Abstract

Named entities (NE) and Out-Of-Vocabulary (OOV) words in text are treated differently for application like Machine Translation and Information Retrieval. While normal text in the Source Language (SL) is translated on the basis of translation mapping, named entities and out-of-vocabulary are written in the script of the Target Language (TL) without any change to the articulation of the word in both the languages. The process is transliteration. This paper shows the exploitation of Phoneme-Based model for transliteration of English to Kashmiri that can be used in information extraction and machine translation for the language pair. The overall accuracy of the system achieved while tested on medical domain and Wikipedia based English text is 86% (intelligible transliteration).

References

2. CMU. The CMU Pronunciation Dictionary. www.speech.cs.cmu.edu/cgi-bin/cmudict.
   School of Computer Science, Carnegie Mellon University, Pittsburgh, USA, 2006.
   Approach. Forum for Information Retrieval Evaluation, Delhi, India.
   and Contextual Rules. In proceedings of the 19th International Conference on Computational
   Linguistics, Taipei, Taiwan, 758-764.
   521-526.
   International Journal of Information Technology and Knowledge Management, Vol. 4(No. 2),
   459-463.
   on Language & Technology, NUC and ES, Lahore, Pakistan.

Index Terms

Computer Science Pattern Recognition

Keywords

Machine Transliteration, Phoneme-Based Machine Transliteration, English-Kashmiri Machine
Transliteration.