Surveying Question Answering Systems: A Comparative Study

Amr M. Sauber*1, Nada Ali*2, Passent El Kafrawy **3

*Math & Computer Science Department, Faculty of Science, Menofia University, El Menofia, Egypt

1amrmausad@computalityit.com|amr@science.menofia.edu.eg

2nadaali@science.menofia.edu.eg**School of information Technology and Computer Science, Nile University, Egypt

3pelkafrawy@nu.edu.eg|basant.elkafrawi@science.menofia.edu.eg

Abstract:
Question Answering (QA) systems are considered an advanced form of information retrieval that enables asking specific questions by humans and the system infer answers using natural language queries with deeper knowledge understanding mechanisms from huge amount of data instead of having the user search all documents himself. As "what a user really wants" is often a precise answer to a question by consulting documents on the web or using special knowledge base.
Although many systems have been developed over the past years, it remains a challenge that most systems yet require improvements to increase the accuracy for correct interpretation of the question and provide a good exact answer to user questions. This paper presents a short study of the generic QA framework and aims to survey some of the current state-of-the-art Question Answering systems in different domains and makes a comparison between all of these systems based on some identified criteria in literature.

Key words: ***Question Answering Systems, Natural Language Processing, Information Retrieval, Knowledge Representation.***

1 Introduction

Question Answering systems (QAS) have emerged as robust tools for automatically answering questions asked by humans in natural language using either a structured database or a collection of natural language documents. It is considered an advanced form of Information Retrieval (IR) [1]. The demand for this type of system increases tremendously every single day since it delivers short, precise and, question-specific answers rather than full documents. Thus, users do not need to afford seeking one-by-one web pages to find the information needed. Question Answering Systems are classified into two types according to the domain they cover, Closed-domain and Open-domain question answering. Closed-domain question answering deals with questions within a specific domain (for example, automotive maintenance or medicine`), and can take advantage of domain-specific knowledge frequently formalized in ontologies. Open-domain question answering should deal with questions about almost anything, therefore, can only rely on general ontologies and world knowledge. However, these systems often have much more data available from which the answer is to be extracted.

There are many challenges of question answering as diverse question sets require different approaches/strategies to answer. Some answers are spread across different sections of the paragraph, handling different domains and different classes of questions is essential. Researchers still work on these challenges to reach better results in answering questions. This study aims at identifying Question Answering techniques, tools, systems and compares these systems according to many factors.

This paper is structured as follows; section two describes the background of QA systems. In section three literature survey is discussed in detail. Finally, the discussion and future work are concluded in section four.

2 Background

Question answering systems' main task is to answer the natural language questions automatically. It requires understanding of text, linguistics and common knowledge. In order to understand the question answering subject, we first need to define the associated terms. *A Question Phrase* is the term that says what to be searched for an answer. The term *Question Type* refers to a categorization of the question based on its purpose. The term *Answer Type* refers to a class of objects which are sought by the question. *Question Focus* is the property or entity being searched by the question. *Question Topic* is the object or event that the question is about. *Candidate Passage* can broadly be defined as anything from a sentence to a document retrieved by a search engine in response to a question. *Candidate Answer* is the text ranked according to its suitability to be an answer.
Previous studies mostly defined architecture of Question Answering systems in three stages Question Analysis, Document Retrieval and Answer Extraction [1-3], [13] as explained in Fig.1.

A. Question Analysis Stage

The objective of this process is to understand the question posed by the user. It begins with analyzing the user's question then classifying the question and determining its type. After that extracting keywords from the question to form a list of keywords to be passed to the Document Retrieval stage. Finally extracting Names Entities from the question.

B. Document Retrieval Stage

The analyzed question is submitted to the Document Retrieval stage, which in return retrieves a ranked list of relevant documents. The document retrieval stage usually relies on one or more information retrieval systems to gather information from a collection of document corpora which almost always involves the World Wide Web as at least one of these corpora. The documents returned by the information retrieval system is then filtered and ordered. Therefore, the main goal of the document retrieval stage is to search for candidate documents depending on the focus of the question and create a set of candidate ordered paragraphs that contain the answer(s).

C. Answer Extraction Stage

As the final stage in the QA architecture, this stage is the most challenging task in a Question Answering system. It is responsible for processing candidate ordered paragraphs containing the answer, identifying, extracting and validating answers from the set of ordered paragraphs passed to it from the Document Retrieval stage [3]. The answer must be a simple answer for the question, but it might require merging information from different sources, summarizing, and dealing with uncertainty or contradiction.
Most Question Answering Systems follow these three stages. However, they may differ in how they implement every stage.

