

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ

ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО

ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(ДГТУ)

**Отчет по лабораторной работе №1**

Исследование применимости законов Зипфа к русскоязычным текстам

Выполнил:

студент МИН21

Урывский Д.В.

Ростов-на-Дону

2020

**Цель работы**

В ходе лабораторной работы получить практические навыки морфологического анализа текста, применимости законов Зипфа к русскоязычным документам и оптимизации поиска соответствующей информации в Интернет.

**Ход выполнения**

Определим частоту вхождения слов (ссылка на статью http://itno.e.donstu.ru/documents/articles/313-316.pdf).

Таблица 1. Частота вхождения слов

|  |  |  |
| --- | --- | --- |
| Ранг | Частота | Слово |
| 1 | 10 | система |
| 2 | 7 | студент |
| 2 | 7 | изучение |
| 3 | 6 | курс |
| 4 | 4 | работа |
| 5 | 3 | направление |
| 5 | 3 | механики |
| 5 | 3 | механика |
| 5 | 3 | язык |
| 5 | 3 | прикладной |
| 5 | 3 | подготовки |
| 5 | 3 | отдельный |
| 5 | 3 | процесс |
| 5 | 3 | позволять |
| 5 | 3 | компьютерных |
| 5 | 3 | компьютерной |
| 5 | 3 | комплекс |

Определим вероятность вхождения произвольно выбранного слова в текст. Очевидно, она будет равна отношению частоты вхождения этого слова к общему числу слов в тексте. Таким образом, справедливо следующее выражение:

Вероятность = Частота вхождения слова / Число слов

Вероятность слова “изучение” = 7 / 700 = 0,01 = 1%

Если умножить вероятность обнаружения слова в тексте на ранг частоты, то получившаяся величина (С) – константа Зипфа приблизительно постоянна:

С = (Частота вхождения слова х Ранг частоты) / Число слов

С = (“изучение” 7 \* 2) / 700 = 0,02

С = (“курс” 6 \* 3) / 700 = 0,026

С = (“язык” 5 \* 3) / 700 = 0,021

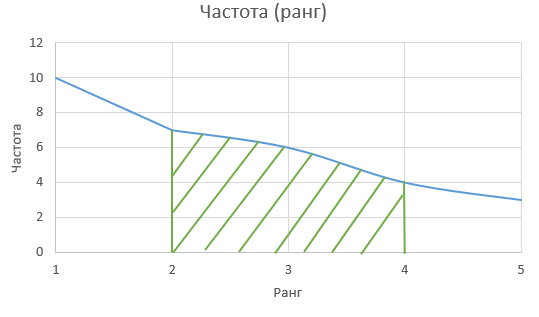


Рис. 1. Диаграмма частота – ранг

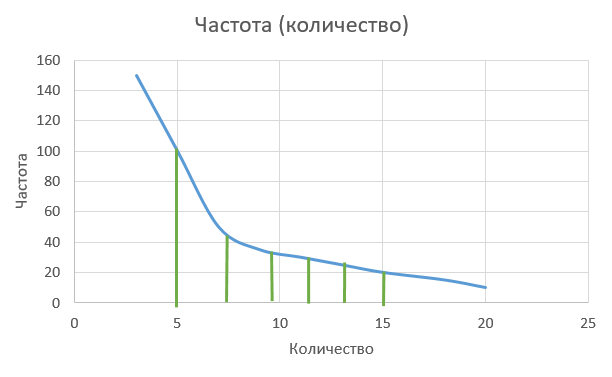


Рис. 2. Диаграмма частота – количество

Исходя из анализа был создан поисковой запрос, состоящий из слов: система студент изучение курс работа направление механики механика язык прикладной.