ISSN: 0976-3104
SPECIAL ISSUE
Mukherjee and Chakraborty

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

NLP BASED HYBRID KNOWLEDGE PROVIDER SYSTEM FOR TEXT AND IMAGE DATA EXTRACTION FROM A USER QUERY

Prasenjit Mukherjee and Baisakhi Chakraborty*

National Institute of Technology, Durgapur, INDIA

ABSTRACT

Syntactic analysis (Parsing) is a main method of analyzing a sentence in natural language. There are several techniques of parsing. This paper proposes a Hybrid Knowledge Provider System (HKPS) where permutation-combination (PC) based parsing technique and Grammatical Rules (GR) based parsing technique have been applied on a single system. HKPS is an automated system that shall be able to extract text and image based knowledge data from database.

Received on: 06th-May-2015
Revised on: 17th-July-2015
Accepted on: 21st–Aug-2015
Published on: 4th-Jan-2016

KEY WORDS

Knowledge Provider; Natural language query; Query-Response model; E-R diagram; HKPS

***Corresponding author: Email:** baisakhichak@yahoo.co.in, **Tel:** +91-9434788065 **Fax:** +91-343-2546406

INTRODUCTION

Computer linguistic refers to analysis of sentence formation in natural language by the computer. The process of analysis of sentence formation is called parsing. There are several techniques of parsing. The importance of Syntactic parsing is mediating between linguistic expressions and their meanings. Many works have been done on usefulness of syntactic representations for subsequent tasks such as relation extraction, semantic role labeling and paraphrase detection as in [1]. We have proposed two techniques of parsing on a Knowledge Management System (KMS) termed as Knowledge Provider System (KPS) in this paper which are, Permutation-Combination (PC) based parsing and Grammatical Rules (GR) based parsing. In the PC based parsing system, KPS reads natural language query and creates a conceptual form of database using the Permutation-Combination (PC) technique. The semi-automated system follows the client-server architecture to handle the queries. Some individuals termed Knowledge Workers (KWs) are associated with the KPS. KPS has a default database with certain queries and responses stored in it. If KPS is not able to generate the response of a NLP query or if data is not present in the database then situation is handled by the KWs. The PC based KPS has been proposed by the authors in [2]. GR based parsing technique is another approach in Knowledge Provider System (KPS) to create the conceptual form of database. GR based KPS follows English grammar rules. KWs are not associated with GR based AKPS. The system itself generates response to the client side. The natural language queries may be in assertive or interrogative sentences. GR based AKPS refers the grammar rules at runtime to extract nouns as an entities and verb as a relationship. The system uses these entities and initializes them in to the semantic table for creating conceptual form of database. Each response is generated from the default database. The default database needs to be updated from reliable sources. The GR based AKPS has been proposed and discussed in [3]. This paper proposes a Hybrid KPS (HKPS) where PC based parsing technique and GR based parsing technique have been applied in a single system. It is an automated system that able to extract text and image based knowledge data from database.

PREVIOUS RELATED WORKS

There are many natural language processing based models. A study of several of them has revealed a wide variety of application based models. A common sense filter system has been evaluated in [4] for the ReVerb Open IE system, applied as a method for answer validation in a Question Answering task applicable to a large database of

THE IIOAB JOURNAL
ISSN: 0976-3104

facts. This system queries a database to find the presence of arbitrary facts. A model of NLP considering the problem of learning commonsense knowledge has been discussed in [5]. The commonsense knowledge is in the form of first-order rules and noisy natural-language extractions. The noisy natural-language extractions are produced by an off-the-shelf information extraction system as in [5]. In complex inference problems involving long, complicated formulae, the Markov logic is used to integrate logical and distributional information in natural-language semantics results. The system proposed a new inference algorithm based on Sample Search. The algorithm computes probabilities of complete formulae rather than ground atoms in [6]. Generating sentences from images is an idea of combining visual and linguistic information that has been gaining traction in the Computer Vision and Natural Language Processing communities over the past several years. The motivation of the model for a combined system is to generate richer linguistic descriptions of images and investigate the performance of several integrated information from language and vision systems as in [7].

