International Journal on Emerging Technologies **11**(5): 567-571(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Using FrameNet to Improve Question Answering System Performance

Ibrahim Mahmoud Ibrahim Alturani[1] and Mohd Pouzi Hamzah[2]

[1]Al-Balqa Applied University, Jordan.
[2]Universiti Malaysia Terengganu, Malaysia.

(Corresponding author: Ibrahim Mahmoud Ibrahim Alturani)
(Received 17 September 2020, Revised 21 October 2020, Accepted 16 November 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Question answering (QA) system is one of the most important information systems models used for information extraction, sharing and reusing. In light of the importance of QA system, different methodologies for building QA systems have been proposed. QA system construction is a difficult and time-consuming process. Many efforts have been made to help designers to construct the QA system and to overcome the bottleneck of answer extraction. The information retrieval stage is a key stage in such a system, which retrieves potential passages based on their relevance to the question. In addition, information retrieval stage more efficient of QA systems to be able of analyzing texts, understanding and extracting semantic relationships between concepts to retrieve the most relevant passages. Accordingly, this paper introduces an approach of answer extraction of information retrieval from these corpus based on FrameNet, which work with ontology and WordNet, in order to recognize the semantic relationship to improve answer accuracy. Furthermore, this study used two other common methods are keywords based search and Ontology-based search using WordNet to compare the results of the proposed method in randomly selected questions from standard test collection. The experimental results of 110 questions show the degree of the performance of the proposed method. The suggested approach showed efficient information retrieval measurement results by contrasting recall and precision with other common approaches that were used in the same test collection. The results of the experiments given a score of 0.883 for precision and 0.818 for recall. We assess our model and demonstrate that it works better than the other two strategies for the answering task.

Keywords: Information retrieval, Question answering systems, Keyword search, Ontology and FrameNet.

I. INTRODUCTION

QA system recently attracts more interest. Because of the growing expectation of technology for Artificial Intelligence, People tend to view systems answering questions as a brand-new technology emerging today. However, the most effective systems use the conventional QA strategies with new techniques [4, 6, 12, 17, 23, 24], Includes a number of joint tasks, like TREC [22] and CLEF [9].

A QA system is designed to take an user's question as the input in natural language. Then on the question, some analysis is done, to find out what is being asked for. A QA system user is interested in a concise, comprehensible, and accurate answer that can refer to a phrase, sentence, or paragraph. Therefore, the modern QA systems rely on ontology to structure data in a machine-readable format and ensure satisfactory efficiency. There are generic ontologies, such as WordNet [14] and Cyc [11], but the majority of the applications need a specific field ontology for describing concepts and relationships in this area. So, users and the system may interact with each other using a shared understanding of a domain provided by an ontology [19].

II. RELATED WORKS

The importance of ontology in the categorization and structuring of domain knowledge is being exploited in QA system. Most ontology-based QA system categorize the relevant domain knowledge into ontology structures and then compose a list of questions and answers based on the ontology that has been created. This can be seen from several studies relating to the creation of QA system based on ontology, including AQUA, QASYO, Pythia and NLQA projects.

AQUA project developed by Vargas-Vera and Motta [21]. It works based on the combination of the techniques of natural language processing, ontology, logic and information retrieval. The system has been tested to answer academic people and organizations' questions. Here, ontology is used to formulate the natural language query in the ontological structures.

Similarly, the Moussa research YAGO query [15], named QASYO. QASYO is a question-answering system at the sentence level that also integrates into a unified framework natural language processing, ontologies, and information retrieval technologies.

Meanwhile, Pythia is a QA framework based on ontology given by Unger and Cimiano [20]. It uses deep semantic analyzes on questions to produce a translation into formal queries with respect to a grammar composed in ontology. NLQA is another QA system that has been suggested by Athira *et al.*, [3]. It uses domain ontology that dynamically populates each document in the collections.

These ontology-based QA systems highlight the essential role of ontology to improve the retrieval accuracy of the answers based on the questions asked in the knowledge domain. Therefore, this research

describes our challenge to extract more accurate answer, which acquires knowledge through the documents and questions, given knowledge drawn from the documents representation method. So, in this work, the new approach is proposed to model knowledge representation with employ WordNet and FrameNet to design a retrieval model for an answer.

III. BACKGROUND: ONTOLOGY, WORDNET AND FRAMENET

A. Ontology

Ontology is a well-known concept in the field of knowledge engineering and AI. Gruber's most common description of ontology in IT and the AI community [8], which notes that: "An ontology is a formal, explicit specification of a shared conceptualization". Ontology was used for large-scale applications such as information retrieval, knowledge management, information integration and e-learning [25][26]. Since ontologies in these areas are significant, various methodologies have been proposed in building ontologies. The manual construction of ontologies is subjective, very hard, unclear, time-consuming, confused and can cost a lot. To overcome these difficulties, the research area known as ontology learning is generated.

