

IMPROVING NATURAL LANGUAGE QUERIES SEARCH AND RETRIEVAL THROUGH SEMANTIC IMAGE ANNOTATION UNDERSTANDING

Haitham Samih*

Faculty of Computer
and Information
Technology, Egyptian
E-Learning University,
Cairo, Egypt
hsamih@eelu.edu.eg

Sherine Rady

Information Systems
Department,
Faculty of Computer and
Information Sciences, Ain
Shams University,
Cairo, Egypt
srady@cis.asu.edu.eg

Manal A. Ismail

Computer Engineering and
Systems Department,
Faculty of Engineering,
Helwan University,
Cairo, Egypt
Manal_shoman@yahoo.com

Tarek F. Gharib

Information Systems
Department,
Faculty of Computer and
Information Sciences, Ain
Shams University,
Cairo, Egypt
tfgharib@cis.asu.edu.eg

Received 2020-11-17; Revised 2020-12-27; Accepted 2021-01-03
Available online 2021-1-5

Abstract: Retrieving images using detailed natural language queries remains a difficult challenge. Traditional annotation-based image retrieval systems using word matching techniques cannot efficiently support such query types. Significant improvements for this problem can be achieved with a semantic understanding for those query sentences and image annotations. This paper presents a two-stage semantic understanding approach for natural language query sentences. At the first stage, the Stanford parser and a designed rule-based relation extraction tool are used in triple extraction process to efficiently extract the objects attributes, instances and natural language annotations relationships involving these objects. The second stage integrates the extracted relations with external commonsense knowledge source in a mapping process to provide high-level semantic meanings to the extracted triples. Experiments are conducted for evaluating the benefit of the proposed semantic understanding against a testing set of natural language sentences from the Flickr8k dataset. The results show that the proposed approach succeeds to extract relational triples with average accuracy value of 97% for the different types of annotations relationships: attributes and instance relations, multiword dependence relations, and semantic relations.

Keywords: Image retrieval, natural language queries, semantic understanding, commonsense knowledge sources

1. Introduction

* Corresponding author: Haitham Samih
Faculty of Computer and Information Technology, Egyptian E-Learning University, Cairo, Egypt
E-mail address: hsamih@eelu.edu.eg

Exponential access to digital technology and continuous advancement in Internet communications results in considerable growth of digital media repositories such as audio, image, and video. Searching these repositories is an exhausting task for the user without the aid of efficient retrieval systems. Efficient systems allow users to search for images by identifying not only the objects but also the attributes and structured relationships involving these objects. E.g. “Two brown dogs playing with a Frisbee in the water”. However, searching for very specific results using such queries remains a challenge [1, 2]. As a consequence, retrieving images while considering the semantic understanding of their content has attracted the attention of many researchers from different communities, such as computer vision (CV) and natural language processing (NLP) [1, 3].

Traditional annotation-based image retrieval systems annotate images with textual descriptions and image retrieval often done using text-matching techniques. Such systems and solution paradigm do not produce appropriate desired results as they are unable to understand the rich semantics of natural language queries [4].

To support natural language queries with rich semantics, some work proposed the mapping of image annotations to multi-modal embedding of images and descriptions [5-8]. However, these models are usually trained on single-sentence annotation which is not enough to capture the high-level semantics of visual scenes. In addition, description coverage depends largely on human perception [9].

A significant improvement to overcome the limitation of these models is to annotate images with a semantic graph-based representation referred to as a scene graph [2]. The proposed scene graph explicitly combines object instances, attributes of objects, and relationships involving these objects. In that work, the scene graph managed to capture image semantics and accordingly enhance the retrieval accuracy. However, the user's query is restricted to be in the form of a scene graph instead of an image description in natural language, which is inconvenient to potential users.

Transforming natural language descriptions into meaningful graph representation has attracted the attention of some computational linguists. A great advantage of such intermediate representation is to allow distinct development, improvement and reuse of NLP and vision subsystems.

