

COMMUNITY

THE ASF CONFERENCE

CODE

**Transforming your
search with
cutting-edge
techniques and LLMs**

Atita Arora

07 October 2023

About me



Started 2008



- Computer Applications
- Strategic Business Management



- Vector / Semantic Search
- Language analysis
- Information retrieval



Mom of 2 boys



Opensource



Loves to travel,
eat, cook

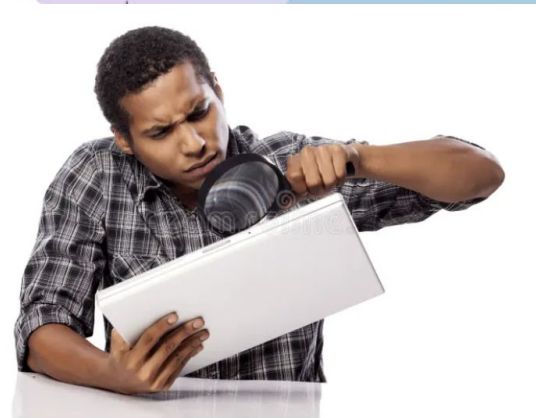


Atita Arora

Search Architect / Engineer /
Enthusiast for NLP, Vectors, ML & AI.



Searching for keywords like it's 1999...



Search apps powered by neural search and LLMs! 🧠 🔍

The Future of Search Engines in a World of AI and LLMs

Building Vector Search in Chorus: A Technical Deep Dive



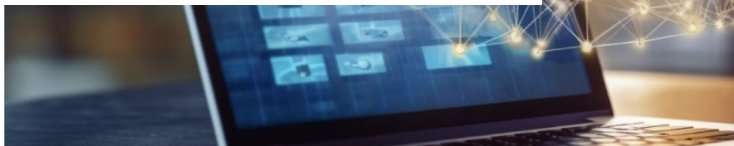
March 22, 2023

[Atifa Arora](#)

Category: [Ecommerce](#)

In my [last blog](#) we established how vectors function internally and how they could be a game-changer in the world of e-commerce. We hope the first blog helped set the tone for a more focussed discussion of the technical details of our implementation of vector search in [Chorus](#), our open source reference implementation of e-commerce search.

So how does vector search in Chorus really work?



LLM-Powered Search Applications: A New Era of Semantic Search



Fred C. Siika · Follow

3 min read · Aug 6



There's no denying that the digital age has made information more accessible than ever before. However, with the information explosion, finding the precise piece of data you need has become increasingly difficult. Enter the world of semantic search, where search engines understand user intent and the contextual meaning of terms, thus offering a new level of

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.

... into the search engine's world.



Home / Articles / Vector Similarity Beyond Search

Vector Similarity beyond Search

When making use of unstructured data, there are traditional go-to solutions that are well-known for developers:

- **Full-text search** when you need to find documents that contain a particular word or phrase.
- **Vector search** when you need to find documents that are semantically similar to a given query.

Sometimes people mix those two approaches, so it might look like the vector similarity is just an extension of full-text search. However, in this article, we will explore some promising new techniques that can be used to expand the use-case of unstructured data and demonstrate that vector similarity creates its own stack of data exploration tools.



JULY 24, 2023 · ADRIEN GRAND

Vector search in Elasticsearch: The rationale behind the design

There are different ways to implement a vector database, which have different trade-offs. In this blog, you'll learn more about how vector search has been integrated into Elasticsearch and the trade-offs that we made.

9 min read · Vector Search, ML Research

The Evolution of Language Models - From Word Embeddings to Multimodal Transformers

