

**From text search and recommendation  
to ads and online dating;  
approximate nearest neighbors in real  
world applications**

Lester Solbakken, Verizon Media

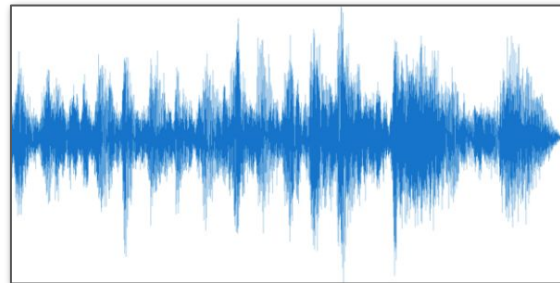
# Vespa.ai

A open-source platform for low latency computations over large, evolving data

- Search and filter over structured and unstructured data
- Query time organization and aggregation of matching data
- Real-time writes
- Elastic and fault tolerant
- Scalable and fast
- Pluggable
- Advanced relevance scoring with tensors as first class citizens
- Built-in support for machine learning models (TensorFlow, ONNX, XGBoost, LightGBM ++)
- **Scalable approximate nearest neighbors**

# (Dense) representations

Vespa is a full-featured text search engine and supports both regular text search and fast approximate vector search (ANN). This makes it easy to create high-performing search applications at any scale, whether you want to use traditional techniques, or a modern vector based approach. You can even combine both approaches efficiently in the same query, something no other engine can do.



[0.27, 0.79, 0.98, ...]

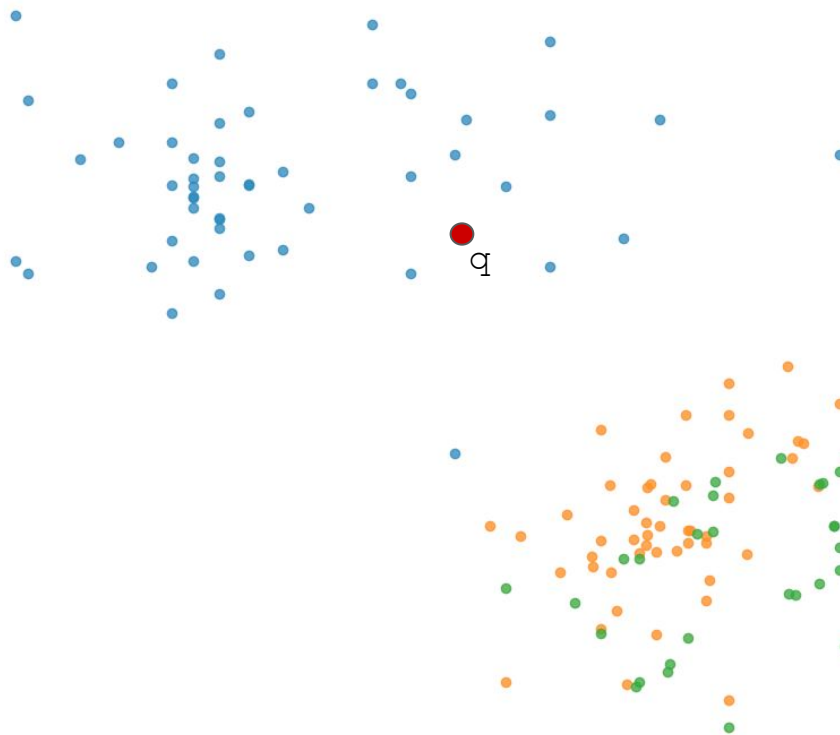


[0.11, 0.14, 0.38, ...]



[0.89, 0.14, 0.52, ...]