Recent work has proposed an abstract meaning representation (AMR) framework for representing the semantics of natural language sentences in a directed graph form [10]. The AMR framework is followed by several parsers to parse natural language sentences into AMR [11-13]. However, AMR parsers have unsatisfactory performance compared to dependency parsers, making the dependency trees a better start for the parsing process [1, 13, 14].

The work in [1, 13] proposed a two-stage approach for parsing image descriptions to scene graphs. In the first stage, dependency parser is utilized to produce dependency trees from natural language sentences, which provide a representation of the syntactic relations between words in the sentence. Next, a rule-based system is developed to extract the semantic relations and objects instances and attributes. The second stage is a mapping process that maps from dependency trees, the output of the first stage, to scene graphs.

Scene graphs have been reused with various systems including 2D-scenes or 3D-scenes synthesis [15, 16] and visual question answering [17] where the input should be represented graphically. Performance of these tasks still needs to be improved, in which more accurate scene graph parsers can contribute to this improvement [18].

The work in [19, 20] considered the use of the K-Parser as a parsing tool for extracting various knowledge from image descriptions. However, the conceptual knowledge of the K-Parser is sometimes

incorrect. In that work a set of rules are developed to transform K-Parser's output and solve some basic parsing mistakes.

The proposed approaches achieve high-level performance with the relation extraction stage but limitations remain with regard to the mapping stage. Mapping relational triples to background/external knowledge has the effect of reducing the heterogeneity problem and thus provides accurate meanings for the extracted relations.

This paper presents a two-stage semantic understanding for natural language sentences. The first stage employs off the shelf and a developed rule-based relation extraction tools to efficiently extract the objects attributes, instances and relationships involving these objects. The second stage integrates the ConceptNet knowledge base as external commonsense knowledge with the output of the first processing stage in significant process named relation mapping process, whose aim is to provide accurate semantic meanings for the natural language sentences.

The rest of this paper is structured as follows. Related work is summarized in Section 2. Section 3 presents the proposed semantic annotation understanding: Stanford parser and the used knowledge base sources. Section 4 reports the experiments and their results. Finally, section 5 concludes the work and offers directions for future work.

2. Related Work

Recently, good attempts have been made to transform image descriptions, represented in the form of natural language, into a graph-based semantic representations and integrating these graphs with commonsense knowledge. The work proved that the graph-based representation is an effective form for representing natural language sentences [1, 2]. The work in [21] utilizes commonsense knowledge to enhance the image retrieval process. The work proves that integrating background knowledge from commonsense knowledge bases significantly enhance the image retrieval process.

Extracting semantics of image contents and/or image descriptions and preserving these semantics using graphs recently receive some attention. The work in [14, 15] introduced a data structure named scene graph for describing the semantics of an image. In that work, a two-stage parser is designed for parsing natural language sentences into scene graphs. The first stage includes the use of a dependency parser to establish syntactic dependencies between different words in the sentence. Next, a rule-based system is designed to extract various types of relations including object instances and attributes from the generated dependency parse tree. The second stage involves a mapping process that provides an accurate understanding of the extracted triples.

In [22] a knowledge generation approach is proposed. First, NLP techniques are utilized to extract relation triples (subject, predicate, object) from natural language sentences. Next, the extracted triples are transformed into resource description framework (RDF) statements using its designed ontology.

The work in [23, 24] presents open information extraction systems. Grammatical and lexical patterns are defined to extract various relation triples from natural language sentences. These systems don't consider the entity mapping process, thus ambiguity may be present in the extracted triples.

The work in [20] made a promising effort in knowledge extraction and applying commonsense knowledge in various language technologies for improving image understanding and retrieval. In that work, image annotations are parsed using the K-Parser and newly developed parsing rules and converted to RDF statements. From the other hand, natural language queries are processed using the same parsing tools and the generated representation is translated into SPARQL. Such semantic understanding and representation for queries and image annotations managed to increase the searchability of the user query over the annotated images. The main problem with this system is the heavy dependence on the output of the K-Parser which is sometimes inaccurate due to parsing and conceptual mistakes.

