ISSN 2047-3338

A Strategy for Question Interpretation in Question Answering System

Syarilla Iryani A. Saany, Ali Mamat, Aida Mustapha and Lilly Suriani Affendey

***Abstract*—A Question answering (QA) system aims to pull exact and precise answer for Natural Language (NL) question. To extract answer from the text corpus or knowledge base, the QA system has to understand exactly what the question is. Ambiguity may arise when NL question contains modifier term which needs to compare and evaluate its semantic dimension. Therefore, a strategy for an evaluation metrics of the associated modifier term is proposed in this paper. The identified metrics may make possible to regulate using user modeling (UM) and relevance feedback (RF) mechanisms. The performance of the proposed model is evaluated by using the standard information retrieval measurement on the Geoquery datasets.**

***Index Terms*—Question Answering System, Question Analysis, Ontology, Relevance Feedback and User Modeling**

I. INTRODUCTION

QUESTION answering (QA) system is a discipline that has been extensively researched and largely driven by TREC (Text REtrieval Conference-http://trec.nist.gov/) QA track. QA system is within the field of Information Retrieval (IR) and Natural Language Processing (NLP). In IR, the objective of an IR system is to search the elements in the resource that mapped with user's specified need [11]. NLP, on the other hand, creates an easy and friendly environment since it does not require any programming language skills to access the data where users may input natural language (NL) to interact with the system.

In consequence, QA system is an automated tool that can search and retrieve information from a textual document repository or knowledge base (KB). The goal of an effective QA system is to pull exact and precise answer for NL question. QA is a type of information retrieval which requires more complex NLP techniques than other types of information retrieval such as document retrieval [4], [8]. The significance difference lies in the output of IR and QA systems. IR system outputs a list of ranked documents potentially relevant to the input user's query. The user then has to scan the documents in order to find the pertinent information. On the contrary, QA system takes the NL question as an input and the answer of the question is produced in the form of a text fragment or exact answer extracted from the electronic document sources.

The QA systems must provide facilities for analyzing question, be able to quickly and efficiently search for documents or KB relevant to the question, locate the scope of answers and choose the best among them. The process of question analyzing has always been a focus in the research of QA area. The process starts with analyzing and processing the question posted in natural language by a user.

To extract answer from the text corpus or knowledge base, the QA system has to understand exactly what the question is. The question processing component may exploit the morphological analysis, structural analysis, syntactic analysis and semantic analysis. The question is also classified based on the types (i.e., date, location or name of a person). Query is formulated based on the question analysis techniques and the class of the question. If NL question cannot be disambiguated during the process of transforming into query, the user will be asked to clarify his or her question. Ambiguity may arise when NL question contains modifier term which needs to compare and evaluate its semantic dimension. Therefore, a strategy for an evaluation metrics of the associated modifier term is proposed in this paper. The identified metrics may make possible to regulate based on user modeling (UM) and relevance feedback (RF) mechanism.

UM involves the process of developing, retaining and maintaining the user profiles of the systems [12], [13]. The users will be classified, and then the search engine inferred about individual users on the basis of that classification. RF is either applied after the system has produced the results (answers) based on the NL question submitted by the user or is exploited to interpret the questions [5], [6].

In the previous QA systems, several issues are pinpointed:

- Inadequate representations of the user's question. E.g. "How big is Alaska", here *how big* is depending on to the question context or the structure of the KB.
- Free form question format. E.g. the queries "What is the capital of Texas?" and "Can you tell me the capital of Texas?" both will yield similar answer.

Syarilla Iryani A. Saany is with the Faculty of Informatics, Universititi Sultan Zainal Abidin, 21300 Kuala Terengganu, Terengganu, Malaysia (Email: syarilla@ unisza.edu.my).

Ali Mamat, Aida Mustapha and Lilly Suriani Affendey are with the Faculty of Computer Science and Information Technolgy, Universiti Putra Malaysia, Serdang, Selangor, Malaysia (Email: {ali, aida, suriani}@fsktm.upm.edu.my).

