HomeWork-5 Report

1. Methodology
a. Obtaining Queries
- Description of the Assigned Queries

Our queries are: "Brezhnev Doctrine", "Sino-Soviet split", "Cuban Missile Crisis", "Poland Solidarity". Mine is "Sino-Soviet split". These queries are related to significant events in Cold War history, so my vertical search engine focused on historical documents or educational resources.

- Vertical Search Engine

My Elasticsearch queries use a multi_match query, which searches across multiple fields ('title' and 'text'). This approach can catch references where the event or concept might be mentioned differently in the text or title.

The ranking functionality, represented by the get_rank_list method in the ES class, retrieves documents based on their relevance scores, which are crucial for presenting the most relevant documents first. Also implemented highlighting in the Flask app to improve the usability of search results.

b. Assessment Graphical Interface

The class supports various metrics at different cutoff points (e.g., precision@k, recall@k, F1@k for k=5, 10, 20, 50, 100), average precision, R-precision, and nDCG. Also in Script.js, use submitAssessments Function collects the assessments of the search results (non-relevant, relevant, or very relevant) and sends them to the /assess endpoint.
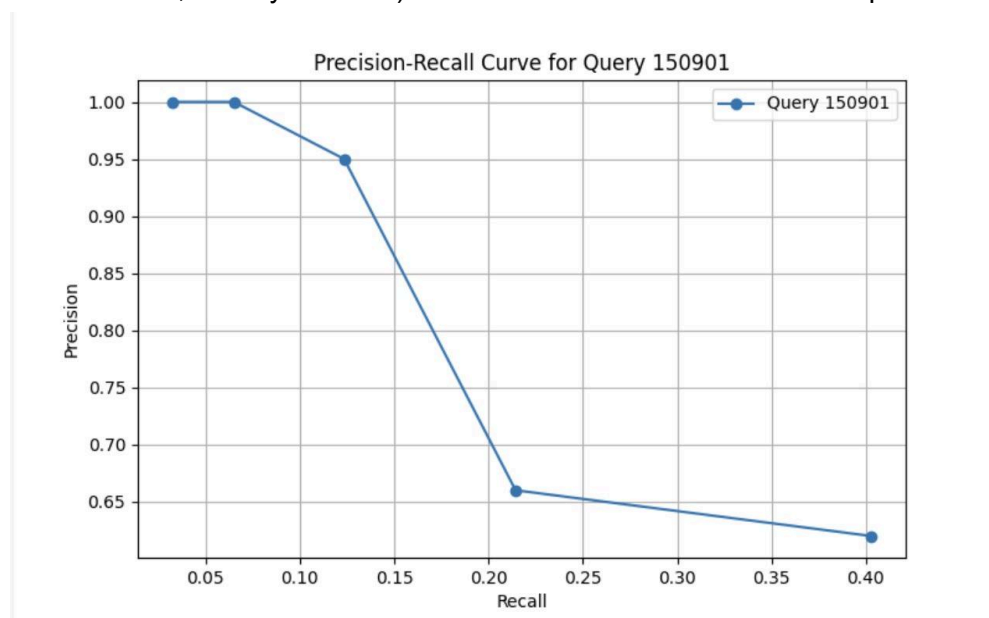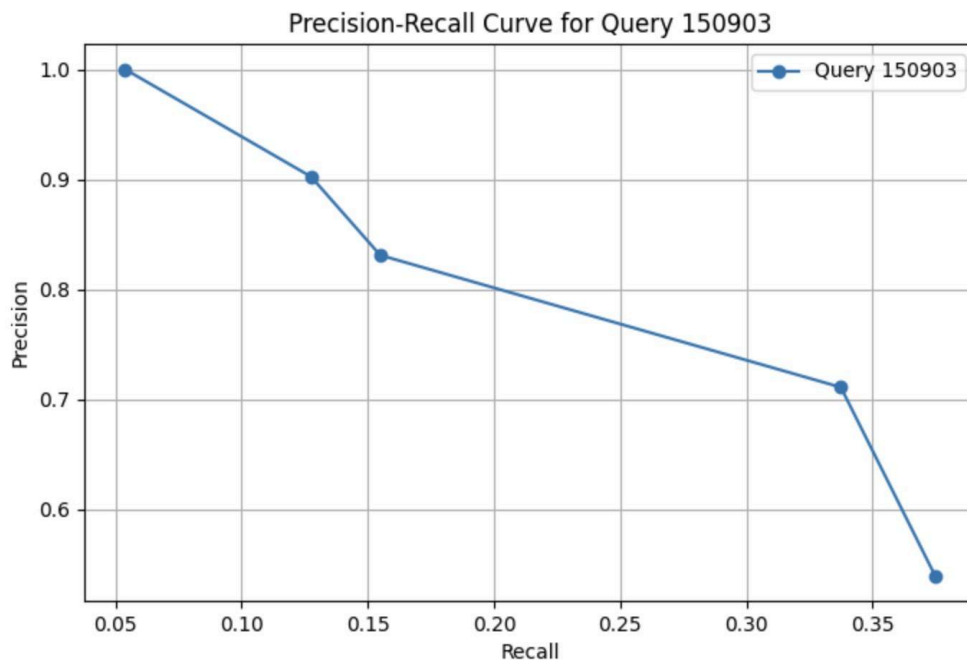

