Abstract

This paper proposes a sorting algorithm for Unicode based Bangla texts using Trie. Bengali texts cannot be sorted using the Unicode character scheme as Unicode character sequence is different from the Bangla Academy character sequence. Moreover, Bengali, an Indo-Aryan language spoken by approximately 200 million people, has some distinct properties with its diacritic signs. In this paper, we have sorted Bangla texts based on the Bangla Academy character order using an efficient information retrieval data structure. Our proposed algorithm is more memory efficient and is applicable to any unicode based Bangla text.

References

An Approach to Sort Unicode based Bengali Text using Trie

6. Aamira Shabnam, Debakar Shamanta Piklu, “An Easily Comprehendible Unicode Based Sorting Algorithm for Bangla Words”
8. Partha Sarathi Kar, Shantanu Mandal, Labiba Jahan, “An Improved Unicode Based Sorting Algorithm for Bengali Words”
10. বাংলা ডাটাসেট (কর্পাস) of http://scdnlab.com/corpus/

Index Terms

Computer Science Pattern Recognition

Keywords

Sorting Algorithm, Unicode Bengali text, Trie