Ontology learning techniques rely on methods from various fields such as machine learning, knowledge acquisition, Natural Language Processing (NLP), statistics, and information retrieval. There are three key approaches for learning ontology from the text: Linguistic, Statistical, and Hybrid approaches. In this paper use Linguistics-based techniques which are mainly dependent on NLP tools. Several of the techniques are part-of-speech (POS) tagging, sentence processing, analysis of the syntactic structure and analysis of the dependency. Other techniques are based on the application of semantic lexicon, lexico-syntactic patterns [10]. Also, the conceptual relationships in ontology can be identified by matching to predefined rules to extract the relationships between terms.

B. Wordnet

Recently, there has been a growing interest in more in-depth semantic analysis for practical NLP tasks, in particular as a basis for open-domain information access. Large-scale lexical-semantic resources, such as WordNet [7] have been developed and put to use for approximate semantic modeling in many applications. Therefore, WordNet [14], an electronic lexical database of English, is considered to be one of the most useful tools in computer linguistics for researchers, text analysis, and several other related areas. The design is prompted by the human lexical memory's psycholinguistic and computational theories. English nouns, verbs, adjectives and adverbs, each reflecting a sub-lexicalized concept, shall be grouped in sets of cognitive synonyms (sunsets). Synsets are interlinked across conceptual-semantic and lexical relationships, resulting in more than 200,000 word-sense pairs of lexico-semantic interlinked concept networks.

C. Framenet

A frame describes a conceptual structure or prototypical situation together with a set of semantic roles, or frame elements (FEs), that are involved in the situation. Using Frames to represent documents, groups words and expressions (lexical units) into semantic classes (frames) and lists semantic roles for each frame. This type of lexical-semantic information is particularly useful for information access tasks, that use in QA. We are currently investigating the use of FrameNet frames for building partial text meaning representations [5], to be used in QA systems. Semantic representations building on frames provide normalization over surface realizations (e.g. verb/nominalization) and thus a sensible granularity for this system.

IV. QUESTION ANSWERING FRAMEWORK

The general architecture of a QA framework consists of three major components as shown in Fig. 1: processing of questions, processing documents and answering [13]. The component of question processing receives the user question form in the form of a natural language. It then runs multiple processes to determine the type of question and to create a computational query. The query is forwarded to the document processing where the relevant data is extracted and the candidate answers are retrieved. The last step is to compare the answers and identify the exact answers based on the type of question.

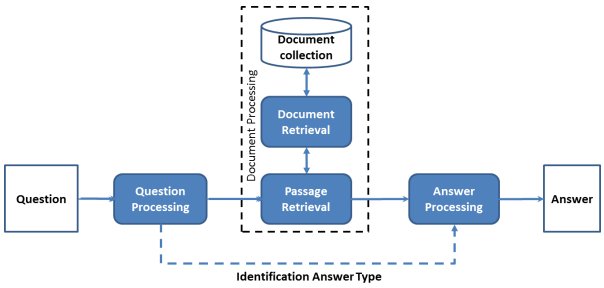

Fig. 1. General architecture of a QA framework.

Most QA researchers were heterogeneous somehow with regard to their system architecture, approaches, scope, assessment metrics, etc. On the other hand, researchers concentrated primarily on the question classification, the retrieval and the answer extraction, which combines information retrieval with information extraction methods for the identification of a group of possible candidates and then for the final approval of the answers using some ranking scheme.

V. PROPOSED SYSTEM

We propose a new approach for information retrieval stage in QA system from documents, or documents representation task, it is composed of several phases, each one made up of one or several resources as shown in Fig. 2.

