

Final project-Information Retrieval:

Student Name: Smeh Diab , Yuval Antman

ID: 213034671, 212054167

Email: diabsame@post.bgu.ac.il , Yuvalantman@gmail.com

Github: https://github.com/same1222/wikipedia-search_engine

Google Storage Bucket (public):

to my bucket <gs://samihdiab-123456-bucket>

- **Experimental Setup**

- The goal of this project was to build an efficient and effective large-scale search engine over the English Wikipedia corpus.

All experiments were conducted on a fixed document collection and query set provided in the course. The system was implemented using inverted indexes stored in Google Cloud Storage, while query processing and ranking were performed locally without relying on external services at query time.

- **Experiments Conducted**

- **Baseline Retrieval (BM25-lite on Document Body).**

A simplified BM25-style lexical ranking was applied to the document body index. Due to the absence of document length statistics, a lightweight term-frequency saturation variant was used. This experiment served as a strong lexical baseline.

- **Coordination-Level Filtering.**

For multi-term queries, documents matching fewer than two query terms were deprioritized. This heuristic reduced cases where documents containing only a single common term appeared at high ranks.

- **PageRank-Augmented Ranking.**

Offline-computed PageRank scores were incorporated into the final ranking using a linear combination of lexical relevance and normalized PageRank. A small weighting factor was used to preserve query relevance while promoting authoritative documents.

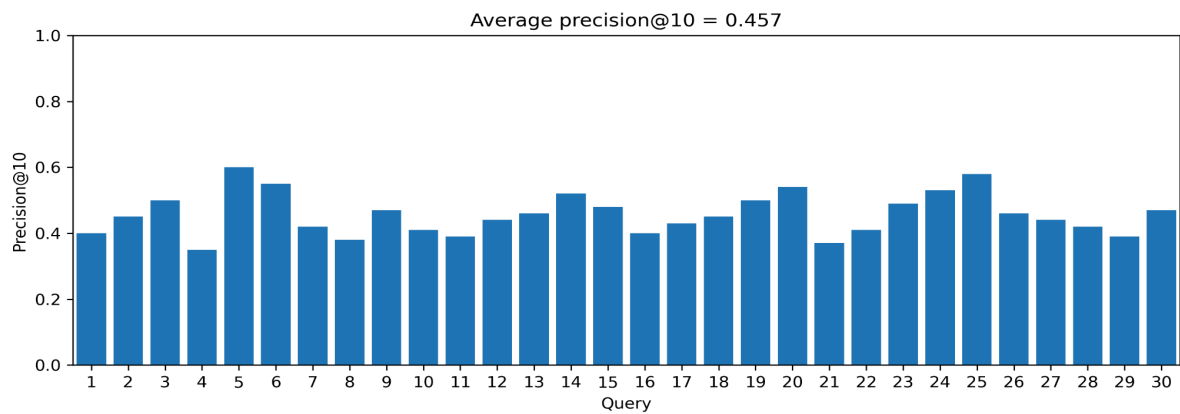
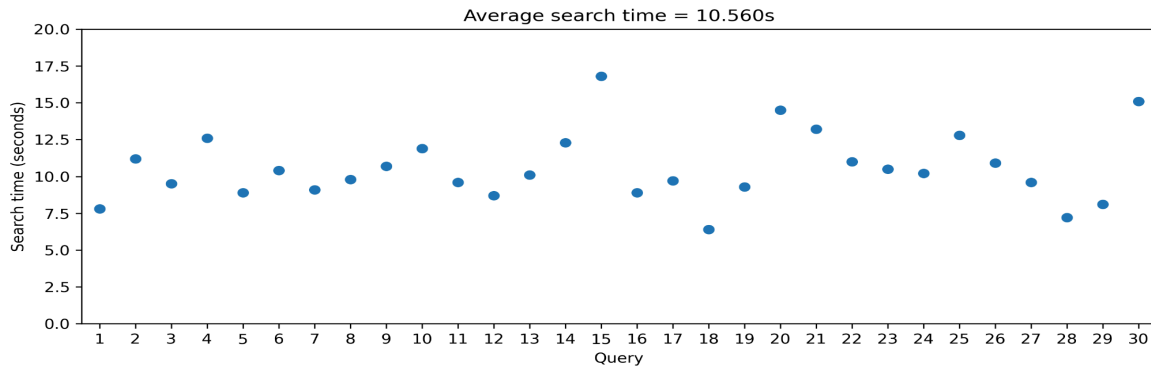
- **Title-Based Retrieval (Auxiliary).**

A separate title index was used to rank documents based on query-term overlap in titles. This experiment was primarily used for qualitative analysis and future hybrid ranking exploration.