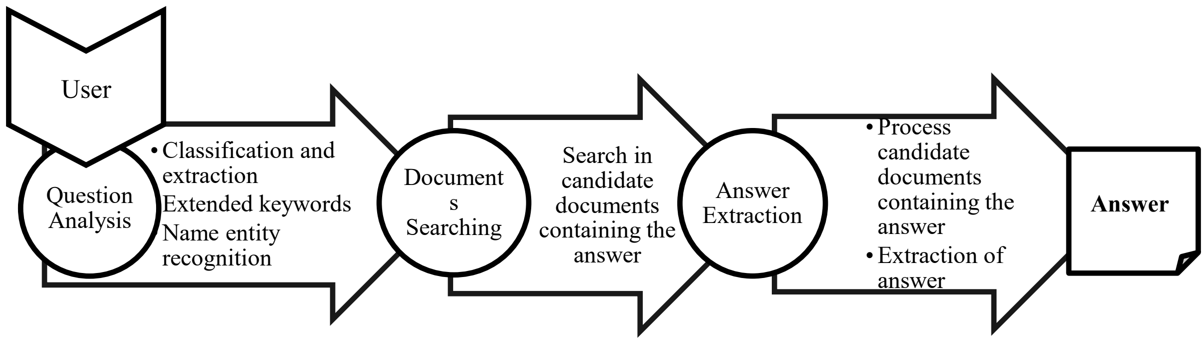

Fig.1 Stages of question answering system for Web of documents and text

3 LITERATURE SURVEY

Besides the main architecture, QA systems could be classified by the paradigm each type implements:

A. Information Retrieval (IR)QA

IR systems use search engines to retrieve a ranked list of relevant documents that contain answers and then apply filters and ranking on the recovered passage. The mission of the IR is to perform a selection of paragraphs that are considered relevant to the input question then rank these paragraphs to extract the correct answer. [14-16]

B. Natural Language Processing (NLP) QA

NLP use linguistic intuitions and machine learning methods to extract answers from retrieved snippet. NLP tools are used to analyze the question in order to make it easy to reach to the correct answer. Name Entity Recognition (NER) is type of NLP tools used to extract entities from questions. Part of speech tagging (POS) is the process of marking up a word based on both its context and its definition [17, 18]. They were often written in the form of regular expressions to increase precision.

C. Knowledge Base (KB)QA

This is to find answers from structured data source (a knowledge base) instead of unstructured text. Standard database queries are used in replacement of word-based searches. This paradigm, make use of structured data, such as ontology. Ontology describes a conceptual representation of concepts and their relationships within a specific domain. Ontology can be considered as a knowledge base which has a more sophisticated form than a relational database. To execute queries in order to retrieve knowledge from the ontology, structured languages are proposed and one of them is SPARQL. Nowadays, there are KBs about media, publications, geography, life-science and more [8]. The core purpose of a QA system over KBs is to retrieve the desired information from one or many KBs, using natural language questions. This is generally addressed by transforming a natural language question into a SPARQL query.

D. Hybrid(H) QA

This type of QA is a high-performance system that makes use of as many types of resources as possible, especially with the prevailing popularity of modern search engines and enriching community contributed knowledge on the web. A hybrid approach is the combination of IR QA, NLP QA and KB QA [19, 20].

In this study we focus on a combination of different techniques and approaches.

Wisesa et. al in [4] described an ontology based QAS capable of answering the questions asked by users in Bahasa Indonesia (Indonesian language) correctly, according to question context, even though with incomplete question sentences input. For example: the sentence with no question word or no question object, or no question adverb. First, it designed an ontology that used a property named keyword to find the question objective; through the keyword property, it can determine which class or instance object that is meant in a user's question. Each class or instance can contain many keyword properties. Second, system architecture was designed to transform questions from natural language into instructions that can be understood by the computer to find answers by applying these steps: Stemming, Stopword Remover, Tokenizing, POS Tagging, Keyword Identification, and SPARQL Query Formation. The stemming process used to eliminate the affix in a word. The stopword remover process used to eliminate unnecessary words. This is done by comparing each word in question sentence with stopwords dictionary. The Tokenizing process will change the question sentence to lowercase, delete characters and symbols that are not needed. The POS Tagging process will split the sentences into the words and classify those words based on its type. POS Tagging stage also identifies the words that might be keywords. SPARQL Query Formation process describes how the SPARQL queries are built, which used to draw answers from the ontology knowledge base.

The system was tested using three groups of question sentences, namely: complete question sentences, incomplete question sentences without question objects, and incomplete question sentences without question objects and question words. The test case shows that the keyword plays an important role in answering a question especially in Bahasa Indonesia, even though the question does not have a question word (QW) or object (O). It also shows that it is important in an ontology design to relate two instances that should have linked using a predicate.

