

Result Diversification



Peter Straßer

Senior Solution Architect, Elastic



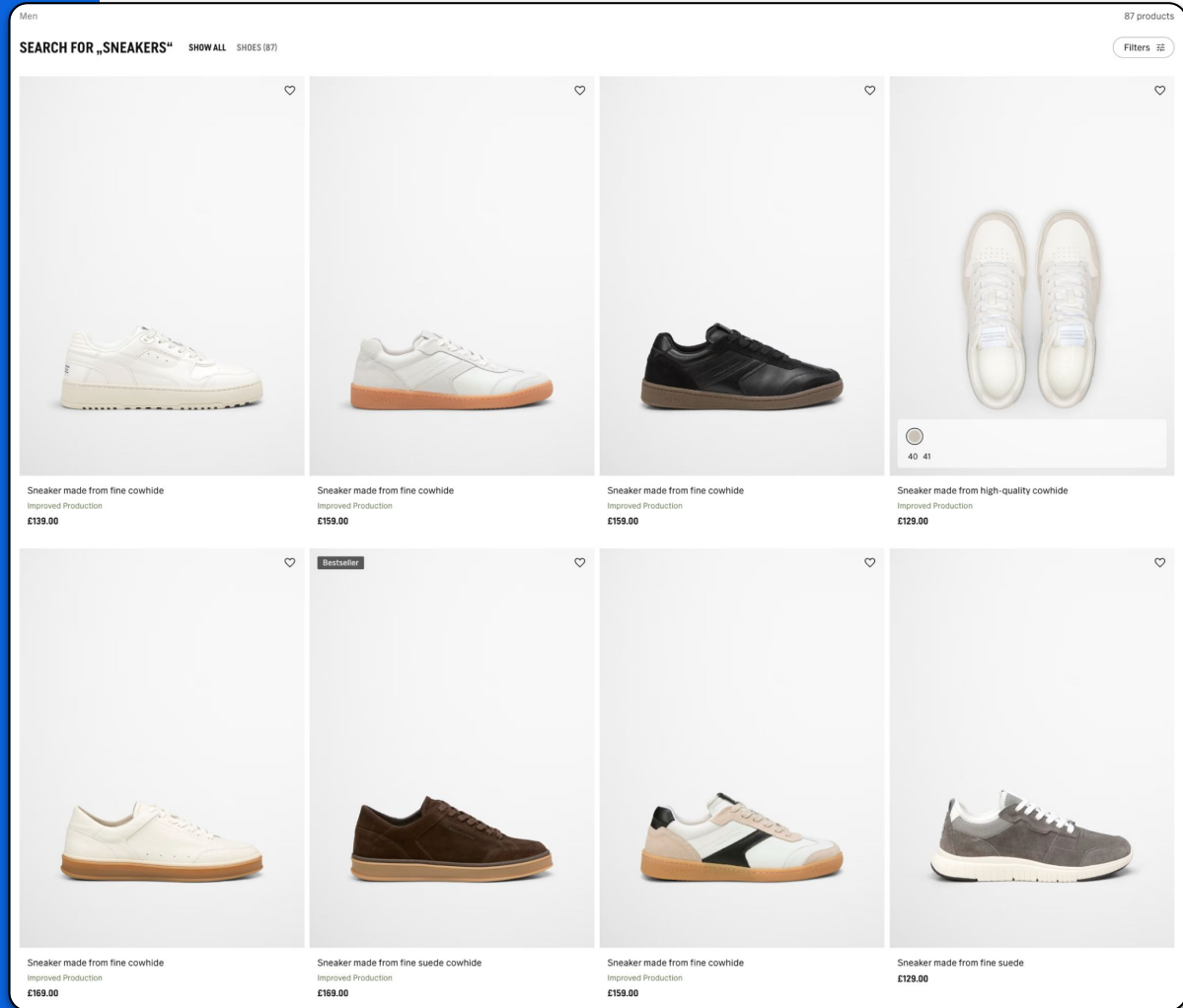
elastic

| The Search
AI Company



The Problem

Exploration is limited













Browsing and discoverability is encouraged

Got a fashion question? Need some good sportswear tips? There's a lot we can chat about.



I am looking for sneakers.

