

# KNOWLEDGE UNDERSTANDING AND ADVANCED SEARCHING

R. Vidya<sup>1</sup>, D. Vigneshwara Raj<sup>2</sup> and K. Sujatha<sup>3</sup>

<sup>1</sup>Department of Computer Science, Thiruvalluvar University, India

<sup>2</sup>Vikiinformatic School of Computing and Soft Skills, India

<sup>3</sup>Department of Computer Science, Government College of Engineering, Tirunelveli, India

## Abstract

*It's a comprehensive fact that millions of people around the world surf the Internet for want of answers for their questions. Generally, the questions are asked in the form of Searching or direct questions which follow perfect ontological directions. It is important that the system understands the questions in the right sense and can provide the best answer for all the questions raised in the web forum. One such pragmatic method is required which is expected to provide optimum solution to achieve best answers for questions that not only perceps language but also follows perfect ontological information in accordance with the cyber law. This Proposed Model presents a new dynamic model called Knowledge Understanding and Advance Searching (KUAS) that studies the importance of Smart Question Answering with other question answering engines like START and proves to give the optimal solution compared to them.*

## Keywords:

*Knowledge Understanding, Advanced Searching, Ontological Information, Optimal Searching Technique*

## 1. INTRODUCTION

“There are many data in this world which exists not based on real knowledge; but based on the perception and understanding of Perceptor Knowledge”. Knowledge is the most important fact that incorporates best ideas for the present and the future. Knowledge Management [1] is the process of capturing, analyzing and sharing of knowledge to have an effective use of knowledge. To manage the knowledge effectively, it must be used with best possible output when it is required. Hence Knowledge Understanding is much required as it helps the preceptor to understand, analyze and give best solution to problems. Generally searching is the way millions of people use to ask questions and get their answers. Ontology [2] is a specification of a conceptualization that is a description of the concepts and relationships that can exist for an agent or a community of agents. The Research Paper focuses on the new and optimal pragmatic technique known as KUAS (Knowledge Understanding and Advanced Searching) as a solution to the long spent problem on Natural Language Question answering to achieve the best answers for the right questions and compared its existing question answering systems like START and WEBSCOOP to prevail its optimistic existence in the ontological web world.

## 2. REVIEW OF LITERATURE

The Technique of searching is in existing for more than a century. Many proposed methods and algorithms are available for the benefit of the internet users to search for ideas in Internet. It's

a prominent fact that almost everyone uses internet in Computers, mobiles for searching information. Hence all of them will be questioning some useful information and it's the responsibility of the search engines to provide accurate results for the questions. The new technologies need a method that could be designed as a combination of knowledge representation [3], knowledge acquisition, deductive inference, Natural Language Processing and Artificial Intelligence. There are various ways of research involved in this discipline so that a best solution occurs with respect to question-answering systems. In general, in the current context of the Natural Language Processing, factual questions can be answered very quickly and easily whereas explanative answers are very hard to find out. Hence a pragmatic system is required to find solutions to all type of questions in the global ontological system. It's not compulsory that all questions must be answered right. There may be right as well as wrong answers. In general, the Question Answer Recall [3] is termed as the “ratio of number of correct answers to the total number of questions given”.

Shanshan Zhao [4] et al. developed a new model for question answering system for solving single-option geography questions. It uses a large scale Chinese dataset containing entrance examination questions set in geography that is observed from internet. The Results showed a high accuracy around 73% that could be further refined using a new model.

