International Journal of Computer Science & Applications
Vol. 5, No. 2, pp 52 - 68

ALGORITHMS TO IMPROVE PERFORMANCE OF NATURAL LANGUAGE INTERFACE

M. R. Joshi,
Department of Computer Science, North Maharashtra University, Jalgaon.
E-mail: joshmanish@yahoo.com

R. A. Akerkar,
Technomathematics Research Foundation, Kolhapur, India
E-mail: raakerkar@yahoo.com

Abstract:

Performance of Natural Language Interface often deteriorates due to linguistic phenomena of Semantic Symmetry and Ambiguous Modification (Katz and Lin, 2003). In this paper we present algorithms to handle problems caused by semantic symmetry and ambiguous modification. Use of these algorithms has improved the precision of Natural Language Interface. Proposed shallow parsing based algorithms reduce the amount of syntactic processing required to deal with problems caused by semantic symmetry and ambiguous modification. These algorithms need only POS (Part of Speech) information that is generated by shallow parsing of corpus text. Results are compared with the results of basic Natural Language Interface without such algorithm. Dealing with linguistic phenomena using shallow parsing is a novel approach as we overcome the usual brittleness associated with in depth parsing. We also present computational results that produced comparative charts based on answers extracted for a same query posed to these two systems.

Keywords: Natural language interface, semantic symmetry, algorithm.

1. Introduction:

(Katz and Lin, 2003) described the phenomena of semantic symmetry and ambiguous modification and further elaborated how syntactic processing (by indexing syntactic relations) and analysis successfully handles these issues, resulting improved precision of Question Answering System. As compared with the simple Information Extraction driven approach for Question Answering System they additionally applied sophisticated NLP approach to improve performance. They obtained results by using syntactic relations, captured in terms of ternary expression. While one of the strong points of this algorithm was that it operates primarily on syntactic information alone, still it has to explicitly build syntactic relations using relations-indexing engine. A functional dependency parser Minipar was used and based on these results, typical subject-verb-object type ternary relations were generated.

In this paper, we present work that further eliminates the need of building ternary expressions to handle the problem caused by semantic symmetry and ambiguous modification. The modifications discussed below makes the algorithm available to a wide range of Natural Language Interfaces, which, due to the lack of full syntactic parsing capability, normally would have been unable to generate precise answer. The work is additionally important, we feel, for the search engines as it would filter out irrelevant documents among from mere keyword matching based fetched documents.

2. Overview

Information needs to be extracted from a vast dat a. Data is stored either in a 'structured manner' or in an 'unstructured manner'. In the structured representation, various attributes of data are identified separately and are maintained separately whereas in the unstructured representation data is stored merely as a sequence of characters. Interfaces to structured as well as unstructured data have been a major area of research since 1960.

LUNAR was the first usable Natural Language Interface to Database (NLIDB) appeared in late 1960s. LUNAR NLIDB contained chemical analyses of moon rocks and had a significant influence on subsequent computational approaches to natural language.

The process of establishing interaction between human being and machine was made successful in 1966 by ELIZA system, which was developed by Joseph Weizenbaum. ELIZA worked by simple parsing and substitution of key words into phrases stored in knowlwdge base. Though ELIZA did not employed any language related phenomena still it remains a milestone simply because it was the first time a programmer had attempted such a human-machine interaction with the goal of creating the illusion (however brief) of human-human interaction.

Since large information is available in unstructured manner, retrieving out relevant documents containing the required information was the primary goal of the interfaces of this category. This task is known as Information Retrieval. Pinpointing exact information called as Information Extraction is the next step and development of Question Answering System is advanced step of user machine interaction.

Information extraction (IE) is the name given to any process that selectively structures and combines data that is found, explicitly stated or implied, in one or more texts. Question answering system establishes non-formal communication with the user and answers to the queries posed by the user.

The concept of combining Natural Language Processing (NLP) techniques with large-scale Information Retrieval and Information Extraction is not new, but yet not successful to the desired extent. Fagan (1987), Croft and Lewis (1987), Smeaton et al. (1994), Strzalkowski et al. (1996), Zhai et al. (1996) and Arampatzis et al., (1998) experimented with various approaches by carrying syntactic analysis. However, none of these experiments resulted in substantial improvement in precision or recall, and on the other hand often resulted in degraded performance. Rolf Schwitter, Michael Hess et al., (2000) developed ExtrAns system for restrictive domain that answer users plain English language query related to UNIX operating system. In this system they took care of synonyms, hyponyms by syntactic parsing.