The work in this paper is close to [14, 15]. The proposed work further improves this work in different aspects: First, external commonsense knowledge is used in a relation mapping process to provide accurate meaning for the extracted triples to solve the problem of semantic heterogeneity. Second, the work considers an additional type of semantic relations, the multiword dependence relation that improves the performance of different text mining tasks by reserving the semantics of the multiword expressions.

3. Proposed Semantic Annotation Understanding

The proposed approach consists of two consecutive processing steps: Relation triple extraction and Relation mapping as presented in Figure. 1. Relation triple extraction process involves the use of off the shelf and a developed rule-based relation extraction tools that are described in the subsequent

subsections. These tools guarantee the extraction of various types of semantic relations from the input natural language sentence that support the semantic annotation understanding.

Extracted relations and associated arguments are represented in the traditional triplet form: subject, predicate, and object. Next, extracted triples are integrated with the ConceptNet knowledge base in the relation mapping process to provide semantic accurate meanings of the extracted triples.

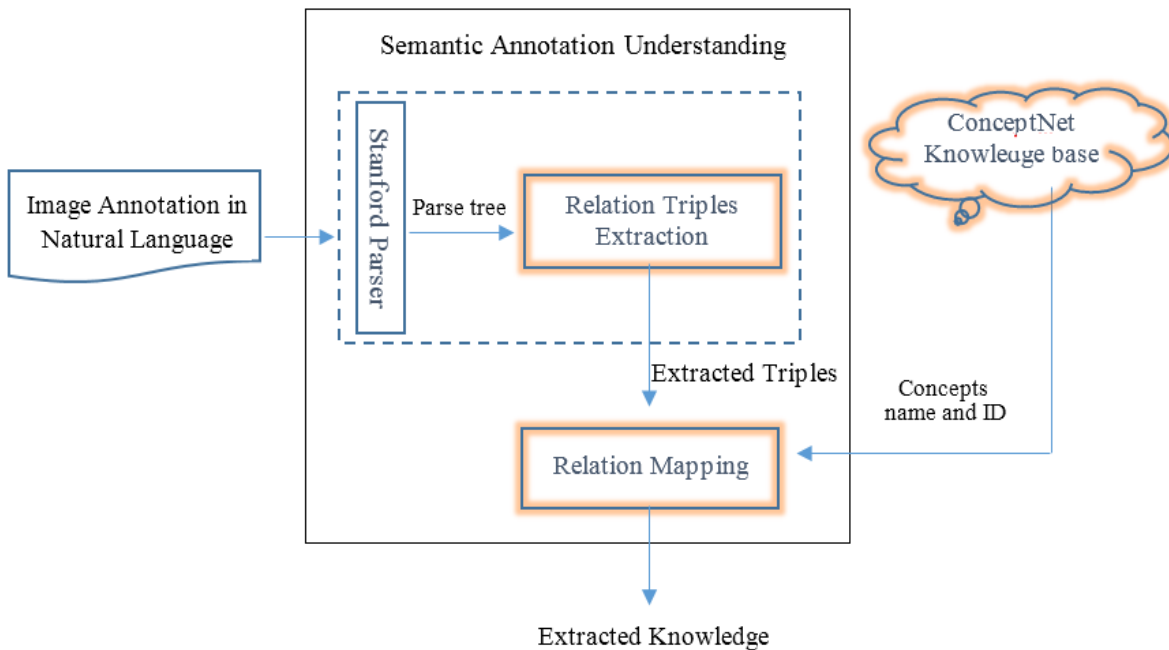


Figure. 1: Proposed Semantic Annotation Understanding architecture

3.1 Relation triple extraction

The main purpose of the proposed relation triple extraction is to interpret the meaning of the input natural language sentence by considering different relationships and the arguments associated with them. Almost, related works assume that the relation represents the predicate and its subject and object arguments [25]. The relation extraction is extended to include not only those mentioned, but also adjectives, instances and multiword dependencies. In this work, three types of relations are tracked: attribute and instance relation, multiword dependence relation and semantic relation. Regardless of the different types of relations, dependency trees are commonly used as a better start for parsing natural language sentences. Hence, the Stanford Parser is employed firstly to generate a dependency parse tree for the input sentence. Figure. 2 shows the parse tree for a given example sentence.

