Question Answering Systems Performance Evaluation – To Construct an Effective Conceptual Query Based on Ontologies and WordNet

Santosh Kumar Ray[1], Shailendra Singh[2], B.P.Joshi[3]

[1]Birla Institute of Technology, Muscat, Oman, santosh@waljatcolleges.edu.om,
[2]Samsung India Software Center, Noida, India, shailendra.s@samsung.com,
[3]Birla Institute of Technology, Noida India, bp_joshi@yahoo.com

Abstract. Question Answering Systems play a significant role to retrieve exact answers for user's specific questions. In answer retrieval process, they employ query expansion methods which play a major role to expand scope of original questions in correct sense. In this paper, we have carried out an extensive survey of few popular web-based open domain Question Answering Systems and critically evaluated their performances on a set of 300 questions from 30 different domains collected from standard resources including TREC to conclude our results. On the basis of findings, we have suggested an efficient query expansion framework that uses multiple ontologies retrieved from semantic web search engine such as Swoogle and combines them with WordNet to disambiguate the context. The proposed approach successfully constructs a conceptual query for user's questions to retrieve relevant answers. We have experimented on a set of 300 questions to judge the effectiveness of the proposed approach.

Keywords: Question Answering System, Ontology, Swoogle, Query expansion, Performance Evaluation

1 Introduction

Today, World Wide Web has become the chief source of information for everyone from general users to experts, students to researchers, to fulfill their domain specific needs. Search engines like Google help the users to find the relevant information based on the keyword searching and retrieve a large number of links. However, in many cases none of the retrieved web pages contain the relevant answer. In last few years, Question Answering Systems have emerged as a good alternative to provide relevant answers of the user queries in succinct form.

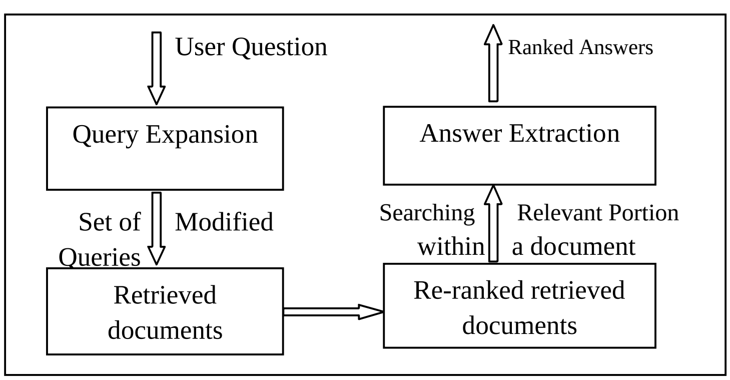

Fig. 1. A typical Question Answering System architecture

A typical Question Answering System, as shown in Fig. 1, takes user's question in some natural language as an input. This question is then optionally modified using some query modification technique (also called query expansion) and output of this modification process is a set of queries similar in meaning to the original question. These modified questions are fed into knowledge repository which may be either predefined corpus as in START [16], or entire World Wide Web as in AnswerBus [1], and Inferret [9]. Documents containing answers of the user query or modified queries are retrieved from the knowledge repository and re-ranked based on their relevance to the user query. Finally the most relevant portions of the predefined number of documents along with links are presented as answers to the user's question.

In this paper, we are performing comparative analysis of some popular web-based open-domain Question Answering Systems using a corpus of 300 questions as described in TREC [17], world fact book 2008 [21], Worldbook [22], and other standard sources. We observed that for web-based open domain Question Answering Systems, one or two ontologies are not sufficient to identify the correct sense of the words. Therefore, we are using multiple ontologies and combining them with WordNet [18] to disambiguate and identify the correct senses of the concepts in the user query. At present, it is very difficult to find the suitable set of ontologies. There are few semantic web search engines available on World Wide Web such as Swoogle [5], OntoSearch [20], AKTiveRank [2], OntoClean [8], and OntoKhoj [13] which are maintaining repositories for a large number of domain ontologies. We are using Swoogle as Ontology database for our query expansion method because Swoogle is having the largest number of ontologies and updates its ontology base periodically.