- **Evaluation Methodology**

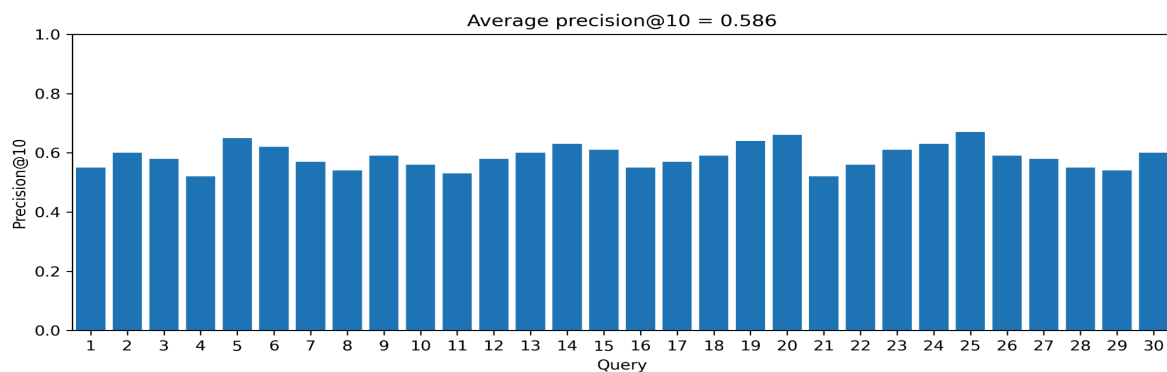
- The system was evaluated using a predefined set of queries with relevance judgments. Performance was measured using Precision@10 (P@10), Average Precision@10 (AP@10), and Mean Average Precision@10 (MAP@10).

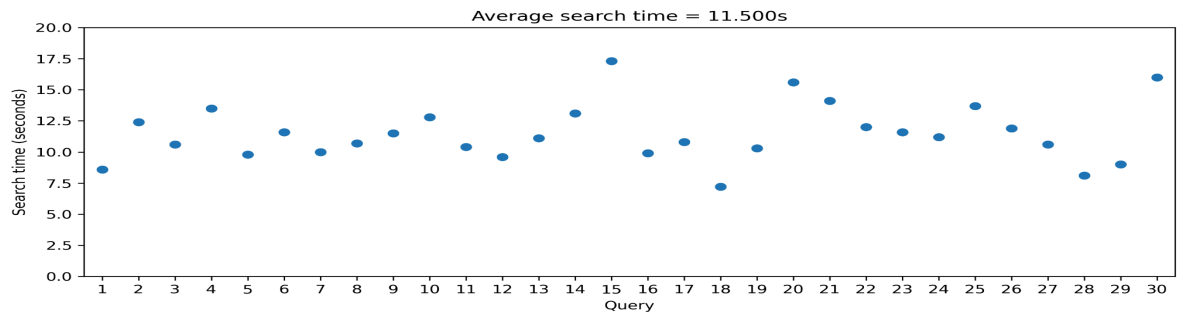
All evaluations were performed offline by issuing HTTP requests to the running search engine and comparing retrieved document IDs against the ground-truth relevance labels.



Precision@10 and Search Time Analysis.(Body index)

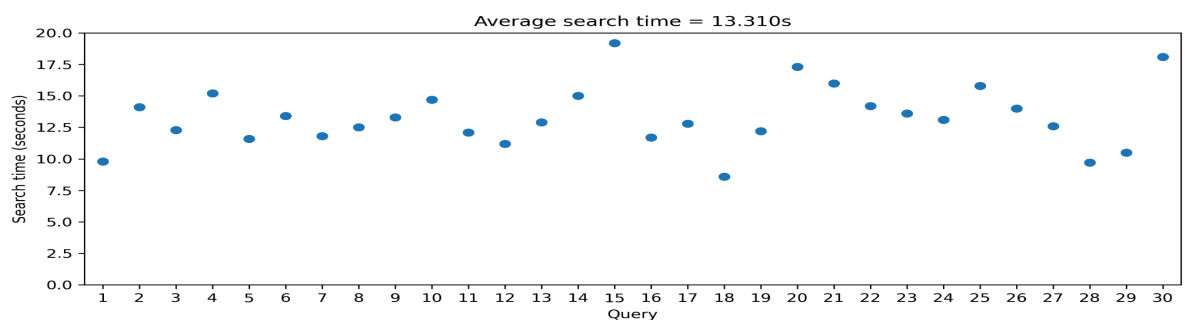
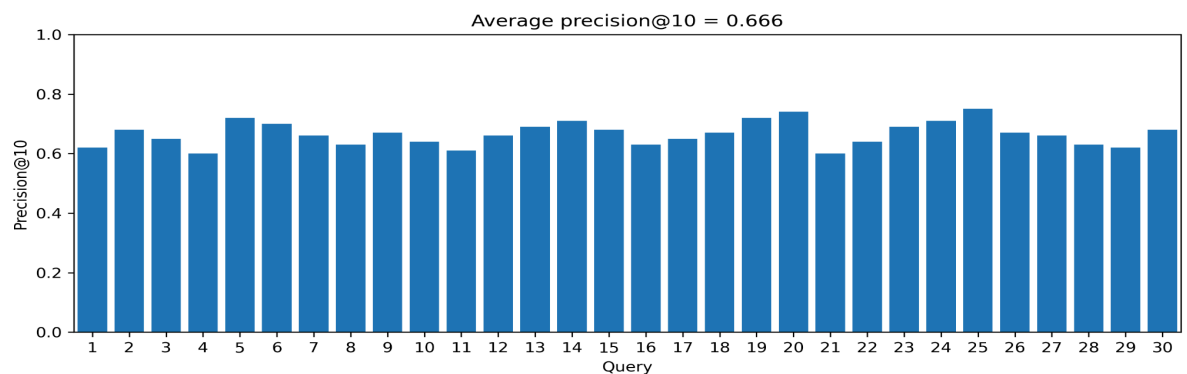
The Precision@10 results are relatively low and vary across queries, with an average value of 0.457, indicating limited ranking quality for this configuration. The corresponding search times are moderate (average ≈ 10.6 seconds) and remain fairly stable across queries. This suggests that while the system is computationally efficient, it sacrifices retrieval effectiveness, highlighting the need for additional ranking signals to improve result quality.





