

Информация о файлах freqrnc2011.csv и freqrnc2011rev.csv
(кодировка UTF8, разрыв строк UNIX, разделитель полей табуляция).

Электронная версия издания:

О. Н. Ляшевская, С. А. Шаров, Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.

<http://dict.ruslang.ru/freq.php>

freqrnc2011.csv

Алфавитный список лемм (общая лексика; имена собственные и аббревиатуры)

Список приводится в алфавитном порядке. Указаны:

Lemma – лемма (исходная форма слова),

PoS – часть речи,

Freq(ipm) – частота леммы на 1 миллион употреблений,

R (range) – число сегментов корпуса, в которых встретилась лемма (из 100 возможных),

D – коэффициент Жуйана,

Dos – число текстов, в которых встретилась лемма.

freqrnc2011rev.csv

Обратный список лемм (общая лексика; имена собственные и аббревиатуры)

Список отсортирован по алфвиту, начиная с конца слова (обратная сортировка). Указаны:

Lemma – лемма (исходная форма слова),

PoS – часть речи,

Freq(ipm) – частота леммы на 1 миллион употреблений,

R (range) – число сегментов корпуса, в которых встретилась лемма (из 100 возможных),

D – коэффициент Жуйана,

Dos – число текстов, в которых встретилась лемма.

Len - длина слова

Reverse - перевернутая запись слова

Словарь создан на основе современного подкорпуса Основного корпуса и Устного корпуса НКРЯ (1950-2007), общий объем 92 млн. словоупотреблений.

Более подробная информация - во "Введении к словарю" <http://dict.ruslang.ru/freq.pdf>.

При использовании ресурса просьба ссылаться на первоисточник:

О. Н. Ляшевская, С. А. Шаров, Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.

(с) О. Н. Ляшевская, С. А. Шаров, 2009-2011.

(с) Национальный корпус русского языка, 2009-2011.

(с) Издательство "Азбуковник" (печатная версия словаря), 2009.

(с) А. В. Санников (создание сайта dict.ruslang.ru, техническая поддержка), 2008–2011.