IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 26, 2017, accepted May 18, 2017, date of publication June 8, 2017, date of current version July 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2710315

Illustrate It! An Arabic Multimedia Text-to-Picture m-Learning System

ABDEL GHANI KARKAR, (Member, IEEE), JIHAD MOHAMAD ALJA'AM, AND ARIF MAHMOOD
Department of Computer Science and Engineering, College of Engineering, Qatar University, Doha 2713, Qatar

Corresponding author: Abdel Ghani Karkar (a.karkar@qu.edu.qa)

ABSTRACT Multimedia learning is the process of building mental representation from words associated with images. Due to the intuitiveness and vividness of visual illustration, many texts to picture systems have been proposed. However, we observe some common limitations in the existing systems, such as the retrieved pictures may not be suitable for educational purposes. Also, finding pedagogic illustrations still requires manual work, which is difficult and time-consuming. The commonly used systems based on the best keyword selection and the best sentence selection may suffer from loss of information. In this paper, we present an Arabic multimedia text-to-picture mobile learning system that is based on conceptual graph matching. Using a knowledge base, a conceptual graph is built from the text accompanied with the pictures in the multimedia repository as well as for the text entered by the user. Based on the matching scores of both conceptual graphs, matched pictures are assigned relative rankings. The proposed system demonstrated its effectiveness in the domain of Arabic stories, however, it can be easily shifted to any educational domain to yield pedagogical illustrations for organizational or institutional needs. Comparisons with the current state-of-the-art systems, based on the best keyword selection and the best sentence selection techniques, have demonstrated significant improvements in the performance. In addition, to facilitate educational needs, conceptual graph visualization and visual illustrative assessment modules are also developed. The conceptual graph visualization enables learners to discover relationships between words, and the visual illustrative assessment allows the system to automatically assess the performance of a learner. The profound user studies demonstrated the efficiency of the proposed multimedia learning system.

INDEX TERMS Multimedia systems, text-to-picture, learning technologies, conceptual graph matching, ontology.

I. INTRODUCTION

Multimedia learning applications have a broad prospect in enhancing user realization, education and training [1], [2]. Different research works in education support the assertion that textual illustrations can broadly improve the performance of learners in terms of diverse cognitive outcomes [2]. A user can learn more efficiently from words associated with pictures than words only. Experimental analysis, that compares learning from text with illustrations [3], strongly shows that depiction with illustration helps learning of textual content. However, just adding pictures and symbols to words may not be an efficient manner to achieve improved multimedia learning. The objective is to provide instructional illustrations in the light of mental representation and the context of words [4]. Many multimedia systems have been proposed to visually explain a given topic [5] by depicting inspirational pictures [6], or enriching textual content [7]. There are some common limitations in the existing systems which have not been properly addressed:

1) Existing Text-to-Picture (TTP) systems can mine pictures automatically from the Internet and generate illustrations [5], [8], [9]. However, the obtained illustrations might not be appropriate for the required learning tasks.
2) Many pictures are available on the web, but manually finding the required pedagogic illustrations is difficult and time consuming.
3) Most of existing TTP systems can illustrate simple text based on sentence selection or keywords selection. However, these systems may not consider all important terms resulting in incorrect illustrations.

From a research-oriented perspective, multimedia instructional message elements are illustrations of material using words and pictures that are designed to promote learning.

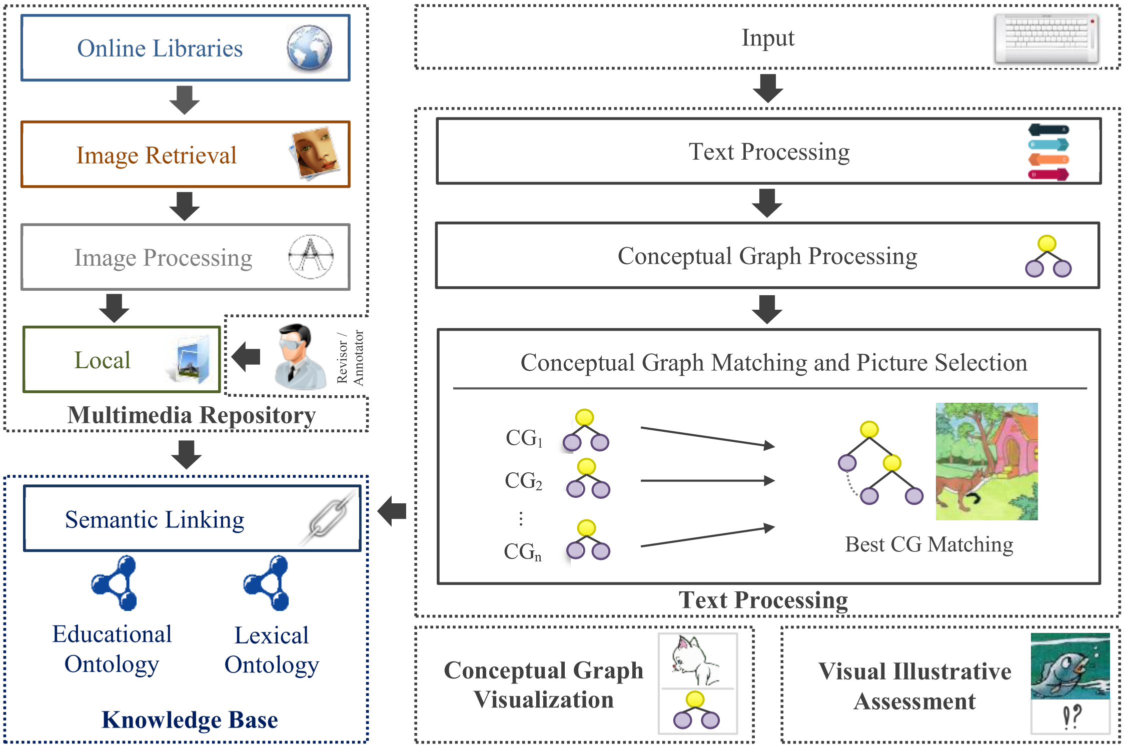

FIGURE 1. Architecture of the proposed system. It has five components: (1) multimedia repository, (2) knowledge base, (3) text processing, (4) conceptual graph visualization, and (5) visual illustrative assessment.

The pictures can be static illustrations such as images, symbols, figures, tables, charts, and maps or dynamic illustrations such as animation, or video clips. Multimedia learning occurs if one builds a mental representation of the system founded on words and pictures in the scope of multimedia instructional messages. Due to the intuitiveness and vividness of visual illustration, more effective systems need to be developed. Encouraged by the above findings, we investigate the hypothesis that automatic retrieval of pedagogic illustrations can help both instructors and students.

In the current work, we present a novel mobile educational system, which we call "Illustrate It!". It enables a user to access learning material and mine illustrative pictures for sentences or paragraphs for educational use. The system stresses the usage of conceptual graphs to facilitate the retrieval of the most relevant educational pictures. It matches the computed conceptual graph of the textual content with the graph of the picture to overcome the limitations of mining with small sentences and selected keywords. In addition to that, a multimodal visual illustrative assessment theme is also proposed. An instructor can select a picture and ask his/her students to describe the picture. The system performs automatic assessment and provides him/her an appropriate grade. The system would also assist the instructor to prepare educational materials or to illustrate them in the class. When the sentence "fox is walking near the house" is used for students, they will directly see its related illustration. In fact, the proposed multimedia system can keep users engaged for longer time [10] and also has a significant impact on their way of learning.