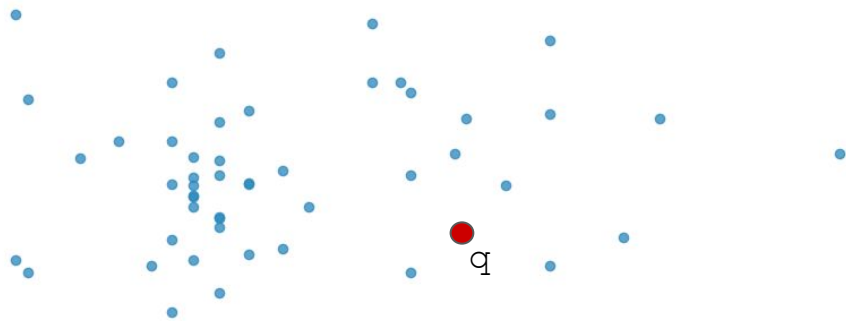
# Nearest Neighbor Search



## Distance functions:

- Euclidean
- Cosine angle
- Manhattan
- Hamming
- Inner (dot) product

# Nearest Neighbor Search



## Distance functions:

- Euclidean
- Cosine angle
- Manhattan
- Hamming
- Inner (dot) product

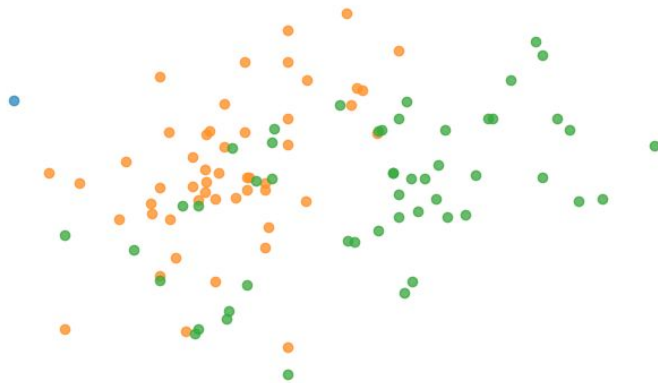
## Approximate nearest neighbor

### Inverse index

- K-means
- Product quantization (and relatives)
- Locality-sensitive hashing

### Non-inverse index

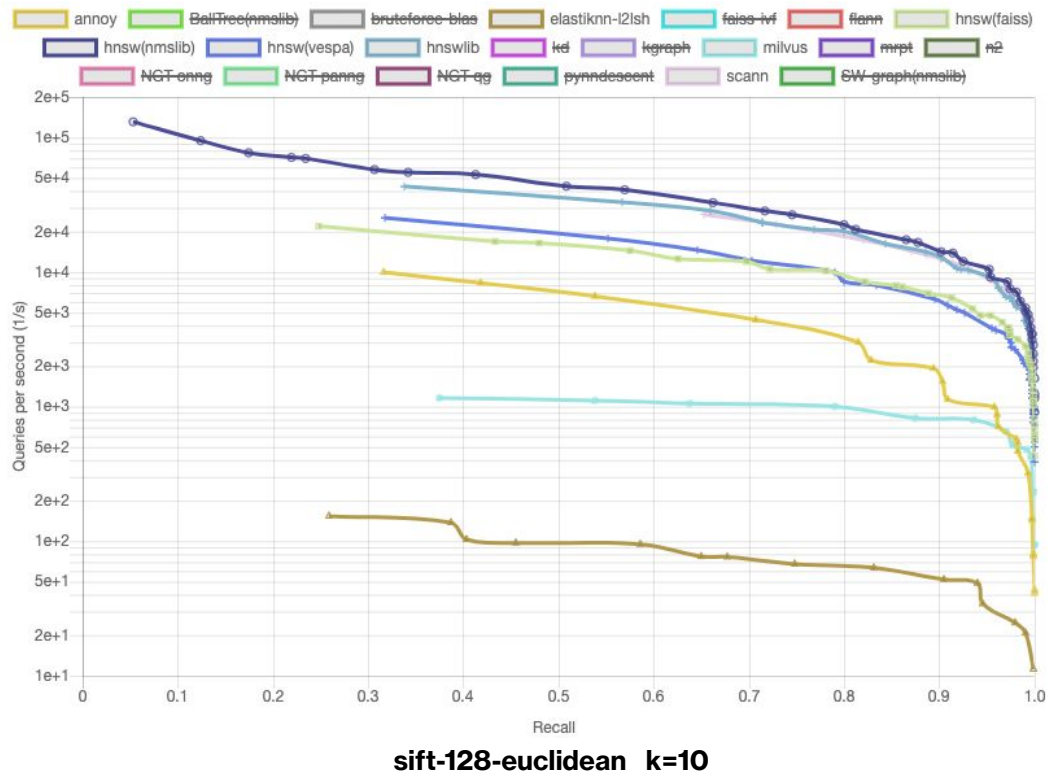
- Hierarchical Navigable Small World (HNSW)



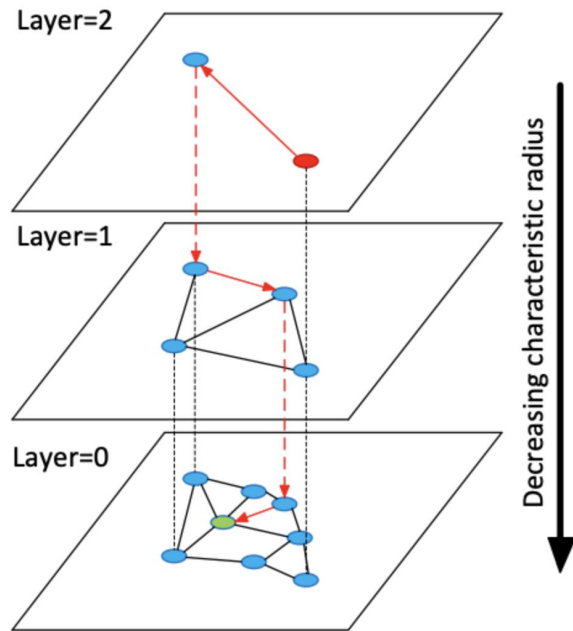
# ANN-benchmarks.com

Notable ANN libraries:

- ScaNN
- FAISS
- Annoy
- Vespa (not a library)



# Hierarchical Navigable Small World (HNSW)



- Based on graph structures
- Very efficient
- **Allows incremental modification**



Important for many search applications

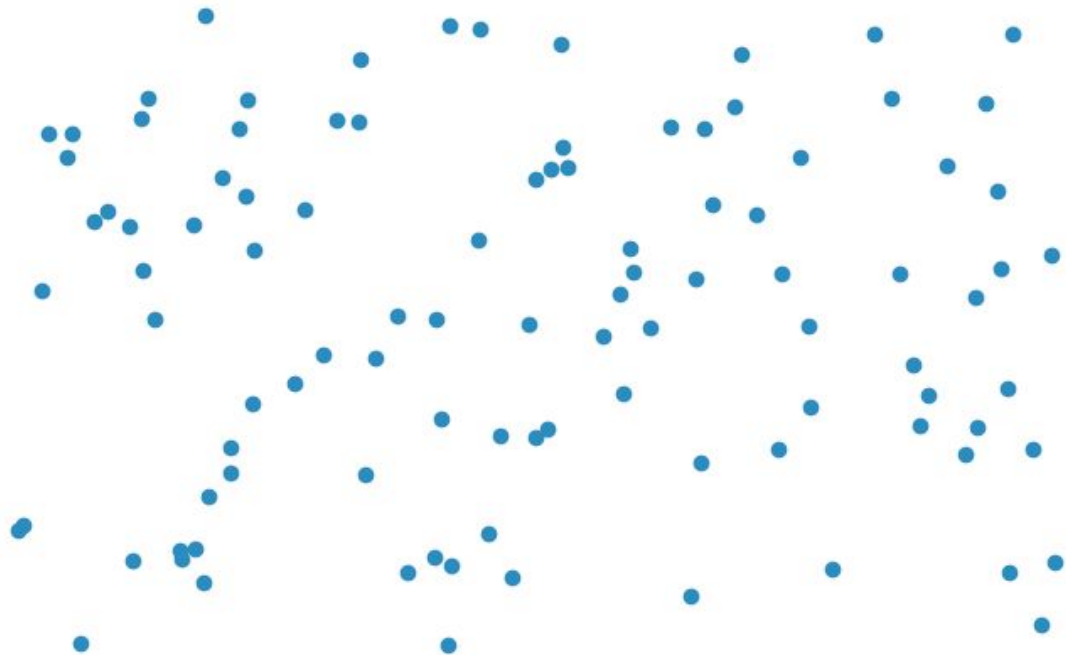
*Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs, Yu. A. Malkov and D. A. Yashunin*