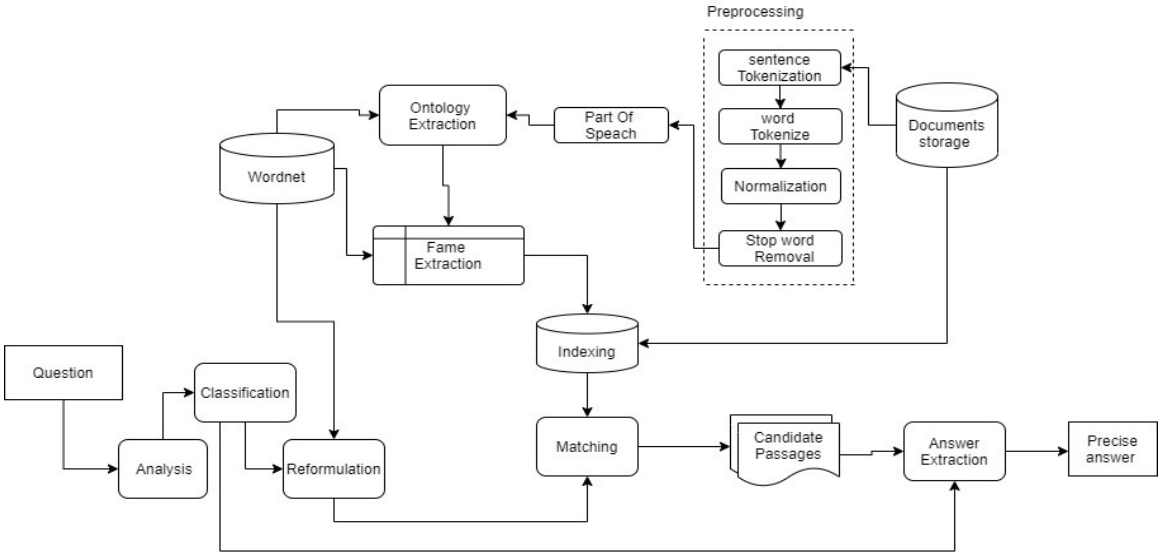

Fig. 2. Proposed QA system architecture.

FrameNet and WordNet help QA systems in the comprehension of the semantics of texts, and therefore, finding procedure to link FrameNet and WordNet involves challenges related to a better understanding of the natural language. The researchers use deep analyzes to improve the performance of different QA systems. So, the coherence and integration are achieved by exploiting the specific characteristics of each lexico-semantic resource, with particular emphasis on the explicit, formal semantic relationships within each.

The proposed approach mechanism was clarified by the pseudocode, which explained as follows:

Input

Documents collections, and General questions;

Output

Retrieve the passage that relevant the question;

For each question;

Begin:

Phase1: Question processing

1. *Extract the question terms;*
2. *Determine the question type;*
3. *Determine the question domain;*

Phase2: Documents processing

Go to indexed documents, and then:

1. *Extract all candidate keywords (concepts) from each document;*
2. *Compute the similarity between document keywords using WordNet;*

Frames process

3. *Representing the document as a set of classes (frames) and connections between them;*
4. *Indexing each document frame;*

Phase3: Answer processing

1. *While similarity (question domain, Doc domain) do: Search in all the documents in the same domain; Match (question Term, Document keywords);*
2. *Retrieve all of related passages;*
3. *Select the most relevant passage;*

Three main phases are included in the proposed approach, as shown in Fig. 2; each phase comprises subtasks as follows:

1) In documents processing

- The pre-processing phase, including normalization, tokenization, and stop-words removal.
- Using POS tagging, ontology and WordNet to extract the keywords in each document.
  It is very important to refer to the key topics (frames) of the document, where the search process is so efficient. The extraction of keywords also provides the target document domain that enables the system to extract the relationships between concepts and knowledge.
- Classify the concepts to extract general semantic FrameNet based on WordNet, which in class hierarchies, several concepts linked to each other are included. It enables the determination of the relevant concepts in ontology and their semantic relations.
- Indexing each document FrameNet

2) In user question processing

- Apply the preprocessing techniques like document preprocessing.
- Extract the question terms.
- Determine the question type

3) In Answer processing

- Using the VSM to compute the similarity between the terms of the question with related documents through determining the semantic domain in both, documents and user question.
- Use Cosine similarity to match documents relating to the user query.
- Rank the related documents.

A. Preprocessing Phase

Text pre-processing is a crucial component of any NL application processing, such as information retrieval. The input of this stage is a text of the standard NL. It used to make a text more comprehensible and computer-readable. Text pre-processing includes the following tasks:

1. Tokenization is the separation of the plain text into phrases, words, symbols or other important parts known as tokens. The list of tokens is used as input for different processing, such as parsing or text mining.
2. Normalization: It is the process of transforming text into a single canonical form, by applying certain linguistic models to text tokens.
3. Stops words removal: This method includes eliminating terms that are meaningless and recur very often. Stop words are used to connect words in a sentence together. They are common words, such as 'and' and 'are' and 'this' frequently used.