Input Sentence: "A black and white dog runs through the white water."

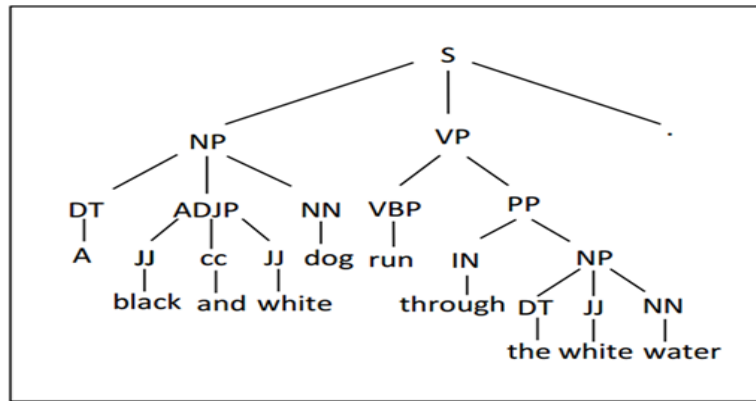


Figure. 2: Stanford parse tree for a given sentence

a) *Attribute and instance relation*

An attribute relation is a triple relation that has one or more adjective in its object part describing the subject part of the relation. An instance relation is a triple relation where the object part represents the number of instances of the noun given in its subject part. Starting from the dependency tree of the input sentence, two dependency patterns are then followed to detect ADL+NOUN and CD+NOUN POS patterns. Dependency patterns are defined using Tregex expressions [26]. Tregex is an effective matching tool used to match an individual word, regular expression or group of POS tags such as a Noun Phrase (NP) or Verb Phrase (VP).

b) *Multiword Relation*

A multiword dependency relation states the dependency between two words to form one multiword concept. Such relation provides accurate understanding of natural language sentences and improves the performance of different text mining tasks as it retains the semantic interpretation of the multiword concepts. [27]. Starting from the parse tree of the input sentence, Tregex expressions are employed to detect NN+NN POS pattern. Such a pattern will generally consider any two consequence nouns as a single multiword concept (e.g. "A brown dog plays with a tiny soccer ball"). The "soccer ball" will be extracted as a single multiword concept. However, other inaccurate relations may be exist (e.g. "A tan dog runs in a grassy field"). The "tan dog" will be extracted as a single multiword concept due to some parsing mistakes. Wrong relations will be fixed using another part of the design (Relation Mapping).

c) *Semantic relation*

A semantic relation triple is a relational triple describes a semantic relationship between its subject and object parts. Different Open Information Extraction (IE) techniques can be used to detect and extract such type of relations among them OLLIE [28] and Stanford Open IE [29]. In this work, the Stanford Open IE was selected.

Stanford Open IE has the advantage of dividing a long sentence into short and coherent sentences, allowing better awareness of the context of each extraction and providing accurate triples to other applications

The relation extraction algorithm presented in [30] is also utilized to improve the extraction process and generate additional triples. Figure. 3 shows the pseudocode for the triplet extraction algorithm.

3.2 Relation Mapping

The relation mapping process associates elements of the extracted triples with their relevant concepts in ConceptNet knowledge base. Such process provides accurate interpretations to the extracted triples and consequently, provides a precise understanding of the natural language sentence. Triple elements are sent as a query to the ConceptNet knowledge base, when a match occurs, the matched concept will substitute the triple element. Otherwise, relevant concepts and their semantic weights are retrieved in response to the query and the most related concept will be selected.