Boris and Lin (2003) of MIT showed that NLP technology in general is not powerless. Performance drop or performance improvement of the inter face, on the contrary depends on the manner in which NLP techniques has been applied. Catz and Lin (2003) have identified two broad linguistic phenomena that are difficult to handle with the information extraction driven paradigm. They proposed to build ternary expressions in order to tackle the linguistic phenomena of semantic symmetry and ambiguous modification.

Boris Katz at MIT's Artificial Intelligence Laboratory developed the START (SynTactic Analysis using Reversible Transformations) Natural Language System. It is a software system designed to answer questions that are posed to it in a natural language. START uses several language dependant functions like parsing, natural language annotation to present the appropriate information segments to the user.

KUQA system presented in TREC-9 (2000) developed by Soo-Min Kim and his colleagues categorized questions based on expected answer and then used NLP techniques as well as Wordnet for finding candidate answers which suits in corresponding category. They however not handled any linguistic phenomena.

In TREC-13 (2004) Michael Kaisser and Tilman Becker presented QuALiM system that used complex syntactic structure. Based on certain syntactic description question patterns were identified. Syntactic description of prospective answers was also maintained and accordingly system generated answer from documents retrieved using google search engine.

3. Semantic Symmetry

One of the phenomena that cannot be easily handled by linguistically uninformed Natural Language Interface System is **semantic symmetry.** Semantic Symmetry occurs when an entity is used as subject as well as an object in different sentences. As selectional restriction (keyword matching) in different sentences is based on such entities, system generates wrong answer. Following Example illustrates the phenomenon of semantic symmetry and demonstrates problems caused thereof.

Question : Who killed militants ?

Candidate Answer 1 : National army soldiers killed six militants.
Candidate Answer 2 : Militants killed 13 bus passengers.

In above sentences ‘Militants’ is an entity (POS – Noun) which acts as subject in sentence 2 and as an object in sentence 1. The selectional restriction for the subject of ‘kill’ is word ‘Militants’ in one sentence and the selectional restriction for the object is also word ‘Militants’ in another sentence.

Hence the question fetched two candidate answers on the basis of keyword matching and both sentences has different meaning altogether. System wrongly returns both sentences as answers if such questions are not handled explicitly. Such questions are referring to the sentences, which contains semantic symmetry relationship.

Two sentences 'National army soldiers killed six militants.' and 'Militant killed 13 bus passengers.', are similar at the word level, but they have very different meanings and should be presented as answer appropriately by considering meaning of these sentences. In these cases, lexical content is insufficient to determine the meaning of the sentence.

Table 1 shows additional sentences showing semantic symmetry.

(1) The bird ate the snake. (2) The Germans defeated the French. (3) President of India visited flood affected area.	(1) The snake ate the bird. (2) The Germans were defeated by the French. (3) Bill Gates visited President of India

Table 1: Shows some more examples of Semantic symmetry

4. Ambiguous Modification

Adjectives are often ambiguous modifiers. If a paragraph contains a pool of adjectives and nouns, any particular adjective could potentially modify many nouns. Under such circumstances, a Natural Language Interface System cannot achieve high precision without exactly identifying the association between adjective and nouns. Following Example illustrates the phenomenon of Ambiguous Modification and demonstrates problems caused thereof.

Question : What is the largest volcano in the Solar System?
Candidate Answer 1: Mars boasts many extreme geographic features; for example, Olympus Mons, the *largest volcano in the solar system*.
Candidate Answer 2: The Galileo probe's mission to Jupiter, the *largest* planet *in the Solar system*, included amazing photographs of the *volcanoes* on Io, one of its four most famous moons.
Candidate Answer 3: Even the *largest volcanoes* found on Earth are puny in comparison to others found around our own cosmic backyard, *the Solar System*.
Candidate Answer 4: Olympus Mons, which spans an area the size of Arizona, is the *largest volcano in the Solar System*.
Candidate Answer 5: I*n the Solar System,* Olympus Mons, which spans an area the size of Arizona, is the *largest volcano*.
Candidate Answer 6: I*n the Solar System, the largest planet Jupitor has more than 50 volcanos*.

