BOOKS AND REVIEWS

SEARCH ENGINE

Ana Sofia Teixeira (up201806629)
Ricardo Nunes Ribeiro (up202310095)
Sérgio Manuel Carvalhais (up202007544)
Tomás Manuel Martins (up201704976)

Project Overview

o   The objective of our project is to provide the user with an advanced book search so that they can find the reading they are looking for based on several factors.

o   Thus, the reader can base their search not only on the book category and author, but also on the topics it covers, reviews, ratings and many other topics.

Main improvements

New Schema

Improved search performed by adding 2 filters, the EdgeNGramFilterFactory and the PorterStemFilterFactory.

Synonyms System

Improved search system that can identify and react to different words that mean the same things or refer to the same groups.

Main improvements

Embeddings System

Semantic Search that enables systems to understand query context using compact dimensions to represent document meaning, allowing for efficient nearest neighbor searches.

Keywords System

Searches the words of the query introduced in the top 10 words (new field keywords) of all book reviews combined. The sample of this system only contains 5000 books.

User Interface
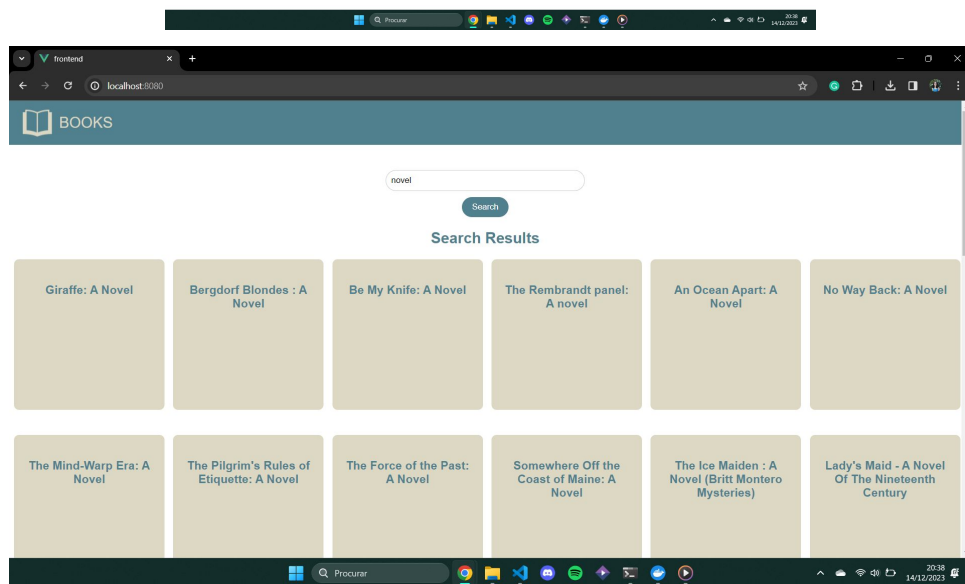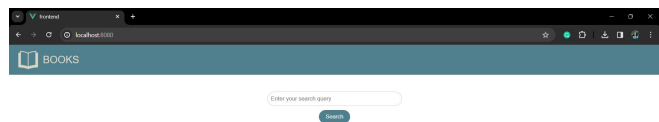
Information Needs

Queries | 🔍

- Retrieve books that talk about artificial intelligence.

- Retrieve books that support fights against racism.

- Retrieve novels set in post-apocalyptic worlds.

- Retrieve books related to American Civil War.

Search Systems

Queries	Systems	Results

"query": "artificial intelligence"

"query": "fight fights against racism discrimination"

"query": "novel AND post-apocalyptic"

"query": "american civil war"

"name": "Enhanced System",
"query_url": "http://localhost:8983/solr/books_m2/select?defType=edismax&indent=true&q
"query_params":
  "qf": "book_title^2 reviews.text^2 reviews.summary description categories",
  "rows": 40

"name": "New Schema System",
"query_url": "http://localhost:8983/solr/books_m3_new_schema/select?defType=edismax&in
"query_params":
  "qf": "book_title^2 reviews.text^2 reviews.summary description categories",
  "rows": 40

"name": "Synonyms System",
"query_url": "http://localhost:8983/solr/books_m3_synonyms/select?defType=edismax&inde
"query_params":
  "qf": "book_title^2 reviews.text^2 reviews.summary description categories",
  "rows": 40

"name": "Keywords System",
"query_url": "http://localhost:8983/solr/books_m3_keywords/select?defType=edismax&inde
"query_params":
  "qf": "book_title^2 reviews.text^2 reviews.summary description categories keywords^5
  "rows": 40

