

אחזור מידע מהאינטרנט (67782) תרגיל 2: בניית האינדקס - ניתוח

מוגש על-ידי: מיכאל גרינבאום (211747639), גיא לוי (211744636)

31 במאי 2021

1 תיאור כללי

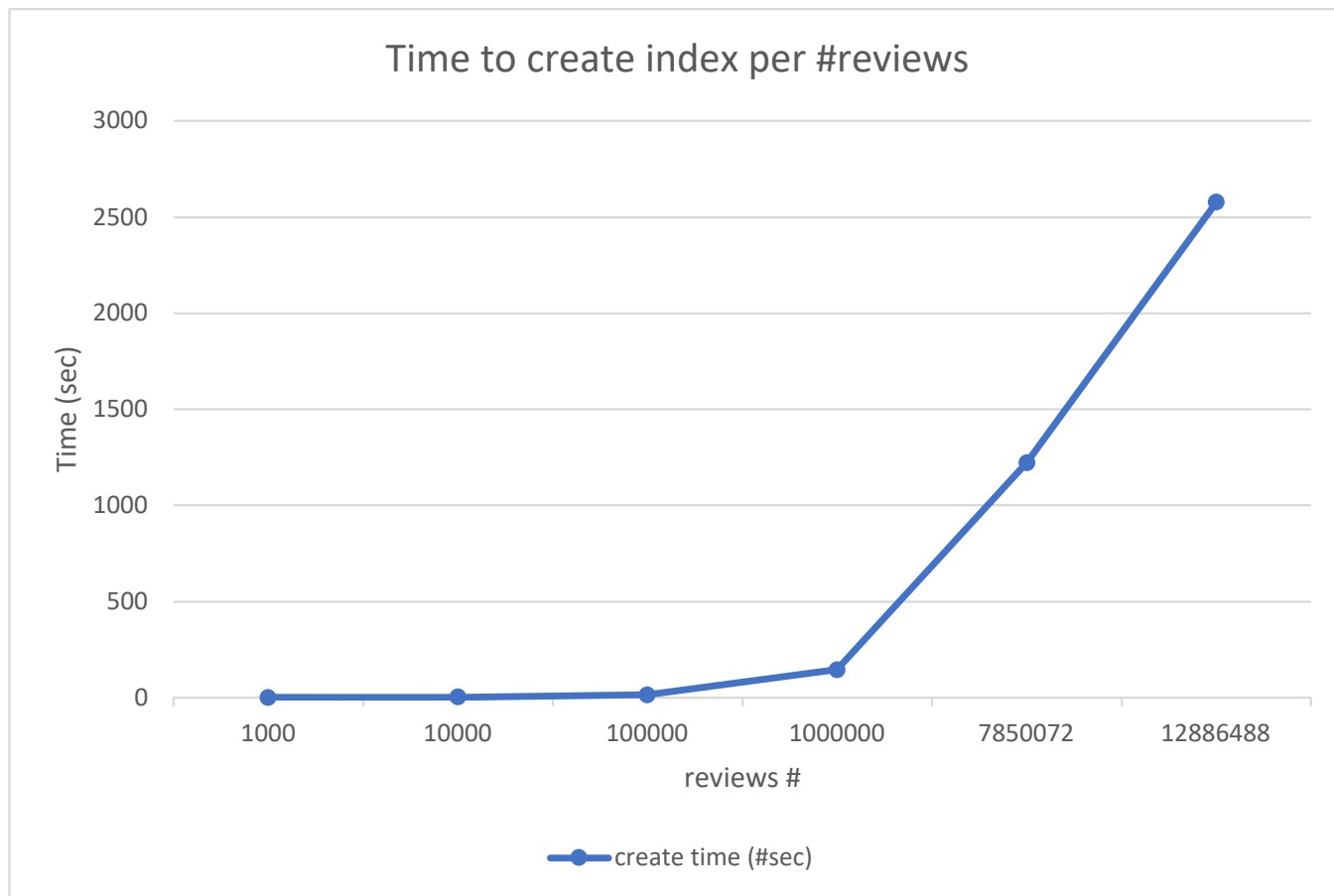
בניית האינדקס שלנו נעשית באופן הבא:

1. אנחנו יוצרים אינדקסים בסיסיים על $K (= 500,000)$ משובים רצופים עד סיום קריאת כל האינדקסים
 2. אנחנו מתחילים תהליך מיזוג של עד $L (= 20)$ אינדקסים כל פעם מהאינדקסים שנוצרו בשלב האחרון וחוזרים על שלב זה עד שנוותר אינדקס אחד.
- המספרים הספציפיים מבטיחים פרקטית שאנחנו לא חורגים מגודל heap size בזמן הריצה.

