Indian Journal of Science and Technology, *Vol 9(23), DOI: 10.17485/ijst/2016/v9i23/91711, June 2016*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Framework for Improved Question Answering System

Muthukrishnan Ramprasath[1]* and Shanmugasundaram Hariharan[2]

[1]Computer Science and Engineering, Sathyabama University, Chennai - 600119, Tamil Nadu, India; mramprasath@gmail.com

[2]Department of Computer Science and Engineering, TRP Engineering College, Trichy - 621105, Tamil Nadu, India; mailtos.hariharan@gmail.com

Abstract

Background/Objectives: The proposed frame work uses statistical based similarity measure for QAS, which gives short and exact answer to user query. The entire frame work is focusing similarity computation between question and answer pair. **Methods/Statistical Analysis:** The involvement of similarity computation between question and answer pair should be investigated in this work QAS. Statistical based similarity measure is an important issue as compared to similarity computation in many fields such as, Natural Language Processing, Information Retrieval, Ontology Mapping, Knowledge Acquisition and Question Answering System. Semantic web has been used to present the structure knowledge representation between questions and answer keywords. Ontology can be used as important proposal to enable semantic similarity between keywords present in the question and answer pairs. **Findings:** The proposed frame work investigate, how the statistical based similarity measure used in these system helps to improve the performance of the QA system for all Wh-questions. We find the keywords having most similar meaning will return the answer as the final answer. The best performance is achieved by extracting the relevant snippets information form Google search engine. The performance of the system will change as the retrieval of the document is increases beyond 15; this could indicate that restriction in the retrieval of document is helps to optimize the performance of the system. **Application/Improvements:** This frame work is applicable to all type of search engines which helps to finds most relevant answer to user questions. In future, we implement our frame work for questions having more than five keywords.

Keywords: Information Retrieval, Ontology, Question Answering System, Semantic Similarity.

1. Introduction

The goal of QA system is to directly communicate with the user through Natural Language, which brings more simplicity to user who submits their information need. Semantic based question answering system play vital role to provide accurate answer to user questions. Identifying the semantic meaning of the keyword present in the user question and estimating the statistical based similarity between the question and answer term adds better understanding of the textual resources. Semantic similarity computation is an significant issue which has many direct applications, such as, Natural Language Processing, Information Retrieval based on conceptual distance[1], Sentence retrieval based on machine translating system, word-sense disambiguation[2], word spelling correction[3], ontology learning[4] and document clustering[5].

To establish agreement between different knowledge domains, ontology matching and alignment can be used. In language technology, several efforts have been made in the recent past, to define text and word similarity distance based on words overlap for various applications.

**Author for correspondence*

Knowledge interoperability in the semantic web in which ontology specifies the semantic meaning of the keywords explicitly and logically[6]. However, previous ontology-based approaches having limitations such as, it is not easy to find what the users exactly look for; it will take much time to the user to retrieve the answer in precise manner.

WordNet database that contain large number of lexical and semantic hierarchical relation and classes[7]. These knowledge bases are used to construct taxonomy of word senses or concepts, with various types of relations and to map the text fragments in the question to be compared with the taxonomy. Unsupervised method used to construct the semantic representation of documents and of words, by analyzing co-occurrences relationship in the corpus. Explained one common measure of text semantic similarity, by defining the word similarity and word specificity by compared with several knowledge based and corpus based methods[8].

The knowledge sources covers only fraction of vocabularies, technical terms and few proper names. Numerous methods can be used in the path of the semantic network or WordNet to compute the lexical semantic relatedness. Relation between the nodes is one of the visible criteria to measure the semantic relatedness between the words. Information-theoretic definition for similarity that applicable to any type of entities[9]. Random graphs walk methods uses a graph to compute word relatedness using Personalized Page Rank (PPR)[10]. It uses a graph built from WordNet having 400,000 nodes and 5 million links. It is used to find the shortest path and exploit all possible links between the words in the graph. In Wikipedia, each word is mapped to the corresponding articles by using the titles.

Semantic role based QA system can be used in the answer extraction module but their system focus NE-based question during evaluation[11]. Probabilistic model for information theoretic definition of similarity which expresses how the definition can be used to measure the similarity in different domain. Roget's thesaurus used as a knowledge base for calculating semantic distance between two words and also allows an easy implementation of edge counting approach[12].

Lexical semantic relatedness based on random graph walk principal that incorporates information form explicit or implicit path connecting the two words in the entire graph. The graph can be treated as Markov chain and compute word-specific stationary via Page Rank algorithm also discuss the application of Markov chain theory to measuring lexical semantic relatedness. Numerous existing algorithms calculate the relatedness only by traversing the hypernymy taxonomy. Conversely WordNet provides other type of relationship such as meronymy, antonymy and verb entitlement.

The formal representation of concepts and their relationship is very much helpful for QA systems to extract the relevant answer to questions. In closed domain QA system, where they use a domain ontology which formally represents concepts and their relationship to provide all the information required by user. Conversely, in open domain QA systems, where we use very extended semantic network called WordNet, which contain large number of lexical and semantic hierarchical relations and classes[13]. Recently WordNet has been extended to other language and so has advantage to provide multilingual information.