# Metadata filtering

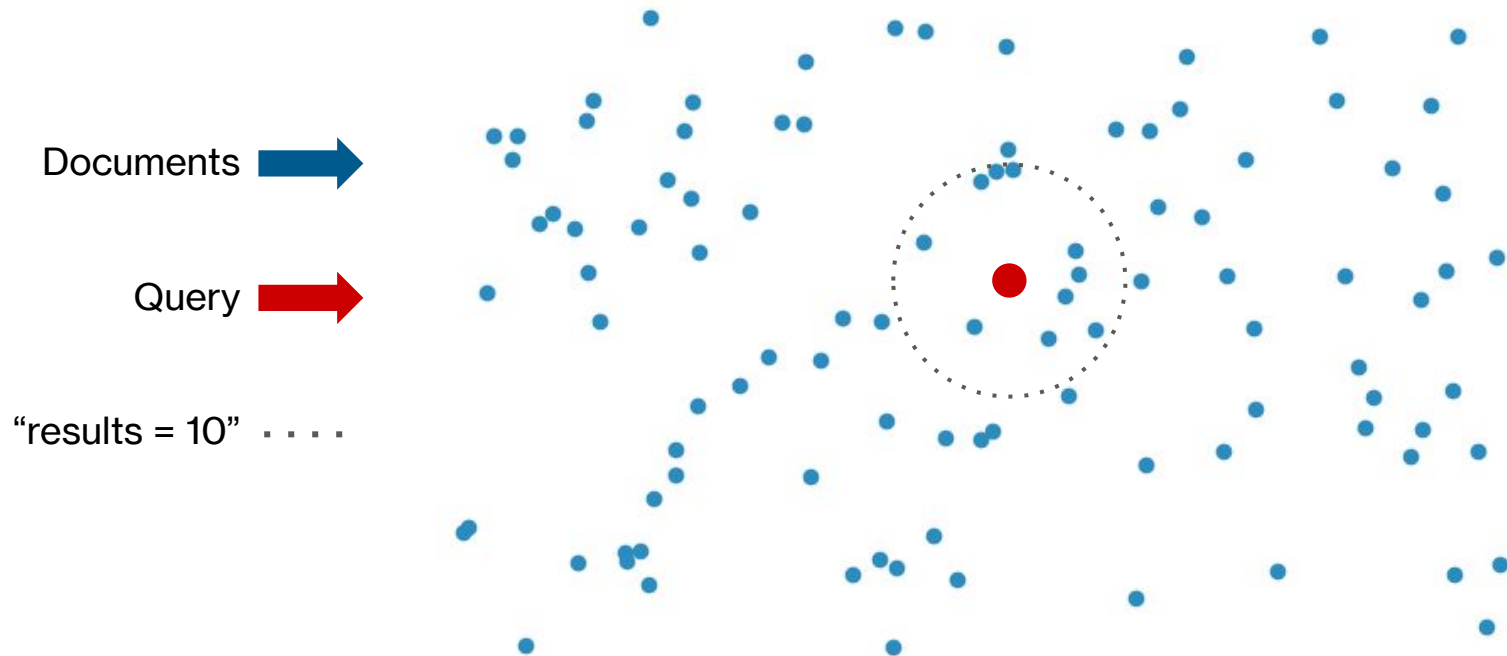


# Semantic search for text

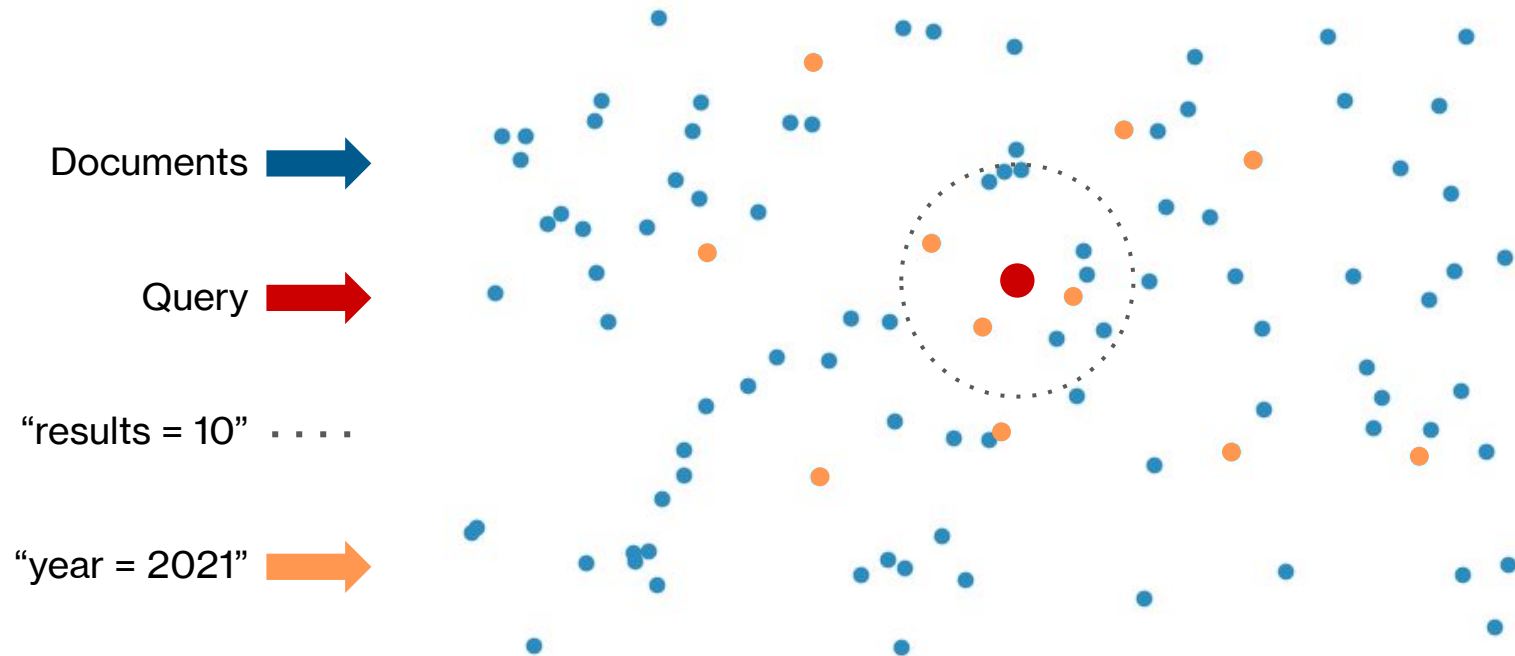
Documents



# Semantic search for text

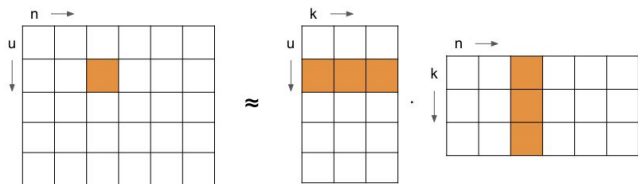


# Semantic search for text

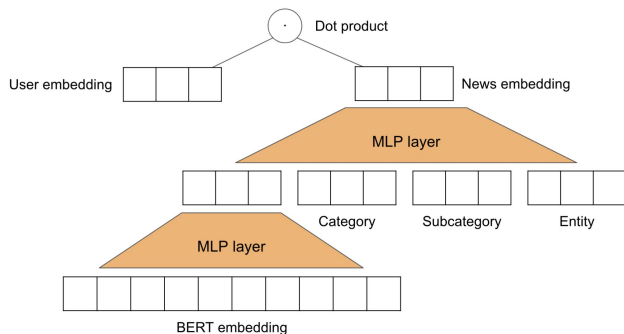


# Recommendation

## Matrix factorization:

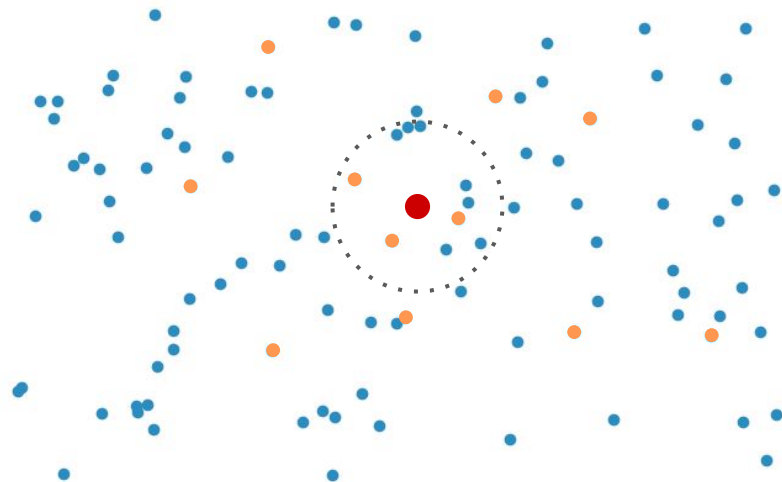


## Neural collaborative filtering:



<https://docs.vespa.ai/en/tutorials/news-4-embeddings.html>

Convert MIPS  
to euclidean



**Metadata filters:** appropriateness for age, banned, region availability, language, ++

**Business rule filters:** diversity, deduplication, already seen, ++

# Ad search



“Relevance”



[0.58, 0.59, 0.13, ...]



[0.51, 0.61, 0.23, ...]



Metadata:

- Target age group
- Target location
- **Budget**
- +++

**Can't exceed  
budget of ad or  
campaign!**

**Experts: Djokovic's gift to boy could fetch big money**  
The generous gift that newly crowned men's champion Novak Djokovic gave to a young fan could fetch between \$25,000 and \$35,000 today, sports memorabilia experts say.  
'This kid's reaction ... is everything' »

**What \$200,000 homes look like in 3 different states**

**Why cars, gas, airfare, meat are getting more expensive**

**After collapsing on field, Danish midfielder 'was gone'**

**Whatever it's called, snow shouldn't be this red**

**Nets lose Game 4 to Bucks and perhaps more**

**U.S. Fox News**  
**Bay Area couple accused of running multi-million dollar 'brothel' from apartment: report**  
A Bay Area couple is accused of trafficking women from overseas and operating a "brothel" in Milpitas located in a complex where units can fetch \$5,000-a-month for rent, according to reports.

**Ad Monivator**  
**Can Elon's Next Breakthrough Make You Rich?**  
One of Wall Street's most famous investors says this radical new technology will change nearly EVERYTHING about our daily lives.

**Style Yahoo Life**  
**Elizabeth Hurley wows in sheer gown days after turning 56: 'Goddess'**  
The English star was a vision in yellow Versace.