B. Document Representation

The proposed model uses NL processing techniques to analyze the contents of a document works as follow:
1. Based on its context and definition, POS tagging is the process of marking up a word in a text to a corresponding part of a speech tag.
2. Extracting keywords is a way of discovering the terms that represent the document and the content of the text.
3. Convert the words found in the sentence into concepts using WordNet ontology (word-level knowledge).
4. Use the parser to access the relation between words, to candidate the keywords. Ontology includes several concepts which related with each other in class hierarchies. It concerns to determine the relevant concepts in an ontology, and semantic relations between of them.
5. Deriving FrameNet, which a resource that contains information about different semantic classes (frames), in which semantic relations are syntactically realized in natural language sentences (i.e lists semantic roles). For example, the phrase "low fat milk" should be analyzed with "milk" evoking the food frame, where "low fat" fills the descriptor of that frame and the word "milk" is the actual physical_object.

Example:

Frame: Food

Semantics: Physical_object
noun: almond, apple, banana, …

Frame Element: constituent_parts, descriptor, type

C. *Semantic Similarity Calculation*

Semantic similarity is a method of measuring the degree of similarity between sentences. In other words; Semantic similarity is a process of measuring the distance between two documents by extracting the semantic distance between terms or concepts in those documents. In fact, it is not only measured at the document level but can also be used in the various sentence and word levels. WordNet, ontologies and other thesauruses have been used in semantic similarity to get the semantic similarity between documents [2]. So, semantic similarity is used to improve the accuracy of search by understanding the intent of the user and the meaning of the terms in the searching sentence. As a result, it uses to produce highly relevant searching results.

VI. EVALUATION METRICS

The evaluation of QA systems is performed according to the criteria used to judge the correctness of the answer. Relevance, correctness, conciseness and completeness are the most criteria for answer evaluation [16]. Based on those criteria, an answer extracted from a document contains three different judges: Correct, Unsupported or Inexact. Different measurements have been applied over the years, but the measures used are the most commonly used in evaluations [1]:

$$\text{precision} = \frac{\text{number of correct answers}}{\text{number of questions answered}} \quad (1)$$

$$\text{recall} = \frac{\text{number of correct answers}}{\text{number of questions to be answered}} \quad (2)$$

VII. EXPERIMENTAL RESULT

The proposed model tests and validates on 110 questions that selected randomly from smith dataset [13] that was generated from the three datasets [18]: Yahoo Non-Factoid Question Dataset, TREC 2007 QA Data and a Wikipedia dataset. Also, at the same time, we have applied two approaches to the same sample questions: keywords-based search and ontology-based search using WordNet. Table 1 shown below contains the experimental results that retrieved from the proposed model using FrameNet and different QA approaches.

Table 1: Experimental Results of Three types of model.

	Measures	
Type of model	**Precision**	**Recall**
Proposed model	0.883	0.827
keywords based search	0.832	0.764
Ontology-based search using WordNet	0.857	0.818

Table 1 is shown the obtained results from applying the keywords based search. The experimental results were about 110- questions. The precision was 0.832. Also, the same sample is conducted by the Ontology-based search using WordNet the precision was 0.857. In contrast, the same sample is conducted by the proposed method, where the precision was 0.883. The experimental results of 110- questions show the degree of the performance of the proposed method. The proposed approach showed a good result of QA system measurement through compare recall, and precision with other studies that have been applied in this area. This indicates a good improvement of the performance of QA system. In addition to that, building relational FrameNet shows a significant enhancement than other common models that conducted in QA system.

VIII. CONCLUSION

In recent years, there are many studies on QA system research to meet the challenges due to information explosion in this information and communication technology era. It is important to notice that one of the most important features of QA Systems is their ability to produce exact answers. So, to solve these challenges, should enhance the QA systems accuracy, and recall measurement. One of these suggested solutions to improve the efficiency of QA system is a build FrameNet. A building FrameNet helps QA systems to access the target information domain. The experimental results of 110 questions showed the degree of the performance of the proposed model. The proposed model showed an efficient result of information retrieval measurement through compare recall and precision with

two different QA model that has been applied in this same test collection.

IX. FUTURE SCOPE

This paper's major contribution is that using the FrameNet and Wordnet can support the process of answering the question in resolving word disambiguation through extracting new semantic relation between concepts in order to access more correct answers.

Since only a prototype of the proposed system is implemented, in a future work we look forward to implementing a complete system. We focused on designing and developing IR stage of QA systems. As future work, the same can be applied for the other two stages of QA systems; that is "Question processing" and "Answer processing". The success of our proposed prototype also encourages us to look for ways to extend the ontology by adding more data and semantic information And therefore to cover the largest amount of questions and obtain more accurate results.

REFERENCES

[1]. Allam, A. M. N., & Haggag, M. H. (2012). The question answering systems: A survey. *International Journal of Research and Reviews in Information Sciences (IJRRIS), 2*(3).

