

Language identification system for the Tatar language

Khusainov A., Suleymanov D.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

This paper describes a speech identification system for the Tatar, English and Russian languages. It also presents a newly created Tatar speech corpus, which is used for building a language model. The main idea is to investigate the potential of basic phonotactic approaches (i.e. PRLM-approach) when working with the Tatar language. The results indicate that the proposed system can be successfully employed for identifying the Tatar, English and Russian languages. © 2013 Springer International Publishing.

http://dx.doi.org/10.1007/978-3-319-01931-4_27

Keywords

language identification, speech corpus, the Tatar language