International Journal of Current Trends in Engineering & Research (IJCTER)
e-ISSN 2455–1392 Volume 2 Issue 4, April 2016 pp. 93 -101
Scientific Journal Impact Factor : 3.468
http://www.ijcter.com

Knowledge Base Structure Description Language for Text Summarization with 6-W Information Network

Piyush Pratap Singh , Dr. Jayashri Vajpai, Prof. Dr Yogesh Sharma

[1]*Research Scholar, Faculty of Computer Application Jodhpur National University ,Jodhpur*
[2]*Professor, M.B.B Engineering College, Jodhpur*
[3]*Professor and Dean, Faculty of Computer Application, Jodhpur National University ,Jodhpur*

Abstract—Knowledge based structures definition language is the advance application of Computational Linguistics where different structures and techniques are applied for Natural Language Processing. This paper is in continuation of Implementation of 6-W based Structure Description Language for Text Summarization using VB.NET where text summarization and information retrieval system was implemented using a 6-W based structures definition language. In which a Knowledge base is used with respect to 6-W where the W's are who, what, when, whom, why, where. It was seen that all the time W's aren't a single attribute and must de described. This paper updates the details of the 6-W's making an information network. This structure can represent W as single attribute or the detailed entity.

Key Words—NLP, CL AI, PNL, SDL, KB, KBSDL

I. INTRODUCTION

In advanced Computational Linguistics (CL) focus of research is on Knowledge based structures definition language (KBSDL) and its application. KBSDL is a combination of Knowledge base (KB) which is pre-stored information and Structures Definition Language is the system in which the technique of Natural Language Processing is structured. The most advanced areas of research in KBSDL are

1. Development of a improved structure of KBSDL : There can be many type of SDL but efforts are made to improve them to achieve maximum efficiency

2. Precisiation of Natural Language Text (NLT): Precisiation or Text summarisation is a important task of the KBSDL

3. Global Translation of Natural Language Text: Translation is the most commercial and necessary application, in which different languages of the world are processed for translation.

4. Concurrent Question-Answering (QA): In addition of simple translation of world language QA system can be used for different languages query system, and QA for NLT is also promising application are of KBSDL.

5. Image processing using 9-Square (9-S):

   a. Representation of Images
   b. Knowledge extraction and recognition of images using 9-S technique where S1 is the central editing of images focus S2..........S9 are neighbourhood squares

S9	S2	S3
S8	S1	S4
S7	S6	S5

6. Transformation of poetry to prose.
7. Transformation of prose to poetry points 6 and 7 and the most complicated applications.

This paper improves the SDL structure implemented by Vajpai et al []

In the former paper implements 6-W based Structure Description Language for Text Summarization using VB.NET. The six W's are who, what, whom, when, where and why. This concept will be elaborated in section III to develop a Knowledge Base Structural Description Language (KBSDL). The paper also includes the state of art as understood from the study of literature. The details of implementation of KBSDL and 6-W information network are presented in section IV, followed by an illustrative example and conclusion.

II. STATE OF ART

As discussed in section I KBSDL is most advanced research area of Computational Linguistics Zadeh [] proposed the concept of fuzzy or generalized constraints and integrated it with the PNL to develop Generalized Constraints Language (GCL). The concept of protoform analysis was used to develop the structure of PNL which was pioneer step for developing KBSDL but was not implemented in its true form.

Thint et al [] have implemented the SDL in question answering system but it was more a rule based than a structure based system

Galley et al [] have designed an accurate Non-Hierarchical Phrase-Based Translation system that overcomes the principal weakness of conventional, i.e., non-hierarchical phrase-based statistical machine translation methodology that can handle only continuous phrases which gives a hint of SDL but Galley himself was not aware of implementing it using SDL.

Research in KBSDL is still in its infancy phase so not much is done in NLP using SDL even Google Translate, Microsoft Word use a pattern matching technique.

Vajpai et al [] have proposed a 6 W's based SDL concept that overcomes many drawbacks of the above approaches. This classifies the entire NL text in six categories, facilitating the summarization of text and uses the KBSDL. The implementation and further details of this approach are discussed in the next section.

