

Phonetic-based Sindhi spellchecker system using a hybrid model

[Zeeshan Bhatti](#), [Imdad Ali Ismaili](#), [Dil Nawaz Hakro](#), [Waseem Javid Soomro](#)

Digital Scholarship in the Humanities, Volume 31, Issue 2, June 2016, Pages 264–282, <https://doi.org/10.1093/llc/fqv005>

Published:

30 March 2015

Abstract

This article presents a novel architecture using a hybrid model for developing a Sindhi spellchecker system which has yet not been developed prior to this work. The compound textual forms and glyphs of Sindhi language presents a substantial challenge for developing a Sindhi spellchecker system and generating a similar suggestion list for misspelled words. In order to implement such a system, phonetic-based Sindhi language rules and patterns must be taken into account for increasing the accuracy and efficiency. In this research work, a simple and efficient combinational hybrid system is proposed, using three different algorithms, the Edit Distance algorithm to find the measure of similarity between two Sindhi strings. The phonetic-based SoundEx and ShapeEx algorithms are developed for pattern or glyph matching, generating accurate and an efficient suggestion list for incorrect or misspelled Sindhi words. The proposed system is established with a blend between Phonetic-based SoundEx algorithm and ShapeEx algorithm for pattern or glyph matching, generating accurate and efficient suggestion list for incorrect or misspelled Sindhi words. In this article, a table of phonetically similar-sounding Sindhi characters is presented which are grouped together along with another table containing similar glyph or shape-based character groups. The system has been successfully integrated into a pre-developed Sindhi word processor application. The Sindhi word segmentation methodology and algorithms required for the spellchecker has already been published and so are not discussed in detail in this article.