Generation Rules to Deconvert UNL Expressions to Bangla Sentences

Mohammad Zakir Hossain Sarker, Md Nawab Yousuf Ali, Jugal Krishna Das

Abstract— This paper discusses the design and implementation of Generation Rules for deconverting Universal Networking Language or UNL expressions to the Bangla Language. The Universal Networking Language (UNL) facilitates translation between Natural Languages across the world. The paper also focuses on the linguistic aspects of Bangla required for the deconversion process. We have developed a set of generation rules for converting UNL expression to Bangla sentences. Our experiment shows that these rules successfully generate correct Bangla sentences from UNL expressions. These rules can currently produce simple Bangla sentences. It is being enhanced to handle more types of sentences. The present system is a small achievement towards attaining a complete translation facility for the Bangla language.

Index Terms— Generation Rules, UNL, UNL Expression, Deconverter, Converter, Generation Window, Condition Window

1 Introduction

The UNL system allows people to communicate with peoples of different languages in their mother tongue. The UNL is a common language to exchange information through computers which can deal with natural languages. The UNL system basically consists of language servers, UNL editors and UNL viewers [1].

A conversion system from native languages into UNL is called "enconverter", and the one that deconverts from UNL into native languages is called "deconverter". Information "enconverted," from any language is exchanged in UNL format via networks. Information represented in UNL is "deconverted" into each native language on the terminal network [1].

In this paper, we have emphasized on “Deconverter” especially on the Generation/Deconversion rules that are required to deconvert UNL expressions to Bangla sentences.

2 How Deconverter Works

The DeConverter is a language independent generator, which provides a framework for syntactic and morphological generation synchronously. It can convert UNL Expressions into a variety of natural languages using respective word dictionaries and sets of grammar rules of deconversion of the languages. A word dictionary contains the information of words that correspond to UWs included in the input of UNL Expressions and grammatical attributes (features) that describe the behaviors of the words. Deconversion/Generation rules (grammar rules of deconversion) describe how to construct a sentence using the information from the input of UNL Expressions and defined in a word dictionary. The DeConverter converts UNL Expressions into sentences of a target language following the descriptions of Generation rules.

Co-occurrence relation-based word selection for natural collocation can also be carried out synchronously. For this purpose, a co-occurrence dictionary of the target language is necessary. The UNL Ontology is also helpful when no correspondent word for a particular UW exist in a language. In this case, the DeConverter consults to the UNL Ontology to try to find a more general (upper) UW of which a correspondent word exists in its word dictionary and use the word of the upper UW to generate the target sentence instead.

The DeConverter works in the following way. It first transforms the input of a UNL expression – a set of binary relations - into a directed graph structure with hyper-nodes called node-net. The root node of a node-net is called entry node and represents the head (e.g. the main verb) of a sentence. Deconversion of a UNL Expression is carried out by applying Generation Rules to the nodes of node-net. It starts from the entry node, to find an appropriate word for each node and generate a word sequence (a list of words in grammatical order) of a target language. In this process, the syntactic structure is determined by applying syntactic rules, and morphemes are similarly generated by applying morphological rules. The deconversion process ends when all words for all nodes are found and a word sequence of target sentence is completed.
Fig 1. shows the structure of the DeConveter. “G” indicates generation windows, and “C” indicates condition windows of the DeConverter. The DeConverter operates on the node-list through generation windows. Condition windows are used to check conditions when applying a rule. In the initial stage, in opposite to the EnConveter, the entry node of a UNL Expression exists in the node-list. At the end of deconversion, the node-list is the list of all morphemes, with each as a node, that are converted from the node-net and constitute the target sentence.

