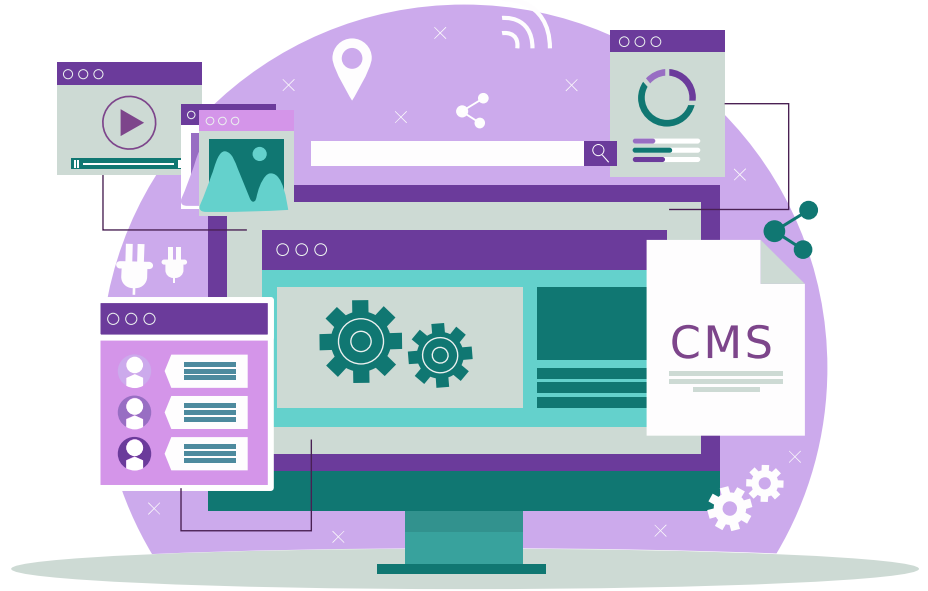


# Goodreads Books and Reviews

January 18th, 2022



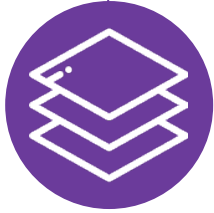
**PRI, Group 2144**

Inês Silva, up201806385

Mariana Truta, up201806543

Rita Peixoto, up 201806257

# Search System Improvements



## Data

Enrichment of information about **authors**.  
Elimination of **inconsistencies** due to use of multiple languages.  
Sentiment analysis improvement.



## Information Retrieval

Search system enhancement using Solr's search components.



## User interface

Improvement of user experience with development of a frontend.

# New Dataset

## Books



csv file



13 columns

**100k**

lines

from Kaggle, originally from  
Goodreads

## Books' Reviews



csv file



3 columns

**510k**

lines

scraped from Goodreads  
Now with translations...

## Authors



csv file



3 columns :  
name, image,  
description

**63k**

lines

scraped from Goodreads

# Translations



## Reviews Dataset

link	review	sentiment
<a href="https://goodreads.com/book/show/1001090.T...">https://goodreads.com/book/show/1001090.T...</a>	<b>Muy buen libro</b> para los profesionales que se d...	0
...	...	...
<div>↓ langdetect + googletrans</div>		
<a href="https://goodreads.com/book/show/1001090.T...">https://goodreads.com/book/show/1001090.T...</a>	<b>Very good book</b> for professionals who are dedic...	1
...	...	...

# Solr's Search Components

## Suggester

Implements auto-suggest feature in the application

## Spell checker

Inline query suggestions based on similar terms

## Learning to rank

Runs trained learning models on query results

## More like this

Query for documents similar to the ones in the result list

## Result grouping

Group documents by common field value

## Faceting

Arranges search results into categories by indexed terms

## Synonyms

Can be used to configure synonyms for use during both indexing and querying of textual data