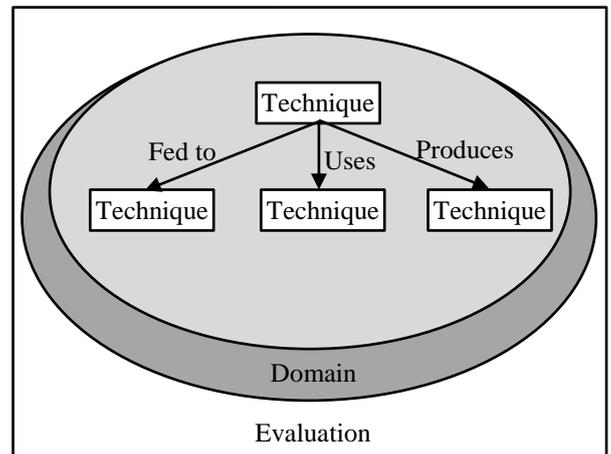


Fig.1. Approaches of Question Answering Technique

Sean Gallagher [5] et al. conducted a test to measure the contribution of searching in internet to answer jeopardy question answering. The research compared the watsonism architecture with other methods using recall with and without internet search data from Microsoft's Bing. The final conclusion revealed that in question answering, the network enabled system could alter all other systems to provide better results. It is important to have a decent look at the need for question answering systems that are

available at present so that the new system fits into it. There are various types of question methods available in this world.

In general, the domain holds the answering system with proper algorithms and techniques. The questions will be fed into the system which uses appropriate algorithm based on the types of questions asked by the user. The questions are analysed and evaluated with the global ontological information from the cyberlaw and then it produces the resultant answer.

## 2.1 DIMENSIONS OF QUESTION ANSWERING

Various types of Questions can be formulated as follows:

- **Factual** - It is based on facts. Hence the answer is straight-forward and no second choice.
- **Simple Reasoning** - It is simple logic or incident oriented
- **Synthesis-based** - It is analysis oriented to study the facts
- **Dialogue-based** - It is based on conversational analysis and decision making
- **Speculative** - It is based on discussions and decision making

## 2.2 AI WITH NLP IN KNOWLEDGE UNDERSTANDING

It's a general customary aspect that artificial intelligence plays a vital role to make a technology to understand information much easier than normal humans. The linguistic ability of human beings can be matched with technologies with the help of Natural Language Understanding and hence it's the best solution for creating a competent solution framework for better question answering systems.

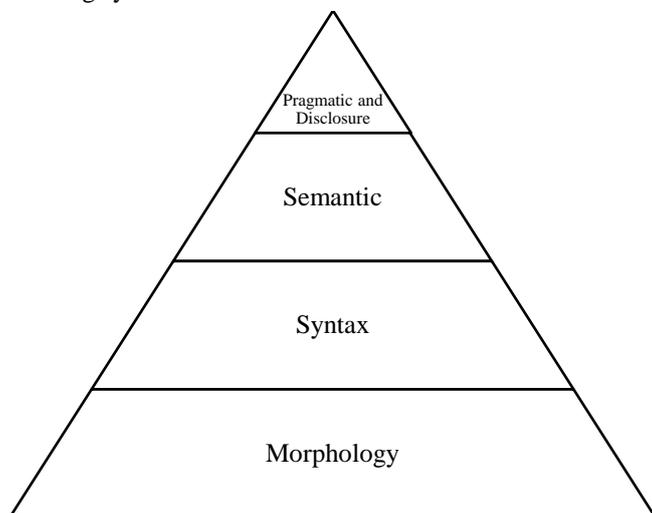


Fig.2. Stages of Analysis in NLP

### 2.2.1 Morphological Analysis:

Encompasses the knowledge gained through individual word analysis and non-word tokens such as punctuation are separated from words. E.g. Part of speech tagging (noun, adjective, verbs, articles etc.)

### 2.2.2 Syntactic Analysis:

Formulates the study of elements within sentences and converting it into a structure that defines the units represented by that list called sentence parsing.

### 2.2.3 Semantic Analysis:

Creates meanings defined by syntactic analyzers to the output structures.

Search Engines plays a vital role in normal search methods employed by millions of people around the globe. There are many Question Answering systems also that counteracts the needs of the people or researchers who are in need of some useful information over net. START and WEBCOOP are two such best question answering systems. START provides users with "just the right information," a system that provides intelligent cooperative responses to Web queries. Besides a heavy use of language (processing queries, generating responses, extracting knowledge from web pages), such a system requires the integration of knowledge representation and the use of advanced reasoning procedures. However, there are few restrictions in generating appropriate results that matches the question raised by the user. They are,

- There is a lack of innovative architecture that could integrate natural language understanding and advanced searching methods to solve the problems related to proper question Answering system.
- Sufficient semantic analysis has not been made to be used for effective answering in all domains at the same time
- Limited to rule-based learning rather than reasoning techniques.