As shown in Figure 1, the proposed system is composed of five components: (1) multimedia repository, (2) knowledge base, (3) text processing, (4) conceptual graph visualization, and (5) visual illustrative assessment. In order to enrich the system, a multimedia repository is constructed offline. Illustrations have been collected mainly from Scribd[1] online book library. Image filtering and clustering is done to eliminate junk images and to semantically index them in the knowledge base. The knowledge base uses a developed educational ontology and a lexical ontology.

We compared the results of our system with two former approaches used in previous TTP systems including selection of the best sentence [7], and selection of the best keywords [6], [11]. Moreover, we also integrated our proposed solution with these former approaches. In conclusion, the obtained results showed significant improvement to select the most appropriate picture.

The rest of the paper is organized as follows: in Section 2 we discuss some related work. In Section 3, we give details about our system and its features. In Section 4, we present a user study. And finally in Section 5, we conclude the paper.

II. LITERATURE REVIEW

Many multimedia systems illustrating textual content by pictures have been proposed in the last decade [5], [8], [9].

[1]Scribd: http://www.scribd.com

Zhu *et al.* [5] proposed a TTP system that generates pictures for unrestricted natural text. The system combines the extraction of key-phrases, searching for the informative picture, and drawing the pictures at appropriate positions. The evaluation of the system performance was conducted in two scenarios: visual summary of news articles and illustrations of children books. Results showed that TTP synthesis has a considerable prospect in increasing the performance of human-human and human-computer communication instruments. Goldberg *et al.* [8] developed a general-purpose TTP synthesis to ameliorate the comprehension of users with special needs, such as disabled or elderly people, through examining documents and representing pictorial summaries. The developed system targeted the general English language and used machine learning techniques. The pictorial representation of natural text passes through phases: 1) extraction of keywords with picturability, 2) usage of semantic role labeling (SRL), 3) retrieving images using image search engines and ranking them using machine learning methods, and 4) optimizing the layout of the final picture according to the designed layout.

Mihalcea and Leong [12] proposed a system that generates pictorial representations for simple sentences. The analysis of score denoted interesting aspects regarding the achieved improvement of understanding. Li *et al.* [9] developed a multimedia system called Word2Image. The system uses Flickr to explore images from the web in order to provide visual explanation for the given words. Their experiments highlighted the usability of the system and the required accuracy improvement for image retrieval. Agrawal *et al.* [7] proposed techniques for retrieving images from the Internet to enrich textbooks with images while avoiding redundant images in the same chapter. Authors used a corpus of high school textbooks used in India. They used the Amazon Mechanical Turk platform to evaluate their proposed system. Ustalov [13] developed a TTP synthesis system for the Russian language. It combines a natural language processing subsystem, stage processing and a renderer. The system uses graphics gallery, thesaurus, ontology, and drawing rules as predefined information resources. Ustalov and Kudryavtsev [14] presented an ontology-based technique to TTP synthesis. It uses an RDF/XML based ontology for loose coupling of semantics, gallery, thesaurus, and depiction rules. Bui *et al.* [15] developed an application that automatically converts text to pictures, called Glyph. It is based on natural language processing and computer graphics which aims to illustrate patient instructions. It has five processing phases: pre-processing, medication annotation, post-processing, image construction, and image rendering. Aletras and Stevenson [16] proposed an approach to provide images that are useful to represent topics. The approach generates a collection of images for topics using online search engine. The selection of images is based on graph-based methods which enable the usage of both visual information and text. Jiang *et al.* [17] proposed a novel assisted instant messaging program, called Chat with Illustration (CWI). It displays illustrations linked with textual messages. The CWI searches for images in an offline database based on keywords. The final representation of the picture is constructed from a set of most representative images. Jain *et al.* [18] proposed a Hindi natural language processing called Vishit. It visualizes the text to help the communication between cultures that use different languages at university. Authors prepared an offline image repository module consisting of semantic feature tags. Semantic features serve in the selection and representation of appropriate images. Li *et al.* [6] proposed a system for children called VizStory. It transforms text to visual form of fairy pictures. It uses a web search engine to explore and finds suitable pictures to compose the visualizations. Aramini *et al.* [11] proposed an approach to automatically represent illustrations of short texts using Google image search engine. Authors proposed five methods to extract keywords from text. Methods include, extraction of information attached to the retrieved images from the web, comparison of keywords in the semantic spaces to measure their relations, re-ranking of the retrieved images, and displaying appropriate images.

Hong *et al.* [19] proposed a multimedia system called Mediapedia. It attempts to show multimedia elements such as images and videos by fetching resources from the web and assembling them for presentation. Li *et al.* [20] proposed an image search system called iSearch. It considers features of images during the search process such as local style. The system uses triangle relations chains to improve filtering consistency. Experimental evaluations showed effectiveness of the proposed system. Boato *et al.* [21] proposed a system which stresses the usage of visual saliency to get detailed information about parts of images. The re-rank of images using the saliency-based method showed improvements in image retrieval in the context of textual content.

While diverse multimedia systems have been proposed, Arabic based solutions that can automatically provide complete illustration for educational sentences and stories are still missing. In this paper, we propose a handled Arabic Ontology-based educational system that uses an educational ontology to fetch resources. The system allows teachers to address logical queries to get additional resources that enrich learning materials and enhance children skills. The system can be used to improve children educational content using illustrations, to ease the depiction of stories, and to assess students. The proposed system has demonstrated better results than existing text-to-picture systems. The usability of the system proved it can assist instructors, and can promote educational technologies in educational centers.

III. THE PROPOSED SYSTEM

In this section, we present the main components of the proposed system. First, we introduce the construction of the multimedia repository. Second, we present the content of the knowledge base. Third, we elaborate on the processing phases of the text to select appropriate images. Next, we show the conceptual graph visualization. Finally, we present a novel visual illustrative assessment scheme.

A. THE MULTIMEDIA REPOSITORY

1) THE EDUCATIONAL ILLUSTRATIONS

In order to present pedagogic illustrations, a multimedia repository is constructed to facilitate the access to the pedagogic content. Though, it is difficult to search for pedagogic illustrations on the web, we used mainly the Scribd online book library in order to fetch educational books and stories. We input various criteria in order to get intended educational contents, such as: a) the stories should target the age between 3 and 12 years, b) it must contain numerous pictures, and c) should match selected search keywords. Since Scribd library is bounded to syntactic search, we used different combinations of search keywords to get the required educational content, such as children's stories (Kesas atfal – قصص اطفال'') and child's stories (Kesas Tefl – قصص طفل''). We store these illustrations locally and mark them as ready for text extraction.