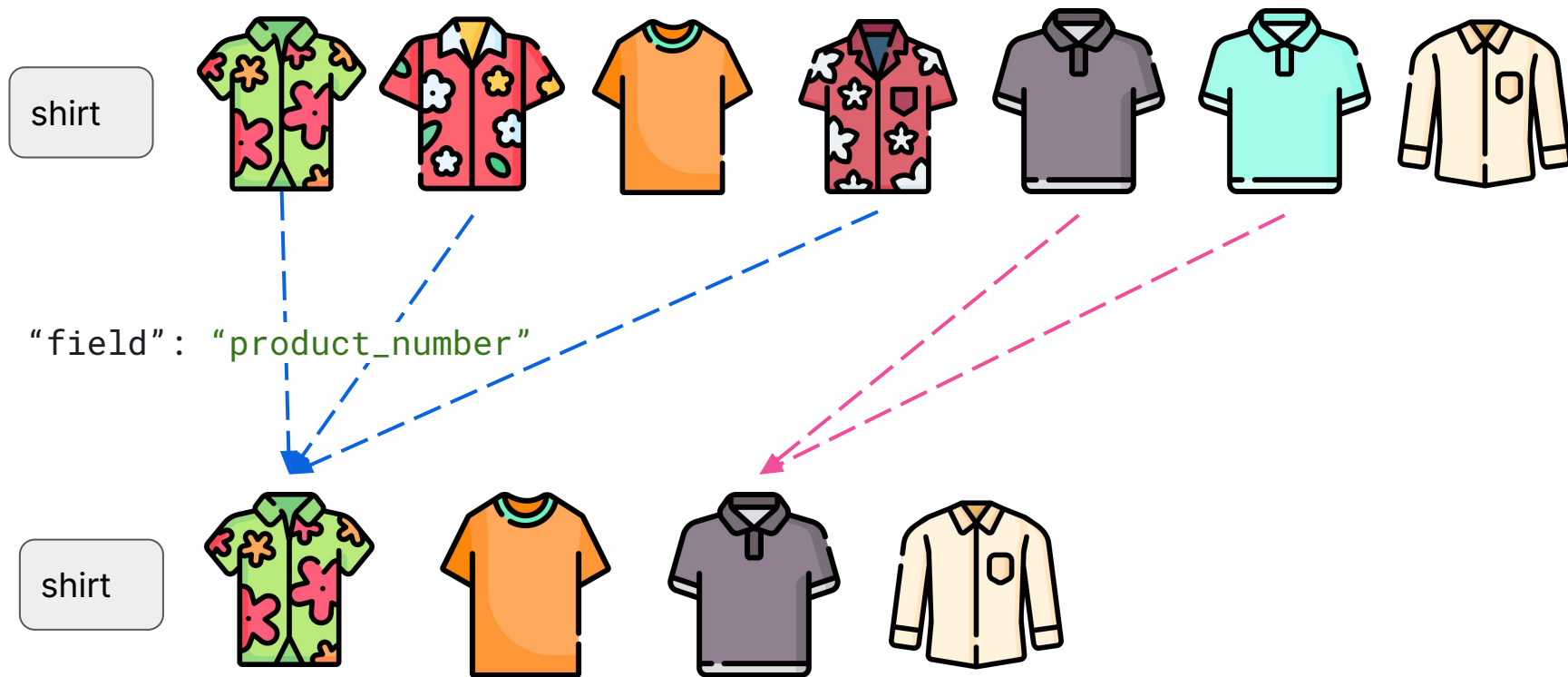
Here are some men's sneakers including trainers and running shoes from brands like Polo Ralph Lauren, adidas Performance, Jack & Jones, Nike Performance, Antony Morato, and Tommy Hilfiger. They come in various colors and styles suitable for leisure and sports. Would you like to see sneakers for specific activities or styles?

 	 	 	 	 
Designer			Designer Exclusive	New
Polo Ralph Lauren MASTERS CRT TOP UNISEX - Trainers - white/navy	adidas Performance GALAXY - Road running shoes - core black	Jack & Jones JFWJORDAN - Trainers - navy/bright white/cognac	Polo Ralph Lauren TRAIN 89 LACE UNISEX - Trainers - black/gold	Nike Performance JOURNEY RUN - Road running shoes - black/anthracite...
€105.00 Regular price: €150.00 -30%	€54.95	€20.99 Regular price: €59.95 -65%	€113.00 Last lowest price: €150.00 -25%	€99.99
Plus Premium delivery	Plus Premium delivery	Plus Premium delivery	Plus Premium delivery	