"name": "Embeddings System",
"query_url": "",
"query_params":
  "q": "{!knn f=vector topK=40}[-0.024350997060537338,0.016706684604287148,0.037720501
  "fl": "book_id,book_title,description,authors,categories,publishedYear,reviews.summa
  "rows": "40",
  "wt": "json"

"name": "Combinated System",
"query_url": "http://localhost:8983/solr/books_m3_combined/select?defType=edismax&in
"query_params":
  "qf": "book_title^2 reviews.text^2 reviews.summary description categories",
  "rows": 40

"book_title":"Spycatcher: The Candid Autobiography of a Senior Intelligence Officer",
"authors":["Peter Wright"],
"publisher":"Musson Book Company",
"categories":["Great Britain"],
"publishedYear":1987,
"reviews.summary":["The ultimate insider's view of Britain's MI5","Addictive!","British counterintelligenc
"book_id":170537,
"id":"3282c832-db44-4bf8-a433-0d3ef779ba44",
"reviews.text":["Reading \"Spycatcher\", I was continually struck by this thought: how was it possible for
"_version_":1785119397922209792

"book_title":"Prison Camps of the Civil War (Civil War (Management Concepts))",
"description":"*Includes pictures*Includes accounts of the prison written by surviving prisoners*Includes footnotes and a biblio
"authors":["Charles River Editors"],
"publisher":"Createspace Independent Pub",
"categories":["History"],
"publishedYear":2015,
"reviews.summary":["This is a 32 page book intended for children."],
"reviews.text":["My comments should not be construed by anyone to be an assessment of this book. It is merely an advisory to tho
"book_id":201366,
"keywords":["camps","prison","book","children","civil"],
"id":"f32da4c2-1239-4b22-8b25-240a73152c17",
"_version_":1785110438500892672

"book_title":"Devil's Island: A Novel (Apocalypse Diaries)",
"description":"The apostle John pushed aside the incense. \"I will not make your sacrifice,\" he announced
"authors":["John Hagee"],
"publisher":"Thomas Nelson",
"categories":["Fiction"],
"publishedYear":2002,
"reviews.summary":["A Brilliant Work of Fiction","Best Seller List","Awesome read!","A Good book to read.
"reviews.text":["Devil's Island is the first installment of the Apocalypse Diaries, and I have to admit I
"book_id":77870,
"id":"b621d69f-a180-418f-a2c6-ce1c0a057c55",
"_version_":1785111235105128450

Books About AI

Evaluation

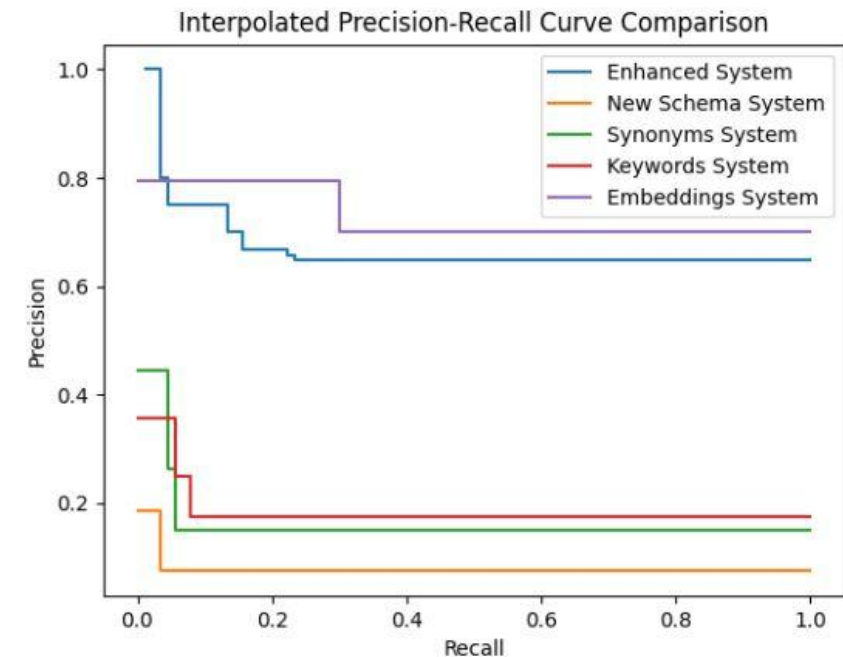| System	Average Precision	Precision at 10 (P@10)
Enhanced System	0.858932	0.900000
New Schema System	0.777723	0.800000
Synonyms System	0.000000	0.000000
Keywords System	0.565476	0.400000
Embeddings System	0.774552	0.700000