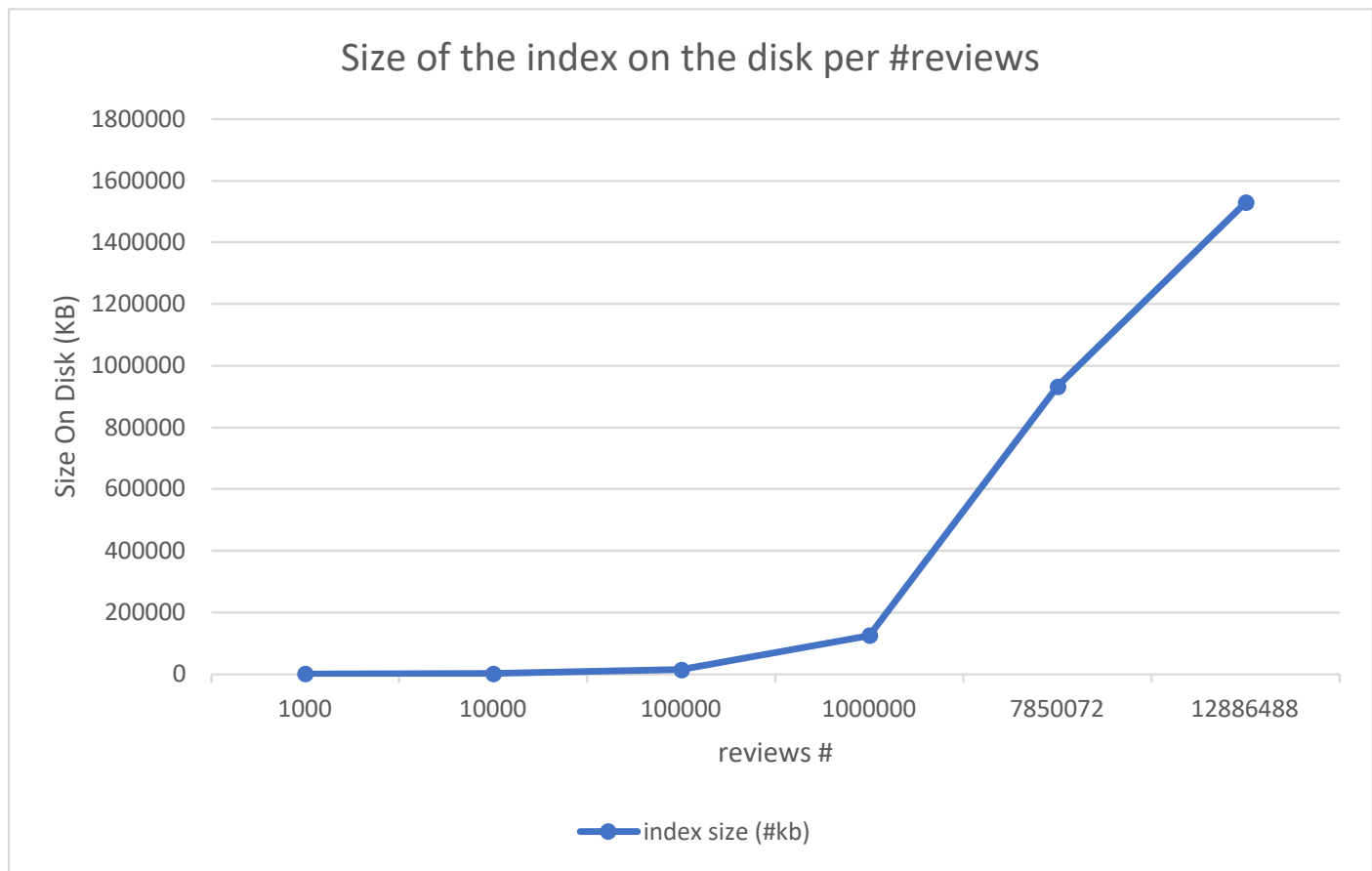
2 תוצאות

ערכנו את הניסויים המבוקשים על המחשב beh מרחוק ויש לציין כי תנאי הבדיקה לא היו אופטימליים (בדרך כלל זמן יצירת אינדקס של 500,000 משובים לקחה 50 שניות, היו זמנים בזמן הבדיקה שזה לקח במקום דקה וחצי, ככל הנראה כי הופעלו תהליכים נוספים על המחשב).