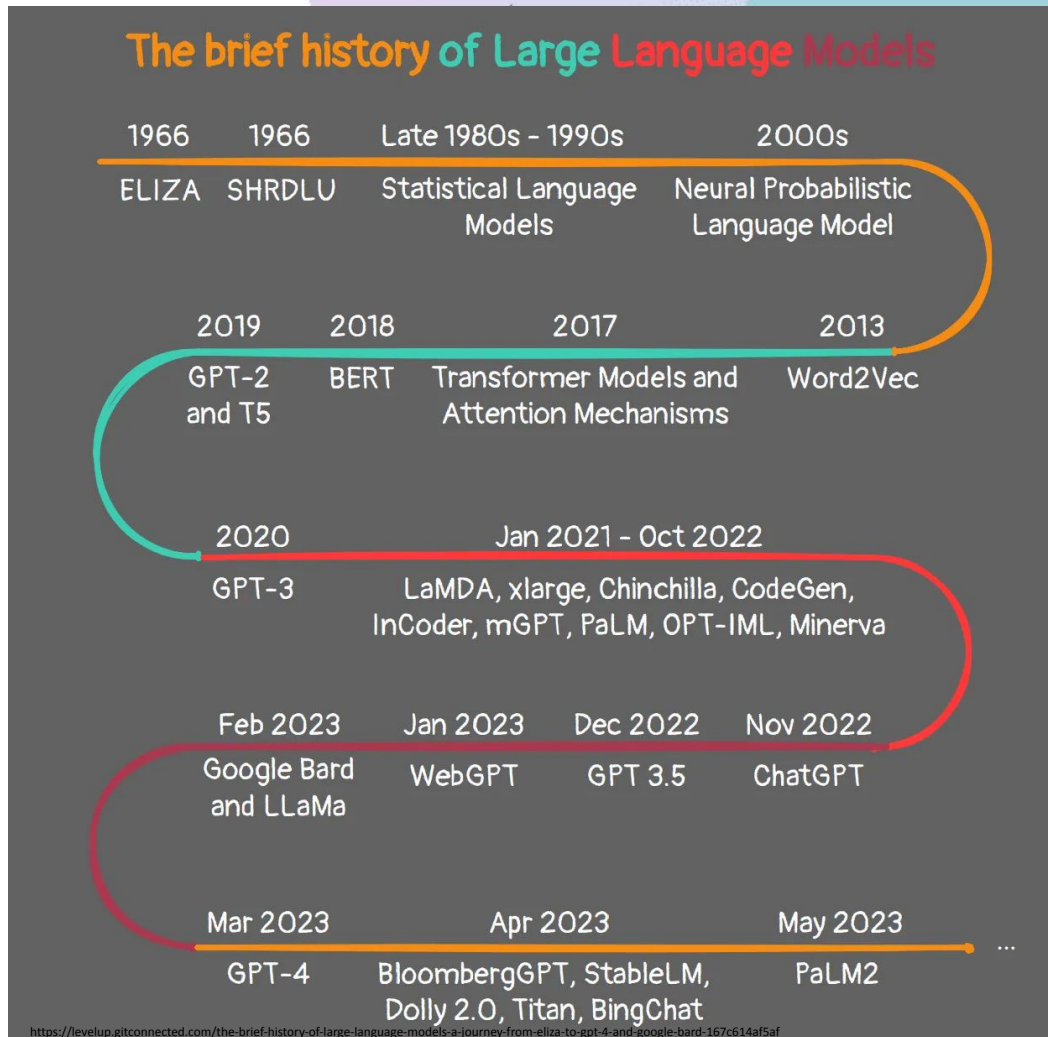
Key Time periods

Pre 2010 - Focus on statistical models and predefined rules

2010 - Introduction of Word Embeddings

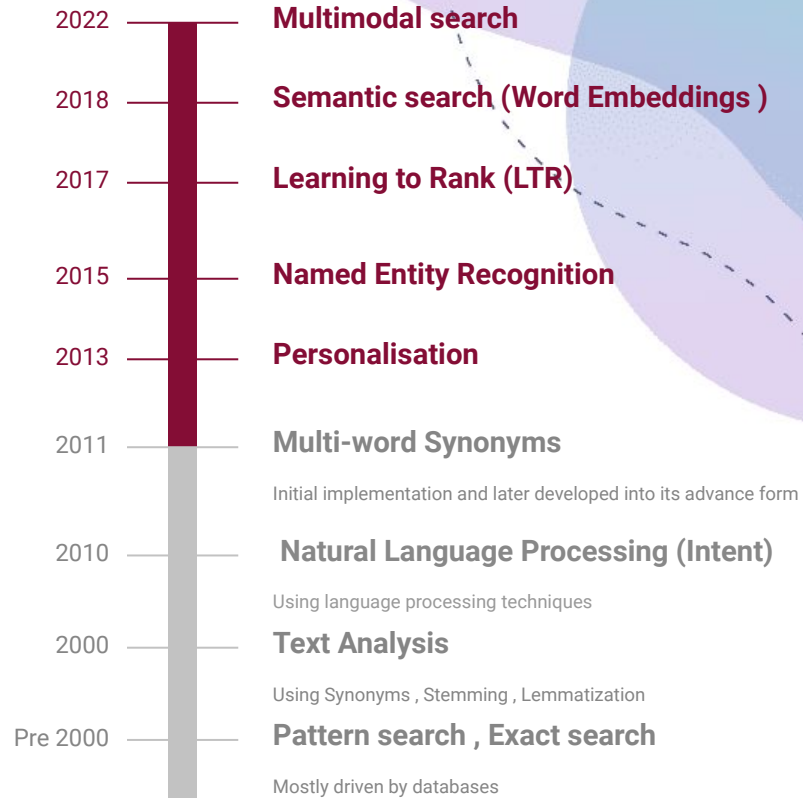
2018 - Introduction of Transformers

And rest is just the history !!



The Evolution of Search Applications

1. Pattern / Exact Search (Early Years)
2. Synonyms + Stemming (Mid-2000s)
3. Natural Language Processing (2010)
4. Multi-word Synonyms
5. Personalisation
6. Entity Recognition
7. Semantic search
8. Multimodal / Multilingual search



"Hemdjacke mit Brusttaschen " (23)

Sortierung : Beliebtheit ▾ Produktansicht ▾

Marke ▾ Größe ▾ Farbe ▾ Preis ▾ Sale Store ▾ Material ▾ Saison ▾



We Care Together

Armedangels

Hemdjacke mit Brusttaschen Modell 'AL...

189,99 €



45%

Minimum

Hemdjacke mit Brusttaschen Modell 'PR...

59,99 € ~~109,99 €~~



We Care Together

Armedangels

Hemdjacke mit Brusttaschen Modell 'AL...

189,99 €



44%

Drykorn

Hemdjacke mit Brusttaschen Modell 'EN...

99,99 € ~~179,99 €~~

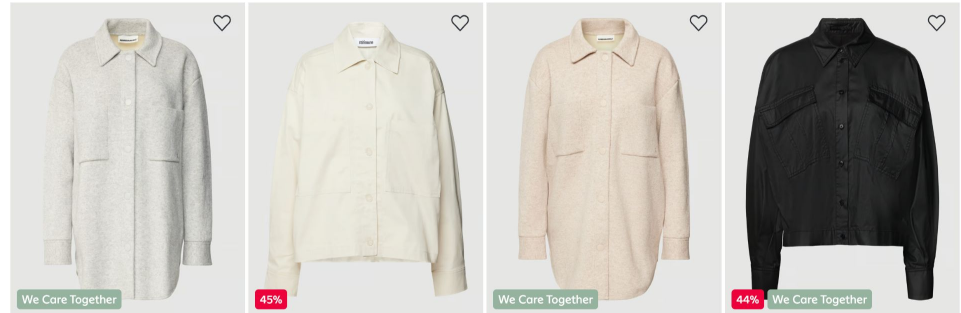


Shirt Jacket 'with" Chest Pockets

"Hemdjacke ohne Brusttaschen " (3347)

Sortierung : Beliebtheit ▾ Produktansicht ▾

Marke ▾ Größe ▾ Farbe ▾ Preis ▾ Sale Store ▾ Material ▾ Saison ▾



We Care Together



XS S M L XL
189,99 €

45%

Minimum

Hemdjacke mit Brusttaschen Modell 'PR...

59,99 € ~~109,99 €~~



We Care Together

Armedangels

Hemdjacke mit Brusttaschen Modell 'AL...

189,99 €



44%

Drykorn

Hemdjacke mit Brusttaschen Modell 'EN...

99,99 € ~~179,99 €~~

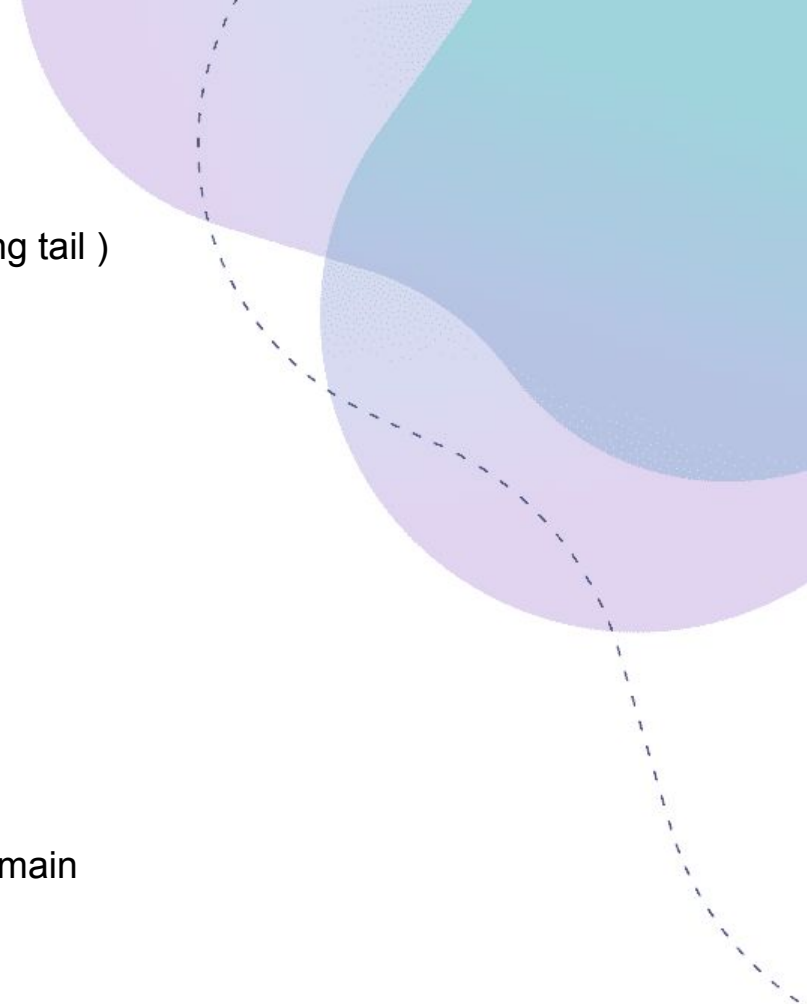


Shirt Jacket 'without' Chest Pockets

Challenges of the Keyword Search

1. Actual user query intent is still a mystery! (language , long tail)
2. Vocabulary mismatch
3. Missing standardization
4. Variety of attributes
5. Trends and Popularity
6. Multimodal Content (Text , Image, Audio , Video)

.....And there are certainly many more domain level challenges!!



