

# Vectorize your E-commerce Search

[opensourceconnections.com](https://opensourceconnections.com)

---



*Atita Arora*

# About me



- Working in domain of Search since 2008
- Open source enthusiast and active contributor on various project Quepid , Chorus , OpenNLP , Solr , Opensearchserver etc
- Primary interest is 'Search Relevance' in Ecommerce Search
- Polyglot Developer who loves everything about Language analysis , Information retrieval et al
- Master in Computer Applications + Master in Strategic Business Management
- Mother of 2 boys :)
- Loves to travel and cook !!

# OpenSource Connections

- “Empowering The World’s Search Teams” by:
  - Meetups & conferences, Relevance Slack - join 2000 others at [www.o19s.com/slack](http://www.o19s.com/slack)
  - Books, blogs and reports
  - Training - “Think Like a Relevance Engineer”, LTR, NLP
  - Consulting - Elasticsearch, Solr, Relevance Engineering, Learning to Rank, NLP, Building Search Teams, E-commerce Search...
- Offices in USA, UK and Germany



[www.o19s.com/books](http://www.o19s.com/books)

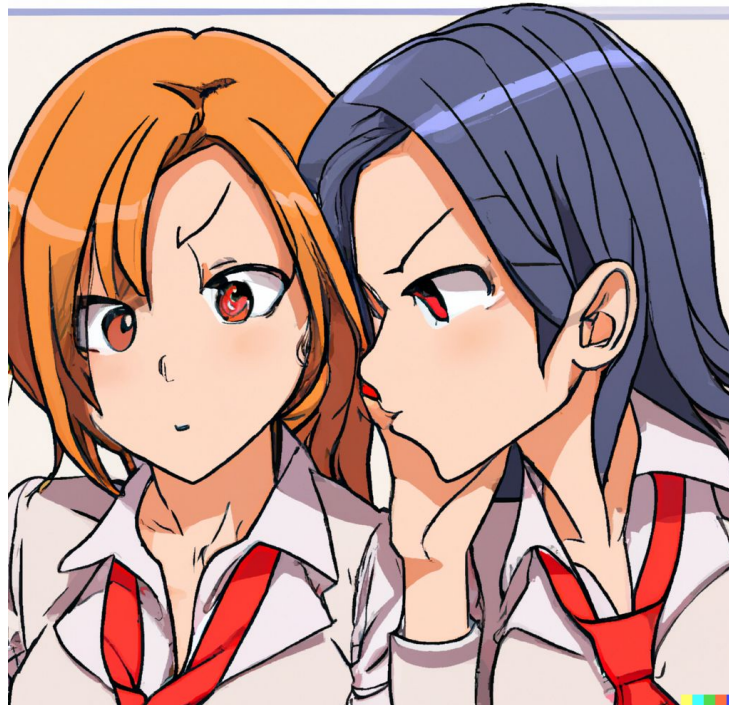


[www.haystackconf.com](http://www.haystackconf.com)

The Search Relevance Conference  
(Haystack LIVE! Meetups every few weeks)

# What the heck is Vector Search?

---



# What the heck is Vector Search?

- Works by converting text into a vector (numerical) embeddings using a desired transformer (like bert , minilm etc) then index and search through those vector embeddings.
- The idea of encoding is to capture the semantic similarity of documents in the given space.
- Traditional search revolves around keywords, while vector search leverages semantic meaning, to find documents most similar to the query (encoding).



# How can Vectors help me in Ecommerce Search?

---

- Very active area of research interest
- Fixing recall issues caused by:
  - Handling long-tail queries
  - Handling zero hits
  - Handling multilingual search
- Exposes a lot of potential in terms of :
  - Detecting similar users (marketing and promotions)
  - Detecting similar products (recommendations and suggestions)

# Benefits

---

- Leverage the power of deep learning / machine learning in your search engine
- Leverage other means (images /audio / videos) to deliver better results
- Boosting precision by leveraging semantic meaning

# Exciting .. How to get started?

[Overview](#)[Documentation](#)[Examples](#)[Playground](#)

## Get started



Enter an instruction or select a preset, and watch the API respond with a **completion** that attempts to match the context or pattern you provided.