Table 2: Examples illustrating ambiguous modification

Ambiguous modification related with adjective occurs when an entity behaves in *un*restrictive manner and can associate to more than one noun in a particular sentence. Since the adjective *largest*, involved in an adjective-noun modification relation, is not very constrained in its possible choices of head nouns, and hence is free to "float" among nouns.

In above example 'largest' is the adjective which acts as modifier in adjective-noun modification relation, whereas 'In the Solar System' is used to specify scope. Candidate answer 2 and 3 are correct at lexical level, but wrong at meaning level. In Candidate answer 2, *largest* modify the incorrect head noun. In Candidate answer 3, *in the Solar System* does not modify the correct head noun.

5. Algorithms

5.1. Pattern of Sentences and Semantic Symmetry

All sentences are categorized as Active voice sentences or Passive voice sentences. Sentences in Active voice follow the structure of SVO that is Subject followed by Verb followed by Object, whereas Passive voice sentences have OVS structure that is Object followed by Verb followed by Subject. The sentence in Active voice can also be presented in Passive voice by changing position of Subject and Object and changing verb to past participle form.

Questions may be of type XVO where X is the subject entity, which we want to find out as an answer, V is the verb and O is the object entity. 'Who killed Militants? ', is XVO type question. Now only <u>those active sentences</u> are correct in which object entry do not come before verb entry that is the order 'VO' is maintained. If this correct sentence available in passive voice form (OVS) then sequence of Verb and Object in question shall not match, but we can understand that this sentence is in Passive voice form by looking at the verb. If the verb in the sentence is in Past Participle form it indicates that the sentence is in Passive voice form. Hence sequence is not matched and POS of verb not matched points to the correct answer. So, we can formulate rules that work upon sequence of entities and tense of verb.

This same logic is applicable for the questions of type – SVX that is X is the Object entity which we want to obtain as an answer, V is the verb and S is the subject entity. 'Militants killed whom?' is an example of such category. Any sentence in which word 'Militants' appears after Verb 'killed', is an object entity whereas we are looking for sentences in which word 'Militants' is placed as Subject. It emphasizes the importance of sequence of words in sentence. In Passive voice sentence SVO sequence is not followed but the passiveness of the sentence can be determined by the tense of the verb, which helps in deciding the correctness of the sentence. This fact underlines importance of tense of verb (Past tense or Past Participle).

Even if the question is in Passive voice form, the above-mentioned logic works. XOV* type of questions falls in this category. 'By whom Militants were killed?' is the example of this type of questions. X – is the Subject which we expect from the Question

Answering System, O is the object and V* is the past participle tense verb used in the question.

5.2. Algorithm

Based on the study of patterns of questions and the sentences having semantic symmetry we have formulated rules that pick up exact sentences as answer among from number of candidate answer sentences. Following algorithm shows how the problem caused by semantic symmetry is solved with only shallow parsing applied on the corpus.

Two important factors considered are -

1. Sequence of keywords in question and in the candidate answer sentence
2. POS of keywords

The algorithm scans each candidate answer sentence and applies following rule to check whether that sentence is correct answer sentence or not.

Rule 1 -
If (sequence of keywords in question and candidate answer **matches**) then
    If (POS of verb keyword **are same**) then
        Candidate answer is Correct
Rule 2 -
If (sequence of keywords in question and candidate answer **do not match**) then
    If (POS verb keyword **are not same**) then
        Candidate answer is Correct
Otherwise -
    Candidate Answer is wrong

Following example shows how these rules can extract out correct sentences as an answer by handling semantic symmetry problem for these three questions.

Question 1 : Who killed Militants ?	XVO
Question 2 : To Whom Militants killed?	SVX – XSV
Question 3 : By whom Militants were killed?	XOV*

Seven Candidate answer sentences are fetched by the system which works merely on keyword matching principal. These candidate answer sentences are fetched from the corpus. Details about the corpus are given in next session. Annexure-1 shows how application of rules filters out exact sentences for example question.

5.3. Pattern of Sentences and Ambiguous Modification:

After studying structure of sentences we have formulated rules that are based on shallow parsing. A candidate answers fetched based on keyword matching can be tested for correctness using these rules.

Every sentence that is amenable for problem due to Ambiguous Modification contains one adjective and more than one noun. One of these nouns is used for defining the scope whereas the other is pointing to the identifier which we are looking for. These nouns can easily be distinguished and sequence of these two nouns and adjective is the important factor that leads to the rules.

