of all the concepts in its cluster. The ICeLeT hubs include: (1) Algorithm, which includes methods applied to improve learning processes; (2) Personalization, which includes characteristics and functions of user profiling; (3) Learning, which is a representative of some actions involved in the training and

<table>
<thead>
<tr>
<th>Table 2. Semantic networks characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>ICeLeT</td>
</tr>
<tr>
<td>ICALT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Novel e-Learning concepts removed in the low score removal step.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICeLeT</td>
</tr>
<tr>
<td>Dyslexic</td>
</tr>
<tr>
<td>Knowledge</td>
</tr>
<tr>
<td>Experimental learning</td>
</tr>
<tr>
<td>Decision support system</td>
</tr>
</tbody>
</table>
learning processes; (4) **Collaboration**, which is a delegate for various tools and methods involved in social or collaborative learning; (5) **Infrastructure**, which is the representative of some technical resources; (6) **Application**, which represents several appliances of eLearning; (7) **Resource**, which represents some types of resources and learning objects; and (8) **Higher education**, which includes some important concepts related to education in universities, academies, seminaries, and institutes of technology.

Organizing the concepts of the ICeLeT semantic network in several clusters is an indicant of the extent of Iranian research. More coherent relations between concepts in each cluster than the relations between the clusters demonstrates more independent research in the areas related to the cluster hubs and less interdisciplinary research. Some interdisciplinary research areas appear in the form of impurities, which exist in several clusters. In other words, it is expected that nodes belonging to a cluster relate meaningfully to the group hub. These meaningful relations are defined in the domain of eLearning. The cluster with the hub named “resource” is one of the impure clusters. In addition to learning resources and objects, several concepts related to eLearning algorithms and infrastructures have also been included in this cluster. One reason for this issue is that these topics are probed together in the Iranian studies. It is so justifiable, since developing eLearning resources and their usage requires further research in technical infrastructures and algorithms. Another impurity is the inclusion of the concept “game” in this cluster. In the Iranian research, “game-based learning” is a topic discussed recently, but most current research is focused on “gamification.” Gamification is the use of game thinking or game elements in non-game contexts to engage users in solving problems (Kapp, 2012). This concept was only studied in one paper in which the title was related to learning content. Therefore, the two terms have been grouped in the same cluster.

<table>
<thead>
<tr>
<th></th>
<th>ICeLeT</th>
<th>ICALT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error rate of automatically edge removing</td>
<td>0.0358</td>
<td>0.0126</td>
</tr>
</tbody>
</table>
Another gist obtained from the clustering results is that concepts such as “culture” and “Islam” are located in the cluster related to “application.” This is because religious beliefs of Iranians are associated with their academic studies. In the eLearning domain, some research focused on integrating these concepts into learning environments and applications. If these concepts had to be joined in one of the clusters, “personalization” would be the best choice. Although these concepts could be members of “user profile,” accomplished research does not apply them as user tags. By increasing the number of studies about these concepts, they will probably be placed in their appropriate group.

We anticipated having a group with a hub named “role” that represents all the actors of eLearning systems. However, the input corpus consists of only three actors (i.e., student, teacher, and professor), which have been grouped in the “personalization,” “learning,” and “higher education” clusters, respectively. These are groups with the most relevance to the context of these concepts, and naturally, they have the maximum number of links to these roles. Perhaps this is not surprising given that eLearning is still a relatively young field of research in Iran, with the age of less than 10 years. Promoting eLearning studies is expected to lead to deployment of other actors such as “instructor,” “mentor,” “assistant,” “author,” etc.

“Assessment” and “evaluation” are the other aspects of eLearning, which has been weakly addressed in the Iranian articles. The reason for this claim is that the extracted terms related to “evaluation” could not form an independent cluster. In addition, these nodes have low scores in the generated graph. “Assess” and “exam” are among the nodes with the lowest scores in the ICeLeT graph. Reviewing the non-English articles also demonstrates that “evaluation” has not been technically considered. Only in three Persian papers, these concepts were studied from the perspective of psychology. Accordingly, the “learning” cluster may be the best group for them as happened.

The high score of node “model” cannot be an indicator for a strong focus of Iranian researchers on user modeling and personalization alone. This term cannot be restricted to “user modeling” and might be used variously in the manuscripts. However, the high score of three nodes including “model,” “learningstyl,” and “personal” simultaneously indicates a high number of studies on user modeling with the aim of personalization. Therefore, we can say that personalization is a main subject of the Iranian studies.