2) TEXT EXTRACTION FROM IMAGES

Images marked for processing may contain educational text. In order to extract that text, the image is first transformed to binary format. We use color threshold in order to assign a binary image pixel black (0) or white (255) values according to the average color intensity. For a pixel at index $B(i, j)$, we use the average value of red $R(i, j)$, green $G(i, j)$, and blue $B(i, j)$ to binarize that pixel. We use a threshold γ to set the value of a pixel to 0 or 255.

$$B(i, j) = \begin{cases} 0, & \text{if } \frac{R_{i,j}+G_{i,j}+B_{i,j}}{3} \leq \gamma \\ 255, & otherwise \end{cases}$$

Then we detect lines in the binary image B. A line splitter S_p is used to separate possible textual lines. It is defined with minimum number of rows $\delta = 20$ that should not have any black pixels. Each row B_r is identified as a splitter if all its columns j have no black pixels. It can be defined as follows:

$$B_r \leftarrow true \iff \{\forall B(i, r) \in B_r,\ B(i, r) = 255\}$$

For this purpose, we employ the library Tess4J[2] for optical character recognition (OCR) to transform the textual content in the image into characters. Tess4J library uses the Tesseract OCR interface [22]. A training dataset is employed to find optimal matching characters.

3) LOCAL REPOSITORY FILES

Information in the repository is stored in unstructured form as shown in Figure 2. Each story has a set of directories. Each directory can have one picture and a set of descriptive files. These files include textual content (.content), annotation (.tag), and revision file (.rev). The textual content stores the extracted text that was elaborated on in the previous section. An annotation file is written manually to precisely describe the content of the image. It is used for automatic assessment purposes. A revision file stores the revision information pertaining to files, such as revision number and author details.

[2]Tess4j OCR library: http://tess4j.sourceforge.net/

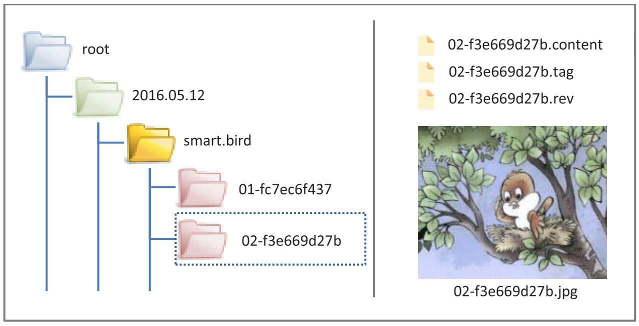

FIGURE 2. Multimedia repository sample directories.

B. THE KNOWLEDGE BASE

An ontology defines relationships between concepts and entities belonging to a particular domain. Each concept has different properties which describe all viable facts that can be associated with it. Each property can have specific settings, such as multiplicity and role restriction. An example of multiplicity restriction is: an animal cannot have more than one blood type. An example of role restriction is: an animal cannot simultaneously be a herbivore and a carnivore. An entity is an instance of a concept. For instance, the entity *lion* is an instance of the concept *carnivore*. Since ontologies supply systems with knowledge and information about a particular domain, it has become a prime technology for semantic knowledge extraction and reasoning. For example, the USDA-National Animal Genome Research Group has developed a farm animal ontology [23], which includes semantic information about cows, chicken, cattle, and other animals. As our proposed system targets Arabic education for children, we have surveyed existing educational ontologies [24] and created our own ontology on top of Arabic WordNet [25]. This is done in order to benefit from the existing knowledge base, with respect to the most commonly used domains, such as animals, foods, and sports.

We created our ontologies using Protégé [26], which is an open source tool for editing ontologies. In this application, declaring a new instance of a specific concept requires selection of the parent concept, creation of an instance, and filling values of the instance. Our knowledge base consists of two ontologies, an educational ontology and a taxonomy ontology, as described in the following sections. Figure 3 shows portions of the developed ontologies in the knowledge base.

1) THE EDUCATIONAL ONTOLOGY

The educational ontology is designed to provide educational resources that cover different domains in real life. It defines a grammatical tree structure [27], questions semantic structure [28], and stories structure. Each story contains title, content, objects, purpose, and moral. The text that we

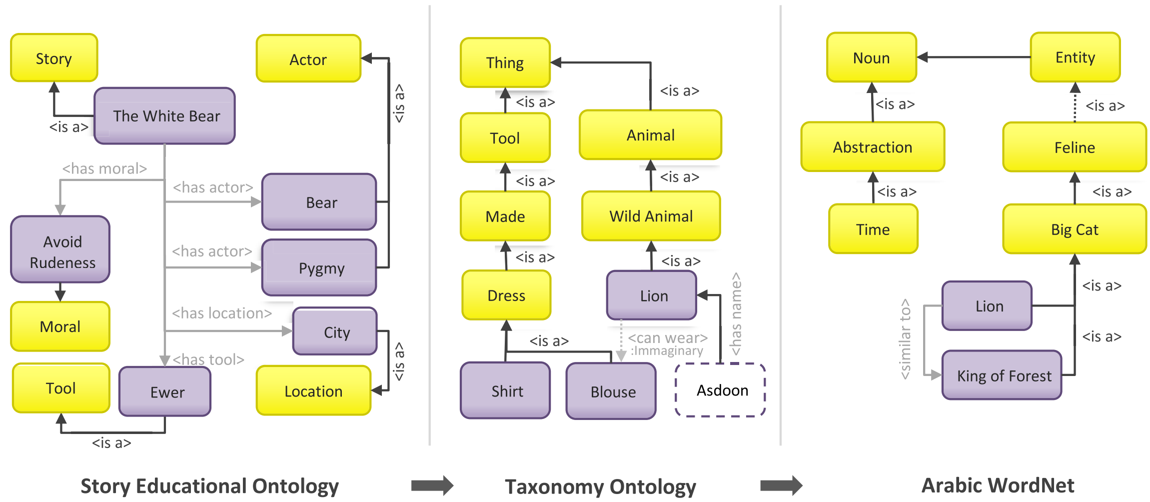

FIGURE 3. Sections for the content of the knowledge base.

extracted from images is used to fill the title and the content of the story. The object entities in a story, such as actors and tools, are linked with the taxonomy ontology and used to provide comprehensive information in the story. For instance, the Arabic text "The elephant Filo was playing with the football– الفيل فيلو كان يلعب بالطابة" can enable the user to get all semantic information about "Filo – فيلو " and "football – طابة ".

2) THE LEXICAL TAXONOMY ONTOLOGY

The taxonomy ontology is a multi-domain ontology which has been created to provide detailed structure about the relations between concepts and entities. We defined a set of concepts and associated them with their types. We considered contradictory terms through the declaration of negation terms, such as *not playing* is a negation of *playing*. We used the Arabic WordNet [25] lexical ontology to enable fetching synonymy semantic entities. We defined a set of semantic rules to associate entities instead of linking them manually. For instance, the semantic rule of "all herbivorous animals eat plants" will automatically associate all herbivorous entities with grass entities through the property "can eat", as follows:

$$\forall(x, y) : \text{herbivorous}(x), \text{grass}(y) \Rightarrow \text{can-eat}(x, y) \quad (1)$$

Consequently, if the system attempts to get information about a concept which does not exist in the knowledge base, then it will ask the user to specify it. For instance, for the sentence "Filo is playing – فيلو يلعب ", the system will ask the user to specify the type of term "Filo – فيلو ".