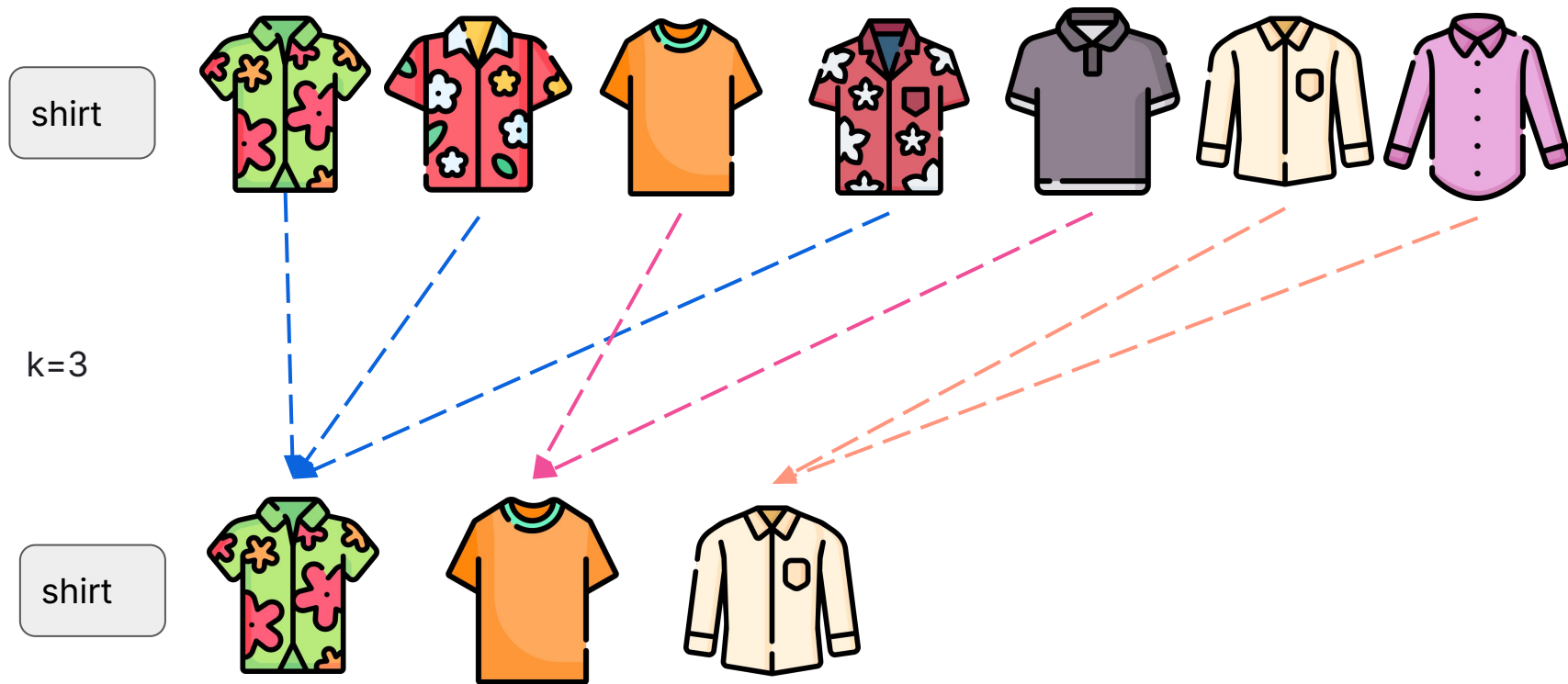
The background is a solid blue color. It features several abstract, 3D-style geometric shapes in a lighter shade of blue. These shapes are composed of multiple concentric, slightly offset lines that create a sense of depth and perspective. They are positioned in the corners and along the edges of the frame, framing the central text.

How can we achieve it?

Approach 1: collapse, group_by, etc.



Approach 2: kmeans aka. "semantic grouping"



Approach 3: Maximum Marginal Relevance

shirt



shirt



Maximum Marginal Relevance

High Level Definition

MMR is an iterative algorithm that selects for each position how similar the current document is to the user query and how dissimilar to the already selected results.

