IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED SOFTWARE AND DATA ENGINEERING FOR SECURE SOCIETIES

Received January 12, 2019, accepted January 25, 2019, date of publication February 11, 2019, date of current version February 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896713

Improving Arabic Text to Image Mapping Using a Robust Machine Learning Technique

JEZIA ZAKRAOUI[1], SAMIR ELLOUMI[2], JIHAD MOHAMAD ALJA'AM[1], AND SADOK BEN YAHIA[3]

[1]Computer Science and Engineering Department, Qatar University, Qatar
[2]Faculty of Sciences of Tunis, University of Tunis El Manar, LR11ES14, Tunis 2092, Tunisia
[3]Department of Software Sciences, Tallinn University of Technology, 12618 Tallinn, Estonia

Corresponding author: Jihad Mohamad Alja'am (jaam@qu.edu.qa)

This work was made possible by NPRP Grant #10-0205170346 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

ABSTRACT In this paper, we introduce an approach to automatically convert simple modern standard Arabic children's stories to the best representative images that can efficiently illustrate the meaning of words. It is a kind of imitating the imaginative process when children read a story, yet a great challenge for a machine to achieve it. For simplification issues, we apply several techniques to find the images and we associate them with related words dynamically. First, we apply natural language processing techniques to analyze the text in stories and we extract keywords of all characters and events in each sentence. Second, we apply an image captioning process through a pre-trained deep learning model for all retrieved images from our multimedia database as well as the Google search engine. Third, using sentence similarities, most significant images are retrieved back by selecting top-k highest similarity values. It is worth mentioning that using the captioning process, to rank top-k images, has shown reasonable precision values as per our preliminary results. The option to refine or validate the ranked images to compose the final visualization for each story is also provided to ensure a flexible and safe learning environment.

INDEX TERMS Robust machine learning, automated Arabic text illustration, mapping text to multimedia, visualization, deep learning model.

I. INTRODUCTION

A text-to-picture system is a system that automatically converts a natural language text into pictures representing the meaning of that text. The conversion of a general text to its visual representation requires a dynamic mapping process, which is an important step in many computer vision applications such as story picturing [1], natural language visualization [2], etc. Indeed, there are many other application areas for text-to-picture systems such as summarization of news articles [3], data visualization, games, visual chat [4], and learning for children with learning difficulties, to name but a few. We aim in this work to convert Arabic children's stories to visual static or dynamic representations. However, the transformation from one representation to another needs many requirements and challenges that have to be analyzed and investigated. In particular, Arabic language, unlike the English one, has complex morphological aspects and lacks both linguistic and semantic resources [5], yet another challenge to be addressed accordingly.

There are main challenges on mapping natural text to multimedia in general. First, according to Hassani [6] difficulties in mapping text to multimedia root in characteristics of natural languages such as being semi-structured, ambiguous, context-sensitive and subjective besides the technical issues. So, such mapping requires at first tools for text processing and text analysis in order to understand the semantics behind it and then proceed with fetching appropriate image resources.

Second, as highlighted in [7], the association between images and texts in multimedia-rich content can hardly be established using traditional methods since alone the scale of the text can cover the entire natural language vocabulary. Therefore, there is a need for more powerful methods and techniques. Coelho and Ribeiro [8] argue to tackle an increased difficulty in managing large multimedia sources in order to explore, retrieve, filter and rank relevant information

The associate editor coordinating the review of this manuscript and approving it for publication was Mahmoud Barhamgi.

in particular images. Indeed, associated textual information to images, such as image tags, is noisy and insufficient to describe the rich content of images comprehensively and substantially as admitted in [9]. Finally, returned images from Google search engine (referring to JSON response object) lacks appropriate captions or at least meaningful tags about the images. All obtained images cannot be considered in our approach due to missing captions. To overcome this limitation, we propose to use deep learning captioning model to complete the missing information.

With regard to the Arabic language there exist additional challenges; the first one to tackle is multimedia search using Arabic keywords. For English, there are enough annotated image collections used for information retrieval. However, they are rarely translated into Arabic, and consequently not directly reusable for processing Arabic.

The second challenge is related to open multimedia resources; the lack of open image resources from the Arab research community makes the problem of sourcing multimedia difficult. High-quality pictures associated with well-maintained metadata and tags, freely usable and containing a diverse set of concepts are the keys for achieving our initiated goal.

Considering this situation, we simplified the problem by restricting our image search to Google image search. The latter has been shown through literature to often produce appropriate images after several filtering steps. We also applied an automatic captioning process for all retrieved images using English captions then translated these into Arabic. Thus, we obtained different versions which we successively evaluate in the remainder. To do, we propose the following:

1) To use Machine Translation (MT) to translate Arabic text to English text to overcome the problem with the lack of image resources annotated into Arabic.
2) To use a convolutional neural network (CNN) as a pre-trained model to automate the captioning process for all images to be included in our approach.
3) To investigate semantic aspects of text matching

TABLE 1. Description of four cases used for comparative study.

Cases	Descriptions
AW	Arabic Words without captions
AWC	Arabic Words with captions
EW	English Words without captions
EWC	English Words with captions

Therefore, we propose to investigate four cases depending on using MT and image captioning, as presented in Table 1.

- The AW case: In this case, we use Arabic keywords to retrieve relevant images. Retrieved images do not have captions as per return from Google image search. The selection and ranking of top-k images are handled through an image scoring evaluation *getUserEval* function. The pseudocode of the latter is given by Algorithm 2.
- The AWC case: As in AW case, we use Arabic keywords to retrieve relevant images. All retrieved images are piped through a captioning process based on CNN. An image subset consisting of images with captions is returned for further processing. These captions are translated into Arabic using MT tool. So, we obtain Arabic captions. The selection and ranking of top-k images are handled by a captioning function *getCaptionByDeepLearningModel*, whose pseudo-code is given by Algorithm 1.
- The EW case: In this case, we use English keywords to retrieve relevant images. Retrieved images do not have captions as per return from Google image search. The selection and ranking of top-k images are handled by an image scoring evaluation function *getUserEval*, see Algorithm 2.
- The EWC case: In this case, we use English keywords to retrieve relevant images. All retrieved images are piped through a captioning process based on CNN. An image subset consisting of images with captions is returned for further processing. The selection and ranking of top-k images is handled by the getCaptionByDeepLearningModel function.

Subsequently, our approach prepares a set of candidate images for each story by querying a local image database or Google image search engine with all relevant keywords. The top-k images are generated based on a deep learning model that in turn generates an English caption each time a new image is downloaded. The evaluation is done using a test data set created for this study that was annotated automatically by the mentioned captioning process. Results of the evaluation show that the proposed method is promising when considering some improvements.

