Домашняя работа. LogL. Сухова Ксения, 132 гр

Я выбрала два корпуса:

1. Собрание сочинений братьев Стругацких (первые 8 томов)
2. НКРЯ

Было выбрано 4 слова: два специфичных и два общеупотребительных (интуитивно, без предварительных расчётов). Специфичные: планета, звездолет; общеупотребительные: окно, стол.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| wi | Тип | CountSpecC | CountRefC | LogLikelihood | Ранг | Вес 2 (хи2) | Ранг |
| планета | спец | 937 | 13 603 | 2604.97 | 1 | 7231.3 | 1 |
| звездолет | спец | 174 | 551 | 948.74 | 2 | 5865.5 | 2 |
| окно | общ | 1338 | 82 128 | 777.09 | 3 | 1077.5 | 3 |
| стол | общ | 1298 | 106 216 | 317.13 | 4 | 457.79 | 4 |
| SpecC: |  | 1798486 | 265 401 717 |  |  |  |  |

Вывод:

Ранжирование слов по разным методам оценки специфичности не отличается. Однако для LogL “окно” ближе по специфичности к “звездолету”, чем к планете. Более логично смотрится распределение чисел в тесте хи квадрат.

Также, интуитивно, казалось, что “звездолет” должен быть более специфичен для Стругацких, чем “планета”. Однако оба теста показывают, что это не так.