



19 - 21 July 2024

London, UK

Transliteration of Non-Latin Texts: From Everyday Practice to Linguistic Technologies

Maksym Vakulenko

Darmstadt University of Applied Sciences, Darmstadt, Germany

Abstract

This paper discusses various transcoding systems designed to convert non-Latin texts into Latin script. Particularly significant is Romanization for Slavic languages. In general, the Latinization systems are categorized into two classes: those based on the transliteration approach and the ones based on practical transcription. During transliteration, the main attention is paid to simple correspondence (mutual unambiguity) between original and converted characters that allows for text reversibility, i. e. the ability to restore the original text after re-transliteration. During practical transcription, the primary concern is the sound of words in the original or another language, mostly English. In the last scenario, it is not always possible to restore the original text. The significance of transliteration extends to historical texts written in non-Latin scripts as well. Latinization systems are broadly utilized in multilingual Natural Language Processing tasks, resulting in their wider use and enhanced need.

Keywords: transliteration, practical transcription, simple correspondence, historical texts, Natural Language Processing