You can control which **model** completes your request by changing the model.

### KEEP IN MIND

- 🚩 Use good judgment when sharing outputs, and attribute them to your name or company. [Learn more.](#)
- 📄 Requests submitted to our models may be used to train and improve future models. [Learn more.](#)
- 📅 Our default models' training data cuts off in 2021, so they may not have knowledge of current events.

## Playground



Vector search is great, how do I implement it on my E-commerce Search engine?

There are a few steps involved in implementing vector search on an e-commerce search engine.

1. Gather the data: Depending on the type of e-commerce search engine, you will need to collect data from the e-commerce store, product databases, and customers. This data will be used to create a vector space model.
2. Create the vector space model: Using the data collected, create a vector space model. This will involve creating vectors for each item and assigning a numerical value to each vector based on how closely related it is to other items.
3. Implement the search algorithm: Once the vector space model has been created, you can implement the search algorithm. This will involve querying the vector space model to find the most relevant results for the search query.
4. Test and optimize: Finally, you should test and optimize the vector search engine to ensure that it is providing the most accurate and relevant results. This will involve testing different queries and adjusting the vector space model accordingly.

Good luck!

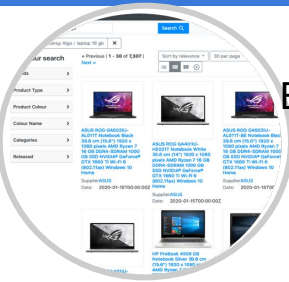
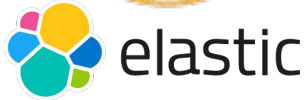


# The Devil is in the details ;)

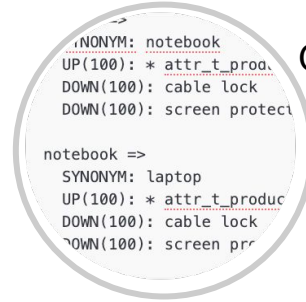
---



# Chorus for Ecommerce Search is....



Blacklight



Queryer



SMUI

v1.0	v1.1	v1.2	Δ	v1.0
0.5230	0.4759	0.6459	0.0471	0.7000
v1.0	v1.1	v1.2	Δ	v1.0
0.1546	0.1540	0.1798	-0.0006	0.0268
v1.0	v1.1	v1.2	Δ	v1.0
0.1546	0.1540	0.1798	-0.0006	0.0268
v1.0	v1.1	v1.2	Δ	v1.0
0.1546	0.1540	0.1798	-0.0006	0.0268
v1.1	v1.2	Δ		