2. Relationship between Words, Texts and Sentences in Nature

In linguistic, semantic relations between the words and texts have been well studied and categorized. It includes classical relation such as synonymy (identity of senses, 'fortunate' and 'auspicious') are the terms whose meaning are regard as same in a wide range of context. Antonyms (opposition of senses, 'happy' and 'unhappy'), hyponymy ('education' and 'school') and holonymy (part-whole relation such as 'RAM' and 'CPU'), homonym is a kind of semantic relation, where the two word spelled same way but have different meaning (letter–correspondence, letter-alphabet), polysemy words that has several sub senses which are related to one another.

These classical relations are showed in lexical dictionary and ontological resources, such as WordNet, Encyclopaedia and they can be inferred from distributional data. It also helps to quantify the semantic relation between the words. In addition to these classical relations, non-classical relations are described in lexical types, for example based on relationship in similar classes or on association by location, but these non-classical relations does not help to quantify similarity between keywords.

The comparison of the similarity measure work in terms of algorithm task and different database used is shown in Table 1. Paraphrase or logical relations can give some notions regarding this semantic relatedness in the sentence part. Commonly, two sentences can be related by a similar topic. Although, the idea of the topic is difficult to define, also an idea that applies to multi-sentences as well. It is frequently expressed in conditions of the Continuity of Themes. In cohesive devices linguistic have analyzed these types of cohesiveness, which consist of lexical cohesion, similarity chain based on classical lexical relation and identity of reference[14]. From various distributional approaches, the semantic relatedness of words, texts and sentences are noticed by researchers in Natural Language Processing (NLP).

In this approaches some limitations are identified by[15]. Using Latent Semantic Analysis (LSA)[16] the sentences and texts from coherent units which is used to infer the lexical relation and word similarity by the hidden topical parameters. Corpus and knowledge measure for calculating semantic similarity of short text segments. In corpus based measure uses information entirely derived from large corpora to identify the degree of similarity between words. Its uses two metrics namely: 1) PMI (Point wise Mutual Information, 2) Latent Semantic Analysis. PMI based on the word co-occurrence method uses counts collected from large corpora. Term co-occurrence is captured in LSA by means of dimensionality reduction operated by Singular Value Decomposition (SVD) on TD matrix representing the corpus. Knowledge based measure uses WordNet hierarchy to find the similarity between the concepts. He presented six measure based on the experimental performance in other language processing application and for their reasonably high computational efficiency.

2.1 Semantic Web Knowledge for Semantic Relatedness

In linguistics, the process of relating syntactic structure, from the point of phrase, clause, sentence and paragraph to the level of writing as a whole is called semantic analysis. In our approach, we represent the text fragment present in the user question as a weighted combination of a prearranged set of natural concepts or keywords. To represent the text in question as a set of natural concepts, we use the concepts related to the keywords defined by semantic web articles. An important advantage of our semantic approach is thus the use of huge amount of highly prearranged human knowledge encoded in semantic web.

Machine learning methods helps to construct the semantic interpreter that maps text fragment in the user question into a weighted sequence of concepts ordered by their relevance to their text fragment in user question. For instance, find the meaning of the word "cat"? The explicit interpretation of the word: A cat is a mammal with four legs, which belongs to the feline species, one more way to interpret the meaning of "cat" is by the strength of the association of the concepts relate strongly to "pet" and "feline", a bit less strongly to the concepts "Tom and Jerry" and mouse etc. Semantic roles and ontology's have been used as semantic information to improve the performance of the QA system or even deal with particular type of questions.

Table 1. Comparison of similarity measure work in terms of algorithm, task and different datasets

Author	Knowledge source	Algorithm Proposed	Work	Data set	Factor used
Mihalcea R, Corley C8	Internet	LSA	Paraphrase	Microsoft	Short Text similarity
Hughes and Ramag10	WordNet	PPR	Similarity between words	M&C, R&G, WS-353	Random walk in WordNet
Jarmasz M, Szpakowicz12	Roget thesaurus	Shortest path	Similarity between words.	M&C, R&G, Synonyms	Semantic distance
Miller A, Walter Charles G18	Word net	Context based measure	Degree of synonyms between words.	M&C 30-Noun pairs	Semantic similarity
Rubenstein and John 24	Internet	Context based measure	Synonyms judgements	65- Noun pairs	Semantic similarity
Gabrilovich and Markovitch25	Wikipedia	ESA: TF-IDF + Cosine sim.	Word and doc similarity	WS-353, Lee	Degree of similarity between text

Semantic role contribution has received more attention in QA system[17]. Each predicate in question sentence will helps to identify all the constituents, determining the roles. In this manner, semantic role represent "WHO did WHAT to WHOM, WHEN, WHERE, HOW and WHY?" in a sentence, which specifies that they may be extremely use full in information extraction. Similarity measure that focused the degree of synonyms between the words in question and snippet information[18]. This information does not help for the user to identify the correct answer. So we proposed statistical based similarity measure between question and answer pairs. For instance, the following example shows that, semantic relationship between the keywords in user question and does not provide sufficient information for identifying exact answer. For instance, consider the following question.