- *Author Mohammad Zakir Hossain Sarker, is currently pursuing PhD degree from Jahangirnagar University, Bangladesh, PH-+8801732233300. E-mail: zakir.publications@gmail.com*
- *Co-Author Md. Nawab Yousuf Aki is currently working as the Chairman, CSE Department of East West University, Bangladesh. PH-+8801xxxxxxxxx. E-mail: nawab@ewubd.com*
- *Co-Author Jugal Krishna Das is a professor of CSE Department, Jahangirnagar University, Bangladesh. E-mail: drdas64@gmail.com*

International Journal of Scientific & Engineering Research Volume 4, Issue 12, December-2013
ISSN 2229-5518

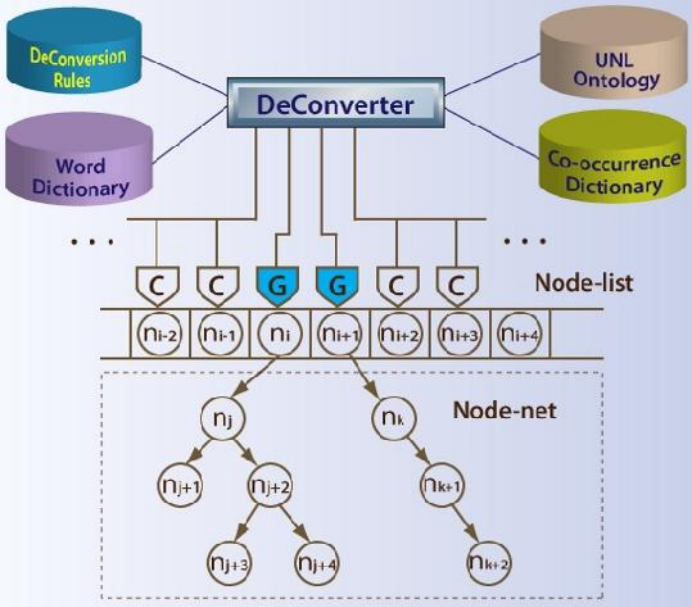


Fig 1: Deconversion process

3 Literature Review

For deconverting UNL expressions to Bangla sentence firstly, we have gone through Universal Networking Language (UNL) [2, 3, 4, 5, 6] where we have learnt about UNL expression, Relations, Attributes, Universal Words, UNL Knowledge Base, Knowledge Representation in UNL, Logical Expression in UNL, UNL systems and specifications of Enconverter. All these are key factors for preparing Bangla word dictionary, deconversion rules in order to deconvert UNL expressions to a natural language sentence (here Bangla sentence). Secondly, we have rigorously gone through the Bangla grammar [7, 8, 9, 10], Verb and roots (Vowel and Consonant Ended), Morphological Analysis [11, 12, 13, 14, 19], suffixes [15, 16, 17, 18], construction of Bangla sentence based on semantic structure. Using above references we extort ideas about Bangla grammar for morphological and semantic analysis in order to prepare Bangla word dictionary (for verb root, verbal inflections, etc) in the format of UNL provided by the UNL center of the UNDL Foundation.

4 Proposed Generation rules

In this section, we have proposed the formats of some generation rules that are to be used for converting UNL expressions to Bangla sentence.

Format 1: Format of rules to insert subjective pronouns of not alternative roots for agent (agt) relation

:"HPRON,(x)P,SUBJ::agt:"{ROOT,VEND,^ALT,#AGT,^(x)p:(x)p::}P9;

where, grammatical attributes HPRON denotes human pronoun, SUBJ for subject of a sentence, agt for agent relation, ROOT for verb root, VEND for vowel ended root, ^ALT for not alternative root, #AGT indicates that the corresponding root involves with agent relation. P indicates person and when the value of x=1, 2 and 3 then 1P denotes first person, 2P for second person and 3P for third person respectively and p is the temporary attribute for person to prevent recursive operations. Some examples of rules for not alternative roots are as follows:

:"HPRON,1P,SUBJ::agt:"{ROOT,VEND,^ALT,#AGT,^1p:1p::}P9;

:"HPRON,2P,SUBJ::agt:"{ROOT,VEND,^ALT,#AGT,^2p:2p::}P9;

