

# PRIMED

## A Medicine Search System

Milestone #3: Search System



Pedro Simões, up202403063@up.pt  
Miguel Garrido, up202108889@up.pt  
Emanuel Maia, up202107486@up.pt  
Guilherme Martins, up202403106@up.pt

# Table of Contents

**01**

Milestone #2 Review

**02**

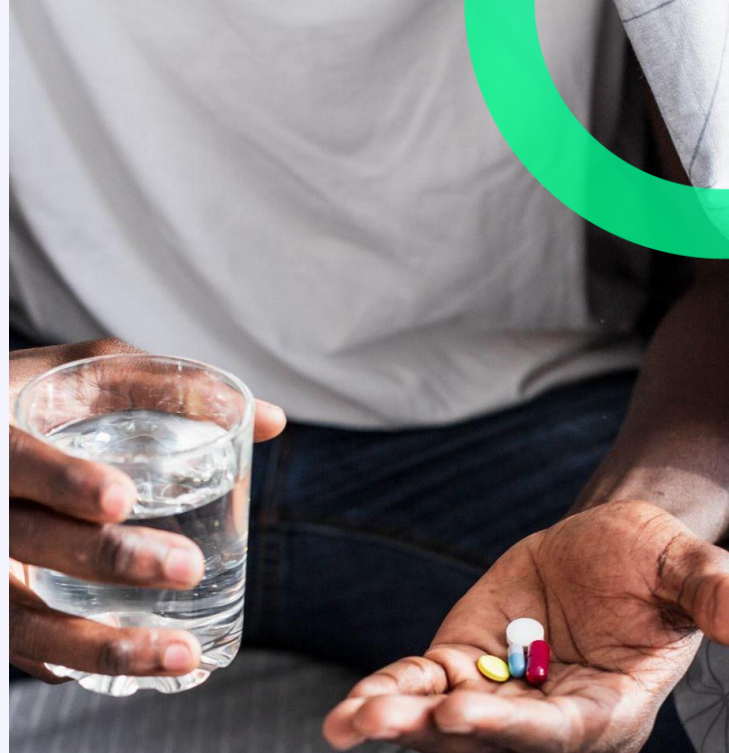
Re-ranking + Polarity

**03**

Semantic Search

**04**

Search System UI





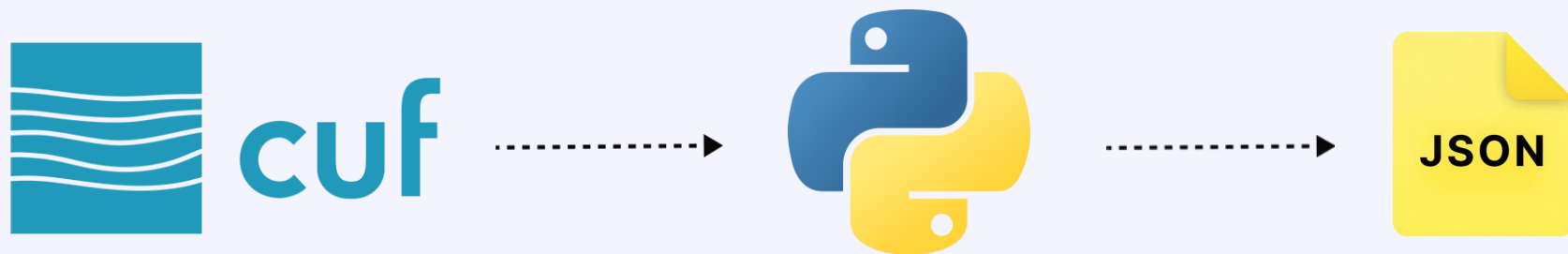