- Ambiguous relevance answers in defining proper clarification. E.g. both statements "California State has 45 million people live in there" and "14% from 324 million people in USA live in California" provide the same information but are defined in different forms.
- No user profile such as user specific information, experiences, common goals and requirement behaviors.
- Little knowledge on the contents and structures of the knowledge base.

To overcome the shortcomings of previous QA systems, a strategy to process a user's NL question is proposed. This paper presents a strategy to interpret and convert a NL question containing modifier term in order to get answers from the KB. The proposed QA system model utilizes the combination of UM and RF approach. The paper is organized as follows: Section 2 discusses the related works in QA system. The proposed model will be presented in Section 3. Section 4 discusses the experimental analysis. Section 5 sums up the paper with conclusions and future works.

II. RELATED WORKS

In QA system, the user will provide natural language question (NL question) to the system. Normally, there are three main components involved in the QA system; *Question analysis*, *Document processing* and *Answer processing*. The submitted NL question will be processed through all these stages based on the standard rules of natural language processing mechanism. Knowledge base (KB), ontology, WordNet and some others NL related database are the sources of answer; will be tied with answer specification based types of phrases generated by the question analysis component. The document analysis component extracts several answer candidates in the form of text fragments or passages and passes to the answer processing component. Figure 1 depicts the general components of QA system. Several approaches have been developed for QA systems.

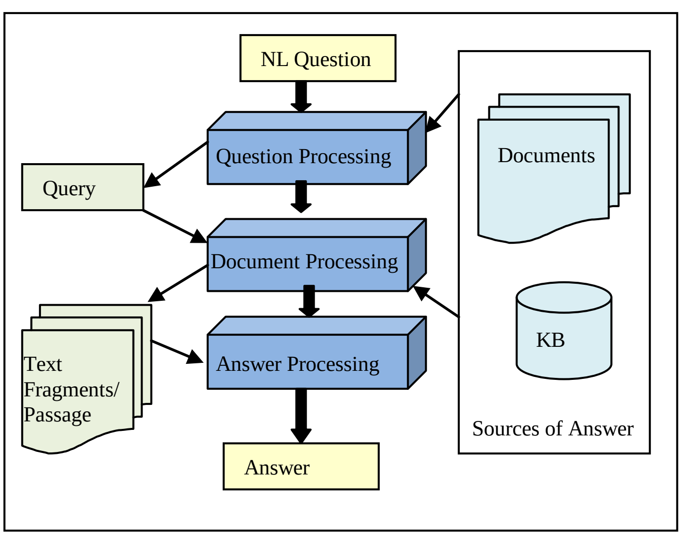

Fig. 1. General Components for QA system

As in AquaLog [14], two main components are introduced; the Linguistic Component which transforms the NL-question into NL Query-triple format form; and the Relation Similarity Service (RSS) Component that maps the Query-triple format with ontology used to produce the Ontology-Compliant queries known as Onto-Triples. AquaLog uses triple-based data model, which takes the form of <subject, predicate, object>. Also, AquaLog utilizes a portable and contextualized learning mechanism to learn user's jargon in order to improve the system's performance.

QACID [10] relies on the ontology, a collection of user queries, and an entailment engine that associated new queries to a cluster of queries. Each query is considered as a bag of words, and the mapping is done through string distance metrics and an ontological lexicon to map words in NL queries to instances. This system depends on the variety of questions collected in the domain. The process is domain-dependent, lack of portability and can only be applied to domains with limited coverage.

In [9] proposed a technique to handle comparative and evaluative question answering for business domain. A procedure to identify the terms in the question is introduced which later will be used in comparison or evaluative process of the queries.