5.4 An Algorithm:

The algorithm scans each candidate answer sentence and applies following rule to check whether that sentence is correct answer sentence or not.

We have identified the adjective as Adj, Scope defining noun as S_N and the Identifier noun as I_N.

Rules –

If the sentence contains keywords in following order –

Adj a S_N Where a indicate string of zero or more keywords.

Then

Rule$_1$-a → If a is I_N ==➔ Correct Answer Or

Rule$_1$-b → If a is Blank ==➔ Correct Answer

Else

Rule$_2$ → If a is Otherwise ==➔ Wrong Answer

If the sentence contains keywords in following order –

S_N a Adj ß I_N, where a and ß indicate string of zero or more keywords.

Then

Rule$_3$ → If ß is Blank ==➔ Correct Answer

Value of a Does not matter

Else

Rule$_4$ → If ß is Otherwise ==➔ Wrong Answer

Annexure-2 shows how rules Rule $_1$ to Rule $_4$ helps to find out correct answer among from number of candidate answers shown in Table-2.

6. Experiment using the System – ENLIGHT

In order to demonstrate our ideas, we have developed a natural language question answering system ENLIGHT (intelligEnt Natural Language Interface). Architecture of this system is given in annexure-3.

The modules developed in ENLIGHT are as follows -

1. Pre-processor

This module prepares platform for the Intelligent and Effective interface. This module transfers raw format data into well-organized corpus with the help of following activities.

- Sentence segmentation
- Removal of stop-words
- Keywords identification

2. Shallow Parsing

Using a third party software tool 'part of speech' of each keyword as well as grouped local words was identified. This information is used in improving the precision of the interface.

- Identifying group of words with specific meaning (Local Word Grouping)
- POS tagging

3. Question Processing

Using following sub-modules analysis of question is done and information regarding question is identified and maintained for future use.

- Keyword separation
- Factoid questions / List Questions / WH Questions
- Question Tagging

4. Candidate Answer Retrieval

Finding out sentences from corpus which has more number of keyword matches as with the keywords of the question is the task performed in this module. Presentation of the answer to the user is done using this module.

- Keyword Matching
- Answer Presentation

5. Answer Rescoring

Mere Part -of-speech (POS) of all keywords are obtained earlier using third party software tool. This shallow parsing works as the basis for improvement of the answer by eliminating out sentences which looks correct at the word level but wrong at the meaning level.

- Identifying appropriate and precise answer by tackling **Semantic Symmetry** and **Ambiguous Modification**

6. Incorporating Intelligence

Based on the feedback given by the user to the interface certain keywords from question were associated to suitable sentences. Different options of feedback train system appropriately. Interface is programmed to identify questions which need explicit action of execution of shallow based algorithms.

- Getting feedback from user
- Identifying questions required to be processed through Answer Rescoring module

The ENLIGHT system retrieves answers by Information Extraction driven approach and further improves precision by filtering out sentences containing semantic symmetry.

Algorithm to handle problems caused by semantic symmetry is implemented in ENLIGHT system as Answer Rescoring module. We are using QTAG – a probabilistic

Parts-of-Speech tagger - for getting POS of keywords. We have integrated QTAG as a module in our system.

7. Results

We do not come across with any claim of resolving certain linguistic phenomena using only shallow parsing. We are first to provide algorithms that deal with semantic symmetry and ambiguous modification, the frequently occurring phenomena in English Language. We have implemented these algorithms in ENLIGHT system and obtained results are tested on following criterions.

7.1 Preciseness

ENLIGHT System with Answer Rescoring module (Handling Semantic Symmetry and Ambiguous Modification) and without Answer Rescoring Module, are the two modes of the system. Results obtained from these two modes of ENLIGHT system are studied and compared.

The test corpus used in our experiment is electronic versions of news from newspapers. Approximately 2000 news extracts and information broacher of few institutions was provided to the ENLIGHT system.