III. KNOWLEDGE BASED STRUCTURAL DESCRIPTION LANGUAGE USING 6-W

Rudyard Kipling, an indefatigable globe trotter, discovered that all the knowledge of global activities can be described by a small group of information carriers. The six principal elements of this group are [2]

Who (W1) – The actors of activities
What (W2)- The actions carried out by the actors
Whom (W3) -The object to which the activities are directed
When (W4)- The time of activities
Where (W5)- The place of activities
Why (W6) - The reasons which prompted the occurrence of activities

But the W's are not always a single attribute they are combination of attributes which form an entity of respective W as follows

Who (W1): Actor
W1a – Gender (Male/female)
W1b- Name
W1c- Age

What (W2): Action
W2a – type of activity
W2b –Severity of activity

When (W3): Time
W3a- Date
W3b- Time
W3c- Season
W3d-session

Where (W4): Place
W4a – locality
W4b- City
W4c- Country

Whom (W5): Object of action
W5a – Gender
W5b –Name
W5c- Age
Why (W6): Reason
W6a –Actor
W6b –Action
W6c-Time
W6d –place
W6e- object of action

This paper discuss the technique of information retrieval from the natural language sentence

IV. DEVELOPMENT OF 6-W BASED KNOWLEDGE BASE STRUCTURE DESCRIPTION LANGUAGE

As discussed KBSDL is a combination of Knowledge base (KB) which is pre-stored information and Structures Definition Language. Components and structure of the proposed KBSDL is discussed below

Main title	Sub Title				
Who (W1)	W1a	W1b	W1c		

What (W2)	W2a	W2b			
When (W3)	W3a	W3b			
Where (W4)	W4a	W4b	W4c		
Whom (W5)	W5a	W5b	W5c		
Why (W6)	W6a	W6b	W6c	W6d	W6e

Fig 1 represents the detail W structure of the KB which is linked to the main KB this structure stores the details of the W's ,it is always not necessary that the W's have any details but if present it should me mentioned in the detail table by the developer and while retrieving the information the user automatically get the desired details of corresponding W's.

Sentence description

Sentence No.	W1					W2					W3					W4					W5					W6				
	a	b	b	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e

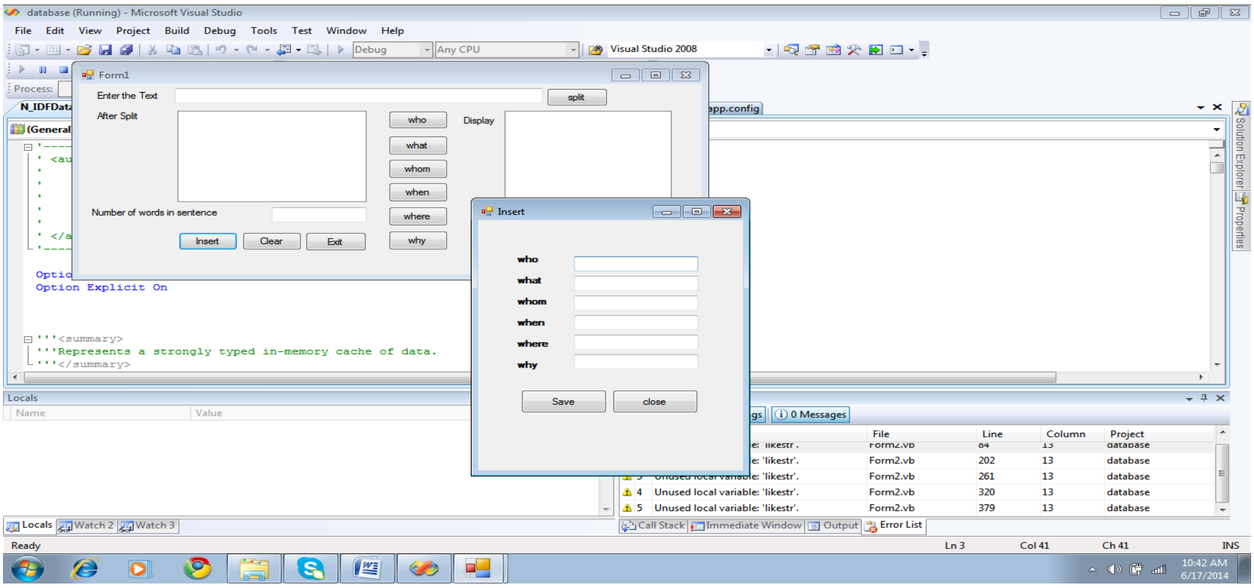
Fig 2 represent the structure of the sentence which is used for information extraction.