01



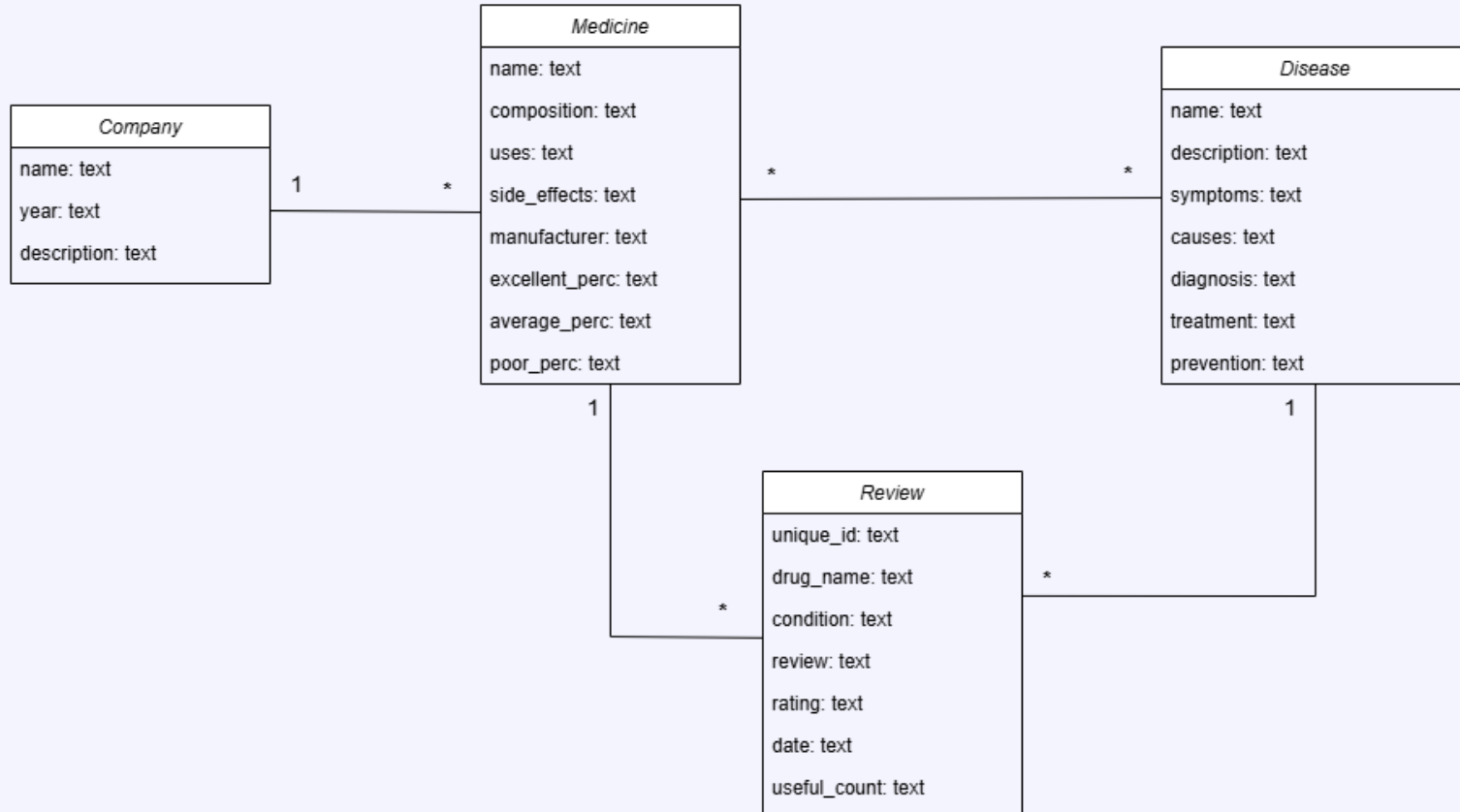
# Milestone #2 Review



## Additional Information Source




# Conceptual Data Model

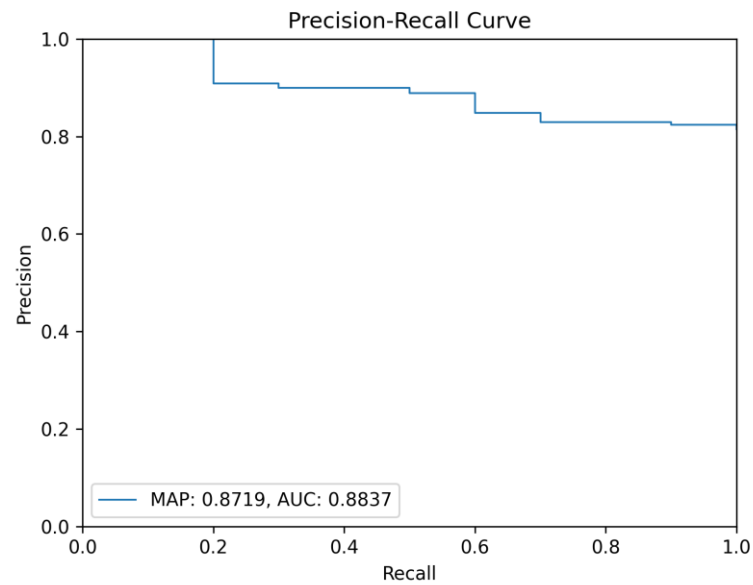
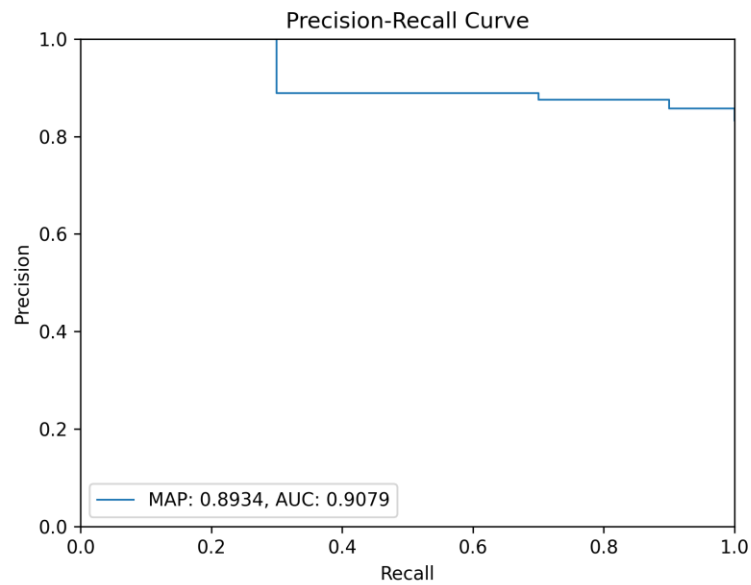