Subburaj et. al in [5] proposed a QAS system designed where ontology was constructed to retrieve the exact answer for the users' question. This system works on a domain specific ontology that was incorporated to improve e-learning applications. It used tree-tagger parser to extract word patterns (noun/verb/adverb/adjective) from the question. Then it checked the question patterns for the definition type questions and stored them in the question database. After that it extracts noun/verb/adverb/adjective found in the nodes of the ontology to find Class, Subclass, etc. The ontology file is converted to JSON-LD (Java Script Object Notation-Linked Data) to reduce the time complexity of searching in the ontology during retrieval of answers as JSON-LD is easy for machines to parse and generate. It finds the synonyms of the noun/verb/ adverb/adjective using WordNet and check whether the synonyms of noun/verb/adverb/adjective of the question match with the ontology Class/ Sub Class. Then query is built and retrieves the answer from JSON-LD file. Finally check the relevant answer by using semantic similarity ranking if found, then return the results to the user otherwise return "answer not extracted". The listed questions and its answers are stored in database for future reference. If another user inputs the same question in different type, the proposed system identifies the question as well as the answer from the database. This model was tested with definition type of questions of about 200 patterns of data set. The accuracy rate of the system was 86%.

Nyberg et. al in [6] described an attempt to address the Ideal Answer Generation task of the sixth edition of the annual BioASQ challenge, which is a large-scale semantic indexing and question answering challenge in the biomedical domain. This was done using ontology-based retrieval and a neural learning-to-rank approach, combined with extractive and abstractive summarization techniques. The hypothesis explored that a combination of these summarization techniques will provide better overall performance on the ideal answer task. The used dataset includes 2,251 questions accompanied by a list of relevant documents and a list of relevant snippets extracted from each of these documents. The first step of the system starts by pre-processing the question to enrich it with features derived from standard NLP techniques such as Part of Speech (POS) tagging, Named Entity Recognition (NER) and Medical Entity Recognition (MER). An ontology based retrieval system is used in order to retrieve relevant snippets for the question. It used entity and relation extraction techniques to represent and compare the content of questions and candidate answers that improved the answer selection

process and in turn provide more relevant answers. The retrieved snippets are combined with the given BioASQ snippets for the question, and passed to the ranking module. In this module Learning to Rank (LETOR) method was used to rank these snippets. The ranked snippets are then inputted to the sentence selection module which used similarity measures to select the most relevant and least redundant snippets. The selected sentences were then passed to the summarization module, which produced the final summary using extractive and abstractive summarization techniques.
This model was tested using different question types by both the pre-trained model and fine-tuned model in the BioASQ dataset. Some answers generated by the fine-tuned model are more readable than the pre-trained model.

Another type of Question answering systems is explained in [7] which is based on Quran ontology. This question answering framework consists of several components as pre-processing, morphology analysis, question classification, query expansion, document processing, and answer extraction. The input for the system was user queries in a factoid question. This query was processed by pre-processing phase. The aim of automatic pre-processing phase was to prepare the words with an appropriate form by tokenization, disposing the punctuation, and applying stop words removal. The output from this phase was used as the input for the morphology analysis phase where stemming operation was applied with stemmer algorithm for Indonesian text. The output from this phase was a set of keywords in root word form. In question classification stage questions were classified and the answer type was determined using Radial Basis Function Networks (RBFN) algorithm. The keywords from the morphology analysis stage were used in query expansion phase. This phase aimed to expand the keywords using Indonesian WordNet. The output from query expansion stage was the origin keywords and their synonym. The document processing stage performed execution of SPARQL query with some parameters, i.e. the answer type and keywords from the previous phase. This answer type would access the class of thematic topics within Quran ontology, and the keywords were used to extract the instances based on the class (Quran verses). This execution generated answers candidate, i.e., Quran verses and their Tafsir. At the answer extraction stage, word matching scoring technique was applied to rank the verses and their Tafsir. This technique computed the number of similar words between the expanded query and the Tafsir of the verses, then ranked all these verses score. The best answer was determined based on the verse with the highest score. This framework worked only for small training corpus like Quran which was one of its drawbacks