2.1.1 User Question: Which is the longest river in the world? Question words = {longest, river, world}

From the above observation we conclude that similarity between keywords based on the semantic relations does not provide the exact answer to user question. To overcome this problem, in this paper we propose a statistical based similarity measure between questions and answer pairs and it provides best result to user question. In Section 4 we discussed statistical based similarity measure.

3. Statistical based QA Architecture

Statistical based similarity measure play significant role in many of the research fields, such as short text similarity, corpus and knowledge based similarity, sentence similarity measure, Information Retrieval and question answering and more.

In QA system finding the semantic relationship between the question keyword and snippets information keywords is crucial task. Statistical based similarity measure play important role for retrieving relevant information to user questions. In this paper we investigated the use of the similarity measure as discussed by[19] in Section 2 for our proposed system. The following figure shows the statistical based QA architecture and its components of proposed work.

It has been implemented as part of statistical based Information Retrieval module of the general QA system. It follows the steps used in[20]. In contrast to that, our system

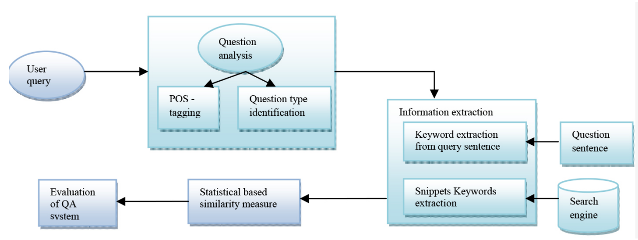

Figure 1. Statistical based QA system architecture.

is structured into four functional components. Question analysis, Information extraction, statistical based similarity measure and evaluation of system. The user can post the query to search engine, its checks the type of the question. After identifying question type the keywords in the user question can be extracted. Ontology based question answering system[21] address the user queries collections and its types. [22]Ontology based knowledge access used in health care system for retrieving precise result to user question.

3.1 Question Analysis and Snippet Keyword Extraction

The requirement of question analysis component in our QA system is helps to identify the type of question post by the user. Identification of the question types helps the user to retrieve the relevant information to user question. Question analysis component involves question type identification and POS-tagging. POS-tagging provides necessary information for extracting semantic meaning from the word net dictionary.

Semantic meaning for the Keyword present in the user question aid to express the kind of information requested. The POS method helps to provide necessary information while computing the semantic similarity between the keyword. The Wh-term in user questions, maps into the one or more several categories such as person, location, reason, time etc. All these categories are interconnected with WordNet top concepts. Table 2 shows the Wh-terms and corresponding semantic relationship.

Table 2. Question type and semantic role

WH –term	Semantic relationship
Where	Location
How	Quantity
Why	Reason
When	Time

For better understanding of IR module based on the semantic relationship between keywords is now briefly described. The goal of this module is to extract the semantic meaning for the keywords in the user questions and snippet information. Depending on the Wh-term and keywords in user question, different semantic relationship may be considered as answer. For example, “HOW”, in “HOW + quantity expression” must give quantity as an answer to user question. The following table shows the summary of semantic relationship between the keywords.

3.1.1 Similarity Measure based on Statistical Analysis

Step 1: Post the user question to search engine.
Step 2: Perform the Syntactic analysis on user question sentence.
Step 3: Pulling out of keywords form question sentence.
Step 4: Retrieving snippets information appropriate to question keywords.
Step 5: Extraction of keywords for snippets information.
Step 6: Statistical based similarity measure between keywords in question sentence and snippets keywords.
Step 7: Evaluation of proposed system using Precision, recall and F-Measure metrics.

Because of having extra information which is not related to our user questions. It makes the user, spending much time to get the answer. But in our proposed work, based on the semantic relationship between the keywords, the retrieved snippets information is validated. Using semantic similarity between keywords aids the user to get relevant answer and also in short time. We use statistical based similarity measure which only considers the semantic meaning between the set of keywords. After finding the semantic meaning of keywords in user sentence and snippets information our system uses all type questions for evaluation.

4. Statistical based Similarity Measure

The methodology that has been adapted to develop a QA system based on statistical based similarity measure for providing relevant answer to user question. Semantic based approach[23] was used in health seekers and health care knowledge portal for accessing correct information from database. The principle of the statistical based similarity measure is a linear combination of semantic similarity and statistical similarity. We consider word co occurrence between question and answer pair for statistic similarity and utilize the word for semantic similarity. Initially we calculate two parts separately and then do the linear combination. Distinct to other methods, statistical similarity calculation is based on dynamically formed vector space which avoid sparse vector space problem. Bipartite mapping is used for semantic similarity calculation between question and answer pairs.

Traditional Information Retrieval system uses high dimensional space to represent the documents and each word in the document collection is treated as one dimension. This retrieval system work well in large document retrieval application because document share lot of words. Since there are only a small amount of words in question and answer pair’s sentence, because of that the high dimensional representation method will lead to a poor performance of the system. So we use low dimensional vector space to represent question and answer pairs. The low dimensional vector space is formed based on the word set of question and answers. This formation helps to avoid sparse vector problem.