Script signifies the sequence of knowledge of stereotypical event and it can aid text understanding. The Initial statistical methods have been developed in this model to learn probabilistic scripts from raw text corpora in [8]. The distributional models do not succeed to distinguish between semantic relations and the distributional models cannot be a valid model of conceptual representation. Then the approach is to use the Distributional Inclusion Hypothesis, which states that hyponyms tend to occur in a superset of contexts in which their hyponyms are found and thus propose a robust supervised approach that achieves accuracies of .84 and .85 on two accessible datasets as in [9]. An interactive text to 3D scene generation system which learns the expected spatial layout of objects from data has been discussed in [10]. As per the system, a user provides input in natural language text from which the system can extract explicit constraints on the objects and it should appear in the scene as in [10]. A natural language model has been discussed in [11] that integrate NLP with computer vision. The model has proposed a strategy for generating textual descriptions of videos and has used a factor graph to combine visual detections with language statistics which has improved recognition and description of entities in real-world videos [11]. TOKENSREGEX is a NLP based system which follows a framework for defining cascaded regular expressions over token sequences. TOKENSREGEX is a part of the Stanford CoreNLP software package and it is used for various tasks which require reasoning over tokenized text as in [12]. Two structured prediction models for joint parsing and multiword expression identification have been developed and proposed in [13] wherein earlier, syntactic analysis and multiword expression identification had been proposed as alternative methods of NLP. The experiments on prediction models in [13] show that both models can identify multiword expressions with much higher accuracy than a state-of-the art system based on word co-occurrence statistics.

This paper proposes an application of PC based Parsing technique and Grammatical Rules based Parsing technique in to a single system. Section 2 discusses on previous related works, section 3 stated the principle of the architecture where PC based Parsing algorithm is associated into GR based Parsing technique and generates text based and image based responses by the algorithm. Section 4 describes Algorithm and its Rules, Section 5 describes Methodology and Section 6 winds up with conclusion and future works.

THE PRINCIPLE OF HYBRID KNOWLEDGE PROVIDER SYSTEM (HKPS)

Knowledge Management System is domain specific and needs an organized knowledge database for extraction of knowledge. The KMS is meant to work for e-service oriented organizations providing regular services to clients. The authors have proposed a PC based KPS model in [2] and GR based AKPS model in [3] to handle Natural language client queries. Both systems follow the request-response model and extract knowledge data from knowledge database. Both of the systems use queries in natural language which may be assertive or interrogative. The role of the KWs in the KPS has been discussed in [2]. PC based KPS may not able to manage complex queries whereas GR based AKPS system may able to manage these complex queries where complexity refers to handling of auxiliary verb as main verb. The parts of speech table (POS) has been used by the KPS to identify the unique words from the query sentences. KPS applies the PC based parsing technique to select the right combination of two entities and relationship of these entities generating a conceptual form of database which creates the sql-query for the extraction of data from the default database as a response. If records are not present into the database, then Knowledge Workers (KWs) handles the requested query and updates the database. The parsing technique of KPS has been described greatly in the paper [2].

THE IIOAB JOURNAL
ISSN: 0976-3104

Process of PC based Parsing Algorithm

i.User posts Query in Natural Language.
ii. Query is checked for validity. If valid then go for further Processing, if not, then prompt again.

iii. The Query in Natural Language is tokenized. Query (String) = {S1, S2, S3, S4,…… Sn.} Where, S1, S2 are tokens.

iv. Each token is checked with English Grammar, placed in a Grammar Table (except for unique Tokens) and removed from the Token List after its placement in the grammar Table.

```
For i = 0……n
{
If( Si== wh[])
            // Insert s1 into the grammar table in its position.
                  Else
If(Si == Pre[])
           // Insert S3 into the grammar table in its position.
                  Else
If(Si == Aux[])
// Insert S3 into the grammar table in its position.
.
.
If(Sn== Con[])
// Insert S3 into the grammar table in its position.
}
```

v.Check again for all unique Tokens (Words) and if the Unique Token matches with the content of Grammar table, then it is removed from the Token List. The Algorithm is applied to the rest of the Token List in a similar manner.
If unique Token does not match with the grammar table, then apply the Algorithm on Token List.
vi. After applying the Algorithm will get the Conceptual form of Database (E-R representation) from the Natural Language Query.
vii. If Database and Tables already created response the Query else create the Database and Tables and send query to the Knowledge Workers (KWs). Knowledge Workers will update the Database from where the response will be generated.
viii. Resultant Database stores the response from where the user will get the Query Response.