FREyA [1], [2] is a Feedback Refinement and Extended Vocabulary Aggregation system. It combines syntactic parsing with knowledge in ontology for reducing customization effort. The rules are not used in this system, instead the knowledge encoded in ontology is given to understand the user's question. The syntactic parsing is used to get a precise answer. The ontology concepts are identified and verified initially. Mapping of user query with ontology concept is implemented in two ways, either automatically or by the help of user.

[13] developed a tool called YourQA which utilizes user modeling technique and depends on a web search engine to generate answers from a KB. The tool is an open-domain interactive QA system and provides answers including descriptions and definitions. Dialog model and dialog manager are used to implement the open-domain interaction. YourQA applies user modeling technique to filter the documents in KB and re-rank the answers based on the degree of match with the user's profile. Three parameters used in the user's profile: age range, reading level, and interests. The interactive YourQA version is tested under user-centered environment and the results show that user's satisfaction is better than baseline version of YourQA.

Many researchers consider RF technique in information/image retrieval systems, and it is proven that RF could improve the answers ranking process [3], [5], [6]. In order to improve a QA system, [6] experimented a pseudo relevance feedback for a probabilistic information retrieval system. The model assumed that the document/information which contains the correct class name entities is having more potentiality of relevance than those are not containing them. In other words, although the documents have the correct topic but with no named entity of the expected answers category will not give high probability of relevancies. By using named entity parameter, the pseudo relevance feedback can help the users to target the relevant documents at the top rank and eliminate the non-relevant documents more effectively.

This research focuses on incorporating the UM and RF approach which act as a general strategy in interpreting and converting user question into further-processed queries for QA system.

III. THE MODEL

The general context of the model can be formalized as follows: U, a user that provide a NL query and has the specific requirement of retrieving a non-empty set of answer from knowledge base; a knowledge base that contain a set of information labeled as $I = \{i_1, i_2, \ldots, i_n\}$ and a set of $S_i \subseteq I$ which is contains information that totally (or partially) relevant to the user's requirement. To produce S_i, the QA system is depended to the user requirement and all the information in the knowledge base, which can be represented as the function *f: (U, I)*.

In the proposed QA system model, the user requirement is defined as the combination of three elements and can be denoted as $U = (U_{\text{nl-query}}, U_{\text{user-modeling}}, U_{\text{relevance-feedback}})$. Within this context, each element can be defined as:

- $U_{\text{nl-query}}$ concerning the goals/answers to be retrieved
- $U_{\text{user-modeling}}$ is a set of user interest, and can be defined as set of user understanding domain, $U_{\text{user-modeling}}$ = {knowledge base concept, question context, language theory}
- $U_{\text{relevance-feedback}}$ is a set of interactions with the system, and can be defined as a set of users action activities, $U_{\text{relevance-feedback}}$ = {modification of term weight, query expansion, query simplifying}

Based on the definition given above, therefore a complete function of the proposed QA system model is as follows:

$$f: (U_{nl\text{-}query}, U_{user\text{-}modeling}, U_{relevance\text{-}feedback}, I) \rightarrow (S_i, L_i) \quad (1)$$

Where L_i is the quality of S_i to the query that can be calculated by using the standard information retrieval measurement.

A. The Answer Type Identification Architecture

Question interpretation strategy aims to extract some terms which will likely be used in finding the answer. The extracted terms are the focus, head focus, the modifier of the head focus and the focus complement. For each question, the focus, its modifiers (e.g. adjective, noun complement, etc.) and complement contribute an important part in finding the answer. The rules, which determine the question focus, depend essentially of the syntactic form of the question, and very often the question focus corresponds to the subject of the question. In later part of this section, definition of these extracted terms or also known as lexical elements will be defined.

Figure 2 is the Answer Type Identification Architecture (ATI). In the proposed model, user modeling technique is used to get user interest and profile in order to filter and classify the answers of the user question. To improve the query specifications provided by the user, relevance feedback technique is applied in the second stage of answer type identification process. Syntactic and semantic knowledge are used to interpret these terms in order to get the expected type of the answer.