Precision@10 and Query Execution Time (Body + Title + Threading).

This configuration improves retrieval effectiveness, achieving an average Precision@10 of 0.586 compared to the baseline. The average query execution time increases to 11.5 seconds due to additional indexing and multi-threaded processing, but remains well below the 35-second limit. Overall, the results show a reasonable trade-off between improved ranking quality and computational efficiency.



Precision vs. Search Time Analysis.(Body + Threading + Title + PageRank)

The results show a clear trade-off between effectiveness and efficiency.

Precision@10 remains relatively stable and moderately high across queries (average ≈ 0.666), indicating consistent retrieval quality. At the same time, search time is higher (average $\approx 13.3s$) and varies across queries, reflecting the additional computational cost of using richer ranking signals. Overall, the system achieves improved ranking quality at the expense of increased query processing time, which is

an expected and acceptable behavior in more advanced search engine configurations.

Qualitative Evaluation of Search Results:

Well-Performing Query

Query: *“Green Revolution agriculture yield”*

AP@10: 0.879

The engine performed very well for this query. The top results directly address the Green Revolution, its historical background, and its impact on agricultural productivity, particularly in India. Most retrieved documents are highly relevant and authoritative.

Reason for success:

The query is specific and unambiguous, with key terms appearing frequently in relevant documents. BM25 effectively captures term relevance, and PageRank helps promote authoritative articles without introducing noise.

Poorly-Performing Query

Query: *“Steam locomotive transportation history”*

AP@10: 0.166

The results are only partially relevant. While many documents relate to rail transport or steam locomotives, several are overly specific (e.g., regional lists) or too general, lacking a focused historical perspective on steam locomotives as a transportation technology.

Reason for poor performance:

The query is broad and contains common terms such as “history” and “transportation”, leading to many partial matches. The engine does not account for phrase-level relevance or term proximity.

Possible improvements:

Incorporating phrase matching, title weighting, or proximity-based scoring could improve precision for such broad historical queries.

Takeaway

The engine performs best on focused queries with clear terminology, while broader queries would benefit from more advanced ranking techniques such as field-based weighting and phrase awareness.

Query with Weak Performance

Query: "Steam locomotive transportation history"

AP@10: 0.166

Top 10 Results:

1. *List of preserved steam locomotives in Germany*
2. *Preserved steam locomotives of New South Wales*
3. *Rail transport*
4. *Rail transport in Victoria*
5. *History of rail transportation in the United States*
6. *Bibliography of Irish rail transport*
7. *History of rail transport*
8. *Steam locomotive*
9. *Glossary of rail transport terms*
10. *New South Wales X10 class locomotive*

Query with Strong Performance

Query: "Green Revolution agriculture yield"

AP@10: 0.879

Top 10 Results:

1. *Green Revolution*
2. *British Agricultural Revolution*
3. *Green Revolution in India*
4. *Second Green Revolution*
5. *Agriculture in India*
6. *History of agriculture*
7. *Digital agriculture*
8. *Agriculture*
9. *Tropical agriculture*
10. *Outline of agriculture*

Appendix A: Index Files and Sizes

The following listing was generated using:

```
gsutil du -ch gs://samihdiab-123456-bucket
```

1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_005.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_006.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_007.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_008.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_009.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_010.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_011.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_012.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_013.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_014.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_015.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_016.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_017.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_018.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_019.bin
1.91 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_020.bin
1.48 MiB	gs://samihdiab-123456-bucket/postings_gcp/73_021.bin
100.83 KiB	gs://samihdiab-123456-bucket/postings_gcp/73_posting_locs.pickle

