DSP 2

# נתונים

מספר זוגות הkey-value שעברו בין הmapper הראשוני שביצע את החלוקה לשני קורפוסים, ובנוסף סידר את המידע בצורה שבה שאר הjobs יכולים להתבצע הוא 382051404, ומשקלם הוא 537180832 בתים.

# שלשות מילים מעניינות

1. The-united
   1. The-united-forces 5.18E-08
   2. The-united-states 3.72E-08
   3. The-united-efforts 2.95E-08
   4. The-united-force 2.62E-08
   5. The-united-armies 2.50E-08
   6. The-united-army 2.04E-08
2. My-name
   1. Harry-Potter-books 2.18E-07
   2. Harry-Potter-series 6.84E-08
   3. Harry-Potter-book 4.17E-08
   4. Harry-Potter-novels 1.80E-08
   5. Harry-Potter-stories 1.73E-08
3. If-we
   1. If-We-Must 1.17E-07
   2. If-We-Can 1.91E-08
   3. If-We-Had 1.76E-08
   4. If-We-Are 1.48E-08
   5. If-We-Do 1.09E-08
4. We-love
   1. We-Love-You 5.70E-08
   2. We-Love-Our 1.50E-08
   3. We-Love-Glenda 8.00E-09
   4. We-Love-Can 7.91E-09
   5. We-Love-Each 6.14E-09
5. School-football
   1. School-football-team 6.36E-08
   2. School-football-coach 2.90E-09
   3. School-football-game 2.83E-09
   4. School-football-field 2.78E-09
   5. School-football-games 2.46E-09
6. Your-personal
   1. Your-Personal-Library 2.69E-08
   2. Your-Personal-Life 1.40E-08
   3. Your-Personal-Finances 1.25E-08
   4. Your-Personal-Computer 1.14E-08
   5. Your-Personal-Guide 9.65E-09
7. When-Israel
   1. When-Israel-went 5.38E-08
   2. When-Israel-came 3.32E-08
   3. When-Israel-saw 6.47E-09
   4. When-Israel-entered 5.48E-09
   5. When-Israel-does 4.94E-09
8. Walt-Disney
   1. Walt-Disney-World 8.42E-07
   2. Walt-Disney-Productions 1.62E-07
   3. Walt-Disney-Company 1.53E-07
   4. Walt-Disney-Studios 1.05E-07
   5. Walt-Disney-Co 8.41E-08
9. What-price
   1. What-Price-Glory 1.16E-06
   2. What-Price-Hollywood 5.69E-08
   3. What-Price-Vigilance 1.26E-08
   4. What-Price-Israel 1.23E-08
   5. What-Price-Relieving 9.09E-09
10. My-name
    1. My-Name-Is Infinity
    2. My-Name-shall 1.14E-08
    3. My-Name-Was 8.67E-09
    4. My-Name-Address 5.52E-09
    5. My-Name-Good 4.94E-09

המערכת הגיע לתוצאות צפויות, אף שהיינו צריכים לבדוק מאיפה הגיעה השלשה we-love-glenda (מסתבר שיש ספר כזה).  
למשל, ניתן לראות כי הסלוגן walt-disney-world פופולרי יותר מאשר walt-disney-co.  
נציין גם שבstop-words סיננו גם מילים באורך אות אחת מתוך מחשבה שאלו שגיאה סטטיסטית ולא מילים אמיתיות.

# הוראות הרצה

יש לפתוח את תיקיית הRunner כאן נמצאת תכנית ההרצה, יש לפתוח את קובץ הJava, ואז להריץ את הmain.

# קובץ הפלט של הריצה

קובץ הפלט נמצא בקישור <https://resultproject2dsp2023.s3.amazonaws.com/result/part-r-00000>.