C. TEXT PROCESSING AND PICTURE SELECTION

1) ENTITY-RELATIONSHIPS EXTRACTION

In order to extract the relationship between entities, we employed a basic formal concept analysis approach based on the composition of entity-property matrices. It is based on the multiplication of binary matrices [29]. The entity-sentence matrix M_S is constructed according to the occurrence of the entities (e.g., lion, tiger, etc.) and their corresponding sentences. The property-sentence matrix M_P is constructed according to the occurrence of property values (e.g., brown $\rightarrow$ color, walk $\rightarrow$ behavior) and their corresponding sentences [28], [30]. The composition of the two matrices M_E will provide the contextual-relation of all entities and their properties. The matrix multiplication can be formulated as:

$$M_E = M_S \times M_P \quad (2)$$

As an example, for the Arabic sentence "the brown fox walked near the house – مشى الثعلب البني قرب البيت ", will provide the contextual relations {fox: [behavior: walk]} and {fox: [color: brown]}. Thus the variable r_{value} of the first row in the matrix contains *fox*. It has two properties (or columns). The name of the first property p_{name} is *behavior* and the second is *color*. Consequently, the value of the first property p_{value} is *walk*, and the second is *brown*.

2) CONCEPTUAL GRAPH CONSTRUCTION

The construction of the graph passes through several phases. First, the text is split into paragraphs. Each paragraph is split into sentences. Therefore, each term t of a sentence is linked with its relevant entity e defined in the knowledge base. The linkage is based on the minimum distance between the term and the name of the entity e_n of the ontology. We consider concepts and their entities, both as nodes in the obtained graph.

The Jaro-winkler string similarity [31] is used to compute the similarity between the two strings. The Jaro similarity for a term with length t and an entity name with length s computes the common characters c and the number of transpositions p.

$$S_j(t, s) = \frac{1}{3}\left(\frac{c}{t} + \frac{c}{s} + \frac{c-p}{c}\right) \quad (3)$$

Consequently, Winkler [31] enhanced the Jaro similarity in order to improve the measure. It considers the beginning matching characters of the two strings m. Based on Jaro-winkler string similarity, we define the Jaro-winkler string distance as:

$$d_{jw}(t,s) = 1 - S_j(t,s) + \frac{m}{10}(S_j(t,s) - 1.0) \tag{4}$$

A lower value of the normalized distance $d_{jw}(t,s)$ means the two strings are more similar. Thus, the value is 0 if the two strings are equal, and 1 if the two are completely different.

We consider a matching similarity between two strings if $d_{jw}(t,s) < \alpha$, where α is a threshold to cater to small differences between two strings, for example the addition of "Al" in Arabic such as *Madina* and *Al-Madina* should not make the strings different. In order to select the best matching between the entity e_n and term t, we apply the threshold α on the lowest distance $l = d_{jw}(t,s)$. We use *getEntityByName()* function to select the best matching entity which is most similar to the term t, see Algorithm 1. After identifying the required entities, we use the ontological conceptual graph G to create the conceptual graph G_E. We denote the value $e.parent$ as the parent entity of the entity e, and the function $add(arg_1, arg_2..arg_n)$ to add entities to the graph. Eventually, the obtained entity-relationship matrix M_E is used to associate entities with each other using the same approach. Algorithm 2 shows the required steps to construct graph G_E for a given text.

Algorithm 1 Select Best Match Entity $e \in G$

```
Input: Graph G, term t.
Initialization: selected ← null, min ← α.
for each entity e in G do
    if (l ← d(e_n, t)) ≤ min then
        selected ← e
        min ← l
    end
end
Output: selected
```

3) INTEGRATION WITH FORMER RETRIEVAL APPROACHES

We have integrated our proposed approach with two former retrieval approaches, the retrieval based on the selection of the best sentence, and the retrieval based on the best keywords selection. We have employed the *Term Frequency - Inversed Term Frequency* (TF-IDF) algorithm [32] in order to rank words by relevant scores. Thus, a word with a higher TF-IDF value indicates it is more important and can be used to summarize a document. For a term t and document d in corpus C, where t occurs in number n of N documents in a corpus, the TF is computed as:

$$TF(t,d) = \sum_{w \in d} \begin{cases} 1, & \text{if } w = t \\ 0, & otherwise \end{cases} \tag{5}$$

Algorithm 2 Build Conceptual Graphs of Text

```
Input: Ontology graph G, sentences S, Entity Relation
       matrix M_E.
Initialization: G_E ← empty Graph.
▷ creating hierarchical relations
for each term t in S.split() do
    selected ← getEntityByName(G, t)
    prnt ← selected.parent ▷the parent entity
    prev ← selected ▷the previous entity
    ▷ adding the hierarchical relations of entities
    while prnt is not null do
        G_E.add(prev, "is a", prnt)
        prev ← prnt
        prnt ← prnt.parent
    end
    G_E.add(prev, "is a", "root")
end
▷ linking entities between each other
for each row r in M_E.rows do
    ▷ passing over properties
    for each column p in r.cols do
        entity ← r.value
        value ← p.value
        G_E.add(entity, p.name, value)
    end
end
Output: G_E
```

and IDF is calculated as:

$$IDF(n,N) = \log(\frac{N-n}{n}) \tag{6}$$

Thus, the TDF-IDF is formulated as:

$$TFIDF(t,d,n,N) = TF(t,d) \times IDF(n,N) \tag{7}$$

A sentence with the maximum sum of weighted terms will be selected as the best sentence. The best keywords are selected according to their highest weight in a paragraph. We use the content of the collected stories as a corpus. Thus, we build conceptual graphs for the selected best sentence and for the best selected keywords.

4) CONCEPTUAL GRAPH MATCHING AND PICTURE SELECTION

The obtained graph G_E is used in order to select the best picture. The selection of the best picture I_b is based on the maximum intersection between G_E and the conceptual graphs G_p of the pictures in the multimedia repository R.

$$I_b = \{ \max_{G_p \in R, I_b \Leftrightarrow G_p} (G_E \cap G_p) \} \tag{8}$$

We have adopted the method proposed by Jaro [31] in order to compute the intersection score. The intersection between edge e_E and edge e_P is considered if all their elements are equivalent, as shown in Figure 4. The matching score $S(E,P)$ is obtained by normalizing the number of intersecting edges

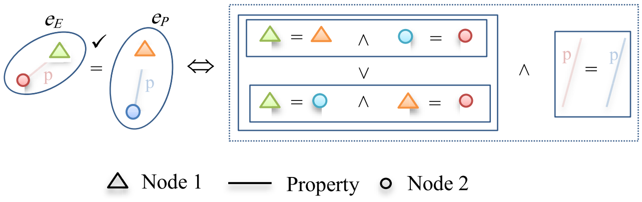

FIGURE 4. Equivalence of edges $e_E \in G_E$ and $e_P \in G_P$ of the two conceptual graphs.