## 3. PROPOSED METHOD

In Master's Research, the best solution to question answering system over web is very hard to be identified. It's quite optimistic and non-deterministic nature too. Hence it is very important to set scope for setting dimensions to our problem. The following problems are present in the existing system

1. Low-level knowledge understanding
2. Minimalist choice of representation,
3. Formalism and conventional reasoning without advanced features will pose difficulty in scaling across domains and to real life natural language text.

Based on the problems and scope defined, three aspects were discussed:

1. Full-disclosure Knowledge Understanding
2. Representing Authoritative and sensitive presentation like semantic network.
3. Reasoning that is network based which supports advanced reasoning techniques.

The following hypothesis constraints were proposed,

1. The problem must adopt a solution that adopts itself to any questions from any source to give best answering system considering the crucial aspects of discourse analysis [6].
2. The Result must have an Authoritative and sensitive method of representing semantic network, data provided by sensitive formalism like NLP that included all intrinsic information and objects provided by Cyber Law [7].
3. Network based ontological perspectives, representations and rule-based speculations can be integrated to a major

domain and implemented for a best question answering system.

After formulating the hypothesis, the proposed design was constructed on the basis of finding a most effective method for formulating and identifying a best solution for Question Answering technique in Knowledge Understanding.

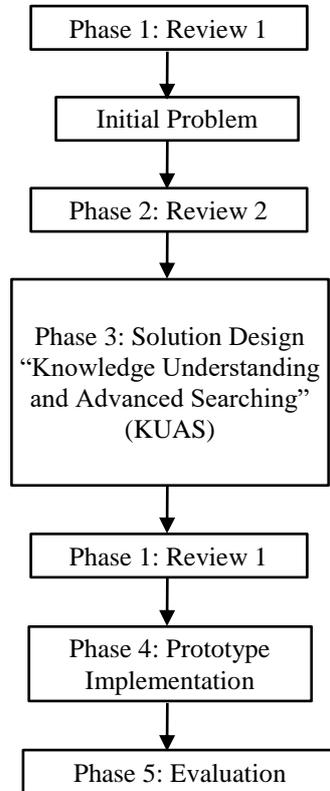


Fig.3. KUAS Methodology

It is proposed to have 5 Phases as indicated in the diagram. In the initial phase, a question is identified and taken for query analyser for identification of pattern of question. Then in phase 2, a review is made with corrections from review 1 and then moved to KUAS system for answer identification. Hence due to two reviews, the system will have more understanding about the question so that it can answer exactly to the point of the question in KUAS system. Then based on available solutions a Prototype will be build and hence a system can be implemented in phase 5 based on the success of the prototype.

#### 4. ARCHITECTURE OF KUAS

KUAS (Knowledge Understanding and Advanced Searching) is a proposed Prototype model designed and to be implemented based on the most complex algorithms and data structures that can encompass all best techniques that will enable the best solution for answering any complex questions over networks and semantics of the cyber law ontology. As a collective system, the application architecture interacts with environment and forms references based on Cyber Law to understand and respond to question raised from the user in a prompt and specific manner.

In general, the storage structures are segregated as three major groups as news repository, knowledge base that contains ontology along with semantic network and gazetteer.

#### 4.1 KNOWLEDGE SEARCHING SUBSYSTEM

The functionalities of KUAS are nicely packed into two major subsystems namely Natural Language Understanding and Network-Based Advanced Searching. This sub system is responsible to read and understand two things:

1. Questions from users and
2. Sentences of processed Cyber Law news articles from news repository

The System will propose a subsystem that will convert the representations in English questions or news contents into network representations. The System activities is carried out under four major phases as follows:

- Sentence parsing,
- Named-entity recognition,
- Relation inference and
- Discourse integration.