- Precision-Recall Curve for Query 150901 graph shows a strong start in precision at low recall levels, which indicates that the first documents retrieved are highly relevant. There's a steep decline as recall increases, suggesting that as the system tries to

retrieve more relevant documents, it also retrieves more non-relevant ones, reducing precision.

Precision-Recall Curve for Query 150902
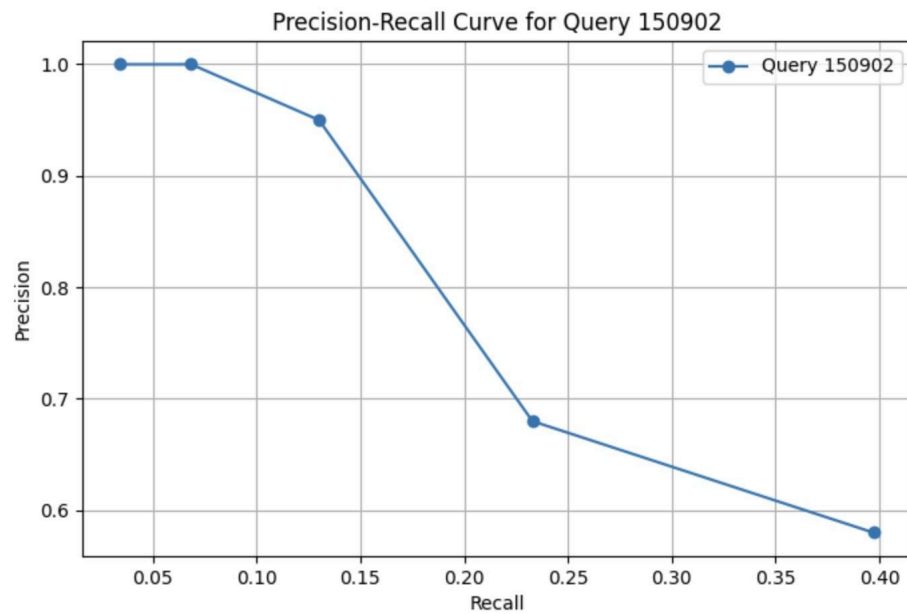
-
- Precision-Recall Curve for Query 150902 graph similar to the first, there's high precision at the outset, which drops off as recall grows. The decrease in precision is more gradual compared to Query 150901, implying a better balance between retrieving relevant documents and avoiding non-relevant ones.