The test set consisting questions that illustrate the linguistic phenomenon of semantic symmetry and Ambiguous Modification is formed; some of these questions are shown in Table 3. Questions like 'Who did France beat for the World Cup?' (Q- 69.2), 'Who did Foreman defeat for his first heavyweight championship?'(Q- 77.4) and 'What Shiite leaders were killed in Pakistan? ' (Q – 136.7) are the questions taken from **TREC-2005 Question Database.**

Who killed militants?
Who did Foreman defeat for his first heavyweight championship?
What do frogs eat?
Who visited Bill Gates?
Who did France beat for the World Cup?
What Shiite leaders were killed in Pakistan?
What is the largest volcano in the Solar System?
Which is the longest river in the world?

Table 3: Sample questions used in the study

Table 4 shows a sample output of the basic system that does not handles the problem caused by semantic symmetry.

Q : Who killed militants?
AS1- National army soldiers killed 6 militants. AS2- Militant killed 13 bus passengers. AS3- Militant killed 4 policemen. AS4- BSF soldiers in an encounter killed 3 militants in Rajauri district. AS5- In bomb attack in Srinagar militant killed 7 peoples, 25 injured. AS6 - …

Table 4: Sample output from basic system

System returned 10 sentences as answer that contain the word 'killed' and 'militants'. Out of these sentences only 4 sentences are correct. ENLIGHT system with above mentioned algorithm (Answer Rescoring Module) has displayed exactly these 4sentences as answer. This result is shown in Table 5.

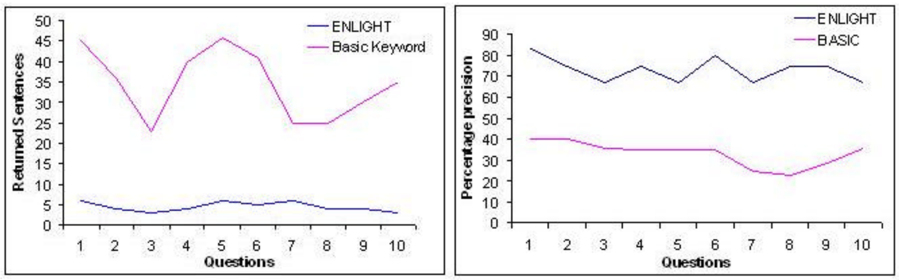
Q : Who killed militants?
AS1 – National army soldiers killed 6 militants. AS4 - BSF soldiers in an encounter killed 3 militants in Rajauri district. AS6 - 3 Militants were killed after 12 hour long shoot out in Poonch sector of Jammu and Kashmir. …

Table 5: Sample output from ENLIGHT system

The comparative results are displayed in Table 6. Average number of sentences returned, average number of correct sentences and average precision are the parameters for comparison.

	ENLIGHT	Basic Keyword Matching
Average Number of sentences returned as Answer	3	34.6
Average Number of correct sentences	2.63	6
Average precision	84 %	32 %

Table 6: Comparison between ENLIGHT and basic system



(a) (b)

Fig-1 (a) shows comparative chart based on number of sentences returned as answer for a query between ENLIGHT system and basic keyword matching system Fig-1 (b) shows improvement in percentage precision between ENLIGHT system as compare to basic keyword matching system.

Above figure represents comparison between two modes of ENLIGHT system. For various questions ENLIGHT system achieved a precision of 0.84, while the basic keyword matching system achieved only a precision of 0.32. As the ENLIGHT system returned less number of sent ences, obtaining proper answer among from these sentences is easier for the user. Hence, improvement in preciseness is claimed.

7.2 Response Time

We have compared the response time of ENLIGHT system with Sapere system developed Katz and Lin. As the exact data provided to Sapere system was not available, we have to follow a different approach for the comparison. Both systems carry out syntactic parsing of the corpus as well as question. Both systems used third party software for parsing and these intermediate results are further used in processing to obtain answer for the query. Therefore, we have decided to compare time taken to carry out parsing. Sapere system uses Minipar functional dependency parser whereas ENLIGHT system uses a Part of Speech tagger - QTAG tagger.

The result of comparison is shown in Table 7.

Type of Data and No. of words	Time Required by QTAG (Used in ENLIGHT)	Time Required by Minipar (Used in Sapere)
News extract Times of India. 202 Words	1.71 s	2.88 s
Reply START QA System 251 Words	1.89 s	3.11 s
University Information NMU Broacher 274 Words	1.55 s	2.86 s
Brazil Information Source: Wikipedia 226 Words	1.67 s	3.13 s
AVERAGE	**1.705 s**	**2.995 s**

Table 7: Time required for syntactic parsing

Above results are obtained after executing both QTag Tagger and Minipar Parser from an application developed using Visual Basic 6.0.