The input is sentence or word based as given by the user who needs some useful information from internet. The major output of this subsystem will be sentence based or paragraph based components that can be the answer for the query raised by the user. The Sentence parsing module accepts a simple English sentence and outputs the grammatical groups and relations of the sentence. Named-entity recognition module implements both the noun phrase chunking and category assignment algorithm to get an optimal solution. The Relation inference will enable the user identify the difference between the systems and also the relationship between them. Discourse integration will combine the results and integrate the components with knowledge based system. The Gazetteer is a person who is in need of the system. When the user uses the keyword to search for a word, the named-entity recognizer will read the content and parsed the sentence and the relationship between the grammatical sequences will be analyzed using relation influence. The relationship will be integrated and combined in the discourse integration machine.

#### 4.2 NETWORK BASED ADVANCED REASONING SYSTEM

The Network-based advanced reasoning subsystem implements ideas in the reasoning mechanisms of the solution framework and is responsible to discover a valid answer and generate unambiguous response or generate an explanation for user's questions.

The entire system is presented as five phases with five modules like

- Network-to-path reduction,
- Selective path,
- Relaxation of event constraint,
- Explanation on failure and
- Template based response generation.

Network-to-path reduction module enables the user to get into a network and find a global solution that matches with the query of the user. Hence the output implements global knowledge and identifies best and optimal solution from various sources and maps it to the constraints of the cyber law. To give a most

extraordinary response, various ontological responses and constraints were also mapped along with the sequences. If any question could not be matched with proper responses over network, a failure state is proclaimed and during that stage, an

explanation containing the information of question and the reason for failure will be indicated to the user. Sometimes the constraints may be relaxed and a new solution can be identified.

