דו"ח עבור פרוייקט מנוע אחזור בקורס אחזור מידע

b. **https://github.com/ayemoy/RetrieveProject**

c. **Link to anchor postings_gcp bucket**
   Link to title posting_gcp bucket
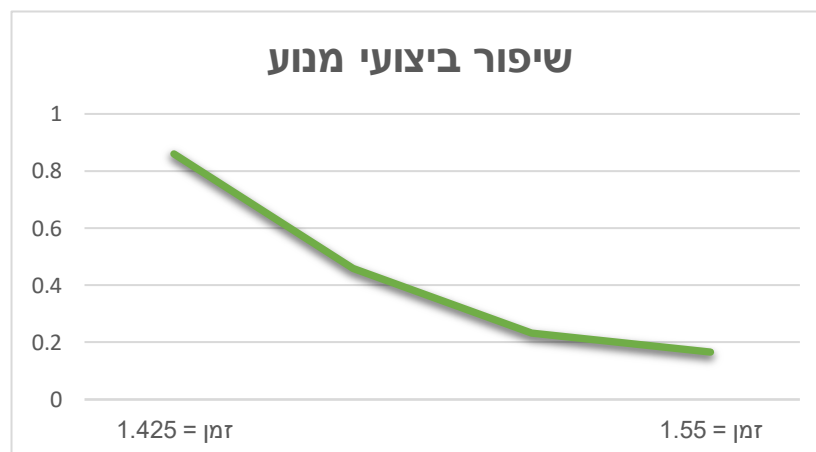   **Link to body postings_gcp bucket**

d. **Link to list all index files**

e. While building the search engine we performed a number of key experiments in order to evaluate and improve the performance of the retrieval results.
   At first, we tried to use the BM25 algorithm to optimize the results coming out of the queries, but changing the parameters in the class did not yield us satisfactory results. (We had 46% fail in the search function and 86% fail in the body search function).
   We then tried to use the similar cosine formula to optimize the search results, and we saw that the results improved miraculously.
   Therefore, we decided to combine the two search functions for the body of the article and the title of the article, give them weights, and use the cosine function, thus optimizing the search engine (we have 16% failures in the main search function of the engine, averaging 1.7 for 30 queries, and an average of 0.526 for map @ 40).

f.

g. **screen shoots:**

```
⇥  30
   failures : 14
   time avg:  2.5088390111923218
   0.4666666666666667
```

```
30
failures : 26
time avg:  1.4252774794896443
0.8666666666666667
```

```
⇥  30
   failures : 5
   time avg:  1.5858024915059408
   0.16666666666666666
```

```
⇥  30
   failures : 7
   time avg:  1.5558143615722657
   0.23333333333333334
```
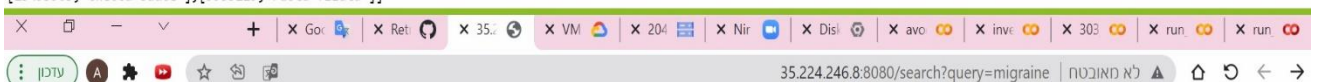
h. We ran the query "migraine" and the query "Does pasta have preservatives?",
We saw that for the query "migraine" we got the score 0.949, and passed the
query "Does pasta have preservatives" we got the score 0.121.
It can be seen in the first results that queries are repeated that are relevant to
"pasta" but not to "preservatives".
However, in the query "migraine" more relevant results are repeated.
This is because the more words the query contains, the harder it is to pinpoint
the results that need to be returned. When a query consists of only one word
then the documents dealing with exactly that word can be returned.