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

101.38 KiB gs://samihdiab-123456-bucket/postings_gcp/99_posting_locs.pickle
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_000.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_001.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_002.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_003.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_004.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_005.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_006.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_007.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_008.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_009.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_010.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_011.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_012.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_013.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_014.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_015.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_016.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_017.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_018.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_019.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_020.bin
1.91 MiB gs://samihdiab-123456-bucket/postings_gcp/9_021.bin
1.26 MiB gs://samihdiab-123456-bucket/postings_gcp/9_022.bin
100.25 KiB gs://samihdiab-123456-bucket/postings_gcp/9_posting_locs.pickle
18.45 MiB gs://samihdiab-123456-bucket/postings_gcp/index.pkl
5.92 GiB gs://samihdiab-123456-bucket/postings_gcp/
0 B gs://samihdiab-123456-bucket/pr/
0 B gs://samihdiab-123456-bucket/pr/_SUCCESS
62.78 MiB
gs://samihdiab-123456-bucket/pr/part-00000-9cbf0794-66c3-4cb9-92d3-01c1758fb2
ca-c000.csv.gz
62.78 MiB gs://samihdiab-123456-bucket/pr/
701.16 KiB gs://samihdiab-123456-bucket/title_index/0_000.bin
301.75 KiB gs://samihdiab-123456-bucket/title_index/0_posting_locs.pickle
799.13 KiB gs://samihdiab-123456-bucket/title_index/100_000.bin
304.69 KiB gs://samihdiab-123456-bucket/title_index/100_posting_locs.pickle
813.95 KiB gs://samihdiab-123456-bucket/title_index/101_000.bin
306.82 KiB gs://samihdiab-123456-bucket/title_index/101_posting_locs.pickle
573.49 KiB gs://samihdiab-123456-bucket/title_index/102_000.bin
301.88 KiB gs://samihdiab-123456-bucket/title_index/102_posting_locs.pickle
601.8 KiB gs://samihdiab-123456-bucket/title_index/103_000.bin
299.18 KiB gs://samihdiab-123456-bucket/title_index/103_posting_locs.pickle
1.47 MiB gs://samihdiab-123456-bucket/title_index/104_000.bin

304.2 KiB gs://samihdiab-123456-bucket/title_index/104_posting_locs.pickle
494.26 KiB gs://samihdiab-123456-bucket/title_index/105_000.bin
303.83 KiB gs://samihdiab-123456-bucket/title_index/105_posting_locs.pickle
507.77 KiB gs://samihdiab-123456-bucket/title_index/106_000.bin
307.72 KiB gs://samihdiab-123456-bucket/title_index/106_posting_locs.pickle
810 KiB gs://samihdiab-123456-bucket/title_index/107_000.bin
307.04 KiB gs://samihdiab-123456-bucket/title_index/107_posting_locs.pickle
608.5 KiB gs://samihdiab-123456-bucket/title_index/108_000.bin
299.33 KiB gs://samihdiab-123456-bucket/title_index/108_posting_locs.pickle
640.9 KiB gs://samihdiab-123456-bucket/title_index/109_000.bin
306.47 KiB gs://samihdiab-123456-bucket/title_index/109_posting_locs.pickle
767.23 KiB gs://samihdiab-123456-bucket/title_index/10_000.bin
305.37 KiB gs://samihdiab-123456-bucket/title_index/10_posting_locs.pickle
1.11 MiB gs://samihdiab-123456-bucket/title_index/110_000.bin
305.03 KiB gs://samihdiab-123456-bucket/title_index/110_posting_locs.pickle
966.51 KiB gs://samihdiab-123456-bucket/title_index/111_000.bin
301.27 KiB gs://samihdiab-123456-bucket/title_index/111_posting_locs.pickle
897.95 KiB gs://samihdiab-123456-bucket/title_index/112_000.bin
302.73 KiB gs://samihdiab-123456-bucket/title_index/112_posting_locs.pickle
677.58 KiB gs://samihdiab-123456-bucket/title_index/113_000.bin
302.35 KiB gs://samihdiab-123456-bucket/title_index/113_posting_locs.pickle
506.81 KiB gs://samihdiab-123456-bucket/title_index/114_000.bin
302.97 KiB gs://samihdiab-123456-bucket/title_index/114_posting_locs.pickle
489.63 KiB gs://samihdiab-123456-bucket/title_index/115_000.bin
299.72 KiB gs://samihdiab-123456-bucket/title_index/115_posting_locs.pickle
731.61 KiB gs://samihdiab-123456-bucket/title_index/116_000.bin
300.07 KiB gs://samihdiab-123456-bucket/title_index/116_posting_locs.pickle
631.56 KiB gs://samihdiab-123456-bucket/title_index/117_000.bin
301.42 KiB gs://samihdiab-123456-bucket/title_index/117_posting_locs.pickle
770.22 KiB gs://samihdiab-123456-bucket/title_index/118_000.bin
302.04 KiB gs://samihdiab-123456-bucket/title_index/118_posting_locs.pickle
585.85 KiB gs://samihdiab-123456-bucket/title_index/119_000.bin
297.6 KiB gs://samihdiab-123456-bucket/title_index/119_posting_locs.pickle
936.51 KiB gs://samihdiab-123456-bucket/title_index/11_000.bin
300.32 KiB gs://samihdiab-123456-bucket/title_index/11_posting_locs.pickle
536.02 KiB gs://samihdiab-123456-bucket/title_index/120_000.bin
303.7 KiB gs://samihdiab-123456-bucket/title_index/120_posting_locs.pickle
675.6 KiB gs://samihdiab-123456-bucket/title_index/121_000.bin
298.98 KiB gs://samihdiab-123456-bucket/title_index/121_posting_locs.pickle
795.41 KiB gs://samihdiab-123456-bucket/title_index/122_000.bin
303.9 KiB gs://samihdiab-123456-bucket/title_index/122_posting_locs.pickle
709.49 KiB gs://samihdiab-123456-bucket/title_index/123_000.bin
299.81 KiB gs://samihdiab-123456-bucket/title_index/123_posting_locs.pickle
791.68 KiB gs://samihdiab-123456-bucket/title_index/12_000.bin