Vector / Semantic Search comes to the Rescue !!

Text , Image,
Audio , Video

- Involves a way of representing "an object" using a list of characteristics or features.
- Vector search involves converting a "source object" into a vector embedding using a desired transformer, then index and search through the vector embeddings.

Encoding Example :

[-0.06275664, -0.038876858, -0.031730175,.....-0.027633812]

- The goal of semantic search is to provide search results that align with the user's intent, even if the query and indexed content do not match precisely in terms of keywords.

iphone with
transparent backcover

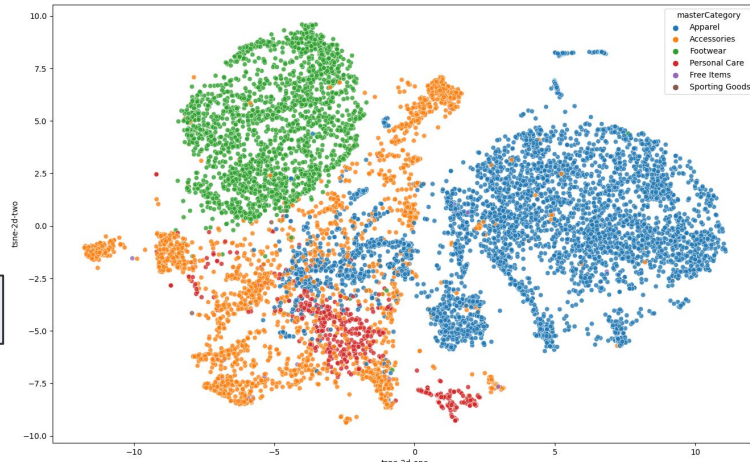


transparent
backcover for iphone

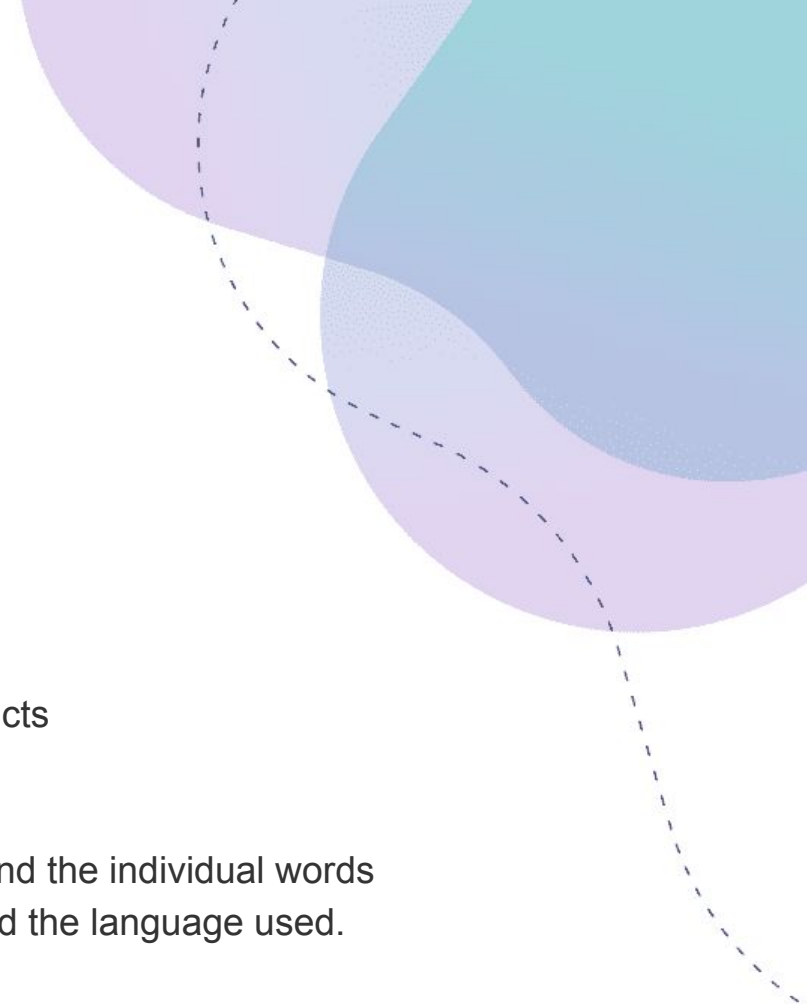
Convertible laptop
with touchscreen

==

Touchscreen laptop
with 360-degree
hinge



How can it help ?

- Fix recall issues caused by:
 - long-tail queries
 - vocabulary mismatch
 - multilingual search
 - Multimodal (Use text , images , audio , video)
 - Also exposes a lot of potential in :
 - Marketing and Promotions - similar users
 - Recommendations and Suggestions - similar products
 - Serchandising - filter , facets , synonyms
 - Boosting precision by leveraging meaning that goes beyond the individual words themselves and focuses on the **context and intent** behind the language used.
- 

Query : black shirt to go with blue jeans men

<- Results with default algo

blacklight History

Default Algo black shirt to go with blue jeans men Search Q

Start Over black shirt to go with blue jeans men

Limit your search << Previous | 1 - 30 of 4,011 | Next >> Sort by relevance 30 per page

Gender >

Product Type >

Product Colour >

Categories >

Total hits: 3936.0 | Entropy total: 3.7878891516650826

tshirt shoes shirt casual

Myntra Men Blue Dying To Ride T-shirt Myntra Men Back To The 70s Black T-shirt Spykar Men Ep Jeans Blue Jeans Spykar Men Ep Jeans Blue Jeans

Pepe Jeans Women Blue Jeans Locomotive Men Blue Jeans Spykar Men Blue Jeans Denizen Men Blue Jeans

Horsefly Men Blue Jeans Denizen Men Blue Jeans Myntra Men Back To The 70s Cream T-shirt Locomotive Men Washed Blue Jeans

Results with match by image vector ->

blacklight History

Query match by black shirt to go with blue jeans men Search Q

Start Over Query match by image vector black shirt to go with blue jeans men

Limit your search << Previous | 1 - 30 of 200 | Next >> Sort by relevance 30 per page

Gender >

Product Type >

Product Colour >

Categories >

Total hits: 203.0 | Entropy total: 1.932059927736722

tshirt shirt top jean

Scullers Men Black Shirt Wrangler Men Rider Solid Black Shirts John Players Men Solid Black Shirt United Colors of Benetton Men Black T-shirt

Indigo Nation Men Young Lords Black T-shirt UCB Men Ben Logo Black Shirt Locomotive Men Black Shirt Indigo Nation Men Black Shirt

Quiksilver Men Black Shirt Classic Polo Men's Cari Black Polo T-shirt UCB Men's Johnny Collar With Two Tone Black T-shirt Indigo Nation Men Bling Blue Shirt

Query : date night dress for a girl

blacklight History

Default Algo | date night dress for a girl | Search Q

Start Over | date night dress for a girl

Limit your search | « Previous | 1 - 30 of 222 | Next » | Sort by relevance | 30 per page