Given Q_s and A_s we define a word set as follows: QAS $= Q_s \cup A_s$.The question answering set QAS contain all distinct words in Q_s and A_s. Each question and answer can be represented by vector *v*. The dimensionality is equal to the word number of QAS. Each component in the vector represents the corresponding word in QAS. The value of each word set is determined as follows:

- If the keyword present in the question is does not match with snippets keywords, the value of the component is said to be zero;
- If the keyword present in the question is match with snippets, the value of the component is said to be one;

Given two vectors, we calculate the statistical similarity of Q_s and as by calculating their cosine product:

$$Sim_{statistic} = \frac{\mathrm{v}_1 \ v_2}{|v_{1|.|v_{2|}}} \tag{1}$$

Where $Sim_{statistic}$ is the statistic similarity of Qs and A_s.
For instance consider the following word set:

Q: What is the colour of rose?
Question key words: {what, colour, rose}

Snippet Keywords: {Colour, rose, flower, colour, plant, red, roses}

Word sets: {what, colour, rose, colour, plant, red, roses}

Vector dimension: 7 X_1={1,1,1,0,0,0,0} X_2={1,1,1,1,1,1,1}

$$sim_{statistic}\frac{x1.x2}{|x1||x2|} = \mathbf{0.6547}.$$

4.1 Precision, Recall and F-Measure

Several metrics have been used for evaluating the result of the Question Answering System. Precision and recall metrics are used to measure the performance of the system. In the field of QAS precision is the fraction of retrieved documents that are relevant to the user question. Recall is the probability that a relevant document is retrieved in the search. Accuracy is used as a one of the major evaluation metrics, for which the answers are judged to be a globally correct.

$$Preciion = \frac{|R_d| \cap |R_r|}{|R_d|} \tag{2}$$

$$Recall = \frac{|R_d| \cap R_r}{|R_d|} \tag{3}$$

Where, denotes relevant document and retrieved documents related to user queries. In statistical based measure F-Measure can be used to test the accuracy of the system. It considers both precision and recall of the test to compute the score.

$$F - measure = 2.\frac{precision.recall}{precision + recall} \tag{4}$$

While evaluating the QAS recall, precision and F-measure metrics are used to evaluate the performance of the system.

5. Experimental Setup

We presented our proposed system based on the approach discussed in the previous sections 3, 4, 5. We dealt with what, where, who, when, how, why, type of questions. Here we collected 25 questions from each type. So we evaluated 150 questions totally in this part. By using search engines we retrieved answer which is related to the keyword in user question. Google and Yahoo search engine used as a knowledge base for answer all type of questions posted by user. The answer retrieved from these search engines are unique, trustable and bounded.

Now, the selected questions are posted to search engine. The results given by the search engine, which is short text segment called snippets. Question keywords are extracted from user question and answer keywords are extracted from snippets information retrieved from search engine are inputted to the system. Text processing of question and answer helps for removing unnecessary stop-words, punctuations and mark-up. We consider 58 frequently occurring stop-words for evaluation purpose.

The collected sample questions are used to construct the probability model for each type of question separately. Initially the system checks the question types by using the keyword with the class information in the WordNet. After identifying the type of questions, the statistical based Information Retrieval component take question keywords as input to the system and gives more relevant text portion as a result of user question. This extraction of the answer is done by matching the keywords between question and answer sentences.

5.1 Results

In result section, the proposed system was evaluated with different WH-type questions. The lengths of these questions are restricted from 3 words to 10 words. Based on the keywords in the user question we manually construct 25 questions for each type. For each question type the top N text units are retrieved and passed to snippet keyword extraction module, which extract keywords in the web snippet information. Let N be 5, 10, 15, 20 and 25. The accuracy of proposed QA system is calculated and presented in the following tables from Table 3 to Table 14 based on the number of keywords present in the user question and snippets information given by search engine.

5.2 Study 1

Here we discusses about the experiments result carried out in Section 6. (1) How the statistical based similarity measure used in these system helps to improve the performance of the QA system for all Wh-questions. (2) We used different evaluation metrics (Precision, Recall, F-measure) for compare the performance system. We find the keywords having most similar meaning will return the answer as the final answer. The best performance is achieved by extracting the relevant snippets information form Google search engine, when retrieving top 15 documents for all "Wh-type" questions gives with a precision in the range from 0.63 to 0.75 and recall in the range from

Table 3. Illustrate precision recall comparison for top 25 snippets documents (What type questions)

	Google							
What type question	What is family?		What is the colour of horse?		What are the standards needs for hospital?		What do student like in college life?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	4/25=0.16
Top 10	8/10=0.8	8/25=0.32	7/10=0.7	7/25=0.28	8/10=0.8	8/25=0.32	7/10=0.7	6/25=0.24
Top 15	12/15=0.8	12/25=0.48	13/15=0.86	13/25=0.52	12/15=0.8	12/25=0.48	13/15=0.86	8/25=0.32
Top 20	13/20=0.65	13/25=0.52	14/20=0.7	14/25=0.56	13/20=0.65	13/25=0.52	14/20=0.7	10/25=0.4
Top 25	13/25=0.52	13/25=0.52	15/20=0.75	15/25=0.57	13/25=0.52	13/25=0.52	15/20=0.75	11/25=0.44

