

Cell Phone Reviews

PRI · Information Processing and Retrieval, 2021/22

M.EIC · Master in Informatics and Computing Engineering

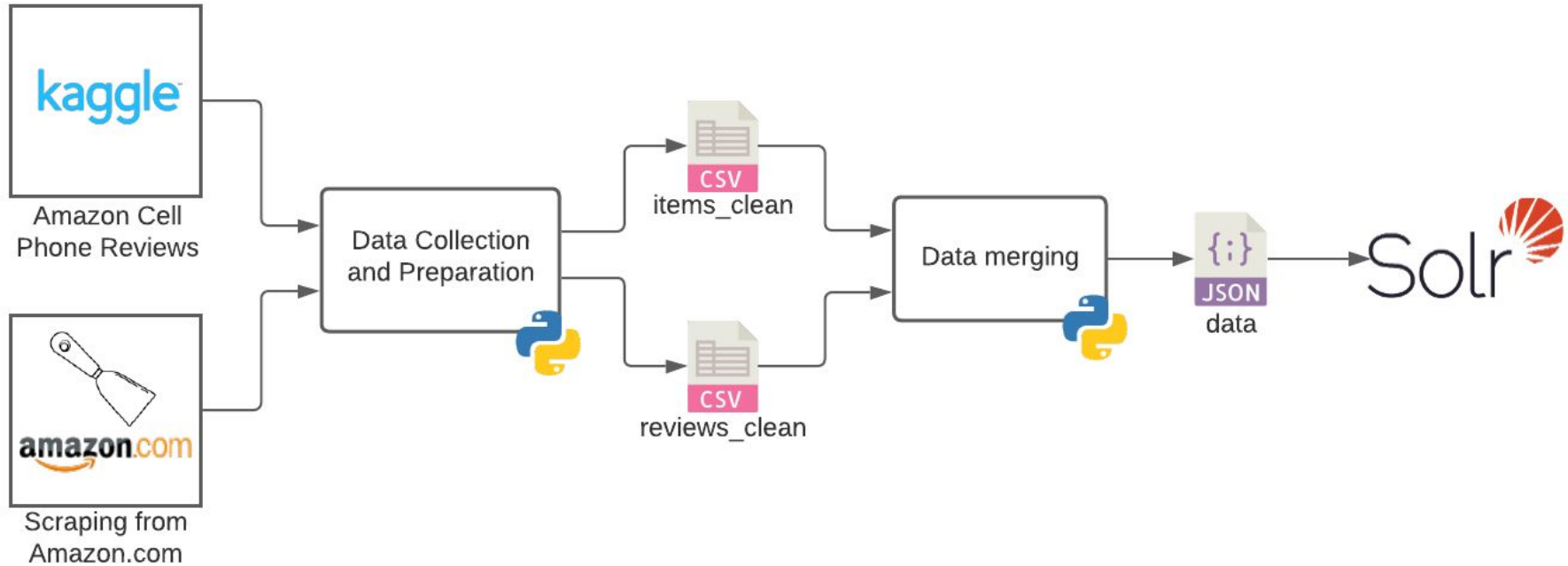
T4G5 · Milestone #2

André Gomes - up201806224

André Nascimento - up201806461

Catarina Fernandes - up201806854

Pipeline



Collection and Documents

- Solr core with a single collection;
- Items and Reviews documents, where the Reviews are child documents of the respective Item;
- Relevant fields:
 - brand
 - ratings
 - titles
 - body
 - date
 - country

```
{
  "brand": "Motorola",
  "title_item": "Motorola Q Black Phone (Verizon Wireless, Phone Only, No Service)",
  "rating_item": 3.0,
  "totalRatings": 14,
  "price": 0.0,
  "originalPrice": 0.0,
  "id": "B000MQFG02",
  "image": "https://m.media-amazon.com/images/I/81dPV5oyW8S._AC_UL320_.jpg",
  "content_type": "item",
  "_version_": "1719034631149322240",
  "reviewUrl": "https://www.amazon.com/product-reviews/B000MQFG02",
  "asin": "B000MQFG02",
  "url": "https://www.amazon.com/Motorola-Black-Verizon-Wireless-Service/dp/B000MQFG02",
  "reviews": [
    {
      "name": "Kim",
      "rating_review": 3,
      "date": "2008-03-14T00:00:00Z",
      "verified": false,
      "title_review": "An OK smartphone with some nice features and poor battery life",
      "body": "After owning a Motorola Razr V3c for two years on the Verizon network I was ready to",
      "helpfulVotes": 1,
      "country": "United States",
      "id": "B000MQFG02-r0",
      "content_type": "review",
      "_version_": "1719034631149322240",
      "asin": "B000MQFG02"
    },
    {
      "name": "Amazon Customer",
      "rating_review": 4,
```