:"HPRON,2P,SUBJ,@respect,^@contempt,HON,^NGL::agt:"{ROOT,VEND,^ALT,#AGT,^2p:2p::}P9;

:"HPRON,2P,SUBJ,^@respect,@contempt,^HON,NGL::agt:"{ROOT,VEND,^ALT,#AGT,^2p:2p::}P9;

:"HPRON,3P,SUBJ::agt:"{ROOT,VEND,^ALT,#AGT,^3p:3p::}P9;

Format 2: Format of rules to insert subjective pronouns alternative roots for agent (agt) relation

:"HPRON,(x)P,SUBJ, [^] @respect, | [^]@contempt, | [^]HON,[^]NGL::agt:"{ROOT,VEND,[^]@present|@progress|@complete,VEG(y),ALT,#AGT,^(x)p:(x)p::}P10;

where, UNL attributes @respect, @contempt, @present, @past, @future, @progress, and @complete denotes for respected person, neglected person, present tense, past tense, future tense, continuous tense and perfect tense respectively. ALT indicates alternative roots and when the value of y=1 then VEG1 is to be used for vowel ended group 1 and when y=2 then VEG2 for vowel ended group 2 and so on. '^' expresses negation of the following value. Some examples of rules for not alternative roots are as follows:

:"HPRON,1P,SUBJ::agt:"{ROOT,VEND,@present,^@progress,@complete,VEG1,ALT,#AGT,^1p:1p::}P10;

:"HPRON,1P,SUBJ::agt:"{ROOT,VEND,@past,^@progress,^@complete,VEG1,ALT,#AGT,^1p:1p::}P10;

:"HPRON,2P,SUBJ,^@respect,^@contempt,^HON,^NGL,^PL::agt:"{ROOT,VEND,@present,^@progress,^@complete,VEG4,ALT,#AGT,^2p:2p::}P10;

:"HPRON,2P,SUBJ,@respect,HON::agt:"{ROOT,VEND,@present,^@progress,^@complete,VEG4,ALT,#AGT,^2p,^hon:2p,hon::}P10;

:"HPRON,2P,SUBJ,@contempt,NGL::agt:"{ROOT,VEND,@present,^@progress,^@complete,VEG4,ALT,#AGT,^2p,^ngl:2p,ngl::}P10;

:"HPRON,3P,SUBJ::agt:"{ROOT,VEND,@present,^@progress,^@complete,VEG2,ALT,#AGT,^3p:3p::}P10;

:"HPRON,3P,SUBJ::agt:"{ROOT,VEND,@past,^@progress,^@complete,VEG1,ALT,#AGT,^3p:3p::}P10;

Format 3: Format of rules to insert subjective pronouns of not alternative roots for thing with attribute (aoj) relation

:"HPRON,(x)P,SUBJ::aoj:"{ROOT,VEND,^ALT,#AOJ,^(x)p:(x)p::}P9;

Examples:

:"HPRON,1P,SUBJ::aoj:"{ROOT,VEND,^ALT,#AOJ,^1p:1p::}P9;

:"HPRON,2P,SUBJ::aoj:"{ROOT,VEND,^ALT,#AOJ,^2p:2p::}P9;

:"HPRON,3P,SUBJ::aoj:"{ROOT,VEND,^ALT,#AOJ,^3p:3p::}P

International Journal of Scientific & Engineering Research Volume 4, Issue 12, December-2013
ISSN 2229-5518

9;