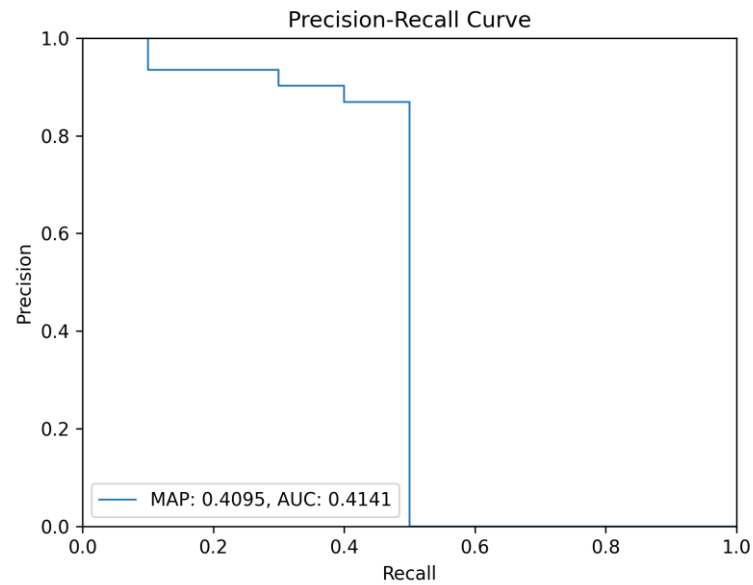
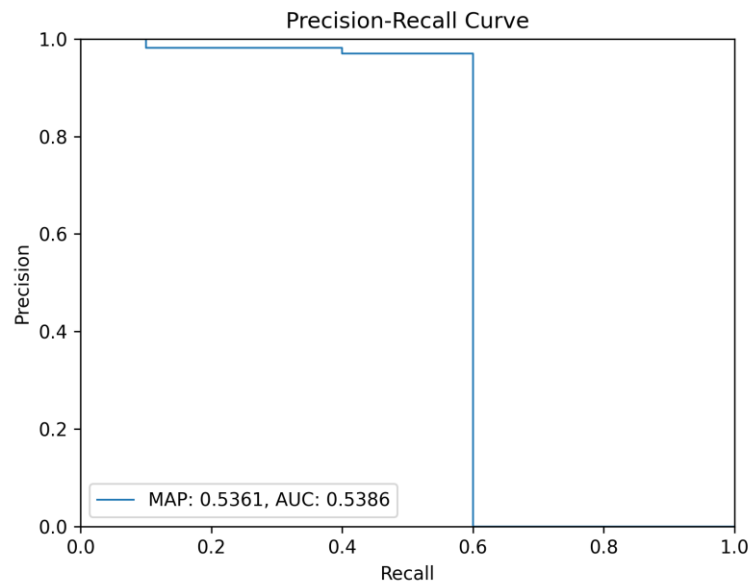
# Information Needs

- Identify pharmaceutical companies or organizations that develop, produce, or provide treatments for malaria.
  - Find customer reviews, opinions, or feedback regarding Eli Lilly's products, services, or the company's reputation.
  - Obtain information about the weight gain side effect caused by the use of antidepressant medications, including data on specific antidepressants.
  - Search for medication or treatments that are effective for managing or curing migraines.
- 

# Evaluation



# Evaluation

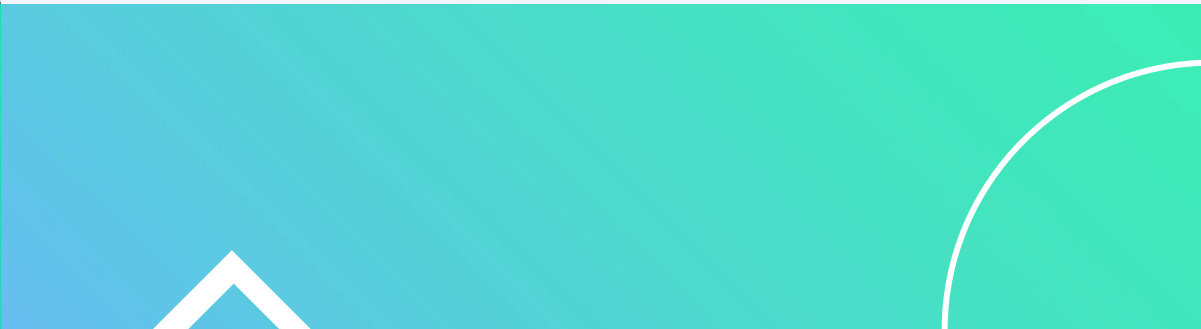


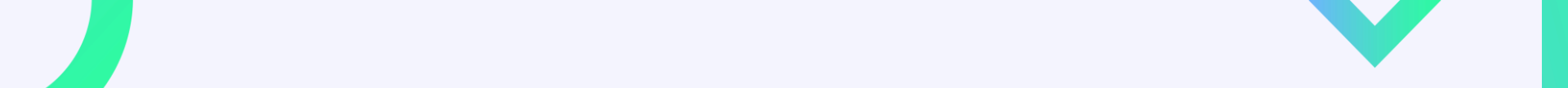




# 02

## Re-ranking + Polarity





To make use of a polarity analysis, we utilized the *NLTK* package's VADER sentiment analysis tool.

