amazon Cell Phone Reviews

PRI · Information Processing and Retrieval, 2021/22

M.EIC · Master in Informatics and Computing Engineering

T4G5 · Milestone #3
André Gomes - up201806224
André Nascimento - up201806461
Catarina Fernandes - up201806854

Item data improvement

- Not enough information in the textual fields to satisfy more complex information needs
- More scraping from products page, retrieving new textual fields:
  - **Description** - long text field, with some of the item's characteristics and catchy features
  - **About** - main points of the product and some details that the buyer should be aware of
  - **More** - multi-valued key-value field that contains the items features
- Extracted main features from the *More* field, to be later used on filtering the results
  - wireless carrier
  - screen size
  - color
  - memory storage capacity
  - cellular technology
  - operating system

Item data improvement: Schema

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

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
	brand	true
	url	false
	image	false
	reviewUrl	false
	name	false
	country	true
	wireless_carrier	true
	operating_system	true
	color	true
	screen_size	true
	memory_storage_capacity	true
	cellular_technology	true
title	title_item	true
	title_review	true
	more	true
text_body	body	true
	description	true
	about	true

Information Retrieval Tool Improvements

- Analyze and explore Solr's searching features, like: *Faceting, Spell Checking, Query Re-Ranking, Suggester, MoreLikeThis,* and *Pagination of Results.*

☑ Faceting
☐ Spell Checking
☑ Query Re-ranking
☐ Suggester
☐ MoreLikeThis
☑ Pagination of results

IR Tool Improvements - Faceting

- Using *Term Facets*, to return buckets with unique values of a field and their counts

- Need to specify *field* and *limit* parameters in the API

- Applied to brand, *wireless_carrier*, *operating_system*, *color*, *screen_size*, *memory_storage_capacity*, *cellular_technology*, for the items, and *country* and *verified*, for the reviews

- Useful to implement filter checkboxes in the UI

```
"facet_counts":{
  "facet_queries":{},
  "facet_fields":{
    "wireless_carrier":[
      "Unlocked",705,
      "T-Mobile",226,
      "AT&T",171,
      "Verizon",166,
      "Sprint",34,
      "Straight Talk",20,
      "Tracfone",17,
      "3",12,
      "Cricket",12,
      "Boost Mobile",11,
      "Verizon Wireless",9,
      "Net10",7,
      "Simple Mobile",7,
      "WhatsApp SIM",7,
      "Virgin Mobile",6,
      "MVNO",4,
      "Telcel",4,
```

IR Tool Improvements - Pagination

- Define *start* and *rows* parameters:

    - **start** - set the the current "page", starting in 0, by defining the offset of the documents. start = page_number * rows

    - **rows** - specifies how many results we want to display per page

- Useful to present the search results in several pages

- *"Re-Ranking allows you to run a simple query for matching documents and then re-rank the top N documents using the scores from a different, more complex query"*

- Learning To Rank (LTR) module we can configure the features and the ranking models.

- 2 linear models developed, one for the items and another for the reviews

IR Tool Improvements - LTR

Original query score used in both models

Item's features:
- Rating
- Number of total ratings
- Discount (%)

Review's features:
- Number of helpful votes
- Length of the text
- Recency

Feature	Weight
originalScore	0.4
itemRating	0.1
itemTotalRatings	0.25
itemDiscount	0.25

Feature	Weight
originalScore	0.4
reviewVotes	0.3
reviewRecency	0.1
reviewBodyLength	0.2

User Interface

Frontend:
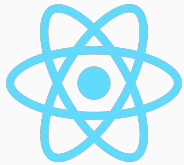
Backend:

Implemented interfaces:

- Item search page
- Item page
- Review search page

Item search interface

Item page

phoneprime   Phones   Reviews

Tracfone Samsung Galaxy A11 4G LTE
Prepaid Smartphone (Locked) - Black -
32GB - Sim Card Included - CDMA

★★★★⯪ 9126 ratings

91.07$ 99.99$ Save 8.92$ (9%)

ASIN	B08J4LFDSR
Carrier	Tracfone
Operating System	Android
Color	Black
Screen Size	6.4"
Storage Capacity	32 GB
Cellular Connectivity	LTE

About this product

- Introducing the new A Series. The Samsung Galaxy A11 combines smartphone essentials with the trusted reliability of Samsung. Take crisp, clear photos and videos with our powerful triple lens camera. Enjoy cinematic clarity on our 6.4" edge-to-edge screen.
- 6.4" HD Screen; 1.8 GHz plus 1.8 GHz Octa-Core Processor; Android 10.0; Triple 13MP plus 5MP plus 2MP Rear Cameras/8MP Front Facing Camera; Internal Memory 32 GB (device only); Supports microSD Memory Card up to 512GB (NOT INCLUDED)
- 4G LTE; Wi-Fi Capable; Bluetooth 4.2 wireless technology; GPS Capable
- CARRIER: This phone is locked to Tracfone, which means this device can only be used on the Tracfone wireless network.