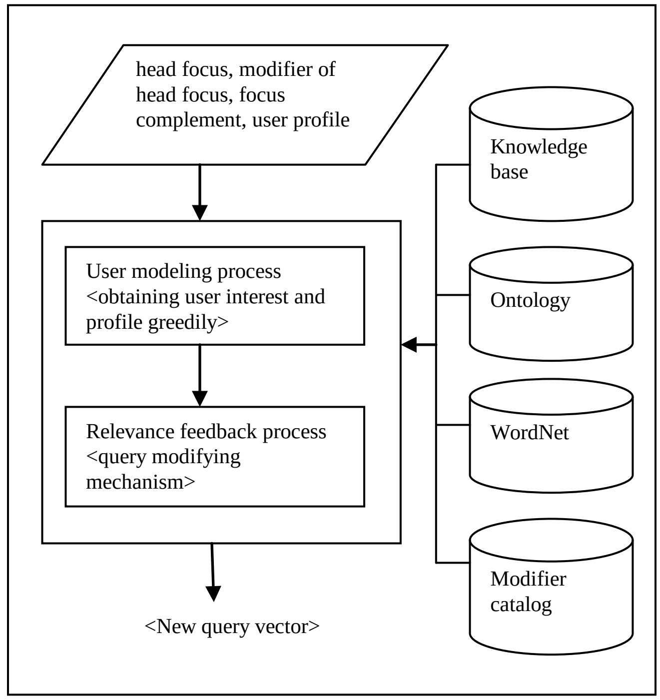

Fig. 2. Answer type identification (ATI) architecture

In this model, four (4) lexical elements are considered and can be explained as follows: (Example of lexical elements terms are depicted in Table 1)

- *Focus:* A word or a sequence of words in the question which indicates the interest of the question, disambiguates and emphasizes the type of answer expected. Commonly focus is the head of the first noun or verb in the question after removing stop words, auxiliary and copulative verbs.
- *Head-focus:* The question focus which can be extracted from focus terms which assigns modifier to modifier-head-focus focus and/or complement respectively
- *Modifier-head-focus:* Terms which modify a noun or a pronoun by describing, identifying, or quantifying words. Modifier-head-focus often precedes the noun or the pronoun which it modifies.
- *Focus-Complement:* Noun with complement.

Apart of the lexical elements, several pre-requirements components will also be needed and they are defined below. Table 2 is the example of the Modifier Catalog (it is not an exhaustive catalog).

- *Knowledge Base (KB)/Ontology:* Currently the proposed QA model works with a Raymond Mooney's (http://www.cs.utexas.edu/~ml/nldata/geoquery.html) Geobase ontology.

- *WordNet:* Lexical resource
- *Modifier Catalog:* A list of modifier terms based on categories. A modifier term which extracted from the KB will be matched with modifier terms from the catalog.

TABLE 1

EXAMPLE OF FOCUS, HEAD-FOCUS, MODIFIER-HEAD-FOCUS AND FOCUS COMPLEMENT FROM NL QUESTION

NL Question	Focus	Head Focus	Modifier Head Focus	Focus Complement
How big is texas?	big, texas	texas	big	texas
What are the cities in california?	Citie,cali fornia	cities	-	california
What are the cities of the state with the highest point?	cities, state, highest, point	cities, point	highest	-
What are the highest points of all the states?	highest, points, all, states	point	highest, all	-
What is the capital city in texas?	capital, city, texas	capital	-	texas
How tall is the highest point in montana?	tall, highest, point, montana	point	tall, highest	montana
What are the cities in california?	cities, california	cities	-	california