Interpolated Precision-Recall Curve Comparison
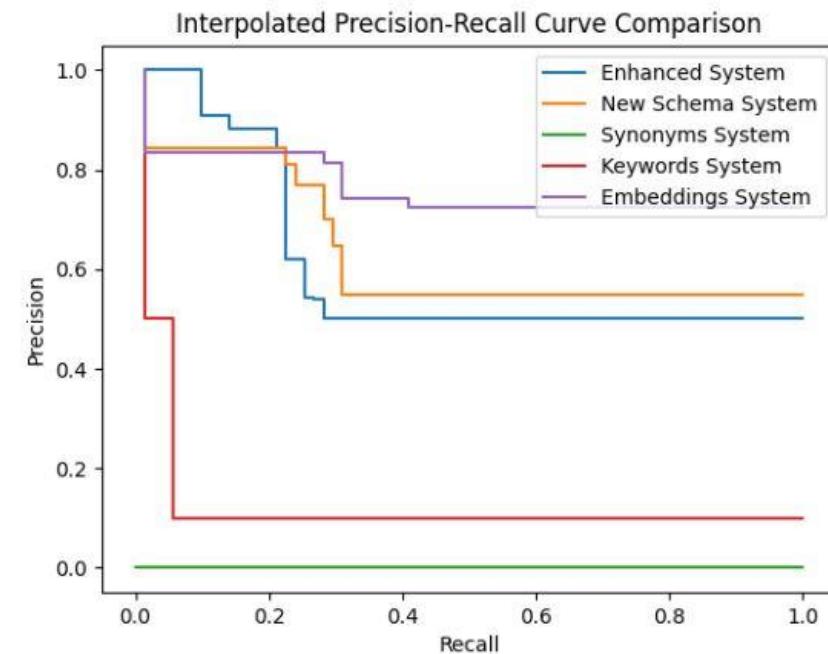
Books That Fight Against Racism

Evaluation

System	Average Precision	Precision at 10 (P@10)
Enhanced System	0.707906	0.700000
New Schema System	0.140422	0.000000
Synonyms System	0.303251	0.400000
Keywords System	0.254040	0.200000
Embeddings System	0.700967	0.600000

Interpolated Precision-Recall Curve Comparison

Post-Apocalyptic World Novels

Evaluation

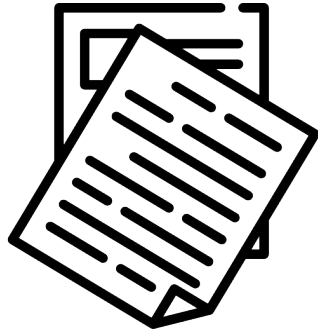| System	Average Precision	Precision at 10 (P@10)
Enhanced System	0.845534	0.800000
New Schema System	0.716314	0.700000
Synonyms System	0.616292	0.600000
Keywords System	0.805556	0.300000
Embeddings System	0.529027	0.600000
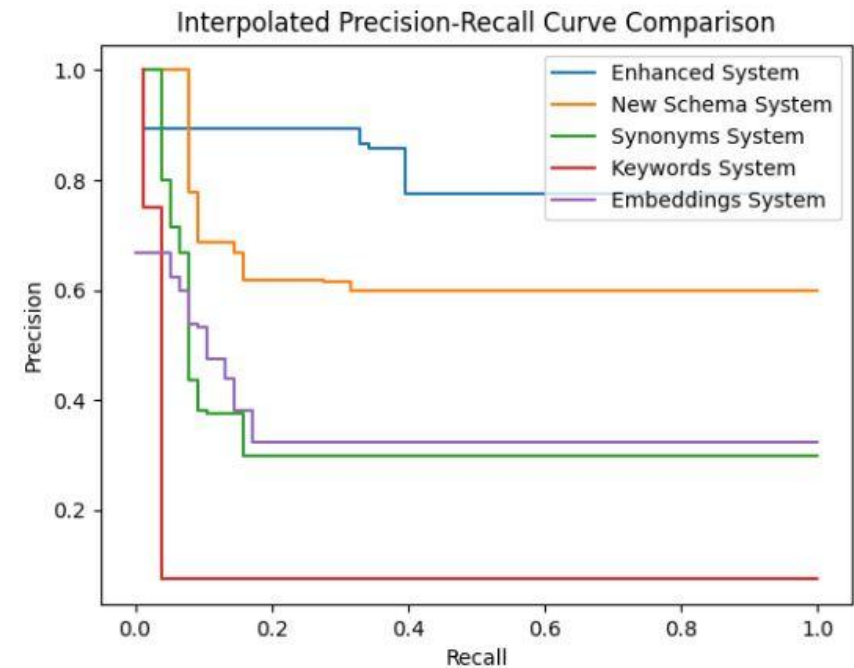
Interpolated Precision-Recall Curve Comparison

American Civil War Books

Evaluation

System	Average Precision	Precision at 10 (P@10)
Enhanced System	0.968378	1.000000
New Schema System	1.000000	1.000000
Synonyms System	1.000000	1.000000
Keywords System	0.385366	0.400000
Embeddings System	0.931871	1.000000

