Steam Game Data Search Engine: Search System

Bruno Mendes - up201906166@edu.fe.up.pt
Fernando Rego - up201905951@edu.fe.up.pt
Joel Fernandes - up201904977@edu.fe.up.pt

PRI 2022/2023 T1G8

Where we are

- ❏ *M1*
  - ❏ Dataset with Steam Games' data
  - ❏ Data analysis and cleanup
  - ❏ Enrichment with Wikipedia description
- ❏ M2
  - ❏ Indexing in Apache Solr
    - ❏ Small textual fields with stemming and phoneme injection
    - ❏ Long textual fields with synonyms, stemming, stop word removal
  - ❏ Querying in Apache Solr
    - ❏ eDisMax queries exploring field boosts, fuzziness…
- ❏ **M3**
  - ❏ Enrichment with extracted entities' Wikipedia description
  - ❏ Use of Solr's More Like This feature based on textual fields
  - ❏ Using query style as a signal to modify the boost function
  - ❏ Wrapping it up in a friendly command line interface

+   evaluation

+   evaluation

Entity Exploration

- WikiData
  - People
  - Locations
  - Organizations

Example:

Counter-Strike (2000)
- Counter-Strike: Source
- Jess "Cliffe" Cliffe
- Minh
- Ritual Entertainment
- Counter-Strike Nexon: Studio
- Counter-Strike Neo
- Valve
- Counter-Strike: Condition Zero
- Hidden Path Entertainment
- Xbox 360
- the Counter-Strike Online
- Half-Life
- Turtle Rock Studios

More Like This

- WikiData
- PromotionalDescription
- EntityWikiData

Example:

Counter-Strike (2000)
- Counter-Strike: Source
- Counter-Strike: Condition Zero
- Strike Suit Zero
- Counter-Strike: Global Offensive
- Bang Bang Racing
- Pineapple Smash Crew
- Commandos: Behind Enemy Lines
- Catch a Falling Star
- Sherlock Holmes: The Mystery of the Mummy
- Mimpi

New Signals

- Query length
  - Longer queries likely to mean more experienced player
  - More experienced players likely to want better-reviewed games
- "Gamer terms" usage
  - Using terms or word groups such as "open world" or "pvp" signal experienced player
  - More experienced players likely to care less about game popularity
- "Gamer mode"
  - 2 or more "gamer terms" employed trigger the "gamer mode"
  - "Gamer mode" switches the query operator to AND

Query	Metacritic bf weight	q.op	First result
shots	0.55	OR	Counter-Strike
shooter	0.65	OR	Counter-Strike
open world shooter	0.85	AND	Call of Duty

CLI Live Demo

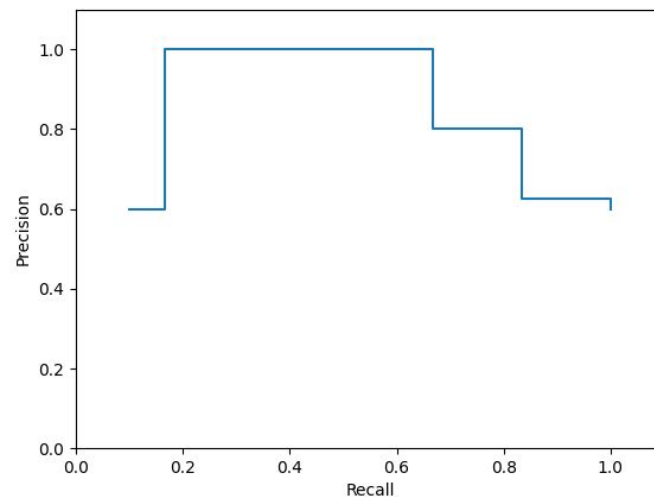Steam Game Data Search Engine: Search System