RRE



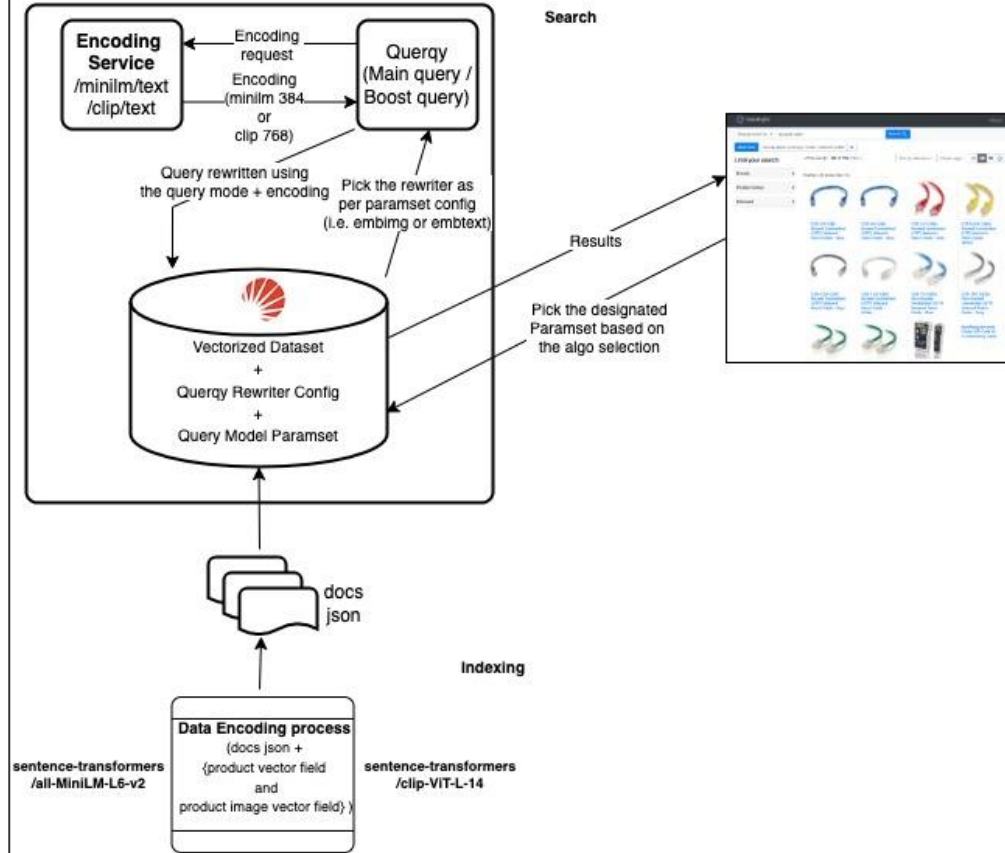
Quepid

# Chorus for Ecommerce Search

---

- Chorus is a set of tools that helps you to get started or improve your e-commerce search.
- Consists of :
  - A fully qualified search engine (Solr / Elastic)
  - Querqy and SMUI - Search Management
  - Quepid for relevance ratings & judgments
  - Metrics / Dashboards for monitoring
- Available at <https://github.com/querqy/chorus>.

# Chorus with Vectors



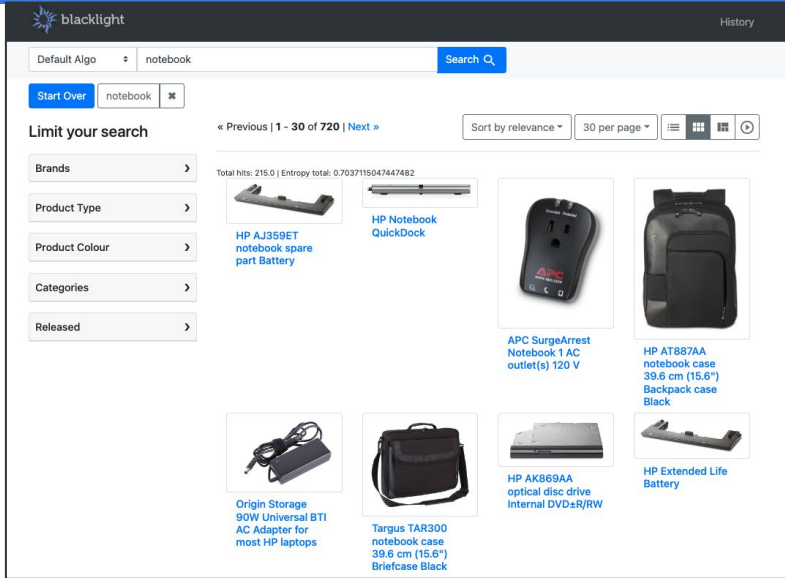
- Embedding service using FastApi
- Query rewriting using Querqy for respective models (CLIP , MiniLM and OpenAI)
- Vectorized ICECAT dataset using CLIP (image field) and MiniLM (text field - title + brand)
- Query models support
  - Match by Text Vector Search
  - Match by Image Vector Search
  - Boost by Text vectors
  - Boost by Image vectors