Maximum Marginal Relevance

$$MMR^{def} = \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j)) \right]$$

Maximum Marginal Relevance

shirt



0.91



0.88



0.85



0.83



0.82



$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j)) \right]$$



$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j)) \right]$$



$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j)) \right]$$

Maximum Marginal Relevance

shirt



0.91



0.88



0.85



0.83



0.82

Results (R)

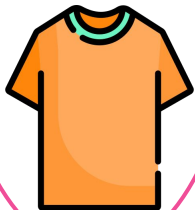
$$= \arg \max_{D_i \in \mathbf{R} \setminus S} \left[\lambda \cdot \text{Sim}_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (\text{Sim}_2(D_i, D_j)) \right]$$

Maximum Marginal Relevance: **Similarity Function**

shirt



S



0.91



0.88



0.85



0.83



0.82

R \ S

$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j)) \right]$$



Candidates for position 3 (D_i)

Maximum Marginal Relevance: **Similarity Function**

Q shirt

$$\lambda \cdot \text{Sim}_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (\text{Sim}_2(D_i, D_j))$$



$$\begin{aligned} &0.5 * 0.88 \\ &- 0.5 * \max(0.95, 0.70) \\ &=-0.035 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.83 - \\ &0.5 * \max(0.85, 0.70) \\ &=-0.01 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.82 - \\ &0.5 * \max(0.60, 0.70) \\ &=0.06 \end{aligned}$$

Maximum Marginal Relevance: Similarity Function

What is Sim()

Sim() can be any numerical value that we would like:

- Ranking model
- Document property
- BM25 or TF/IDF
- **Vector Similarity**

Maximum Marginal Relevance: **Max Dissimilarity Function**

Selected (S) =



$$\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j))$$



$$\begin{aligned} &0.5 * 0.88 - \\ &0.5 * \max(0.95, 0.70) \\ &= -0.035 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.83 - \\ &0.5 * \max(0.85, 0.70) \\ &= -0.01 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.82 - \\ &0.5 * \max(0.60, 0.70) \\ &= 0.06 \end{aligned}$$

Maximum Marginal Relevance: **Lambda** Parameter

$$\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (Sim_2(D_i, D_j))$$



$$\begin{aligned} &0.5 * 0.88 - \\ &0.5 * \max(0.95, 0.70) \\ &= -0.035 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.83 - \\ &0.5 * \max(0.85, 0.70) \\ &= -0.01 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.82 - \\ &0.5 * \max(0.60, 0.70) \\ &= 0.06 \end{aligned}$$

Maximum Marginal Relevance: **Selected Result**

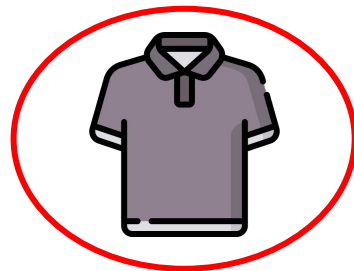
$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot \text{Sim}_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (\text{Sim}_2(D_i, D_j)) \right]$$



$$\begin{aligned} &0.5 * 0.88 - \\ &0.5 * \max(0.95, 0.70) \\ &= -0.035 \end{aligned}$$



$$\begin{aligned} &0.5 * 0.83 - \\ &0.5 * \max(0.85, 0.70) \\ &= -0.01 \end{aligned}$$