This paper is organized as follows: section 2 provides details of Question Answering Systems and ontologies related work. In section 3, we explain the proposed method of comparative analysis to evaluate performance of selected Question Answering Systems and performance of these Question Answering Systems based on the proposed method. In section 4, we explain the proposed method for query expansion using multiple ontologies and WordNet. We have shown our observations and results in section 5. In the last section, we have stated our conclusion and future directions to build an Intelligent Question Answering System.

2 Question Answering System Related Work

There are number of Question Answering Systems like START [16], AnswerBus [1], BrainBoost [3], PowerAnswer [14], Inferret [9], and Yahoo Answering System [19] etc. running on Web to fulfill user needs. One important factor for Question Answering Systems is to judge the correctness of retrieved answers against questions fed by the user. TREC [11] adopted a method to judge the correctness of the answers which has been accepted widely by several Question Answering Systems. Evaluation in TREC is essentially based on the F-measure to assess the quality of response in terms of precision and recall. There are several inherent requirements to compute F-measure, as discussed in [7], that makes the TREC method inappropriate for evaluations of domain-independent Question Answering Systems. [7] has done an extensive useful research for the evaluation of Question Answering Systems but their experiment didn't include Question Answering Systems based on statistical approaches. Also, there has been no consideration to the user's ease of answer interpretation in available literature.

Use of ontologies for query expansion has become popular in recent years. However more focus has been given on the use of single domain ontology while use of multiple ontologies is quite rare. [15] has combined Web ontology and WordNet together. However focus of their work is to create web document representation rather than query expansion. In [10], WordNet is used for word sense disambiguation. However, they have shown a small improvement in word sense disambiguation by combining WordNet with Google. [4] has used multiple ontologies for query expansion. However, their experiments are using limited number of ontologies. In contrast, our proposed approach is using multiple ontologies accessed from Swoogle which dynamically includes the ever increasing ontologies on the semantic web. Further, they have not included WordNet for their query expansion method. Our proposed system is similar to system discussed in [6] that uses ontologies pool obtained from Swoogle and other lexical resources such as WordNet. The aim of their system is to find all possible keyword sense using ontology pool. On the other hand, our system aims to find only those senses of keywords that are closer to the domain of the other keywords existing in the user question. Keeping in view the time complexity and large resources provided by Swoogle, we have limited ourselves to Swoogle for searching of ontologies.

3 Performance Evaluation of Web-based Question Answering Systems

Question Answering System analysis needs inception of an evaluation approach to provide a fair opinion about each Question Answering System and about its performance. To carry out performance evaluation, we have done exhaustive search for Question Answering Systems available on World Wide Web using Google and selected top five systems like START, AnswerBus, PowerAnswer, BrainBoost, and Inferret based on their working methods so that the analysis can cover each possible approaches used by Question Answering Systems. None of these Question Answering Systems are supposed to be specialized in any particular domain and hence domain specific bias has been avoided. We carried out our survey in two components: Question Collection and Answer Evaluation. The detailed description of each component is given below.

3.1 Question Collection

In Question Answering System performance analysis, the selection of Question Set draws major attention. We tried to avoid any bias in the evaluation while preparing the Question Set for evaluation of Question Answering Systems. We have chosen a set of 300 questions from 30 different domains and categorized into six groups: *Factoid Questions, List Questions, Contextual Questions, Textual Image Retrieval Questions, Biographical Questions,* and *Other Complex Questions* as described in [11] [12]. We have collected as many as possible questions of different formats from each of the question types. This set of questions includes all types of wh-questions, number (How many), and non-wh-questions (starting with verbs, auxiliary verb). We also ensured that all the questions in our collection have definite answer(s) in top 10 documents returned by Google.

