

CEUR Workshop Proceedings 2014 vol.1197, pages 54-58

Zipf's law for LiveJournal

Trifonov N.

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

Abstract

The paper provides an overview of research of frequency of language units on the material of the LiveJournal corpus. The corpus includes texts on Russian language from 2002 to 2014 year, totaling more than 5 million words of articles written by 2 thousand authors. Research was held in the following main directions, represented in the present work: estimation of coefficients for the Zipf's law for different authors, estimation of coefficients for the Zipf's law for the total number of words in all the analyzed articles.

Keywords

Frequency of words, LiveJournal corpus, Rank distribution, Zipf's law