TABLE 2

EXAMPLE OF MODIFER CATALOG

Category	Modifier Term	Rules
size	big, biggest, huge, immense, enormous, massive, vast, large, largest, wide, widest, high, highest, long, longest, tall, tallest , gigantic	max
	small, smallest, little, tiny, low ,lowest, short, shortest, teeny , petite	min
	same as	equal
	medium, intermediate	average
Age	old, ancient, aged, elderly, ageless, overage	max
	young, juvenile, adolescent, teenage, underage, youthful,	min
	middle	average
Time	late, behindhand, delayed, ill-timed, , tardy, unearthly, traditional, old-fashioned, old, swift , long,	max
	brief, short, early, premature, modern, young, new, quick, rapid, slow,	min
	punctual,	definite
Quantity	one, two, three, first, second, third, empty, heavy, numerous,	definite
	Abundant, heavy, substantial, significant,	max
	light, insignificant	min
	How many	count

B. User Modeling and Relevance-Feedback Approach for Question Interpretation Strategy

The aims of using UM approach are to increase readability of answers and to filter the huge of potential relevance answers from the knowledge base. In the stage of UM, the model attempts to capture user interest based on three (3) aspects that can be labeled as $U_{user\text{-}modeling}$ = {knowledge base concept, question context, language theory}. These aspects should be the set of class concept and properties that construct the knowledge base and could fashion the contents of the answers retrieved. All the information provided by the user will be saved in a user profile which can be obtained when the user decides to post another query to the QA system.

For the knowledge base concept, question context and language theory aspects, the model may consider instances to the class concept, concepts, properties, query focus or/and query focus complement. For instance name of place, name of river, point of area, mountain etc. The defined concepts/terms of the query are represented in a vector space model as follows:

$$Q' = \{(C_1, W_1), (C_2, W_2), \ldots (C_n, W_n)\} \quad (2)$$

Where, C represents the instances to the class concept of query or terms and W is the weight of C. The weight of each term can be determined by the formula *tf* × *idf* where *tf* is the term frequency in knowledge base and *idf* is the inverse term frequency in the whole knowledge base collections.

For the relevance feedback approach, the model considers three (3) aspects of modifying the query representation and can be defined as $U_{relevance\text{-}feedback}$ = {modification of term weight, query expansion, query simplifying}. The QA system can propose to the user to modify the question by reformulate the weight of each terms exist in the query. The modification is done based on the requirement domain of the user. The users can also be suggested to expand the query by inserting new terms such as adding new head focus, modifier head focus, focus complement and category of expecting answers. Simplifying the query into several sub-queries is also considered in RF stage in the proposed model. The sub-query produced will lessen the complexity of the query therefore providing an easier answer classification.

In this model, the submitted NL query is represented into vector space model and can be defined as follows:

$$Q = \{(T_1, W_1), (T_2, W_2), \ldots, (T_n, W_n)\} \quad (3)$$

Where, T represents the terms that exist in the query which include {head_focus, modifier_head_focus, focus_complement} and W is the weight of T. The weight of each term can be determined by the formula *tf* × *idf* where *tf* is the term frequency in knowledge base and *idf* is the inverse term frequency in the whole knowledge base collections.

When the RF mechanism is implemented to the Q by modifying the query (modification of weight, adding new modifiers, simplifying the query), then the following set of solutions (new query vector) will be produced:

$$Q''_i = \{\{(T_{1,i}, \Delta W_{1,i}), (T_{2,i}, \Delta W_{2,i}), \ldots, (T_{n,i}, \Delta W_{n,i})\}, \{(T_{n+1,i}, W_{n+1,i}), (T_{n+2,i}, W_{n+2,i}), \ldots, (T_{k,i}, W_{k,i})\}\} \quad (4)$$

Where, ΔW represents the change of weight value, and $T_{n+1,i}$,

$T_{n+2,i}$, $T_{k,i}$ are the new inserted terms that associated with their weight $W_{n+1,i}$, $W_{n+2,i}$, $W_{k,i}$ respectively and i > 0, k > n