3.2 Answer Evaluation

The evaluation criteria do a major job in the completeness of performance analysis. We have divided all the domains into 10 groups and each group is having 3 domains. We came up with evaluation strategy, in which each group will be evaluated by two persons so that correctness of the answers of the questions could be judged properly. We feed each question in listed Question Answering Systems and analyze retrieved answers. The evaluation is performed in two phases. In the first phase of evaluation, we identify whether Question Answering Systems are able to retrieve answers of the asked questions or not and judge the quality of retrieved answers using carefully chosen 11 criteria as shown in table 1. These criteria are based on two parameters which are as follows:

- How easily answers are available to the user?
- How easily they are understood by an average user?

To measure criteria’s importance, we introduce a weighted scheme where weight can be assigned from [0-1] with an interval difference of 0.1. The weight 0 shows non-retrieval of answers while weight 1 shows that retrieved answers are highly relevant and higher weight was assigned in case of conflict as shown in table 1.

In the second phase of evaluation, modified form of the same set of questions were asked from Question Answering Systems. We have modified original questions using *Changing keywords positions*, *Word Substitution, Changing the voice, Sentence formatting, Expansion/Contraction of acronyms, and Factual Rephrasing.* The aim of the modification of the question was to test whether or not the Question Answering Systems were able to answer the rephrased questions as different users generally use varied format to ask the same thing. Since these questions were already answered by them in the first phase of evaluation, it is obvious that answers of these questions were already present in the knowledge base of the Question Answering Systems.

Table 1. Measuring criteria for weight assignment.

S. N	Measuring Criteria	Wt
1	Exact answer in the expected format displayed on one of the link.	1
2	Correct answer displayed on one of the link but it requires small interpretation. (e.g. interpreting proper nouns)	0.9
3	Correct answer in expected format found in the first paragraph of the web page pointed by one of the top five links	0.8
4	Correct answer in expected format found in the first paragraph of the web page pointed by one of the next top 5 links	0.7
5	Correct answer in expected format in one of the documents pointed by any of the top 10 links and the document requires careful study	0.6
6	Correct answer in expected format found in the first paragraph of the web page pointed by one of the links ranking 11 and more	0.5
7	Correct answer found in one of the links with depth 2 or more than 2	0.4
8	Answer found in first paragraph of web pages pointed by one of the links ranking 1-10 but finding answers require some manual interpretation of sentences	0.3
9	Answer found in web pages pointed by one of the links ranking 1-10 but finding answers require some manual interpretation of sentences	0.2
10	Correct answer in expected format found in the first paragraph of the web pages pointed by one of the links ranking 11 and more and finding answers require some manual interpretation of sentences	0.1
11	No answer at all	0

3.3 Answer Evaluation Results

The results of the first phase of evaluation are shown in column 2 of table 2 where the highest percentage match of correct answers is retrieved by Inferret which uses statistical approaches to retrieve answers. These results show that recall is more important than precision in the case of very large corpus like World Wide Web. AnswerBus and PowerAnswer have performed consistently well in TREC evaluations because of sophisticated NLP techniques for query and document processing. In case of large knowledge repository like World Wide Web, overall performance of the Question Answering System is influenced more by recall than by precision. This explains the lower performance of NLP-based Question Answering Systems such as PowerAnswer as compared to Inferret. Still their performance is much higher than BrainBoost and START mainly due to two factors: Using World Wide Web as knowledge repository and very sophisticated processing of the retrieved documents resulting into the high precision. The BrainBoost is on 4^{th} position because it first tries to find the answer from the set of similar questions prepared by answers.com and it searches the web up to a certain degree. START retrieves the lowest percentage match of correct answers because it does not perform search on the entire web rather searches in its own created knowledge database. The next important question is judging the quality of retrieved answers. Are Question Answering Systems able to retrieve correct answers in a simple manner? Therefore, we performed evaluation of the answers in terms of quality using criteria's discussed in section 3.2. In table 2, N_1 represents the no. of retrieved answers which are having weight 1.0 while P shows the precision value of answers having weight 1.0 out of total number of correct answers (N).

Table 2. Performance of Question Answering Systems for the original set of 300 questions.