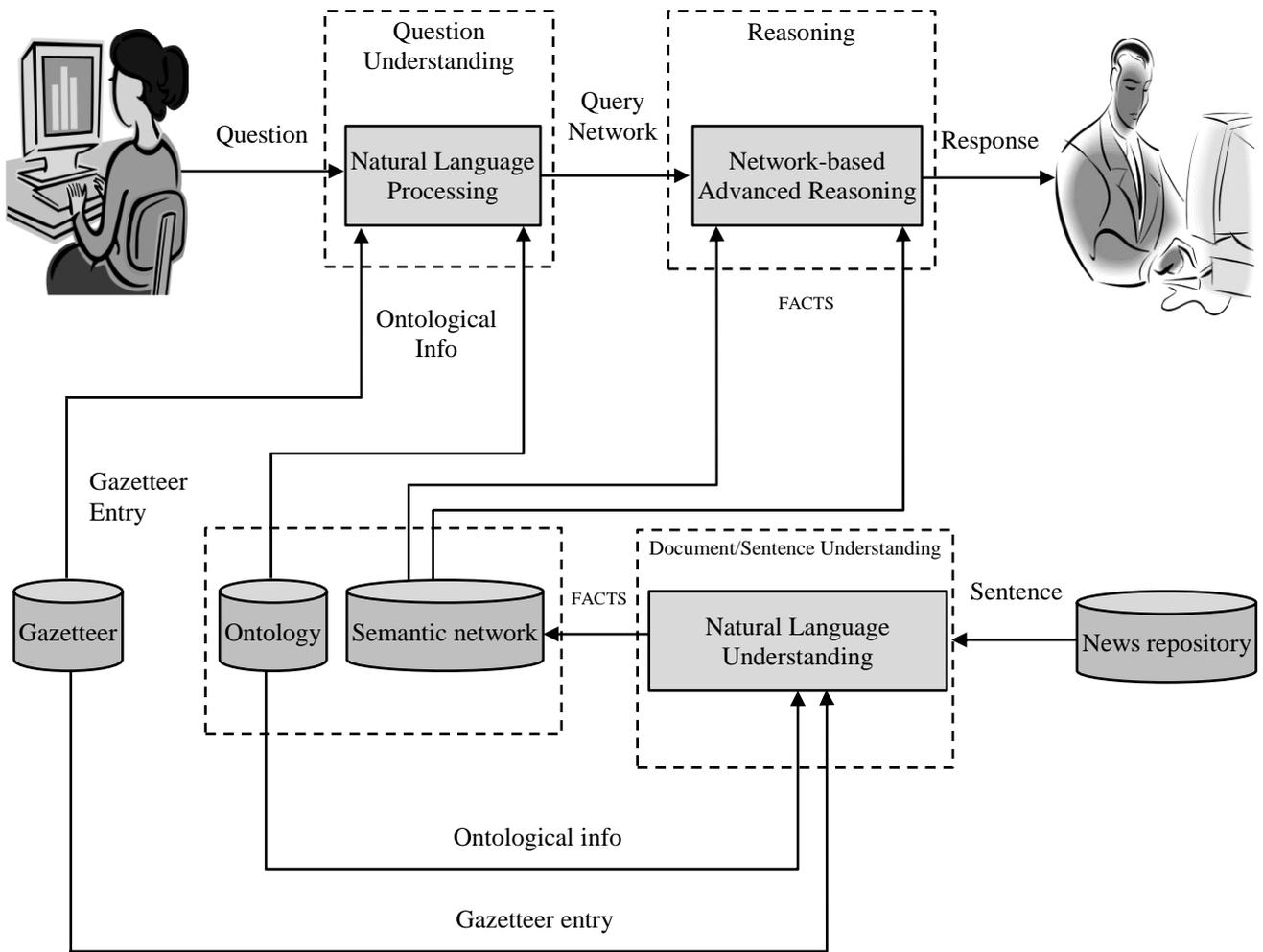


Fig.4. General Architecture of KUAS system

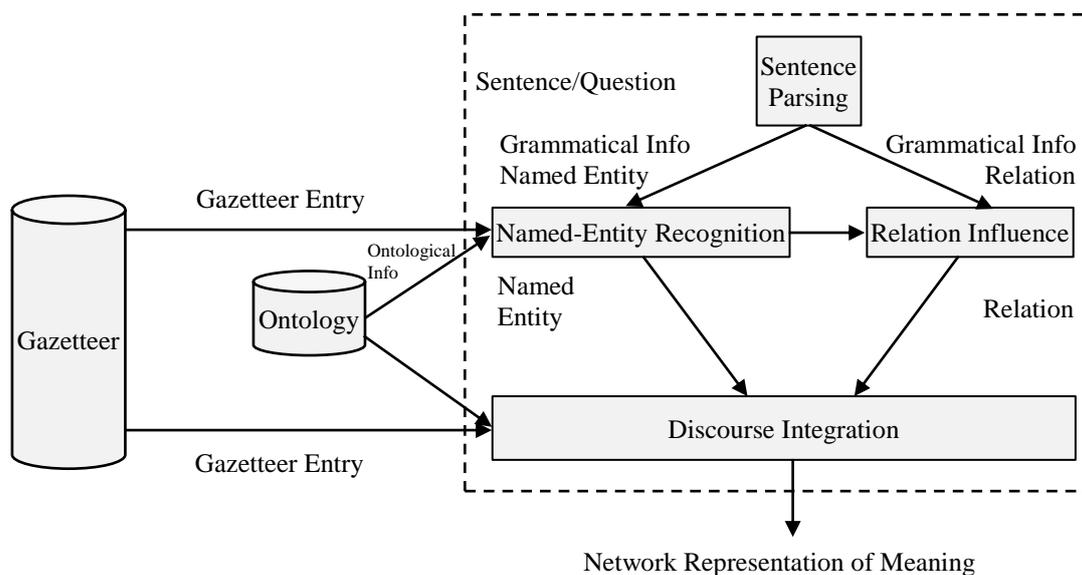


Fig.5. Architecture of KUAS subsystem based on Natural Language Processing

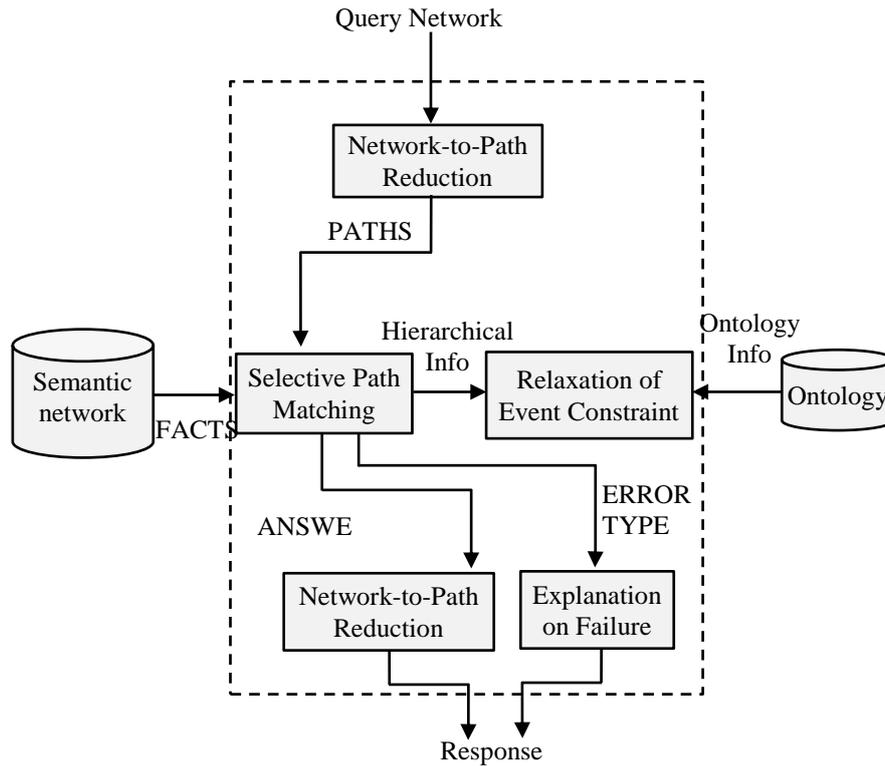


Fig.6. Architecture of KUAS Sub-system - Network based advanced searching system

**5. PROTOTYPE IMPLEMENTATION**

The Software and web implementation is done with PERL and MySQL. PERL is more suitable for Network based implementations and hence it's the best choice for creating the software. Ontology consists of two types of entries that collectively define the ontology of KUAS namely class entry and attribute entry. The entire class entries collection defines the hierarchy of the ontology. The Prototype is only the entry level system and hence it requires further technological promotions in the future.

**5.1 EVALUATION CRITERIA**

The evaluation criteria namely response time and response quality are used for evaluation in this research. These two criteria attempt to look at the practicality issue of question answering approaches from two dimensions namely time and quality of answer. Response time concerns with how much time is required to provide an answer or response to a question.

Table.1. Comparative Analysis of START and KAUS

System	Max time	Min Time	Avg. Time	Standard Deviation
START	19.3377	2.4234	3.2798	6.2859
KUAS	18.3106	2.3921	3.0957	4.5986

Response Quality ensures that the question answering approach generates accurate responses that directly answer the

question and not merely passages or snippets extracted from the full document.