# Chorus Demo with Vectors

# Query : Notebook

<- Results with default algo



blacklight History

Default Algo notebook Search Q

Start Over notebook

Limit your search

Brands >

Product Type >

Product Colour >

Categories >

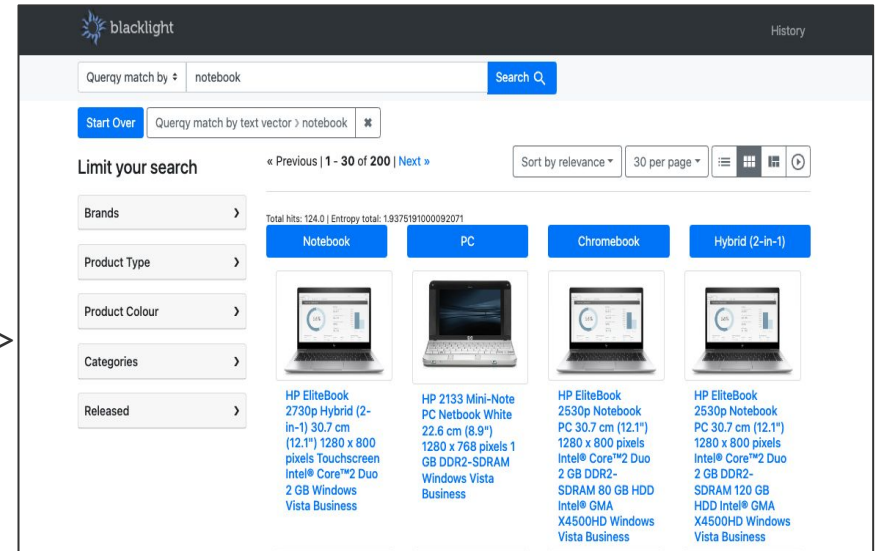
Released >

« Previous | 1 - 30 of 720 | Next » Sort by relevance 30 per page

Total hits: 215.0 | Entropy total: 0.703715047447482

- HP AJ359ET notebook spare part Battery
- HP Notebook QuickDock
- APC SurgeArrest Notebook 1 AC outlet(s) 120 V
- HP AT887AA notebook case 39.6 cm (15.6") Backpack case Black
- Origin Storage 90W Universal BTI AC Adapter for most HP laptops
- Targus TAR300 notebook case 39.6 cm (15.6") Briefcase Black
- HP AK869AA optical disc drive Internal DVD±R/RW
- HP Extended Life Battery

Results with match by text vector ->



blacklight History

Query match by notebook Search Q

Start Over Query match by text vector notebook

Limit your search

Brands >

Product Type >

Product Colour >

Categories >

Released >

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

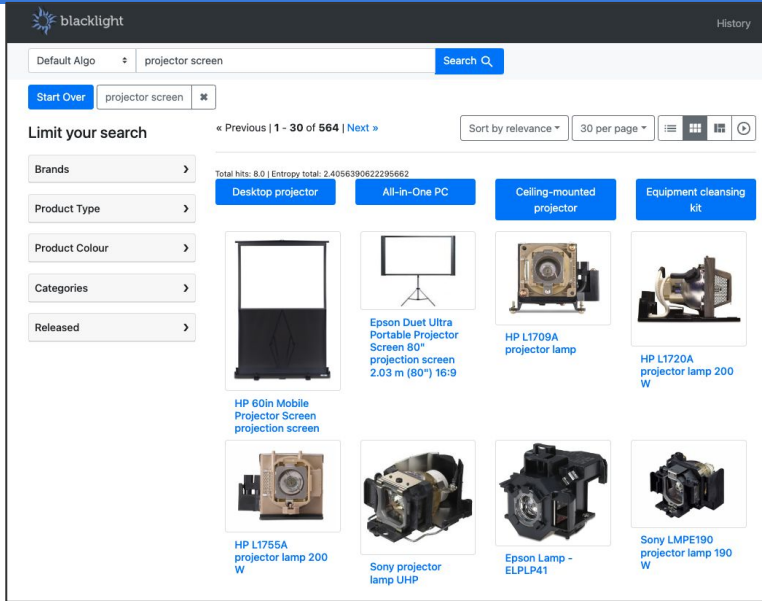
Total hits: 124.0 | Entropy total: 1.937519100092071