Interpolated Precision-Recall Curve Comparison
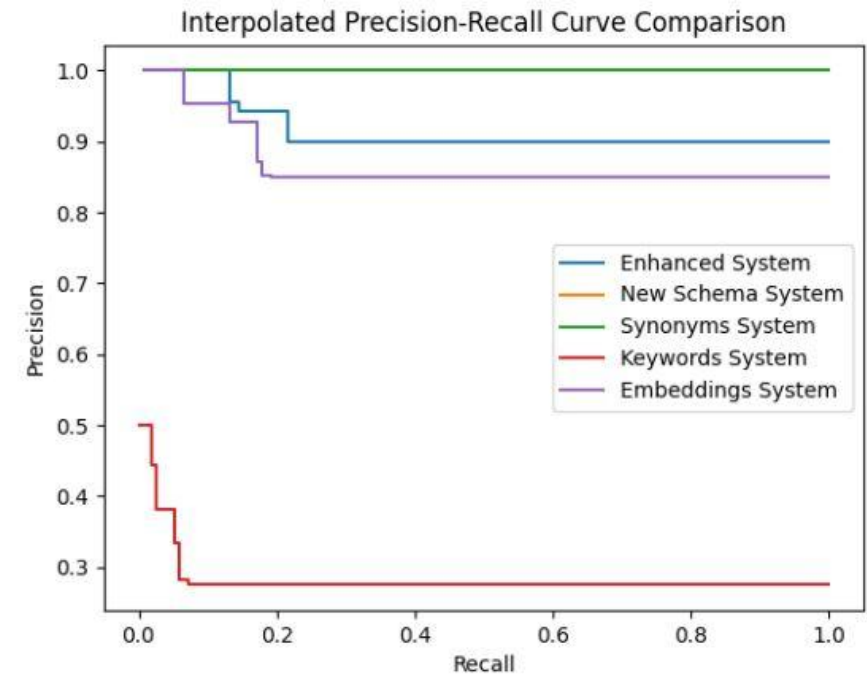
Combined System

New Schema

+

Synonyms System

Example: American Civil War Books

Evaluation

System	Average Precision	Precision at 10 (P@10)
Enhanced System	0.968378	1.000000
New Schema System	1.000000	1.000000
Synonyms System	1.000000	1.000000
Combined System	1.000000	1.000000

Interpolated Precision-Recall Curve Comparison
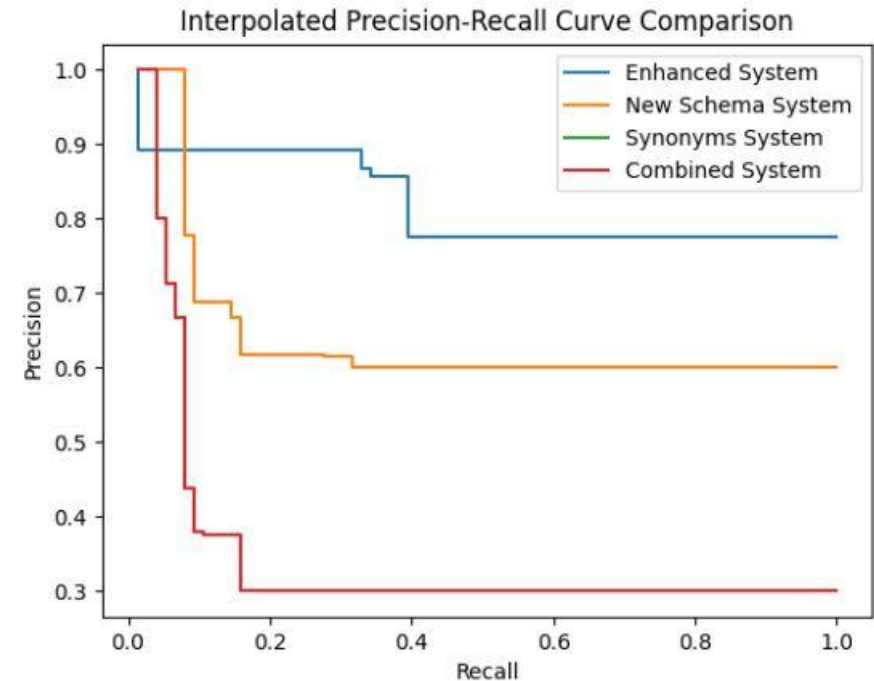
Mean Average Precision

System	Mean Average Precision
Enhanced System	0.845187
Base System	0.811680
Embeddings System	0.734105
New Schema System	0.658615
Keywords System	0.502610
Synonyms System	0.479886
Combined System	0.479886