The contributions of this paper are as follows: (i) describes an automatic mapping of Arabic text to images: (ii) makes use of deep learning model to include eventually all relevant images that do not have captions; (iii) to evaluate generated images and the proposed approach as a whole.

The remainder of this paper is organized as follows: Section 2 scrutinizes state-of-the-art of the text-to-picture approaches. Section 3 presents in deep our approach. Especially, it describes the general architecture and details the different algorithms. Section 4 discusses the results and the evaluations. Finally, we sketch future work avenues.

II. STATE OF THE ART

Mapping general text to pictures has been a major subject for many approaches and systems. In particular, text-to-picture systems have been developed to date to achieve this task. For instance, story picturing [1] that attempts to find representative pictures for a fragment of text performs text illustrations by using Wordnet [10], an annotated picture database, as well as a mutual reinforcement-based ranking algorithm.

Some text-to-picture systems are viewed as a translation approach from a text language to a visual language [11] with excessive manual efforts. For instance,

Mihalcea and Chee [12] find images for dictionary words as a kind of visual linguistic representations of machine translation using an in-house image database, PicNet, and other resources. Worthy of mention, other text-to-picture systems are being seen as image retrieval and ranking problem [2]. For instance, Agrawal *et al*. [13] presented techniques for finding images from the Web that are most relevant for augmenting a section of the textbook under predefined constraints.

Whereas some text-to-picture systems rely on many filtering algorithms and techniques in order to get appropriate materials from Web image searches, other systems create their own multimedia datasets, which has revealed the excessive manual efforts behind these systems. For example, WordsEye [14] is an interesting system for automatically converting text into representative 3D scenes, but it relies on its huge offline rule-base and data repositories containing different geometric shapes and types, which have been annotated manually. This reveals its lacking for an automatic annotation task.

A worth mentioning text-to-picture system for general, unrestricted texts by defining a picturability measure for words is proposed by Zhu *et al*. [15]. This system evolved and used semantic role labeling for its latest version. A TextRank summarization algorithm [16] is applied to compute probabilities, and the top 20 keywords are selected and used to build the key phrases, each having an assigned importance score.

A promising system using a domain ontology is proposed by Dmitry and Aleksandr [17] and designed for Russian language processing. It operates with natural language analysis component, a stage processing component, and a rendering component. The system evolved from its previous version to convey the gist of general, semantically unrestricted Russian language text. Huang *et al*. [18] proposed VizStory as a visualization of fairy tales by transforming the input texts to suitable pictures while also considering the narrative structures and the semantic contents of stories. In this work, keywords are selected from segments in the stories, relevant pictures are searched from online repositories based on their tags, and finally, the pictures are composed for showing the main ideas of the original segments.

Storytelling systems have also been proposed in many works [19]–[24], [25]. A recent multimedia system for Arabic stories based on conceptual graph matching, is proposed in [26]. Worth mentioning approaches [3], [27], [28] in the domain of news streaming have been proposed that are useful to represent emotions and breaking news. More recently, a medical record summary system was recently developed by Ruan *et al*. [29]. The latter enables users to briefly acquire the patient's medical data which are visualized spatially and temporarily based on the categorization of multiple classes consisting of event categories and 6 physiological systems.

Table 2 glances an overall comparison focusing on syntax analysis, semantic analysis and input/output modalities of the functional text-to-picture systems.

TABLE 2. Comparison of functional text-to-picture systems focusing on NLP, NLU and IO-modalities.

System	Syntax Analysis (NLP)	Semantic Analysis (NLU)	Input Modality	Output Modality
Story picturing engine [1]	Bag-of-words	semantic association	Paragraph	2D picture
Text-to-picture synthesizer [15]	POS tagging	association, semantic role labeling	Typed text	2D picture collage
Utkus [30]	Chunking (verb, noun phrases)	dependency parsing	Text	Icons
WordsEye [14]	Statistical parsing	dependency parsing	Story	3D pictures
Vishit [31]	POS tagging	dependency parsing	Text	Scene

As Table 2 indicates, some systems follow shallow semantic analysis such as semantic role labeling, whereas other ones rely on deep semantic analysis or linguistic approaches that investigate deeper semantic parsing such as dependency parsing.

Although there are many real working text-to-picture systems that automatically map a given sentence to images systems for Arabic text are very limited which reflects the current technical difficulties in understanding Arabic natural language. Yet, none of them consider the mapping process using an automatic captioning based on deep learning model to annotate retrieved images with English and Arabic sentences, which is what we target at and exploit for presenting the Arabic story through suitable pictures retrieved from Google search engine.

III. PROPOSED APPROACH

Considering the fact that we are dealing with simple stories, we propose to use keyword-based image search from our local database and eventually from Google image safe search to retrieve educational multimedia representative for simple stories in the domain of animals. The current version of our proposed system is built on multiple open resources to enable faster advancement by exploiting larger community contributions.

A. GENERAL ARCHITECTURE

In this section, we describe thoroughly introduced approach for mapping Arabic simple stories to images. This approach is split into two main parts, represented in Figure 1:

(1) Story text processing and image retrieval using keyword-based search, containing the steps 1, 2 and 3.

(2) Image ranking using automatic captioning process and sentence similarity, containing the steps 4, 5, and 6.

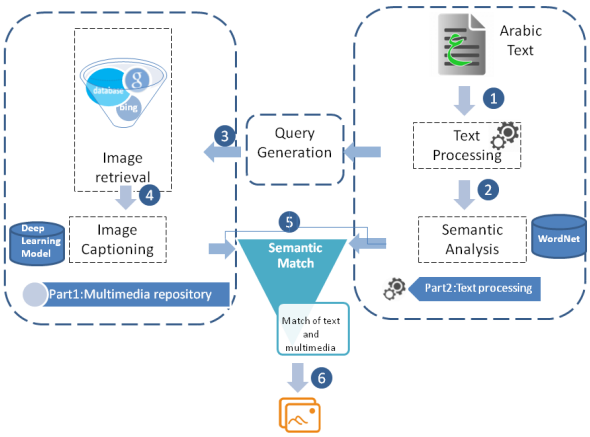


FIGURE 1. The architecture of the proposed system at a glance.

To achieve this functionality, we performed the following main tasks:

1. Collecting Arabic simple stories in the domain of animals;
2. Processing Arabic text using different NLP tools for Arabic;
3. Generating SQL queries and search engine queries to retrieve images and creating a database to store the mapping of keywords to images to serve as preliminary image pool;
4. Captioning retrieved images using a Convolutional Neural Network (CNN) pre-trained model;
5. Matching semantic aspects by retrieving a set of images with initial input text;
6. Validating and ranking of the retrieved images using sentence similarity.

To perform those tasks, we used several open source tools such as MT tool [32], CNN/LSTM image captioning model [33], etc. In the following, we present in detail the different steps.