Notebook PC Chromebook Hybrid (2-in-1)

- HP EliteBook 2730p Hybrid (2-in-1) 30.7 cm (12.1") 1280 x 800 pixels Touchscreen Intel® Core™2 Duo 2 GB Windows Vista Business
- HP 2133 Mini-Note PC Netbook White 22.6 cm (8.9") 1280 x 768 pixels 1 GB DDR2-SDRAM Windows Vista Business
- HP EliteBook 2530p Notebook PC 30.7 cm (12.1") 1280 x 800 pixels Intel® Core™2 Duo 2 GB DDR2-SDRAM 80 GB HDD Intel® GMA X4500HD Windows Vista Business
- HP EliteBook 2530p Notebook PC 30.7 cm (12.1") 1280 x 800 pixels Intel® Core™2 Duo 2 GB DDR2-SDRAM 120 GB HDD Intel® GMA X4500HD Windows Vista Business

# Query : Projector Screen

← Results with default algo



blacklight History

Default Algo projector screen Search Q

Start Over projector screen

Limit your search

« Previous | 1 - 30 of 564 | Next » Sort by relevance 30 per page

Total hits: 8.0 | Entropy total: 2.4056390622295662

- Desktop projector
- All-in-One PC
- Ceiling-mounted projector
- Equipment cleansing kit

HP 60in Mobile Projector Screen

Epson Duet Ultra Portable Projector Screen 80" projection screen 2.03 m (80") 16:9