The Relation Mapping process is also used to fix inaccurate extraction of multiword dependence relations. The multiword dependence relation is sent as a query to the ConceptNet knowledge base. If a match occurs, then the multiword relation will be accepted, otherwise it will be ignored.

```

Input: Stanford parse tree for a natural language sentence
Output: Semantic relation in traditional triplet form
1. Subj ← first noun in NP_subtree
2. Pred ← deepest verb in VP_subtree
3. siblings ← NP, PP and ADJP siblings of VP_subtree
   for each S in siblings do
       if S = ADJP
           Obj ← first adjective in S
       else
           Obj ← first noun in S
   return Subj, Pred, Obj

```

Figure. 3: Semantic triple extraction pseudocode

4. Experimental Study

4.1 Dataset and experiment setup

Flickr8k dataset [31] is a prevalent dataset frequently used as a benchmark for assessing the performance of image retrieval approaches. It includes 8000 images of people and animals doing various activities. These images are obtained from the Flickr website. The dataset also includes 5 general and detailed annotations of the content of each image collected through crowdsourcing (using Amazon Mechanical Turk).

The proposed semantic annotation understanding is evaluated against a set of 200 natural language sentences synthesized by considering various annotations extracted from the Flickr8k dataset and related concepts from the ConceptNet knowledge base, to finally produce a random combination testing set.

Relational triples are extracted manually from the testing set constituting a total set of 1000 triples involving 575 attributes and instance relations, 35 multiword dependence relations and 390 semantic relations. Table 1 gives some examples of the testing set.

Table 1. Testing set samples

No.	Natural language sentence	Manual triples extraction		
		Attributes and instance relation	Multiword dependence relation	Semantic relation
1	A brown dog with a tiny soccer ball in its mouth be stand on a blanket	dog; has_property; brown ball; has_property; tiny blanket; has_count; 1 dog; has_count; 1 ball; has_count; 1	ball; has_dependent ; soccer	ball; in; mouth dog; with; soccer_ball dog; stand; blanket
2	A black and white dog be play with a large brown cow.	dog; has_property; black:white cow; has_property; large:brown cow; has_count; 1 dog; has_count; 1	-	dog; play_with; cow dog; be; play
3	A black dog and a brown dog run around in the sand	dog; has_property; black dog; has_property; brown dog; has_count; 2	-	dog; run; sand

4.2 Experiments and Results

The conducted experiments validate the extraction quality of the proposed semantic annotation understanding against the human manual extraction that simulates the ground truth.

In the experimentation, the testing set is processed through the proposed semantic annotation understanding to generate a knowledge-based semantic understanding for each sentence.

Table 2 shows some examples for such processing. The table lists different and complex set of natural language sentences. For the first sentence, the proposed approach successfully extracts all explicitly included relations. Regarding the extraction rules, the proposed approach managed to extract the implicit instance relation found in the second sentence for the keyword "dog" which is 2. The third sentence includes more complex natural language relations, one of which is the multiword dependence relation which states the relationship between the keyword "fence" and the keyword "wire".

Table 2. Examples of annotation extraction using the proposed automated approach

No.	Natural language sentence	Automated triple extraction		
		Attributes and instance relation	Multiword dependence relation	Semantic relation
1	Two brown dogs play with a Frisbee in the water.	dog; has_property; brown dog; has_count; 2 frisbee; has_count; 1	-	dog; play_with; water frisbee; in; water dog; play; Frisbee
2	A white dog chases a black dog across the grass.	dog; has_property; white dog; has_property; black dog; has_count; 2	-	dog; chase; grass dog; chase; dog
3	A black dog runs parallel to a wire fence in a grassy meadow	dog; has_property; black meadow; has_property; grassy dog; has_count; 1 fence; has_count; 1	fence; has_dependent; wire	dog; run_to; fence fence; in; meadow

Three types of relations are considered: attribute and instance relation, multiword relation and semantic relation. For each type, the number of extracted triples is recorded and accuracies are calculated.