Indexing: Schema

Solr's default types:

- pint
- pfloat
- boolean
- string

Custom types:

- title
- text_body
- text_standard

Indexing fields relevant to textual queries and filtering

Type	Field	Indexed
pint	totalRatings	true
	rating_review	true
	helpfulVotes	true
pfloat	rating_item	true
	price	true
	originalPrice	true
boolean	verified	false
string	asin	false
	content_type	true
	url	false
	image	false
	reviewUrl	false
title	title_item	true
	title_review	true
text_body	body	true
text_standard	brand	true
	name	true
	country	true

Indexing: Custom field types

Class: TextField

Tokenizer: StandardTokenizerFactory

Type	Filter	Index	Query
title	EdgeNGramFilterFactory	true	false
	ASCIIFoldingFilterFactory	true	true
	EnglishMinimalStemFilterFactory	true	true
	LowerCaseFilterFactory	true	true
text_body	StopFilterFactory	true	true
	ASCIIFoldingFilterFactory	true	true
	EnglishMinimalStemFilterFactory	true	true
	LowerCaseFilterFactory	true	true
text_standard	ASCIIFoldingFilterFactory	true	true
	LowerCaseFilterFactory	true	true

Retrieval Process

- eDisMax Parser
- 5 information needs
- 3 different system configurations:
 - S1 - a baseline, filterless system, where all the documents have either Solr's default types, or our custom types but without any filters;
 - S2 - a system with our complete specified custom schema;
 - S3 - a system with the schema, and also with weights where the relevant fields are boosted in the retrieval process;

Field	Weight
title_item	1.5
title_review	1.7
body	1.3

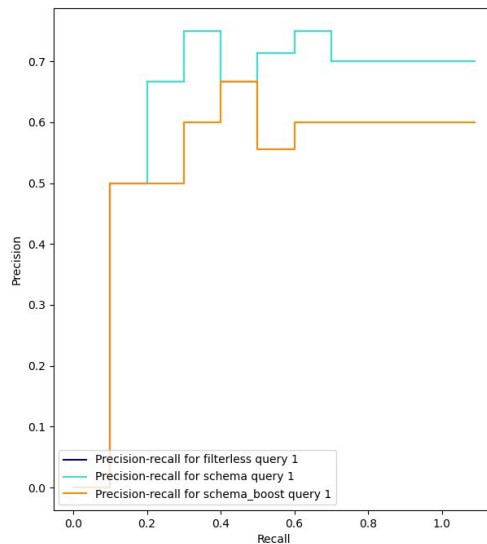
IN: Mexican Reviews

Reviews made in Mexico or
reference that country

q: Mexico

qf: country title_review body

fq: content_type:review



	AvP	P@10
S1	0.70839	0.7
S2	0.70839	0.7
S3	0.624603	0.6

IN: Christmas Gifts

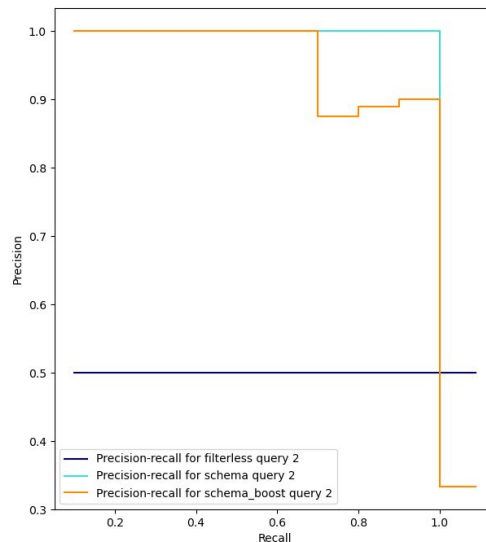
See several opinions on phones
in the 2020 Christmas season,
so we can choose a good gift
for this Christmas.

