Professional Video Game Reviews

Milestone #2: Information Retrieval (PRI)

Gonçalo de Abreu Matias — up202108703@up.pt
Henrique Filipe Pereira da Silva Caridade — up202108817@up.pt
Pedro Miguel Marques da Silva Pinto — up202108826@up.pt

Project Overview

01

Documents

02

Configuration

03

Indexing

04

Retrieval

05

Evaluation

00 M1 Overview

- IGN website data preparation

- Number of columns: 5 ['Title', 'Subtitle', 'Subheader', 'Content', 'Score']

- Number of rows: 6937

- JSON file with data

```
{
  "Title": "Red Dead Redemption 2 Review",
  "Subtitle": "A meticulously polished open-world ode to the outlaw era.",
  "Subheader": "By Luke Reilly Updated: Jul 12, 2024 6:40 pm",
  "Content": "Red Dead Redemption 2 is a sprawling Western tale of loyalty, conviction,
  "Score": 10,
  "id": "353bb75c952c53bf1869b31adeb3b34b"
},
```

01 Documents

- Game reviews serve as the primary documents, from IGN

- Each document contains:

  - Title: Game name and Review Indicator

  - Subtitle: Brief tagline

  - Subheader: Author and date

  - Content: Full review text

  - Score: Numerical rating (0 - 10)

  - ID: Individual Identification

02 **Configuration**

- Solr configuration
- Using *text_simple* field type for textual content
- Fields types and usage:
  - **text_simple:** Used for textual content (Title, Content, Subtitle, Subheader)
    - Standard Tokenization, Lowercase Filtering, Stop Word removal
  - **pfloat:** Used for Score field
    - Numeric field type, Supports range queries

```
"add-field": [
  {
    "name": "Title",
    "type": "text_simple",
    "indexed": true,
    "stored": true
  },
  {
    "name": "Content",
    "type": "text_simple",
    "indexed": true,
    "stored": true
  },
  {
    "name": "Subtitle",
    "type": "text_simple",
    "indexed": true,
    "stored": true
  },
  {
    "name": "Subheader",
    "type": "text_simple",
    "indexed": true,
    "stored": true
  },
  {
    "name": "Score",
    "type": "pfloat",
    "docValues": true,
    "indexed": true,
    "stored": true
  }
]
}
```

03 Indexing Implementation

- CSV to JSON conversion
- Docker container setup
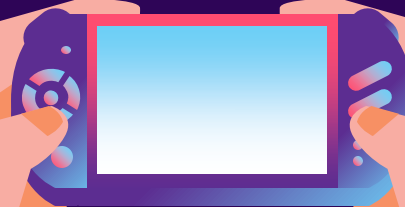- Document indexing

03 Indexing : Simple Schema

Field	Type	Indexed
Title	text_simple	Yes
Content	text_simple	Yes
Subtitle	text_simple	Yes
Subheader	text_simple	Yes
Score	pfloat	Yes

- The *text_simple* type includes:
  - Standard tokenization
  - Lowercase filtering
  - Stop word removal

04 Retrieval

4 Retrieval Process in Solr

Query (q): Focuses on most valuable terms for retrieval.

Query Operator (q.op): AND/OR operations to refine results.

Query Filter (fq): Additional conditions to limit results.

Filter List (fl): Limits fields returned in results.

Sort Field (sort): Specifies sorting order of results.

Start (start): Sets pagination starting point.

Rows (rows): Limits number of documents per query.

Query Parser (defType): Specifies query parser (e.g., edismax).

Boost Query (bq): Increases relevance of specific terms.