Question Answering System	N (Correct Answers Retrieved out of 300)	N_1 (Correct Answers Retrieved with weight 1.0 out of 300)	$P = (N_1 / N)$ (Precision Value of correct answers having weight 1.0)
Inferret	225	153	0.680
AnswerBus	206	156	0.757
PowerAnswer	200	171	0.855
BrainBoost	162	137	0.845
START	144	139	0.965

In table 3, we show that the quality of correct answers retrieved by START is highest. START is having 139 correct answers of weight 1.0 (out of 144 correct answers) while Inferret is on the bottom because quality of answers was not appreciated by the user. We also classified the retrieved answers, based on the weighted scheme, into three categories that is good, average, and poor to get a better picture of each Question Answering Systems' performance. We consider an answer with score 1-0.8 as "good", from 0.7-0.5 as "average" , and "poor" (less than 0.5) as shown in table 3.

Table 3. Categorization of correctly retrieved answers as good, bad, and poor answers.

Question Answering System	Good Answers	Average Answers	Poor Answers	Total
Inferret	194	13	18	225
PowerAnswer	190	7	3	200
AnswerBus	192	8	6	206
BrainBoost	154	3	5	162
START	142	2	0	144

In the second phase of evaluation, we have modified 300 original questions and retrieved results. We found a steep degradation in the retrieval of correct answers. In table 4, we show the correct retrieval data for each Question Answering System when the user asked the same question in a different format. We are calculating decline in recall percentage using following formula:

$$F = (N-N_2)*100 / N \quad . \qquad (1)$$

where N_2 represents the number of correct answers retrieved for modified questions. Table 4 shows that that BrainBoost is very sensitive in the identification of questions and very much dependent on question language while Inferret shows the lowest decline which can be mainly attributed to its efficient usage of large corpus like World Wide Web.

Table 4. Retrieval performance of each Question Answering System after question modification.

Question Answering System	N (as in Table3)	N_2	F
Inferret	225	173	23.11%
AnswerBus	206	147	28.64%
PowerAnswer	200	138	31%
START	144	89	38.19%
BrainBoost	162	82	49.38%

The results of table 2, table 3, and table 4 clearly prove the inefficiency of the existing Question Answering Systems in question interpretation and retrieval of correct answers. Their retrieval of correct answers is very much dependent on "question form" and not on the meaning of the question. We are showing this in fig. 2 by considering 20 questions which are modified by changing parameters discussed in section 3.2 for each Question Answering Systems. In case of changing keyword positions, AnswerBus and Inferret were remained mostly unaffected with a retrieval result of result of 80% and 85% while performance of START, BrainBoost, PowerAnswer went down drastically with a retrieval result of 50%, 45%, and 60% respectively. This is very much obvious in the given example question like "How far is the Mars from the Sun?" where the question is answered by BrainBoost but it failed to answer the modified question like "How far is the Sun from the Mars?”. Even very sophisticated Question Answering Systems like START and PowerAnswer were affected by this kind of question modification.

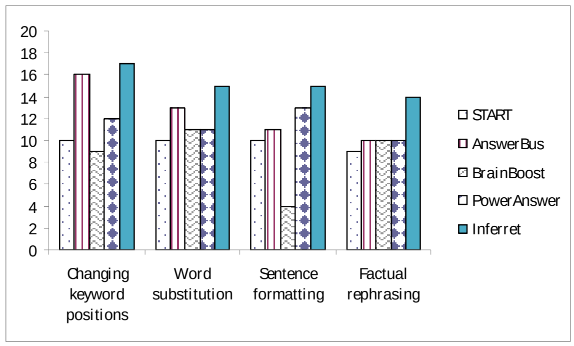

Fig. 2. Performance of Question Answering Systems with respect to various modifications.

In case of factual rephrasing, we observed a sudden decline in the performance of each Question Answering System. Their performance went down from 30% to 55% with Inferret and START respectively. These modifications imply about the reasoning capabilities of Question Answering Systems, as the question "How far is the Mars from the Earth?" was modified into “How far is the Mars from our planet?” with the hope that Question Answering System will modify “our planet” with “Earth”. With these observations, this is very much clear that existing Question Answering Systems need an efficient methodology to interpret asked question using discussed parameters to improve user satisfaction.