- **Average Review Rating Normalization:** Normalization of the *review\_average\_rating* field's range for every document in the collection (the range became [-1, 1]).
- **Review Polarity Score:** For each review in a specific document, get a polarity score.
- **Average Polarity:** Calculate the mean polarity score for the reviews in a specific document.
- **New Field:** For each document, a new *polarity\_rating* field is created, which represents the average between the normalized average review rating and the mean polarity score, for each document.
- **Re-ranking:** The *polarity\_rating* field is then used to re-rank the top 30 documents returned from a user query to the system.

# ReRank Query Parser

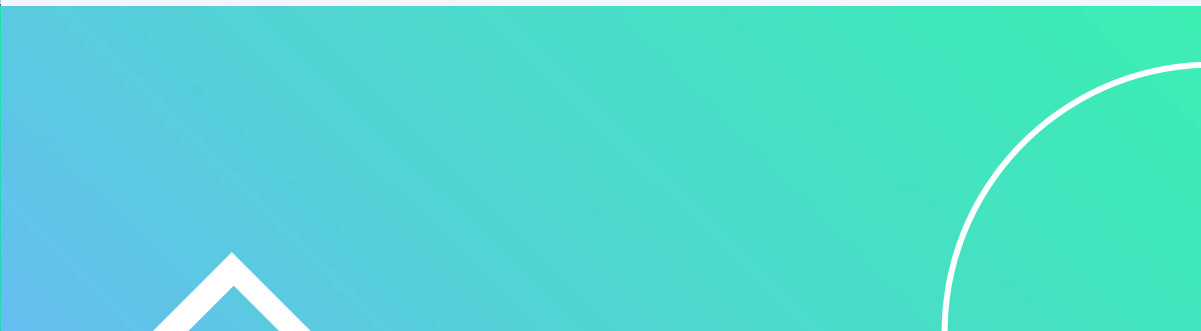
Parameter	Value
reRankQuery	\$rqq
reRankDocs	30
reRankWeight	2.0

Variable	Value
\$rqq	{!func}sum(product(reviews_average_rating, 4), product(polarity_rating, 2))



03

# Semantic Search



# Semantic Search

To make use of the semantic search, the following steps were taken:

- **Conversion of the valued data to vectors:** Conversion of diseases info, side effects and reviews into vectors.
- **Conversion of the user's query into a vector:** For the query to match the vectors it was converted into one vector itself.
- **Advanced ranking:** Ranking of the results, basically searching, not for matching words, but for matching semantics and intents.
- **Possible integration with the query itself:** It can be used both the actual query and the semantic obtained from it.

# Semantic System

Parameter	Value
q	{{!knn f=vector topK=100}}{\$embedding_str}
q.op	AND
start	0
rows	100

# Mixed Semantic System

Parameter	Value
q	{!bool should=\$lexicalQuery should=\$vectorQuery}
q.op	AND
start	0
rows	100

Variable	Value
\$lexicalQuery	{!type=edismax qf=\$fields}{\$query}
\$vectorQuery	{!knn f=vector topK=100}{\$embedding_str}



# 04

## Search System UI



## PRIMED

Enter your query:

cancer treatment

Select search system:

- ☐ Baseline Search
- ☒ Advanced Search
- ☐ Semantic Search

Search

## PRIMED Search Results

Query: **cancer treatment**

**Drug:** Eleftha 440mg Injection

**Composition:** Trastuzumab (440mg)

**Applicable Diseases:** Breast cancer, Stomach cancer

**Average Rating:** 9.75

► [Reviews](#)

**Drug:** Eleftha 150mg Injection

**Composition:** Trastuzumab (150mg)

**Applicable Diseases:** Breast cancer, Stomach cancer

**Average Rating:** 9.75

► [Reviews](#)