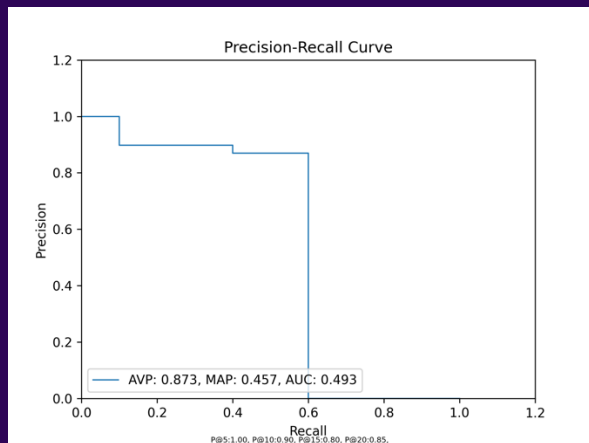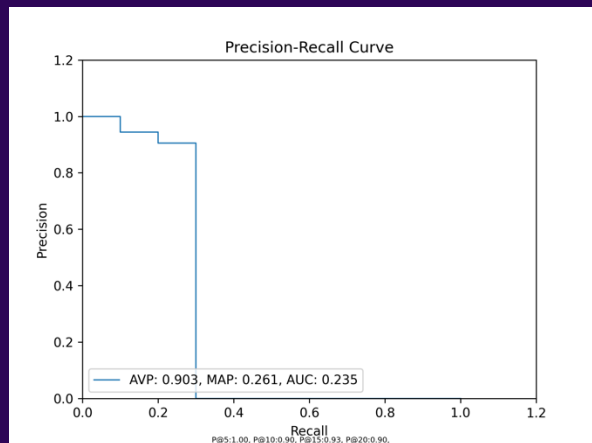**Highlighting (hl)**: Highlights matched terms in results.

5 Evaluation Process

- Games with Tight Controls

- Games good for Relaxing

- Story Narrative games

- Visually impressive Games
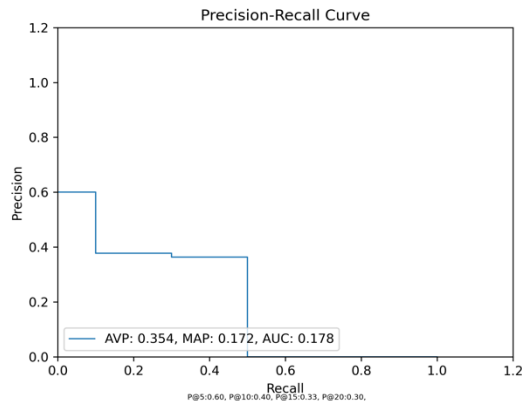
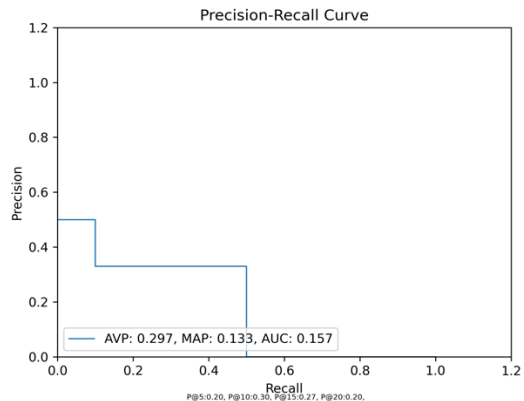- Multiplayer Games