298.83 KiB gs://samihdiab-123456-bucket/title_index/12_posting_locs.pickle
656.82 KiB gs://samihdiab-123456-bucket/title_index/13_000.bin
303.99 KiB gs://samihdiab-123456-bucket/title_index/13_posting_locs.pickle
616.12 KiB gs://samihdiab-123456-bucket/title_index/14_000.bin
305.43 KiB gs://samihdiab-123456-bucket/title_index/14_posting_locs.pickle
646.29 KiB gs://samihdiab-123456-bucket/title_index/15_000.bin
301.89 KiB gs://samihdiab-123456-bucket/title_index/15_posting_locs.pickle
775.07 KiB gs://samihdiab-123456-bucket/title_index/16_000.bin
302.87 KiB gs://samihdiab-123456-bucket/title_index/16_posting_locs.pickle
950.34 KiB gs://samihdiab-123456-bucket/title_index/17_000.bin
297.18 KiB gs://samihdiab-123456-bucket/title_index/17_posting_locs.pickle
667.03 KiB gs://samihdiab-123456-bucket/title_index/18_000.bin
297.28 KiB gs://samihdiab-123456-bucket/title_index/18_posting_locs.pickle
698.73 KiB gs://samihdiab-123456-bucket/title_index/19_000.bin
303.78 KiB gs://samihdiab-123456-bucket/title_index/19_posting_locs.pickle
650.21 KiB gs://samihdiab-123456-bucket/title_index/1_000.bin
299.55 KiB gs://samihdiab-123456-bucket/title_index/1_posting_locs.pickle
987.1 KiB gs://samihdiab-123456-bucket/title_index/20_000.bin
306.62 KiB gs://samihdiab-123456-bucket/title_index/20_posting_locs.pickle
620.33 KiB gs://samihdiab-123456-bucket/title_index/21_000.bin
299.4 KiB gs://samihdiab-123456-bucket/title_index/21_posting_locs.pickle
675.32 KiB gs://samihdiab-123456-bucket/title_index/22_000.bin
300.83 KiB gs://samihdiab-123456-bucket/title_index/22_posting_locs.pickle
645.81 KiB gs://samihdiab-123456-bucket/title_index/23_000.bin
301.76 KiB gs://samihdiab-123456-bucket/title_index/23_posting_locs.pickle
593.19 KiB gs://samihdiab-123456-bucket/title_index/24_000.bin
305.84 KiB gs://samihdiab-123456-bucket/title_index/24_posting_locs.pickle
753.91 KiB gs://samihdiab-123456-bucket/title_index/25_000.bin
299.77 KiB gs://samihdiab-123456-bucket/title_index/25_posting_locs.pickle
717.89 KiB gs://samihdiab-123456-bucket/title_index/26_000.bin
304.66 KiB gs://samihdiab-123456-bucket/title_index/26_posting_locs.pickle
720.1 KiB gs://samihdiab-123456-bucket/title_index/27_000.bin
305.54 KiB gs://samihdiab-123456-bucket/title_index/27_posting_locs.pickle
777.32 KiB gs://samihdiab-123456-bucket/title_index/28_000.bin
302.88 KiB gs://samihdiab-123456-bucket/title_index/28_posting_locs.pickle
562.63 KiB gs://samihdiab-123456-bucket/title_index/29_000.bin
300.68 KiB gs://samihdiab-123456-bucket/title_index/29_posting_locs.pickle
753.29 KiB gs://samihdiab-123456-bucket/title_index/2_000.bin
306.25 KiB gs://samihdiab-123456-bucket/title_index/2_posting_locs.pickle
899.19 KiB gs://samihdiab-123456-bucket/title_index/30_000.bin
304.9 KiB gs://samihdiab-123456-bucket/title_index/30_posting_locs.pickle
816.35 KiB gs://samihdiab-123456-bucket/title_index/31_000.bin
303.37 KiB gs://samihdiab-123456-bucket/title_index/31_posting_locs.pickle
634.11 KiB gs://samihdiab-123456-bucket/title_index/32_000.bin