Table 4. Illustrate precision recall comparison for top 25 snippets documents (Where type questions)

	Google							
Where type question	Where is the earth?		Where is the Tajmahal?		Where do French people live?		Where Microsoft corporate headquarters is located?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2
Top 10	9/10=0.9	9/25=0.36	10/10=1	10/25=0.4	9/10=0.9	9/25=0.36	10/10=1	10/25=0.4
Top 15	13/15=0.86	13/25=0.52	15/15=1	15/25=0.6	13/15=0.86	13/25=0.52	15/15=1	15/25=0.6
Top 20	18/20=0.9	18/25=0.72	20/20=1	20/25=0.8	18/20=0.9	18/25=0.72	20/20=1	20/25=0.8
Top 25	21/25=0.84	21/25=0.84	23/25=0.92	23/25=0.92	21/25=0.84	21/25=0.84	23/25=0.92	23/25=0.92

Table 5. Illustrate precision recall comparison for top 25 snippets documents (Who type questions)

	Google							
Who type questions	Who are aliens?		Who is the president of America?		Who first circumnavigated the globe?		Who invented the road traffic cone?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2
Top 10	9/10=0.9	9/25=0.36	6/10=0.6	6/25=0.24	10/10=1	10/25=0.4	8/10=0.8	8/25=0.32
Top 15	11/15=0.73	11/25=0.44	11/15=0.73	11/25=0.44	14/15=0.93	14/25=0.56	9/15=0.6	9/25=0.36
Top 20	12/20=0.6	12/25=0.48	14/20=0.7	14/25=0.56	14/20=0.7	14/25=0.56	11/20=0.55	11/25=0.44
Top 25	13/25=0.86	13/25=0.86	16/25=0.64	16/25=0.64	14/25=0.56	14/25=0.56	12/25=0.48	12/25=0.48

Table 6. . Illustrate precision recall comparison for top 25 snippets documents (When type questions)

Search engine	Google							
When type questions	When will the world end?		When type questions		When will the world end?		When type questions	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	2/5=0.4	2/25=0.08	5/5=1	5/25=0.2	3/5=0.6	3/25=0.12	5/5=1	5/25=0.2
Top 10	4/10=0.4	4/25=0.16	9/10=0.9	9/25=0.36	6/10=0.6	6/25=0.24	10/10=1	10/25=0.4
Top 15	9/15=0.6	9/25=0.36	13/15=0.86	13/25=0.52	8/15=0.53	8/25=0.32	12/15=0.8	12/25=0.48
Top 20	11/20=0.55	11/25=0.44	18/20=0.9	18/25=0.72	10/20=0.5	10/25=0.4	14/20=0.7	14/25=0.56
Top 25	12/25=0.48	12/25=0.48	21/25=0.89	21/25=0.89	11/25=0.44	11/25=0.44	15/25=0.6	15/25=0.6

Table 7. Illustrate precision recall comparison for top 25 snippets documents (how type questions)

Search engine	Google							
How type question	How is the army		How type question		How is the army		How type question	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	2/5=0.4	2/25=0.08	5/5=1	5/25=0.2	3/5=0.6	3/25=0.12	5/5=1	5/25=0.2
Top 10	4/10=0.4	4/25=0.16	9/10=0.9	9/25=0.36	6/10=0.6	6/25=0.24	10/10=1	10/25=0.4
Top 15	9/15=0.6	9/25=0.36	13/15=0.86	13/25=0.52	8/15=0.53	8/25=0.32	12/15=0.8	12/25=0.48
Top 20	11/20=0.55	11/25=0.44	18/20=0.9	18/25=0.72	10/20=0.5	10/25=0.4	14/20=0.7	14/25=0.56
Top 25	12/25=0.48	12/25=0.48	21/25=0.89	21/25=0.89	11/25=0.44	11/25=0.44	15/25=0.6	15/25=0.6

Table 8. Illustrate precision recall comparison for top 25 snippets documents (why type questions)

Search engine	Google							
Why type question	Why MBA?		Why is English required?		Why India is called democratic country?		Why did David ask FBI for a word processor?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	3/5=0.6	3/25=0.12
Top 10	10/10=1	10/25=0.4	9/10=0.9	9/25=0.36	8/10=0.8	8/25=0.32	6/10=0.6	6/25=0.24
Top 15	13/15=0.86	13/25=0.52	11/15=0.73	11/25=0.44	10/15=0.6	10/25=0.4	8/15=0.53	8/25=0.32
Top 20	18/20=0.9	18/25=0.72	12/20=0.6	12/25=0.48	12/20=0.6	12/25=0.48	9/20=0.45	9/25=0.36
Top 25	18/25=0.72	18/25=0.72	12/25=0.48	12/25=0.48	13/25=0.86	13/25=0.86	11/25=0.44	11/25=0.44

Table 9. Illustrate precision recall comparison for top 25 snippets documents (what type questions)