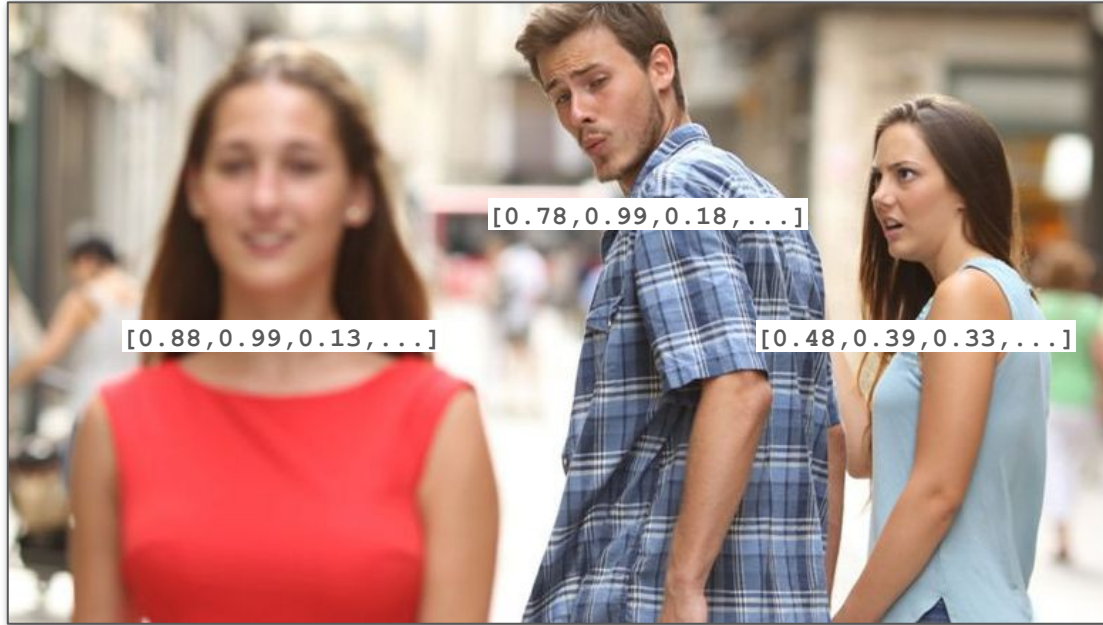
**U.S. Reuters**  
**Judge tosses Houston hospital workers' lawsuit over vaccine requirement**  
A U.S. federal judge has dismissed a lawsuit brought by 117 workers at a Texas hospital over its requirement that they be vaccinated against COVID-19. In a ruling issued on Saturday, U.S. District Judge Lynn Hughes upheld Houston Methodist Hospital's policy mandating employee...

**Business MarketWatch**  
**Musk says Tesla sold about 10% of bitcoin to test market, and will 'resume allowing crypto transactions' when 50% of miners use clean...**

**Reuters**  
**Musk says Tesla will accept bitcoins when miners use more clean energy**

**Reuters**  
**Bitcoin jumps after Musk says Tesla could use it again**

# Online dating



## Filters:


- Location
- Gender
- Age range
- Height
- Already seen
- ++

# Shopping

“sup” → [0.08, 0.99, 0.43, ...]

## Metadata:

- Price
- Category
- Brand
- Rating
- In stock?



[ABOUT](#) [DOCUMENTATION](#) [VESPA.AI](#)





### Categories

- Sports & Outdoors (19)
- Boating & Water Sports (18)
- Standup Paddleboarding (8)
- Surfing (4)
- Standup Paddleboards (3)
- Kayaking (2)
- Paddles (2)
- Surfing Fins (2)
- Diving & Snorkeling (2)
- Diving Suits (2)
- Wetsuits (2)
- Standup Paddles (2)
- Canoeing (1)
- Paddles (1)
- Leashes (1)
- Team Sports (1)
- Football (1)
- Footballs (1)
- Boating (1)
- Safety & Flotation Devices (1)

### Brands

- Pelican (1)
- Futures Fins (1)

### Rating

	5 (6)
	4 (5)
	1 (2)
	0 (6)


### Price range


\$8 to \$8 (1)
\$18 to \$23 (5)
\$30 to \$30 (1)
\$50 to \$92 (3)
\$100 to \$165 (4)
\$220 to \$450 (4)
\$944 to \$944 (1)


## Results for 'sup'


Results per page: 10 | 25 | 100      Order by: relevance | lowest price | highest price | lowest rating | highest rating


**Found 19 results**





Pelican Sup Flow 106 Board, Fade Red/Yellow  
**\$ 449.99**  
 (6)





'STORM' SUP Leash 10' straight by Bullet Proof Surf - with Double Stainless Steel Swivels and Triple Rail Saver  
**\$ 19.99**  
 (1)





Boardworks SUP SHUBU 10'7" Inflatable SUP  
**\$ 944.00**  
 (2)



Inflatable SUP Stand Up Paddle Board and 3PC w/ Paddle 9' 9"  
**\$ 399.00**  
 (4)



Rave Performance 3-Piece Aluminum SUP Paddle  
**\$ 91.66**  
 (1)



air7 Stainless Steel 316 fin screw [ 2 pack] for Longboard Surfboard , Stand Up Paddle (SUP) and airSUP - most US box fin boxes  
**\$ 18.49**



# Local search

“Restaurants near me”