As shown in fig 1 And 2 the KBSDL uses two tables one is the structure and second is for details

V. IMPLEMENTATION OF THE PROPOSED METHOD

The user can pre-store initial knowledge in the SDL format in the KB as per requirement, by using the insertion window as shown in Fig 3.The insertion button is used to directly add new 6W attributes, which helps in continuously updating the Knowledge base, thus progressively enhancing its content, wherever required.

The success of the software also depends on this step, i.e. the bigger the knowledge base better the efficiency of the software. Three buttons have been created in the insertion window of the software for ease of operation. The ‘Clear’ button- clears the entire window and resets the process, the ‘Exit’ button- exits the software and the ‘Close’ button closes the insertion window(Fig3).



```
Option Explicit On

'''<summary>
'''Represents a strongly typed in-memory cache of data.
'''</summary>
```

	File	Line	Column	Project
e: 'likestr'.	Form2.vb	84	13	database
e: 'likestr'.	Form2.vb	202	13	database
3 Unused local variable: 'likestr'.	Form2.vb	261	13	database
4 Unused local variable: 'likestr'.	Form2.vb	320	13	database
5 Unused local variable: 'likestr'.	Form2.vb	379	13	database

Fig 3. Insertion Window

Major Steps in the implementation of the proposed method described in section IV are as follows:

1. The initial window of the software asks the user to enter a sentence in the text box.
2. After entering the sentence, the user clicks on the split button .
3. The software then splits the text into words by searching for ‘.’ & ‘blank’ character and arranges the sentence into different words.
4. The software displays the splitted words in a display box for further search and also displays the total count of words.
5. Pattern matching is carried out for these splitted words by matching them with the pre-stored classified words or group of words in the knowledge base with respect to 6-W. Semantic extraction is then applied on these matched words in accordance with SDL comprising of the six W’s organized as follows:

Who (W1): Actor
W1a – Gender (Male/female)
W1b- Name
W1c- Age

What (W2): Action
W2a – type of activity
W2b –Severity of activity

When (W3): Time

W3a- Date
W3b- Time
W3c- Season
W3d-session

Where (W4): Place
W4a – locality
W4b- City
W4c- Country

Whom (W5): Object of action
W5a – Gender
W5b –Name
W5c- Age
Why (W6): Reason
W6a –Actor
W6b –Action
W6c-Time
W6d –place
W6e- object of action

This is the most important step which classifies the entire sentence in the 6 fields.

6. The software extracts the words or group with their details to the respective W attributes and displays them in related fields of the result window.

The SDL representation hence obtained comprises of a details with the original sentence. This leads to effective text summarization using 6-W based structure description language.

VI. ILLUSTRATIVE EXAMPLE

In order to understand the proposed methodology consider the following natural language sentence

> **"Zadeh proposed the concept of PNL to his students in 2004 at UC Berkeley for better processing of Natural Language".**

Applying the methodology described in section V, the step by step implementation is elaborated as follows:

1 The sentence mentioned above is entered by user in initial window as shown in Fig 4.

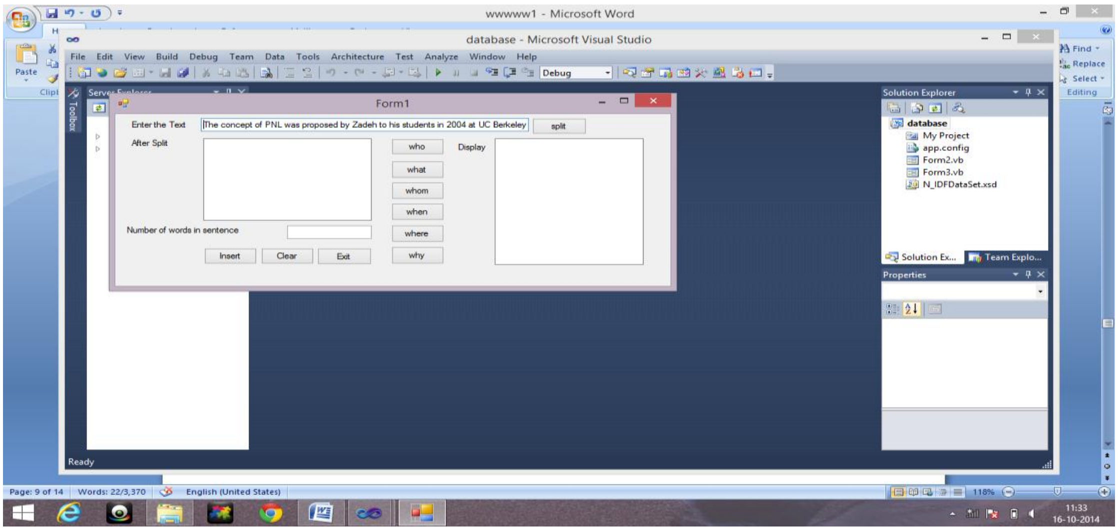
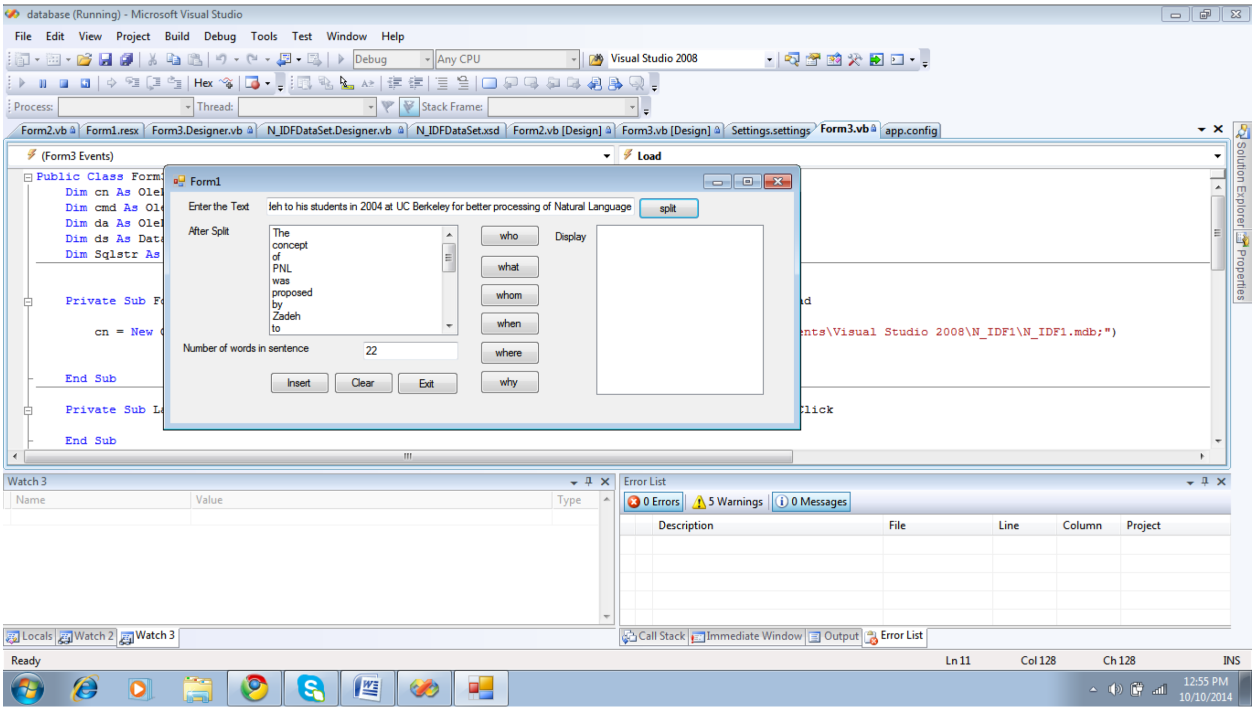