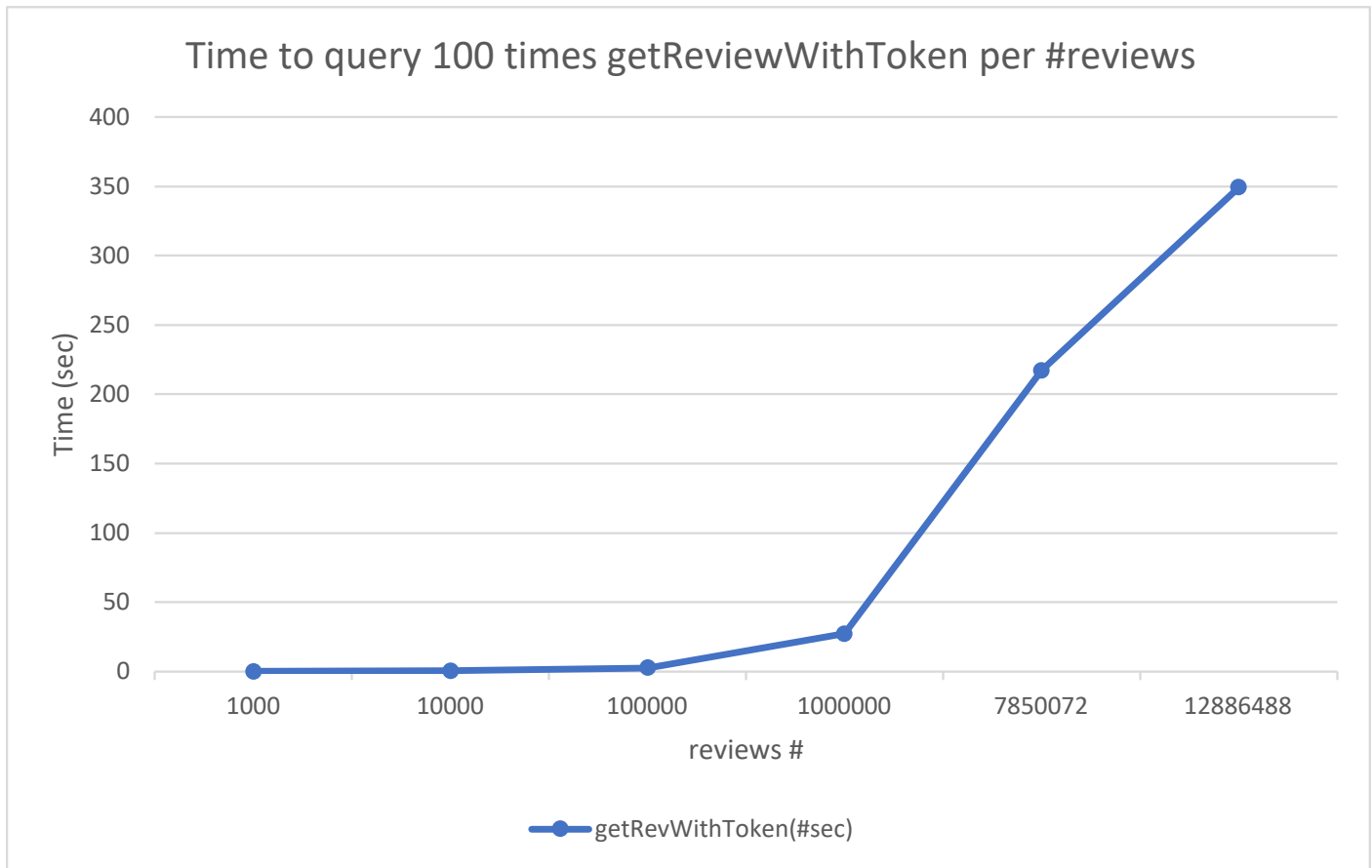
1. גרף של זמן יצירת האינדקס כפונקציה של כמות המשובים:



2. גרף של גודל האינדקס כפונקציה של כמות המשובים:



3. גרף של זמן מענה ל- 100 שאילתות :getReviewsWithToken



4. גרף של זמן מענה ל- 100 שאילתות getTokenFreq:

