Instructions:

1. Create Credentials File –  
   run from cmd\shell: “aws configure”, you will need to fill the following:  
   aws\_access\_key\_id = your\_access\_key\_id  
   aws\_secret\_access\_key = your\_secret\_access\_key
2. Run the main jar file – java -jar main.jar <language>  
   <language> = heb for Hebrew, eng for English

Technical Information:

* Node used: M4.large()
* Number of Instances: 15
* Number of Jobs: 4

Jobs:

**Job1:**

Main objective: Counting words and merging similar keys.

Input: 1gram & 2 gram.

Output – Occurrences of each show(1 gram\2 gram\decade)

**Job2:**

Main objective: counting word1 and decade occurrences for each 2gram show

Output: Occurrences of each show(1 gram\2 gram\decade) + setting 1word and decade occurrences for 2gram shows.

**Job3:**

Main objective: counting word2 for each 2gram show

Output: only 2 grams shows after calculating the likelihood.

**Job4:**

Main objective: merging all files into one

Output: 1 file with all needed results

Results + Output:

\*Output files for each run can be found under outputs dir.

**Time:**

Results:

|  |  |
| --- | --- |
| Run Type: | Elapsed Time: |
| Heb with local aggregation | 20 minutes |
| Heb without aggregation | 21 minutes |
| Eng with local aggregation | 1 hour 55 minutes |
| Eng without aggregation | 2 hours 21 minutes |

**Records count:**

Eng with local aggregation:

|  |  |
| --- | --- |
|  | number of keys sent from mappers to reducers: |
| Job1 | 443168450 |
| Job2 | 347144353 |
| Job3 | 311391418 |
| Job4 | 4293 |

Eng without aggregation:

|  |  |
| --- | --- |
|  | number of keys sent from mappers to reducers: |
| Job1 | 2096289448 |
| Job2 | 347144353 |
| Job3 | 311391418 |
| Job4 | 4293 |

Heb with local aggregation

|  |  |
| --- | --- |
|  | number of keys sent from mappers to reducers: |
| Job1 | 43251909 |
| Job2 | 35265062 |
| Job3 | 26858901 |
| Job4 | 3300 |

Heb without aggregation:

|  |  |
| --- | --- |
|  | number of keys sent from mappers to reducers: |
| Job1 | 215234250 |
| Job2 | 35265062 |
| Job3 | 26858901 |
| Job4 | 3300 |

Good examples:

Decade: 1560, Collocation: בית דין, LikeliHood: 206.3462372567124

Decade: 1670, Collocation: בית המקדש, LikeliHood: 74.57860000590745

Decade: 1750, Collocation: עדי ראייה, LikeliHood: 117.09316334173215

Decade: 1830, Collocation: צרורה בצרור, LikeliHood: 120.19533041362001

Decade: 1670, Collocation: גוג ומגוג, LikeliHood: 75.64737483364209

Decade: 1580, Collocation: domini nostri, LikeliHood: 53.94528926141686

Decade: 1600, Collocation: human nature, LikeliHood: 16.282999169243286

Decade: 1720, Collocation: holy ghost, LikeliHood: 33.989230054972246

Decade: 1620, Collocation: eternal life, LikeliHood: 16.257316045398625

Decade: 1590, Collocation: inner parts, LikeliHood: 24.319998702275257

Bad examples:

Decade: 1530, Collocation: ועכשו אתה, LikeliHood: 22.419749362041898

Decade: 1680, Collocation: ענה הזקן, LikeliHood: 31.832312176424466

Decade: 1700, Collocation: תראה איך, LikeliHood: 39.03322585113085

Decade: 1810, Collocation: לחשוב בתקופות, LikeliHood: 14.453165332037232

Decade: 1870, Collocation: שמו מגת, LikeliHood: 25.339299259189318

Decade: 1600, Collocation: wast wont, LikeliHood: 14.717094424908474

Decade: 1720, Collocation: john gibbs, LikeliHood: 22.717749756989846

Decade: 1590, Collocation: celebrated physician, LikeliHood: 8.5241699939699

Decade: 1590, Collocation: sir philip, LikeliHood: 65.81951778163528