L E N T A . R U W O R D S D A T A S E T

+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

This database contains sets of words extracted from the Internet resource's news feeds "Lenta.ru". The data was extracted in the period from 16 to 22 March 2019. The data extraction interval was 15 minutes.

The file "dictionary.txt" contains a dictionary that includes 1000 of the most common words during the data collection period. Each word or phrase occupies exactly one line.

Files are stored in the "word sets" folder. Files names are in format "setN.txt", where N is a timestamp (from 0 to 551). For example, in the file "set1.txt" data extracted 15 minutes after collection starts, in the file "set2.txt" data extracted 30 minutes after collection starts, and so on.

Each file "setN.txt" stores a list of words encountered at the specified time. Each word or phrase occupies exactly one line.