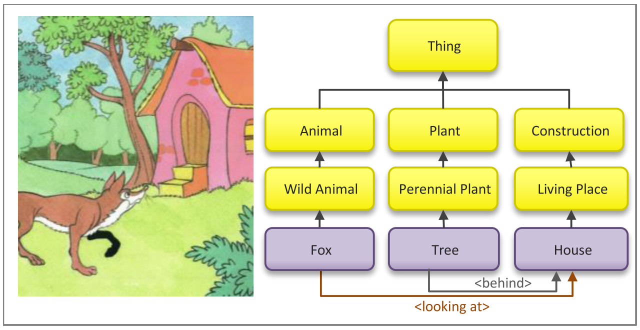

FIGURE 5. Conceptual Graph Visualization of a Picture.

with the size (total number of edges) of the two graphs, as shown in Algorithm 3. The score is equal to 1 if the two graphs are exactly similar.

Algorithm 3 Compute Graph Intersection Score

Input: Picture graph $\mathbf{G_p}$, text graph $\mathbf{G_E}$.
Initialization: $score \leftarrow 0,\ c \leftarrow 0.$
▷ *getting the intersection of the two graphs*
for *each edge* e_P *in* $G_P.edges$ **do**
  for *each edge* e_E *in* $G_E.edges$ **do**
    if $e_P = e_E$ **then**
      ▷ *counting intersecting edges*
      $c \leftarrow c + 1$
    end
  end
end
▷ *computing the score based on the intersection*
$S(E,P) \leftarrow \frac{1}{2} \times (\frac{c}{G_E.size} + \frac{c}{G_P.size})$
Output: score

D. CONCEPTUAL GRAPH VISUALIZATION AND ASSESSMENT

1) CONCEPTUAL GRAPH VISUALIZATION

Each illustrative picture is associated with semantic conceptual graphs. The conceptual graph representation is a useful tool for children to comprehend the relation between entities in the picture. The entities in the pictures are presented as nodes of the graph. The hierarchical relation is obtained from the defined taxonomy ontology. The root node is always the abstract node. Figure 5 shows an example of the conceptual graph visualization of a picture.

2) VISUAL ILLUSTRATIVE ASSESSMENT

In order to build a complete educational multimedia system, we enabled our system with three different student assessment models. First, the instructor can select a picture and ask his/her students to describe it. The description can be in form of sentences, or separate words to name entities in the picture. Second, the instructor can ask students to name entities including their hierarchical relations. Therefore, a graph G_S computed from the input of the student is compared with the picture conceptual graph G_P. A grade for a student is automatically calculated according to the maximum number of intersecting edges between the terms chosen by the student and the items in the picture. And third, the instructor can ask the student to write only a summary of a picture. For instance, in the example of {apple}, {banana}, and {peach}, the student has to name the entity {fruit}. The grade G, over total score T is computed as:

$$G = \frac{T}{2}\left(\frac{\text{count}(G_S \cap G_P)}{G_P.size} + \frac{\text{count}(G_S \cap G_P)}{G_S.size}\right), \quad (9)$$

where the intersection $(\cdot \cap \cdot)$ is defined as the number of equivalent edges in the two graphs. The equivalence of two edges is defined in Figure 4.

IV. EXPERIMENTS

The proposed system named as "Illustrate It!" has been implemented as discussed in the previous sections. The system is compared with the *Best Sentence Selection* and the *Best Keyword Selection* based systems. Also the combinations of conceptual graph approach in both methods are implemented and compared. The comparisons are performed for general retrieval, order of relevance, immediate illustration, and mining performance in terms of execution time. In addition to that, usability assessments and learning improvement comparisons in children with special needs is also performed. The performance evaluation of the proposed system have depicted significant improvements in retrieving images illustrating the text.

A. GENERAL RETRIEVAL COMPARISONS

In order to conduct evaluations, we compared five different retrieval methods. These methods include:

1) The proposed Conceptual Graph (CG): The conceptual graph is constructed for the entered text as described in previous sections.
2) Best Sentence Selection (BSS): The text is split into sentences. The sentence with the highest weight is selected. The Jaro-winkler [31] string similarity is used during the retrieval process.
3) Best Sentence - Conceptual Graph (BS-CG): The best sentence is selected as described in the second method. A conceptual graph is constructed for the selected best sentence and used during the retrieval process.
4) Best Keywords Selection (BKS): Keywords of the entered text are weighted. Top five keywords are selected and used for the retrieval of illustrations.

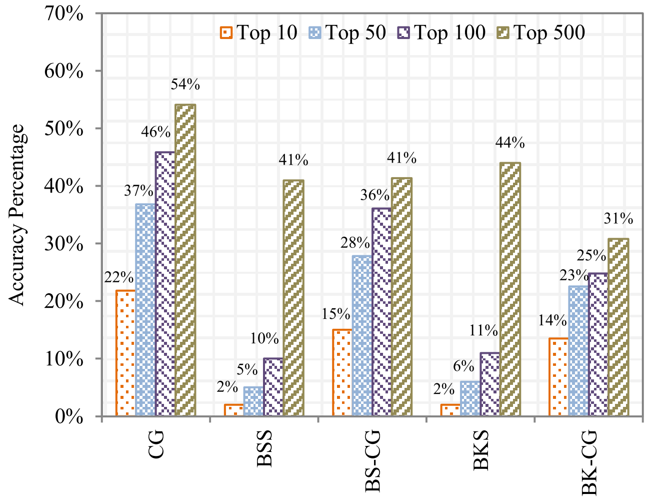

FIGURE 6. Comparison for the general retrieval between retrieval methods.

5) Best keywords - Conceptual Graph (BK-CG): A conceptual graph is constructed for the selected best keywords. The conceptual graph is employed to retrieve the relevant pictures.

In some cases, the same picture can be used appropriately to describe several paragraphs. In order to clarify the evaluation process, we exclude such pictures from the test dataset. This is to make sure the evaluation is considering only the intended picture that is associated with its accompanied paragraph. In addition, a retrieval failure is counted if the index of the required image is greater than the specified rank β. We used five β values to assess different retrieval methods, $\beta = 500$ or top five hundred images, $\beta = 100$ or top hundred images, $\beta = 50$ or top fifty images, $\beta = 10$ or top ten images, and $\beta = 1$ for immediate retrieval.

Figure 6 shows the results of the general retrieval experiment for the different methods. In this study, the order of images is not considered. Only retrieved images are considered. The proposed method, CG, achieved the highest score compared to the other methods. Moreover, the retrieval using BSS and BKS improved significantly when the CG is integrated with these methods. This shows the effectiveness of the proposed CG approach.