Gender | Product Type | Product Colour | Categories

Total hits: 234.0 | Entropy total: 3.3679003217425154

dresse top night suit

Doodle Girls Pink Printed Dress
Doodle Girls Blue Printed Dress
Doodle Girls Orange Printed Dress
Doodle Girls Pink Printed Dress

Doodle Girls Pink Printed Dress
Doodle Girls Blue Washed Dress
Doodle Kids Girl Printed White Dress
Doodle Kids Girl Printed Lavender Dress

Palm Tree Girls Beyonce White & Blue Dress
Doodle Kids Girl Printed Pink Dress
Doodle Kids Girl Printed Pink Dress
Doodle Kids Girl Printed Pink Dress

<- Results with default algo

blacklight History

Query match by | date night dress for a girl | Search Q

Start Over | Query match by text vector > date night dress for a girl

Limit your search | « Previous | 1 - 30 of 200 | Next » | Sort by relevance | 30 per page

Gender | Product Type | Product Colour | Categories

Total hits: 248.0 | Entropy total: 3.5766036802340637

dresse watche nightdress kurta

Forever New Women Multi Coloured Dress
Avirate Women Black Dress
AND Women Pink Dress
Avirate Black Dress

ONLY Women Peach Dress
Avirate Black Formal Dress
AND Women Red Dress
Sepia Women Black Dress

Elle Women Black Dress
109F Women Multi Coloured Dress
Elle Black Sheath Dress
AND Women Solid Grey and Black Dress

Results with match by text vector ->

Query : smartwatch by samsung

blacklight History

Default Algo smartwatch by samsung Search Q

Start Over smartwatch by samsung ✕

Limit your search « Previous | 1 - 30 of 141 | Next » Sort by relevance ▾ 30 per page ▾

Total hits: 3.0 | Entropy total: 1.584962500721166

Brands >

Product Type >

Product Colour >

Categories >

Released >

Notebook PC Renewal

Eaton 68604 warranty/support extension

Samsung RS68N8240B1 side-by-side refrigerator Freestanding Black 617 L A+

ASUS P5Q SE/R LGA 775 (Socket T) ATX

Samsung Galaxy Watch Active2 smartwatch SAMOLED 3.05 cm (1.2") Black GPS (satellite)

Denver SW-350 smartwatch IPS 3.3 cm (1.3") Black

Samsung Galaxy Watch Active2 smartwatch SAMOLED 3.43 cm (1.35") Black GPS (satellite)

Denver SW-510BLACK smartwatch 3.3 cm (1.3") Black GPS (satellite)

ASUS P7P5D-E Deluxe LGA 1156 (Socket H) ATX Intel® P55

<- Results with default algo

Results with match by image vector ->

blacklight History

Query match by smartwatch by samsung Search Q

Start Over Query match by image vector > smartwatch by samsung ✕

Limit your search « Previous | 1 - 30 of 200 | Next » Sort by relevance ▾ 30 per page ▾

Total hits: 16.0 | Entropy total: 1.6737949406953985

Brands >

Product Type >

Product Colour >

Categories >

Released >

Headphones PC Processor 1

Samsung Galaxy Watch Active2 smartwatch SAMOLED 3.05 cm (1.2") Black GPS (satellite)

Samsung Galaxy Watch Active2 smartwatch SAMOLED 3.56 cm (1.4") Black, Pink gold GPS (satellite)

Samsung Galaxy Watch Active2 smartwatch SAMOLED 2.79 cm (1.1") Silver GPS (satellite)

Huawei WATCH GT 2e smartwatch AMOLED 3.53 cm (1.39") Metallic GPS (satellite)

Huawei WATCH GT 2e smartwatch AMOLED 3.53 cm (1.39") Black GPS (satellite)

Huawei WATCH GT 2e smartwatch AMOLED 3.53 cm (1.39") Silver GPS (satellite)

Huawei WATCH GT 2e smartwatch AMOLED 3.53 cm (1.39") Black GPS (satellite)

Samsung Galaxy Watch Active 2

Query : Men shorts in size m

<- Results with default algo

blacklight History

Default Algo | men shorts in size m | Search

Start Over | men shorts in size m

Limit your search | < Previous | 1 - 30 of 2,622 | Next > | Sort by relevance | 30 per page

Gender > | Total hits: 2895.0 | Entropy total: 3.830474796864112

Product Type > | shoes | tshirt | shirt | casual

Product Colour >

Categories >

ADIDAS Men M Basic White Shorts

Boss Men In Motion Perfume

Boss In Motion Men Perfume

M tv Men Grey Sunglasses

Jacques M Women Fragrances

Quechua Arpenaz 200 M

ADIDAS Men Duramo 4 M Black Sports Shoes

M tv Men My Fab Eyewear Brown Sunglasses

ADIDAS Men Furano 3 M Black Sports Shoes

ADIDAS Men Adi Fiero M White Sports Shoes

M tv Men My Fab Eyewear Black Sunglasses

ADIDAS Men Black Shorts

blacklight History

Query match by | men shorts in size m | Search

Start Over | Query match by text vector > men shorts in size m

Limit your search | < Previous | 1 - 30 of 200 | Next > | Sort by relevance | 30 per page

Gender > | Total hits: 221.0 | Entropy total: 3.3109454333008776

Product Type > | short | watche | brief | shoes

Product Colour >

Categories >

ADIDAS Men Black Shorts

Status Quo Men Solid Black Shorts

ADIDAS Men Blue Shorts

ADIDAS Men's Cont Black Short

ADIDAS Men's Black Football Shorts

ADIDAS Men Solid Navy Blue Shorts

ADIDAS Men Cfc H Sho Blue Shorts

Kalenji Men Essential Baggy Black Shorts

ADIDAS Men's Bas Exs Black Short

Puma Men Black Shorts

Puma Men Black Shorts

Nike Men Training Black Shorts

Results with match by text vector ->

Query : Fashionable formal footwear for women

<- Results with default algo

blacklight History

Default Algo fashionable formal footwear for women Search Q

Start Over fashionable formal footwear for women ✖

Limit your search « Previous | 1 - 30 of 3,126 | Next » Sort by relevance ▾ 30 per page ▾

Total hits: 3442.0 | Entropy total: 3.901779264478818

shoes casual sport kurta

Gender >

Product Type >

Product Colour >

Categories >

Scullers For Her Women Pink Shirt

Scullers For Her Women Grey Trousers

Scullers For Her Blue Shirt

Scullers For Her Black Trousers

Enroute Men Formal Black Formal Shoe

Provogue Men Formal Black Formal Shoes

Franco Leone Men Formal Black Formal Shoes

Franco Leone Men Formal Black Formal Shoes

Franco Leone Men Formal Black Formal Shoes

Franco Leone Men Formal Brown Formal Shoes

Carlton London Men Formal Brown Formal Shoes

Nike Men's Zoom For Black Shoe

blacklight History

Query match by fashionable formal footwear for women Search Q

Start Over Query match by image vector > fashionable formal footwear for women ✖

Limit your search « Previous | 1 - 30 of 200 | Next » Sort by relevance ▾ 30 per page ▾

