Abstract

Differently written words may have similar pronunciations. In computerized textual analysis for English vocabularies, a code mapping is available, accessible in terms of a database function named ‘Soundex’. Depending on pronunciation, codes are generated which are utilized for word suggestion in searching, similarity detection, text mining etc. However, this Soundex is not multilingual and supports English language only. This paper describes designing and development of Soundex for Assamese language.

References

Design and Development of Soundex for Assamese Language

- Nedjma Djouhra Ousidhoum, Asma Bensalah, and Nacéra Bensaou, "A New Classical Arabic Soundex Algorithm", Proc. of Int. Conf. on Advances in Communication and Information Technology 2012, ACEEE.

Index Terms

Computer Science
Programming Language

Keywords

Text Mining  Soundex  Similarity  Plagiarism.