HP L1709A projector lamp

HP L1720A projector lamp 200 W

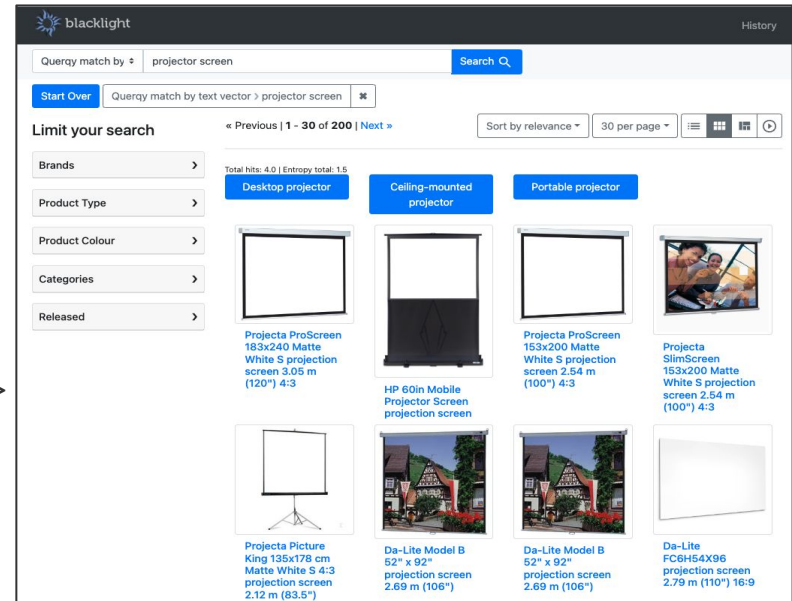
HP L1755A projector lamp 200 W

Sony projector lamp UHP

Epson Lamp - ELPLP41

Sony LMPE190 projector lamp 190 W

Results with match by text vector ->



blacklight History

Query match by projector screen Search Q

Start Over Query match by text vector > projector screen

Limit your search

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

Total hits: 4.0 | Entropy total: 1.5

- Desktop projector
- Ceiling-mounted projector
- Portable projector

Projecta ProScreen 153x200 Matte White S projection screen 3.05 m (120") 4:3

HP 60in Mobile Projector Screen

Projecta ProScreen 153x200 Matte White S projection screen 2.54 m (100") 4:3

Projecta SlimScreen 153x200 Matte White S projection screen 2.54 m (100") 4:3

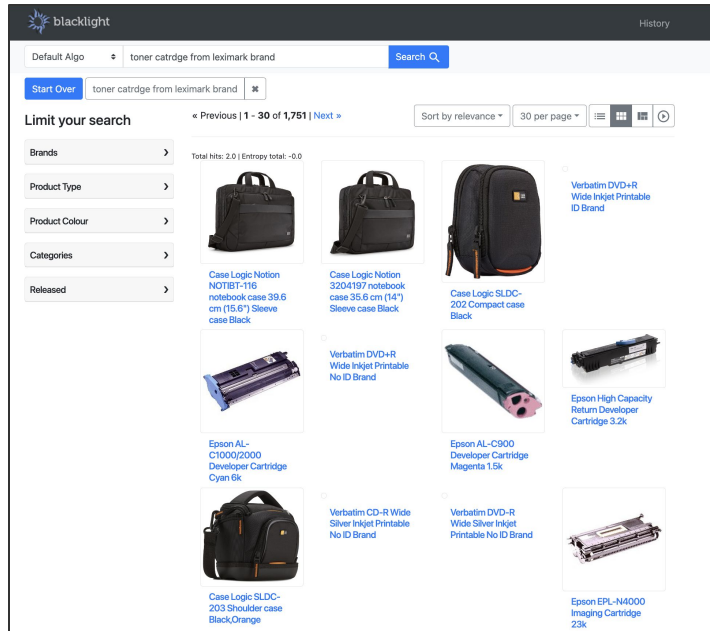
Projecta Picture King 135x178 cm Matte White S 4:3 projection screen 2.12 m (83.5")

Da-Lite Model B 52" x 92" projection screen 2.69 m (106")

Da-Lite Model B 52" x 92" projection screen 2.69 m (106")

Da-Lite FC6H54X96 projection screen 2.79 m (110") 16:9





blacklight History

Default Algo toner cartridge from leximark brand Search Q

Start Over toner cartridge from leximark brand

Limit your search < Previous | 1 - 30 of 1,751 | Next > Sort by relevance 30 per page