The Question answering system in [8] is different as its approach was querying five KBs: Wikidata, DBpedia, MusicBrainz, DBLP and Freebase in five different languages namely English, German, French, Italian and Spanish.
In Question Answering Systems Field less than 10% of the approaches were applied to more than one language and 5% to more than one KB. This happens because of the heavy use of NLP tools or NL features.
This system addressed the challenge of using more than one KB and multilingual. It depended on that many questions can be understood from the semantics of the words in the question. By knowing the semantics of the words the intention of the user was deduced. They could be correctly interpreted without considering the syntax as the semantics of the words was sufficient for them. The system consisted of four stages: question expansion, query construction, query ranking and response decision. In question expansion stage all entities, properties and classes were identified, which the question could refer to. This was done by finding an n-gram from question and excluding stop words from it. Then resources of these n-grams were extracted. In query construction stage a set of queries that represent interpretations of the question were constructed. The semantics of the words were used into the particular KB. It could tell what the most probable interpretation of the question was. A set of possible SPARQL queries were computed (candidates). In ranking stage the candidates from previous stage needed to be ordered by their probability of answering the question correctly. They were ordered in descending order with respect to F-measure. The top ranked query was executed against a SPARQL endpoint, and the result was computed. In answer decision stage an additional confidence score was computed. A model was constructed consisted of SPARQL queries (training set) and the labels true or false. True indicated if the F-score of the SPARQL query was bigger than a threshold and false otherwise. Once the model was trained, it could compute a confidence score. The question was answered when its confidence score was above a threshold otherwise it was not answered. This approach could also be extended to multiple KBs. In the query expansion step, one had just to take in consideration the labels of all KBs. In the query construction step, one could consider multiple KBs as one graph having multiple unconnected components. The query ranking and answer decision step were literally the same. Using more than KB took more time to answer the question which was one of its limitations. This system needs to work in parallel to achieve acceptable response time. The system should contain a multi-language translator to respect the meaning of words in different language
Alaoui et. al in [9] described a fully automated semantic biomedical QA system named SemBioNLQA which accepted a variety of natural language questions as input, and outputs both short precise answers and summaries as results. It was able to handle the kinds of yes/no questions, factoid questions, list questions and summary questions that were commonly asked in the biomedical domain. This system consisted of four stages: question classification and query reformulation, document retrieval, passage retrieval and answer extraction stages.
Question classification and query reformulation aimed to identify the question type and therefore determine the expected answer format based on handcrafted lexico-syntactic patterns and support vector machine (SVM), to see whether the answer should be a biomedical entity name or a short summary [“yes” or “no”]. To build the query for document retrieval, MetaMap was used for mapping terms of the question to a unified medical language system (UMLS) meta-

thesaurus in order to extract biomedical entity names. Then the extracted entities were concatenated with the “+” operator. The UMLS is a repository of biomedical concepts.

Document retrieval stage was executed by retrieving the set of relevant documents that contained the answer to a given query constructed by the previous stage. The correct answers could be found when the set of retrieved documents was determined correctly. The constructed query was sent to PubMed search engine by calling E-utilities Web service from PubMed to retrieve relevant documents from the MEDLINE database. This database was considered as the authoritative source of medical evidence for medical professionals, biomedical researchers, and many other users. The returned documents were re-ranked based on the UMLS similarity. The idea was to compute the sum of the semantic similarity scores between biomedical concepts of a given question and each title of the returned documents. After re-ranking the returned documents, the D top-ranked ones were kept and used as an input data to the following stage.

Passage retrieval stage retrieved and fixed the P top-ranked passages from the retrieved documents. These top-ranked passages were served as answer candidates and the SemBioNLQA system retrieves the answer from them. Abstracts of the D top-ranked documents were forwarded to Stanford CoreNLP sentence splitter so as to segment them into sentences. A passage in SemBioNLQA was considered as one sentence. These set of sentences were preprocessed by applying tokenization, removing stop words and applying Porter’ stemmer. The MetaMap program was used for mapping both biomedical questions and candidate passages to UMLS concepts in order to extract biomedical concepts. Using stemmed words and UMLS concepts as features, the set of candidate passages were ranked and kept the P top-ranked ones using BM25 as retrieval model. The P top-ranked passages (p1, p2... pp) was used as input to answer extraction.

Answer extraction was the stage where the precise answer had to be extracted from the candidate answers retrieved by the passage retrieval component. The output from this module was a short accurate answer to the user's question. The appropriate answer was selected according to the question type that was automatically detected by the question classification stage. After retrieving the P candidate answers and identifying the question category to a given biomedical question, the SemBioNLQA system applied the appropriate answer extraction method to extract final answers as SemBioNLQA deals with four types of questions.

1) *Yes/no answer extraction method*: was based on SENTIWORDNET: a lexical resource for sentiment analysis and opinion mining. The Stanford CoreNLP tools were used for tokenization and POS tagging one by one the P retrieved candidate answers. Then, each word of the candidate answers was assigned its SENTIWORDNET score. Finally, the decision to output “yes” or “no” depended on the number of positive or negative candidate answers: “yes” for a positive final sentiment candidate answers score and “no” for a negative one.
2) *Factoid answer extraction method:* was based on UMLS meta-thesaurus, BioPortal synonyms and term frequency metrics. The P candidate answers were first mapped to the UMLS meta-thesaurus to extract the set of biomedical entity names Es. Next, the obtained set of biomedical entity names were ranked based on term frequency metrics TF (ei, Es); ei the number of times entity name Es appeared in the set of biomedical entity names. The answers were located in the first and second candidate answers. Then, synonyms for each of the T top-ranked entity names were retrieved using Web services from BioPortal. Finally, the T top-ranked biomedical entity names and their T top synonyms were displayed as answers, excluding entities mentioned in the question as most entities that appear in questions were not part of the answers.
3) *List answer extraction method:* was similar to the one described for factoid questions. The main difference between factoid and list questions was that the former required a single list of answers while the latter expect a list of lists of entity names, numbers, or similar short expressions.
4) *Summary answer extraction method*: required for formulating short summaries (ideal answers) of relevant information. These ideal answers of questions were formed by concatenating the two top-ranked passages which were retrieved by the proposed passage retrieval approach.