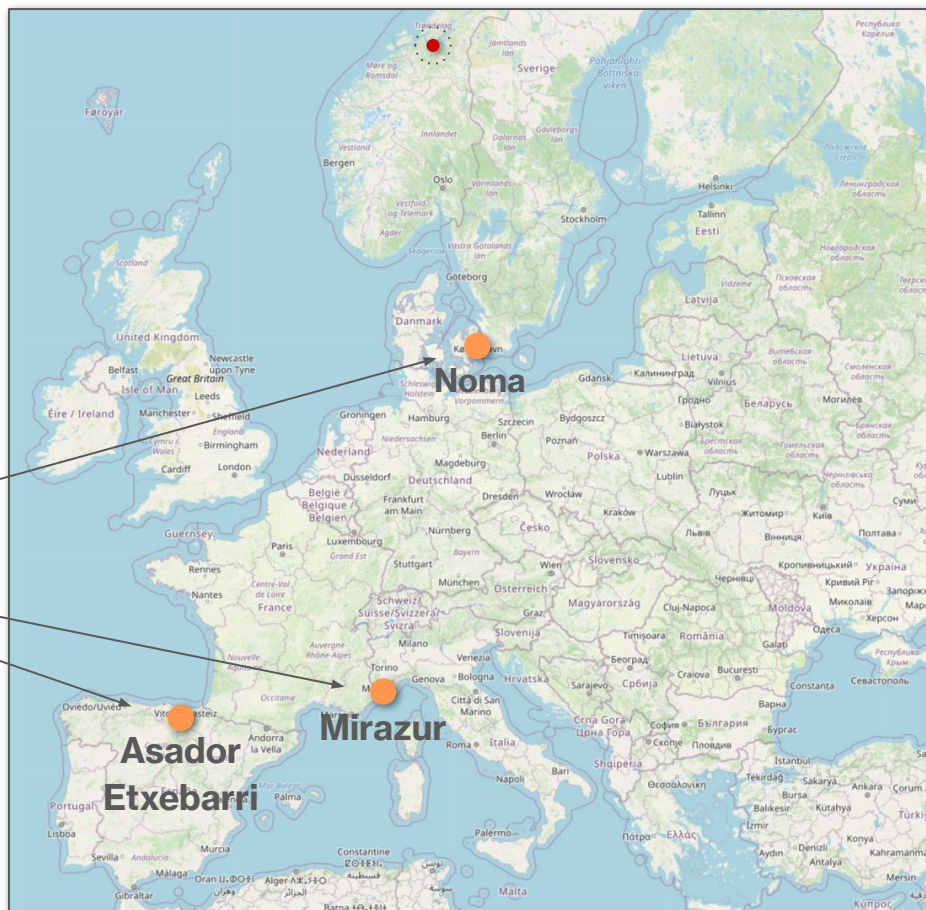


Preference  
profile



[0.58, 0.59, 0.13, ...]

Location filter: 5km ?



