Лабораторная работа №4

«Статистический анализ текстов из НКРЯ»

1. Изучить информацию о Национальном корпусе русского языка на сайте корпуса: [http://ruscorpora.ru/ (](http://ruscorpora.ru/)разделы

«что такое корпус?», «состав и структура», «статистика», «параметры текстов»).

*Изучил указанные выше разделы*

1. 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»:

|  |  |  |  |
| --- | --- | --- | --- |
| ***Слово*** | ***Частотность (ЧС)*** | ***Абсолют. частота* (НКРЯ)** | ***Частотность (НКРЯ)*** |
| ***В*** | **0,0406** | 26079 | 0.030569 |
| ***И*** | **0,0343** | 30081 | 0.03526 |
| ***НЕ*** | **0,0182** | 5946 | 0.006969 |
| ***Я*** | **0,0131** | 1756 | 0.002058 |
| ***БЫТЬ*** | **0,0126** | 4520 | 0.00529 |
| ***ОДИН*** | **0,0031** | 1762 | 0.0020654 |
| ***ГОД*** | **0,0021** | 5051 | 0.00592 |
| ***БОЛЬШОЙ*** | **0,0020** | 1221 | 0.001431 |
| ***ОЧЕНЬ*** | **0,0013** | 740 | 0.0008674 |
| ***Объем подкорпуса*** |  | **853099** |  |

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** | 8491 | 0.349366359 |
| **Глагол** | 2846 | 0.117100066 |
| **Прил агательное** | 3061 | 0.125946346 |
| **Наречие** | 699 | 0.028760698 |
| **Числительное** | 966 | 0.039746544 |
| **Числительное-прилагательное** | 91 | 0.00374424 |
| **Местоимение-существительное** | 924 | 0.038018433 |
| **Местоимение-прилагательное** | 1150 | 0.047317314 |
| **Местоименное наречие** | 282 | 0.011603028 |
| **Местоимение-предикатив *(некого, нечего)*** | 1 | 4.11455E-05 |
| **Союз** | 1773 | 0.072950955 |
| **Предлог** | 2868 | 0.118005267 |
| **Частица** | 506 | 0.020819618 |
| **Причастие** | 515 | 0.021189928 |
| **Вводи ое слово** | 69 | 0.002839039 |
| **Междометие** | 3 | 0.000123436 |
| **Инзциал** | 59 | 0.002427584 |
| **ИТОГ О (общая сумма)** | 24304 |  |

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***

|  |  |  |
| --- | --- | --- |
| 4 | Веренич Владислав Николаевич | реклама |