Product details

With the Samsung Galaxy A11, you get exactly what you need with the performance and reliability from a brand you can trust. Do more with a long-lasting battery, keep all of your stuff with a 32GB built-in memory, and experience it on a 6.4"HD Screen. Finally, the phone you need to keep connected with friends and family—and at just the right price. With a long-lasting battery, the Galaxy A11 gives you more time to post, talk, text and share with friends and family. Don't worry, keep it all Keep more with 32GB of built-in memory, so you can keep all of your photos, shows, movies and apps, without having to delete to make room. Go edge-to-edge Enjoy a bold, bright viewing experience on a 6.4" screen. With an Infinity Display that goes edge-to-edge, you can see more story before you scroll. Capture the moment The Galaxy A11 lets you make everyday moments special. Take crisp, clear photos with an 13MP rear camera and perfect selfies with the kids on a 8MP front camera.

Check out this product's reviews

Review search interface

Retrieval Process - Re-evaluation

- eDisMax Parser
- Same 5 information needs from M2 (only 3 will be presented)
- 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, with weights where the relevant fields are boosted in the retrieval process and with use of LTR

Field	Weight
title_review	1.7
body	1.3

Query fields for Reviews

| Field	Weight
title_item	2
description	1.5
about	1
more	1

Query fields for Items

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]

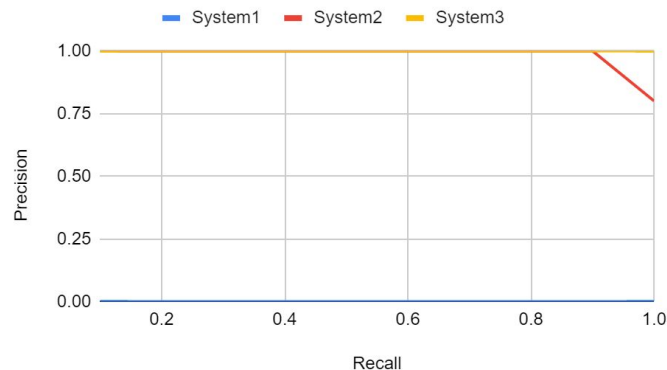|	System1	System2	System3		Precision @				Recall @		
					System1	System2	System3		System1	System2	System3
1	1	1	1		1.00	1.00	1.00		0.50	0.25	0.14
2	1	1	1		1.00	1.00	1.00		1.00	0.50	0.29
3	0	0	1		0.67	0.67	1.00		1.00	0.50	0.43
4	0	0	1		0.50	0.50	1.00		1.00	0.50	0.57
5	0	0	1		0.40	0.40	1.00		1.00	0.50	0.71
6	0	0	1		0.33	0.33	1.00		1.00	0.50	0.86
7	0	1	0		0.29	0.43	0.86		1.00	0.75	0.86
8	0	0	0		0.25	0.38	0.75		1.00	0.75	0.86
9	0	0	0		0.22	0.33	0.67		1.00	0.75	0.86
10	0	1	1		0.20	0.40	0.70		1.00	1.00	1.00
				AvP	1.00	0.71	0.96				

Precision-Recall Curve - Improved Systems - Q2
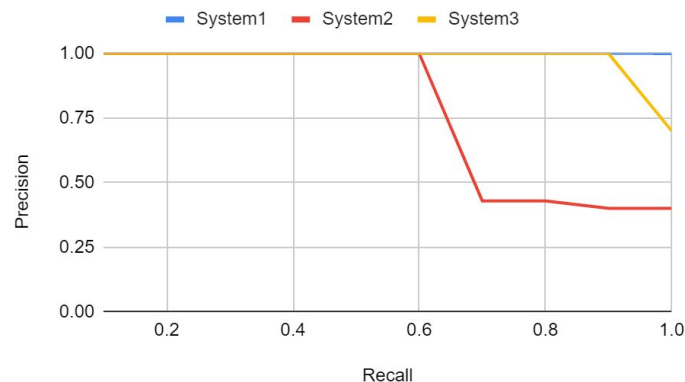
IN: Google or Apple Phones

