



M.EIC PRI 2022/2023 G55

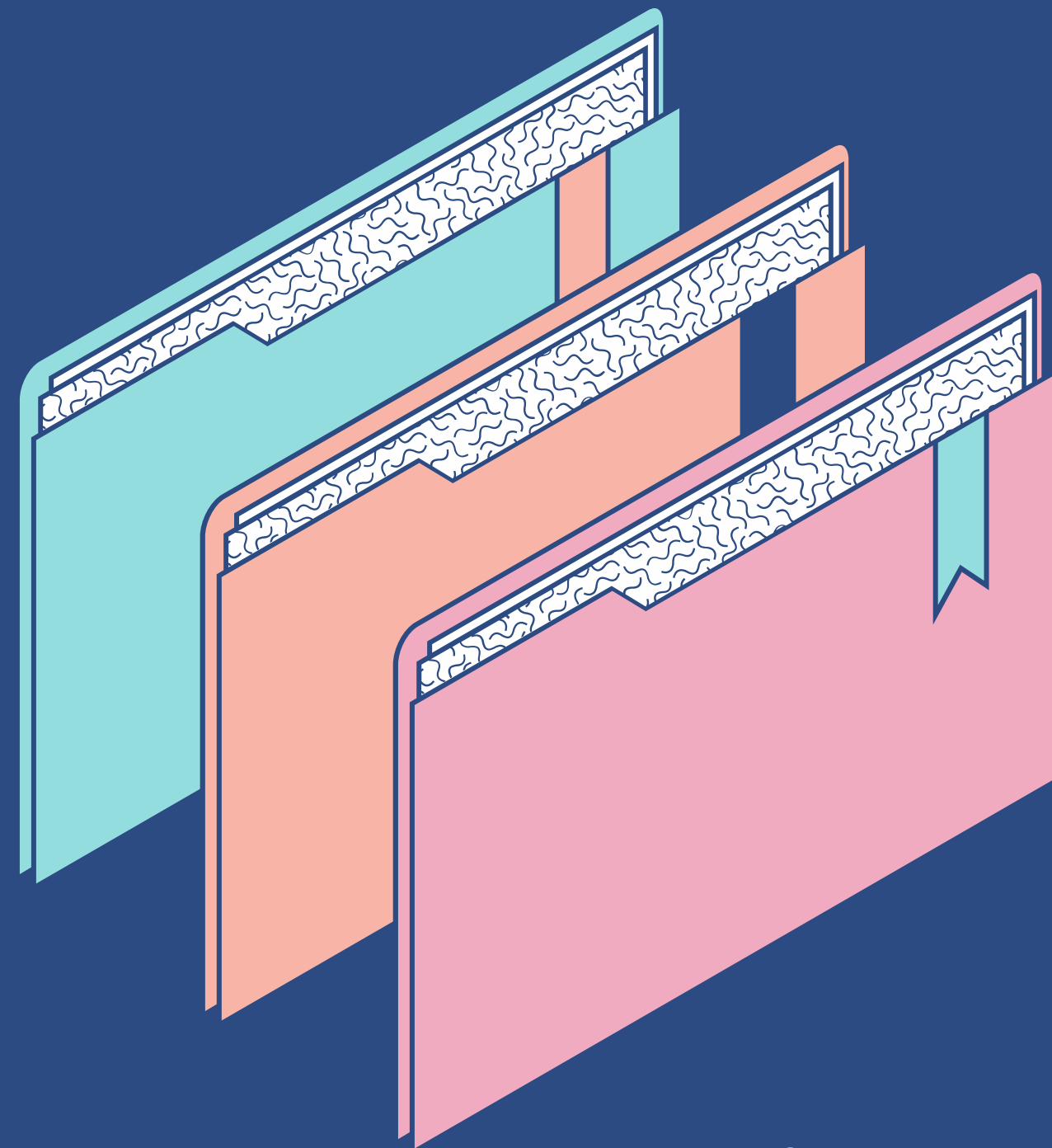
arXiv - Scientific Papers

Scientific Articles Search System

Beatriz Santos, up201906888

Sérgio Estêvão, up201905680

Sérgio da Gama, up201906690

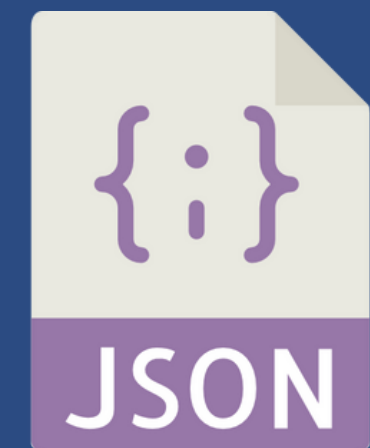


Dataset

Data

kaggle

Data



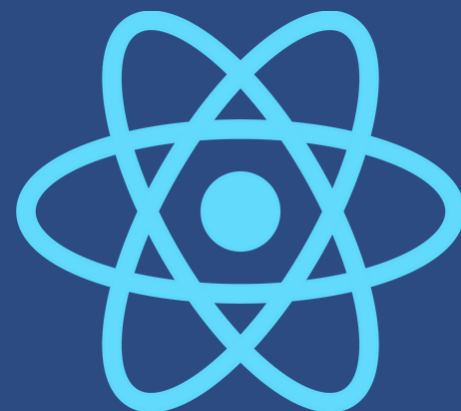
Information Retrieval

Collection of scientific papers and their corresponding information from the website ARXIV

Solr

Information Retrieval

User Interface





Collection

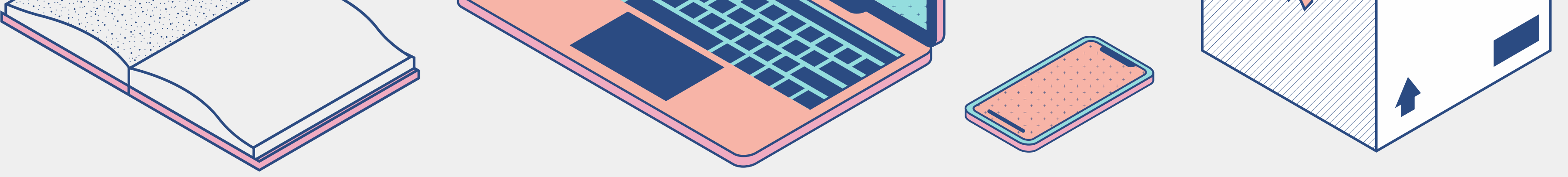
A paper defines a document which has the paper's information.

Most relevant fields:

- Title
- Summary
- Authors
- Areas
- Fields
- Subjects



System Improvements



Highlighting

When a user searches for a specific term, Solr will highlight the term within the search results, making it easy for the user to see where the term appears within the documents.

This feature is useful for quickly locating the relevant information within the search results.

In our case, we use `` and `` to highlight these results



MoreLikeThis

MoreLikeThis feature in Solr works by analyzing the text of the input document and then finding other documents in the index that have a similar text of the fields specified. The documents with the highest similarity scores are then returned as the results of the MoreLikeThis query.



Schema

In order to obtain a better search, we looked for more filters that would be possible to add to the schema and decided to test EdgeNGramFilterFactory.

However, we came across schemas that did not obtain the result that we wanted.



Schema "improved" but not really

Upon searching black-holes...

```
"13649":{  
  "summary":["We <b>have</b> introduced two crossover operators, MMX-<b>BLXexploit</b> and\nMMX-<b>BLXexplore</b>, for"]},  
"38328":{  
  "summary":[" Americans. In response\nto #BlackLivesMatter, other Twitter users <b>have</b> adopted #AllLivesMatter, a\ncounter"]},  
"31800":{  
  "summary":["In this paper, we formulate a novel problem for finding <b>blackhole</b> and volcano\npatterns in <b>a large</b>"]},
```