The results of the evaluation had shown that SemBioNLQA achieved better results and succeeded at answering the majority of randomly selected BioASQ questions when compared with other biomedical systems. However, there were some mistakes that the system still could not fix. The current form of the system was not able to provide answers to some questions especially for those expected a number as an answer instead of biomedical entities. On the other hand, if the set of retrieved documents, passages and the type of a given question were not identified correctly, further processing steps to extract the answers would inevitably fail too.

Soesanti et. al in [10] presented a literature review analyzed the state-of-the-art methods used by Question answering systems in question analysis, document retrieval and answer extraction stages in recent years.

1) *Methods Used in Question Analysis stage:* In [21] Question classification was considered as one of the most important phases of Question analysis stage. It helped the system find an accurate answer. There were two main approaches for question classification, which were rule-based approach and learning based approach. However, in this review approaches were categorized into Natural Language Processing (NLP) approach (lexical, semantic, and pattern-based), machine learning approach, and hybrid approach (the combination of NLP and machine learning).

- In NLP approach the question was classified by a rule based method into coarse-fine categories [22]. As the system extracted a list of keywords and then labeled them into a primary keyword, secondary keyword, and question object. Next, the question classification phase matched the keyword with the pre-defined pattern. In [23] another type was classified using a semantic approach to create a query expansion in English Quran Question answering system. They used available WordNet synonyms and Islamic synonyms to create several questions that would be used in verses retrieval.

- In Machine Learning Approach the question was classified using Machine learning algorithms [24] like Support Vector Machine (SVM), Naïve Bayes and Random Forest. SVM is a more complex algorithm but can model nonlinear decision boundaries. Furthermore, it is also quite robust especially in high-dimensional space, like in the question and text classification problems both.
- In the hybrid approach the question was classified by the combination of NLP and machine learning approaches. The results validated that this combination enhanced the classification accuracy. Moreover, it helped the machine learning algorithms to better differentiate between different class types.

2) *Methods Used in Document/Passage Retrieval:* Document retrieval stage was the process of choosing the candidate document or passage that would be used as an input to the answer extraction stage. The existing search engine systems still face problems like retrieving many irrelevant documents and word mismatch, especially when the query given by the user was not specific enough [25 , 26]. To find the correct documents or passages, researchers used two types, document similarity score approach and machine learning approach.

- In document similarity score approach, the query was inputted into the search engine to get relevant documents based on similarity between concepts of the query and each title of the returned documents. Then the candidate passages were re-ranked and the N top-ranked ones were kept [27].
- In machine learning approach, an example was about Al-Baqarah Surah [23]. A Neural Network classifier labeled the Quran verses into two labels (Pilgrimage and Fasting). In the document retrieval stage, the purpose not to find actual verses to the question, but to extract relevant verses, before sending them to the answer extraction stage. The system then classified the question label (either pilgrimage or fasting) and chose the verses which had the same label. Since the research scope focuses on the users' questions that refer to the two pillars of Islam: Fasting and Pilgrimage. Therefore, Al-Baqarah Surah will be classified into two classes Fasting and Pilgrimage

3) *Methods Used in Answer Extraction:* This stage was typically applied to the top-ranked passages rather than to all available documents [7]. Researchers used a word matching score technique. The N-gram techniques; unigram and bigram were used to extract a list of words from the question. The candidate answers that contained similar words with the list of extracted words would be retrieved. The system performed word-matching scoring to calculate the number of similar words between the expanded questions and these candidate answers. Lastly, the system ranked all these candidate answers where the top ones were returned as the answer.

Socher et. al in [11] presented a CO-Search which consisted of a retriever-ranker semantic search engine that used search queries (including questions in natural language) and retrieved scientific articles over the coronavirus literature. CO-Search displayed content from over 128,000 coronavirus-related scientific papers made available through the COVID-19 Open Research Dataset Challenge (CORD-19).

CO-Search consists of a retriever and a ranker, with an offline pre-processing step to split and create a document index.