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

|	System1	System2	System3		Precision @				Recall @		
					System1	System2	System3		System1	System2	System3
1	0	1	1		0.00	1.00	1.00		#DIV/0!	0.13	0.11
2	0	1	1		0.00	1.00	1.00		#DIV/0!	0.25	0.22
3	0	1	1		0.00	1.00	1.00		#DIV/0!	0.38	0.33
4	0	1	1		0.00	1.00	1.00		#DIV/0!	0.50	0.44
5	0	1	1		0.00	1.00	1.00		#DIV/0!	0.63	0.56
6	0	1	1		0.00	1.00	1.00		#DIV/0!	0.75	0.67
7	0	1	1		0.00	1.00	1.00		#DIV/0!	0.88	0.78
8	0	0	1		0.00	0.88	1.00		#DIV/0!	0.88	0.89
9	0	0	1		0.00	0.78	1.00		#DIV/0!	0.88	1.00
10	0	1	0		0.00	0.80	0.90		#DIV/0!	1.00	1.00
				AvP	#DIV/0!	0.98	1.00				

Precision-Recall Curve - Improved Systems - Q4

IN: Samsung discounts

Samsung phones, in discount, ordered by highest absolute discount

q: samsung
qf: title_item description about more
fq: content_type:item
   -originalPrice:0.0
sort: dist(2, originalPrice, 0, price, 0) desc

	System1	System2	System3		Precision @				Recall @		
					System1	System2	System3		System1	System2	System3
1	0	1	1		0.00	1.00	1.00		#DIV/0!	0.10	0.10
2	0	1	1		0.00	1.00	1.00		#DIV/0!	0.20	0.20
3	0	1	1		0.00	1.00	1.00		#DIV/0!	0.30	0.30
4	0	1	1		0.00	1.00	1.00		#DIV/0!	0.40	0.40
5	0	1	1		0.00	1.00	1.00		#DIV/0!	0.50	0.50
6	0	1	1		0.00	1.00	1.00		#DIV/0!	0.60	0.60
7	0	1	1		0.00	1.00	1.00		#DIV/0!	0.70	0.70
8	0	1	1		0.00	1.00	1.00		#DIV/0!	0.80	0.80
9	0	1	1		0.00	1.00	1.00		#DIV/0!	0.90	0.90
10	0	1	1		0.00	1.00	1.00		#DIV/0!	1.00	1.00
				AvP	#DIV/0!	1.00	1.00				

Precision-Recall Curve - Improved Systems - Q5
System1 System2 System3
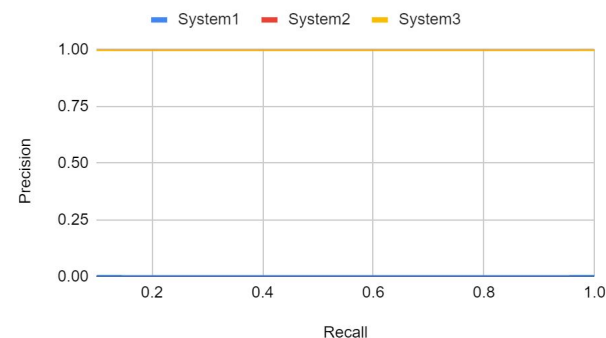
Final Results

Query	System 1		System 2		System 3	
	AvP	P@10	AvP	P@10	AvP	P@10
1	0.79	0.4	0.77	0.6	1	0.8
2	1	0.2	0.71	0.4	0.96	0.7
3	0.39	0.3	0.46	0.4	0.8	0.6
4	0	0	0.98	0.8	1	0.9
5	0	0	1	1	1	1
MAP	0.436		0.784		0.952	

- Improved System Results

- Initial System Results

Query	System 1		System 2		System 3	
	AvP	P@10	AvP	P@10	AvP	P@10
1	0.70	0.6	0.73	0.7	0.63	0.5
2	1	0.1	0.99	0.9	0.98	0.9
3	0.73	0.4	1	0.8	1	0.7
4	0	0	0.7	0.6	0.7	0.6
5	0	0	1	1	1	1
MAP	0.286		0.884		0.862	