Regardless of user (student) modeling and personalization, many studies focus on the teacher as an important user in eLearning systems. eLearning is still not accepted by many teachers. Several teacher-centered articles studied the feasibility of automating various teaching processes and the readiness of teachers to accept and use this technology.

By surveying the concepts of the ICeLeT graph, their score, and edges weights, and comparing them with the similar items in ICALT, we conclude that:

(1) There are some gaps in the Iranian studies in the following areas:

- Lifelong learning and its consequences;
- Open educational resources (OERs) and free access to education and learning;
- Massive open online courses (MOOCs);
- Clouds as eLearning infrastructures;
- Big data and its effect on eLearning;
- Smart phones and ubiquitous technologies for learning; and
- Disabilities like intellectual, emotional, and physical disabilities, and eLearning problems and solutions.
Iranians are at the starting point of research in several sub-domains of eLearning. These sub-domains include:

- Intelligent learning and agents’ roles in this domain;
- Digital game and intelligent toy-enhanced learning;
- Knowledge management in eLearning;
- Children characteristics and child-specific eLearning;
- Adults’ learning and eLearning as a tool for vocational training;
- Decision support systems in eLearning environments; and
- Experience-based learning and using technology to support it.

Quality and number of the published papers in ICeLeT in the following sub-domains are considerable:

- Facilitating the training of blinds and adjusting teaching methods to dyslexics;
- Providing security in eLearning environments (e.g., plagiarism and fraud detection in learning activities, and authorization and access levels in infrastructures); and
- Collaborative learning (concepts related to this topic are strongly connected and have high scores in the ICeLeT graph).

In addition to technology, the scope of ICeLeT covers other tracks such as psychology, case study, and review. Therefore, some of the papers accepted in this conference do not contain any technological innovations. These articles can be divided in two categories. In one category, research findings are all repetitive, with no significant added values, and low-level. But in the other category, findings are helpful for solving current issues of eLearning or beneficial for designing future strategies.

The high score of term “Iran” in the ICeLeT graph shows high number of articles focused on eLearning studies about Iran. In other words, many articles of the input corpus contain studies considering this country as the case. Most of them investigate “advantages and barriers,” “strengths and weaknesses,” or “present situation” of eLearning. Integrating them with the research focused on “eLearning acceptance,” “e-readiness,” and “eLearning readiness” in this country provides future national decision makers with a base to make decisions regarding policies in these areas.

eLearning in Iran is delivered by both the private sector and the governmental organizations. Several universities such as Amirkabir University of Technology, Iran University of Science and Technology, Shiraz Virtual University, and learning centers like Islamic virtual centers and Faculty of the Science of Hadith incorporated virtual learning into their educational programs. As we know, eLearning is the interpretation of the slogan “anyone, anytime, and anywhere.” However, courses offering in the virtual learning environments in Iran are only restricted to the “anywhere” concept (it is also limited to places with proper internet access). Not only recent developments have not been applied in these environments, but also experimentally validated solutions such as personalized services have not been employed yet. The weaknesses in Iran’s technology infrastructure might be a reason for this issue. Internet bandwidth, which is a basic necessity in eLearning deployment, is faced with serious problems. This makes virtual courses to be provided and presented in accordance with certain conditions. In addition to software and hardware limitations, eLearning in this country does not have a high acceptance rate. As we said, eLearning is still a young field in Iran. Therefore, the promotion of virtual learning environments in this country should be parallel with the development of software, hardware, and appropriate strategies.
The country’s development status as well as its information and communication technology (ICT) foundations is the motive for researchers to develop an integrated system of education and application. Developing the national learning network was a collaborative effort between universities and industries in this respect. The goal of this network was to provide an integrated infrastructure for educational activities in the four ministries, “Education,” “Health and Medical Education,” “Science, Research and Technology,” and “Labour and Social Welfare.” Several studies have been conducted based on this network in attempt to develop its dimensions. Articles presented in ICeLeT with the titles “challenges of eLearning in Iran,” “human factors in the eLearning acceptance,” and “critical success factors for eLearning establishment” are some examples.

Regarding this, the studies focused on merging eLearning with the current systems can also have added values. Using eLearning in initial training (Ministry of Education), skill training (Ministry of Labour and Social Welfare), higher education (Ministry of Health and Medical Education, and Ministry of Science, Research and Technology), and non-formal education (Ministry of Culture and Islamic Guidance), and its potential abilities to realize learning and training in each of these areas are the basis of some academic and industrial research. Furthermore, Iran’s background on religion attracted research on special topics, which can be incorporated into the eLearning domain. Concepts such as “Islam” and “epistemology” in the semantic network are some indicants of the claim.