Table 3 shows a comparison of the extracted triples when performing manual extraction versus the extraction of the proposed automated approach.

Table 3. Comparison of the extracted triples- proposed versus manual extraction

Relation type	Manual Extraction	Proposed Automated approach	Accuracy
Attributes and instance relation	575	560	0.97
Multiword dependence relation	35	40	1
Semantic relation	390	367	0.94

The results show that the proposed semantic understanding extracts about 97% of the Attributes and instance relations, 94% for the semantic relations and 100% for the multiword dependence relations compared to the manual triples extraction. The results also show that compared to the manual extraction of the multiword dependence relations, the proposed approach extracts 5 incorrect dependence relations. Unintended multiword relations are generated due to some parsing mistakes.

Experiments are extended to assess the effect of the proposed automated natural language understanding on improving the performance of natural language queries search and retrieval.

For this context, the accuracy at k performance measure has been tracked for a testing set of 50 natural language queries. In this experiment, the proposed semantic understanding approach is used as intermediate structure for the retrieval process. Firstly, user's query and image annotations are processed using the proposed semantic understanding approach. Next, for the retrieval, the obtained semantic understandings are tested using two recent relation-based matching methods, which are the semantic relation similarity [32] and the binary relation similarity [33].

Figure. 4 shows the retrieval accuracy for the semantic relation similarity against the binary relation similarity for different top k values over the proposed automated natural language understanding. The figure shows that higher retrieval accuracies are obtained due to the use of the proposed natural language understanding as intermediary structure for the semantic relation matching process. The semantic relation similarity achieves an average accuracy value of 95% compared to 50% for the binary relation similarity. This counts for 45% increase in the accuracy at k values. This proves the efficiency of the proposed semantic understanding approach for representing the semantics of the user's query and image annotations. It also highlights the superiority of semantic relation matching over binary relation matching in improving natural language query search and retrieval.

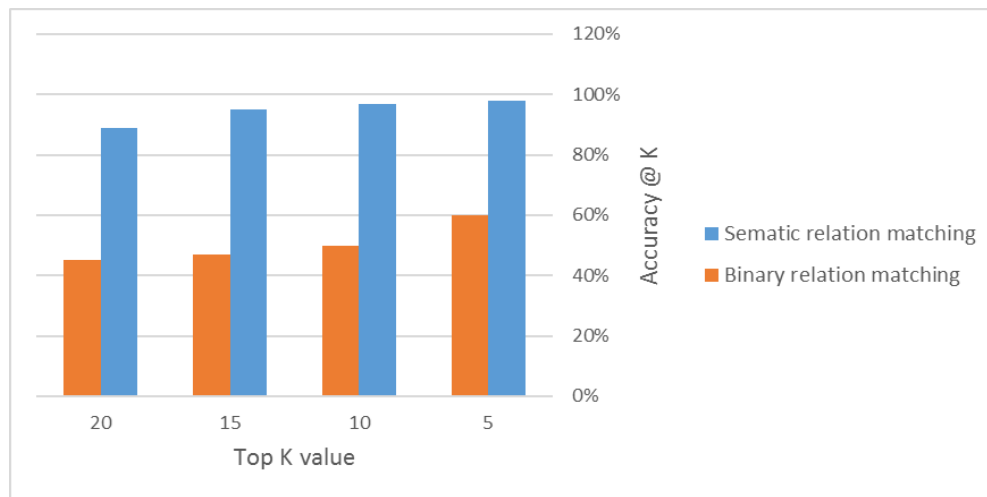


Figure. 4: Comparison of the retrieval accuracy-semantic versus binary relation matching