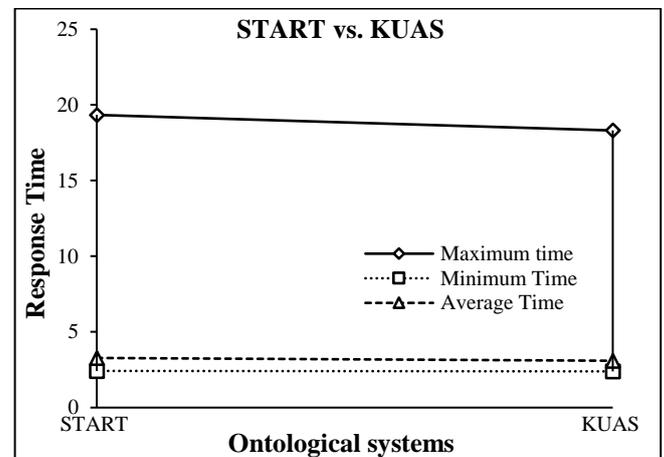


Fig.7. Performance Level Comparison of START with KAUS

Table.2. Result of KAUS before analysis

Category	? <sub>1</sub>	? <sub>2</sub>	...	? <sub>k</sub>	
	? <sub>x1</sub>	? <sub>x2</sub>	? <sub>y2</sub>	? <sub>xk</sub>	? <sub>yk</sub>
BQ_?					
LQ_?					
Oj_?					
Total					

where,

**BQ\_N**: responses that provide accurate and direct answers to questions

**LQ\_N**: not relevant

**O1\_N**: informative responses in the event of no answers

**O2\_N**: responses that suggests possible spelling mistakes.

Table.3. Comparative Results of START and KAUS

Category	START Vs KUAS	
	START	KUAS
BQ_	0	0
LQ_	0	0
O_1	0	1
O_2	0	0
<b>Total</b>	<b>0</b>	<b>1</b>

## 6. CONCLUSION

The Natural Language Processing has always been the best solution for linguistic and semantic problems raised by web and ontological surfers over net. Hence this pragmatic method Knowledge Understanding and Advanced Searching (KUAS) will sure enhance and provide a better solution to all question answering systems that are meant to dominate and promote better standards to searching and acquirement of knowledge in the future.

## REFERENCES

- [1] Thomas H. Davenport, "Saving IT's Soul: Human Centered Information Management", *Harvard Business Review*, Vol. 72, No. 2, pp. 119-131, 1994.
- [2] Tom Gruber, "What is an Ontology?", Stanford University, Retrieved 2009-11-09. Available at: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [3] Cody Kwok, Oren Etzioni and Daniel S. Weld, "Scaling Question Answering to the Web", *ACM Transactions on Information Systems*, Vol. 19, No. 3, pp. 242-262, 2001.
- [4] Shanshan Zhao, Yuqing Zheng, Conghui Zhu, Tiejun Zhao and Sheng Li, "Semantic Computation in Geography Question Answering", *Proceedings of 12<sup>th</sup> International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pp. 1572-1576, 2016.
- [5] Sean Gallagher and Wlodek Zadrozny, "Leveraging Large Corpora using Internet Search for Question Answering", *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 532-535, 2016.
- [6] P.M. Athira, M. Sreeja and P.C. Reghuraj, "Architecture of an Ontology-based Domain-Specific Natural Language Question Answering System", *International Journal of Web and Semantic Technology*, Vol. 4, No. 4, pp. 31-39, 2013.
- [7] Nguyen Tuan Dang, Do Thi and Thanh Tuyen, "Natural Language Question Answering Model Applied to Document Retrieval System", *World Academy of Science, Engineering and Technology*, Vol. 51, pp. 36-39, 2009.
- [8] Abhijit Kumar and Lavit Rawtani, "Question Answering System using Artificial Intelligence and Fuzzy System", *MES Journal of Technology and Management*, pp. 12-15, 2011.
- [9] David Gunning, Vinay K. Chaudhri and Chris Welty, "Introduction to the Special Issue on Question Answering", *Association for the Advancement of Artificial Intelligence*, pp. 11-12, 2010.
- [10] L. Hirschman and R. Gaizauskas, "Natural Language Question Answering: The View From Here", *Natural Language Engineering*, Vol. 7, No. 4, pp. 275-300, 2001.
- [11] Zhengyou Zhang, "Autonomous Mental Development: A New Interdisciplinary Transactions for Natural and Artificial Intelligence", *IEEE Transactions on Autonomous Mental Development*, Vol. 1, No. 1, pp. 1-11, 2009.
- [12] R. Vidya and G.M. Nasira, "Exploring Clinical Reasoning in Novices: Knowledge Sharing System between Social Media and Medical Professionalism", *Data Mining and Knowledge Engineering*, Vol. 6, No. 9, pp. 349-353, 2014.