

عنوان مقاله:

Language Identification from Text Using HMMs

محل انتشار:

دوازدهمین کنفرانس سالانه انجمن کامپیوتر ایران (سال: 1385)

تعداد صفحات اصل مقاله: 6

نویسندگان:

Oktie Hassanzadeh - *Department of Computer Science, University of Toronto, Toronto, ON M5S-3G6, Canada*

Ehsan Zamiri - *Department of Computer Engineering, Ferdowsi University of Mashad, Mashad, Iran*

خلاصه مقاله:

Language identification from text has received less attention than identification from other forms of input. This is due to the fact that it is considered an easy problem. Several techniques exist and it is possible to gain perfect accuracy in identifying the language of the text in some methods. Nevertheless, there has been very few works on the accuracy and performance of different techniques with limited input, i.e., accurate detection of the language with less input length and identification of the language of a short sentence or a single word. In this paper, we present a method based on Hidden Markov Models (HMMs) for language identification from text. We use the power of HMMs for detecting language of character strings and show the benefits of using this model over a simple model. We will show how an extremely simple realization of this model outperforms simple models in accurately identifying languages of short input strings.

کلمات کلیدی:

Language Identification, Text-to-Speech Systems, Machine Learning, Hidden Markov Models

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/44492>