Concerning the ICALT Semantic Network

As expected, the number of concepts extracted from the ICALT corpus is more than ICeLeT. However, the number of clusters in the ICeLeT graph is more than the graph of ICALT. The number of concepts in each cluster and their density in the ICeLeT graph are also lower than the similar cases in the graph of ICALT, which demonstrates the weaknesses of the Iranian studies. This issue can indicate a lower number of intra-cluster research in Iran. Figure 3 displays the semantic networks extracted from the ICALT papers.

We obtained seven clusters from partitioning the ICALT graph. Again, we introduce each cluster with a hub. The hubs are: (1) Higher education, which is the representative of many types of

![Figure 3. ICALT semantic network.](image-url)
eLearning concepts related to post-secondary education; (2) Resource, which is a delegate of many learning resources; (3) Collaboration, which is a hub for social learning activities and their tools; (4) Personalization, which is a hub and the purpose of user profiling; (5) Algorithm, which is a representative of algorithms applied in sub-domain of eLearning systems; (6) Assess, which is a representative of some concepts applied for measuring or evaluation; and (7) Role, which is the head for various actors of eLearning systems. Although an eLearning system typically contains various actors, the number of articles about student is not comparable with those about the other actors.

Although the number of clusters in the ICALT network is lower than in the ICeLeT, they cover more important sub-domains of eLearning. In addition to the common clusters in both networks, “role” and “assess” (obtained from the set of ICALT hubs minus the ICeLeT hubs) are two clusters with the most significant shares of research in this domain. However, due to little research related to these concepts in Iran, and consequently lower number of concepts with low score in the ICeLeT graph, they have not formed separate clusters. On the other hand, the concepts related to three clusters including “infrastructure,” “learning,” and “application” (obtained from the set of ICeLeT hubs minus the ICALT hubs) were also included in the ICALT graph, but have been classified among the other clusters. This shows that the global research in these sub-domains is growing and entering the interdisciplinary phase.

Comparing the clusters obtained from the ICALT graph with the ICeLeT also shows that some clusters in these two networks have the same hubs, but different constituent elements. For example, in the “personalization” cluster, in addition to concepts in common between ICALT and the corresponding cluster in ICeLeT, there exists concepts which are related to cognitive styles and learning theory (e.g., “cognition,” “regulation,” “pedagogy,” “psychology”). Adding the instances of these concepts to this group can enhance the efficiency of user modeling. For example, the concept “disable” is a member of this group. Instead of answering yes or no to questions about disability in student profiling, type of disability can also be defined. Similarly, this can be applied for learning styles. Same as the ICeLeT graph, high score nodes such as “model” and “personalization” are indicants of a high number of studies focused on user modeling and personalization. So, we can conclude that “personalization” is a hot topic in eLearning research worldwide.

Unlike the ICeLeT graph, there exists no cluster in the ICALT network with a hub named “infrastructure.” One reason might be that research topics focused on learning infrastructures are already saturated on the global level. In other words, the progress in learning infrastructures has been more rapid than their expected usages. However, since Iran is a developing country and does not have sufficient technological infrastructures, this concept was regarded as a main research topic, hence a hub in the corresponding network. In the ICALT network, the node “infrastructure” was placed in the “resource” cluster, because learning resources were considered as infrastructures in most papers. Although some concepts such as “wireless” and “cloud” exist in the ICALT graph, their low scores show that there are few studies on these subjects.

Another difference with the ICeLeT network is the concepts related to “semantic web.” In the ICALT graph, these nodes are members of the “algorithm” cluster, whereas in the ICeLeT, they are members of “resource.” “LOM” is the only concept which can be related to semantic web and is a member of the “resource” group in the ICALT network. The reason is clear. LOM (learning object meta-data) is a descriptive meta-data useful for sharing contents. Membership of other concepts related to semantic web in the “algorithm” cluster demonstrates that this technology is not restricted to learning contents. Unlike the ICeLeT graph, the “algorithm” cluster in the ICALT graph is densely populated. There also exists some systematic eLearning concepts in this cluster, such as “framework,” “context-aware,” “MOOC,” and “lifelong learning.” It seems that concepts related to eLearning algorithms and eLearning applications are combined in this cluster. Meaningfully representing these concepts would be possible through hierarchical clustering. Figure 4 shows the result of applying our graph partitioning approach on the “algorithm” cluster in the ICALT network.