5 Games with Tight Controls

Boosted Query

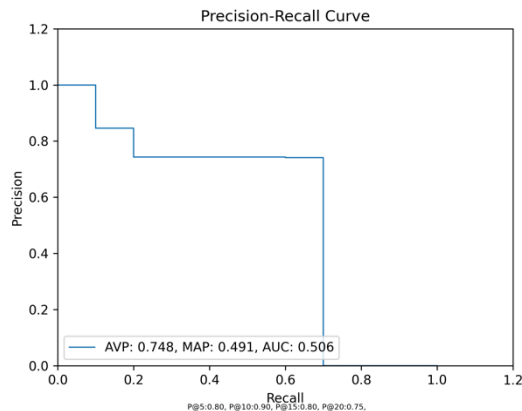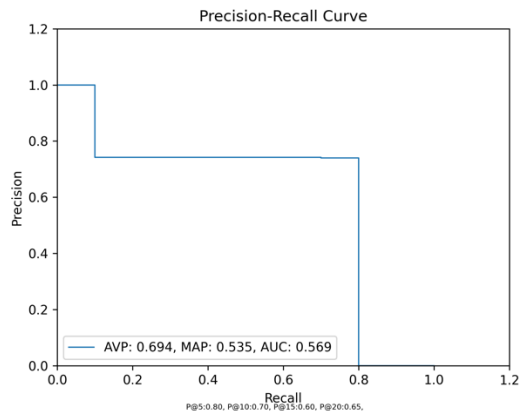|	Query
q	(Title:(*controls* *responsive* *precise* *tight*) OR Content:(*controls* *responsive* *precise* *tight* *accurate* *fluid* *precision*) OR Subtitle:(*controls* *responsive* *precise* *tight*))
qf	Title^2.0 Content^3.0 Subtitle^2.5
pf	Title^6.0 Content^3.0
bq	Subtitle:(*precise*)^5.0 Content:(*fluid* *responsive*)^3.0
fl	id,Title,Content,Score

5 Games good for Relaxing

Precision-Recall Curve

AVP: 0.297, MAP: 0.133, AUC: 0.157

P@5:0.20, P@10:0.30, P@15:0.27, P@20:0.20,

Precision-Recall Curve

AVP: 0.354, MAP: 0.172, AUC: 0.178

P@5:0.60, P@10:0.40, P@15:0.33, P@20:0.30,

Boosted Query

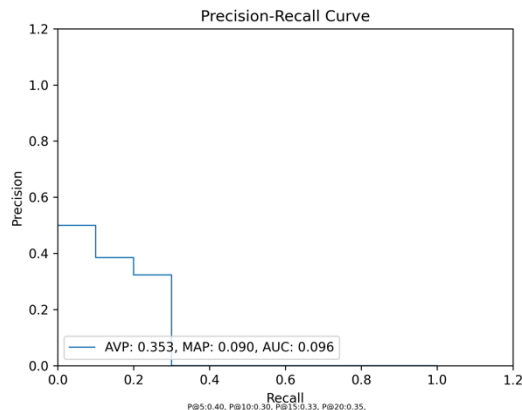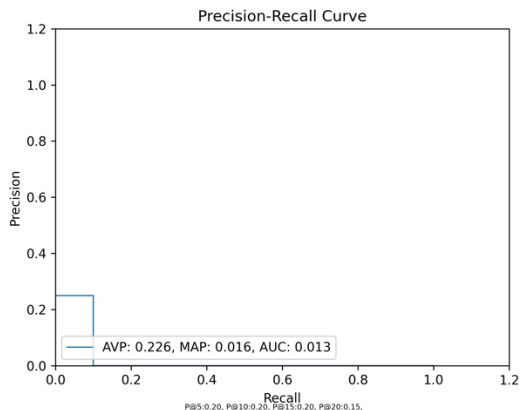| Query	
q	Title:(*relax* *peaceful* *calm* *chill*) OR Content:(*relax* *peaceful* *calm* *meditat* *chill* *mindful* *simple* *serene* *enjoy* *soft*)
qf	Title^4.0 Content^2.0 Subtitle^1.5
pf	Title^8.0 Content^4.0
bq	Subtitle:(*peaceful*)^2.0 Content(*calm* *relax*)^3.0
fl	id,Title,Score