Brands > Total hits: 2.0 | Entropy total: -0.0

Product Type >

Product Colour >

Categories >

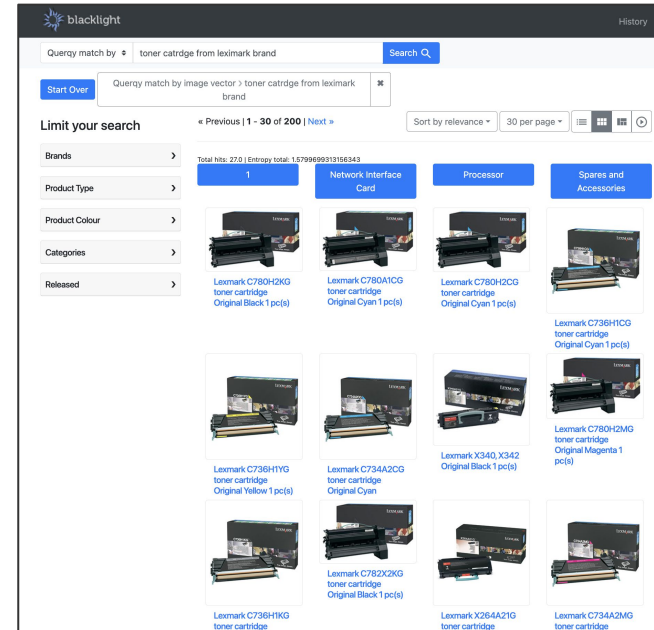
Released >

- Verbatim DVD+R Wide Inkjet Printable ID Brand
- Case Logic Notion NOTBT-116 notebook case 39.6 cm (15.6") Sleeve case Black
- Case Logic Notion 3204197 notebook case 35.6 cm (14") Sleeve case Black
- Case Logic SLDC-202 Compact case Black
- Verbatim DVD+R Wide Inkjet Printable No ID Brand
- Epson High Capacity Return Developer Cartridge 3.2k
- Epson AL-C900 Developer Cartridge Magenta 1.5k
- Epson AL-C1000/2000 Developer Cartridge Cyan 0k
- Verbatim CD-R Wide Silver Inkjet Printable No ID Brand
- Verbatim DVD-R Wide Silver Inkjet Printable No ID Brand
- Epson EPL-N4000 Imaging Cartridge 23k
- Case Logic SLDC-203 Shoulder case Black/Orange

# Query : toner catrdrge from leximark brand (with 2 misspells)

<- Results with default algo

Results with match by image vector ->



blacklight History

Query match by toner cartridge from leximark brand Search Q

Start Over Query match by image vector > toner cartridge from leximark brand

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

Brands > Total hits: 27.0 | Entropy total: 1.57996993136343

Product Type > 1 Network interface Card Processor Spares and Accessories

Product Colour >

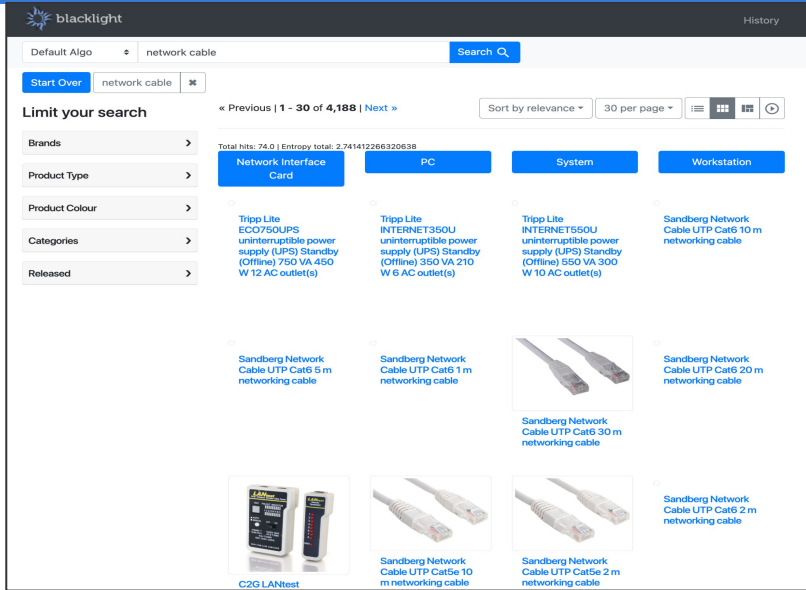
Categories >

Released >

- Lexmark C780H2KG toner cartridge Original Black 1 pc(s)
- Lexmark C780A1CG toner cartridge Original Cyan 1 pc(s)
- Lexmark C780H2CG toner cartridge Original Cyan 1 pc(s)
- Lexmark C736H1CG toner cartridge Original Cyan 1 pc(s)
- Lexmark C780H2MG toner cartridge Original Magenta 1 pc(s)
- Lexmark C736H1YG toner cartridge Original Yellow 1 pc(s)
- Lexmark C734A2CG toner cartridge Original Cyan
- Lexmark X340, X342 Original Black 1 pc(s)
- Lexmark C782X2KG toner cartridge Original Black 1 pc(s)
- Lexmark C736H1KG toner cartridge
- Lexmark X264A21G toner cartridge
- Lexmark C734A2MG toner cartridge

# Query : Network Cable

<- Results with default algo



blight network cable

Search

Start Over network cable

Limit your search

« Previous | 1 - 30 of 4,188 | Next »

Sort by relevance 30 per page

Brands

Product Type

Product Colour

Categories

Released

Total hits: 74.0 | Entropy total: 2.741412266320638

Network Interface Card PC System Workstation

Tripp Lite ECO750UPS uninterruptible power supply (UPS) Standby (Offline) 750 VA 450 W 12 AC outlet(s)

