A Comparison of Ontology based and Keyword based Query Processing System for Urdu Language

Rukhsana Thaker[#1], Dr. Ajay Goel[*2]

[#1] *Lecturer, CSE Department , CoET, BGSB University*
Rajouri, J&K, India

[*2] *Head, CSE Department, Baddi University*
Baddi, H.P, India

***Abstract*— Query Processing is a field of Artificial Intelligence in which a user asks a question in natural language and the system replies with an answer in the same language. A lot of work has been done in English, Chinese and many other languages but very limited work is done in Urdu Language. Urdu is one of the widely spoken and written languages of south Asia. Due to unstructured format of Urdu language information retrieval of information is a big challenge. This paper compares two query processing systems, Ontology based and keyword based systems, to process a user query in natural language. Ontology based system uses ontology for knowledge building, rule generations and for defining relationships. Keyword based system, on the other hand, uses keywords for answering using query. This paper applies both the systems on the same data sets and analyses their performance on the basis of recall and precision value of the results**.

Keywords*— Ontology, Knowledge Representation, Query.

I. INTRODUCTION

Question-answering (QA) study emerged as an effort to deal this in sequence-excess problem. QA[1] systems are classified in two main parts: namely open domain QA system and closed domain QA system. Question which deals with nearly everything and can only relies on worldwide information, such type systems are called as open domain question answering system. On the other hand, closed-domain question answering deals with questions under a particular domain (music, weather forecasting etc.) The domain specific QA[2] system involve deep use of natural language processing systems formalized by building a domain specific ontology[3,4].

From the last decade Information is present not only in English language but also in many other languages. This is due to the advent of Unicode scheme that user can contribute their knowledge over the web in their own language. Various scientists have worked on QA with English [5], French, and Chinese[6] etc languages. But from the last few years scientists move their research into regional language like Punjabi, Tamil, and Hindi etc. Among variety of languages, the Urdu language is among the largest spoken languages of the world and state language of J&K state. Due to the availability of large amount of Urdu language documents over the web, it is required to develop a sophisticated information retrieval system to utilize the information efficiently.

In our previous work we have proposed a query processing system for Urdu language which is based on ontology. Ontology is used for knowledge representation of the domain. It is used to give meaning to the information of that particular domain. For our previous work we worked on Car ads in Urdu language. For that we have proposed an algorithm to answer user query in natural language. In this paper we are comparing two query processing system for Urdu language. One is Ontology based and another is keyword based.

The outline of the paper is as follows. The paper is organized as follows In Section II we discussed related work of this object using Ontology Section III discussed the same using Keyword search. Section IV presents Comparisons using Experiments and Section VI presents the conclusion.

II. QUERY PROCESSING USING ONTOLOGY

This section presents the architecture and functionality of Ontology based system proposed in our previous work [8]. The working of the proposed model is as follows. In this system user can answer in his own natural answer and system will return the answer in the same language. The architecture proposed for System is as revealed in Figure 1.

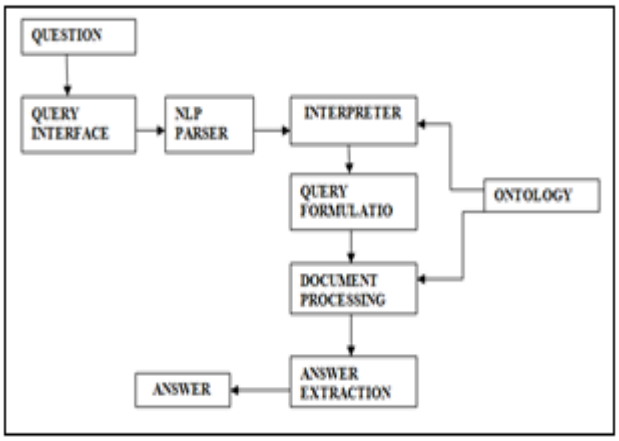

Figure 1: Architecture of Ontology based QA System.

Following steps are followed for processing the query.

Step 1: The first step is to select the domain for processing which will act as dataset for our system

Step 2: Ontology is created for that particular domain.

Step 3: User enters the query in natural language in the interface provided by the system

Step 4: Query is processed using the ontology to find the query elements like what information is needed, whose

information is to be processed and what the conditions of processing are.
Step 5: Dataset is searched for answer using the Ontology and the query elements.
Step 6: Finally answers are framed and displayed to the user.