Results

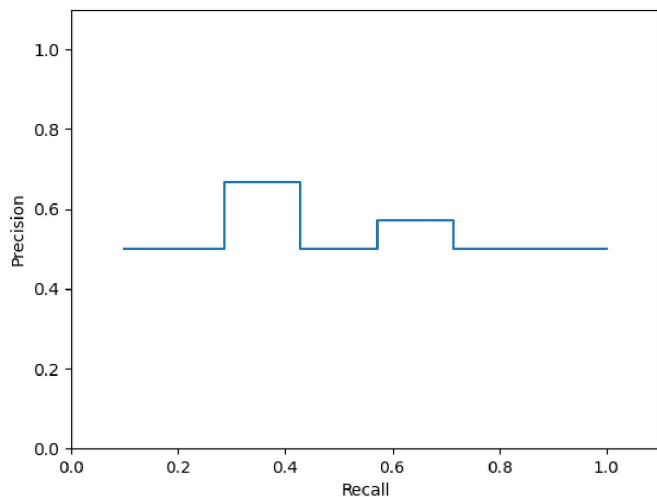- Which are the best games based on Vietnam?

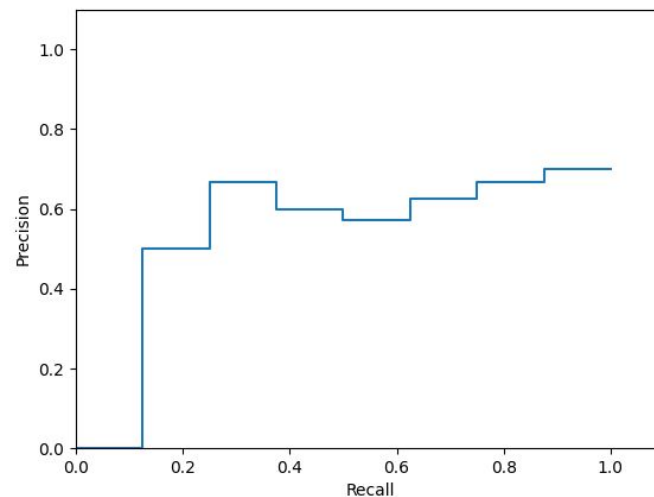M2 System
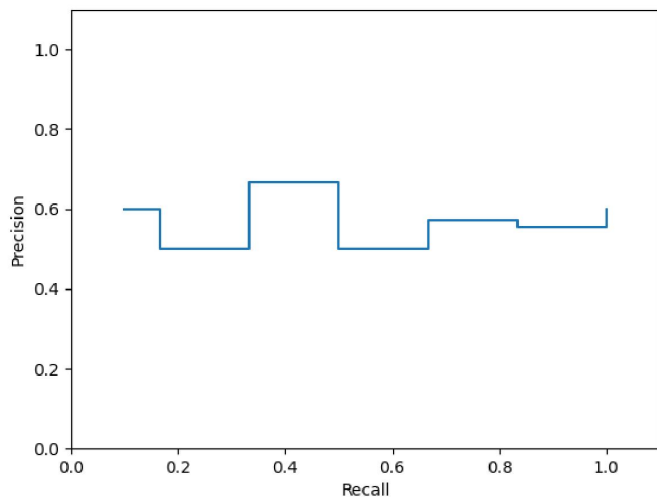
M3 System

Results

- Which are the best games to solve riddles?