Architecture of Hybrid Knowledge Provider System

The proposed HKPS model is based on the client server architecture where requests are sent to the HKPS which generates corresponding responses. The modular architecture of HKPS is shown in **Figure– 1**. Clients access the HKPS through a user interface where queries are posted in natural language. The Parsing Module runs two processes that generate the conceptual form of database from the user's queries in natural language. The database module defines three types of Databases, namely, Temporary database, Default database and Result database. The Temporary database stores the unmanaged queries where request data is not present in Default database. Default database is a main database from where the response is generated to the client side. The Result database temporarily stores the response for the clients. The Context diagram in **Figure– 2** indicates the working principle of HKPS which is self explanatory.

ALGORITHM AND ITS RULES

In The HKPS reads the queries in natural language. The query sentence may be of assertive or interrogative. The HKPS creates rules at runtime using the phrases of assertive and interrogative sentences.

THE IIOAB JOURNAL
ISSN: 0976-3104

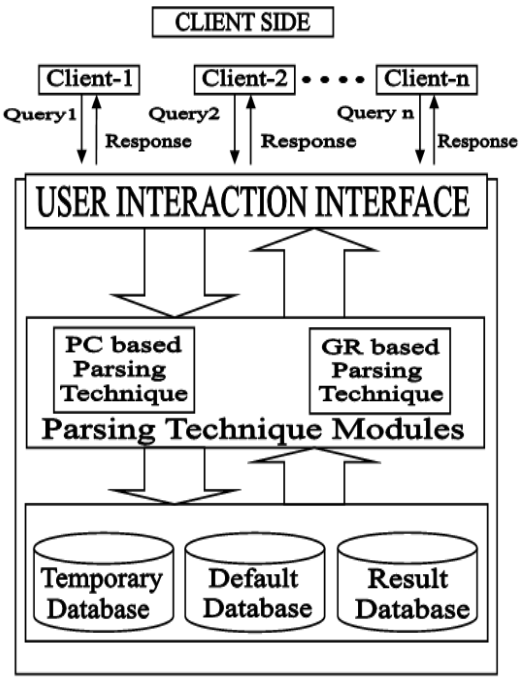

Fig: 1. Architecture of HKPS

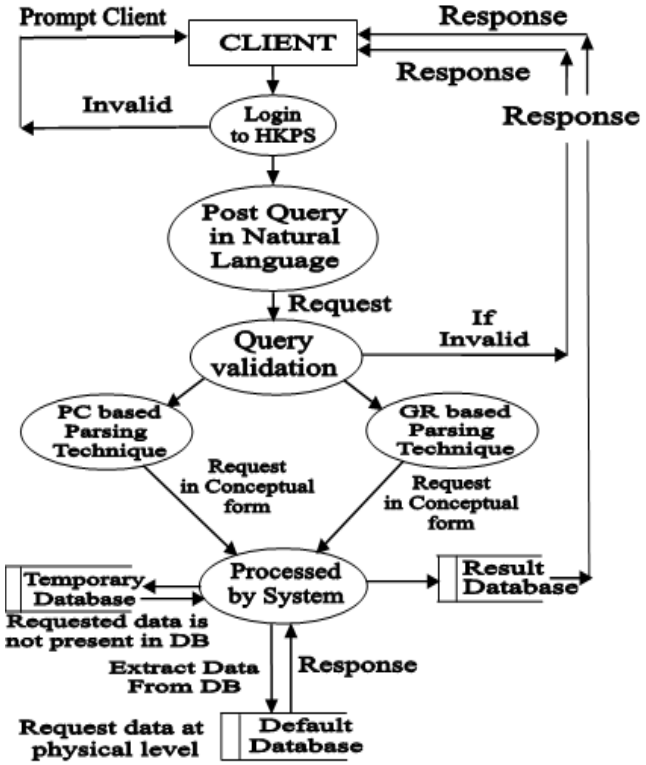

Fig: 2. Context diagram of HKPS

THE IIOAB JOURNAL
ISSN: 0976-3104

Assertive Sentence formation

Assertive Sentence - Noun phrase + Verb phrase + Complement.
The formation of rules of Assertive Sentence at runtime of the HKPS system will be
"Noun + Verb + Noun", "Determiners + Noun + Aux + Adverb + Verb + Preposition + Noun "or
"Adjective + Noun + Preposition + Verb + Preposition + Determiners + Noun".