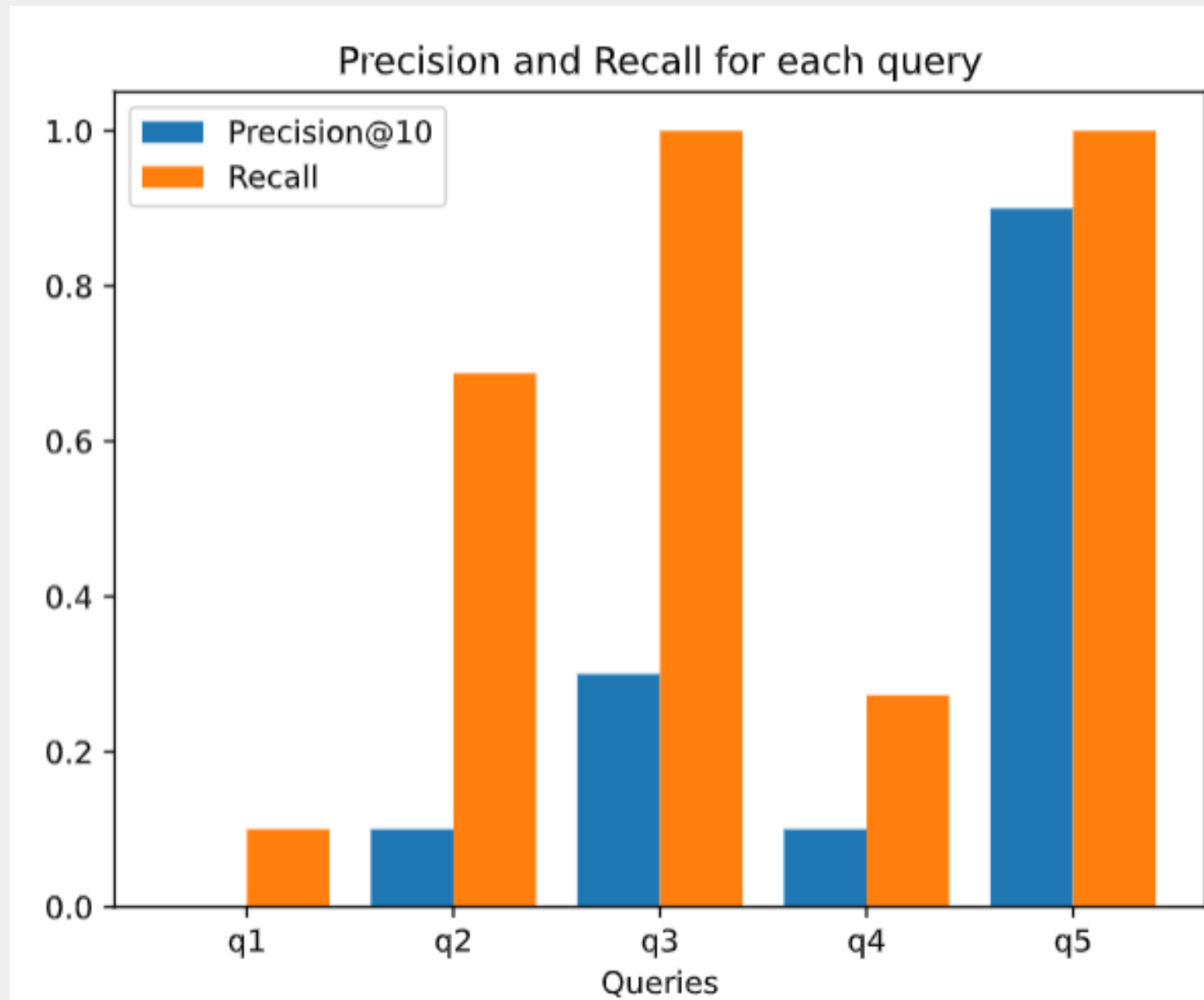

Current Schema

Upon searching black-holes...