Ontology based system give more accurate answers and number of answers are also increased.

III. QUERY PROCESSING USING KEYWORDS

In Keyword based system user query is processed based on the keywords present in the user query. It searches for the answers word by word by scanning all the dataset.
Following steps are followed for processing the system.
Step 1: User enters the question in the interface provided in his natural language.
Step 2: Query is processed to find the keywords present in the question.
Step 3: System searches the dataset for the keywords filtered out from the query.
Step 4: if the keywords are not present in the dataset, system will simply ignore those keywords.
Step 5: if the keywords are present, system will frame the answer using those keywords and finally displayed them to users.

IV. COMPARISONS USING EXPERIMENTS

In this Section results of all the implementation steps of the proposed methods are highlighted. Implementation is done in visual studio 2010 using C#. Experimental results of two methods are compared. To conduct the experiment a data set of 200 car ads is created. The evaluation is done on the basis of recall and precision value of properties: Year, Price, Color, Mileage, and Engine. The values are computed as

Recall= C/N
Precision = C/C+I

Where N is the number of ads in which property value is found, C is the number of correctly extracted values and I is the number of incorrectly values.
The Evaluation of each property is done one by one using proposed methodology and comparison is done with Key word based search. For each property fifteen queries are evaluated and its recall and precision are calculated and compared with keyword based search.

A. YEAR PROPERTY

Sample Query: " ہنڈا سٹی کار کی کلر بتاؤ جسکا ماڈل 2005 اور 2010 کے بیچ کا ہو"

Ontology Answers:

ہنڈا سٹی کار ماڈل 2006 کلر وائٹ
ہنڈا سٹی کار ماڈل 2010/ کلر وائٹ
ہنڈا سٹی کار ماڈل 2008 رنگ لال
ہنڈا سٹی کار ماڈل 2009 کلر سفید
ہنڈا سٹی کار ایئرآفم ٹوفکچرنگ 2010 کلر سفید
ہنڈا سٹی کار ء 2007 کلر وائٹ
ہنڈا سٹی کار ایئرآفم ٹوفکچرنگ 2005- کلر وائٹ

Keyword Answers:

ہنڈا سٹی کار ماڈل 2006 کلر وائٹ
ہنڈا سٹی کار ماڈل 2010/ کلر وائٹ
ہنڈا سٹی کار ماڈل 2008
ہنڈا سٹی کار ماڈل 2009 کلر سفید
ہنڈا سٹی کار ماڈل 2006

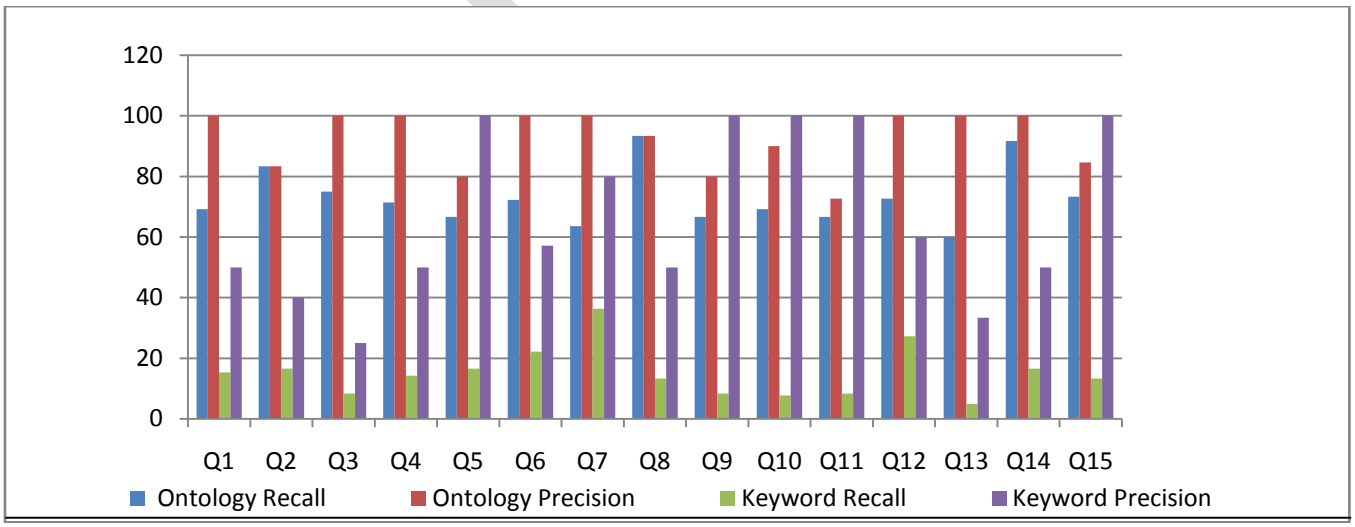

Figure 2: Recall and precision for Year Property

B. PRICE PROPERTY

Sample Query: "SUZUKI KHYBER کار کا کلر، ماڈل بتاؤ جسکی قیمت 5لاکھ سے کم ہے"