4 Multiple Ontologies and WordNet Based Query Expansion Methodology

Ontologies and WordNet are having rich information about domains and semantic relations between concepts. Query expansion methods based on only selected relations in WordNet have resulted into degradation of Question Answering Systems’ performance. On the other hand, if we use all the relations in WordNet in an uncontrolled manner then we get more number of semantically related words which forms a large number of modified queries against user’s single question. So, this method is not viable because of the requirement of more computational resources.

We are proposing multiple ontologies and WordNet based query expansion method. This method takes user question as an input and extracts key concepts from the question and then automatically finds the most relevant senses for key concepts of the question using WordNet. To disambiguate the correct sense, the method uses multiple domain ontologies retrieved from Swoogle semantic search engine. When key concept(s) from the user’s question is fed into Swoogle then it returns ontology classes describing the concept(s). In the proposed method, we compute semantic distance of the key concepts from retrieved ontology class, super class, and its subclasses and consider the class with the lowest semantic distance for the further processing. The complete proposed algorithm is given as follows:

Algorithm: Query_Expansion_MultipleOntologies
Input: User’s Question considered as Query (Q)

Output: Expanded query (Q_E)

Step 1: *Let T be a set of quadruples and defined as* $T=\langle C, O, W, R\rangle$, *where C denotes concept in user's question, O represents ontology for the concept C, W represents weight of an ontology O, and R represents one of the semantic relations retrieved from WordNet. Initially, T is empty.*

Step 2: *User enters a query Q.*

Step 3: *Extract key concepts* $C_1, C_2 \ldots C_k$ *from Q.*

Step 4: *User assigns* $W_1, W_2 \ldots W_k$ *weights to the concepts* $C_1, C_2 \ldots C_k$ *on the scale of 1-10. The concepts with higher weights are considered as important concepts.*

Step 5: *Search Swoogle for the combination of concepts using term dropping strategy. Query for the Swoogle is fed into Conjunctive Normal Form. All ontologies describing a concept combination are put into one group. Let us assume 'n' ontology groups defined as* $OG_1, OG_2, \ldots, OG_n$.

Step 6: *Let* $WN_{c1}, WN_{c2}, \ldots, WN_{ck}$ *be the domain set in WordNet for concepts* $C_1, C_2 \ldots C_k$. *.Elements in* WN_{ci} *are denoted by couple (S, R) where S is synonymous set for concept* C_i *and R is relation in WordNet that connects* C_i *to S.*

Step 7: *for (i=n; i>0; i--) do following for each ontology of* OG_i *group.*

$T= T \cup (\langle C_1, x, W_{OGi}, R_1\rangle, \forall x \in (O_{ij} \cap WN_{c1})) \cup (\langle C_2, x, W_{OGi}, R_2 \rangle, \forall x \in (O_{ij} \cap WN_{c2})) \cup \ldots \cup (\langle C_k, x, W_{OGi}, R_k\rangle, \forall x \in (O_{ij} \cap WN_{ck}))$ *(where Oij is* j^{th} *ontologies of ontology group* OG_i *)*

Step 8: *If T is empty for each* C_i *add one sense from all relations available in WordNet to T. We select most frequently occurring sense of the word and assign zero weight to the ontologies.*

Step 9: $Q_E= (C_1 \text{ OR } O_{11} \text{ OR } O_{12} \ldots \text{OR } O_{1m}) \text{ AND } (C_2 \text{ OR } O_{21} \text{ OR } O_{22} \ldots \text{OR } O_{2n}) \text{ AND } \ldots\ldots (C_k \text{ OR } O_{k1} \text{ OR } O_{k2} \ldots \text{OR } O_{kr})$ *where* O_{ij} *is the common ontology for concept* C_i *found in previous steps.*

We explain proposed algorithm with the help of an example. Let us consider a query in the form of question i.e. "*What is Jupiter's atmosphere made of?*". The key concepts found in this question are written as C_1 = "Jupiter", C_2 = "atmosphere", and C_3 = "made". The key concepts in the question are obtained by removing the wh-words and the stop words. Our system analyses the question and derives the expected answer type before actual removing of the wh-words and the stop words takes place. The key concepts are then presented to the user and the user assigns weights to each concept as W_1= 9, W_2=9, W_3=3 respectively. We do start searching of relevant ontologies from Swoogle using term dropping strategy. For the example query, the search strings are "Jupiter AND atmosphere AND made", "Jupiter AND (atmosphere OR made)" , "(Jupiter OR made) AND atmosphere" , "(Jupiter OR atmosphere) AND made" , "Jupiter", "atmosphere", and "made". Swoogle provides many advanced meta-tags for specific searches. We are using meta-tag called *"desc: term1"* to retrieve ontologies which are having "*term1*" in the description of the document, generally in the annotations. We are passing queries to Swoogle in Conjunctive Normal Form to retrieve ontologies which are relating given concepts in some meaningful way. Swoogle uses "AND" as a default logical operator. In step 7, we find common concepts between ontology group and WordNet group like in 5^{th} query, concept *planet* is common in both OG_5 and $WN_{Jupiter}$. Hence quadruple for this query will be < Jupiter, Planet, 9, hypernym> and will be added in T. We do step 7 and step 8 for all queries and get final set T = {<atmosphere, air, 18, synonym> <Jupiter, Planetary Object, 18, hypernym>, <Jupiter, Planet, 9, hypernym > <atmosphere, weather, 9, hypernym>, <make, constitute, 9, synonym>}. Therefore, final expanded query is defined as Q_E = [(Jupiter OR Planet OR Planetary Object) AND (atmosphere OR air OR weather) AND (make OR constitute)]. In the next section, we will run our proposed algorithm for large number of questions and will also compare results with existing Question Answering Systems.

5 Query Expansion Results

To judge the accuracy of proposed query expansion method, we have considered a set of 300 questions collected from standard sources like TREC etc. (same set as discussed in Sec 1) which are covering almost 30 different domains. We are representing top 20 questions and their corresponding expanded queries in table 5.

Table 5. Listing top 20 original questions and their corresponding expanded queries based on multiple ontologies and WordNet

1	What is Jupiter's atmosphere made of?	(Jupiter OR Planet OR Planetary Object) AND (atmosphere OR air OR weather) AND (make OR constitute)
2	Explain the reason for sky's blue color	(reason OR cause) (sky OR rainbow OR cloud OR lightning) AND (blue OR sky-blue) AND color
3	Which planet has the least	planet AND (least OR smallest OR minimum) AND surface AND
4	What capital is on the Susquehanna River?	(capital OR "state capital" OR means OR Centre OR "Graphic symbol") AND Susquehanna AND river
5	How far is Mars from our planet?	(far OR distant) AND (Mars OR "Red Planet") AND our AND (Planet OR "terrestrial planet")
6	What is famous invention by Marconi?	(famous OR celebrated OR known OR notable) AND (invention OR creativity OR creativeness OR "creative thinking" OR " creating by mental act") AND (Marconi OR Guglielmo Marconi**)**
7	What is the life expectancy of the average woman in Nigeria?	("Life Expectancy" OR "Life Expectancy at Birth") AND average AND woman AND (Nigeria OR Lagos OR Zaria OR "Yerwa Maiduguri" OR Niger OR Africa)
8	Give me the countries that border India.	(Countries OR country OR land) AND (border OR "has border" OR "borders on") AND(India OR Indian)
9	Which is the deepest sea?	(deepest OR deep) AND (sea OR Ocean)
10	What continent is India in?	(continent OR subcontinent OR landmass OR Asian OR African) AND (India OR Indian)
11	Which state has the longest coastline on the Atlantic Ocean?	(state OR province) AND (longest OR long OR length) AND coastline AND Atlantic AND Ocean
12	What fraction of the ozone layer is destroyed?	Fraction AND (Ozone OR "Ozone layer" OR stratosphere Or oxygen) AND layer AND depleted
13	What year was Beethoven born?	Year AND (Beethoven OR Ludwig van Beethoven OR music OR composer) AND(born OR "born in")
14	Who is the composer of opera semiramide?	Composer And (opera OR "comic opera" OR "opera bouffe" OR bouffe OR "opera comique") AND semiramide
15	What music did Debussy compose?	(music OR Bach) AND (Debussy OR Claude Debussey, Claude Achille Debussy) AND (composed OR compose OR composer OR "composed for " OR " is composed of" OR " composed for" OR " music composed by")
16	In what year did Arundhati Roy receive a Booker Prize?	Year AND ("Arundhati Roy" OR Arundhati) AND (get OR receive) AND Booker AND(prize OR award)
17	Who was the seventh president of	(Seventh OR 7^{th}) AND president AND (India OR Indian)
18	What Indian state has the highest life expectancy?	Indian AND (state OR province) AND highest AND ("life expectancy" OR " life expectancy at birth")
19	What does the Taiwan flag look like?	(Taiwan OR Taiwanese OR Taipei OR "South China sea") AND (flag OR "national flag") AND (look OR appear) AND (like OR "likes of")
20	What famous communist leader died in Mexico City?	Famous AND communist AND leader AND (died OR "died in year" OR death) AND Mexico AND (city OR Leon OR "Acapulco de Juarez" OR Tepic OR Culiacan OR Matamoros OR "Tuxtla