Total hits: 272.0 | Entropy total: 2.2710816381099406

heel shoes formal flat

Gender >

Product Type >

Product Colour >

Categories >

Catwalk Women Silver Heels

Rocia Women Brown Heels

Carlton London Women Classy Silver Sandals

Inc 5 Women Casual Gold Heels

Catwalk Women Silver Heels

HM Women Silver Heels

Inc. 5 Women Casual Silver Heels

Rocia Women Brown Heels

Clarks Women Black Leather Flat Shoes

HM Women Blue Heels

Catwalk Women Black Heels

Carlton London Women Black Heels

Results with match by image vector ->

Query : Bag for College

blacklight History

Default Algo bag for college Search

Start Over bag for college

Limit your search

« Previous | 1 - 30 of 102 | Next » Sort by relevance 30 per page

Total hits: 131.0 | Entropy total: 3.270353773849941

handbag bag laptop duffel

Be For Bag Women Pink Bag
Be For Bag Women Blue Tote Bag
Be For Bag Women Zoli Black Bag
Be For Bag Women Cream And Black Tote Bag

Be For Bag Women White And Red Striped Tote Bag
Scullers For Her Blue Shirt
Scullers For Her Beige Shirt
Scullers For Her Printed White Shirt

Scullers For Her Women Pink Shirt
Scullers For Her Black Trousers
Quechua Black Bag
FNF Red & Gold Collection for Wedding Sari

<- Results with default algo

blacklight History

Query match by bag for college Search

Start Over Query match by image vector > bag for college

Limit your search

« Previous | 1 - 30 of 200 | Next » Sort by relevance 30 per page

Total hits: 224.0 | Entropy total: 2.16351369113403

backpack handbag bag laptop

Nike Unisex Swoosh Black Backpacks
Nike Unisex Classic Turf Navy Blue Backpack
ADIDAS ESS Black Unisex Backpack
ADIDAS Classic Blue White Backpack

ADIDAS Unisex Bess Bp Black Backpacks
Puma Unisex Pink Backpack
Nike Unisex Blue Hayward Backpack
Nike Unisex Black Max Air Backpack

Puma Unisex Black Backpack
ADIDAS Men Tiro Black Red Backpack
ADIDAS Unisex Black BTS Backpack
ADIDAS Unisex Black Backpack

Results with match by image vector ->

Query : Blakdress with something shimmery

No Results with 'Default Algo'

The screenshot shows the blacklight search interface. At the top, the search bar contains the query "blakdress with something shimmery". Below the search bar, there are filters for "Gender", "Product Type", "Product Colour", and "Categories". The search results are displayed in a grid format, showing 12 items. Each item includes a small image of the dress and its name. The items are:

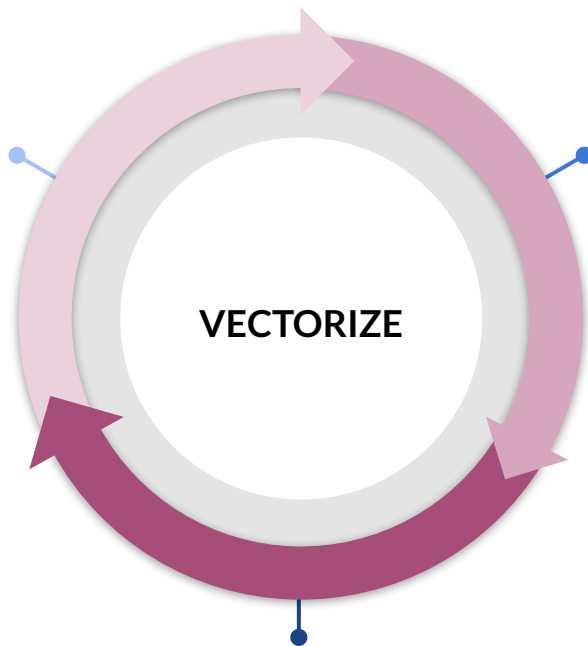
- Avirate Women Black Dress
- Avirate Black & Silvery Grey Dress
- Avirate Black Dress
- Elle Black Sheath Dress
- Avirate Black Formal Dress
- Sepia Women Black Dress
- Elle Women Black Dress
- Arrow Woman Black Dress
- Remanika Women Black Dress
- French Connection Black Dress
- French Connection Black Dress
- Elle Women Black Sheath Dress

Results with text vector ->

From Keywords to Semantic Search in 3 steps

1. Vector Extraction

- * Data gather
- * Config changes & fields selection
- * Model Selection
- * Adding Vectors to the data



2. Conversion and Transformation

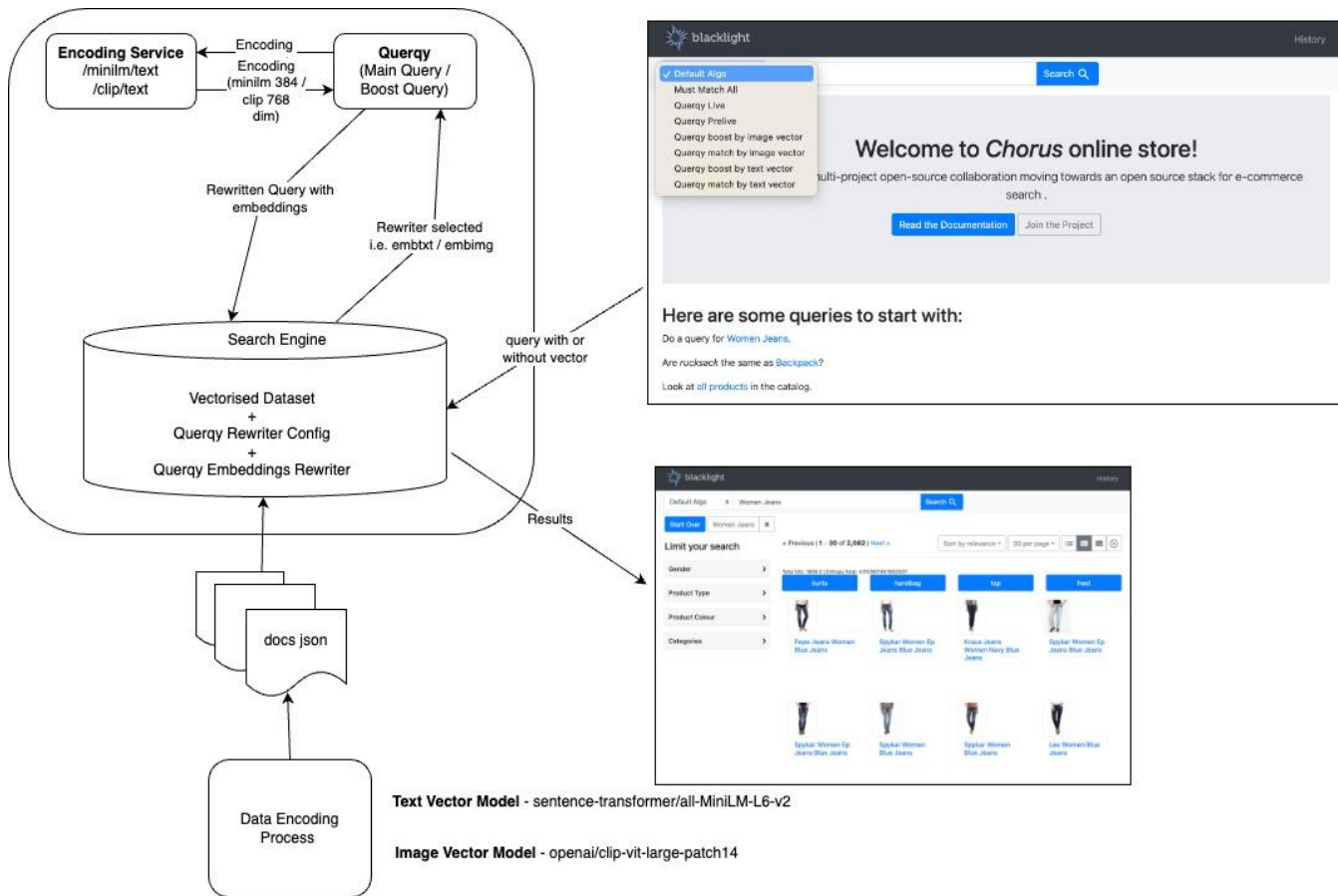
- * Encoding Service - Conversion of text query -> Text / Image Vector
- * Transformation of Search API to accommodate Vectors queries