The index is created by embedding documents in three ways. Firstly a semantic method (BERT embeddings) was used to embed paragraphs and image captions, and two keyword methods (TFIDF, BM25) were used to embed entire documents. Documents were split into paragraphs, and then the titles of the citations of each paragraph were extracted. A bipartite graph from paragraphs and their cited articles was created to generate over 2.2 million (paragraph, title) tuples that used to train a Siamese-BERT (SBERT) model on the binary task of classifying if a citation was contained in a given paragraph. SBERT was used to embed queries and documents into the same latent space.

The retrieval step takes an input query, embeds it using SBERT. Then combines these embeddings with TFIDF and BM25. SBERT paragraph-level retrieval scores are combined linearly with TF-IDF document-level retrieval scores to generate a document list, and then reciprocal ranked fusion is used to combine this list with that obtained from BM25 retrieval. Ranking takes this set of documents and query, runs them through a question answering module (QA) and an abstractive summarizer, then ranks the documents by a weighted combination of their retrieval scores; the QA output, and the summarizer output. Question answering module takes the query, the retrieves documents, and uses a sequential paragraph selector model to filter for paragraphs that could answer the query. Once filtered, the paragraphs set were fed into an extractive reading comprehension model to extract answer candidates.

The summarizer takes the retrieved documents and generated a single abstractive summary.

This system was evaluated using the CORD-19 document dataset and relevance judgments provided by the TREC- COVID competition.

TABLE I
Comparative study of QA systems

Ref. Feature	Question answering paradigm	Language used to display Result	Open / Closed domain	Ontology support	Limitations
[4]	H (NLP, KB)	Indonesian language	Open	Yes	The system cannot answer questions containing more than two instances.
[5]	H (NLP, KB)	English language	Closed	Yes	The system uses only definition type of questions.
[6]	H (NLP, KB)	English language	Closed	Yes	In factoid type of questions, the answer is not readable enough. It also answers some questions correctly, but describe the facts very differently. In summary type of questions, the last sentence is repeated in some of questions answer.
[7]	H (NLP, KB)	Indonesian language	Closed	Yes	This framework works only for small training corpus.
[8]	H (NLP, KB)	English, German, French, Italian and Spanish language	Open	Yes	The system takes more time to answer the question because of using more than KB
[9]	H (IR, NLP, KB)	English language	Closed	Yes	The current form of the system was not able to provide answers to some questions which expect a number as answer instead of biomedical entities. If the type of a given question, the set of retrieved documents and passages are not identified correctly, processing steps to extract the answers will fail too.
[11]	IR	English language	Closed	No	In this time of crisis (COVID-19.), tens of thousands of documents are being published, only some of which are scientific, and rigorous. This may lead to the inclusion of misinformation and the potential rapid spread of scientifically disprovable or otherwise false research and data.
[12]	H (NLP, KB)	English language	Closed	Yes	The difficulties in Arabic language made the system unable to build Arabic question answering system and use Arabic Dbpedia ontology

ElKafrawy et. al in [12] presented an architecture of factoid question answering system using Dbpedia ontology. It consisted of three stages: Question Classification stage, Question Processing stage, Query Formulation and Execution stage. In Question Classification stage questions were categorized into one or more classes using SVM. It also determined the answer type which facilitated answering the question. A combination of syntax, syntactic and semantic features was chosen to be used to reach acceptable accuracy of SVM classifier, which made the classifier more accurate in determining the answer type. In Question Processing stage resources and keywords were extracted from the question. Dbpedia spotlight was utilized for extracting resources. To extract keywords from a question, these keywords should be nouns or verbs or non-stop words or complex nominals. After the extraction of keywords from the question, the synonyms of these keywords were extracted from a website (merriamwebster.com). Also, the keywords were enriched by extracting synonyms from WordNet After determining answer type, resources, keywords, and its synonyms. Query Formulation was done by determining ontology classes and properties of the given question. To make this, a Sparql query was built with the resource. The result of the query was an RDF file which contained all ontology properties and classes of that resource. Similarity between the keywords (and synonyms) in the question and ontology properties and classes was computed using Levenshtein distance algorithm and got the ontology class with the highest similarity to be utilized in the final Sparql query. The final Sparql query was built to answer the question from Dbpedia server. This query consisted of the resource and the ontology class determined in similarity stage. The dataset which was used in this work consisted of 5500 labeled questions which was utilized as a training set and 500 independent labeled questions which was utilized as a test set.

Table II includes a comparison between these systems.

1 CONCLUSION AND FUTURE WORK

This study collects the current state of different domains of QAS. The study showed that the Hybrid approach is the most flexible to independent language implementations which is noteworthy.
Even though QAS has been developed in various domains and techniques, the available QAS is mainly built for English. Therefore, it still needs to be evaluated in other languages. This study is done mainly to improve the system in [12].