5 Story Narrative Games

AVP: 0.694, MAP: 0.535, AUC: 0.569
P@5:0.80, P@10:0.70, P@15:0.60, P@20:0.65,

AVP: 0.748, MAP: 0.491, AUC: 0.506
P@5:0.80, P@10:0.90, P@15:0.80, P@20:0.75,

Boosted Query

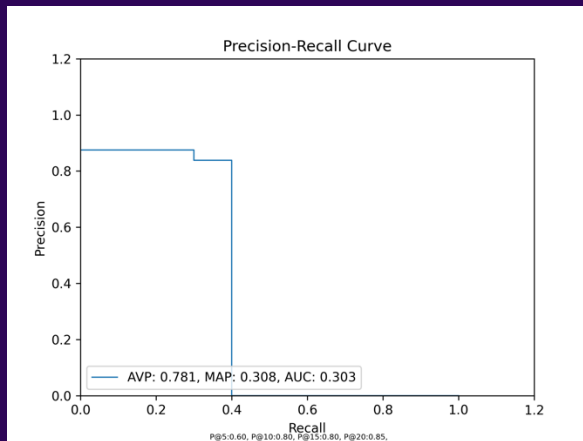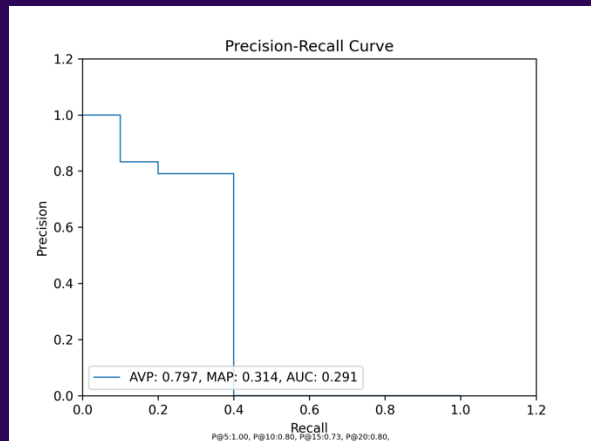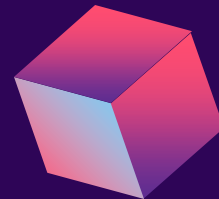| Query	
q	Title:(*story* *narrative* *emotional* *character*) OR Content:(*story* *narrative* *emotional* *character*) OR Subtitle:(*story* *narrative* *emotional* *character*)
qf	Title^2.0 Content^3.0 Subtitle^2.5
pf	Title^6.0 Content^3.0
bq	Subtitle:(*story*)^5.0 Content:(*character* *emotional*)^4.0
fl	id,Title,Content,Score

5 Visually Impressive Games

Boosted Query

Query	
q	Title:(*visually* *graphically* *visuals* *realistic*) OR Content:(*visually* *graphically* *visuals* *realistic*) OR Subtitle:(*visually* *graphically* *visuals* *realistic*)
qf	Title^2.0 Content^3.0 Subtitle^2.5
pf	Title^6.0 Content^3.0
bq	Subtitle:(*visuals*^4.0 Content:(*realistic*)^3.0
fl	id,Title,Content,Score

5 Multiplayer Games

Precision-Recall Curve

AVP: 0.797, MAP: 0.314, AUC: 0.291

P@5:1.00, P@10:0.80, P@15:0.73, P@20:0.80,

Precision-Recall Curve

AVP: 0.781, MAP: 0.308, AUC: 0.303

P@5:0.60, P@10:0.80, P@15:0.80, P@20:0.85,

Boosted Query

Query	
q	Title:(*multiplayer* *online*) OR Content: (*multiplayer* *team*) Subtitle: (*multiplayer* *online* *cooperative* *team*)
qf	Title^4.0 Content^2.0 Subtitle^1.5
fq	Content:multiplayer
bq	Subtitle:(*online*)^3.0 Content:(*multiplayer*)^5.0
fl	id,Title,Content,Score

Conclusion and Future Work

All objectives were accomplished for this milestone:

- CSV file converted into JSON file
- Indexed Documents
- Created Queries for various scenarios

Aspects to improve in the future:

- Improve quality of query results
- Improve overall M2 based on feedback
- Finalize search system

Thanks!

Do you have any questions?