Precision-Recall Curve for Query 150903

-

- Precision-Recall Curve for Query 150903 graph starts off with high precision, but there's a noticeable dip early as recall increases. The precision reduces at a moderate rate, indicating a fair performance of the retrieval system for this query.
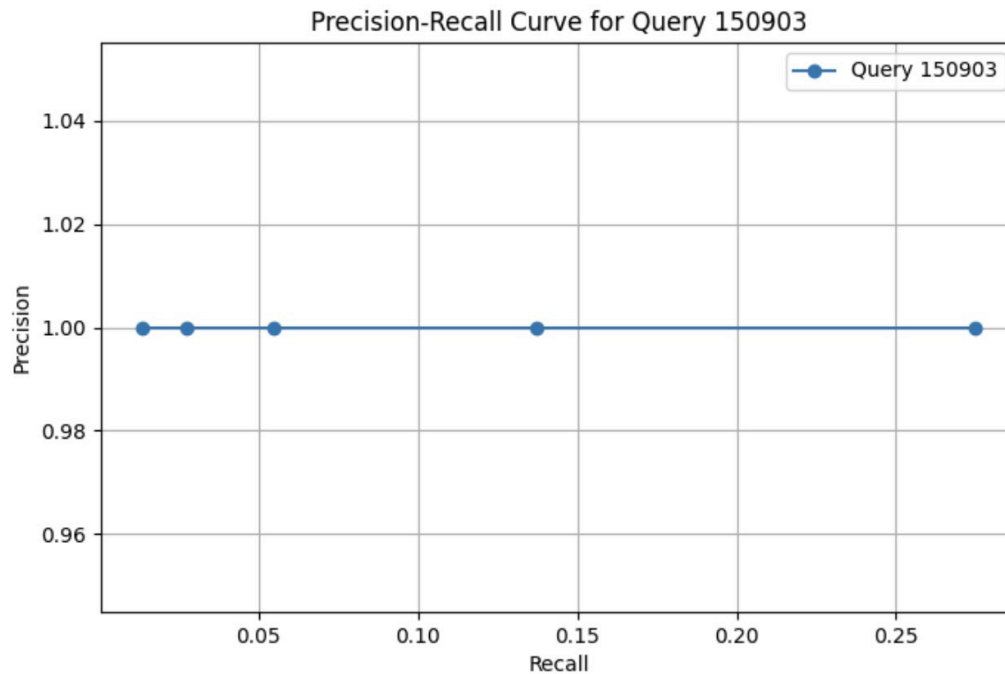
-
- The Precision-Recall Curve for Query 150904 graph is more steady than the previous queries, and it declines at a steady rate as recall.

c. Manual Assessments

The assessor must decide how well the document meets the information needed expressed by the query by clicking inside the url to browse the content.

Title: Democracia cristiana - Wikipedia, la enciclopedia libre
Snippet: influencia en part de europa y latinoamérica lo movimiento son círculo puro de pensamiento y de doctrina *sino*
Url: https://es.wikipedia.org/wiki/Democracia_cristiana

◉ Non-Relevant ○ Relevant ○ Very Relevant

Title: Crisis de los misiles en Cuba | Portal Académico del CCH
Snippet: que constituía una amenaza en la región sólo como difusor de lo ideal de la revolución proletaria *sino*
Url: https://e1.portalacademico.cch.unam.mx/alumno/historiauniversal2/unidad3/principales-conflictos-de-la-guerra-fria/crisis-de-los-misiles

◉ Non-Relevant ○ Relevant ○ Very Relevant

Title: Bolsheviks - Wikipedia
Snippet: большинство bolshinstvo led vladimir lenin faction marxist russian social democrat labour parti rsdlp *split*
Url: https://en.wikipedia.org/wiki/Bolshevik

◉ Non-Relevant ○ Relevant ○ Very Relevant

Title: Veintiocho bolcheviques - Wikipedia, la enciclopedia libre
Snippet: varían según la fuent tampoco est nombr designa un cuerpo organizado de lídere trabajando conjuntament *sino*
Url: https://es.wikipedia.org/wiki/Veintiocho_bolcheviques

◉ Non-Relevant ○ Relevant ○ Very Relevant

Title: República Socialista Soviética de Lituania - Wikipedia, la enciclopedia libre
Snippet: independencia de lituania reconoció que fueron la fuerza soviética la culpabl que provocaron aquella muert *sino*
Url: https://es.wikipedia.org/wiki/Rep%C3%BAblica_Socialista_Sovi%C3%A9tica_de_Lituania

◉ Non-Relevant ○ Relevant ○ Very Relevant

Title: Informbiro period - Wikipedia
Snippet: articl talk languag watch edit redirect informbiro informbiro period era yugoslavia histori follow *split*
Url: https://en.wikipedia.org/wiki/Informbiro