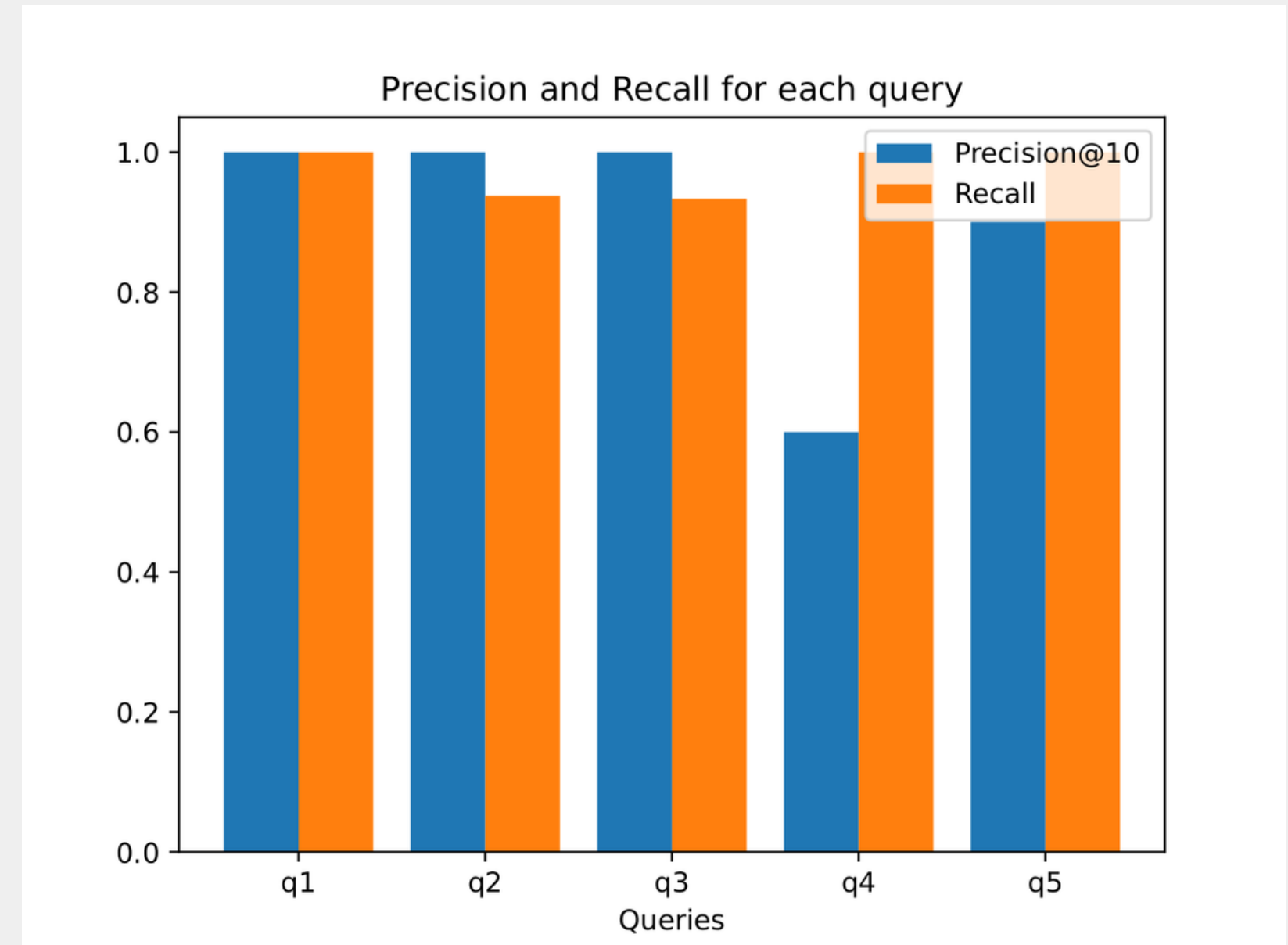
```
"40142":{
  "summary":["Supermassive <b>black</b> <b>holes</b> at centers of clusters of galaxies strongly interact\nwith their host"],
  "title":["Towards understanding feedback from supermassive <b>black</b> <b>holes</b> using\n  convolutional neural networks"]},
"3955":{
  "summary":["The Building <b>Block</b> Hypothesis suggests that Genetic Algorithms (GAs) are\nwell-suited"],
  "title":["Overcoming Hierarchical Difficulty by <b>Hill</b>-Climbing the Building <b>Block</b>\n  Structure"]},
"40112":{
  "summary":["", for quickly evolving sources,\nsuch as the galactic center's supermassive <b>black</b> <b>hole</b> (Sgr A*) targeted by the\nEHT"]},
```



Worst schema results (boosted)