B. ORDER OF RELEVANCE COMPARISONS

In the experiment depicted in Figure 7, the impact of different methods is considered on sorting the required images in the retrieval process. The retrieval score S_m is used to rank the retrieval methods. Thus, if a retrieval method sorted the required image with an index i lower than all other indices i_o of other retrieval methods, then, the score S_m will be incremented by 1. If more than one indices i_s of retrieval methods are equal and lower than other indices i_o of other retrieval methods, then the scores of all retrieval methods S_k with low indices will be incremented by 1. The order of relevance

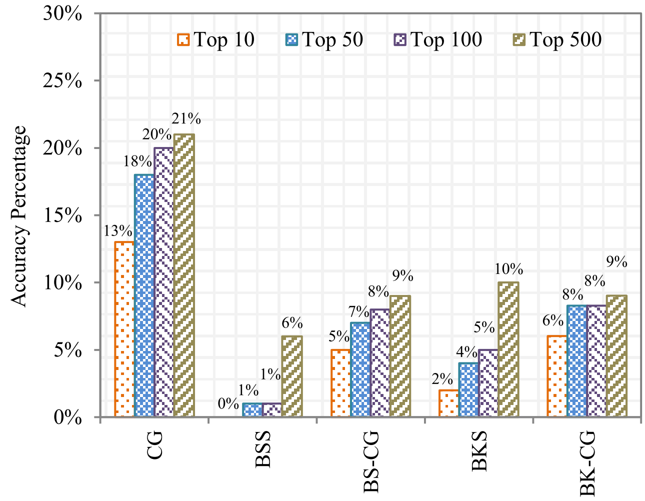

FIGURE 7. Order of relevance comparison between retrieval methods.

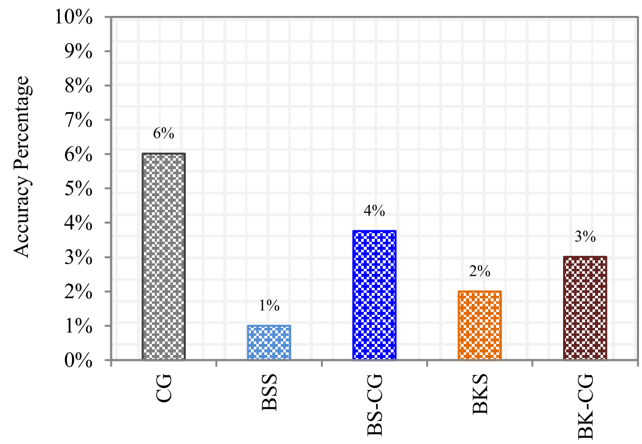

FIGURE 8. Immediate illustration comparison between retrieval methods.

conditions can be summarized as follows:

$$\begin{cases} S_m \leftarrow S_m + 1, & \text{if } (i < i_o) < \beta \\ \forall (S_i \in S_k): \ S_i \leftarrow S_i + 1, & \text{if } (i_s < i_o) < \beta \\ skip, & otherwise \end{cases}$$

The retrieval methods using CG got the highest percentage in comparison with other retrieval methods. Moreover, our proposed approach has shown significant improvement when integrated with previous methods. However, it decreased from 22% to 13% when the order of relevance is considered in the selection of the top ten images. This can also be seen for the retrieval with best sentences integrated with conceptual graph, and best keywords integrated with conceptual graph, where the percentage decreased from 15% to 5% and from 14% to 6% respectively.

C. IMMEDIATE ILLUSTRATION COMPARISONS

In Figure 8, only the desired illustration that should be selected for the paragraph is considered. Our proposed approach showed better results in general when compared with other approaches.

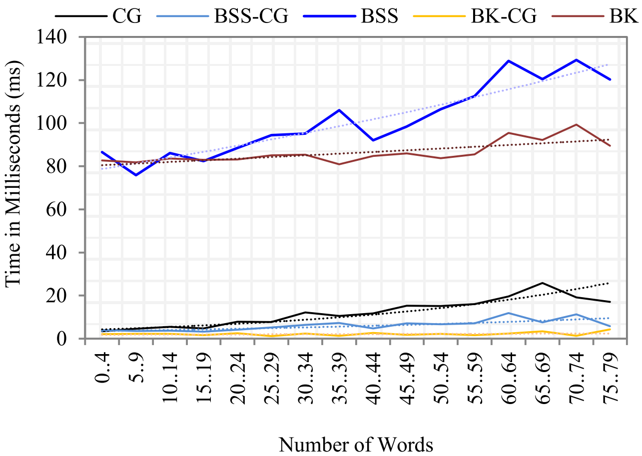

FIGURE 9. Mining performance of the different retrieval methods with number of words representation.

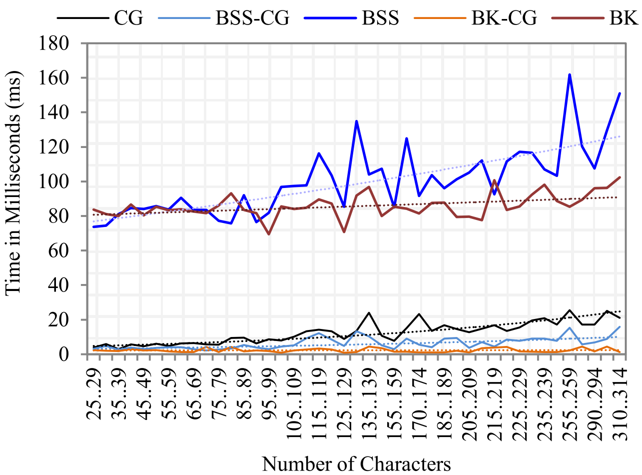

FIGURE 10. Mining performance of the different retrieval methods with number of characters representation.

D. MINING PERFORMANCE COMPARISON

In Figure 9 and Figure 10, the mining performance of the different retrieval methods is demonstrated. Results depict that processing performance is correlated with the number of words or number of characters. Thus, a text with a higher number of words or characters requires more processing time. The proposed CG method achieved better performance in comparison with other methods. Moreover, the performance has been significantly improved after the integration of the proposed approach with existing retrieval methods. The selection using BSS required the highest computational processing, but it did not give the most optimal results. The time complexity of the developed CG algorithms is $O(n^2)$ where n is the number of times statements will be executed. The number of nodes (concepts and entities) in the knowledge base notably affects the matching between terms and their nodes in the graph. Based on the experiments conducted, we are using a total number of 90285 nodes with a hierarchical depth of 17 in Arabic WordNet; along with 1000 nodes and a hierarchical depth of 7 in our Taxonomy ontology. Therefore, in order to reduce the graph construction complexity, we are mapping matched terms with their concepts in hash tables. Thus, the time complexity of graph construction becomes $O(n\ log(n))$. Figure 11 summarizes the overall average of the processing performance. Mining using CG achieved the most optimal performance in comparison with other tested retrieval methods. In addition, the performance improved significantly after the integration of BSS and BKS with CG. This is because during the process of matching terms with the BSS and BKS methods, the distance of all possible terms must be examined with the provided text. However, the process of matching two graphs requires checking only if edges are equal.

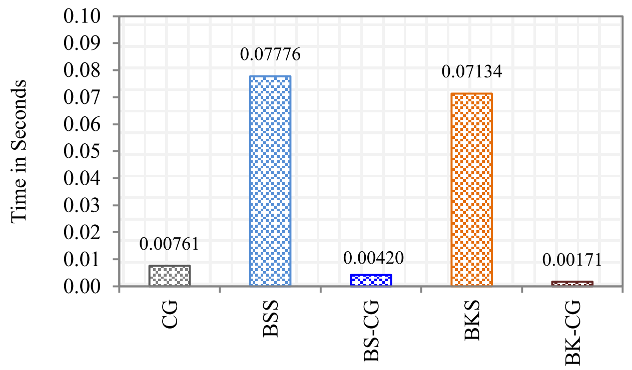

FIGURE 11. Average execution time of the various retrieval methods over 1000 experiments.

