ДЗ №1

Специфичные слова (тема - спорт) – арбитр, тайм

Общеупотребительные слова – стол, книга

По данным НКРЯ:

1. SpecC: Нехудожественные тексты

тематика текста – спорт

3 282 582 слова

1. RefC: НКРЯ

Нехудожественные тексты

Тематика текста – искусство и культурв

18 849 410 слов

Метод 2 : Weirdness (wi) = frs(wi)/frr(wi) = (Ws / Ts) / (Wr / Tr)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| wi | Тип (специфичное vs. общеупотребитель-ное) | CountSpecC  (частота в Specific Corpora) | CountRefC (частота в контрастном корпусе  (Reference Corpora)) | LogLikelihood | Ранг | Weirdness | Ранг |
| арбитр | специфичное | 233 | 12 | 797.37 | 1 | 110.91 | 1 |
| тайм | специфичное | 186 | 47 | 490.72 | 2 | 22.76 | 2 |
| стол | общеупотребительное | 360 | 4247 | - 211.14 | 3 | 0.49 | 3 |
| книга | общеупотребительное | 539 | 20722 | - 3685.00 | 4 | 0.15 | 4 |

Комментарий:

В обоих методах слова получили одинаковые ранги: специфичные слова занимают верхние строки в обоих методах. Однако, в weidness достаточно большое различие между специфичными словами, и маленькое между общеупотребительными. В LL – обратная ситуация.