

## Лабораторная работа №4 «Статистический анализ текстов из НКРЯ»

**Работу оформить в виде doc-файла и загрузить на EDU.  
Срок выполнения - до 16 ноября**

Работа состоит из нескольких этапов:

1. Изучить информацию о Национальном корпусе русского языка на сайте корпуса: <http://ruscorpora.ru/> (разделы «что такое корпус?», «состав и структура», «статистика», «параметры текстов»).
2. Исследовать частоту встречаемости наиболее употребляемых слов русского языка в текстах НКРЯ (создав "Мой подкорпус" в соответствии с тематикой вашего варианта).

В таблице приведены наиболее употребимые 9 слов в русском языке согласно *Частотному словарю русского языка* (под ред. Л.Н. Засориной). Для выполнения работы необходимо исследовать подкорпус в соответствии с вашим вариантом в Национальном корпусе русского языка и заполнить столбец «НКРЯ»:

<i>Слово</i>	<i>Частотность (ЧС)</i>	<i>Абсолют. частота (НКРЯ)</i>	<i>Частотность (НКРЯ)</i>
<b><i>В</i></b>	<b>0,0406</b>		
<b><i>И</i></b>	<b>0,0343</b>		
<b><i>НЕ</i></b>	<b>0,0182</b>		
<b><i>Я</i></b>	<b>0,0131</b>		
<b><i>БЫТЬ</i></b>	<b>0,0126</b>		
<b><i>ОДИН</i></b>	<b>0,0031</b>		
<b><i>ГОД</i></b>	<b>0,0021</b>		
<b><i>БОЛЬШОЙ</i></b>	<b>0,0020</b>		
<b><i>ОЧЕНЬ</i></b>	<b>0,0013</b>		
<b><i>Объем подкорпуса</i></b>		<b>N</b>	

Проведите анализ распределения полученных частот (для каждого слова), сравнив их с результатами по *Частотному словарю русского языка* (под ред. Л.Н. Засориной, 1977 г.).

3. Провести статистический анализ текстов, представленных в Национальном корпусе русского языка

Работа заключается в проведении статистического анализа текстов, представленных в Национальном корпусе русского языка, по грамматическим категориям.

Для выполнения работы необходимо выбрать подкорпус со снятой омонимией (в соответствии с вашим вариантом) в Национальном корпусе русского языка и используя лексико-грамматический поиск Национального корпуса, определите частоту употребления каждой грамматической категории.

Заполните таблицу:

Г рамматические категории	Абсолютная частота (F)	Относительная частота (f)
Существительное		
Глагол		
Прил агательное		
Наречие		
Числительное		
Числительное-прилагательное		
Местоимение-существительное		
Местоимение-прилагательное		
Местоименное наречие		
Местоимение-предикатив ( <i>некого, нечего</i> )		
Союз		
Предлог		
Частица		
<i>в том числе: частицы не, ни</i>		
Причастие		
Вводи ое слово		
Междометие		
Инзциал		
ИТОГ О (общая сумма)		

Проведите анализ распределения полученных частот грамматических категорий, сравнив полученные данные с данными таблицы № 4 приложения 4 частотного словаря (ЧС) и выявив особенности рекламного жанра.

Частотный словарь русского языка [Текст]: Около 40000 слов / [Сост. В. А. Аграев, В. В. Бородин, Л. Н. Засорина и др.]; под ред. Л. Н. Засориной; Ленинградский гос. ун-т им. А.А. Жданова, Науч.-исслед. ин-т прикл. математики и кибернетики при Горьковском гос. ун-те им. Н. И. Лобачевского. - Москва : Рус. яз., 1977. - 935 с. : ил.; 21 см. - **есть электронный вариант**

**ВАРИАНТЫ**

1	Андриевский Максим Дмитриевич	информатика
2	Белотелов Евгений Витальевич	фантастика
3	Васил'ючек Даниил Александрович	искусство и культура
4	Веренич Владислав Николаевич	реклама
5	Голубович Маргарита Михайловна	юмор и сатира
6	Кулешевич Тимофей Витальевич	религия
7	Курочкин Александр Владимирович	реклама
8	Нестюкович Яна Александровна	спорт
9	Новик Арина Вячеславовна	детская
10	Орлович Алексей Юрьевич	история
11	Поживилко Фёдор Андреевич	детектив, боевик
12	Полывяный Глеб Андреевич	право
13	Радюк Арина Юрьевна	здоровье и медицина
14	Рудой Андрей Игоревич	драматургия
15	Третьяков Дмитрий Александрович	новости
16	Уминский Антон Петрович	приключения
17	Черненк'ов Артем Витальевич	фольклор
18	Чернявский Никита Сергеевич	медицина