E. USABILITY ASSESSMENT COMPARISON

We have interviewed fifteen instructors teaching in different places including five elementary school instructors, five instructors for children with special needs, and five private-home teachers. We assumed that instructors that teach at home and at school, are only instructors at school. This is to distinguish the obtained results of school instructors from home instructors. We asked each instructor to fill a survey after using the system. The answer in the survey considers five points rating scale; strongly agree, agree, undecided, disagree, and strongly disagree. Table 1 shows the obtained average score of all instructors.

Moreover, we asked them to select the appropriate place where the application can be used. We asked them to choose between "Schools", "Centers for people with special needs", "Home", and "All the above". Figure 12 shows the obtained results.

Eventually, private tutors were asked to use the system with their students. A total number of 25 students are asked to describe the content of the pictures. Therefore, after reading 10 stories, and seeing illustrated pictures with associated CGs, students used more words in their descriptions. Figure 13 demonstrates the obtained results before and after using the system. Before using the system, the average grade of 25 students was 6.24 ± 0.597, while after using the system the grade changed to 8.48 ± 0.770. The corresponding

TABLE 1. Usability analysis.

Usability Statement	Average
The system is useful for education	4.47
The system was easy to use	3.93
You fairly got pictures that describe your entered text	2.87
Obtained Illustrations were clear	3.93
The conceptual graph visualization feature helped students to memorize more words	4.13
The conceptual graph feature helped students to understand more the relations between words	3.87
The conceptual graph feature assisted students to increase their ability while describing pictures	4.27
The system raised the motivation of your students	4.07
The automatic visual assessment was useful	4.73
The automatic visual assessment gave the right score for students	3.47

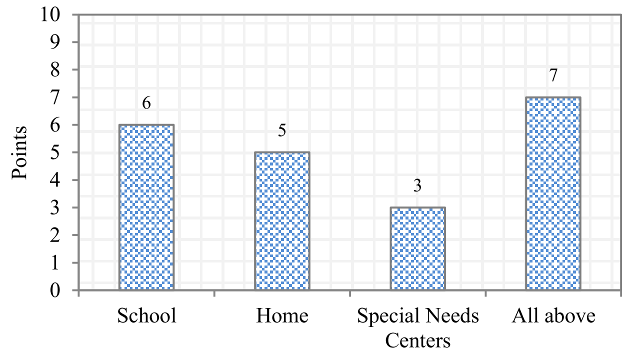

FIGURE 12. Appropriate 'place of use' assessment of the proposed system by 15 instructors.

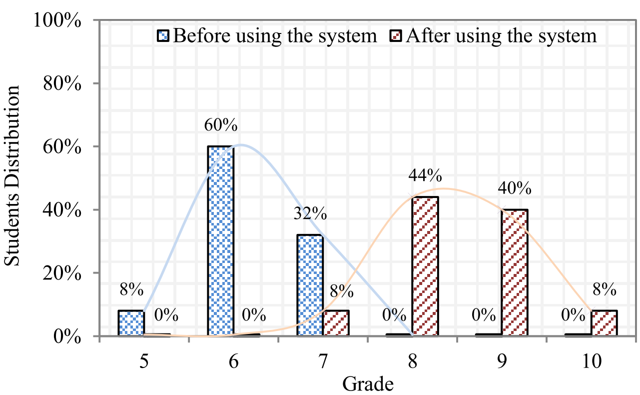

FIGURE 13. Improvement in student scores after using the proposed system. The overall scores of 25 students in the experiment have shifted towards higher side.

p-value is $p = 0.00112$, that shows the hypothesis that the proposed system will improve learning may be incorrect in only 1 in 893 cases. That shows the results of the experiment are statistically significant.

F. ASSESSMENT WITH CHILDREN WITH SPECIAL NEEDS

In this section, we have employed the proposed system to assess its outcome on children with intellectual disabilities (ID) and children with down syndrome (DS). The system has been used by 100 children from Shafallah Center in Doha.

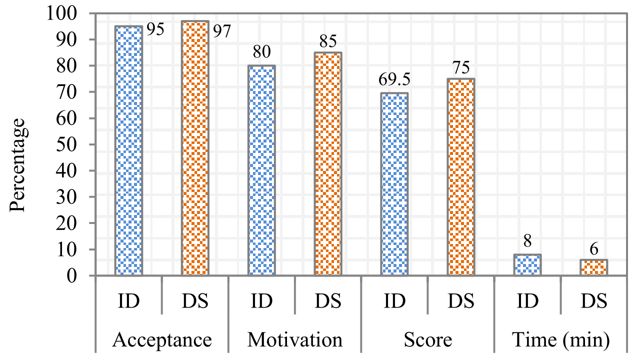

FIGURE 14. Performance of children with ID vs DS.

These children have an average mental age of 8 years, and mildly disabled. Half of these children are children with (DS) and the others are children with intellectual disabilities (ID). Figure 14 shows the obtained results in terms of acceptance, motivation, scores, and average time to solve exercises. Five of children with ID and three children with DS refused to use the system. Children with DS obtained a higher performance score by a margin of 10% over children with DS. In addition, children with IDs took longer time in order to complete the exercises. Both groups had high motivation levels, although children who scored higher (DS) are more probably to have higher motivation.

V. CONCLUSION

A multimedia mobile text-to-picture educational system, called "Illustrate It!" is presented in this study. The proposed system solves the limitations of existing systems by providing pedagogic illustrations, considering all concepts including their semantic relations. It is based on conceptual graph matching in order to select the best illustrative picture. Using a knowledge base, a conceptual graph is built from the text accompanied with the pictures in the multimedia repository and the text entered by the user. Evaluation tests demonstrate that the proposed system has significantly outperformed the best keyword selection and best sentence selection based systems. In addition to that, the system performs automatic evaluation of user learning by matching the semantic graph accompanied with the picture in the multimedia repository and the semantic graph of students' input. Experiments verified that our proposed system significantly improved the learning capabilities of students. The current system cannot automatically find illustrations that do not have annotations or textual content, and cannot locate annotated elements in the picture. Future research in this direction may be to employ computer vision techniques to automatically identify objects in a picture and to introduce diverse multimedia content, such as 3D virtual reality scenes, videos, and audio.

REFERENCES

[1] R. E. Mayer and R. Moreno, "Aids to computer-based multimedia learning," *Learn. Instruct.*, vol. 12, no. 1, pp. 107–119, 2002.

[2] R. N. Carney and J. R. Levin, "Pictorial illustrations still improve students' learning from text," *Edu. Psychol. Rev.*, vol. 14, no. 1, pp. 5–26, 2002.