Interrogative Sentence formation

The syntactic treatment of Interrogative sentence is different from Assertive sentence in English language. In a sentence, the Noun, Verb, Adjective, Adverb represents unique word(s) and verb does not have any fixed position. Verb may appear before Noun or after Noun. HKPS system cannot maintain any database for these unique words identification from a sentence. HKPS utilizes the phrases of interrogative sentence to create rules at runtime.
Interrogative Sentence - Wh Phrase (attached Auxiliary Verb) + Noun Phrase +Verb Phrase + Complement.
Wh phrase contain the Which, Who, What, etc. with am, are, is, was, were etc.
The Rules created at runtime by the HKPS are-
"Wh + Aux + Determiners + Noun + Verb + Preposition + Noun" or
"Wh + Determiners + Noun + Verb + Preposition + Determiners + Noun + Noun".
The HKPS system may able to increase the rules when phrases of interrogative sentence are increased. Rules may be modified and updated from time to time providing an advantage to the system.

This paper combines PC based parsing technique of KPS proposed and discussed in [2] with GR based parsing technique of AKPS as discussed in [3] to generate the HKPS. This system is fully automated and Knowledge Workers are not involved or associated as had been in PC based KPS discussed in [2]. The HPKS system has ability to retrieve text data as well as image data. In this application the image retrieval algorithm has been used. The HKPS model is domain specific and follows the query response model to extract knowledge from default database. The default database stores the knowledge data either in text mode or image mode. The HKPS generates response from this default database. If the HKPS fails to generate response for any query then it will be stored in a temporary database. Whenever the default database is updated from reliable sources like Database Administrator, System Administrator or any other Administrator, system will generate the response of the pending request. Two effective parsing algorithms work in tandem in HKPS to handle user's request and generate response after selecting the appropriate parsing technique. The selection process of parsing method has been done on the basis of number of unique tokens. If number of unique tokens equal to three (3) then PC based parsing algorithm will be applied otherwise GR based parsing algorithm will be applied but number of unique tokens should be greater than one (1) and less than three (3) or greater than three (3).

Process

i. User posts Query in Natural Language.
ii. Query is checked for validity. If valid then go for further Processing, if not, then prompt again.
iii. The Query in Natural Language is tokenized. Query (String) = {S1, S2, S3, S4,…… Sn.} Where, S1, S2 are tokens.
iv. Now apply the Algorithm on Token List.

Algorithm

1. Read Input Statement S.
2. Each token is checked with English Grammar, and identify that how many unique tokens are there in the query sentence.
3. IF (UniqueTokens == 3) THEN apply the KPS algorithm
3.1. The Query in Natural Language is tokenized. Query (String) = {S1, S2, S3, S4,…… Sn.} Where, S1, S2 are tokens.
3.2. Each token is checked with English Grammar, placed in a Grammar Table (except for unique Tokens) and removed from the Token List after its placement in the grammar Table.
3.3. Check again for all unique Tokens (Words) and if the Unique Token matches with the content of Grammar table, then it is removed from the Token List. The Algorithm is applied to the rest of the Token List in a similar manner.
3.4. If unique Token does not match with the grammar table, then apply the permutation function on the Token List.

THE IIOAB JOURNAL ISSN: 0976-3104

3.5. After permutation function being applied, the algorithm will show the noun-verb-noun combination insert them into the semantic table.
4. ELSE
IF ((UniqueTokens >3) || (UniqueTokens < 3 && UniqueTokens >1)) THEN apply the AKPS Algorithm
4.1. Split Input Statement S into the String Array called STR[n].
4.2. First check the STR[0] equal to the words- "who","which","what"….etc.
IF STR[0]== "who"|| IF STR[0]== "which"|| IF STR[0]== "what" ….. THEN
CALL Interrogative_Sentence(STR[n]) method
ELSE
CALL Assertive_Sentence(STR[n]) method
END IF
5. Interrogative_Sentence(STR[n]) method:
5.1 Create the each Rule and store it on Rules Array.