Enter the Text: The concept of PNL was proposed by Zadeh to his students in 2004 at UC Berkeley

Fig 4 Initial Window

2. When the user clicks on split button, the software reads the entire statement character by character.

3. The ‘blank’ and ‘full stop’ characters are identified by the software to split the sentence into different words.

4.The splitted words are displayed in the splitting window as shown in Fig 5.


database (Running) - Microsoft Visual Studio
File Edit View Project Build Debug Tools Test Window Help
Debug
Any CPU
Visual Studio 2008
Form2.vb Form1.resx Form3.Designer.vb N_IDFDataSet.Designer.vb N_IDFDataSet.xsd Form2.vb [Design] Form3.vb [Design] Settings.settings Form3.vb app.config
(Form3 Events)
Load
Form1
split
who
what
whom
when
where
why
Display
Insert
Clear
Exit
Watch 3
Error List
0 Errors
5 Warnings
0 Messages
Ready

```
Public Class Form
    Dim cn As Ole
    Dim cmd As Ol
    Dim da As Ole
    Dim ds As Dat
    Dim Sqlstr As

    Private Sub F                                   d

        cn = New                                    nts\Visual Studio 2008\N_IDF1\N_IDF1.mdb;")

    End Sub

    Private Sub L                                   lick

    End Sub
```

Enter the Text: leh to his students in 2004 at UC Berkeley for better processing of Natural Language

After Split:
The
concept
of
PNL
was
proposed
by
Zadeh
to

Number of words in sentence: 22

Fig 5 Splitting Window

5. Pattern matching and semantic extraction is then applied to the splitted words by comparing them with the pre-stored words and phrases in the SDL form in the knowledge base. The SDL is useful for semantic extraction of the 6-W’s in their respective fields using KB as follows.

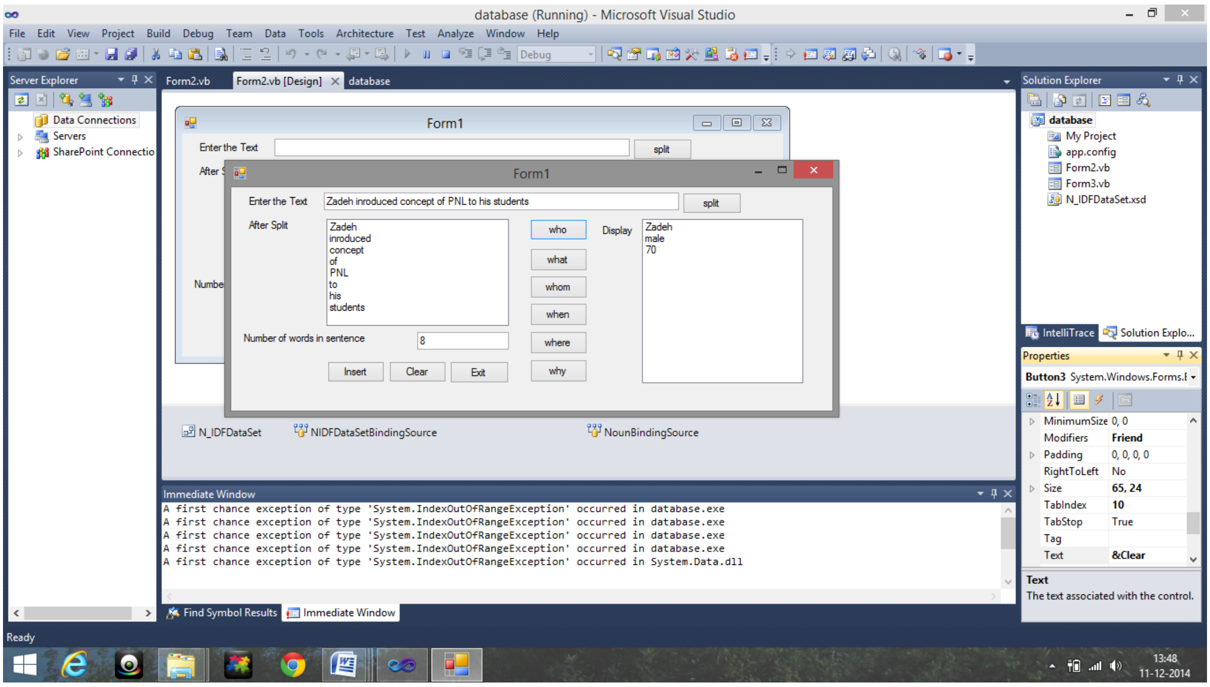
W1 – Zadeh
W2 – Concept of PNL
W3- Students
W4- 2004
W5 –UC Berkeley
W6 – For better processing of Natural Language