[2]. Alturani, I. M. I., & Bin Hamzah, M. P. (2018). A New Approach For Open-Domain Question Answering System. *International Journal of Computer Science and Network Security, 18*(6), 100–103.

[3]. Athira, P. M., Sreeja, M., & Reghuraj, P. C. (2013). Architecture of an ontology-based domain-specific natural language question answering system. *International Journal of Web & Semantic Technology, 4*(4), 31.

[4]. Bian, J., Liu, Y., Agichtein, E., & Zha, H. (2008). Finding the right facts in the crowd: factoid question answering over social media. Proceedings of the 17th In

[5]. ternational Conference on World Wide Web, 467–476.

[6]. Burchardt, A., Erk, K., & Frank, A. (2005). A WordNet detour to FrameNet. *Sprachtechnologie, Mobile Kommunikation Und Linguistische Resourcen, 8,* 408–421.

[7]. Cui, H., Sun, R., Li, K., Kan, M. Y., & Chua, T. S. (2005). Question answering passage retrieval using dependency relations. Proceedings of the *28th Annual International ACM SIGIR Conference on Research and Development in Information* Retrieval, 400–407.

[8]. Fellbaum, C. (1998). WordNet: An electronic lexical database MIT Press. Cambridge, Massachusetts.

[9]. Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies, 43*(5–6), 907–928.

[10]. Harmann, D., Braschler, M., Hess, M., Kluck, M., Peters, C., Schuble, P., & Sheridan, P. (2000). CLIR Evaluation at TREC Cross-Language Information Retrieval and Evaluation.

[11]. Ivanova, T. (2012). Ontology Learning Technologies-Brief survey, trends and problems. Proceedings of the *International Conference on Information Technologies,* 245–255.

[12]. Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 38(11), 33–38.

[13]. Lin, D., & Pantel, P. (2001). Discovery of inference rules for question-answering. *Natural Language Engineering, 7*(4), 343–360.

[14]. Mahmoud Ibrahim Alturani, I., & Pouzi Bin Hamzah, M. (2019). An Efficient Semantic Analysis Technique for the Question Answering Systems. *Journal of Engineering and Applied Sciences, 14*(22), 8289–8292. https://doi.org/10.36478/jeasci.2019.8289.8292

[15]. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*(4), 235–244.

[16]. Moussa, A. M., & Abdel-Kader, R. F. (2011). Qasyo: A question answering system for yago ontology. *International Journal of Database Theory and Application, 4*(2), 99–112.

[17]. Pease, A., Niles, I., & Li, J. (2002). The suggested upper merged ontology: A large ontology for the semantic web and its applications. Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, 28, 7–10.

[18]. Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a question answering system. Proceedings of the *40th Annual Meeting of the Association for Computational Linguistics,* 41–47.

[19]. Smith, N. A., Heilman, M., & Hwa, R. (2008). Question Generation as a Competitive Undergraduate Course Project. 4–6.

[20]. Soo, V. W., & Lin, C. Y. (2001). Ontology-based information retrieval in a multi-agent system for digital library. *6th Conference on Artificial Intelligence and Applications,* 241–246.

[21]. Unger, C., & Cimiano, P. (2011). Pythia: Compositional meaning construction for ontology-based question answering on the semantic web. *International Conference on Application of Natural Language to Information Systems,* 153–160. Springer.

[22]. Vargas-Vera, M., & Motta, E. (2004). AQUA–ontology-based question answering system. *Mexican International Conference on Artificial Intelligence,* 468–477. Springer.

[23]. Voorhees, E. M., & Harman, D. K. (2005). TREC: Experiment and evaluation in information retrieval (Vol. *63*). MIT press Cambridge.

[24]. Xue, X., Jeon, J., & Croft, W. B. (2008). Retrieval models for question and answer archives. Proceedings of the *31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* 475–482.

[25]. Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. Proceedings of the *2003 Conference on Empirical Methods in Natural Language Processing,* 129–136. Association for Computational Linguistics.

[26]. Zouaq, A., Gasevic, D., & Hatala, M. (2011). Towards open ontology learning and filtering. *Information Systems, 36*(7), 1064–1081.

[27]. Zouaq, A., & Nkambou, R. (2010). A survey of domain ontology engineering: Methods and tools. In Advances in intelligent tutoring systems (pp. 103–119). Springer.

How to cite this article: Alturani, I. M. I. and Hamzah, M. P. (2020). Using FrameNet to Improve Question Answering System Performance. *International Journal on Emerging Technologies, 11*(5): 567–571.