[[25659792,"Pasta (disambiguation)"],[23871,"Pasta"],[3080697,"List of pasta"],[34121672,"Dry pasta line"],[53487581,"Banana pasta"],[8245256,"Gemelli (pasta)"],[4275382,"Pasta e fagioli"],[6760389,"Fiori (pasta)"],[8245307,"Lanterne (pasta)"],[43843664,"Ditalini"],[6984468,"Pasta salad"],[61014433,"Wine preservatives"],[29191077,"Bavette (pasta)"],[43911472,"List of pasta dishes"],[9303405,"Mafaldine"],[67279077,"Cascatelli"],[50404581,"Pasta processing"],[13801712,"Pastina"],[23854177,"\u015eehriye"],[48497982,"Capunti"],[1330188,"Al dente"],[54155622,"Pasta mollicata"],[55365859,"Museum of Pasta"],[56383357,"Tortiglioni"],[15974787,"Foglie d'ulivo"],[48461814,"Casarecce"],[47710154,"Fagottini"],[8244710,"Cencioni"],[25215165,"Sacchettoni"],[57252592,"Pasta all'Ortolana"],[45677868,"Pasta con le sarde"],[25215220,"Pillus"],[25215213,"Lasagnotte"],[43911086,"Ciceri e tria"],[1296540,"Pasta al pomodoro"],[31585693,"Pasta alla Norma"],[65360198,"Pasta con i peperoni cruschi"],[57492967,"Pasta alla gricia"],[531849,"Giuditta Pasta"],[13226043,"Rastakraut Pasta"],[2864831,"John Pasta"],[3141956,"Farfalle"],[6728737,"Baked ziti"],[27274354,"New World Pasta"],[19876749,"T\u00f0far\u00f3s csusza"],[19994314,"National Pasta Association"],[43853813,"Macaroni art"],[67634547,"Cimcik eri\u015fte"],[8245125,"Rotelle"],[25215091,"Anelli"],[67624046,"Uzun eri\u015fte"],[34825096,"Pasta Bravo"],[3028364,"Campanelle"],[54009788,"Broccoli Pizza & Pasta"],[8892877,"Radiatori"],[3028377,"Conchiglie"],[2387344,"Bucatini"],[12638661,"Corzetti"],[48461629,"Calamarata"],[8966818,"Capellini"],[28096212,"PASTA domain"],[43981430,"Maccheroni alla molinara"],[35211682,"Cacio e pepe"],[6972293,"Alphabet pasta"],[6696765,"Fusi (pasta)"],[24862,"Preservative"],[63549718,"Pasta Poiatti"],[25215222,"Sagnarelli"],[42102760,"Bigoli in salsa"],[40478583,"Maltagliati"],[53746577,"Crozets de Savoie"],[48846696,"Hilopites"],[3328852,"Strozzapreti"],[16591942,"Neapolitan cuisine"],[15212680,"Passatelli"],[4627535,"Pasta primavera"],[56401501,"Fileja"],[1266369,"Ziti"],[54596819,"Pasta Rigo"],[21281463,"Stringozzi"],[48460535,"Barbine"],[56643734,"Alkaline pasta"],[65419304,"Flomaria"],[8870178,"Robertsons"],[3069815,"Trenette"],[2899729,"Orzo"],[12853101,"Pici"],[3285059,"Fusilli"],[3450096,"Cavatappi"],[6770948,"Gragnano"],[1032674,"Chef Boyardee"],[67726685,"Storci"],[5382150,"Spaghetti and meatballs"],[2436101,"Manicotti"],[25215239,"Tripoline"],[28930609,"Mangia"],[26189868,"Garganelli"],[858120,"Fettuccine"],[13493965,"Checca sauce"],[9955119,"Pasta filata"]]