To determine the best solution to be used, the initial NL query Q is compared with all solutions $Q_i^{''}$, and the similarity scores are computed. The similarity score formula used is as follows:

$$Sim(Q, Q'') = \sum_{m=1}^{k} (W_m \times W_{im}) \tag{5}$$

Then, the $Q_i^{''}$ which has the highest similarity score is direct sum with $Q^{'}$ for producing a new query vector. If Q_{new} is the new query vector that is submitted to the *answering processing* phase, then Q_{new} can be denoted as follows:

$$Q_{new} = (Q^{'} \oplus highest(Q_i^{''})) \tag{6}$$

IV. EXPERIMENTS AND EVALUATION

The Raymond Mooney's Geoquery dataset with 100 annotated user questions and Geobase ontology of US geographical information are used to test the proposed model. The proposed model will be known as QAUF (**Q**uestion **A**nswering system with **U**ser modeling and relevance **F**eedback). In this experiment, QAUF performance is evaluated with Q and Q_{new} in terms of precision, recall and F-measure. Aqualog and FREyA QA systems are also selected for the performance comparison purposes. Table 3 and Table 4 show the results of the experiment conducted respectively.

A. Performance Measurement

QAUF is evaluated using the formal information retrieval metrics calculator called precision (P), recall (F) and F-measure (F). These metrics used three (3) parameters, which are *number of relevant answers retrieved (A), number of irrelevant answers retrieved (B)* and *number of relevant answers not retrieved (C)*. Precision, Recall and F-measure can be defined as follows:

$$P = \frac{A}{A+B} \tag{7}$$

$$R = \frac{A}{A+C} \tag{8}$$

The F is the harmonic mean of P and R, and it is used for giving a summarizing overview and for balancing the precision and recall values.

$$F = 2 \bullet \left(\frac{P \bullet R}{P+R} \right) \tag{9}$$

B. Experimental Result and Analysis

In QAUF, Q is compared with Q' and Q" to find out the similarity of the user's question. Then, $Q_i^{''}$ which has the highest similarity score is direct sum with $Q^{'}$ to produce a new query Q_{new}. Q_{new} is submitted to the answering processing phase for an answer. Table 3 summarizes the comparison result on precision, recall and F-measure between Q and Q_{new}.

TABLE 3
COMPARISON ON QAUF ANSWERS RETRIEVAL PERFORMANCE BETWEEN Q AND Q_{new}

Recall		Precision		F-measure	
Q	Q_{new}	Q	Q_{new}	Q	Q_{new}
0.38	0.91	0.35	0.90	0.36	0.91

Table 3 shows almost 59% increments on Q_{new} Recall and 61% more on Q_{new} Precision. User's NL question which do not contain any modifier term (E.g.: `What is the capital city in Texas?`) is answered successfully by QAUF based on Q representation. However, using Q representation, QAUF fails to give a correct answer for any question containing modifier term (E.g.: `How big is Texas?`).

Q_{new} which denotes the combination of UM and RF manage to help QAUF interpret the question precisely. Q_{new} becomes the new query which is used to query the KB for the correct answer. QAUF is also compared with existing QA systems, AquaLog and FREya. Based on Table 4, QAUF outperformed the AquaLog and FREyA. This is because, AquaLog does not accommodate the type of *how much/how many* questions and questions contain modifier. FREyA, on the other hand, yielded lower precision as reported in [2]. This is because, based on the experiment conducted; a returned answer has to be correct to the intended question without considering partial or incorrect answer.

TABLE 4
COMPARISON ON ANSWERS RETRIEVAL PERFORMANCE

	AquaLog	FREyA	QAUF
Precision	0.42	0.88	0.90
Recall	0.43	0.92	0.91
F-measure	0.42	0.90	0.91

UM is exploited to filter and classify the answers of user NL question based on the KB concept, question context or/and the language theory. Later, RF is applied to modify the query representation by modifying the term's weight, query expansion or query simplifying. Positively, with combination of UM and RF technique, the process of interpreting and converting user question into further processed has proven a significant leap as displayed in the experiments' results.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a strategy to semantically interpret the user's NL question is presented. QAUF manages to solve issues in QA system by combining the user modeling theory and relevance feedback technique. In terms of precision, recall and F-measure metrics, QAUF shows favorable results compared to other QA system. Experiment results on Geobase ontology of US geographical information demonstrate that the proposed QAUF is effective, sensitive and performs significantly better than Aqualog and FREyA.