Finally, the desired SDL of the sentence is obtained as shown in Table 1

Table 1. 6-W based description of the sentence

Who phrase	**What phrase**	**Whom phrase**	**When phrase**	**Where phrase**	**Why phrase**
actor	***action***	***action aider***	***Time***	***Place***	***Reason of action***
W1	**W2**	**W3**	**W4**	**W5**	**W6**
Zadeh	Concept of PNL	Students	2004	UC Berkeley	For better processing of Natural Language

It can be observed that from the initial sentence containing 22 words the SDL representation has been derived with only 14 words thus text summarization has taken place
6. Now consider the query: "Who proposed PNL?" By clicking the Who button on the result window, the answer to this query can readily be obtained from W1, corresponding to Who in the SDL as shown in Fig 6.



Enter the Text: Zadeh inroduced concept of PNL to his students

After Split:
Zadeh
inroduced
concept
of
PNL
to
his
students

Number of words in sentence: 8

Display:
Zadeh
male
70

Button3 System.Windows.Forms.I

MinimumSize	0, 0
Modifiers	Friend
Padding	0, 0, 0, 0
RightToLeft	No
Size	65, 24
TabIndex	10
TabStop	True
Tag	
Text	&Clear

Text
The text associated with the control.

```
A first chance exception of type 'System.IndexOutOfRangeException' occurred in database.exe
A first chance exception of type 'System.IndexOutOfRangeException' occurred in database.exe
A first chance exception of type 'System.IndexOutOfRangeException' occurred in database.exe
A first chance exception of type 'System.IndexOutOfRangeException' occurred in database.exe
A first chance exception of type 'System.IndexOutOfRangeException' occurred in System.Data.dll
```

Fig 6 Result Window

The sentence is updated in the knowledge base in the SDL format. This format of sentence provides quick answer to any query about the activity described in the sentence.

VII.   CONCLUSION

The proposed SDL based software providesVB.NET based implementation of the concept of 6 W's based text precisiation technique. It classifies the natural language into six attributes for classification of language components into primary elements of information, resulting in more efficient representation, searching and question answering. This software and technique can be used for text summarization, information retrieval and automated question answering. Unlike other prevalent techniques, this theory does not use key word or phrase for searching. Work is in progress to further enhance the capability of this software by incorporating a semantic net based detailed structure for further elaboration of the information of the individual W's.