Search engine	Yahoo							
What type question	What is family?		What is the colour of horse?		What are the standards needs for hospital?		What do student like in college life?	
	5/5=1	5/25=0.2	5/5=1	5/25=0.2	4/5=0.8	4/25=0.16	4/5=0.8	4/25=0.16
Top 5	8/10=0.8	8/25=0.32	8/10=0.8	8/25=0.32	5/10=0.5	5/25=0.2	6/10=0.6	6/25=0.24
Top 10	10/15=0.6	10/25=0.4	14/15=0.93	14/25=0.56	8/15=0.53	8/25=0.32	8/15=0.53	8/25=0.32
Top 15	12/20=0.6	12/25=0.48	14/20=0.7	14/25=0.56	10/20=0.5	10/25=0.4	10/20=0.5	10/25=0.4
Top 20	12/25=0.48	12/25=0.48	15/25=0.6	15/25=0.6	11/25=0.44	11/25=0.44	11/25=0.44	11/25=0.44
Top 25	5/5=1	5/25=0.2	5/5=1	5/25=0.2	4/5=0.8	4/25=0.16	4/5=0.8	4/25=0.16

Table 10. . . Illustrate precision recall comparison for top 25 snippets documents (where type questions)

	Yahoo							
Where type question	Where is the earth?		Where is the Tajmahal?		Where do French people live?		Where Microsoft corporate headquarters is located?	
	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	4/5=0.8	4/25=0.16
Top 5	9/10=0.9	9/25=0.36	10/10=1	10/25=0.4	8/10=0.8	8/25=0.32	8/10=0.8	8/25=0.32
Top 10	14/15=0.93	14/25=0.56	14/15=0.93	14/25=0.56	9/15=0.6	9/25=0.36	9/15=0.6	9/25=0.36
Top 15	18/20=0.9	18/25=0.72	18/20=0.9	18/25=0.72	11/20=0.55	11/25=0.44	9/20=0.45	9/25=0.36
Top 20	21/25=0.84	21/25=0.84	21/25=0.84	21/25=0.84	14/25=0.56	14/25=0.56	9/25=0.36	9/25=0.36
Top 25	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	4/5=0.8	4/25=0.16

Table 11. . Illustrate precision recall comparison for top 25 snippets documents (who type questions)

Search engine	Yahoo							
Who type question	Who are aliens?		Who is the president of America?		Who first circumnavigated the globe?		Who invented the road traffic cone?	
	4/5=0.8	**4/25=0.16**	**5/5=1**	**5/25=0.2**	**5/5=1**	**5/25=0.2**	**5/5=1**	**5/25=0.2**
Top 5	8/10=0.8	8/25=0.32	6/10=0.6	6/25=0.24	8/10=0.8	8/25=0.32	7/10=0.7	7/25=0.28
Top 10	9/15=0.6	9/25=0.36	10/15=0.6	10/25=0.4	9/15=0.6	9/25=0.36	9/15=0.6	9/25=0.36
Top 15	10/20=0.5	10/25=0.4	11/20=0.55	11/25=0.44	11/20=0.55	11/25=0.44	10/20=0.5	10/25=0.4
Top 20	11/25=0.44	11/25=0.44	12/25=0.48	12/25=0.48	12/25=0.48	12/25=0.48	10/25=0.4	10/25=0.4
Top 25	4/5=0.8	4/25=0.16	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2

Table 12. Illustrate precision recall comparison for top 25 snippets documents (when type questions)

Search engine	Yahoo							
When	When will the world end?		When is summer?		When did life of the earth begin?		When was London docklans railway constructed?	
	2/5=0.4	**2/25=0.08**	**4/5=0.8**	**4/25=0.16**	**3/5=0.6**	**3/25=0.12**	**5/5=1**	**5/25=0.2**
Top 5	3/10=0.3	3/25=0.12	9/10=0.9	9/25=0.36	4/10=0.4	4/25=0.16	9/10=0.9	9/25=0.36
Top 10	9/15=0.6	9/25=0.36	11/15=0.73	11/25=0.44	7/15=0.46	7/25=0.28	11/15=0.73	11/25=0.44
Top 15	10/20=0.5	10/25=0.4	13/20=0.65	13/25=0.52	8/20=0.4	8/25=0.32	13/20=0.65	13/25=0.52
Top 20	12/25=0.48	12/25=0.48	15/25=0.6	15/25=0.6	10/25=0.4	10/25=0.4	14/25=0.56	14/25=0.56
Top 25	2/5=0.4	2/25=0.08	4/5=0.8	4/25=0.16	3/5=0.6	3/25=0.12	5/5=1	5/25=0.2

Table 13. Illustrate precision recall comparison for top 25 snippets documents (how type questions)

Search engine	Yahoo							
How type	How is the army		How important is the freedom?		How the stock market works?		How many moons does Jupiter have?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2
Top 10	9/10=0.9	9/25=0.36	9/10=0.9	9/25=0.36	8/10=0.8	8/25=0.32	10/10=1	10/25=0.4
Top 15	12/15=0.8	12/25=0.48	10/15=0.6	10/25=0.4	10/15=0.6	10/25=0.4	12/15=0.8	12/25=0.48
Top 20	13/20=0.65	13/25=0.52	11/20=0.55	11/25=0.44	12/20=0.6	12/25=0.48	13/20=0.65	13/25=0.52
Top 25	14/25=0.56	14/25=0.56	13/25=0.86	13/25=0.86	13/25=0.86	13/25=0.86	14/25=0.56	14/25=0.56