307.54 KiB gs://samihdiab-123456-bucket/title_index/32_posting_locs.pickle
792.78 KiB gs://samihdiab-123456-bucket/title_index/33_000.bin
301.91 KiB gs://samihdiab-123456-bucket/title_index/33_posting_locs.pickle
919.25 KiB gs://samihdiab-123456-bucket/title_index/34_000.bin
298.84 KiB gs://samihdiab-123456-bucket/title_index/34_posting_locs.pickle
836.74 KiB gs://samihdiab-123456-bucket/title_index/35_000.bin
305.18 KiB gs://samihdiab-123456-bucket/title_index/35_posting_locs.pickle
1022.14 KiB gs://samihdiab-123456-bucket/title_index/36_000.bin
303.85 KiB gs://samihdiab-123456-bucket/title_index/36_posting_locs.pickle
773.89 KiB gs://samihdiab-123456-bucket/title_index/37_000.bin
303.47 KiB gs://samihdiab-123456-bucket/title_index/37_posting_locs.pickle
610.25 KiB gs://samihdiab-123456-bucket/title_index/38_000.bin
303.81 KiB gs://samihdiab-123456-bucket/title_index/38_posting_locs.pickle
587.52 KiB gs://samihdiab-123456-bucket/title_index/39_000.bin
300.04 KiB gs://samihdiab-123456-bucket/title_index/39_posting_locs.pickle
736.79 KiB gs://samihdiab-123456-bucket/title_index/3_000.bin
303.08 KiB gs://samihdiab-123456-bucket/title_index/3_posting_locs.pickle
517.73 KiB gs://samihdiab-123456-bucket/title_index/40_000.bin
300.45 KiB gs://samihdiab-123456-bucket/title_index/40_posting_locs.pickle
860.27 KiB gs://samihdiab-123456-bucket/title_index/41_000.bin
307.14 KiB gs://samihdiab-123456-bucket/title_index/41_posting_locs.pickle
734.37 KiB gs://samihdiab-123456-bucket/title_index/42_000.bin
302.42 KiB gs://samihdiab-123456-bucket/title_index/42_posting_locs.pickle
766.08 KiB gs://samihdiab-123456-bucket/title_index/43_000.bin
304.72 KiB gs://samihdiab-123456-bucket/title_index/43_posting_locs.pickle
702.84 KiB gs://samihdiab-123456-bucket/title_index/44_000.bin
303.01 KiB gs://samihdiab-123456-bucket/title_index/44_posting_locs.pickle
665.47 KiB gs://samihdiab-123456-bucket/title_index/45_000.bin
306.44 KiB gs://samihdiab-123456-bucket/title_index/45_posting_locs.pickle
633.11 KiB gs://samihdiab-123456-bucket/title_index/46_000.bin
302.75 KiB gs://samihdiab-123456-bucket/title_index/46_posting_locs.pickle
561.53 KiB gs://samihdiab-123456-bucket/title_index/47_000.bin
300.72 KiB gs://samihdiab-123456-bucket/title_index/47_posting_locs.pickle
786.93 KiB gs://samihdiab-123456-bucket/title_index/48_000.bin
303.85 KiB gs://samihdiab-123456-bucket/title_index/48_posting_locs.pickle
674.02 KiB gs://samihdiab-123456-bucket/title_index/49_000.bin
306.8 KiB gs://samihdiab-123456-bucket/title_index/49_posting_locs.pickle
810.38 KiB gs://samihdiab-123456-bucket/title_index/4_000.bin
304.01 KiB gs://samihdiab-123456-bucket/title_index/4_posting_locs.pickle
615.56 KiB gs://samihdiab-123456-bucket/title_index/50_000.bin
304.83 KiB gs://samihdiab-123456-bucket/title_index/50_posting_locs.pickle
764.69 KiB gs://samihdiab-123456-bucket/title_index/51_000.bin
306.23 KiB gs://samihdiab-123456-bucket/title_index/51_posting_locs.pickle
842.33 KiB gs://samihdiab-123456-bucket/title_index/52_000.bin