A third experiment is conducted to validate the quality of the proposed semantic annotation understanding against another parsing tool which is the K-Parser. The latter is utilized in [32] for knowledge extraction from natural language annotations. The retrieval accuracies of the semantic relation similarity when employing the proposed semantic understanding have been compared against the retrieval accuracies when employing the K-Parser, for different respective values of $k = 1, 5, 10$. Table 4 lists the results for such comparison over the Flickr8k dataset. The results show that for the different values of $k = 1, 5, 10$, the proposed semantic understanding approach achieves accuracy values of 91%, 96%, 98% respectively compared to 88%, 98%, 100% respectively for the related work. This counts for corresponding average accuracy at k value of 95% for both approaches. However, there are some inaccuracies in the semantic understanding output from the K-Parser due to parsing and conceptual mistakes. Which requires defining a set of rules or using commonsense knowledge to overcome these mistakes.

Table 4. Comparison of the Accuracy at k values - proposed versus related work in [32]

Approach	K value		
	1	5	10
Related work [32]	88%	98%	100%
Proposed	91%	96%	98%

From the presented experiments and results, the effectiveness of the proposed semantic understanding that integrates off-the-shelf, a developed rule-based relation extraction and external knowledge sources, as the main contribution to this work, has been proved. The accuracy of triple extraction is tracked as an important performance metric for evaluating the performance of the proposed automated triple extraction against manual/human triple extraction. The proposed approach managed to achieve an average semantic extraction accuracy of 97% for the different semantic types of relation triples: attributes and instance relation, multiword relation and semantic relation. The experiments also proved that the proposed semantic understanding managed to achieve an average accuracy at k value of 95% up to the 20th retrieval.

5. Conclusion

This paper presents a two-stage semantic annotation understanding for improving image search and retrieval. The presented approach relies on fusing rule-based relation extraction and external knowledge sources. At the first stage, the proposed approach relies on the Stanford parser and a developed rule-based relation extraction tools for extracting the different relationships from natural language annotations. The second stage integrates the ConceptNet commonsense knowledge base with the parser's output, generated at the first stage, to state accurate meanings to the extracted triples. The proposed approach has been evaluated against a testing set of natural language annotations obtained from the Flickr8k dataset. The results show that the proposed semantic annotation understanding achieves an average semantic extraction accuracy value of 97% for differently proposed semantic concepts and with retrieval accuracy of 95%. As a future work, large-scale testing sets will be used for validating the scalability of the proposed approach.

References

1. Schuster, Sebastian, et al. "Generating semantically precise scene graphs from textual descriptions for improved image retrieval." Proceedings of the fourth workshop on vision and language. 2015.
2. Johnson, Justin, et al. "Image retrieval using scene graphs." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
3. Amin, Ahmed E. "A Novel image retrieval framework based on knowledge based system." *International Journal of Intelligent Computing and Information Sciences* 14.2 (2014): 15-38
4. Liu, Ying, et al. "A survey of content-based image retrieval with high-level semantics." *Pattern recognition* 40.1 (2007): 262-282
5. Plummer, Bryan A., et al. "Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models." Proceedings of the IEEE international conference on computer vision. 2015.
6. Karpathy, Andrej, Armand Joulin, and Li F. Fei-Fei. "Deep fragment embeddings for bidirectional image sentence mapping." *Advances in neural information processing systems*. 2014.
7. Kiros, Ryan, Ruslan Salakhutdinov, and Richard S. Zemel. "Unifying visual-semantic embeddings with multimodal neural language models." arXiv preprint arXiv:1411.2539 (2014).
8. Mao, Junhua, et al. "Deep captioning with multimodal recurrent neural networks (m-rnn)." arXiv preprint arXiv:1412.6632 (2014).
9. Rui, Yong, Thomas S. Huang, and Shih-Fu Chang. "Image retrieval: Current techniques, promising directions, and open issues." *Journal of visual communication and image representation* 10.1 (1999): 39-62.
10. Banarescu, Laura, et al. "Abstract meaning representation (amr) 1.0 specification." Parsing on Freebase from Question-Answer Pairs. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: ACL. 2012.
11. Flanigan, Jeffrey, et al. "A discriminative graph-based parser for the abstract meaning representation." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014.
12. Werling, Keenon, Gabor Angeli, and Christopher Manning. "Robust subgraph generation improves abstract meaning representation parsing." arXiv preprint arXiv:1506.03139 (2015).
13. Wang, Chuan, Nianwen Xue, and Sameer Pradhan. "A transition-based algorithm for AMR parsing." Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015.
14. Lin, Dahua, et al. "Visual semantic search: Retrieving videos via complex textual queries." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