At present, QAUF is still undergoing a continuous enhancement. In the near future, QAUF aims to incorporate a question interpretation strategy on the remaining type of questions that QAUF fails to correctly interpret. Extensive experiments with larger datasets and using different domain datasets will be done to verify the proposed strategy.

ACKNOWLEDGMENT

Special thanks to Universiti Sultan Zainal Abidin (UniSZA), Universiti Putra Malaysia (UPM) and Management of Faculty of Informatics, UniSZA.

REFERENCES

[1] D. Damljanovic, M. Agatonovic and H. Cunningham, FREyA: an Interactive Way of Querying Linked Data Using Natural Language”, Proc of the European Semantic Web Conference, 2010.

[2] D. Damljanovic, M. Agatonovic and H. Cunningham. “Identification of the Question Focus: Combining Syntactic Analysis and Ontology-based Lookup through the User Interaction”, Proc. of the 7th Language Recourses and Evaluation Conference, 2010.

[3] D. L. Lee, H. Chuang and K. Seamons, “Document Ranking and the Vector-Space Model”, IEEE Software, 1997, 67 – 75.

[4] H. Doan-Nguyen, K. Leila, “The Problem of Precision in Restricted-Domain Question Answering. Some Proposed Methods of Improvement”, The ACL 2004 Workshop on Question Answering in Restricted Domains, 2004, pp. 8-15.

[5] J. J. Rocchio, “Relevance Feedback in Information Retrieval”, In G. Salton (ed.), The Smart Retrieval System: experiments in automatic document processing, Prentice Hall, 1971, pp. 313 – 323.

[6] L. Pizzato, D. Molla and C. Paris, “Pseudo Relevance Feedback Using Named Entities for Question Answering”, Proc. Australasian Language Technology Workshop, 2006, 83 – 90.

[7] M. Diana, P. Wim and L. Yaoyong, “Metrics for evaluation of Ontology-Based Information Extraction”, Proc. Intl. Conference on WWW, May 2006.

[8] M.R. Kangavari, S. Ghandchi, M. Golpour, “A New Model for Question Answering Systems”, Journal of World Academy of Science, Engineering and Technology, 2008.

[9] N. Rose, T. Lim, P. Saint-Dizier, B. Gay, and R.E. Roxas, "A preliminary study of comparative and evaluative questions for business intelligence", Natural Language Processing, 2009. SNLP '09. Eighth International Symposium, Oct. 2009, pp.35-41, 20-22.

[10] O. Fernandez , R. Izquierdo, S. Ferrandez and J. L. Vicedo, “Addressing Ontology-based question answering with collections of user queries”.,Journal of Information Processing and Management, vol. 45, issue. 2, March 2009, pp 175-188.

[11] R. Baeza-Yates, B. Ribeiro-Neto, *Modern Information Retrieval***.** Harlow: Addison-Wesley, 1999.

[12] S. Quarteroni, “Personalized Question Answering”, TAL, Vol 51(1), 2010, 97 – 123.

[13] S. Quarteroni and S. Manandhar, “Designing an Interactive Open-Domain Question Answering System”, Journal Language Engineering, Vol. 15(1), 2009, pp. 73 – 95.

[14] V. Lopez, V. Uren, E. Motta, and M. Pasin, “AquaLog: An Ontology-driven Question Answering System for Organizational Semantic Intranets”. Journal of Web Semantics, vol. 5, 2007, pp. 72-105.