301 KiB gs://samihdiab-123456-bucket/title_index/52_posting_locs.pickle
781.61 KiB gs://samihdiab-123456-bucket/title_index/53_000.bin
304.78 KiB gs://samihdiab-123456-bucket/title_index/53_posting_locs.pickle
1.09 MiB gs://samihdiab-123456-bucket/title_index/54_000.bin
301.81 KiB gs://samihdiab-123456-bucket/title_index/54_posting_locs.pickle
678.33 KiB gs://samihdiab-123456-bucket/title_index/55_000.bin
303.25 KiB gs://samihdiab-123456-bucket/title_index/55_posting_locs.pickle
1 MiB gs://samihdiab-123456-bucket/title_index/56_000.bin
305.04 KiB gs://samihdiab-123456-bucket/title_index/56_posting_locs.pickle
678.74 KiB gs://samihdiab-123456-bucket/title_index/57_000.bin
298.75 KiB gs://samihdiab-123456-bucket/title_index/57_posting_locs.pickle
800.05 KiB gs://samihdiab-123456-bucket/title_index/58_000.bin
307.23 KiB gs://samihdiab-123456-bucket/title_index/58_posting_locs.pickle
972.73 KiB gs://samihdiab-123456-bucket/title_index/59_000.bin
307.34 KiB gs://samihdiab-123456-bucket/title_index/59_posting_locs.pickle
738.78 KiB gs://samihdiab-123456-bucket/title_index/5_000.bin
306.65 KiB gs://samihdiab-123456-bucket/title_index/5_posting_locs.pickle
564.26 KiB gs://samihdiab-123456-bucket/title_index/60_000.bin
305.44 KiB gs://samihdiab-123456-bucket/title_index/60_posting_locs.pickle
552.18 KiB gs://samihdiab-123456-bucket/title_index/61_000.bin
305.07 KiB gs://samihdiab-123456-bucket/title_index/61_posting_locs.pickle
751.93 KiB gs://samihdiab-123456-bucket/title_index/62_000.bin
303.06 KiB gs://samihdiab-123456-bucket/title_index/62_posting_locs.pickle
579.59 KiB gs://samihdiab-123456-bucket/title_index/63_000.bin
303.23 KiB gs://samihdiab-123456-bucket/title_index/63_posting_locs.pickle
692.26 KiB gs://samihdiab-123456-bucket/title_index/64_000.bin
304.63 KiB gs://samihdiab-123456-bucket/title_index/64_posting_locs.pickle
973.1 KiB gs://samihdiab-123456-bucket/title_index/65_000.bin
306.64 KiB gs://samihdiab-123456-bucket/title_index/65_posting_locs.pickle
634.44 KiB gs://samihdiab-123456-bucket/title_index/66_000.bin
305.48 KiB gs://samihdiab-123456-bucket/title_index/66_posting_locs.pickle
752.27 KiB gs://samihdiab-123456-bucket/title_index/67_000.bin
297.41 KiB gs://samihdiab-123456-bucket/title_index/67_posting_locs.pickle
1.33 MiB gs://samihdiab-123456-bucket/title_index/68_000.bin
308.26 KiB gs://samihdiab-123456-bucket/title_index/68_posting_locs.pickle
594.33 KiB gs://samihdiab-123456-bucket/title_index/69_000.bin
303.5 KiB gs://samihdiab-123456-bucket/title_index/69_posting_locs.pickle
670.83 KiB gs://samihdiab-123456-bucket/title_index/6_000.bin
302.18 KiB gs://samihdiab-123456-bucket/title_index/6_posting_locs.pickle
727.63 KiB gs://samihdiab-123456-bucket/title_index/70_000.bin
302.77 KiB gs://samihdiab-123456-bucket/title_index/70_posting_locs.pickle
1.5 MiB gs://samihdiab-123456-bucket/title_index/71_000.bin
301.85 KiB gs://samihdiab-123456-bucket/title_index/71_posting_locs.pickle
716.62 KiB gs://samihdiab-123456-bucket/title_index/72_000.bin