REFERENCES

[1] Y. Ledeneva, "Recent Advances in Computational linguistics", Mexico Informatica 34 2010
[2] K.R. Chowdhary, J. Vajpai, V.S. Bansal, "Natural language text compression using 5-W based presiciation structure" ,ECTN-12, 24 25 March 2012.
[3] J. Vajpai, V.S. Bansal, P.P Singh, "Computer assisted multilingual translation for global communication", National Conference on "New Advances in Programming languages and their implementations", March 15-16, 2013
[4] "Microsoft Word grammar checker" http://office.microsoft.com/en-in/word-help/check-spelling-and-grammar-P010117963.aspx 2011
[5] "Google Translate" http://translate.google.com. 2013
[6] "Apple siri software for iSO phones" http://www.zdnet.com/ applesiri 2013.

[7] L. A. Zadeh, "From computing with numbers to computing with words from manipulation of measurements to manipulation of perceptions", International Journal for Applied Math & Computer Science., Vol. 12/3: pp. 307-324, 2001.

[8] L. A. Zadeh, "A new direction in AI - toward a computational theory of perceptions", A.I. Magazine, Spring 2001.

[9] L. A. Zadeh, "Precisiated natural language," pp. 74-91, AI Magazine, 25(3), 2004.

[10] L. A. Zadeh, "Toward a generalized theory of uncertainty (GTU)" -an outline in Information Sciences, 172, 2005, pp. 1-40.

[11] L. A. Zadeh "From Search Engines to Question-Answering Systems: The Problems of World knowledge Relevance, Deduction, and Precisiation" ∗ http://www.springer.com/978-3-540-34780-4 2006

[12] M. Thint, M S Beg, Z. Qin, "PNL-enhanced Restricted Domain Question Answering System," IEEE International Conference on Fuzzy Systems, London, UK, July 2007.

[13] M Thint, M. S. Beg, Z. Qin, M. "Deduction Engine Design for PNL-based Question Answering System," World Congress of the International Fuzzy Systems Association , 2007.

[14] J. shafi ,A. ali "Defining relations in precisiation of natural language processing for semantic web" IJCSE ISSN:0975-3397 Vol .4 no 05 May 2012.pp 72

[15] Y. Ledeneva, "Effect of Preprocessing on Extractive Summarization with Maximal Frequent Sequences", MICAI-08, LNAI 5317,pp. 123-132, Mexico, Springer-Verlag, ISSN 0302-9743, 2008.

[16] Y. Ledeneva, A. Gelbukh, R. G. Hernández, " Terms Derived from Frequent Sequences for Extractive Text Summarization", CICLing-08, LNCS 4919, pp 593-604, Israel, Springer-Verlag, ISSN 0302-9743, 2008.

[17] Y. Ledeneva, A. Gelbukh, R. G. Hernández, "Keeping Maximal Frequent Sequences Facilitates Extractive Summarization", In: G. Sidorov *et al* (Eds). CORE-2008, Research in Computing Science, vol. 34, pp.163-174, ISSN 1870-4069, 2008.

[18] A. Gelbukh and G. Sidorov "Approach to construction of automatic morphological analysis systems for inflective languages with little effort", Lecture Notes in Computer Science, N 2588, 2003, ISSN 0302-9743, Springer-Verlag, pp. 215–220,2003

[19] A Gelbukh ,G Sidorov., S Y Han, "Evolutionary Approach to Natural Language Word Sense Disambiguation through Global Coherence Optimization", WSEAS Transactions on Communications, ISSN 1109-2742, Issue 1 Vol. 2, pp. 11–19, 2003.

[20] A Gelbukh., I Bolshakov, "Internet, a true friend of translator", International Journal of Translation, ISSN 0970-9819, Vol. 15, No. 2, pp. 31–50, 2003.

[21] "Usage of windows operating system in public domain" http://wikipedia.org/wiki/Usage_share_of_operating_systems 2013

[22] K. Toutanova, D. Klein, C. Manning, and Y. Singer, "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," in Proc. of HLT-NAACL 2003, pp. 252-259.2003

[23] Y Zhang, S Clark. "Shift-reduce ccg parsing. In Proceedings of ACL" 2011

[24] Y Goldberg ,K Zhao, L Huang . "Efficient *Implementation* of Beam-Search Incremental Parsers" Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 628–633, Sofia, Bulgaria, 2013