Above results clearly indicate that time taken by the Tagger is less than parser. Hence the response time of the ENLIGHT system is better than the system developed by Katz and Lin.

System developed by David Ahn et al [7], Hang Cui et al [9] are the systems presented in TREC-2004, also used output generated by Minpar parser for performing further processing to deal with linguistic phenomena. ENLIGHT system used QTAG POS tagger to carry out surface parsing and using these results further processing is done.

7.3 Adaptability

The rules for tackling problems caused by semantic symmetry and ambiguous modification are not based on ternary expressions as in case of system developed by Katz and Lin, ENLIGHT system is able to cope up with questions containing features like –

i) Additional Keyword – Question like 'who killed French Militant?' contains additional keyword 'French'. Sapere system of Katz and Lin builds ternary expression using output obtained by Minipar Parser. For question 'Who killed militant?' Sapere system forms ternary expression ([?x Kill Militant]). Adding in 'French' keyword in this ternary expression is bit difficult and not discussed in Sapere system (Katz, 2004).

ENLIGHT system can cope up with this type of additional keywords as the rules looks for sequence of all matched keywords and POS of these all keywords.

ii) Use of synonym – If the question contains synonym for certain keywords , ENLIGHT system shall not fetch precise answer in first run. But ENLIGHT system provides a facility so that user can train system to associate this synonym with certain keyword. Once it is done ENLIGHT system easily manage such questions.

Conclusion:

Due to the involvement of language related complex phenomena development of complete and comprehensive Natural Language Interface has become a challenging task. The uniqueness of the ENLIGHT interface lies in tackling language related phenomena using shallow parsing based algorithms. As these are the first of its type, the results obtained using shallow parsing may motivate more research in this direction.

System gives good result while it handles semantic symmetry and Ambiguous Modification. The precision of ENLIGHT system has improved with these algorithms. As compared to other systems which work upon detail parsing of the corpus, ENLIGHT generates correct results with improved response time. This complete exercise depends upon the output of a QTAG tagger. But sometime QTAG a POS tagger do not differentiate between past tense verb and third past participle verb. Because of this sometime system displays additional sentences as answer though they are not.

We are testing various other POS tagger for getting correct results as an output of shallow parsing. A CLAWS POS tagger was used to tag 100 million words of the British National Corpus (BNC) is one of the taggers that have been tested. We have tested some sentences, which are not correctly tagged by QTAG, using CLAWS system and found that CLAWS POS tagger gives better results. We hope performance of ENLIGHT system shall be improved further with the use of a CLAWS system.

References:

A. Arampatzis, Th.P. van der Weide, C.H.A. Koster, and P. van Bommel. , (1998) Phrase-based information retrieval. *Information Processing and Management*, 4(6), 693

A. Smeaton, R. O'Donnell, and F. Kelledy.,(1994), Indexing structures derived from syntax in TREC-3: System description. In *TREC-3*.

B. Croft and D. Lewis.,(1987), An approach to natural language processing for document retrieval. In *SIGIR-1987*.

Boris Katz & Jimmy Lin., (2003), Selectively using relations to improve precision in Question Answering, in Proceedings of EACL 2003.

Boris Katz & Jimmy Lin.,(2004), Sapere: From Keywords to Key Relations. MIT Computer Science and Artificial Intelligence Laboratory.

C. Zhai, X. Tong, N. Milic-Frayling, and D. Evans. ,(1996), Evaluation of syntactic phrase indexing— CLARIT NLP track report. In *TREC-5*.

Donald Wardhaugh., (2003) , Understanding English Grammar – A Linguistic Approach. Second Edition. 7(4), Blackwell 123.

J. Fagan. , (1987), *Experiments in Automatic Phrase Indexing for Document Retrieval: A Comparisons of Syntactic and Non-Syntactic Methods*. Ph.D. thesis, Cornell University.

Jim Cowie, Yoric Wilks., (**2002**), Information Extraction. In Handbook of Natural Language Processing, Dale, Moisl & Somers (Eds.), 241.

L.Hirschman, R.Gaizaukas.,(**2001**), Natural Language Question Answering: The view from Here, Natural Language Engineering, 7(4).