Table 14. Illustrate precision recall comparison for top 25 snippets documents (why type questions)

Search engine	Yahoo							
Why type question	Why MBA?		Why is English required?		Why India is called democratic country		Why did David ask FBI for a word processor?	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Top 5	5/5=1	5/25=0.2	5/5=1	5/25=0.2	5/5=1	5/25=0.2	3/5=0.6	3/25=0.12
Top 10	8/10=0.8	8/25=0.32	10/10=1	10/25=0.4	8/10=0.8	8/25=0.32	4/10=0.4	4/25=0.16
Top 15	10/15=0.6	10/25=0.4	10/15=0.6	10/25=0.4	9/15=0.6	9/25=0.36	7/15=0.46	7/25=0.28
Top 20	12/20=0.6	12/25=0.48	11/20=0.55	11/25=0.44	11/20=0.55	11/25=0.44	8/20=0.4	8/25=0.32
Top 25	13/25=0.86	13/25=0.86	11/25=0.44	11/25=0.44	12/25=0.48	12/25=0.48	10/25=0.4	10/25=0.4

0.42 to 0.56. The performance of the system will change as the retrieval of the document is increases beyond 15. This could indicate that restriction in the retrieval of document is helps to optimize the performance of the system.

The same experiment has been conducted in Yahoo search engine for all "Wh-type" questions. The precision value lies in the range from 0.58 to 0.85 and recall in the range from 0.28 to 0.45. In Yahoo search engine, the retrieval of the document increases beyond 15 will affect the performance of the system. The experiment results shown in the Section 6.1. By observing the above result tables we conclude that Google search engine provides relevant answer for how type question when compare to Yahoo search engine.

5.3 Discussion

Our frame work also calculates the average Precision, Recall and also F-Measure to test the accuracy of our proposed approach using different search engines. Google search engine provide highest average precision @ 0.72, recall @ 0.41 and F-measure @ 0.51.for all "Wh- type" questions as compare to Yahoo search engine. From the experimental section, we conclude that, if number of keywords increases in user question, it reduces the performance of the system. It shows that Google provide better results as compared to Yahoo search engine. Tables 15 and 16 shows average precision recall and F-score for "Wh-type" type question based on keywords used in Google and Yahoo search engines.

6. Conclusion

In this paper, an effective measure to find the similarity between keywords in the question sentence and Snippets information has been proposed. In beginning we took different Wh-type questions having length as two, three, four and five keywords. After that, the keyword present in the question sentence and keywords in the snippets

Table 15. Illustrate average precision recall and F-score for "Wh-type" type question based on keywords (Google)

Question types	**Google search engine**					
	Number of keywords					
	No of keywords	Two	Three	Four	Five	Average
What	Precision	0.84	0.68	0.67	0.54	0.68
	Recall	0.44	0.36	0.42	0.34	0.39
	F-Measure	0.57	0.47	0.51	0.41	0.49
Why	Precision	0.89	0.72	0.77	0.52	0.72
	Recall	0.51	0.39	0.45	0.29	0.41
	F-Measure	0.64	0.50	0.56	0.37	0.51
How	Precision	0.82	0.75	0.77	0.84	0.79
	Recall	0.45	0.40	0.43	0.46	0.43
	F-Measure	0.58	0.52	0.55	0.59	0.56
When	Precision	0.91	0.48	0.53	0.82	0.68
	Recall	0.53	0.30	0.30	0.44	0.39
	F-Measure	0.66	0.32	0.38	0.57	0.48
Who	Precision	0.81	0.73	0.83	0.68	0.76
	Recall	0.46	0.41	0.45	0.36	0.42
	F-Measure	0.58	0.52	0.58	0.47	0.48
where	Precision	0.91	0.93	0.70	0.60	0.78
	Recall	0.53	0.54	0.37	0.31	0.43
	F-Measure	0.66	0.68	0.48	0.40	0.55

Table 16. Illustrate average precision recall and F-score for "Wh-type" type question based on keywords (Yahoo)

Question types	**Yahoo search engine**					
	Number of keywords					
	No of keywords	Two	Three	Four	Five	Average
What	Precision	0.69	0.81	0.55	0.57	0.65
	Recall	0.37	0.34	0.76	0.31	0.44
	F-Measure	0.48	0.47	0.31	0.17	0.35
Why	Precision	0.77	0.71	0.68	0.45	0.65
	Recall	0.45	0.37	0.36	0.25	0.35
	F-Measure	0.56	0.48	0.47	0.32	0.45
How	Precision	0.78	0.78	0.77	0.80	0.78
	Recall	0.42	0.45	0.45	0.43	0.43
	F-Measure	0.54	0.57	0.56	0.55	0.55
When	Precision	0.45	0.73	0.45	0.76	0.59
	Recall	0.28	0.41	0.25	0.41	0.33
	F-Measure	0.34	0.52	0.32	0.53	0.42
Who	Precision	0.62	0.64	0.68	0.64	0.64
	Recall	0.33	0.35	0.36	0.32	0.34
	F-Measure	0.43	0.45	0.48	0.42	0.44
where	Precision	0.91	0.93	0.70	0.60	0.78
	Recall	0.53	0.54	0.37	0.31	0.43
	F-Measure	0.66	0.68	0.48	0.40	0.55

information has been extracted based on the statistical based similarity measure. We contact the experiment for all Wh-type questions with two different search engines by means of Precision, Recall and F-measures. Then by analyzing these result, Google search engine gives semantically relevant answer compare to Yahoo search engine.

