**READ ME – ASS2:**

Kfir Rymberg 311374185, Orel Hagag 204501704

**How to Run:**

java -jar extract.collations.using.mr.jar ExtractCollations heb/eng

**Description:**

Step 1 – Counting 1-grams & 2-grams ordered by decade

Step 2 – Insert 1-gram counter into 2-grams by first word into it's value

Step 3 – Insert 2-grams counter and N into 2-grams by second word into it's value

Step 4 – Calculates likelihood and orders the results by Top 100

**Output:**

English output:

<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00000>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00001>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00002>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00003>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00004>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00005>  
<https://eceng.s3.amazonaws.com/RatioCalculator/part-r-00006>

|  |  |  |
| --- | --- | --- |
| Steps: | RecordsMapToReducer  with / without Combiner | |
| 1 Counter | 2198655767 | 3402817509 |
| 2 Recorder | 2265154731 | 2265154731 |
| 3 Recorder2 | 2000113116 | 2000113116 |
| 4 RatioCalculator | 108226 | 88050066 |

Hebrew output:

<https://orelkfir.s3.amazonaws.com/output/part-r-00000>  
<https://orelkfir.s3.amazonaws.com/output/part-r-00001>  
<https://orelkfir.s3.amazonaws.com/output/part-r-00002>

|  |  |  |
| --- | --- | --- |
| Steps: | RecordsMapToReducer  with / without Combiner | |
| 1 Counter | 98964175 | 227005257 |
| 2 Recorder | 104173071 | 104173071 |
| 3 Recorder2 | 80833957 | 80833957 |
| 4 RatioCalculator | 7400 | 150321984 |

**Times: (M4\_LARGE, 3 Instances)**

|  |  |
| --- | --- |
| **Type (Combiner)** | **Duration** |
| Hebrew with | 1 hour 29 minutes |
| Hebrew without | 1 hour 37 minutes |
| English with | 13 hours 27 minutes |
| English without | 14 hours 2 minutes |

**Examples:**

10 Good:

1530 בן אלוהים : -0.006363628269753008  
1530 רוח אלוהים : -0.24849954783934436  
1780 לחיי המלך : -2.3243850218271973  
1690 God gives : -0.02743614600532851  
1740 By Heavens : 111.90253944943694  
1740 doing Good : 104.8666630895253  
1740 free Liberty : 103.99723472601598  
1740 better Order : 99.62620971209552  
1740 right Hand : 92.73559265403648  
1740 look forward : 96.50825653159252

10 Bad:

1790 רק ואפילו : -0.07087721201514796  
1820 שהוא היי : -0.3974338547967591  
1930 אלא כצורת : -0.4575541795434219  
1910 גם שוט : -0.25723721773897523  
1760 שהוא בן : -0.045366255962164814  
1690 God daily : -0.02148017943954983  
1760 He suddenly : -0.0010621812048157153  
1740 8 Good : 104.04582126974358  
1740 John 4 : 96.34514930442214  
1620 The comfort : -0.009175493663519774

**Usernames:**

we used both users:

Kfir's user:   
[kfirrym@post.bgu.ac.il](mailto:kfirrym@post.bgu.ac.il)

Orel's user:  
[hagagor@post.bgu.ac.il](mailto:hagagor@post.bgu.ac.il)