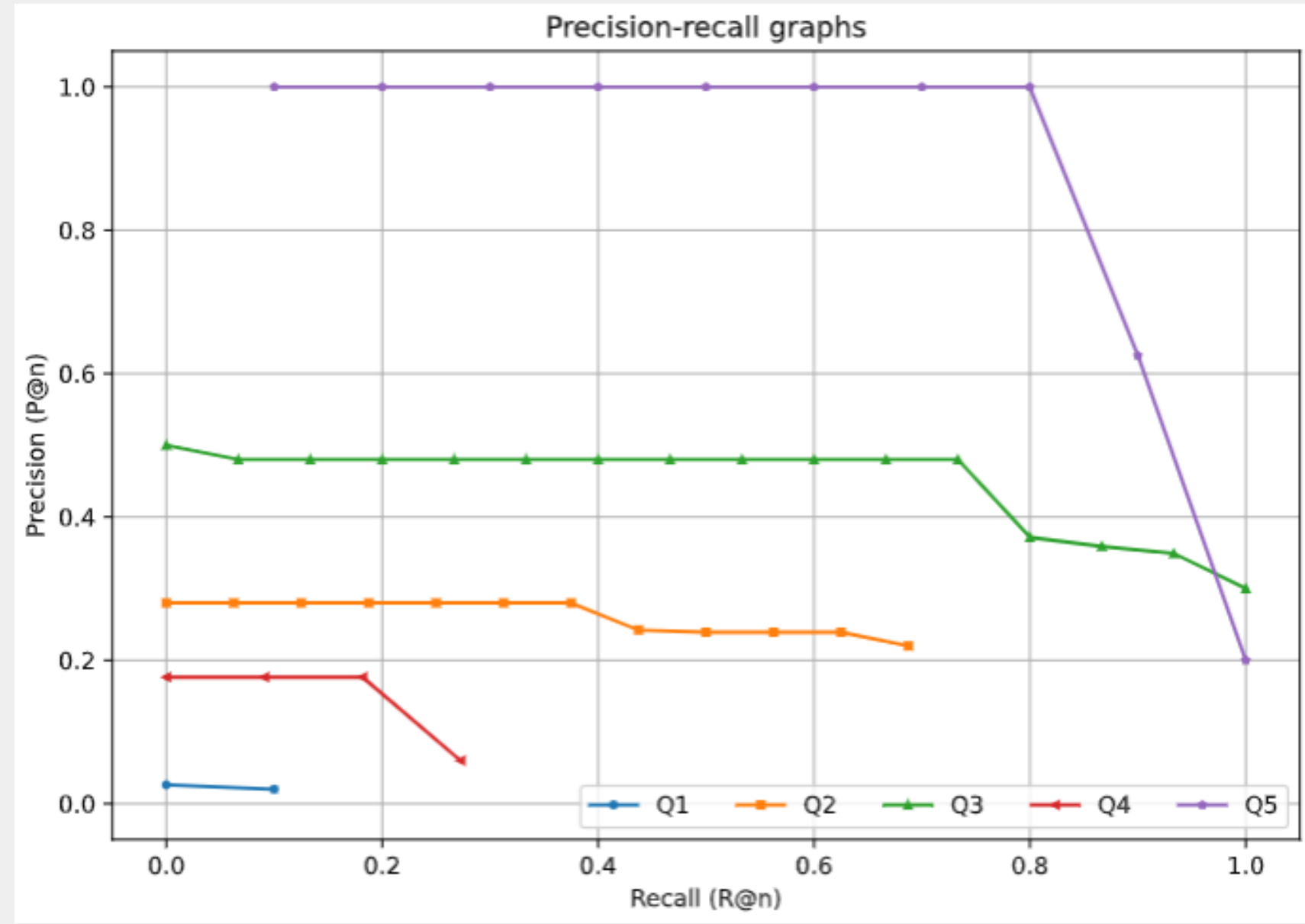
New (worst) schema



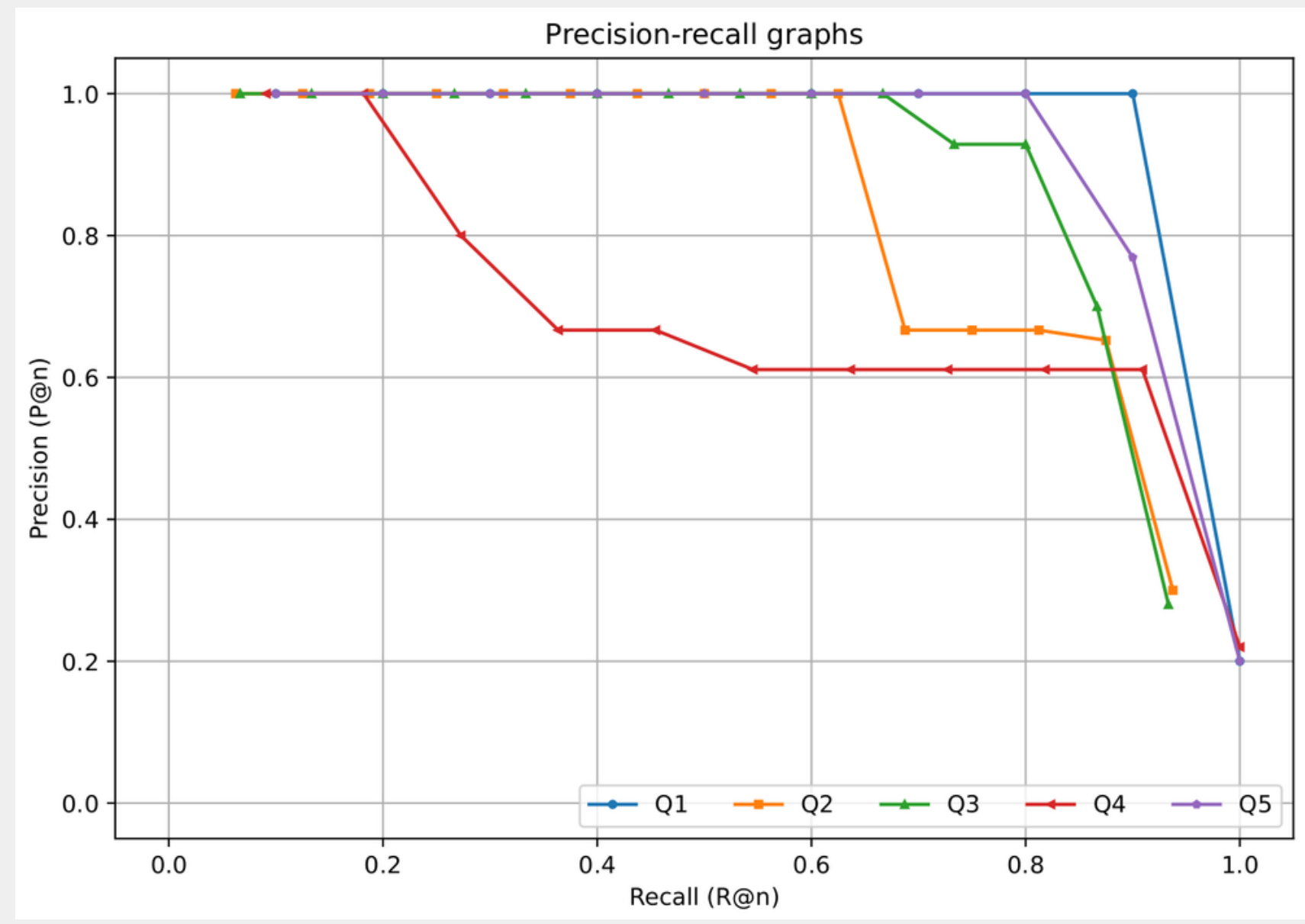
Best schema



Worst schema results (boosted)



New (worst) schema



Best schema


Not always having a schema with many filters lead to a good search



Interface



Search System



black-box

FiltersSearch

Oldest01/01/1993

Earliest12/13/2022

Statistics x x v

Select Fields v

Select Subjects v

Relevance v

MoreLike...

Areas: **Statistics** ; Fields: Statistics ; Subjects: Machine Learning

Black-box Importance Sampling

Qiang Liu,Jason D. Lee

Oct 17 2016

this problem by studying **black-box** importance sampling methods that calculate importance weights for samples generated from anyposals for which the importance weights can be tractably calculated. We address this problem by studying black-box importance sampling methods that calculate importance weights for samples generated from any unknown proposal or black-box mechanism. Our method allows us to use better and richer proposals to solve difficult problems, and (somewhat counter-intuitively) also has the additional benefit of improving the estimation accuracy beyond typical importance sampling. Both theoretical and empirical analyses are provided.

LEARN MORE

Areas: **Statistics** ; Fields: Statistics,Computer Science ; Subjects: Machine Learning,Computation,Methodology,Machine Learning

Black Box Variational Inference

Dejagh Bengioeth Sean Carriah David M. Blai

☐This!

☐This!



Questions?