q: Gift

qf: body title_review

fq: content_type:review

date:[2020-11-01T00:00:00Z TO
2021-01-31T00:00:00Z]



	AvP	P@10
S1	1.0	0.1
S2	1	1
S3	0.969798	0.9

IN: Relevant 5-star reviews

Positive and relevant 5-star reviews to the community.

q: Great Good

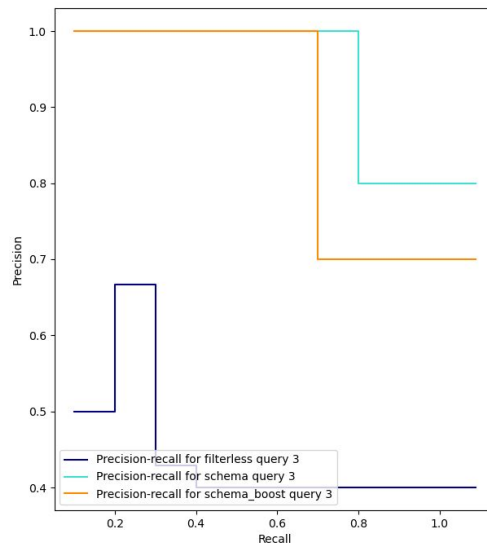
qf: title_review body

fq: content_type:review

verified:true

rating_review:5

helpfulVotes:[10 TO *]



	AvP	P@10
S1	0.729167	0.4
S2	1	0.8
S3	1	0.7

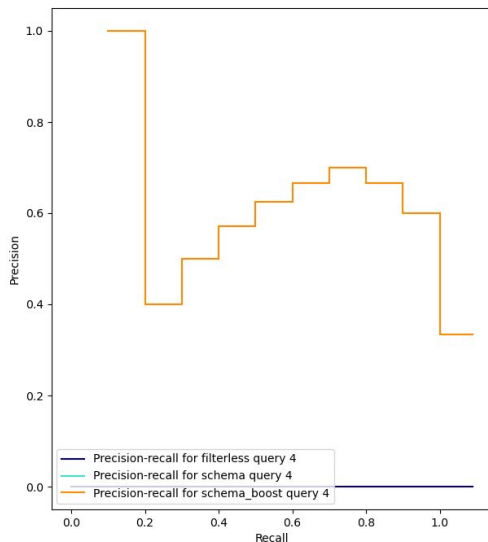
IN: Google or Apple Phones

Search Google phones with
4g connectivity or refurbished
iPhones

q: (Google AND 4g) OR
(iphone AND refurbished)

qf: brand title_item

fq: content_type:item



	AvP	P@10
S1	0	0
S2	0.710768	0.7
S3	0.710768	0.7

IN: Samsung discounts

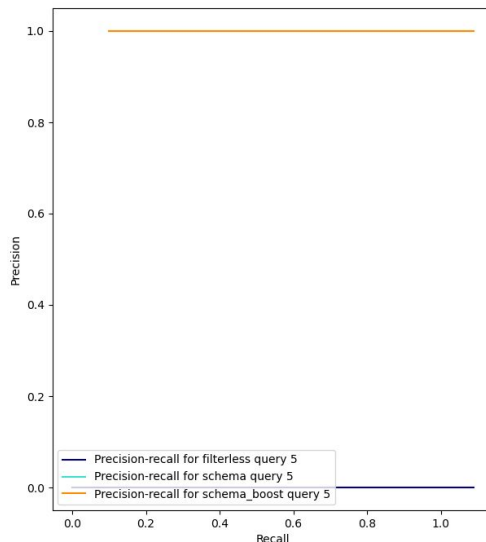
Samsung phones, in
discount, ordered by
highest absolute discount

q: samsung

qf: title_review body

fq: content_type:item
-originalPrice:0.0

sort: dist(2, originalPrice, 0, price, 0)
desc



	AvP	P@10
S1	0	0
S2	1	1
S3	1	1

Conclusions and Questions

