Гончарова Катерина, 132

Домашнее задание 1

1. Объем SpecC – количество словоупотреблений: 3 020 362, тематика / жанр корпуса: нехудожественные тексты, тематика «биология», источник корпуса: НКРЯ
2. Объем RefC – количество словоупотреблений: 1 655 521, источник: НКРЯ (тематика / жанр корпуса: нехудожественные тексты, тематика «физика»)
3. Метод 2: chi-squared, источник: http://corpora.lancs.ac.uk/sigtest/
4. Таблица 1.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| wi | Тип (специфичное vs. общеупотребитель-ное) | CountSpecC  (частота в Specific Corpora) | CountRefC (частота в контрастном корпусе  (Reference Corpora)) | LogLikelihood | Ранг | Вес 2 | Ранг |
| кровь | специфичное | 1454  вхождений | 105  вхождений | 719.65 | 1 | 558.96 | 1 |
| лимфоузел | специфичное | 147  вхождений | 0 вхождений | 128.49 | 3 | 79.03 | 3 |
| есть | общеупотребительное | 4243 вхождений | 2096  вхождений | 15.34 | 4 | 15.06 | 4 |
| что | общеупотребительное | 35 200  вхождений | 21 726  вхождений | 187.51 | 2 | 187.10 | 2 |

1. Комментарий: Методы поставили специфичные слова выше в ранге, но общеупотребительное слово «что» оказалось на втором месте, выше специфичных слов, что неудивительно, так «что» является союзом.