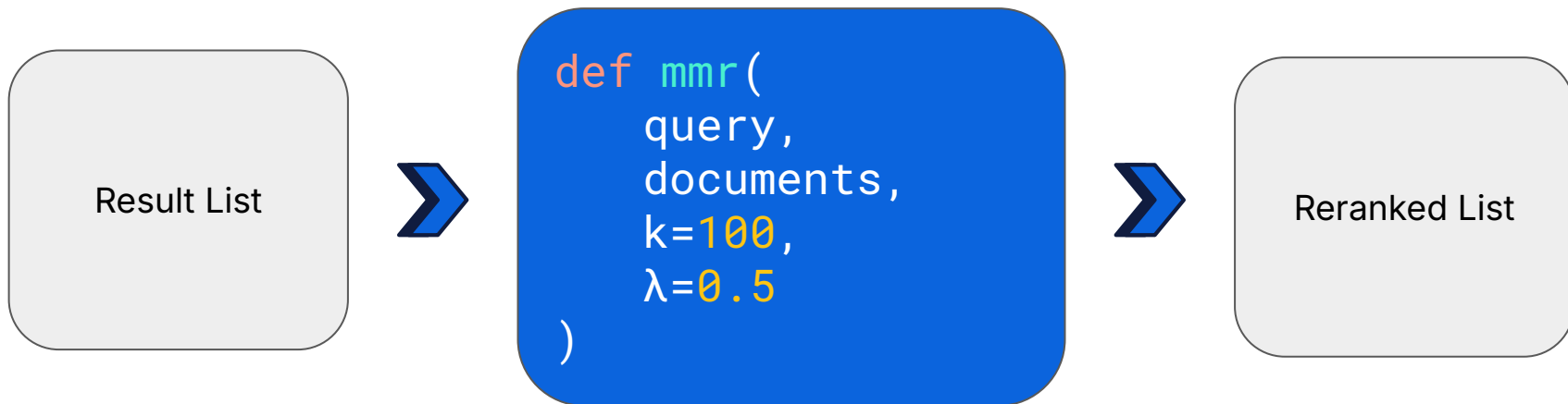


$$\begin{aligned} &0.5 * 0.82 - \\ &0.5 * \max(0.60, 0.70) \\ &= 0.06 \end{aligned}$$



Maximum Marginal Relevance

What do I need to know?



Remember: $\lambda=1.0 \Rightarrow$ Relevancy, $\lambda=0.0 \Rightarrow$ Diversity

Pants

Pure Relevance



ID: 9785

Urban Yoga Women Summer Bottoms - Navy Blue Track Pants

Score: 0.862



ID: 7128

Urban Yoga Women Bottoms - Black Track Pants

Score: 0.861



ID: 19242

Puma Women Grey Capri Pants

Score: 0.858



ID: 3921

Urban Yoga Men's Bottoms - Black Track Pants

Score: 0.857



ID: 52529

Pepi Jeans Men Grey 3/4 Length Shorts

Score: 0.856



ID: 4826

ADIDAS Men's Woven Dark Navy Blue Track Pants

Score: 0.855



ID: 44664

Wills Lifestyle Women Charcoal Track Pants

Score: 0.854



ID: 7133

Urban Yoga Men Bottom Grey Track Pants

Score: 0.854



ID: 43522

French Connection Women Navy Blue Trousers

Score: 0.854



ID: 18869

Puma Women Black Core Track Pants

Score: 0.854

MMR($\lambda=0.5$)



ID: 9785

Urban Yoga Women Summer Bottoms - Navy Blue Track Pants

Score: 0.862



ID: 41163

Allen Solly Woman Khaki Trousers

Score: 0.839



ID: 13255

Palm Tree Kids Boys Checkered Shorts

Score: 0.835



ID: 4774

ADIDAS Women 3S Pink Track Pants

Score: 0.837



ID: 52529

Pepi Jeans Men Grey 3/4 Length Shorts

Score: 0.856



ID: 22466

Myntra Women Cream Palazzo Leggings

Score: 0.836



ID: 44906

Puma Men White 3/4 Length Shorts

Score: 0.853



ID: 32406

Arrow Woman Black Trousers

Score: 0.853



ID: 57824

United Colors of Benetton Green Trousers

Score: 0.842



ID: 30919

Fabindia Women Pink Harem Trousers

Score: 0.840

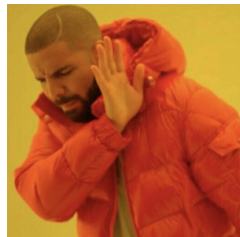
The background is a solid blue color. It features several abstract, 3D-style geometric shapes in a lighter blue shade. These shapes are composed of multiple concentric, slightly offset lines that create a sense of depth and perspective. They are positioned in the corners and along the edges of the frame, framing the central text.