7. References

1. Lee JH, Kim MH. Information Retrieval based on conceptual distance in is - A hierarchy. Journal of Documentation.1993; 49(2):188–207.
2. Resnik P. Semantic similarity in taxonomy. An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research. 1999; 11:95–130.
3. Budanitsky A, Hirst G. Evaluating WordNet-based measures of semantic distance. Computational Linguistics. 2006; 32(1):13–47.
4. Sanchez D, Batet D. Domain ontology learning from the web: An unsupervised, automatic and domain independent approach. 2009.
5. Cilibrasi RL, Vitanyi PMB. The Google similarity distance. IEEE Transactions on Knowledge and Data Engineering. 2006; 19(3):370–83.
6. Lanzenberger M, Sampson J, Kargl H, Wimmer M, Conroy C. Making ontology talk: Knowledge interoperability in the semantic web IEEE Intelligent Systems. 2008; 23:72–85.
7. Hirst G, St-Onge D. Lexical chains as representations of context for the detection and correction of malapropisms. C. Fellbaum (Ed). WordNet: An Electronic Lexical Database. MIT Press; 1995. p. 305–32.
8. Mihalcea R, Corley C, Strapparava C. Corpus-based and knowledge-based measures of text semantic similarity. Proceedings of AAAI 21st National Conference on Artificial Intelligence; Boston, MA. 2006. p. 775–80.
9. Lin D. An information-theoretic definition of similarity. Proc of Fifteenth International Conference on Machine Learning, ICML; Madison, Wisconsin, USA. 1998. p. 296–304.
10. Hughes T, Ramage D. Lexical semantic relatedness with random graph walks. Proceedings of Emprical Methods in Natural Language Proceesing; Prague, Czech Republic. 2007. p. 581–9.
11. Kaisser M. Question answering based on semantic roles. Proceedings of The Deep Linguistic Processing Workshop in 45th Annual Meeting of the Association for Computational Linguistics; 2007. p. 41–8.
12. Jarmasz M, Szpakowicz S. Roget's thesaurus and semantic similarity. Proceedings of RANLP Conference on Recent Advances in Natural Language Processing; Borovetz, Bulgaria. 2003. p. 212–9.
13. Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. Five papers on WordNet. CSL Report 43: Technical Report. USA: Cognitive Science Laboratory, Princeton University; 1990.
14. Halliday M, Hasan R. Cohesion in English. London: Longman; 1976.
15. Budanitsky A, Hirst G. Evaluating WordNet-based measures of semantic distance. Computational Linguistics. 2006; 32(1):13–47.
16. Green BF, Wolf AK, Chomsky C, Laughery K. Baseball: An automatic question-answerer. IRE-AIEE-ACM '61 Western Papers presented at the Western Joint IRE-AIEE-ACM Computer Conference; 1961. p. 219–24.
17. Gildea D, Jurafsky D. Automatic labelling of semantic roles. Computational Linguistics. 2002; 28(3):245–88.
18. Miller A, Walter Charles G. Contextual correlates of semantic similarity. Language and Cognitive Processes. 1991; 6(1):1–28.
19. Bollegala D, Matsuo Y, Ishizuka M. A web search engine-based approach to measure semantic similarity between words. IEEE Transactions on Knowledge and Data Engineering. 2011; 23(7):977–90.
20. Pizzato LAS, Molla-Aliod D. Extracting exact answers using a Meta question Answering system. Proceedings of the Australasian Language Technology Workshop. ALTW05; 2005. p. 1–8.
21. Ferrandez O, Izquierdo R, Ferrandez S, Vicedo J. Addressing ontology-based question answering with collections of user queries. Information Processing and Management. 2009: 45(2):175–88.
22. Mohan K, Aramudhan M. Ontology based access control model for healthcare system in cloud computing. Indian Journal of Science and Technology. 2015 May; 8(S9):213–7. DOI: 10.17485/ijst/2015/v8iS9/53617.
23. Gayathri R, Rajalakshmi V. Connecting health seekers and health care knowledge portal by a semantic approach. Indian Journal of Science and Technology. 2016 Mar; 9(10):1–7. DOI: 10.17485/ijst/2016/v9i10/88889.
24. Rubenstein H,Goodenough JB. Contextual correlates of synonymy. Communications of the ACM. 1965 Oct; 8(10):627–33.
25. Gabrilovich E, Markovitch S. Wikipedia-based semantic interpretation for Natural Language Processing. Journal of Artificial Intelligence Research. 2009; 34:443–98.