35.224.246.8:8080/search?query=migraine

[[21035,"Migraine"],[24224679,"ICHD classification and diagnosis of migraine"],[31556991,"Migraine Action Association"],[11495285,"Migraine (disambiguation)"],[2702930,"Acephalgic migraine"],[20038918,"Migraine Boy"],[234876,"Migraine (book)"],[40748148,"Genetics of migraine headaches"],[62871079,"Menstrual migraine"],[7362738,"Sporadic hemiplegic migraine"],[22294424,"Migraine-associated vertigo"],[23174077,"Migraine Aura Foundation"],[7362700,"Hemiplegic migraine"],[22352499,"Retinal migraine"],[18741438,"Migraine Trust"],[55115883,"Messoud Ashina"],[36984150,"Migraine treatment"],[26085681,"Ted Pecchio"],[16982268,"Migraine surgery"],[57672434,"Ubrogepant"],[25175287,"Anne MacGregor"],[44627466,"Edwin Bickerstaff"],[25045060,"Prevention of migraines"],[61386909,"Marcia Wilkinson"],[54077917,"Rimegepant"],[12410589,"Antimigraine drug"],[39458542,"Hubert Airy"],[26584776,"Migralepsy"],[64269900,"Migraine (Moonstar88 song)"],[5690287,"Scintillating scotoma"],[40428462,"Migraine Specific Quality of Life"],[50083054,"Eptinezumab"],[46919927,"CP-122,288"],[26585811,"Intercalated seizure"],[3108484,"Neurogenic inflammation"],[15712244,"Mixed tension migraine"],[739855,"Visual snow"],[11193835,"Cortical spreading depression"],[50977642,"Galcanezumab"],[3868233,"Eletriptan"],[51078678,"Erenumab"],[4206029,"Familial hemiplegic migraine"],[14001660,"Onze Danses Pour Combattre la Migraine"],[36666029,"Occipital nerve stimulation"],[36579642,"Abdominal migraine"],[36579839,"Calcitonin gene-related peptide receptor antagonist"],[55121208,"Danish Headache Center"],[2555865,"Aura (symptom)"],[692988,"Frovatriptan"],[21594484,"Telcagepant"],[3332410,"Methysergide"],[67754025,"Peter Goadsby"],[54027296,"William J. Binder"],[47403613,"Michel Ferrari"],[12652799,"Naratriptan"],[14725943,"ATP1A2"],[3867847,"Almotriptan"],[42561292,"Ergotamine/chlorcyclizine/caffeine"],[24361010,"Migraine Disability Assessment Test"],[24155104,"C16H21N3O2"],[655650,"Petasites hybridus"],[18264585,"Dimetotiazine"],[3053003,"Osmophobia"],[36026903,"PNU-142633"],[66749629,"AMG-333"],[20799837,"Cyclandelate"],[2139889,"Nociceptive trigeminal inhibition tension suppression system"],[3602651,"Photopsia"],[6052114,"Buclizine"],[36575397,"Postdrome"],[48896644,"Pipethiaden"],[5692662,"Calcitonin gene-related peptide"],[14201682,"Persistent aura without infarction"],[41769313,"Frank Clifford Rose"],[28126369,"C14H12ClNO2"],[843361,"Triptan"],[910244,"Rizatriptan"],[60471253,"ROSAH syndrome"],[8235440,"Pituitary adenylate cyclase-activating peptide"],[9400334,"Proxibarbital"],[57374577,"James W. Lance"],[1141900,"Zolmitriptan"],[47816999,"Teenage Jesus and the Jerks (album)"],[46422015,"Edward Liveing"],[1015919,"Excedrin (brand)"],[61962436,"Migrainous infarction"],[48334533,"5-HT1D receptor agonist"],[18600765,"Dotarizine"],[22901459,"Cephalalgiaphobia"],[63936555,"Patrick Humphrey"],[21381229,"NIH classification of headaches"],[16115433,"LY-334370"],[35736568,"Tropanserin"],[26060699,"Carotidynia"],[21213682,"Calciumopathy"],[3244730,"Caffeine/ergotamine"],[6154091,"Pizotifen"],[44909160,"Selurampanel"],[3105039,"Live Forever or Die Trying"],[4128619,"Dichloralphenazone"]]

0.646
0.698
0.549
0.628
0.608
0.121
0.393
0.913
0.299
0.614
0.045
0.536
0.843
0.333
0.924
0.441
0.025
0.258
0.799
0.711
0.587
0.613
0.074
0.799
0.333
0.094
0.558
0.684
0.659
0.599