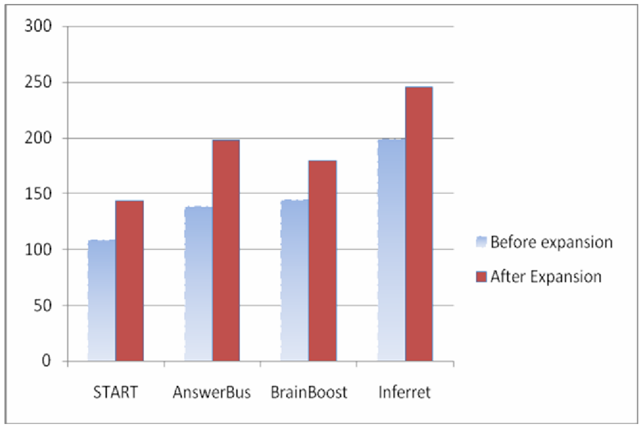

Fig. 3. Performance of Questions Answering Systems with respect to original questions and their expanded questions.

We fed 300 original questions and their expanded queries in Google separately and evaluated top 10 retrieved results for each original question as well as for each expanded query. With the set of original questions, we find satisfactory answers for 258 out of 300 questions while with the set of expanded queries satisfactory answers retrieved for 276 out of 300 questions. The proposed approach has retrieved 92% answers correctly which is an overall improvement of about 8% in comparison with answers retrieved for existing set of original questions.

To measure the performance of proposed approach, we have experimented with some existing popular web-based automatic Question Answering Systems like START, AnswerBus, BrainBoost, and Inferret. We reformulated expanded queries as per the Question Answering System specific format and fed them in all Question Answering Systems. The overall performance of each Question Answering System shows a significant increase in retrieving correct answers. The performance bar chart has been shown in figure 3 where BrainBoost and Inferret indicate an improvement of 25%, START exhibits an improvement of 33% while AnswerBus records maximum improvement of 44%. The overall average improvement is 31.75% on Question Answering Systems which is very significant as these systems are already using very sophisticated information retrieval techniques to retrieve correct answers. On the basis of experimented results, we can say that proposed approach is working reasonably quite well.

6 Conclusion and Future Direction

In this paper, we have evaluated the performance of START [16], AnswerBus [1], BrainBoost [3], PowerAnswer [14], and Inferret [9] open domain based Question Answering Systems. We have observed that Natural Language Processing (NLP) plays an important role in improving the quality of retrieved answers and their presentations. Very high quality of the answers retrieved by sophisticated Question Answering Systems such as START reflects that annotation of retrieved documents positively affect the quality of answers. While other Question Answering Systems based on NLP techniques such as PowerAnswer shows relatively higher percentage of good answers compared to statistical approach based Inferret. We found that Inferret is performing quite well even when questions are modified. This suggests that questions posed by user may match with different vocabularies used for World Wide Web content. Therefore, the designing of future open domain Question Answering Systems must be efficient to handle different forms of original questions.