◉ Non-Relevant ○ Relevant ○ Very Relevant

```
≡ assessments.qrel
1    150902 chenxuan_xu https://en.wikipedia.org/wiki/Sino%E2%80%93Soviet_split 2
2    150902 chenxuan_xu https://www.wikidata.org/wiki/Q506079 2
3    150902 chenxuan_xu http://en.wikipedia.org/wiki/Sino-Soviet_split 2
4    150902 chenxuan_xu http://en.wikipedia.org/wiki/Sino%E2%80%93Soviet_split 2
5    150902 chenxuan_xu http://en.wikipedia.org/wiki/Template_talk:Campaignbox_Sino-Soviet_Split 1
6    150902 chenxuan_xu http://en.wikipedia.org/wiki/Template:Campaignbox_Sino-Soviet_Split 1
7    150902 chenxuan_xu http://en.wikipedia.org/wiki/Talk:Sino-Soviet_relations 1
8    150902 chenxuan_xu https://www.thoughtco.com/the-sino-soviet-split-195455 2
9    150902 chenxuan_xu http://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates/Sino-Sov
10   150902 chenxuan_xu https://doi.org/10.1080%2F14682745.2017.1315923 2
```

2. Implementation of trec_eval
a. Pseudocode

```
# Procedure Calculate_TREC_Evaluation
#   Read the qrel_file into qrel_data
#   Read the trec_file into trec_data
#   Initialize metrics dictionary

#   For each query in trec_data
#     Sort documents by score in descending order
#     Initialize variables: num_relevant_docs, num_retrieved_relevant_docs

#     For each document in sorted documents
#       If document is relevant
#         Increment num_retrieved_relevant_docs
#       Compute precision and recall at this rank
#       If current rank is in cutoffs
#         Store precision and recall in metrics

#     Compute and store average precision, R-precision, and nDCG for the query
#     If print_all_queries is true
#       Print detailed metrics for the query

#   Compute mean of each metric across all queries
#   Print summary of metrics

#   If graph is true
#     Plot Precision-Recall Curve

# End Procedure
```

b. Code snippets

```python
def main(argv):
    try:
        opts, args = getopt.getopt(argv, "hqg", ["help", "query-level", "graph"])
    except getopt.GetoptError:
        print('Usage: trec_eval.py -q -g <qrel_file> <trec_file>')
        sys.exit(2)

    print_all_queries = False
    graph = False
    for opt, arg in opts:
        if opt == '-h':
            print('Usage: trec_eval.py -q -g <qrel_file> <trec_file>')
            sys.exit()
        elif opt in ("-q", "--query-level"):
            print_all_queries = True
        elif opt in ("-g", "--graph"):
            graph = True

    if len(args) != 2:
        print('Usage: trec_eval.py -q -g <qrel_file> <trec_file>')
        sys.exit(2)

    qrel_file, trec_file = args
    evaluator = TrecEval(qrel_file, trec_file, print_all_queries, graph)
    evaluator.calculate_metrics()
```

3. Evaluation Results
   a. Results of Manual Assessments

≡ assessments.qrel M ✕

≡ assessments.qrel

```
 1   150902 chenxuan_xu https://en.wikipedia.org/wiki/Sino%E2%80%93Soviet_split 2
 2   150902 chenxuan_xu https://www.wikidata.org/wiki/Q506079 2
 3   150902 chenxuan_xu http://en.wikipedia.org/wiki/Sino-Soviet_split 2
 4   150902 chenxuan_xu http://en.wikipedia.org/wiki/Sino%E2%80%93Soviet_split 2
 5   150902 chenxuan_xu http://en.wikipedia.org/wiki/Template_talk:Campaignbox_Sino-Soviet_Split 1
 6   150902 chenxuan_xu http://en.wikipedia.org/wiki/Template:Campaignbox_Sino-Soviet_Split 1
 7   150902 chenxuan_xu http://en.wikipedia.org/wiki/Talk:Sino-Soviet_relations 1
 8   150902 chenxuan_xu https://www.thoughtco.com/the-sino-soviet-split-195455 2
 9   150902 chenxuan_xu http://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates/Sino-Soviet_Spl
10   150902 chenxuan_xu https://doi.org/10.1080%2F14682745.2017.1315923 2
11   150902 chenxuan_xu http://en.wikipedia.org/wiki/Wikipedia:Featured_article_removal_candidates/Sino-So
12   150902 chenxuan_xu https://www.marxists.org/history/international/comintern/sino-soviet-split/ 2
13   150902 chenxuan_xu https://www.marxists.org/history/international/comintern/sino-soviet-split/index.h
14   150902 chenxuan_xu http://en.wikipedia.org/wiki/Talk:Albanian%E2%80%93Chinese_split 1
15   150902 chenxuan_xu https://simple.wikipedia.org/wiki/Sino-Soviet_split 2
16   150902 chenxuan_xu https://en.wikipedia.org/wiki/Third_Indochina_War 1
17   150902 chenxuan_xu https://simple.wikipedia.org/wiki/Mao_Zedong 1
18   150902 chenxuan_xu http://en.wikipedia.org/wiki/Third_Indochina_War 0
19   150902 chenxuan_xu https://sv.wikipedia.org/wiki/Kinesisk-sovjetiska_schismen 2
20   150902 chenxuan_xu http://en.wikipedia.org/wiki/Talk:Albanian%E2%80%93Soviet_split 2
```