B. DETAILED EXPLANATION OF THE ARCHITECTURE

This section describes in detail our proposed approach.

1) KEYWORD EXTRACTION (STEP 2)

First, story text is split into sentences and some preprocessing steps are made such as segmentation, stopword removal, and part-of-speech tagging, etc. Second, we select relevant single word tokens as keywords for each sentence and we translate them to English using this MT translation tool [32], [34]. We organize the keywords in the order that they appear in the text to preserve their role within the text.

2) QUERY FORMULATION (STEP 3)

In this step, we formulate keyword-based queries to retrieve candidate images for them. A useful heuristic for finding better representative images for the characters and the events in search engines is to concatenate the extracted keywords including verbs as a single query for each sentence. A standard method of multiple queries is also employed. The combination of single query and multiple queries is also employed.

3) IMAGE SELECTION (STEP 3)

The retrieved images using the concatenation of keywords are downloaded, saved locally and thus prepared for further pre-filtering and captioning process in the next step.

4) IMAGE CAPTIONING (STEP 4)

The prepared set of images is ready for going through this step in order to be captioned by a deep learning model. Image captioning via deep CNN, recurrent neural network RNN and long short-term memory LSTM have witnessed significant improvements in recent years [35]. Deep CNNs can fully represent an image by embedding it into a fixed-length vector. Then, RNN, especially LSTM [36], decodes the fixed-length vector to a desired output sentence by iterating a recurrence relation [37]. We used a pre-trained model as a fine-tuned checkpoint which has been trained over 3 million iterations using the MSCOCO dataset [38].

5) SENTENCE SIMILARITY (STEP 5)

The sentence similarity is requested in order to make the matching between initial keywords and captions. Depending on the considered version, cf. Table 1, we apply sentence similarity after the MT process to guarantee basic textual information for the matching process.

The obtained English captions are then MT processed into Arabic. The resulting Arabic captions are compared with keywords to find out which images are kept for final representation. We use sentence similarity to estimate values for similarity. At this early stage, is not clear yet which similarity function best fit within our case. It is a hard task and most of the metrics fail in identifying the similarity between all variations of text as argued in [39] and [40].

For sentences similarity task, we adopt a standard approach to compare the similarity between sentence pairs by computing a cosine similarity [41] between two sentences. Besides, we employed semantic similarity using WordNet published in these open resources [42], [43] for English sentences. Only images with similarity greater than a user-defined threshold are considered for further selection.

6) IMAGE RANKING (STEP 6)

Based on obtained similarity values, we select sentences that have higher similarity values with the keywords. Associated top-10 images are ranked and eventually shown, to the user/teacher, whenever (s)he clicks the option for that.

7) IMAGE EVALUATION (STEP 6)

This step is based on manual evaluation for each relevant image where the user can give a rank from 1 to 5. The system uses this value as an initial score for the final score calculation, as shown by Algorithm 2.

Algorithm 1 StoryToImages Retrieve Relevant Images to a Story

```
Input = TS: Arabic Text story, Idata: Image database,
        s,t:Thresholds
Output = ImSet: set of images
Begin
1-  keywords = processText (TS)
2-  trKeywords = Translate(keywords,En)
3-  SetEnImages = getImages(keywords, trKeywords,
    Idata)
4-  for each e in SetEnImages
5-    caption = getCaptionByDeepLearningModel(e)
6-    trCaptions = Translate(caption,Ar)
7-    if(similarity(e.caption,keywords) >= s) or
8-      (similarity(trcaption,keywords) >= t) then
9-        add(e,Idata)
10-   end
11-  end
12-  Idata = RankImages()
13-  return Idata
End
```

C. ALGORITHMS

The Algorithm 1 sketches the different steps as we presented in general system architecture. As input, we have an Arabic story text defined as *TS*, our local image-database and the threshold *s*. As output, we got *Idata* updated with a set of relevant images.

- Line 1: process Arabic keywords using NLP tools including text segmentation, pos-tagging and stop words removal;
- Line 2: translate extracted keywords to English keywords;
- Line 3: retrieve image from a local image database if no relevant image found, then Google image search is asked;
- Lines 4-9: caption each image by a deep learning model, if the sentence similarity is greater than a threshold then the image is inserted in our database
- Line 12: rank image.

The expert will also evaluate N images (in our experiment selected $N = 10$).

Algorithm 2 computes the score for each reviewed relevant image. As input, we have a set of N images.

- Line 1: initialize score
- Line 2: get$N$images from a local image database
- Line 3: set a score for each image
- Line 4: return the computed score

For the AW and EW cases, the ranking of relevant top-10 images is handled based on user evaluation, i.e., a user can give for each relevant image a ranking score and based on that the system returns the top-10 images. For the other cases, the ranking is based on similarity values results.

Algorithm 2 EvalImages Calculate Score to Every Relevant Image

```
Input: Idata: Image database, N: number of images
Output: score
Begin
1-  score = 0
2-  Idata = getImages(N)
3-   for each e in N
       score+ = getUserEval(e)
4-  return score /N
End
```

Algorithm 3 RankImages Rank Images According

```
Input : Idata: Image database, N: number of N images,
        s, t:Threshold
Output : ImSet: set of top10 images
Begin
1-  Idata = getImages(N)
2-  for each e in Idata
3-    if(EvalImages(e) >= s or SemSimilarity(e) >= t )
      then
4-      add(e,ImSet)
5-      sort(ImSet, 10)
6-  end
7-  return ImSet
End
```

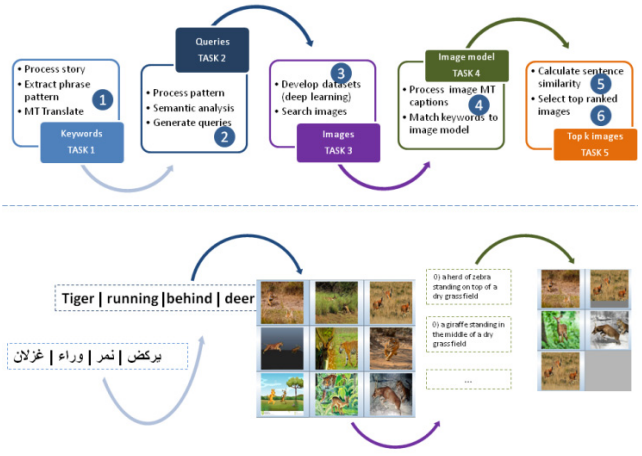


FIGURE 2. The systematic flowchart of the proposed system.

Algorithm 3 ranks input images and returns the top-10 images.