3. Optimization for Relevancy Improvements and Efficiency

- * Offline Experimentation for relevancy improvements
- * Fine-tuning and performance optimizations

Conversion & Transformation

Vector Extraction

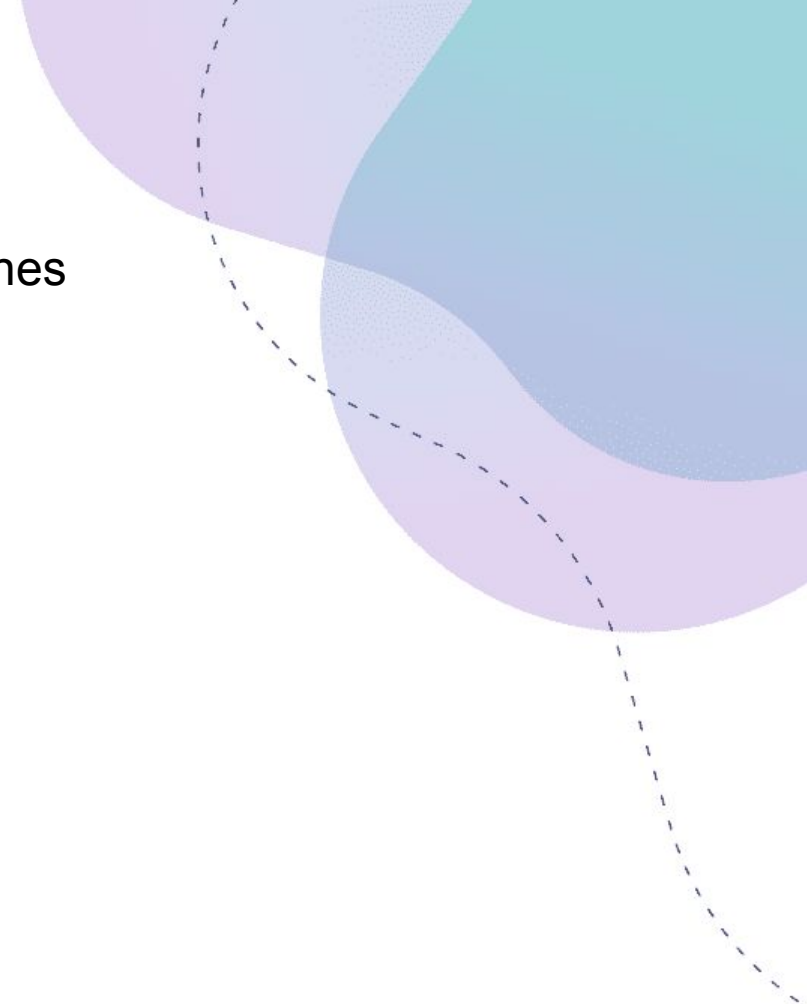


Implementation Architecture

Keep it Simple : From Theory to Practice

- Identify a designated business use-case to try Vector / Semantic Search
- Exemplify the value with a POC:
 - Using existing Lucene-based search or leverage dedicated vector search engines / DBs
 - Leverage pre-existing models (from Hugging Face)
- Experiment to prove improvements
- Focus on building change process understanding

Identifying what to Encode/Use

- Correlated with user queries / types of searches
 - Product Attributes
 - Images / Multimedia
 - Domain specific fields (allergens / occasion)
 - Categorical data
- 

Caveats

- Language model suitable for the business case
- Limitations of the models (Token length , language etc)
- Dimensions are not free - Infrastructural changes
- Fast and scalable vector encoding service
- Persistent 'K' retrievals

Step 3 - How to Evaluate ?

Categorise your user queries and evaluate individual capability of your search / IR

Exact Queries

phillips shb1100
Krupps f08801
Fossil Watch AM4183

Product Type Queries

Bluetooth headset
Backpack / rucksack
Perfumes

Symtomatic Queries

wintercap for women
craft kits for kids
cleaning stuff for equipments

Feature:

glow in dark toy
organising stuff for office
Longsleeve tshirt

Thematic:

Datanight dress for girls
LOL dolls
Organising stuff for office

Compatibility:

power adapter for lenovo
headsets with noise cancellation
Silver Toned Analogue Watch
Gym wear for men

Slang/Acronym:

Kicks
notebook
Beanie

Typos

smartwatch by sumsung
Toner catrdge from leximark brand
Blakdress with shimmery material

- Could be evaluated on # zero results
- nDCG@10 , MAP , MRR etc
- Conversions / ATB
- # Search refinements
- Ctr , abandonment etc

Evaluating the vector / semantic retrieval

Query Type / Algo	Exact	Product Type	Symptoms	Feature	Thematic	Compatibility	Slang/ Acronym	Typos	Avg nDCG@10
BM25	0.76	0.47	0.44	0.44	0.04	0.78	0.34	0.15	0.44
Match by Text Vector	0.26	0.87	0.59	0.67	0.63	1	0.88	0.90	0.70
Match by Image Vector	0.50	0.73	0.50	0.72	0.66	0.80	0.84	0.64	0.66
Boost by Text Vector	0.55	0.65	0.62	0.66	0.10	1	0.85	0.92	0.66
Boost by Image Vector	0.58	0.67	0.50	0.65	0	0.80	0.84	0.78	0.60

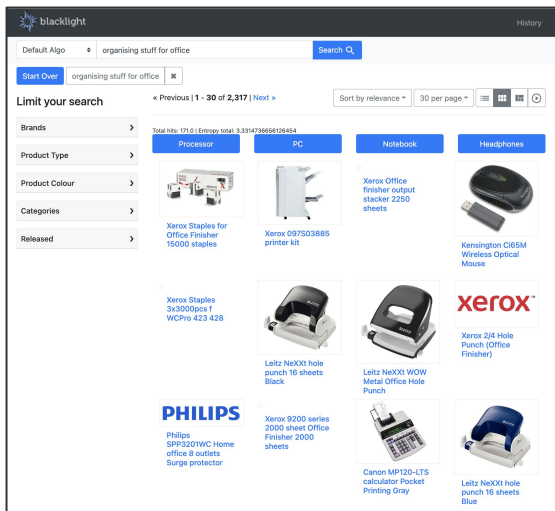
Hybrid

Other things to experiment with

- Fine-tuned / self trained models
- Combining Vectors with managed searches*
- Vector based Re-ranking model
- Combining scores from text + image search results

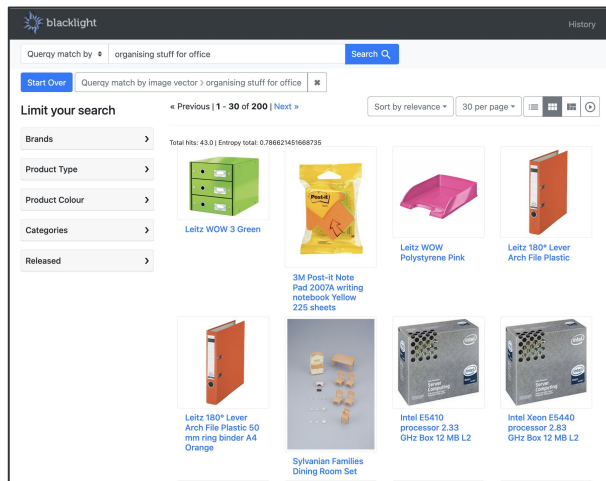
*Promoted searches , legal binding , brand conflicts etc

Getting Better through Fine Tuning..