Tripp Lite INTERNET350U uninterruptible power supply (UPS) Standby (Offline) 350 VA 210 W 6 AC outlet(s)

Tripp Lite INTERNET550U uninterruptible power supply (UPS) Standby (Offline) 550 VA 300 W 10 AC outlet(s)

Sandberg Network Cable UTP Cat6 5m networking cable

Sandberg Network Cable UTP Cat6 1m networking cable

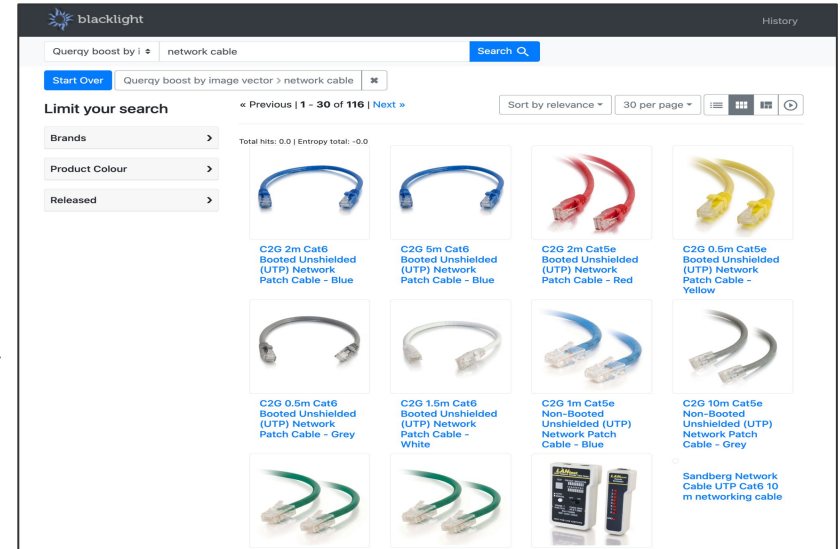
Sandberg Network Cable UTP Cat6 20m networking cable

Sandberg Network Cable UTP Cat6 30m networking cable

Sandberg Network Cable UTP Cat6 2m networking cable

C2G LANtest

Results with boost by image vector ->



blight network cable

Search

Start Over Query boost by image vector > network cable

Limit your search

« Previous | 1 - 30 of 116 | Next »

Sort by relevance 30 per page

Brands

Product Colour

Released

Total hits: 0.0 | Entropy total: -0.0

C2G 2m Cat6 Booted Unshielded (UTP) Network Patch Cable - Blue

C2G 5m Cat6 Booted Unshielded (UTP) Network Patch Cable - Blue

C2G 2m Cat5e Booted Unshielded (UTP) Network Patch Cable - Red

C2G 0.5m Cat5e Booted Unshielded (UTP) Network Patch Cable - Yellow

C2G 0.5m Cat6 Booted Unshielded (UTP) Network Patch Cable - Grey

C2G 1.5m Cat6 Booted Unshielded (UTP) Network Patch Cable - White

C2G 1m Cat5e Non-Booted Unshielded (UTP) Network Patch Cable - Blue

C2G 10m Cat5e Non-Booted Unshielded (UTP) Network Patch Cable - Grey

Sandberg Network Cable UTP Cat6 5m networking cable



# Caveats

---

- Vectors suitable for the business case (wrt model)
- Dimensions are not free
- Limitations on length of text that can be encoded
- Infrastructural changes
- Fast and scalable vector search service

# Further improvements

---

- Capabilities to support custom trained models
- Image vector search results quality
- Add fine-tune capability for Hybrid Search (BM25 + Vectors)
- Combining Vectors with manually manipulated search results
- Add vector based Re-ranking model

# References

---

- Image courtesy : <https://openai.com/dall-e-2/>
- Diagram generated : <https://app.diagrams.net/>

A blue-tinted photograph of an office environment. Several people are seated at desks with computers, and one man is standing and looking at a monitor. The scene is bright and professional.

# Thank you!