Format 4: Format of rules to insert subjective pronouns alternative roots for agent (agt) relation
:"HPRON,(x)P,SUBJ, [^] @respect, [^]@contempt,[^]HON,[^]NGL::aoj:"
{ROOT,VEND,[^]@present,@progress,@complete,VEG(y),ALT,#AOJ,^(x)p:(x)p::}P10;
Examples:
:"HPRON,1P,SUBJ::aoj:"{ROOT,VEND,@present,^@progress,@complete,VEG1,ALT,#AOJ,^1p:1p::}P10;
:"HPRON,1P,SUBJ::aoj:"{ROOT,VEND,@past,^@progress,^@complete,VEG1,ALT,#AOJ,^1p:1p::}P10;
:"HPRON,2P,SUBJ,^@respect,^@contempt,^HON,^NGL,^PL::aoj:"{ROOT,VEND,@present,^@progress,^@complete,VEG4,ALT,#AOJ,^2p:2p::}P10;
:"HPRON,2P,SUBJ,@respect,HON::aoj:"{ROOT,VEND,@present,^@progress,^@complete,VEG4,ALT,#AOJ,^2p,^hon:2p,hon::}P10;
:"HPRON,2P,SUBJ::aoj:"{ROOT,VEND,@past,^@progress,@complete,VEG1,ALT,#AOJ,^2p:2p::}P10;

Format 5: Format of rules for backtracking
Rule 5.1
?{:::}{PRON,SG,@pl:::}P8;
Rule 5.2
?{:::}{PRON,PL,^@pl:::}P8;
Rule 5.3
?{:::}{HPRON,SUBJ,HON,^@respect:::}P8;
Rule 5.4
?{:::}{HPRON,SUBJ,^HON,@respect:::}P8;
Rule 5.5
?{:::}{HPRON,SUBJ,NGL,^@contempt:::}P8;
Rule 5.6
?{:::}{HPRON,SUBJ,^NGL,@contempt:::}P8;

Format 6: Format of rules to insert verbal inflexions at the end of roots for first person
:{ROOT,VEND,(x)p,[^]@present,[^]@progress,[^]@complete,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,(x)P,PRS|PST|FUT,[^]PRGR,[^]CMPL:::"P10;
Examples:
:{ROOT,VEND,1p,@present,^@progress,^@complete,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,1P,PRS,^PRGR,^CMPL:::"P10;
:{ROOT,VEND,1p,@present,@progress,^@complete,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,1P,PRS,PRGR,^CMPL:::"P10;

Format 7: Format of rules to insert verbal inflexions at the end of roots for second person
:{ROOT,VEND,(x)p,[^]@present,[^]@progress,[^]@complete,[^]hon,[^]ngl,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,(x)P,PRS|PST|FUT,[^]PRGR,[^]CMPL,[^]HON,[^]NGL:::"P10;
Examples:
:{ROOT,VEND,2p,@present,^@progress,^@complete,^hon,^ngl,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,2P,PRS,^PRGR,^CMPL,^HON,^NGL:::"P10;
:{ROOT,VEND,2p,@present,^@progress,^@complete,^hon,ngl,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,2P,PRS,^PRGR,^CMPL,^HON,NGL:::"P10;

Format 8: Format of rules to insert verbal inflexions at the end of roots for third person
:{ROOT,VEND,(x)p,[^]@present,[^]@progress,[^]@complete,^hon,|^ngl,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,(x)P,PRS|PST|FUT,[^]PRGR,[^]CMPL,^HON,^NGL:::"P10;
Examples:
:{ROOT,VEND,3p,@present,^@progress,^@complete,^hon,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,3P,PRS,^PRGR,^CMPL,^HON:::"P10;
:{ROOT,VEND,3p,@present,@progress,^@complete,^ngl,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,3P,PRS,PRGR,^CMPL,^NGL:::"P10;
Format 9: Format of rules to insert noun before root
:"N,[^]@pl,^SUBJ:SUBJ:agt:"{ROOT,VEND,#AGT,^3p,[^]sg|pl:3p,sg|pl::}P10;
Examples:
:"N,^@pl,^SUBJ:SUBJ:agt:"{ROOT,VEND,#AGT,^3p,^sg:3p,sg::}P10;
:"N,@pl,^SUBJ:SUBJ:agt:"{ROOT,VEND,#AGT,^3p,^pl:3p,pl::}P10;