# Used Components



## Suggester and spell checker

Applied together to assist user search with dataset information

Consider synonyms, stop words, stemming and token filters

Support misspelled terms by the user

Suggestions identified including entire content of the field

## Facets

Retrieve field filters for genre and book format

Aggregates fields in alphabetic order and limits to 1200 different facets

## Synonyms

File configured with a few synonyms for meaningful words

# User Experience Improvement

01

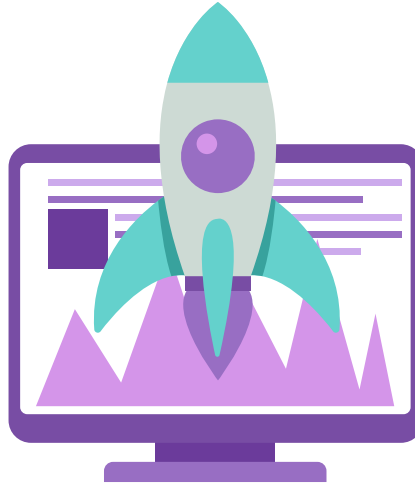
Free text search bar

02

Filter box

03

Facets for genre and  
book format



07

Appealing design

Control over filed  
weights

04

Pagination

05

Info box information  
display

06

<https://localhost:3000>

bookgarden

What are you looking for?

Filter results

genres

bookformats

rating

pages

**Gabriel's Bride**

**Gabriel's Horses**

**Gabriel's Inferno**

**Gabriel's Lady**

**Gabriel's Palace: Jewish Mystical Tales**

**Gabriel's Rapture**

**Gabriel's Redemption**

**Gabriel's Triumph**

**Gabriele D'Annunzio: Poet, Seducer, and Preacher of War**

214 > > >

bookgarden

What are you looking for?

Prioritize: ☒ Title ☒ Description ☐ Author ☐ Positive Reviews ☐ Negative Reviews

Filter results

genres

bookformats

rating

pages

Clear filters

127 results

books

Chelmino and the Holocaust: The History of Hitler's First Death Camp

As the first extermination camp established by the Nazi regime and the prototype of the single-purpose death camps of Treblinka, Sobibor, and Belzec, the Chelmino death camp stands as a crucial but largely unexplored element of the Holocaust. This book is the first comprehensive work in any language to detail all aspects of the camp's history, organization, and operations and to remedy the dearth of information in Holocaust...

Author: Patrick Montague  
Genre: World War II, Holocaust, History, Nonfiction, War

★★★★☆ 21 total ratings 291 pages Hardcover

Holocaust

In "Holocaust", poet Charles Reznikoff's subject is people's suffering at the hand of another. His source materials are the U.S. government's record of the trials of the Nazi criminals before the Nuremberg Military Tribunal and the transcripts of the Eichmann trial in Jerusalem. Except for the twelve part titles, none of the words here are Reznikoff's own: instead he has created, through selection, arrangement, and the rhythms of t...

Author: Charles Reznikoff  
Genre: Poetry, World War II, Holocaust, History, Nonfiction, Academic, Read For School

★★★★☆ 178 total ratings 94 pages Paperback

Filter results

genres

Choose genres

1

19th century

18th century

1st grade

17th century

0 5

pages

Filter results

genres

contemporary romance Choose ge...

bookformats

Paperback Hardcover Choose ...

rating

0 5

pages

230

0 700+

Clear filters



# Results

## System 1

Previous system w/  
custom weights for  
each information  
need

## System 2

Improved search  
system using Solr  
interface w/ global  
weights

## System 3

Improved search  
system using the  
developed user  
interface

### Defined weights

title, desc, author,  
author\_name

1.7

author\_description

1.3

positive\_reviews,  
negative\_reviews

0.8

Remaining fields

default

# Information retrieval 1

## Information Retrieval

One wishes to find New York Times best sellers, since it is a well-known newspaper for its critiques.

### Query

New York Times bestsellers

### Relevance judgement

It is considered relevant if it is a best-seller book or bestseller author and it is possible to verify it according to the retrieved document.

### S1 weights

description

1.2

### Improved System

It was chosen the **authors** option on the search.

Rank	S1	S2	S3
1	2	0	3
2	0	3	2
3	1	3	3
4	3	3	3
5	2	2	2
6	0	0	2
7	0	1	2
8	1	0	3
9	0	2	2
10	1	2	3

# Information retrieval 2

## Information Retrieval

One recently got a sudden interest in wines, therefore it looking for books that will teach him more about wines and wineries.