[3] W. H. Levie and R. Lentz, "Effects of text illustrations: A review of research," *Proc. ECTJ*, vol. 30, no. 4, pp. 195–232, 1982.
[4] R. E. Mayer, "Multimedia learning," *Psychol. Learn. Motivat.*, vol. 41, pp. 85–139, Dec. 2002.
[5] X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock, "A text-to-picture synthesis system for augmenting communication," in *Proc. AAAI*, vol. 7. 2007, pp. 1590–1595.
[6] C.-T. Li, C.-J. Huang, and M.-K. Shan, "Automatic generation of visual story for fairy tales with digital narrative," *Web Intell.*, vol. 13, no. 2, pp. 115–122, 2015.
[7] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, "Enriching textbooks with images," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 1847–1856.
[8] A. B. Goldberg, J. Rosin, X. Zhu, and C. R. Dyer, "Toward text-to-picture synthesis," in *Proc. NIPS Mini-Symp. Assistive Mach. Learn. People Disabilities*, 2009, pp. 1–3.
[9] H. Li, J. Tang, G. Li, and T.-S. Chua, "Word2image: Towards visual interpreting of words," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 813–816.
[10] I. Cheng, A. Basu, and R. Goebel, "Interactive multimedia for adaptive online education," *IEEE Multimedia Mag.*, vol. 16, no. 1, pp. 16–25, Jan. 2009.
[11] S. A. Aramini, E. Ardizzone, and G. Mazzola, "Automatic illustration of short texts via Web images," in *Proc. IVAPP*, 2015, pp. 139–148.
[12] R. Mihalcea and C. W. Leong, "Toward communicating simple sentences using pictorial representations," *Mach. Transl.*, vol. 22, no. 3, pp. 153–173, 2008.
[13] D. Ustalov, "A text-to-picture system for russian language," in *Proc. 6th Russian Young Sci. Conf. Inf. Retr.*, 2012, pp. 35–44.
[14] D. Ustalov and R. Kudryavtsev, "An ontology based approach to text to picture synthesis systems," in *Proc. 2nd Int. Workshop Concept Discovery Un-struct. Data*, 2012, pp. 94–101.
[15] D. D. A. Bui, C. Nakamura, B. E. Bray, and Q. Zeng-Treitler, "Automated illustration of patients instructions," in *Proc. AMIA*, 2012, p. 1158.
[16] N. Aletras and M. Stevenson, "Representing topics using images," in *Proc. HLT-NAACL*, 2013, pp. 158–167.
[17] Y. Jiang, J. Liu, and H. Lu, "Chat with illustration," *Multimedia Syst.*, vol. 22, no. 1, pp. 5–16, 2016.
[18] P. Jain, H. Darbari, and V. C. Bhavsar, "Vishit: A visualizer for hindi text," in *Proc. 4th Int. Conf. Commun. Syst. Netw. Technol.*, Jun. 2014, pp. 886–890.
[19] R. Hong, Z.-J. Zha, Y. Gao, T.-S. Chua, and X. Wu, "Multimedia encyclopedia construction by mining web knowledge," *Signal Process.*, vol. 93, no. 8, pp. 2361–2368, 2013.
[20] H. Li, X. Wang, J. Tang, and C. Zhao, "Combining global and local matching of multiple features for precise item image retrieval," *Multimedia Syst.*, vol. 19, no. 1, pp. 37–49, 2013.
[21] G. Boato, D.-T. Dang-Nguyen, O. Muratov, N. Alajlan, and F. G. Natale, "Exploiting visual saliency for increasing diversity of image retrieval results," *Multimedia Tools Appl.*, vol. 75, no. 10, pp. 5581–5602, 2016.
[22] M. A. Hasnat, M. R. Chowdhury, and M. Khan, "An open source tesseract based optical character recognizer for bangla script," in *Proc. 10th Int. Conf. Document Anal. Recogn.*, 2009, pp. 671–675.
[23] L. Hughes, J. Bao, Z.-L. Hu, V. Honavar, and J. M. Reecy, "Animal trait ontology: The importance and usefulness of a unified trait vocabulary for animal species," *J. Animal Sci.*, vol. 86, no. 6, pp. 1485–1491, 2008.
[24] A. G. Karkar, J. M. Al Ja'am, and S. Foufou, "A survey on educational ontologies and their development cycle," in *Proc. IFIP Int. Conf. Product Lifecycle Manage.*, 2015, pp. 649–658.
[25] A. Abbache, F. Meziane, G. Belalem, and F. Z. Belkredim, "Arabic query expansion using wordnet and association rules," *Int. J. Intell. Inf. Technol.*, vol. 3, no. 12, pp. 51–64, 2016.
[26] M. A. Musen, "Protégé ontology editor," in *Encyclopedia of Systems Biology*. Berlin, Germany: Springer, 2013, pp. 1763–1765.
[27] A. G. Karkar and J. M. Al Ja'am, "A native arabic elearning mobile application to transpose arabic text to illustrations," in *Proc. Int. Conf. Inf. Commun. Technol. Res.*, May 2015, pp. 72–75.
[28] A. G. Karkar and A. Ja'am, and J. Mohamad, "An educational ontology-based m-learning system," *Int. J. Interact. Mobile Technol.*, vol. 10, no. 4, pp. 48–56, 2016.
[29] A. C. Erdal, *Binary Tree Structures for Matrix Multiplication*. East Lansing, MI, USA: Michigan State Univ., 1983.
[30] A. G. Karkar, A. Dandashi, and J. M. Al Ja'am, "A text-to-picture m-learning system in a private wireless mesh network," in *Proc. 19th Conf. Innov. Clouds, Internet Netw.*, 2016, pp. 212–218.
[31] W. Cohen, P. Ravikumar, and S. Fienberg, "A comparison of string metrics for matching names and records," in *Proc. Kdd Workshop Data Cleaning Object Consolidation*, vol. 3. 2003, pp. 73–78.
[32] S. Robertson, "understanding inverse document frequency: On theoretical arguments for IDF," *J. Document.*, vol. 60, no. 5, pp. 503–520, 2004.

ABDEL GHANI KARKAR received the master's degree from the University of Balamand, Lebanon, in 2010. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Qatar University. His Ph.D. thesis was on An Ontology Based Text-To-Picture Multimedia m-Learning System. His research interests include multimedia systems, m-learning technologies, semantic Web, software engineering, artificial intelligence, and distributed systems. Dr. Karkar managed as a Committee Member for several conferences and chaired many sessions.

JIHAD MOHAMAD ALJA'AM received the B.Sc., M.S., and Ph.D. degrees in computing from Southern University (The National Council for Scientific Research, CNRS), France. He is currently with the Department of Computer Science and Engineering, Qatar University. He is also leading a research team in multimedia and assistive technology and collaborating with the Financial Watch and Intelligent Document Management System for Automatic Writer Identification projects. His current research interests include multimedia, assistive technology, learning systems, human–computer interaction, stochastic algorithms, artificial intelligence, information retrieval, and natural language processing.

ARIF MAHMOOD received the master's and Ph.D. degrees (Hons.) in computer science from the Lahore University of Management Sciences in 2003 and 2011, respectively. He then joined The University of Western Australia as a Research Assistant Professor for three years. He is currently a Postdoctoral Researcher with the Department of Computer Science and Engineering, Qatar University. He has been involved in face and object recognition, action recognition, foreground–background detection, emotion mapping and synthesis, image segmentation and composition, community detection in complex networks, and computing superpixels in images and videos.