The manual assessment always goes from a lot of '2' to a lot of '0' since the url shows from the highest score which indicates the highest relevance.

b. trec_eval Results

Query ID: 150901
Precision@5:   1.0000
Recall@5:   0.0325
F1@5:   0.0629
Precision@10:   1.0000
Recall@10:   0.0649
F1@10:   0.1220
Precision@20:   0.9500
Recall@20:   0.1234
F1@20:   0.2184
Precision@50:   0.6600
Recall@50:   0.2143
F1@50:   0.3235
Precision@100:   0.6200
Recall@100:   0.4026
F1@100:   0.4882

Average Precision: 0.6341
R-Precision: 0.6558
nDCG: 0.8255

Query ID: 150902
Precision@5:   1.0000
Recall@5:   0.0342
F1@5:   0.0662
Precision@10:   1.0000
Recall@10:   0.0685
F1@10:   0.1282
Precision@20:   0.9500
Recall@20:   0.1301
F1@20:   0.2289
Precision@50:   0.6800
Recall@50:   0.2329
F1@50:   0.3469

Precision@100:   0.5800
Recall@100:   0.3973
F1@100:   0.4715

Average Precision: 0.4430
R-Precision: 0.5137
nDCG: 0.7728

Query ID: 150903
Precision@5:   1.0000
Recall@5:   0.0137
F1@5:   0.0271
Precision@10:   1.0000
Recall@10:   0.0275
F1@10:   0.0535
Precision@20:   1.0000
Recall@20:   0.0549
F1@20:   0.1042
Precision@50:   1.0000
Recall@50:   0.1374
F1@50:   0.2415
Precision@100:   1.0000
Recall@100:   0.2747
F1@100:   0.4310

Average Precision: 0.4767
R-Precision: 0.4780
nDCG: 0.9109

Query ID: 150904
Precision@5:   1.0000
Recall@5:   0.0625
F1@5:   0.1176
Precision@10:   0.8000
Recall@10:   0.1000
F1@10:   0.1778
Precision@20:   0.7000
Recall@20:   0.1750
F1@20:   0.2800

Precision@50:   0.4800
Recall@50:   0.3000
F1@50:   0.3692
Precision@100:   0.3900
Recall@100:   0.4875
F1@100:   0.4333

Average Precision: 0.3730
R-Precision: 0.4875
nDCG: 0.6213

Mean average_precision: 0.4817
Mean r_precision: 0.5338
Mean ndcg: 0.7826
Mean precision@5:   1.0000
Mean recall@5:   0.0357
Mean f1@5:   0.0685
Mean precision@10:   0.9500
Mean recall@10:   0.0652
Mean f1@10:   0.1204
Mean precision@20:   0.9000
Mean recall@20:   0.1209
Mean f1@20:   0.2079
Mean precision@50:   0.7050
Mean recall@50:   0.2211
Mean f1@50:   0.3203
Mean precision@100:   0.6475
Mean recall@100:   0.3905
Mean f1@100:   0.4560

The system achieved high precision scores across different cutoffs (5, 10, 20, 50, 100), indicating that when it returns results, they are highly relevant. However, recall values are relatively lower, maybe it is because when @5, 10, 20, the relevant docs compare to overall relevant docs is low. The R-Precision is around 0.53 on average, indicating a reasonable level of precision when the recall matches the number of relevant documents. And it achieved a relatively high average precision of around 0.48, which is a

good sign. And also get a relatively high nDCG score of around 0.78, suggesting good ranking performance.

5. Extra Credit
Achieved a detailed and user-friendly interface of vertical search.