Applying the proposed graph partitioning method on the cluster described with “algorithm” splits it into three sub-categories. The hubs for these three parts are “algorithm,” “application,” and
“learning management system (LMS).” Taking a closer look at the resulting sub-categories, it seems that applying a better grouping approach on the whole network could probably provide more meaningful and uniform clusters. In this regard, if the sub-category described by “application” is merged to the “higher education” cluster, we would have better partitions. The current segmentation demonstrates that more concrete research on “higher education” and “eLearning application” domains needs to be done. In addition, since the concept “infrastructure” is placed in the “resource” cluster, it would also be better to place the sub-category “LMS” in the “resource” cluster. This unfavorable segmentation shows that more studies on integrating LMS and LCMS (Learning CMS) need to be conducted.

Another discussion is about the concepts “country” and “foreign.” “Country” was removed in the low score words elimination step. This term can be considered as an indicator of the case studies. China and Taiwan are examples of these country case studies. The number of case studies in ICALT is less than in ICeLeT. This might be due to the fact that the main focus of ICALT is on technological aspects of eLearning. An edge between “foreign” and “language” indicates studies on language learning and teaching. However, these articles are not numerous.

**Discussion**

In Iran’s strategies, eLearning is not considered as an application, but as a comprehensive system. Therefore, one of the main tracks in ICeLeT is focused on eLearning systems and their social, economic, and cultural aspects. Nevertheless, in none of the accepted papers did these aspects appear. Directional and strategic issues of these systems are also the subjects of another track, which has not been addressed by any articles. This means that either researchers have been ignorant to these aspects or studies have not progressed far enough for publication.

As its name indicates, the focus of ICALT is on technological aspects of eLearning. Pedagogy and psychology as well as social, economic, and industrial aspects of eLearning have been less discussed in this conference. It is probably because such issues have less importance on international levels and have more impact on national planes. Nonetheless, some of them were expressed...
in the form of experiences and case studies which contribute to avoid re-inventing the wheel in this domain.

Application of big data in learning analytics is one of the main tracks in ICALT. The semantic network extracted from this conference indicates that considerable research has not been yet accomplished in this area. The same issue is noticeable in the ICeLeT network. This means that big data is a major gap in eLearning research.

**Conclusion**

This article analyzed the Iranian publications in the domain of eLearning during 6 years in proceedings of ICeLeT, the international eLearning conference held in Iran. Comparing Iran’s research status in this field with the world’s status was achieved through perusing the ICALT proceedings. To prevent reviewing all the published articles one by one, information from each paper was represented in the form of domain-specific concepts and their relations. The generated semantic network integrates all the extracted knowledge from separated articles in a uniform resource. Our proposed method for automatically generating semantic network is based on three phases. In the pre-processing phase, key concepts are extracted after four initial steps including candidate word extraction, compound word solidification, word unification, and word standardization. In the network generating phase, after eliminating low score concepts, the weights of edges are calculated using PMI. Network refinement based on the semantic rules is the final step in this phase. Same as document clustering, clustering the semantic network locates similar concepts together, which then facilitates the analyzing process.

Recognizing strengths and weaknesses of eLearning employment in Iran guides educational policies and decision-making process in this country. Focusing on the strengths and reinforcing the weaknesses will improve Iran’s current educational status and will make eLearning a major success to transform the country into a knowledgeable society. The reason why shortcomings exist in this research area might be due to lack of industry or academic needs. Nonetheless, promoting scientific research can be accomplished regardless of local application and only for promoting global knowledge. In addition, by considering multidisciplinary point of view, progress in one domain can lead to improve some other domains and to fuse knowledge or information for solving various problems.

Our methodology in extracting semantic network from eLearning articles and applying it to investigate the current condition can be generalized to other domains as well. This network has various applications in eLearning and recommender systems. Semantic clustering and organizing domain-specific documents are other usages postponed to the future works. Deployment of the semantic network in annotation of educational documents can improve CMS and provide the facility of representing more personalized contents.

**Note**

1. ICeLeT was launched in 2007 in Iran. Back then, only Persian papers were submitted in this conference. However, since 2009, the conference accepts international papers from different parts of the world written in both Persian and English languages.

**References**


Author Notes

Maedeh Mosharraf is a Ph.D. student at the Technology-Enhanced Learning Laboratory at 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 founded in 2005 as well as eLearning Lab founded in 2012 by her. 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.

Sara Alaee is a Master’s student at the Technology-Enhanced Learning Laboratory at University of Tehran. She received her B.Sc. in Information Technology engineering from the University of Tehran in 2011. Her research interests include eLearning, machine learning, knowledge management, semantic web, and data and information retrieval.