Format 10: Format of rules to insert article
Rule 10.1 (For singular)
:{N,^@pl,@def,^boch:boch::}"[[BIV]],BIV,BOCH,DEF,^PL:::"P10;

Rule 10.2 (For plural)
:{N,[^]HUMN,@pl,@def,^boch:boch::}"[[BIV]],BIV,BOCH,DEF,PL,[^]HUMN:::"P10;
Examples:
:{N,^HUMN,@pl,@def,^boch:boch::}"[[BIV]],BIV,BOCH,DEF,PL,^HUMN:::"P10;
:{N,HUMN,VEND,@pl,@def,^boch:boch::}"[[BIV]],BIV,BOCH,DEF,PL,HUMN,VEND:::"P10;

5 Deconversion From UNL Expression To Bangla Sentence

In this section, we have converted UNL expression into Bangla sentence. For this conversion process, we need Dictionary, which we have developed but not discussed in this paper and a set of Generation rules given below.

Rule 1: (Pronoun insertion)
:"HPRON,1P,SUBJ::agt:"{ROOT,VEND,^ALT,#AGT,^1p:1p::}P9
Rule 2: (Right Shift Rule)
R{:::}{SUBJ:::}
Rule 3: (Blank insertion Rule)
:{SUBJ,^blk:blk::}"[],BLK:::"P10;
Rule 4: (Right Shift Rule)
R{:::}{SUBJ:::}
Rule 5: (Noun Insertion Rule)
:"N,^obj:obj:obj:"{ROOT,VEND,^#AGT,#OBJ:::}P9;
Rule 6: (Right Shift Rule)
R{:::}{N:::}
Rule 7: (Blank Insertion Rule)

International Journal of Scientific & Engineering Research Volume 4, Issue 12, December-2013
ISSN 2229-5518

:{N,^blk:blk::}"[ ],BLK:::"P10;

Rule 8: (Right Shift Rule)

R{N:::}{BLK:::}

Rule 9: (Right Shift Rule)

R{BLK:::}{ROOT:::}

Rule 10: (Insertion of Verbal Inflexion)

:{ROOT,VEND,1p,#AGT,^@progress,^@complete,^kbiv:kbiv::}"[[KBIV]],KBIV,VEND,1P,PRS,^PRGR,^CMPL:::"P10;}

Rule 11: (Right Shift Rule)

R{V:::}{:::}

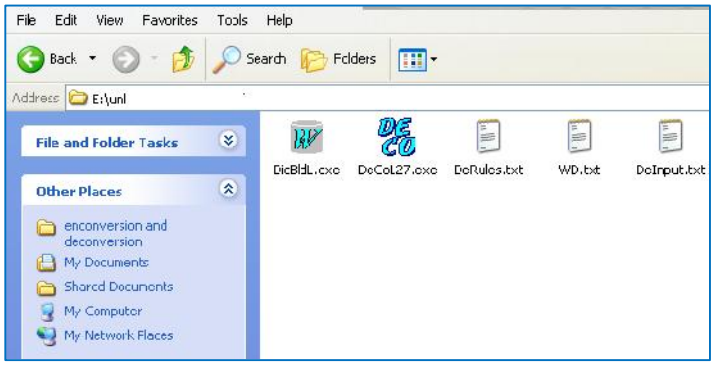
These generation rules will be applied to the nodes in the node-list for operation on them and/or inserting nodes from the Node-net into the Node-list.

Rule 1 is pronoun insertion rule that describes that when root "যা" (ja)is in the RGW, the pronoun "আমি" (aami) to be inserted in the LGW. Rule 2 is the right shift rule which is applied to shift the windows of DeConverter to right. The blank insertion rule (rule 3) is applied to insert a blank between pronoun "আমি" (aami) and root "যা" (ja). Again right shift rule (rule 4) is applied to shift the windows to right. Now, rule 5 to be used to insert noun "বাড়ি" (bari) in the RGW. And if noun "বাড়ি" (bari) is in the RGW then windows will be shifted to the right by applying rule 6. Now, rule 7 will be applied to insert a blank between noun "বাড়ি" (bari) and root "যা" (ja). Rule 8 and 9 both are right shift rules to be applied to shift the windows two steps to right. After that verbal inflexion "ই" (i) to be inserted into the node-list by applying rule 10 . Finally, rule 11 to be applied to place the sentence tail in the RGW to complete the conversion process to get the target sentence - আমি বাড়ি যাই

6 Testing the Genereation Rules

To convert the UNL expression into Bangla language sentence we have used following three files (created by us) and 2 more files (see Screen 1)

1. Input file (unl expressions)
2. Dictionary file (WD.txt)
3. Rules File (DeRules.txt), contains the Generation Rules, described in Section 5
4. Decoder (DeCoL27.exe)
5. A dictionary builder file (DicBldL.exe) provided by the UNDL Foundation of UNL center, which have been downloaded

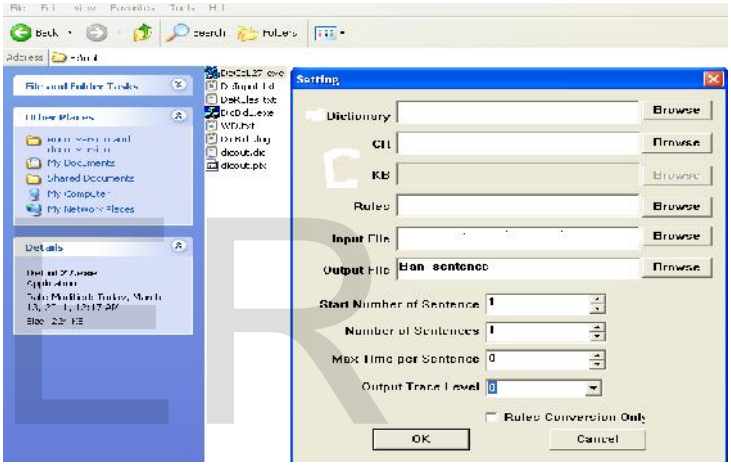


Screen 1: Shows the files for deconversoion of UNL expression to Bangla native sentence

Screen 2 shows the DeConverter (DeCoL27) that can convert UNL expressions to respective native language.

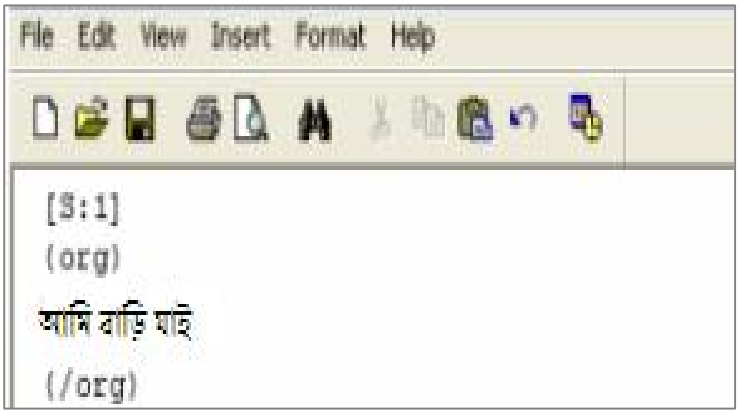
Screen 2: DeConverter file (DeCoL27.exe)

Screen 3 shows the way of selecting files to be used for deconversion



Screen 3: Files to be used for deconversion

Screen 4 shows the contents of output file i.e. the Bangla sentence, "আমি বাড়ি যাই"



Screen 4: Bangla sentence "আমি বাড়ি যাই" from given UNL expression

International Journal of Scientific & Engineering Research Volume 4, Issue 12, December-2013
ISSN 2229-5518

7 Conclusion

In this paper, we have developed few Generation Rules. These rules can currently produce simple Bangla sentences. It is being enhanced to handle more types of sentences. The present system is a small achievement towards attaining a complete translation facility for the Bangla language.

References

[1] http://www.unl.ru/system.html last accessed on October 7, 2013
[2] http://www.undl.org last accessed on October 10, 2013
[3] H. Uchida, M Zhu. T.G. Della Senta. *Universal Networking Language*, 2005/6-UNDL Foundation, International Environment House.
[4] H. Uchida, M. Zhu. *The Universal Networking Language (UNL) Specification Version 3.0 Edition 3* , Technical Report, UNU, Tokyo:, 2005/6-UNDL Foundation, International Environment House, 2004
[5] *EnConverter Specification*, Version 3.3, UNL Center/UNDL Foundation, Tokyo 150-8304, Japan 2002
[6] *DeConverter Specification*, Version 2.7, UNL Center, UNDL Foundation, Tokyo 150-8304, Japan 2002
[7] D.M. Shahidullah. *Bangla Baykaron*, Dhaka: Ahmed Mahmudul Haque of Mowla Brothers prokashani, 2003
[8] D. C. Shuniti Kumar. *Bhasha-Prakash Bangala Vyakaran*, Calcutta : Rupa and Company Prokashoni, July 1999, pp.170-175
[9] Humayun Azad. *Bakkotottyo - Second edition*, Dhaka: Bangla Academy Publishers, 1994
[10] D. S. Rameswar. *Shadharan Vasha Biggan and Bangla Vasha*, Pustok Biponi Prokashoni, November 1996, pp.358-377
[11] M.N.Y. Ali, J.K. Das, S. M. Abdullah Al Mamun, M. E. H. Choudhury. "Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language", *International Conference on Computer and Communication Engineering 2008 (ICCCE'08)*, Kuala Lumpur, Malaysia. pp. 726-731
[12] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A. M. Nurannabi. "Morphological Analysis of Bangla words for Universal Networking Language", *International Conference on Digital Information Management*, icdim, 2008, London, England, pp. 532-537
[13] M.N.Y.Ali, A. M. Nurannabi, G. F. Ahmed, J.K. Das. "Conversion of Bangla Sentence for Universal Networking Language", *International Conference on Computer and Information Technology (ICCIT)*, Dhaka, 2010 pp. 108-113
[14] M. Z. H. Sarker, M.N.Y. Ali, J.K. Das, Dictionary Entries for Bangla Consonant Ended Roots in Universal Networking Language, *International Journal of Computational Linguistics (IJCL)*, Volume (3) : Issue (1) : 2012 pp 79-87
[15] M. Z. H. Sarker, M.N.Y. Ali, J.K. Das, Outlining Bangla Word Dictionary for Universal Networking Language, *IOSR Journal of Computer Engineering (IOSRJCE)*, ISSN: 2278-0661 Sep-Oct. 2012
[16] M. Z. H. Sarker, M.N.Y. Ali, J.K. Das, Development of Dictionary Entries for the Bangla Vowel Ended Roots for Universal Networking Language, *International Journal of Computer Applications 52(19)*, August 2012. Published by Foundation of Computer Science, New York, USA. pp 38-45
[17] M.N.Y. Ali, M. Z. H. Sarker, J.K. Das, Analysis and Generation of Bengali Case Structure Constructs for Universal Networking Language, *International Journal of Computer Applications (IJCA)*, Volume 18, No. 7, March 2011, pp. 34-41
[18] M.N.Y. Ali, M. Z. H. Sarker, G. F. Ahmed, J.K. Das, Conversion of Bangla Sentence into Universal Networking Language Expression (IJCSI), *International Journal of Computer Science Issues*, Vol. 8, Issue 2, March 2011, pp. 64-73
[19] M.N.Y. Ali, M. Z. H. Sarker, G. F. Ahmed, J.K. Das, "Rules for Morphological Analysis of Bangla Verbs for Universal Networking Language", *International Conference on Asian Language Processing (IALP)*, Harbin, China, December 2010, pp. 31-34