M2 System

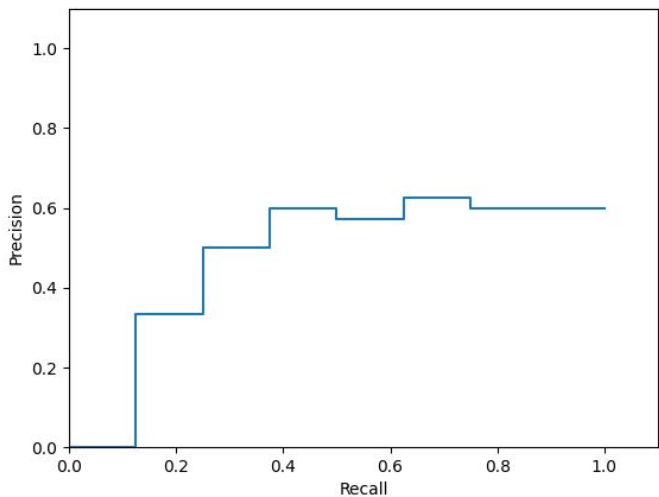M3 System

Results
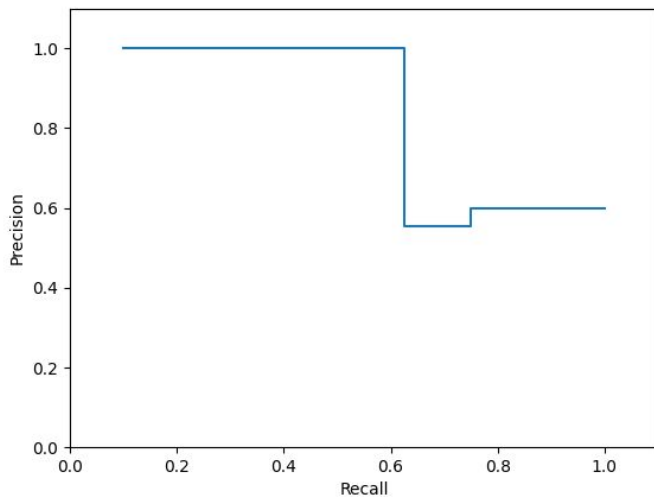
Gamer mode

- Which are the best open-world shooter games?
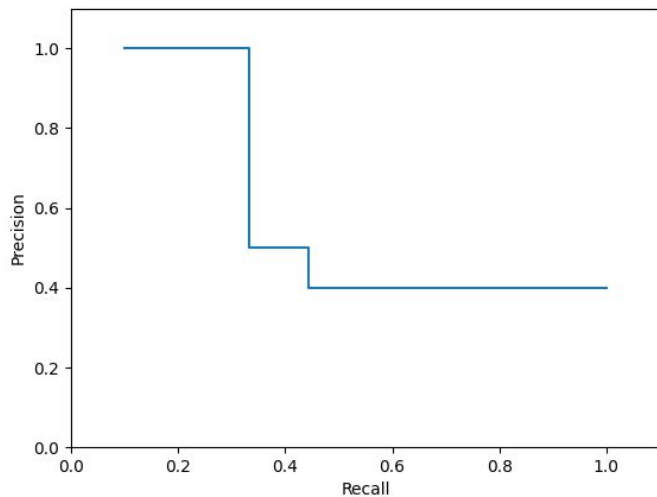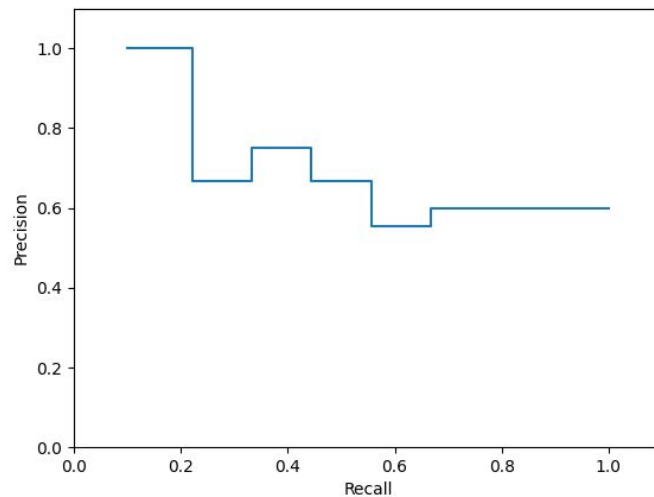

M2 System

M3 System

Results

- Which are the best magic PvP games?

M2 System

M3 System

Conclusions

- More textual data (from entity extraction) enriches the system

- Query signals are not easy

    - Hard to get in the user's head without knowledge of their history

    - Guessing might make the results worse (see M3's opinionated "Gamer mode")

- The search system works, but can be improved

    - Deploying in the real world would give us insights to improve the system

    - Improvements can be manual or self-made (e.g. relevance feedback)