- Line 1: get N images from a local image database
- Lines 2-4: check image evaluation and image semantic similarity and add the image to ImSet
- Line 5: sort the set of images ImSet
- Line 7: return the top10 images

D. ILLUSTRATIVE EXAMPLE

We briefly describe the system workflow. As depicted in Figure 2, our proposed system works as follows. User or

teacher inputs a story text in MSA Arabic using simple sentence structure, here "يركض النمر وراء الغزلان". Then, keywords are extracted using following text processing steps; text segmentation, pos-tagging, and stopwords removal. After that, the MT process to English is applied. Then, the user clicks on "Multimedia search" either local search or Google search. After that, retrieved images are displayed. The user can validate them to be stored in a local database, in case they are not yet validated. Finally, the user can display top-10 ranked images and use them to explain the main ideas and objective of the given each sentence in a story. Table 7 depicts some of the selected sentences.

IV. EXPERIMENTS AND EVALUATIONS

This section discusses the evaluation of main components of the proposed system. These components are (*i*) keyword-based image search; (*ii*) sentence similarities; and (*iii*) the image retrieval and ranking. The goal of the evaluation is to test whether the images generated by our proposed system are able to accurately convey the main characters of the stories. We distinguish here 4 different cases: AW, AWC, EW, and EWC as indicated in Table 1.

A. EXPERIMENTAL SETUP

In the experimental setup, we start by following settings.

- *Input text data set*, we input 30 short simple stories in the domain of animals to our system. In the future, it is planned to set up a corpus for Arabic children's stories.
- *Database* has been set to store images and keywords and their mappings. For each keyword, we store 10 images and set a manual ranking from 1 to 5. Once we rank an image, we validate it to be shown to the learner. This step is necessary to preserve a safe learning environment for the learner. Image captions have been generated using a CNN pre-trained model [33]. Obtained captions are available in English only and been MT translated in Arabic. These data are also stored in the database.
- *Search engine*, we used Google image search [44] to fetch additional images. It is worth mentioning that we use Google image search in our prototype for illustrative purpose only. In the future, it is planned to set up a dataset of appropriate educational images for better learning and understanding.
- *Deep learning model*, we use *im2txt*, which is a pre-trained model for tensorflow published on github [33]. It is a model developed by Google that takes an image as input and creates a caption for it. Basically, a deep CNN is used to encode images to a fixed-length vector representation. It is followed by a LSTM network which in turn takes the encoding and produces a caption.
- *MT translation:* Extracted keywords are MT processed to English using the QCRI MT-tool [32]. Furthermore, captions are MT processed to Arabic using the same tool.

TABLE 3. Input text dataset.

Id	Story	Arabic keywords	Translated keywords
1	يركض النمر وراء الغزلان	[الغزلان, النمر, وراء, يركض]	[Tiger, Behind, Deer, Running]
2	سقطا الفيل و الحمار الوحشي في البحيرة	[سقطا, الحمار الوحشي, البحيرة, الفيل]	[Fell, Zebra, lake, elephant]
3	كان الدب الجائع يجول بالغابة	[الجائع, يجول, بالغابة, الدب]	[bear, forest, Hungry, Tour]
4	الحمار الوحشي يشرب الماء من البحيرة	[الماء, الحمار الوحشي, البحيرة, يشرب]	[water, Zebra, lake, Drink]
5	السنجاب يجري فوق أغصان الشجرة	[يجري, السنجاب, فوق, أغصان, الشجرة]	[it, Being, Branches, Squirrel]
6	سار الجمل في الرمل	[الجمل, الرمل, سار]	[sand, walked, camel]
7	أكلت السلحفاة و الأرنب العشب	[أكلت, الأرنب, السلحفاة, العشب]	[ate, rabbit, turtle, grass]
8	طار العصفور فوق الشجرة	[فوق, طار, العصفور, الشجرة]	[it, flew, bird, tree]
9	أكل الأرنب العشب	[الأرنب, العشب, أكل]	[rabbit, grass, Eat]
10	خرج الأرنب من الجحر	[الأرنب, الجحر, خرج]	[rabbit, Burrow, Out]
11	قفز القرد فوق الشجرة	[قفز, فوق, القرد, الشجرة]	[jumped, it, monkey, tree]
12	اقترب التمساح من الثور قرب النهر	[اقترب, الثور, النهر, التمساح, قرب]	[Approached, bull, river, Crocodile, Near]
13	أكل الجمل العشب	[العشب, أكل, الجمل]	[grass, Eat, camel]
14	حطمت الفيلة بيت الأرانب	[الفيلة, بيت, حطمت, الأرانب]	[Elephants, House, Shattered, Rabbits]
15	سقط الثعلب في البئر	[البئر, الثعلب, سقط]	[well, fox, fell]
16	صارع الفيل التمساح و الحمار الوحشي	[الحمار الوحشي, صارع, التمساح, الفيل]	[Zebra, Struggle , Crocodile, elephant]
17	باضت دجاجة بيضة على الطريق	[بيضة, الطريق, باضت, دجاجة]	[egg, road, eggs, Chicken]
18	كانت الدجاجة تجلس على بيضها	[تجلس, بيضها, الدجاجة]	[chicken, sitting, eggs]
19	يركض الحصان في الملعب	[الحصان, الملعب, يركض]	[horse, stadium, Running]
20	وقف الغراب الأسود على شجرة عالية	[شجرة, وقف, عالية, الغراب الأسود]	[tree, stop, High, crow]
21	يقفز القط فوق الفأر	[يقفز, فوق, الفأر, القط]	[jumps, it, mouse, cat]
22	كان هناك فراشة صغيرة وجميلة تتجول بين الزهور	[صغيرة, فراشة, تتجول, بين, الزهور, جميلة]	[Small, butterfly, Rambling, Between, Flowers, beautiful]
23	الأرانب تعيش في الغابة	[تعيش, الغابة, الأرانب]	[Live, forest, Rabbits]
24	فرس النهر يأكل الخضروات	[فرس النهر, الخضروات, يأكل]	[Vegetables, river, hippopotamus, Eat]
25	وجد الدب شجرة بها خلية نحل	[شجرة, خلية, وجد, الدب, نحل]	[tree, cell, bear, Bees]
26	البقرة تأكل العشب	[تأكل, العشب, البقرة]	[Eat, grass, cow]
27	الضفدع يعيش في البحيرة	[يعيش, الضفدع, البحيرة]	[live, frog, lake]
28	أكل الأسد و شبله اللحم	[شبله, الأسد, اللحم, أكل]	[cub, Lion, Meat, Eat]
29	يجري الكلب والذئب وراء الغزلان	[يجري, الغزلان, والذئب, الكلب, وراء]	[Being, Deer, wolf, dog, Behind]
30	البطريق يأكل السمك	[السمك, البطريق, يأكل]	[Fish, Penguin, Eat]

It is worth mentioning that we set a ranking for images for each keyword manually in the database as an initial setting. This ranking can be updated whenever a user chooses different values.