5.2. Check the STR. length with the each rule length from Rules[n]. If the rule match from Rules[n] then send the selected rule and STR[n] to Parsing method which will either return 0 or 1 to a variable.If variable value is 1 then Insert STR[n] in Sentence Table in Database and Insert selected Rules in Rules Table in Database.

6. Assertive_Sentence(STR[n]) method

6.1. Create the each Rule and store it on Rules Array.
6.2. Check the STR. length with the each rule length from Rules[n]. If the rule match from Rules[n] then send the selected rule and STR[n] to Parsing method which will either return 0 or 1 to a variable. If variable value is 1 then Insert STR[n] in Sentence Table in Database and Insert selected Rules in Rules Table in Database.
7. int Parsing(string Rules, string STR)
7.1. Declare Parts of Speech List: WH[n]={"what","which"…}
AUX[n]={"is","am","are"…}
PRE[n]={"of","to","in","for"…}
DET[n]={"a","an","the"…}
PRO[n]={"I","you",he"….}
UNK[n]={"Noun","Adjective","Verb"….}
7.2. split Rules to R[n] and STR to S[n].
Declare Result[0....n] = NULL. Check each word from S[n] with each word from WH[n], AUX[n], DET[n],PRO[n] and UNK[n] and store result in Result Array.
7.3. Initialize the variable value=0
FOR i=0,j=0, TO Result.length -1 AND R.length-1 STEP 1
IF Result[i] != R[j] THEN
Value INCREMENTED BY 1
END IF
END FOR
7.4. If Value > 0 then the parsing method will return 0 else 1.
8. Select Row one by one from Rules Table and Sentence Table in Database.
8.1. Each Row of Rules Table and Sentence Table initialize in SS1 [n] and SS2 [n].
8.2. Each word from SS1 array inserts it into the grammar table as per SS2 array.
9. From Grammar Table select the values from Noun, Verb columns and properly initialize in Semantic Table.
10. Select values from semantic table and check each word in the synonym database where the actual word is stored with the synonyms.
10.1. Initialize the Entity1, Entity2, Relationship values from Semantic Table to string variables entity1, entity2 and verb.
10.2. Value of entity1 and entity2 variables matches with synonyms words in Relation, Attributes Table and value of verb variable matches with synonyms words in Relationship Table.
10.3. If value matched then reinitialize the values from Tables to entity1,entity2 and verb variables.
10.4. Initialize the Filter Table through the values of entity1, entity2 and verb variables.
10.5. In next step the system further processes the Filter table and create the SQL Query from the Filter table using the database conceptual form.
10.6. The SQL query fetches the knowledge data from the default database as a response and store it into the result database. If the knowledge data not present in the default database then User/Clients query will be stored in temporary database termed as pending request. When default database will update as per the pending request then System will generate the response to the User/Clients.

THE IIOAB JOURNAL
ISSN: 0976-3104

11. The knowledge data can be the text data or image data. If the data is an image into the database then the system runs this algorithm below to retrieve the images from the default knowledge database.
11.1. Declare the byte array to store image bytes
byte[] photo_array;
11.2. SQL-query fetch the text as well as image data from the default database and stored into a dataset call ds.
11.3. In the next step initialize the another bytes array through the dataset variable ds.
byte[] bytes = new byte[ds.Tables[0].Rows[0][3].ToString().Length * sizeof(char)];
11.4. Copies a specified number of bytes from a source array to destination array.
System.Buffer.BlockCopy(ds.Tables[0].Rows[0][3].ToString().ToCharArray(), 0, bytes, 0, bytes.Length);
11.5. Initialize the data of bytes array variable to photo array variable.
photo_array = bytes;
11.6. To represents the bytes data from the variable photo_array, the memory stream instance created.
MemoryStream ms = new MemoryStream(photo_aray);
11.7. Initialize instance of memory stream to Picture Box control or Data Grid control to show the image like-
pictureBox1.Image = Image.FromStream(ms);
pictureBox1.SizeMode = Picture Box Size Mode. Stretch Image;

METHODOLOGY AND TOOLS

The HKPS has been designed using C# at front end and MS-Access at back end as the designed tools. The HKPS has three events. The First event is User Interface event, second is PC based parsing process event, third is Grammatical Rule based parsing process event. Sequence of these events is as follows.

User Interface event

i. The User will Login to the HKPS and post a Query string *"University has opened many Departments"* in natural language into the HKPS.
ii. After clicking the "QUERY SUBMIT" Button, the Query in natural language will be tokenized. Each token will be checked by the Parts of Speech Table to identify the unique tokens. If the number of unique tokens is equal to 3 (three) then PC based parsing process event will be called else, if the number of unique tokens is greater than 3(three) then GR based Parsing process event will be called. GR based Parsing process event can be called if number of unique tokens is less than 3(three) but greater than 2(two).

PC based parsing process event

i. The unique tokens will be extracted from the Query string in natural language. The tokenized query string will be checked by The Parts of Speech Table. Each word is a token and tokens corresponding to the words "has" and "many" in the Parts of Speech Table will be removed. Only three unique tokens "University", "opened" and "Departments" will remain for further processing.
ii. The remaining tokens will be permuted and there will be many combinations from which the system will select the right combination. As per the combinations, the tokens will be inserted into the **Table– 1** (Semantic Table) of the database of the HKPS.

Table: 1. Semantic Table in MS Access Database

Entity1	Relationship	Entity2
University	opened	Departments

iii. The Conceptual Form (E-R Diagram) of Database is created from the semantic table number **Table– 1** where Entity1, Entity2 are the Relations of the Database and Relationship constructs the connection between two Relations.

Grammatical Rules Based parsing process event

i. Assume the Query string from the user is *"What are the Computer courses offered by the Burdwan University".*
ii. Split the string into tokens which correspond to each word of Query sentence. Check the first token if "Wh" word or not. If true, then "Interrogative Sentence" Method will be invoked.
iii. Create each Rule at runtime and store it on Rules Array in Interrogative Sentence Method. Compare Query string length with the each rule length from Rules Array and Insert each token in **Table– 2** (Sentence Table in Database) and Insert each word from matched rule in **Table– 3** (Rules Table in Database).

THE IIOAB JOURNAL
ISSN: 0976-3104

Table: 2. Sentence Table in MS Access Database

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
What	are	the	Computer	courses	offered	by	the	Burdwan	University

Table: 3. Rules Table in MS Access Database

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
Wh	Aux	Det.	Noun	Noun	Verb	Prep.	Det.	Noun	Noun

iv**.** Parsing Method will be called and as per the algorithm and the method would return either 0 or 1.
v. Apply step 8 of the algorithm after selecting row from **Table– 2** (Sentence table) and **Table– 3** (Rules Table) in Database. After completion of step 8 it creates two arrays that will be inserted into **Table– 4** (The Grammar Table).

Table: 4. Grammar Table in MS Access Database

Wh	Det.	Noun	Pronoun	Aux	Adj	Adv	Verb	Preposition
What	the the	Computer courses Burdwan University		are			offered	by

vi. Select the Noun, Aux, Verb, and Preposition from **Table– 4** (Grammar Table) and do the same that is mentioned in STEP 9 of the main algorithm. A Semantic Table is formed with entities and their relationship as shown in **Table– 5**.

Table: 5. Semantic Table in MS Access Database.

Entity1	Relationship	Entity2
Burdwan University	Offered	Computer courses

vii. Using Entity1, Entity2 and Relationship values from **Table– 5** (Semantic Table) will be produced the conceptual form (E-R Diagram) of Database.
viii. The system will construct the SQL Query using Semantic Table. After creating the SQL Query, the system will run this query and generate response to the client side.

CONCLUSION AND FUTURE WORKS

This The HKPS model works on the basis of two main processes, the PC based parsing process and GR based parsing process. Any one of the processes may be chosen depending on the nature of the client query. The model is domain specific. The HKPS is modeled to be used for e-government services. The e- government services may handle several different domains and there might be wide variety of queries. Since this HKPS is dedicated to one domain, future work lies on developing e-government service models to cater simultaneously to different services pertaining to multiple domains on same platform.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

ACKNOWLEDGEMENT

This Research work has been done at Project Research Lab, Dept. of Information Technology, National Institute of Technology, Durgapur, under Visvesvaraya PhD Scheme, Deity, Govt. of India.