Ontology Answers:

سوزوکی خیبر کار ڈیھانڈ 425000 ماڈل 1999 کلر گرے

سوزوکی خیبر کار ڈھانڈ 455000/- ماڈل 2014 کلر گرے
سوزوکی خیبر کار قیمت 425000 ماڈل 2006 رنگ سفید
سوزوکی خیبر کار قیمت 350000 ء 2008 کلر گرے
خیبر کار قیمت 400000/- ماڈل 2008 کلر بلیک

SUZUKI KHYBER کار قیمت
4لاکھ ایئرآفم ٹوفیکچرنگ 2010 کلرسفید
SUZUKI KHYBER کار قیمت 3لاکھ ء 2007 کلر وائٹ
سوزوکی خیبر کار پرائس 3,50,000/- ء 1998 رنگ گرے

SUZUKI KHYBER کار قیمت 3لاکھ ء 07 کلر وائٹ
Keyword Answers:

SUZUKI KHYBER 2010 کار قیمت 4لاکھ ایئرآفم ٹوفیکچرنگ کلرسفید
SUZUKI KHYBER کار قیمت 3لاکھ ء 2007 کلر وائٹ
SUZUKI KHYBER کار قیمت 3لاکھ ء 07 کلر وائٹ

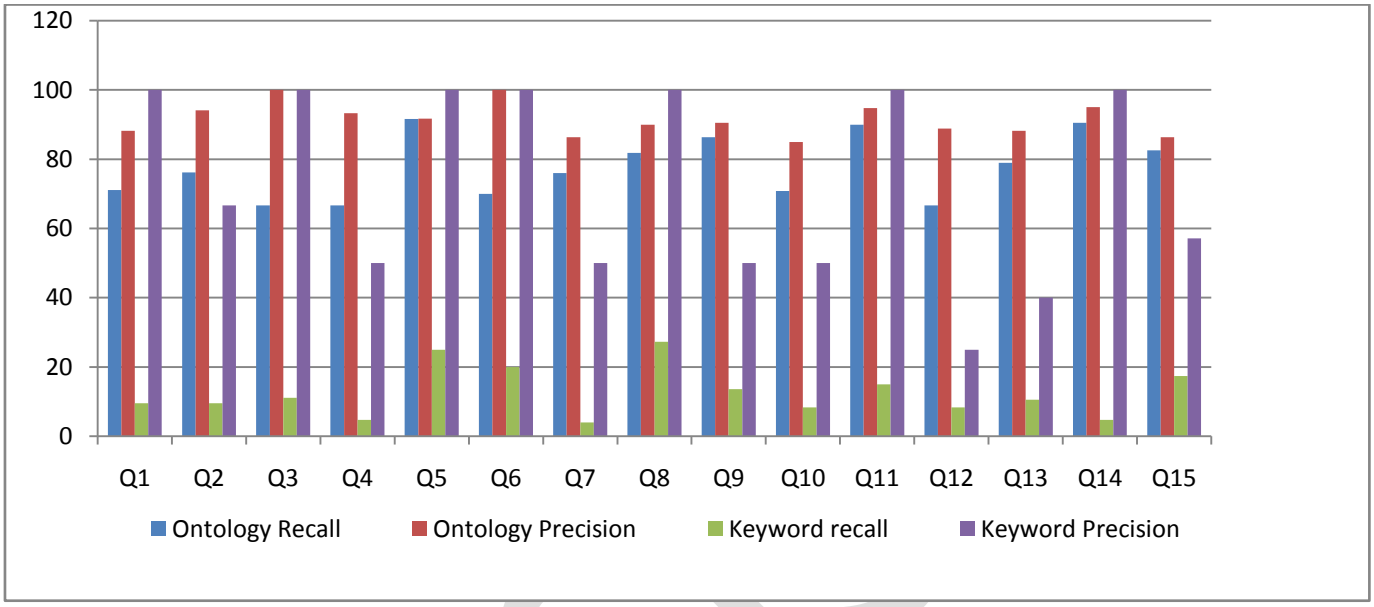

Figure 3: Recall and Precision for Price Property

C. COLOR PROPERTY

Sample Query: " سوزوکی خیبر کار کی ، پرائس ، بتاؤ جسکا رنگ گرے ہو "
Ontology Answers:

سوزوکی خیبر کار کلر گرے ڈھانڈ 425000

455000 سوزوکی خیبر کار کلر گرے ڈھانڈ/-
350000 سوزوکی خیبر کار کلر گرے قیمت
3,50,000 سوزوکی خیبر کار رنگ گرے پرائس

Keyword Answers:

3,50,000 سوزوکی خیبر کار رنگ گرے پرائس

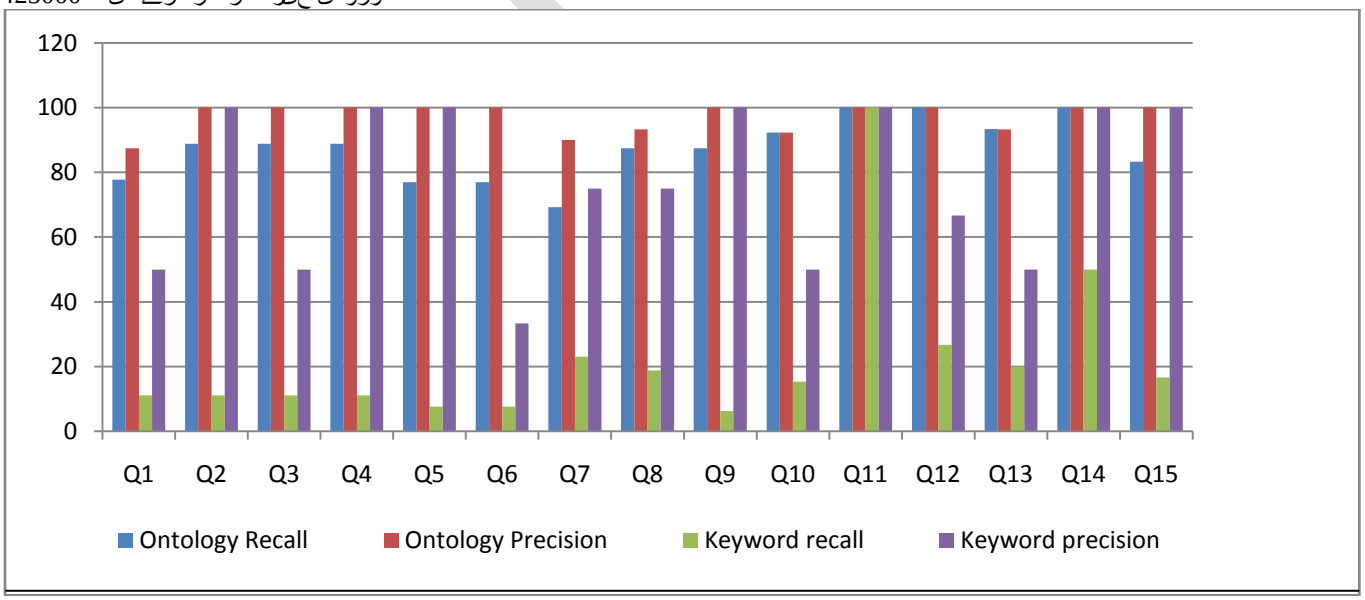

Figure 4: Recall and Precision for Color Property

D. MILEAGE PROPERTY

Sample Query: " DAIHATSU CUORE کار کا کلر ، ماڈل 50000 بتاؤ جسکا مائلیج km سے کم ہے "

Ontology Answers:

ڈائہاٹسو کورے کار km 40000 کلر وائٹ 2010/ ماڈل
ڈائہاٹسو کورے کار km 40000 کلر 10/ ماڈل white

green انٹرآفم ٹوفیچرنگ 2011 کلر km 40000 کورے کار
DAIHATSU CUORE کار km 40000 کلر وائ 2010/ ماڈل

Keyword Answers:

DAIHATSU CUORE کار km 40000 کلر وائٹ 2010/ ماڈل

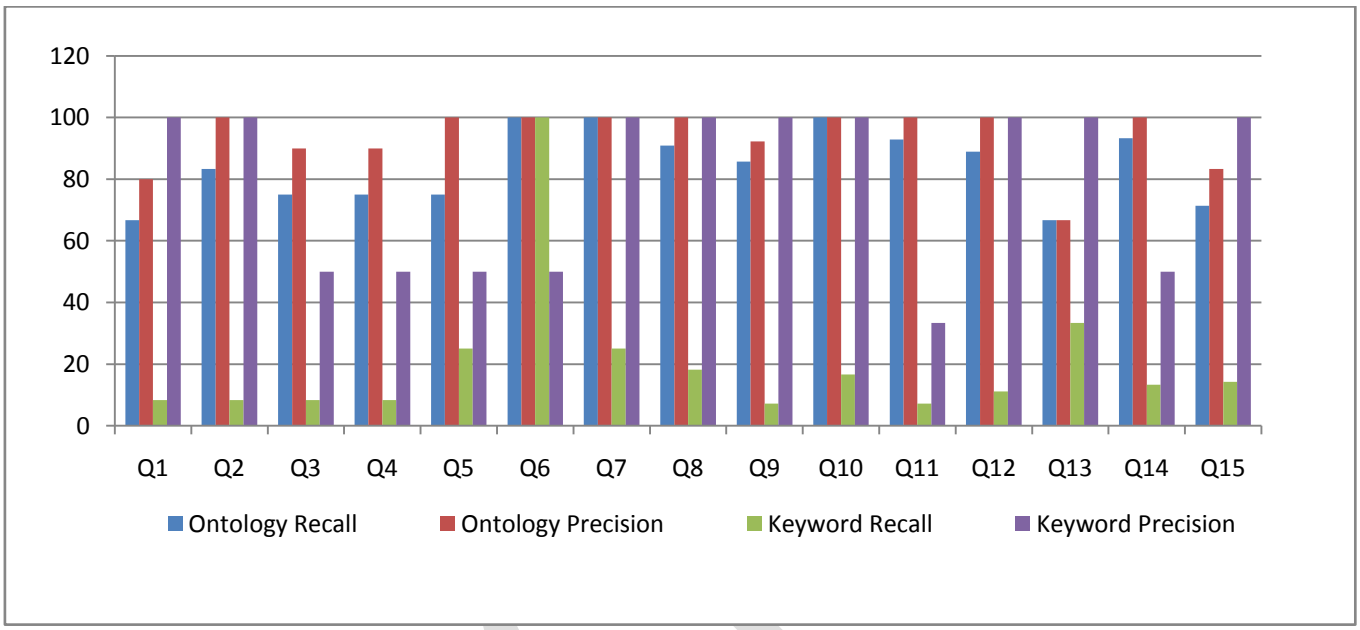

Figure 5: Recall and Precision for Mileage Property

E. ENGINE PROPERTY

Sample Query: "انجن ہنڈا سٹی کار کا کلر، ماڈل بتاؤ جسکی 900cc سے کم ہو ".

Ontology Answers:

ماڈل 2004 رنگ سفید cc 850 ہنڈا سٹی کار
ماڈل 2004 رنگ سفید cc 850 ہنڈا سٹی کار
ماڈل 2013 کلر بلیک cc 850 ہنڈا سٹی کار

ماڈل 2006 رنگ سفید cc 800 ہنڈا سٹی کار

Keyword Answers:

ماڈل 2004 cc 850 ہنڈا سٹی کار
ماڈل 2004 cc 850 ہنڈا سٹی کار
ماڈل 2013 کلر بلیک cc 850 ہنڈا سٹی کار
ماڈل 2006 cc 800 ہنڈا سٹی کار