**Is this
always a
good idea?**

It's about **user intent**

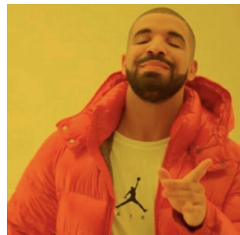
Pixel 10 pro

<KNOWN_ITEM>



Sneakers

<EXPLORATION>

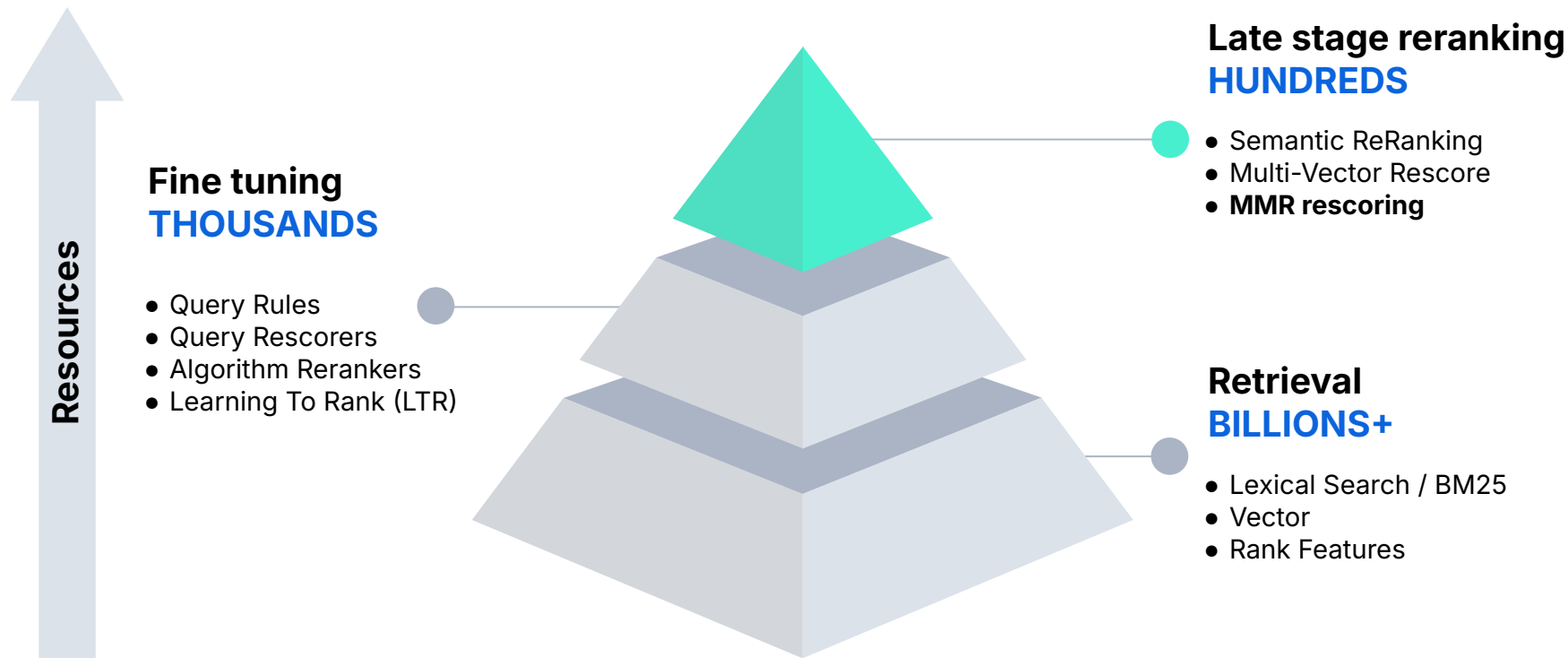


Present for moms birthday?

<INSPIRATIONAL>



It's about **performance**



Improve answer quality of RAG applications

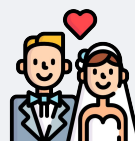
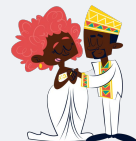
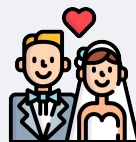
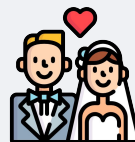
Where should I eat?



Vacation ideas?

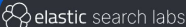


Wedding ceremony?



Want to learn more?



elastic search labs

TutorialsExamplesIntegrationsBlogs


Start free trial

BLOG

Diversifying search results with Maximum Marginal Relevance

Implementing the Maximum Marginal Relevance (MMR) algorithm with Elasticsearch and Python. This blog includes code examples for vector search reranking.

Search RelevanceVector Database

By: **Peter Straßer**
On July 10, 2025

Elasticsearch is packed with new features to help you build the best search solutions for your use case. Dive into our [sample notebooks](#) to learn more, start a [free cloud trial](#), or try Elastic on your [local machine](#) now.

When you search for "pants" in an [e-commerce](#) catalog, do you really want to see 10 variations of the same black capris? Probably not. You'd likely prefer a diverse selection showing different styles, colors, and types of pants. This is where Maximum Marginal Relevance (MMR) comes in — a powerful technique for balancing relevance with diversity in search results.

In this blog, we'll explore how to implement MMR with Elasticsearch to create more diverse and useful search results, using a fashion product catalog as our example.

JUMP TO

The problem: When relevance isn't enough

Enter Maximum Marginal Relevance




How MMR works

Implementing MMR

The impact: Before and after MMR

+ Show more

SHARE



The problem: When relevance isn't enough

Traditional search systems optimize for one thing: relevance. They find items that best match your query and rank them by similarity scores. This works well for many use cases, but it can lead to redundant results.

Thank you!