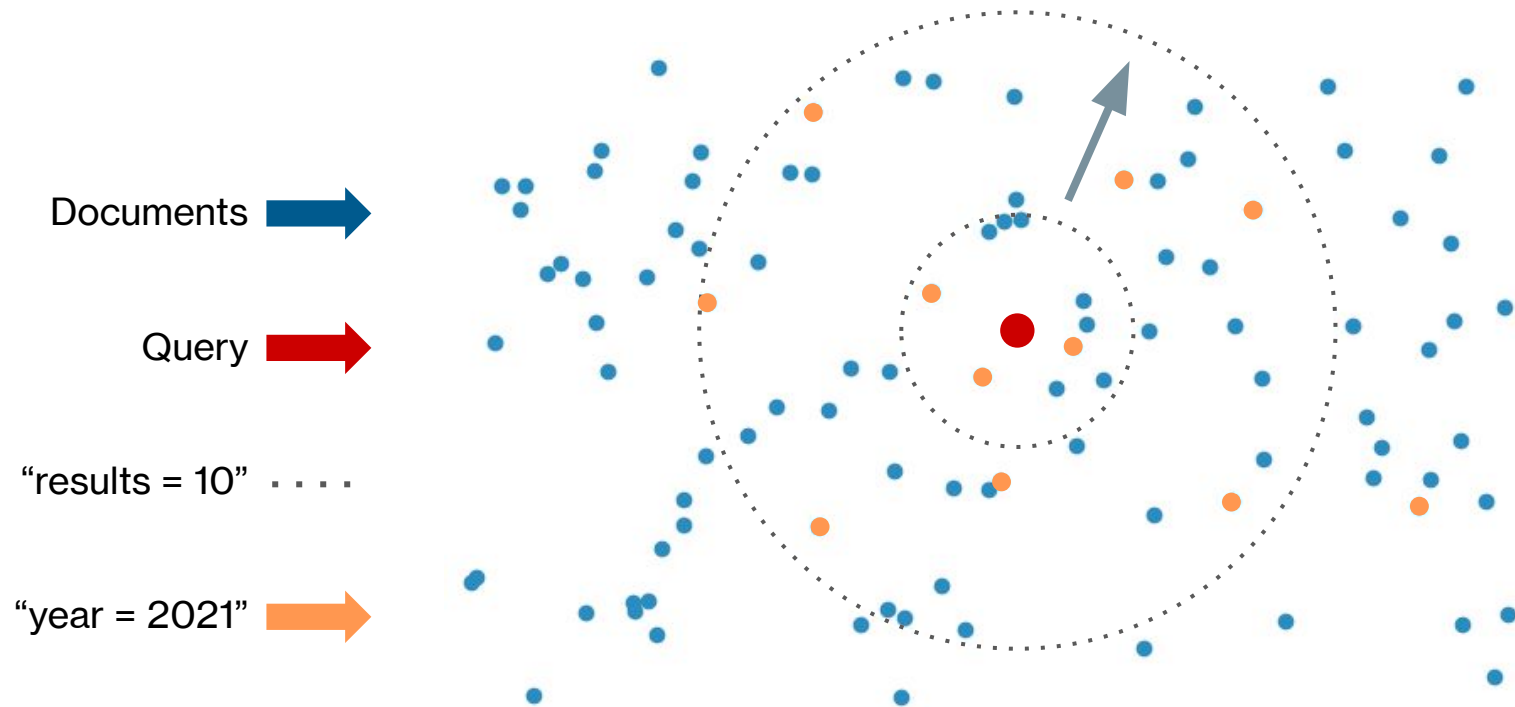
OpenStreetMap, 2021



# Post-filtering does not work

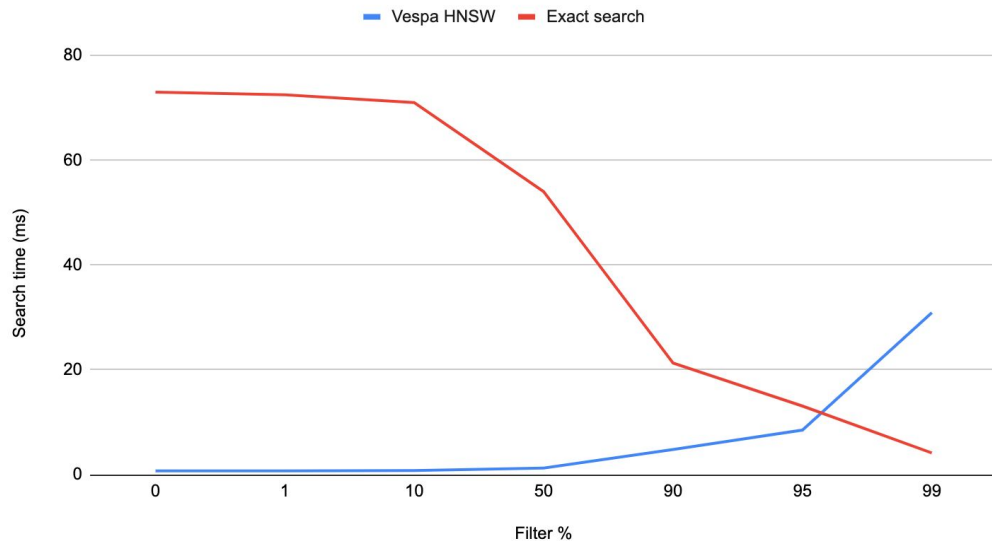
- Recall more
- An index per filter combination
- **Pre-filtering**

# Vespa and pre-filtering



# Pre-filtering vs filter strength

1M SIFT dataset



With **strong filters** (>95%)  
it is more efficient to  
calculate exact nearest  
neighbors **after filtering**.

<https://blog.vespa.ai/approximate-nearest-neighbor-search-in-vespa-part-1/>

# Vespa innovations in approximate nearest neighbors

- **Dynamic modification** of the graph. Vespa supports both adding and removing items in real-time while serving.
- **Multi-threaded indexing** using lock-free data structures and copy-on-write semantics drastically increase the performance of building the index.
- **Metadata filtering** modifies the HNSW algorithm to skip non-eligible candidates.

# Vespa is much more than a vector search engine

## Text search

- Traditional IR: TF/IDF, BM25
- Modern IR: vector similarity, approximate nearest neighbors, Transformers, ++

## Recommendation and personalization

- Fast vector search, filtering, machine learned model evaluation, partial updates

## Question Answering

- Approximate nearest neighbors and answer extraction using BERT models

## Semi-structured navigation

- Text search combined with sorting, grouping, filtering by metadata

## Finding love

- <https://tech.okcupid.com/vespa-vs-elasticsearch/>

# Thank you!

Open source <https://vespa.ai>

Free cloud trial <https://cloud.vespa.ai>