Manish Joshi, Rajendra Akerkar. , **(2006) ,** Shallow parsing based algorithms to improve precision of Question Answering Systems, *NWAI*, SIGAI, C-DAC.

Manish Joshi, Rajendra Akerkar **(2005)** Algorithm to Effectively Handle Semantic Symmetry Question Answering Systems, *CSIT-2005, 5*-8080-631-7, 246.

Michael Kaisser, Tilman Becker, (**2004**) Question Answering by selecting Large Carpora with linguistic methods, In proceedings of the thirteenth TExt RErieval Conference (TREC-2004)

Rolf Schwitter, Diego Mollá Aliod, Michael Hess, (**1999**) ExtrAns - Answer extraction from Technical Documents by Minimal Logical Forms and Selective Highlighting (forthcoming). Presented at The Third International Tbilisi Symposium on Language, Logic and Computation.

Sivaji Bandyopadhya.,(**2000**) Cross Language Database System: Multilingual Human Computer Interface.

Soo-min Kim Dae-ho Baek, Sang-Beom Kim, Hae-Chang Rim, (**2000**) Question Answering considering Semantic categories and co-occurance density. In proceedings of the ninth TExt RErieval Conference (TREC-2000)

T. Strzalkowski, L. Guthrie, J. Karlgren, J. Leistensnider, F. Lin, J. Perez-Carballo, T. Straszheim, J. Wang, and J. Wilding. ,(1996), Natural language information retrieval: TREC-5 report. In *TREC-5*.

http://morphix-nlp.berlios.de/manual/node17.html (For QTAG POS tagger)

http://www.comp.lancs.ac.uk/ucrel/claws/trial.html (For trial of data using CLAWS POS Tagger)

http://trec.nist.gov/data/qa/2005_qadata/QA2005_testset.xml

Annexure- 1:

Result of algorithm for two differently phrased questions

Question - Who Killed Militants?

POS(Killed) – VBD (Past tense)

Candidate Answer	Sequence	POS	Is it an Exact Answer ?
National army soldiers killed 6 militants. POS(Killed)=VBD (Past Tense)	Same	Match	YES
Militant killed 13 bus passengers. POS(Killed)=VBD (Past Tense) Militants killed 4 policemen. POS(Killed)=VBD (Past Tense)	Not Same Not Same	Match Match	NO NO
BSF soldiers in an encounter killed 3 militants in Rajauri district. POS(Killed)=VBD (Past Tense)	Same	Match	YES
In bomb attack in Srinagar militant killed 7 peoples, 25 injured. POS(Killed)=VBD (Past Tense)	Not Same	Match	NO
3 militants were killed by Kashmir Police in combing operation. POS(Killed)= VBN (Past Part.)	Not Same	No Match	YES
5 innocent peoples were killed by militants. POS(Killed)= VBN (Past Part.)	Same	No Match	NO

Annexure- 2:

Application of rules to handle Ambiguous Modification problem.

Question : **What is the largest volcano in the Solar System?**

Adj : largest **S_N = Solar System** **I_N = volcano**

Candidate Answer	Sequence	Rule	Is it an Exact Answer ?
Candidate Answer 1: Mars boasts many extreme geographic features; for example, Olympus Mons, the **largest volcano** in the **solar system**.	**Adj** a **S_N** a = I_N	R_1-a	YES
Candidate Answer 2: The Galileo probe's mission to Jupiter, the **largest** planet in the **Solar system**, included amazing photographs of the **volcanoes** on Io, one of its four most famous moons.	**Adj** a **S_N** a= Otherwise	R_2	NO
Candidate Answer 3: Even the **largest volcanoes** found on Earth are puny in comparison to others found around our own cosmic backyard, the **Solar System**.	**Adj** a **S_N** a= Otherwise	R_2	NO
Candidate Answer 4: Olympus Mons, which spans an area the size of Arizona, is the **largest volcano** in the **Solar System.**	**Adj** a **S_N** a = I_N	R_1-a	YES
Candidate Answer 5: In the **Solar System**, Olympus Mons, which spans an area the size of Arizona, is the **largest volcano**.	**S_N a Adj ß I_N** ß = Blank	R_3	YES
Candidate Answer 6: In the **Solar System**, the **largest** planet Jupitor has more than 50 **volcano**.	**S_N a Adj ß I_N** ß= Otherwise	R_4	NO

Annexure-3:

General Architecture of ENLIGHT System