B. EVALUATION CRITERIA

We define some evaluation metrics based on the general notion of positive and negative human judgments in image retrieval as follows.

- *Relevance/Relevant*is a metric of retrieved images that most represent a specific keyword/story.
- *Precision* [45] is a proportion of relevant retrieved images to all retrieved images.
- *Top10* are relevant images having 10 highest scores.

The similarity score produced by these measures has a normalized real-number standing within the unit interval.

We have performed evaluation experiments using these metrics on a set of 30 stories to analyze user satisfaction with the output and the impact of the automatic captioning process on similarity measures and consequently on the ranking of top-10 images.

C. EVALUATION OF IMAGE RELEVANCE

The relevance and the suitability of each image were obtained through expert user. User was provided with 30 stories together with their output images from our proposed system. He was asked to judge if the images were a representation of the main subject of the topic and provide a rating on a scale of 0 (completely unsuitable) to 1 (very suitable).

The results are shown in Table 4; the average precision of our system is 44% for the AW-version compared to 49% for the EW-version. The user satisfaction with top10 is 38% for the AW-version whereas it is equal to 52% for the EW-version.

We adopt the precision as the evaluation criterion by comparing the output each time a new story is entered. The EW-version showed better average result of 49% for all output and 52% for top10 as shown in Table 4. We think such better accuracy could result from the richness of Web image resources about the availability of images to different forms of English words and to diverse events and actions. This value also reflects the current availability of Arabic open image resources in web search results compared to English image resources. This motivates us to use English translation for representing Arabic text through such diverse and various image resources.

D. EVALUATION OF IMAGE CAPTIONING

We applied an automatic captioning process for top-10 retrieved images which are treated as a preliminary set of images. The results are shown in Table 5; the average precision value for top-10 under the AWC-version is 46% compared to 56% for the EWC-version.

The captioning process is done automatically and saved us time and resources, however, the results are weak; the model creates meaningful captions for images with common objects.

TABLE 4. Precision values and user satisfaction values of the AW and EW-versions.

Story Id	AW			EW		
	Precision %	Satisfaction with		Precision %	Satisfaction with	
		Output	Top10		output	Top10
1	47	1	1	75	1	1
2	35	1	0	50	1	0
3	35	0	-	66	0	-
4	23	0	-	30	1	1
5	20	0	-	30	0	-
6	59	1	1	60	1	1
7	42	0	-	25	1	0
8	37	0	-	55	0	-
9	61	1	1	50	1	1
10	38	0	-	65	0	-
11	20	0	-	45	0	-
12	35	0	-	50	1	1
13	63	1	0	50	1	1
14	35	0	-	25	1	0
15	47	1	0	75	1	0
16	30	0	-	65	0	-
17	55	1	0	45	0	-
18	40	0	-	25	0	-
19	50	1	1	70	1	1
20	35	0	-	20	0	-
21	40	1	0	45	1	0
22	40	0	-	50	0	-
23	44	1	0	30	1	0
24	30	0	-	85	0	-
25	43	1	0	55	1	1
26	52	1	1	50	1	0
27	30	0	-	65	0	-
28	48	1	0	20	1	1
29	40	0	-	62	1	0
30	31	0	-	50	0	-
Avg	**44%**	**43%**	**38%**	**49%**	**56%**	**52%**

However, it fails when more abstract objects or similar objects in shape are present. This could be improved by considering different techniques for training. The obtained captions are stored to be used and processed by the next step, see Algorithm 1, line 7.

Table 5 shows similarity values based on comparing these captions with Arabic keywords and English keywords. Shown values are also weak so we can confirm that the captioning process with this pre-trained version [33] failed. However, even with this version, there is a minimal enhancement in the ranking of top10 images and this underlines the importance of a more accurate captioning process. The impact of this step is high on the ranking process and overall accuracy.

There are two major limitations to this method that need to be addressed in future work:

- The method depends on captions generated in English, where some do not represent either the actual content of the image nor the events or the characters.

TABLE 5. Precision values and user satisfaction values of the AWC and EWC-versions.

Story Id	AWC		EWC			
	Satisfaction with Top10	Sentence similarity sim_{cosine}	Satisfaction with Top10	Sentence similarity		Quality of translation
				Sim $_{cosine}$	Sim $_{sem}$	
1	1	0.12	1	0.05	0.14	0.12
2	0	0.15	0	0.00	0.13	0.07
3	-	0.21	-	0.05	0.08	0.15
4	-	0.13	-	0.05	0.09	0.08
5	-	0.08	-	0.18	0.31	0.07
6	1	0.15	1	0.05	0.11	0.10
7	-	0.00	-	0.00	0.10	0.08
8	-	0.18	0	0.17	0.31	0.17
9	1	0.09	1	0.08	0.08	0.20
10	-	0.14	0	0.07	0.13	0.15
11	-	0.18	-	0.22	0.11	0.12
12	-	0.11	0	0.04	0.15	0.20
13	1	0.14	1	0.07	0.16	0.26
14	-	0.17	-	0.07	0.21	0.32
15	0	0.08	-	0.26	0.15	0.15
16	-	0.17	0	0.15	0.23	0.10
17	0	0.07	-	0.09	0.20	0.20
18	-	0.01	-	0.13	0.19	0.07
19	1	0.21	1	0.15	0.27	0.31
20	-	0.16	-	0.05	0.15	0.10
21	0	0.19	-	0.11	0.26	0.18
22	-	0.19	0	0.11	0.30	0.10
23	0	0.00	-	0.00	0.13	0.05
24	-	0.12	-	0.13	0.19	0.05
25	0	0.16	0	0.19	0.13	0.17
26	1	0.35	1	0.25	0.43	0.26
27	-	0.09	0	0.04	0.24	0.27
28	0	0.07	-	0.14	0.11	0.37
29	-	0.08	1	0.12	0.15	0.10
30	-	0.06	1	0.00	0.14	0.15
Avg	**46%**	**0.12**	**56%**	**0.05**	**0.18**	**0.15**

Moreover, a set of the images were not captioned due to unsuitable image format, image size, resolution, etc.

- Retrain the model with our own image dataset in Arabic and English
- Revise the techniques for automatic image annotations

E. EVALUATION OF IMAGE RANKING AND SENTENCE SIMILARITY

Here, we leverage the semantic correlation between images to filter out some irrelevant ones using sentence similarities.

Table 5 shows values of sentence similarities; however the ability to accurately judge the similarity between sentences based on obtained similarity results from different techniques is difficult. So, in this case, we reconsidered 2 further classes of measures that can be used for identifying the similarity between sentences. We employ Cosine similarity [41], and semantic similarity using WordNet [42] to compute the syntactical similarity respectively semantic similarity between pairs of sentences. Note syntactical similarity has been done on extracted keywords and image captions with minor preprocessing. Cosine similarity is carried out for both languages, while semantic similarity has been done for English only. Nagoudi and Schwab [46] applied semantic similarity measurements for Arabic using word embeddings and preliminary results are promising.