## Query

winery

## Relevance judgement

It is considered relevant if it is a technical book about wineries or vineyards.

## S1 weights

description

1.2

title

default

## Improved System

It was selected the title and description for being prioritized.

Rank	S1	S2	S3
1	3	3	3
2	3	2	2
3	0	3	3
4	0	0	0
5	0	1	1
6	0	2	2
7	2	0	3
8	3	3	2
9	0	0	0
10	0	0	0

# Information retrieval 3

## Information Retrieval

As a typical history lover dad, my dad is reading a book about the Nazi holocaust. I don't know the book's name, I only remember its cover, and I want to tell my sister about it.

### Query

nazi holocaust history

### Relevance judgement

The intention was to retrieve books whose genre contained history and holocaust and whose description and/or title refer to the word Nazi.

### S1 weights

description	1.2
title	default
genre	1.8

### Improved System

It was selected the title and description for being prioritized and filtered by the genre the "holocaust".

Rank	S1	S2	S3
1	3	3	3
2	3	3	3
3	2	0	1
4	3	1	2
5	0	3	3
6	2	0	2
7	1	2	3
8	0	0	2
9	0	0	1
10	0	0	0

# System Comparison

	Query 1			Query 2			Query 3		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
Avg Precision	0.74	0.69	1.00	0.73	0.90	0.91	0.95	0.85	1.00
P@10	0.60	0.70	1.00	0.40	0.60	0.70	0.60	0.50	0.90
nDCG	0.7667	0.7683	0.9656	0.8968	0.9312	0.9359	0.9816	0.9358	0.9622

## System 1

Mean Average Precision

**0.807**

Mean nDCG

**0.882**

## System 2

Mean Average Precision

**0.813**

Mean nDCG

**0.880**

## System 3

Mean Average Precision

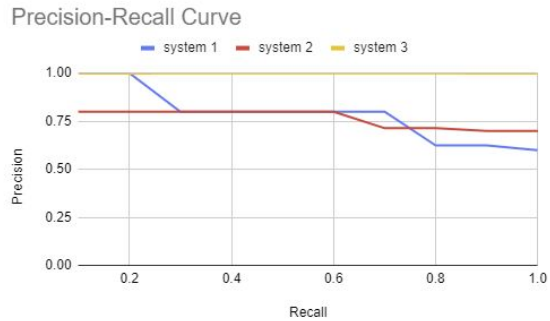
**0.970**

Mean nDCG

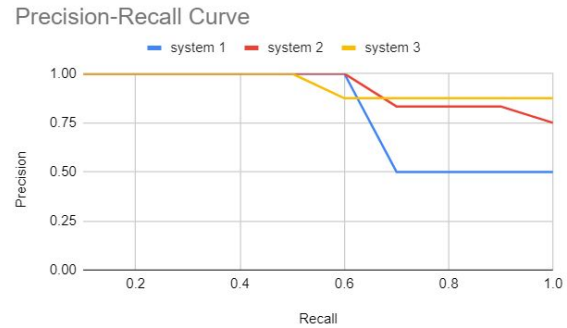
**0.957**

# P-R curves

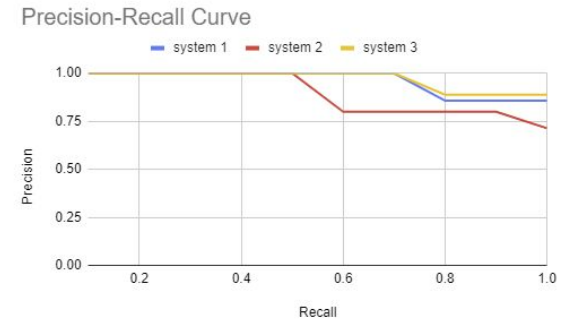
Query 1



Query 2



Query 3



# Conclusion

01

**Aiding the user** to produce better queries was the most meaningful action

02

Measures of **effectiveness** were significantly **improved**



Overall, we are proud of the **work** we developed in this project and the **knowledge** we gained



# Future work

**Spell checker  
improvement**

**AI integration for  
user relevance  
feedback**

**Recommendation  
according to user  
preferences**

