

Web info retrieval ex2 analysis

Eran Hashmonay 308281435

June 23, 2021

Data

All data was gathered on an aquarium computer

1000 reviews:

Index creation: 2 sec.

Index disk size: 3.1M.

runtime for getReviewsWithToken * 100: 6 mili sec.

runtime for getTokenFrequency * 100: 5 mili sec

10,000 reviews:

Index creation: 27 sec.

Index disk size: 28M.

runtime for getReviewsWithToken * 100: 50 mili sec.

runtime for getTokenFrequency * 100: 45 mili sec.

100,000 reviews:

Index creation: 253 sec.

Index disk size: 271.4M.

runtime for getReviewsWithToken * 100: 451 mili sec.

runtime for getTokenFrequency * 100: 434 mili sec

1,000,000 reviews:

Index creation: 2644 sec.

Index disk size: 2.8G.

runtime for getReviewsWithToken * 100: 1251 mili sec.

runtime for getTokenFrequency * 100: 1222 mili sec

5,000,000 reviews:

Index creation: 12,989 sec = 3.6 hours.

Index disk size: 13G.

runtime for getReviewsWithToken * 100: 1.4 sec.

runtime for getTokenFrequency * 100: 1.4 sec

Graphs:

