Generation of UNL Attributes and Resolving Relations for Bangla EnConverter

Md. Nawab Yousuf Ali, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
Md. Sarwar Kamal, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
Md. Shamsujjoha, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
Mohammad Ameer Ali, Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
Ghulam Farooque Ahmed, Computer Village Ltd., Dhaka, Bangladesh

ABSTRACT

Conversion of Bangla language to another native language and another language to Bangla language using Universal Networking Language (UNL) is highly demanding due to rapidly increasing the usage of Internet-based applications. UNL has been used by various researchers as an inter-lingual approach for an Automated Machine Translation (AMT) scheme. This article presents a novel work on construction of EnConverter for Bangla language with a special focus on generation of UNL attributes and resolving relations of Bangla text. The architecture of Bangla EnConverter, algorithms for understanding the Bangla input sentence; resolution of UNL relations; and attributes for Bangla text/language are also explained in this article. This article highlights the analysis rules for EnConverter and indicates its usage in generation of UNL expressions. This article presents the results of implementation of Bangla EnConverter and compares these with the system available at Russian and English Language Server.

KEYWORDS

Analysis Rule, Bangla Language Text, EnConverter, Machine Translation, Morphological Analysis, Universal Networking Language (UNL), Universal Words

DOI: 10.4018/IJALR.2017070102

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
1. INTRODUCTION

Internet has evolved from monolingual into a multilingual World Wide Web for more than 1000 languages shown in (Ali et al., 2012). The Internet usage by the vast number of users now varies from primarily academic purposes to widespread commercial, leisure, education, entertainment etc. purposes. In Internet, vast knowledge and information resources in different languages are scattered all over the world and remain mostly inaccessible due to non-machine representation and language barrier in information distribution as shown in (Ali et al., 2008). It has been showed that there is a great need to translate web pages and electronic mail messages into the native language for overcoming the language barrier (Kumar et al., 2011). Considering this issue, Universal Networking Language (UNL) has been developed by the United Nations University/Institute of Advanced Studies (UNU/IAS) to convert any language to other language presented in (Ali et al., 2013). The UNL system is introduced most of the languages like Arabic, Spanish, Japanese, Chinese, Russian and many more as shown in (Choudhury et al., 2005). A UNL based transfer system needs two components: one for conversion from source language to UNL expression and another for UNL expression to target language. The UNL system has 46 semantic relations and 87 attributes to express the semantic meaning of a sentence of a source language demonstrated by (Uchida et al., 1993). With this increasing pressure of providing information access without language and cultural barriers, Internet today has to deal with immense complexity of multilinguality.

Bangla is the 4th widely spoken language in the world with more than 250 million speakers, most of whom live in Bangladesh and the Indian state of West Bengal which is evidenced (Ali et al. 2008). Bangla text is converted to UNL expressions by the EnConverter of Bangla language. This UNL expression is converted to any other language that has its own DeConverter. This will indeed help to develop a multilingual machine translation system for Bangla language.

A statistical Bangla to English translation engine using only simple Bangla sentences that contain a subject, an object and a verb is developed by (Uddin et al., 2004). A low-cost English to Bangla (E2B)-ANUBAD translating, English text into Bangla text using both rule-based and transformation-based MT schemes along with three-level of parsing is shown in (Saha, 2005). MT Bangla dictionaries that address the organization, contents and details of the information have been developed by (Ali et al., 2002). The research has been done for morphological analysis of Bangla words for UNL by (Choudhury et al., 2005), parsing methodology for Bangla sentences by (Asaduzzaman et al., 2003), and dictionary development of Bangla words by (Ali et al., 2002; Islam, 2009). The suffix, prefix and inflexions of Bangla language are detailed in (Asaduzzaman et al., 2008; Khairunnahar, 2008).

A number of researches have been done in Bangla for UNL based machine translation scheme (Uddin et al., 2004; Khairunnahar, 2008). But so far no attempt has been made for the development of Bangla EnConverte and generation of UNL attributes and resolving the relation of Bangla EnConverter. This motivates us to develop a Bangla EnConverter for UNL expressions for resolving the above issues. A rigorous study on Bangla language grammar presented in (Uddin, 2004; Saha, 2005; Ali et al., 2002; Shahidullah, 2003; Shuniti, 1999; Remeswar 1996; Azad 1994) verb and roots (vowel ended and consonant ended) in (Ali et al., 2017) and morphological
Related Content

Semi-E-Preinvex Functions
[www.igi-global.com/article/semi-preinvex-functions/46026?camid=4v1a](www.igi-global.com/article/semi-preinvex-functions/46026?camid=4v1a)

Heterogeneous Learning Using Genetic Algorithms
[www.igi-global.com/chapter/heterogeneous-learning-using-genetic-algorithms/21133?camid=4v1a](www.igi-global.com/chapter/heterogeneous-learning-using-genetic-algorithms/21133?camid=4v1a)
Reducing Interface Mutation Costs with Multiobjective Optimization Algorithms
www.igi-global.com/article/reducing-interface-mutation-costs-multiobjective/76375?camid=4v1a

Diversity Conserved Chaotic Artificial Bee Colony Algorithm based Brightness Preserved Histogram Equalization and Contrast Stretching Method