Similarity measures values listed in Table 5 are obtained on pairs of sentences in Arabic as well as in English. In our system, these values are low; this is due to the fact that generated captions using current version are not accurate in most cases.

Ranking the top10 images based on these values led to non-satisfactory results, thus a sharp gap is observed between the satisfaction values given by expert and similarity values for those images belonging to each story. Only story "#19" had better user satisfaction for top10 images as well as better similarity values.

We assume that if the expert judges that 2 sentences are similar and the obtained similarity degree between them is weak, then we conclude that the similarity measure might not be appropriate for this task and it is necessary to look for another similarity measure. To overcome this issue, we experimented semantic similarity [42] for the AWC-version. Values shown are also weak, so we can confirm that the captioning process with the pre-trained model [33] has failed.

There are two major limitations to this method that need to be addressed in future work:

- Investigate semantic similarity for Arabic sentence pairs using standard Arabic lexical resources.
- Revise the techniques for word similarity and sentence similarity

F. EVALUATION OF QUALITY OF THE TRANSLATION

For evaluating the quality of MT translation of image captions from English to Arabic, each sentence pair is judged by two humans whether the sentence pair is semantically coherent or not. The users gave a similarity score between 5 and 0, where 5 means a good translation and 0 means that there is no semantic similarity between the pair of sentences.

Table 5 showed overall 15 % of the total sentence pairs are judged to be positive examples means they are semantically equivalent. Consequently, the value indicates relatively poor performance in judging similar pairs. Thus, it adversely affects the overall accuracy.

For the quality of translation, we remark that translations of image captions from English to Arabic have led

TABLE 6. MAP and average sentence similarity values for all versions.

Version	Mean average precision (human judgment)		Sentence similarity measures	
	All	Top10	Syntax	Semantic
AW	44%	38%	0.00	-
AWC	44%	46%	0.12	-
EW	49%	52%	0.00	0.00
EWC	49%	56%	0.05	0.18

to low sentence semantic similarity values with the input stories, except for story 26, which has reached better values overall similarities. These results are not encouraging to use MT translation tool for our future work.

G. EVALUATION OF MAPPING TEXT TO MULTIMEDIA

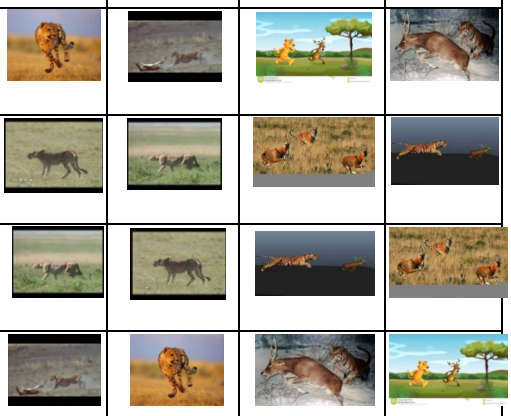
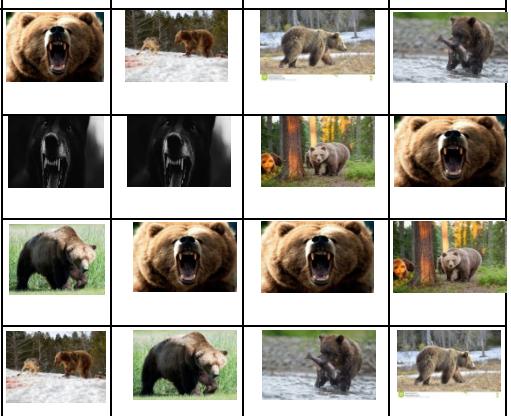
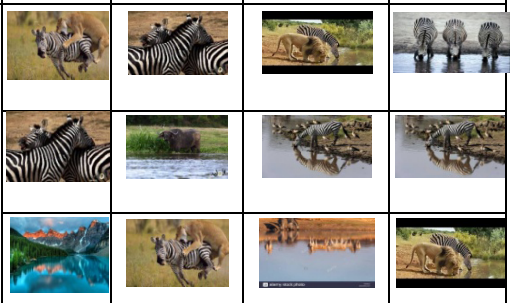
The objective of this work is to map Arabic text to relevant images as first step. We target at enhancing our current proposed system that is still under development and improvement. We are currently processing sentences with simple structure where stop words are eliminated. Then, keywords are extracted and searched first in the local database and their corresponding images are presented. Using the concatenation of all keywords produced often more relevant for delivering the meaning of the whole story and consequently has led to better performance than separate keywords.

We believe that it is owed to the concatenations that capture the interactions between characters, while separate keywords query only lists the characters and other story elements. Top10 images are also an option to save time in looking over all relevant images. Whenever the retrieved images are not relevant, the user can address a query to Google Search. A validation step for new images is required in this case to keep a safe learning environment.

We list below major limitations in our current status that should be improved:

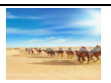
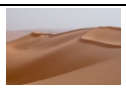
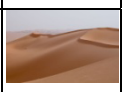
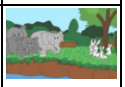
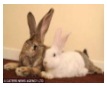
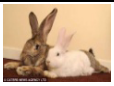
- Search result, to improve search results from our database or from Google search engine, the keyword query should be expanded with synonyms, since some results may be retrieved using synonymous;
- Keyword extraction, we should revise the techniques for keywords extraction;
- NLU, Natural Language Understanding the words of sentences and their roles (verbs, nouns, adverbs, etc.);
- Different heuristics to compute semantic similarity between sentences using different lexical resources can be used;
- Nevertheless, we strongly believe that sentence similarity in this work offers an interesting and useful insight into the performance of these similarity measures which are crucial to such applications;
- Each image's metadata such as link and image name retrieved by the search can also be exploited as an indication of the image's content similar to the work done by Nikolaos and Mark [47];
- A filtering algorithm must be developed to remove inappropriate images;
- An image pre-processing should be employed to bring images to a unified format ready for eventual composition.

TABLE 7. Selected outputs for all versions.

Sample sentences	AW	AWC	EW	EWC
يركض النمر وراء الغزلان				
كان الدب الجائع يجول بالغابة				
الحمار الوحشي يشرب الماء من البحيرة				

We summarize the results in Table 6, which shows enhancements in two main directions. First, using English Keywords instead of Arabic keywords has led to 5% overall in precision (from 44% to 49%). Second, with respect to Arabic and English version with captioning, the

TABLE 7. ***(Continued.)*** **Selected outputs for all versions.**

				
سار الجمل في الرمل				
				
				
				
حطمت الفيلة بيت الأرانب				
				
				
				

values for user satisfaction for top10 reached an overall improvement of 8%.