15. Zitnick, C. Lawrence, Devi Parikh, and Lucy Vanderwende. "Learning the visual interpretation of sentences." Proceedings of the IEEE International Conference on Computer Vision. 2013.
16. Chang, Angel, Manolis Savva, and Christopher D. Manning. "Learning spatial knowledge for text to 3D scene generation." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014.
17. Teney, Damien, Lingqiao Liu, and Anton van Den Hengel. "Graph-structured representations for visual question answering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
18. Wang, Yu-Siang, et al. "Scene graph parsing as dependency parsing." arXiv preprint arXiv:1803.09189 (2018).
19. Yang, Yezhou, Aloimonos Y. EDU U, and Cornelia Fermuller. "Deepiu: an architecture for image understanding." Adv Cogn Syst (2016).
20. Chen, Hua, et al. "Semantic image retrieval for complex queries using a knowledge parser." Multimedia Tools and Applications 77.9 (2018): 10733-10751.
21. Icarte, R. T., Baier, J. A., Ruz, C., & Soto, A. (2017). How a general-purpose commonsense ontology can improve performance of learning-based image retrieval. *arXiv preprint arXiv:1705.08844*.
22. Kríž, Vincent, et al. "Data extraction using NLP techniques and its transformation to linked data." Mexican International Conference on Artificial Intelligence. Springer, Cham, 2014.
23. Fader, Anthony, Stephen Soderland, and Oren Etzioni. "Identifying relations for open information extraction." Proceedings of the 2011 conference on empirical methods in natural language processing. 2011.
24. M. Schmitz, R. Bart, S. Soderland, and O. Etzioni, "Open language learning for information extraction," Proc. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp.523–534, ACL, 2012.
25. Kertkeidkachorn, Natthawut, and Ryutaro Ichise. "An automatic knowledge graph creation framework from natural language text." IEICE TRANSACTIONS on Information and Systems 101.1 (2018): 90-98.
26. Levy, Roger, and Galen Andrew. "Tregex and Tsurgeon: tools for querying and manipulating tree data structures." LREC. 2006.
27. Poria, Soujanya, et al. "Dependency-based semantic parsing for concept-level text analysis." International Conference on Intelligent Text Processing and Computational Linguistics. Springer, Berlin, Heidelberg, 2014.
28. Tsujii, Jun'ichi, James Henderson, and Marius Pasca. "Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning." Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012.
29. Angeli, Gabor, Melvin Jose Johnson Premkumar, and Christopher D. Manning. "Leveraging linguistic structure for open domain information extraction." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.
30. Rusu, Delia, et al. "Triplet extraction from sentences." Proceedings of the 10th International Multiconference" Information Society-IS. 2007.
31. Hodosh, Micah, Peter Young, and Julia Hockenmaier. "Framing image description as a ranking task: Data, models and evaluation metrics." Journal of Artificial Intelligence Research 47 (2013): 853-899.
32. Samih, Haitham, Sherine Rady, and Tarek F. Gharib. "Enhancing image retrieval for complex queries using external knowledge sources." Multimedia Tools and Applications 79.37 (2020): 27633-27657.
33. Anderson, Peter, et al. "Spice: Semantic propositional image caption evaluation." European Conference on Computer Vision. Springer, Cham, 2016.