**Drug:** Gefitinat 250mg Tablet

**Composition:** Gefitinib (250mg)

**Applicable Diseases:** Non-small cell lung cancer

**Average Rating:** 9.6

► [Reviews](#)

## PRIMED Search Results

Query: **cancer treatment**

**Drug:** Eleftha 440mg Injection

**Composition:** Trastuzumab (440mg)

**Applicable Diseases:** Breast cancer, Stomach cancer

**Average Rating:** 9.75

### ▼ Reviews

My husband has been on Herceptin for his salivary gland cancer for 2 years. The fight to keep him on it has been harrowing as it is only formally approved for breast cancer. I truly believe it saved his life. He tested HER-2/neu positive (3) which is normally located in the breast. Dr. Posner (Dana Farber in Boston) recommended a course of treatment that included Herceptin. His Petscans have now been clear and treatment can stop at the end of August. More needs to be done to allow life saving treatment.

I did 7 months of treatment then my Oncologist had to take me off due to Congestive Heart Failure. I personally don't think Herceptin contributed to it at all but was told it was the combo of it and Adryamycin. It was pretty bad, my ejection fraction was down to 28. I was so short of breath that just walking to the bathroom was a real effort. The fatigue was the mind numbing kind that just turning over in bed felt like I had ran a race. As for side effects for the Herceptin, I had frequent nose bleeds, mild nausea and occasional headaches. I would have gladly endured any side effects for the benefits of this life-saving drug. I am so thankful to Dr Slamon for all the work he did and to Genetech. I am a 4 year survivor so I think that says it all.

Had Herceptin along with Perjeta, Carboplatin & Taxotere for 6 months. HER+ hormone negative stage 3a cancer was believed to have completely resolved from biopsy

# References

1. **Sematext**. Getting started with apache solr. [https://sematext.com/guides/solr/.solr introduction and explanation](https://sematext.com/guides/solr/.solr%20introduction%20and%20explanation)
2. **Apache Software Foundation**. Solr's schema file. [https://solr.apache.org/guide/6\\_6/overview-of-documents-fields-and-schema-design.html#solr-s-schema-file,2017](https://solr.apache.org/guide/6_6/overview-of-documents-fields-and-schema-design.html#solr-s-schema-file,2017).
3. **MDN**. Http request methods. <https://developer.mozilla.org/en-US/docs/Web/HTTP/Methods>, 2024.
4. **Apache Software Foundation**. The extended dismax query parser. [https://solr.apache.org/guide/6\\_6/the-extended-dismax-query-parser.html](https://solr.apache.org/guide/6_6/the-extended-dismax-query-parser.html), 2017.
5. **Apache Software Foundation**. Query re-ranking. [https://solr.apache.org/guide/6\\_6/query-re-ranking.html](https://solr.apache.org/guide/6_6/query-re-ranking.html).
6. **Apache Software Foundation**. Function queries. <https://solr.apache.org/guide/solr/latest/query-guide/function-queries.html>.
7. **Sérgio Nunes**. Solr semantic search. [https://gitlab.up.pt/pri/tutorials/-/blob/main/07-semantic-search/README.md?ref\\_type=heads](https://gitlab.up.pt/pri/tutorials/-/blob/main/07-semantic-search/README.md?ref_type=heads), 2024.
8. **CUF**. Saúde de a-z. <https://www.cuf.pt/saude-a-z>, 2024.
9. **E. Fersini**. Sentiment analysis in social networks. <https://www.sciencedirect.com/science/article/abs/pii/B9780128044124000061>, 2017.
10. **NLTK Project**. Natural language toolkit. <https://www.nltk.org/index.html>
11. **Alessandro Benedetti**. Hybrid search with apache solr. <https://sease.io/2023/12/hybrid-search-with-apache-solr.html>, 2023.
12. **Doug Steen**. Precision-recall curves. <https://medium.com/@douglasssteen/precision-recall-curves-d32e5b290248>, 2020.