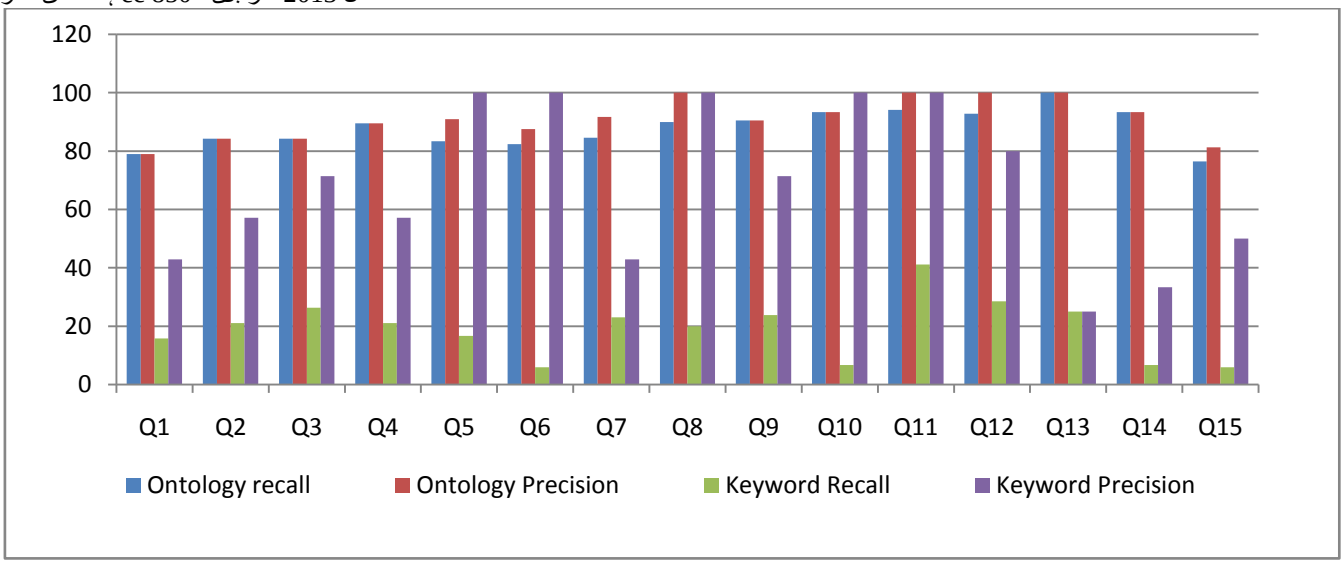

Figure 6: Recall and Precision for Engine Property

After analyzing the recall and precision values for all the properties we came to the conclusion that ontology based system gives better results as compare to the Keyword based system. Numbers of answers in case of ontology based system are much higher as compared to keyword based system. Recall and Precision value of Ontology based system is also higher as compared to keyword based system.

V. CONCLUSION

In this paper comparison is done between keyword based search and keyword based search. Recall and precision values for a set of queries are calculated and comparison is done for recall and precision values of both ontology based system and keyword based system. The result of these comparison show that the ontology based system give better result as compared to keyword based search. The future work will focus on further improvement in the process of semantic annotation to handle complexity and challenges of Urdu language. Further it can be extended by adding other domain ontologies where user can ask query from any domain and the system will answer the question from that particular domain. So before answering the query the system will select the domain for that query from the set of available domains and then answer the question according to the ontology. Moreover present methodology creates ontology manually. Future work intends to create ontology automatically from the user query.

REFERENCES

[1] R. Baeza-Yates & B. Ribeiro-Neto, "Modern Information Retrieval", 2011 second edition.

[2] Breck, E., Burger, J., House, D., Light, M. & Mani, I. ``Question answering from large document collections'', Question Answering Systems: Papers from the 1999 AAAI Fall Symposium, 5-7 November, North Falmouth, MA, AAAI Press, Menlo Park, CA, pp. 26-31.

[3] Muthu krishnan Ramprasath1 and Shanmuga sundaram Hariharan2, 2012, "A Survey on Question Answering System", International Journal of Research and Reviews in Information Sciences (IJRRIS), pp171-179.

[4] Liddy, NY. Marcel Decker "Natural Language Processing" , in Encyclopedia of Library and Information Science in 2001

[5] ROBERT F. SIMMONS, "Natural Language Question Answering Systems: 1969"

[6] Cai Dongfeng, Cui Huan, Miao Xuelei, Zhao, Ren Xiangshi 2004, "A Web-based Chinese Automatic Question Answering System", pp1141-1146, IEEE.

[7] Sallaberry Christian, Etcheverry Patrick, Marquesuzaà Christophe, 2006, "Information Retrieval and Visualization Based on Documents" Geospatial Semantics", International Conference on Information Technology: Research and Education, pp277- 281

[8] Rukhsana Thaker, Dr.Ajay Goel, "QUrdPro: Query processing system for Urdu Language", International Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 5, Issue 6, (Part - 3) June 2015, pp.106-108