V. CONCLUSION

We introduced in this paper an approach to automatically convert simple Arabic children's stories to the best representative images from Google image search using deep learning model for image captioning. Using a captioning process to rank top-k images has shown reasonable precision values as per our preliminary results. Using Google image search and MT process has led to many limitations in our current work such as word ambiguity and malformed MT translations. We aim in the future to look for creating appropriate resources (corpus, well-annotated multimedia etc.) for Arabic text.

In the future, we are planning to follow an alternative technique for Arabic sentences using Part-of-Speech tagging (POS tag) for identification of words that are highly descriptive in each input sentence similar to prior successful works. We are committed to also evaluate in future works the amount of understanding and learning that can be achieved with simple visualizations namely visual representations for basic concrete nouns and verbs.

ACKNOWLEDGMENT

This work was made possible by NPRP Grant #10-0205170346 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] J. Dhiraj, J. Z. Wang, and J. Li, "The story picturing engine—A system for automatic text illustration," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 68–89, 2006.
[2] K. Hassani and W. S. Lee, "Visualizing natural language descriptions: A survey," *ACM Comput. Surv.*, vol. 49, no. 1, p. 17, 2016.
[3] A. Ramisa, F. Yan, F. Moreno-Noguer, and K. Mikolajczyk. (2016). "BreakingNews: Article annotation by image and text processing." [Online]. Available: https://arxiv.org/abs/1603.07141
[4] Y. Jiang, J. Liu, and H. Lu, "Chat with illustration," *Multimedia Syst.*, vol. 22, no. 1, pp. 5–16, 2014.
[5] A. Pasha *et al.*, "MADAMIRA: A Fast, comprehensive tool for morphological analysis and disambiguation of Arabic," in *Proc. 9th Int. Conf. Lang. Resour. Eval.*, 2014, pp. 1094–1101.
[6] W. L. K. Hassani, "Visualizing natural language descriptions: A survey," *ACM Comput. Surv., pp.*, vol. 17, pp. 1–34, 2016.
[7] W. Zhiyu, C. Peng, X. Lexing, Z. Wenwu, R. Yong, and Y. Shiqiang, "Bilateral correspondence model for words-and-pictures association in multimedia-rich microblogs," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 10, no. 4, pp. 34:1–34:21, 2014.
[8] F. Coelho and C. Ribeiro, "Automatic illustration with cross-media retrieval in large-scale collections," in *Proc. 9th Int. Workshop Content-Based Multimedia Indexing (CBMI)*, Madrid, Spain, Jun. 2011, pp. 25–30.
[9] X. Tian, Y. Lu, and L. Yang, "Query difficulty prediction for Web image search," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 951–962, Aug. 2012.
[10] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An on-line lexical database," *Int. J. Lexicogr.*, vol. 3, no. 4, pp. 235–244, 1990.
[11] J. K. Poots and N. Cercone, "First steps in the investigation of automated text annotation with pictures," *Big Data Inf. Anal.*, vol. 2, no. 2, pp. 97–106, 2017.
[12] R. Mihalcea and W. L. Chee, "Toward communicating simple sentences using pictorial representations," *Mach. Transl.*, vol. 22, pp. 153–173, Sep. 2008.
[13] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, "Enriching textbooks with images," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, Glasgow, Scotland, 2011, pp. 1847–1856.
[14] B. Coyne and R. Sproat, "WordsEye: An automatic text-to-scene conversion system," in *Proc. 28th Annu. Conf. Comput. Graph. Interact. Techn.*, New York, NY, USA, 2001, pp. 487–496.
[15] X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock, "A text-to-picture synthesis system for augmenting communication," in *Proc. 22nd AAAI*, vol. 2, Vancouver, BC, Canada, 2007, pp. 1590–1595.
[16] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Barcelona, Spain, 2004, pp. 404–411.
[17] U. Dmitry and K. Aleksandr, "An ontology based approach to text-to-picture synthesis systems," in *Proc. 2nd Int. Workshop Concept Discovery*, Leuven, Belgium, 2012, pp. 94–101.
[18] C.-J. Huang, C.-T. Li, and M.-K. Shan, "VizStory: Visualization of digital narrative for fairy tales," in *Proc. Conf. Technol. Appl. Artif. Intell.*, Taipei, Taiwan, Dec. 2013, pp. 67–72.
[19] A. F. Bobick *et al.*, "The KidsRoom: A perceptually-based interactive and immersive story environment," *Presence*, vol. 8, no. 4, pp. 369–393, 1999.
[20] C. B. Larson and B. C. Petersen, "Interactive storytelling in a multimodal environment," Aalborg Univ., Inst. Electron. Syst., Aalborg, Denmark, Tech. Rep., 1999.
[21] K. Sumi and M. Nagata, "Animated storytelling system via text," in *Proc. ACM SIGCHI Int. Conf. Adv. Comput. Entertainment Technol.*, 2006. Art. no. 55.
[22] M. Udi, "Understanding simple stories through domain-based terms extraction and multimedia elements," M.S. thesis, Qatar Univ., Doha, Qatar, Tech. Rep. 10576/3320, 2013.