REFERENCES

[1] Calijorne Soares, M.A., Parreiras, F.S., “A literature review on question answering techniques, paradigms and systems”. Journal of King Saud University – Computer and Information Sciences vol. 32, no. 6, pp. 635-646, 2020.

[2] Bouziane A, Bouchiha D, Doumi N, Malki M., “Question answering systems: survey and trends”. Procedia Computer Science, vol. 73, pp. 366-375, 2015.

[3] Allam A.M., Haggag M.H., “The question answering systems: A survey”. International Journal of Research and Reviews in Information Sciences (IJRRIS). vol. 2, no. 3, 2012.

[4] Yunmar R.A., Wisesa I.W., “Design of Ontology-based Question Answering System for Incompleted Sentence Problem”. In IOP Conference Series: Earth and Environmental Science vol. 258, no. 1, pp. 012-032. IOP Publishing 2019.

[5] Palaniappan S., Sridevi U.K., Subburaj J. “Ontology based Question Answering system using JSON-LD for Closed Domain”. International Journal of Pure and Applied Mathematics. vol.119, no. 12, pp.1969-1980, 2018.

[6] Kumar A.N., Kesavamoorthy H., Das M., Kalwad P., Chandu K, Mitamura T., Nyberg E., “Ontology-Based Retrieval & Neural Approaches for BioASQ Ideal Answer Generation”. In Proceedings of the 6th BioASQ Workshop A challenge on large-scale biomedical semantic indexing and question answering. pp. 79-89, 2018.

[7] Utomo F. S., Suryana N, Azmi M. S., “New instances classification framework on Quran ontology applied to question answering system”. Telkomnika, vol.17, no. 1, pp. 139-146, 2019.

[8] Diefenbach D., Both A., Singh K., Maret P. “Towards a question answering system over the semantic web”. Semantic Web, (Preprint), vol.11, pp. 421-439, 2020.

[9] Sarrouti M., El Alaoui S.O., “SemBioNLQA: a semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions. Artificial intelligence in medicine. 1;102:101767, https://doi.org/10.1016/j.artmed.2019.101767, 2020.

[10] Thalib I, Soesanti I. A Review on Question Analysis, Document Retrieval and Answer Extraction Method in Question Answering System. In 2020 International Conference on Smart Technology and Applications (ICoSTA) pp. 1-5, IEEE, 2020.

[11] Esteva A., Kale A., Paulus R., Hashimoto K., Yin W., Radev D., Socher R., “Co-search: Covid-19 information retrieval with semantic search, question answering, and abstractive summarization”. arXiv preprint arXiv:2006.09595, 2020.

[12] ElKafrawy P.M., Sauber A.M., Sabry N.A., “Semantic Question Answering System Using Dbpedia”. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems pp. 821-832. Springer, Cham, 2018.

[13] Zhu F., Lei W., Wang C., Zheng J., Poria S., Chua T.S., “Retrieving and Reading: A Comprehensive Survey on Open-domain Question Answering”. arXiv preprint arXiv:2101.00774, 2021.

[14] Salton G., The SMART retrieval system - experiments in automatic document processing. Prentice-Hall, 1971.

[15] Harman D. and Candela G., “Retrieving records from a gigabyte of text on a minicomputer using statistical ranking”. JASIS, vol.4, no. 8, pp.581-589, 1990.

[16] Strohman T., Metzler D., Turtle H., and Croft W.B. Indri: A language model-based search engine for complex queries. In Proceedings of the International Conference on Intelligent Analysis, vol. 2, no. 6, pp. 2-6, Citeseer, 2005.

[17] Ravichandran D., Ittycheriah A., and Roukos S., “Automatic derivation of surface text patterns for a maximum entropy-based question answering system”. In Proceedings of NAACL, short papers, pp. 85-87, 2003

[18] Wu M., Duan M., Shaikh S., Small S., and Strzalkowski T., “ILQUA- An IE-Driven Question Answering System”. In Ellen M. Voorhees and Lori P. Buckland, editors, TREC, volume Special Publication 500-266. National Institute of Standards and Technology (NIST), 2005.

[19] Cucerzan S. and Agichtein E., “Factoid question answering over unstructured and structured web content”. Proceedings of the Fourteenth Text REtrieval Conference, TREC 2005, Gaithersburg, Maryland, November 15-18, pp 72-90, 2005.

[20] Ferrucci D., Brown E., Chu-Carroll J., Fan J., Gondek D., Kalyanpur A.A., Lally A., Murdock J.W., Nyberg E., Prager J., et al., “Building Watson: An overview of the Deep QA project”. AI Magazine, vol.33, no. 3, pp.59-79, 2010

[21] Wasim M., Asim M.N., Ghani Khan M.U., and Mahmood W., “Multilabel biomedical question classification for lexical answer type prediction,” J. Biomed. Inform. , vol. 93, pp. 103-143, 2019.