Further, we have presented that how semantic web and WordNet can be effectively utilized for World Wide Web based Question Answering Systems. We have concluded on the basis of experimented results that combination of semantic web with vast and exhaustive lexical resources like WordNet can greatly improve the performance of the Question Answering Systems. We have proposed query expansion framework using ontologies and WordNet to expand the original question's scope conceptually. Similarly, this can be extended for other phases of Question Answering System. In future, we are intending to develop an efficient content based Question Answering System using ontologies and WordNet.

References

[1] AnswerBus, Question Answering System, website: http://answerbus.com

[2] Alani, H., Brewster, C., Shadbolt, N.: Ranking ontologies with AKTiveRank. In: 5^{th} International Semantic Web Conference (ISWC 2006). LNCS, vol. 4273, pp. 1-15, Springer-Verlag (2006)

[3] BrainBoost, Question Answering System, website: http://www.answers.com/bb/

[4] Dey L., Singh S., Rai R., Gupta S.: Ontology Aided Query Expansion for Retrieving Relevant Texts. In Advances in Web Intelligence, LNCS, vol. 3528, pp. 126—132, Springer, Heidelberg (2005)

[5] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V. C., Sachs, J.: Swoogle: A semantic web search and metadata engine. In Proc. 13th ACM Conf. on Information and Knowledge Management, (2004)

[6] Espinoza, M., Trillo, R., Gracia, J., Mena, E.: Discovering and Merging Keyword Senses using Ontology Matching. In 1st International Workshop on Ontology Matching at ISWC-2006 (2006)

[7] G. O. Sing, C. Ardil, W. wong, and S. Sahib, "Response Quality Evaluation in Hetrogeneous Question Answering System: A Black Box Approach", In Proceedings of World Academy of Science, Engineering and Technology Vol. 9, ISSN 1307-6884 , NOVEMBER 2005 .

[8] Guarino, N.,Welty, C.: An Overview of OntoClean. In Handbook on Ontologies, Springer pp. 151 - 172 (2004)

[9] Inferret, Question Answering System, website:http://asked.jp

[10] Ioannis P., Klapaftis, G., Manandhar S.: Google & WordNet based Word Sense Disambiguation. In Proceedings of the 22^{nd} ICML Workshop on Learning & Extending Ontologies, Bonn, Germany (2005)

[11] J. Lin and B. Katz, "Building a Reusable Test Collection for Question Answering", Journal of the American Society for Information Science and Technology, 2006, pp. 57(7):851-861.

[12] J. Lin, and D. Demner-Fushman, "Methods for Automatically Evaluating Answers to Complex Questions", Information Retrieval, 9(5), 2006, pp. 565–587.

[13] Patel, C., Supekar, K., Lee, Y., Park, E.: OntoKhoj: A semantic web portal for ontology searching, ranking and classification. In: Proceedings of the Workshop on Web Information and Data Management, pp 58 – 61 ,ACM Press (2003)

[14] PowerAnswer Question answering System, website:http://www.languagecomputer.com

[15] Sabrina T., Rosni A., Enyakong, T.: Extending ontology tree using NLP techniques. Proceedings of National Conference on Research & Development in Computer Science REDECS 2001, Selangor, Malaysia, (2001)

[16] START Question Answering System, website:http://start.csail.mit.edu.

[17] Text Retrieval Conference, http://trec.nist.gov

[18] WordNet, website:http://wordnet.princton.edu

[19] Yahoo! Question Answering System, website:http://answers.yahoo.com/

[20] Zhang, Y., Vasconcelos, W., Sleeman. D.: Ontosearch: An ontology search engine. In Proc. 24^{th} SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK, 2004 .

[21] CIA the world Factbook, website: https://www.cia.gov/library/publications/the-world-factbook/

[22] The World Book, Website: www.worldbook.com/