[23] A. Talsania, S. Modha, H. Joshi, and A. Ganatra, "Automated story illustrator," in *Proc. Post Workshops 7th Forum Inf. Retr. Eval.*, Gandhinagar, Gujarat, 2015, pp. 71–73.
[24] D. Ganguly, I. Calixto, and G. Jones, "Overview of the automated story illustration task at FIRE 2015," in *Proc. Post Workshops 7th Forum Inf. Retr. Eval.*, Gandhinagar, Gujarat, 2015, pp. 63–66.
[25] S. Boonpa, S. Rimcharoen, and T. Charoenporn, "Relationship extraction from Thai children's tales for generating illustration," in *Proc. 2nd Int. Conf. Inf. Technol. (INCIT)*, Nakhonpathom, Thailand, Nov. 2017, pp. 1–5.
[26] A. G. Karkar and J. M. Aljai'am, and A. Mahmood, "Illustrate it! An Arabic multimedia text-to-picture m-learning system," *IEEE Access*, vol. 5, pp. 12777–12787, 2017.
[27] H. Eva, K. P. Mc, L. Tom, and C. Joan, "NewsViz: Emotional visualization of news stories," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 125–130.
[28] D. Delgado, J. Magalhães, and N. Correia, "Automated illustration of news stories," in *Proc. IEEE 4th Int. Conf. Semantic Comput.*, Pittsburgh, PA, USA, Sep. 2010, pp. 73–78.
[29] W. Ruan, N. Appasani, K. Kim, J. Vincelli, H. Kim, and W.-S. Lee, "Pictorial visualization of EMR summary interface and medical information extraction of clinical notes," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Ottawa, ON, Canada, Jun. 2018, pp. 1–6.
[30] U. Dmitry, "A text-to-picture system for Russian language," in *Proc. 6th Russian Young Scientist Conf. Inf. Retr.*, 2012, pp. 35–44.
[31] P. Jain, H. Darbari, and V. C. Bhavsar, "Vishit: A visualizer for Hindi text," in *Proc. 4th Int. Conf. Commun. Syst. Netw. Technol.*, Bhopal, India, Apr. 2014, pp. 886–890.
[32] Q. QCRI Doha. *QCRI MT-Tool*. Accessed: Oct. 2, 2018. [Online]. Available: https://mt.qcri.org/api
[33] C. Shallue. *Show and Tell: A Neural Image Caption Generator*. Accessed: May 1, 2018. [Online]. Available: https://github.com/tensorflow/models/tree/master/research/im2txt
[34] A. Abdelali, A. Ali, F. Guzmán, F. Stahlberg, S. Vogel and Y. Zhang, "QAT^2—The QCRI advanced transcription and translation system," in *Proc. INTERSPEECH*, Dresden, Germany, 2015, pp. 726–727.
[35] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 664–676, Apr. 2017.
[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[37] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. (2016). "Show and Tell: Lessons learned from the 2015 MSCOCO image captioning challenge." [Online]. Available: https://arxiv.org/abs/1609.06647
[38] *MSCOCO*. Accessed: Mar. 1, 2018. [Online]. Available: http://cocodataset.org/#home
[39] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proc. 21st Nat. Conf. Artif. Intell.*, Boston, MA, USA, vol. 1, 2006, pp. 775–780.
[40] P. Achananuparp, X. Hu, and X. Shen, "The evaluation of sentence similarity measures," in *Proc. 10th Int. Conf. Data Warehousing Knowl. Discovery*, Turin, Italy, 2008, pp. 305–316.
[41] Wikipedia. *Cosine Similarity*. Accessed: Sep. 4, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Cosine_similarity
[42] Agarwal. *Semantic Similarity Using WordNet*. Accessed: Oct. 1, 2018. [Online]. Available: https://www.kaggle.com/antriksh5235/semantic-similarity-using-wordnet
[43] *WordNet Sentence Similarity*. Accessed: Nov. 1, 2018. [Online]. Available: https://nlpforhackers.io/wordnet-sentence-similarity/
[44] Google Developers. *Google Custom Search*. Accessed: Apr. 1, 2018. [Online]. Available: https://developers.google.com/apis-explorer/#search/customsearch/
[45] Wikipedia. *Precision and Recall*. Accessed: Oct. 1, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Precision_and_recall
[46] D. Schwab, "Semantic similarity of Arabic sentences with word embeddings," in *Proc. 3rd Arabic Natural Lang. Process.*, 2017, pp. 18–24.
[47] A. Nikolaos and S. Mark, "Representing Topics Using Images," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Atlanta, Georgia, 2013, pp. 15–167.
[48] N. UzZaman, J. P. Bigham, and J. F. Allen, "Multimodal summarization of complex sentences," in *Proc. 16th Int. Conf. Intell. User Interfaces*, Palo Alto, CA, USA, 2011, pp. 43–52.
[49] H. Li, J. Tang, G. Li, and T.-S. Chua, "Word2image: Towards visual interpreting of words," in *Proc. ACM Int. Conf. Multimedia, Co-Located Symp. Workshops*, Vancouver, BC, Canada, 2008, pp. 813–816.

JEZIA ZAKRAOUI received the master's and Ph.D. degrees in computer science from the Vienna University of Technology, Austria, in 2012. She is currently with the Department of Computer Science and Engineering, Qatar University. Her research interests include semantic Web, text mining, artificial intelligence, deep learning, and machine learning.

SAMIR ELLOUMI received the Ph.D. degree from the Faculty of Sciences of Tunis, Tunisia, in 2002. Since 2010, he has been a Postdoctoral Researcher with Qatar University, where he focuses on NPRP projects related to information technology such as machine learning, classification, information extraction, and retrieval. He is currently an Assistant Professor with the University of Tunis El Manar. He has focused on the formal concept analysis domain and has contributed in several aspects related to fuzzy and reals sets.

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 focused on the connection machine CM5 with 65 000 microprocessors in USA. He was a Project Manager with IBM, Paris, and an IT Consultant with RTS, France, for several years. He is currently a Full Professor with the Department of Computer Science and Engineering, Qatar University. He is leading a research team in multimedia and assistive technology and collaborating in the financial watch and intelligent document management system for automatic writer identification and MOALEM projects. He is also leading a team in NLP, knowledge extraction, and multimedia. He has collaborated with different researchers in Canada, France, Malaysia, GCC, and USA. He has published so far 159 papers and eight books chapters in computing and information technology which are published in conference proceedings, scientific books, and international journals. His current research interests include multimedia, assistive technology, learning systems, human–computer interaction, stochastic algorithms, artificial intelligence, information retrieval, and natural language processing. He is a member of the editorial boards of the *Journal of Soft Computing*, *American Journal of Applied Sciences*, the *Journal of Computing and Information Sciences*, the *Journal of Computing and Information Technology*, and the *Journal of Emerging Technologies in Web Intelligence*. He acted as a Scientific Committee Member of different international conferences (ACIT, SETIT, ICTTA, ACTEA, ICLAN, ICCCE, MESM, ICENCO, GMAG, CGIV, ICICS, and ICOST). He received the 2015 ACM Transactions on Multimedia Computing, Communications and Applications Nicolas D. Georganas Best Paper Award, the Best Research Paper at the 10th Annual International Conference on Computer Games Multimedia and Allied Technologies, Singapore, in 2016, and the Best Research Paper Award at the IEEE ICe3 Conference on e-Learning, e-Management and e-Services, Malaysia, in 2018. He organized many workshops and conferences in France, USA, and GCC countries. He is a Regular Reviewer for the *ACM Computing Review* and the *Journal of Supercomputing*, and an Associate Editor of the IEEE Access. He is the Main Organizer and the General Chair of the International Conference on Computer and Applications.

SADOK BEN YAHIA received the Ph.D. degree from the Faculty of Sciences of Tunis, Tunisia, in 2001, and the Habilitation degree from the University of Montpellier, France, in 2009. He is currently a Full Professor of computer science technologies and the Head of the Data Science Group, Tallinn University of Technology. His research interests include combinatorial aspects in big data and their applications to different fields, e.g., data mining, e-healthcadre information systems, social computing, information aggregation, and dissemination in smart cities.