FINANCIAL DISCLOSURE

Financial support was received from Visvesvaraya PhD Scheme, Deity, Govt. of India (Order Number: PHD-MLA/4(29)/2014-2015 Dated- 27/4/2015) to carry out this research work.

THE IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] Berant Socher Richard, Bauer John, Manning Christopher D. and Ng Andrew Y. [2013] Parsing With Compositional Vector Grammars. *In Proceedings of the Association for Computational Linguistics (ACL)*.

[2] Mukherjee P, and Chakraborty B. Knowledge Provider System with natural language data modeling. *In Proceedings of ISCA 29th. International Conference on Computers and their Applications, Las Vegas, USA*: 229−236.

[3] Mukherjee P, Chakraborty B, Debnath CH Narayan. [2015] A Natural Language Processing based Automated Knowledge Provider System with Speech Recognition. Accepted in *28th International Conference on Computer Applications in Industry and Engineering*, Hilton San Diego, San Diego, California, USA.

[4] Angeli Gabor, Manning Christopher. [2013] Philosophers are Mortal: Inferring the Truth of Unseen Facts. *In Proceedings of Computational Natural Language Learning (CoNLL)*.

[5] Raghavan Sindhu, Mooney Raymond J. [2013] Online Inference-Rule Learning from Natural-Language Extractions. *In Proceedings of the 3rd Statistical Relational AI (StaRAI-13) workshop at AAAI*.

[6] Beltagy Islam, Mooney Raymond J. [2014] Efficient Markov Logic Inference for Natural Language Semantics. *In Proceedings of the Fourth International Workshop on Statistical Relational AI at AAAI (StarAI-2014), Quebec City, Canada*: 9−14.

[7] MacKenzie Calvin. [2014] Integrating Visual and Linguistic Information to Describe Properties of Objects. *Undergraduate Honors Thesis, Computer Science Department, University of Texas at Austin*.

[8] Pichotta Karl, Mooney Raymond J. [2014] Statistical Script Learning with Multi-Argument Events. *In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), Gothenburg, Sweden*: 220-229.

[9] Roller Stephen, Erk Katrin and Boleda Gemma. [2014] Inclusive yet Selective: Supervised Distributional Hypernym Detection. *In Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland*: 1025−1036.

[10] Chang Angel X. , Savva Manolis , Manning Christopher D. [2014] Interactive Learning of Spatial Knowledge for Text to 3D Scene Generation. *In Proceedings of the Association for Computational Linguistics (ACL, 2014), Workshop on Interactive Language Learning, Visualization, and Interfaces (ACL-ILLVI)*.

[11] Thomason Jesse, Venugopalan Subhashini, Guadarrama Sergio, Saenko Kate, Mooney Raymond. [2014] Integrating Language and Vision to Generate Natural Language Descriptions of Videos in the Wild. *In Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland*.

[12] Chang Angel X. , Manning Christopher D. [2014] TokensRegex: Defining cascaded regular expressions over tokens. *Department of Computer Science, Stanford University, CSTR 2014−02*.

[13] Green Spence , Marneffe Marie-Catherine de , Manning Christopher D. [2013] Parsing Models for Identifying Multiword Expressions. *In the Computational Linguistics journal* 39(1):195−227.

ABOUT AUTHORS

Prasenjit Mukherjee *received the M.Sc. degree in Information Technology In 2005 from Annamalai University, Tamil Nadu, India. In 2011, he received M. Tech in Information Technology at Karnataka State Open University (KSOU), Karnataka, India. Presently, a fulltime Research Scholar in Information Technology in the area of Natural Language Processing from National Institute of Technology (NIT), Durgapur, India under the Visvesvaraya PhD. Scheme.*

Dr. Baisakhi Chakraborty *received the PhD. degree in 2011 from National Institute of Technology, Durgapur, India in Computer Science and Engineering. Her research interest includes knowledge systems, knowledge engineering and management, database systems, data mining, natural language processing and software engineering. She has several research scholars under her guidance. She has more than 30 international publications. She has a decade of industrial and 14 years of academic experience.*