[22] Li X and Roth D., “Learning question classifiers: The role of semantic information,” Nat. Lang. Eng., vol. 12, no. 3, pp. 229–249, 2006.

[23] Hamed S.K. and Aziz M.J.A., “A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification,” J. Comput. Sci., vol. 12, no. 3, pp. 169–177, 2016.

[24] Putter N., ZabidinHusin M., Tahir H.M., and Hussain A., “Building a Question Classification Model for a Malay Question Answering System”, Int. J. of Innovative Technology and Exploring Engineering, vol. 8, no. 5, pp 184-190, 2019.

[25] Imran H and Sharan A, “Thesaurus and query expansion,” J. Comput. Sci., vol. 1, no. 2, pp. 89–97, 2009.

[26] Ishkewy H and Harb H, “ISWSE: Islamic Semantic Web Search Engine,” Int. J. Comput. Appl., vol. 112, no. 5, pp. 975–8887, 2015.

[27] Sarrouti M and Ouatik El Alaoui S, “A passage retrieval method based on probabilistic information retrieval and UMLS concepts in biomedical question answering,” J. Biomed. Inform., vol. 68, pp. 96– 103, 2017.

BIOGRAPHY

Amr Mausad Sauber is an Associate Professor of Computer Science, Faculty of Science, Menofia University. He received his M.Sc. and Ph.D. Degree in Computer Science [Parallel Data Mining] from Faculty of Science, Menofia University. His research interests are Data related Database systems, Big Data and machine learning. He is a member of the Egyptian Society of Language Engineering. He is head of Central Automation Suite (Digital Transformation unite), Menofia University. He is a Treasurer of the Society of Basic and Applied Science at Menofia University.

Nada Ali, She obtained her bachelor's degree in 2012 from Faculty of Science, Menofia University. She is currently a student of master degree in computer science department, Faculty of Science, Menofia University since 2014. Her research interests are Natural Language Processing, semantic web and machine learning.

Passent elKafrawy, is a Full Professor since 2018, she got her PhD from the University of Connecticut, USA in 2006 in Computer Science and Engineering. Then she taught in Eastern State University of Connecticut for one year. In 2007, she worked as an Assistant Professor in Faculty of Science, Menoufia University, Mathematics and Computer Science department, in 2013 she was promoted to Associate Professor and in 2018 to Full professor. Since 2019 she joined Information Technology and Computer Science. School, Nile University. She has over 50 publications and a senior member of IEEE, ACEE, ESOLE, CSTA, IAEng and TIMA research organizations. One of the organizing members of the following conferences: SPIT and ESOLE’13-20. Supervising over 20 research studies between PhD and MSc. Member of the faculty projects for education development, appointed as the Quality Director for the ITCS QA Unit.

استطلاع أنظمة الإجابة على الأسئلة: دراسة مقارنة

عمرو مسعد صابر [1]*، ندى على[2]*، بسنت الكفراوى [3]*

*قسم الرياضيات البحتة وعلوم الحاسب، كلية العلوم، جامعة المنوفية، مصر

[1]amrmausad@computalityit.com|amr@science.menofia.edu.eg

[2]nadaali@science.menofia.edu.eg

[3]pelkafrawy@nu.edu.eg|basant.elkafrawi@science.menofia.edu.eg

ملخص

تعتبر أنظمة الإجابة على الأسئلة (QA) شكلاً متقدمًا من أشكال استرجاع المعلومات التي تتيح طرح أسئلة محددة من قبل البشر ويستنتج النظام الإجابات باستخدام استعلامات اللغة الطبيعية مع آليات فهم أعمق للمعرفة من كمية هائلة من البيانات بدلاً من جعل المستخدم يبحث في جميع المستندات بنفسه. نظرًا لأن ما يريده المستخدم حقًا هو غالبًا إجابة دقيقة على سؤال من خلال استشارة المستندات على الويب أو استخدام قاعدة معرفية خاصة.
على الرغم من تطوير العديد من الأنظمة على مدار السنوات الماضية، إلا أنه لا يزال يمثل تحديًا يتمثل في أن معظم الأنظمة تتطلب تحسينات لزيادة الدقة في التفسير الصحيح للسؤال وتقديم إجابة دقيقة جيدة لأسئلة المستخدم. تقدم هذه الورقة دراسة قصيرة لإطار ضمان الجودة العام وتهدف إلى مسح بعض أحدث أنظمة الإجابة على الأسئلة الحالية في مجالات مختلفة وإجراء مقارنة بين جميع هذه الأنظمة بناءً على بعض المعايير المحددة في الأدبيات.

الكلمات المفتاحية

أنظمة الإجابة على الأسئلة، معالجة اللغة الطبيعية، استرجاع المعلومات ، قاعدة المعرفة