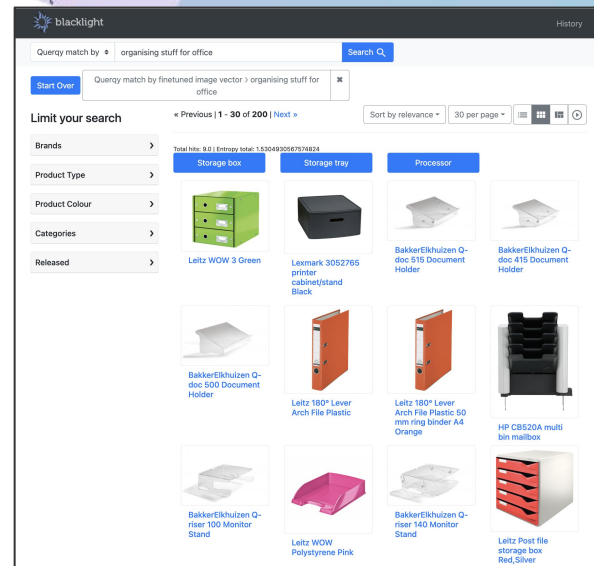


Results with default algo

Query : organising stuff for office



Results with match by image model



Results with fine-tuned image model

Fine-tuned Model courtesy : Jina.ai

What are other implementations ?