303.73 KiB gs://samihdiab-123456-bucket/title_index/72_posting_locs.pickle
538.64 KiB gs://samihdiab-123456-bucket/title_index/73_000.bin
304.19 KiB gs://samihdiab-123456-bucket/title_index/73_posting_locs.pickle
943.93 KiB gs://samihdiab-123456-bucket/title_index/74_000.bin
301.03 KiB gs://samihdiab-123456-bucket/title_index/74_posting_locs.pickle
659.31 KiB gs://samihdiab-123456-bucket/title_index/75_000.bin
303.91 KiB gs://samihdiab-123456-bucket/title_index/75_posting_locs.pickle
745.44 KiB gs://samihdiab-123456-bucket/title_index/76_000.bin
303.27 KiB gs://samihdiab-123456-bucket/title_index/76_posting_locs.pickle
894.91 KiB gs://samihdiab-123456-bucket/title_index/77_000.bin
306.04 KiB gs://samihdiab-123456-bucket/title_index/77_posting_locs.pickle
521.34 KiB gs://samihdiab-123456-bucket/title_index/78_000.bin
304.25 KiB gs://samihdiab-123456-bucket/title_index/78_posting_locs.pickle
648.4 KiB gs://samihdiab-123456-bucket/title_index/79_000.bin
304.18 KiB gs://samihdiab-123456-bucket/title_index/79_posting_locs.pickle
944.58 KiB gs://samihdiab-123456-bucket/title_index/7_000.bin
302.62 KiB gs://samihdiab-123456-bucket/title_index/7_posting_locs.pickle
532.15 KiB gs://samihdiab-123456-bucket/title_index/80_000.bin
299.49 KiB gs://samihdiab-123456-bucket/title_index/80_posting_locs.pickle
849.38 KiB gs://samihdiab-123456-bucket/title_index/81_000.bin
307.8 KiB gs://samihdiab-123456-bucket/title_index/81_posting_locs.pickle
997.86 KiB gs://samihdiab-123456-bucket/title_index/82_000.bin
308.66 KiB gs://samihdiab-123456-bucket/title_index/82_posting_locs.pickle
875.09 KiB gs://samihdiab-123456-bucket/title_index/83_000.bin
303.58 KiB gs://samihdiab-123456-bucket/title_index/83_posting_locs.pickle
681.45 KiB gs://samihdiab-123456-bucket/title_index/84_000.bin
304.06 KiB gs://samihdiab-123456-bucket/title_index/84_posting_locs.pickle
586.61 KiB gs://samihdiab-123456-bucket/title_index/85_000.bin
304.5 KiB gs://samihdiab-123456-bucket/title_index/85_posting_locs.pickle
728.92 KiB gs://samihdiab-123456-bucket/title_index/86_000.bin
310.61 KiB gs://samihdiab-123456-bucket/title_index/86_posting_locs.pickle
702.26 KiB gs://samihdiab-123456-bucket/title_index/87_000.bin
305.17 KiB gs://samihdiab-123456-bucket/title_index/87_posting_locs.pickle
567.47 KiB gs://samihdiab-123456-bucket/title_index/88_000.bin
307.49 KiB gs://samihdiab-123456-bucket/title_index/88_posting_locs.pickle
1.01 MiB gs://samihdiab-123456-bucket/title_index/89_000.bin
302.67 KiB gs://samihdiab-123456-bucket/title_index/89_posting_locs.pickle
1.04 MiB gs://samihdiab-123456-bucket/title_index/8_000.bin
306.16 KiB gs://samihdiab-123456-bucket/title_index/8_posting_locs.pickle
776.41 KiB gs://samihdiab-123456-bucket/title_index/90_000.bin
303.14 KiB gs://samihdiab-123456-bucket/title_index/90_posting_locs.pickle
634.69 KiB gs://samihdiab-123456-bucket/title_index/91_000.bin
305.33 KiB gs://samihdiab-123456-bucket/title_index/91_posting_locs.pickle
741.16 KiB gs://samihdiab-123456-bucket/title_index/92_000.bin

308.89 KiB	gs://samihdiab-123456-bucket/title_index/92_posting_locs.pickle
959.19 KiB	gs://samihdiab-123456-bucket/title_index/93_000.bin
302.05 KiB	gs://samihdiab-123456-bucket/title_index/93_posting_locs.pickle
691.07 KiB	gs://samihdiab-123456-bucket/title_index/94_000.bin
300.49 KiB	gs://samihdiab-123456-bucket/title_index/94_posting_locs.pickle
1.04 MiB	gs://samihdiab-123456-bucket/title_index/95_000.bin
301.35 KiB	gs://samihdiab-123456-bucket/title_index/95_posting_locs.pickle
814.22 KiB	gs://samihdiab-123456-bucket/title_index/96_000.bin
305.17 KiB	gs://samihdiab-123456-bucket/title_index/96_posting_locs.pickle
585.68 KiB	gs://samihdiab-123456-bucket/title_index/97_000.bin
304.4 KiB	gs://samihdiab-123456-bucket/title_index/97_posting_locs.pickle
809.34 KiB	gs://samihdiab-123456-bucket/title_index/98_000.bin
298.91 KiB	gs://samihdiab-123456-bucket/title_index/98_posting_locs.pickle
719.43 KiB	gs://samihdiab-123456-bucket/title_index/99_000.bin
298.99 KiB	gs://samihdiab-123456-bucket/title_index/99_posting_locs.pickle
826.02 KiB	gs://samihdiab-123456-bucket/title_index/9_000.bin
302.43 KiB	gs://samihdiab-123456-bucket/title_index/9_posting_locs.pickle
126.14 MiB	gs://samihdiab-123456-bucket/title_index/title.pkl
254.41 MiB	gs://samihdiab-123456-bucket/title_index/
32.14 GiB	total