Semantic Retrieval

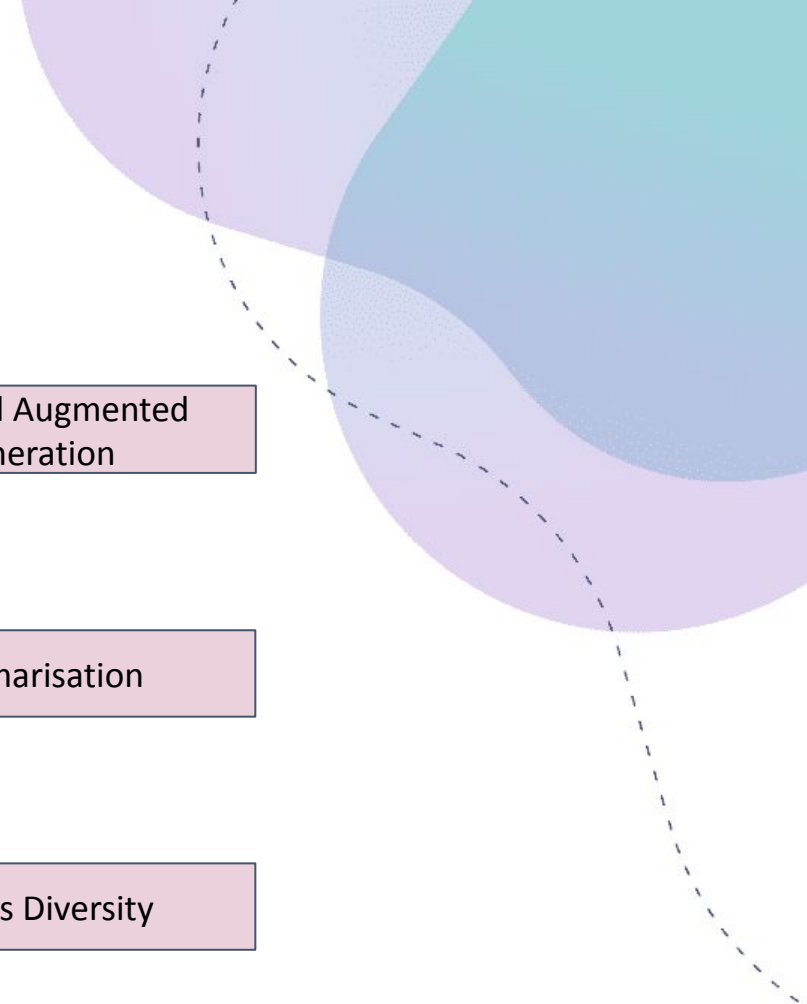
Retrieval Augmented
Generation

Auto Data Labelling

Summarisation

Anomaly Detection

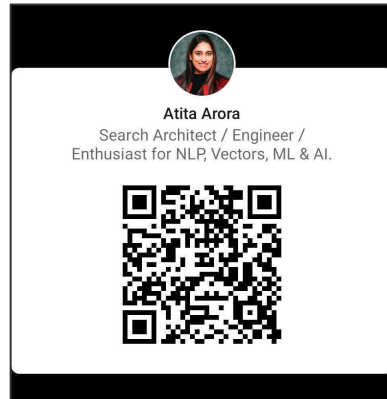
Results Diversity

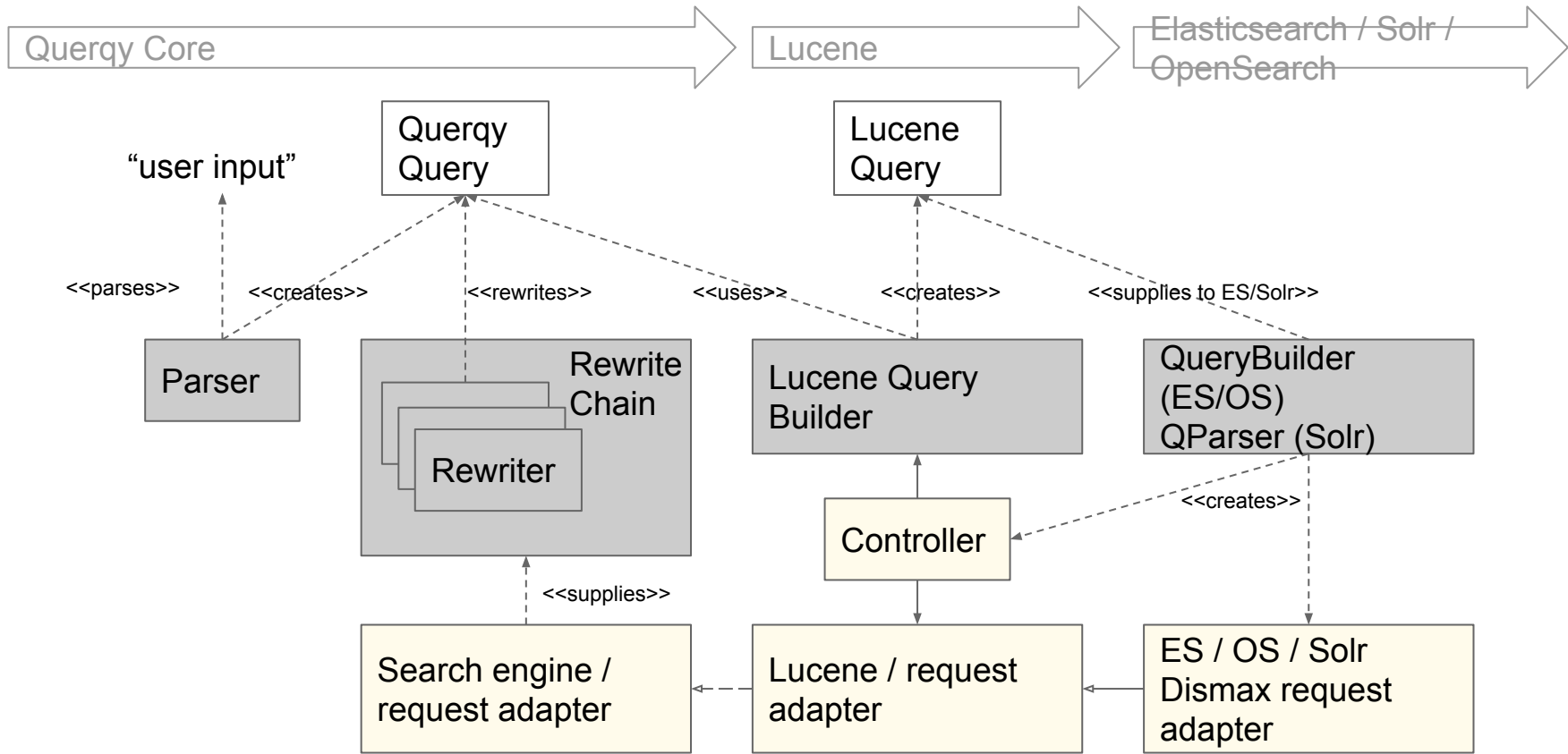


References

- Available at:
 - Solr - <https://github.com/querqy/chorus>.
 - Elastic - <https://github.com/querqy/chorus-elasticsearch-edition>
- Query Types <https://baymard.com/blog/ecommerce-search-query-types>
- Evaluation Tool <https://quepid.com/>
- Model Courtesy <https://huggingface.co/models>

Thank you !!





Women of Search

The image features a white background on the left side where the text is located. On the right side, there is a decorative graphic consisting of several overlapping, semi-transparent shapes in shades of teal, light blue, and lavender. A thin, dark grey dashed line meanders across these shapes, starting from the bottom left and moving towards the top right.

About Women of Search

- Group started in Mar 2021
- A vibrant community dedicated to empowering & celebrating women in the world of search technologies (& related tech)
- Provide a platform for networking, mentorship, and knowledge-sharing among women professionals.
- Showcase, celebrate, brag about the achievements of women
- Practice public speaking and groom your soft skills :)

Group Updates





Updates

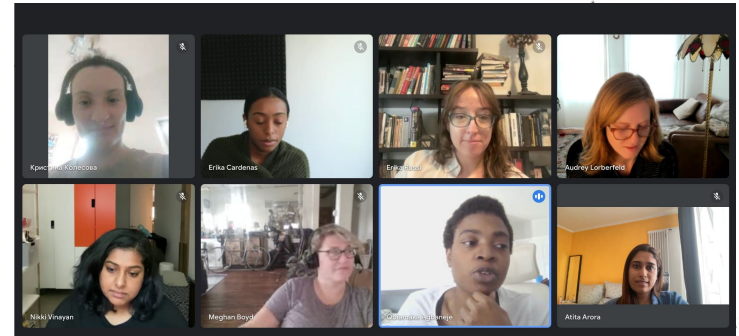
- **'Happy Hour'** continues ...
every **1st Wednesday of each month at 9 am PST.**
- Lots of interesting topics and discussions
- Share useful resources and experiences

Slack : [#relevancy.slack.com / #women-of-search](https://relevancy.slack.com)

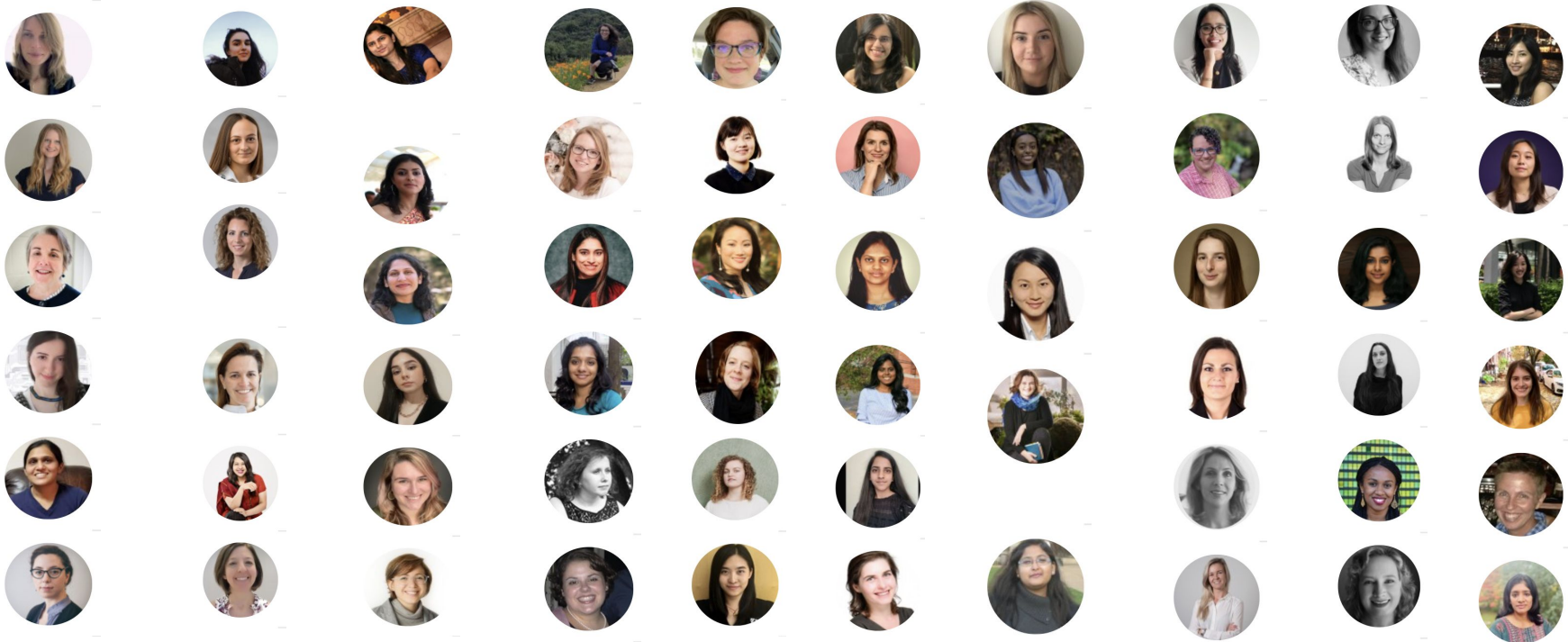
LinkedIn : <https://www.linkedin.com/groups/12659464/>

Calendar: <https://bit.ly/2SywIDh>

Google group: <https://groups.google.com/g/women-of-search>



Thank you !!!



In the realm of generative AI, women's voices echo softly,
but their impact resounds loudly!