

Result Diversification



Peter Straßer
Senior Solution Architect, Elastic



The Problem

Exploration is limited

Men

SEARCH FOR „SNEAKERS“ SHOW ALL SHOES (87)

Filters #

87 products

Sneaker made from fine cowhide
Improved Production
£139.00

Sneaker made from fine cowhide
Improved Production
£159.00

Sneaker made from fine cowhide
Improved Production
£159.00

Sneaker made from high-quality cowhide
Improved Production
£129.00

Sneaker made from fine cowhide
Improved Production
£169.00

Sneaker made from fine suede cowhide
Improved Production
£169.00

Sneaker made from fine cowhide
Improved Production
£159.00

Sneaker made from fine suede
£129.00

elastic

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?

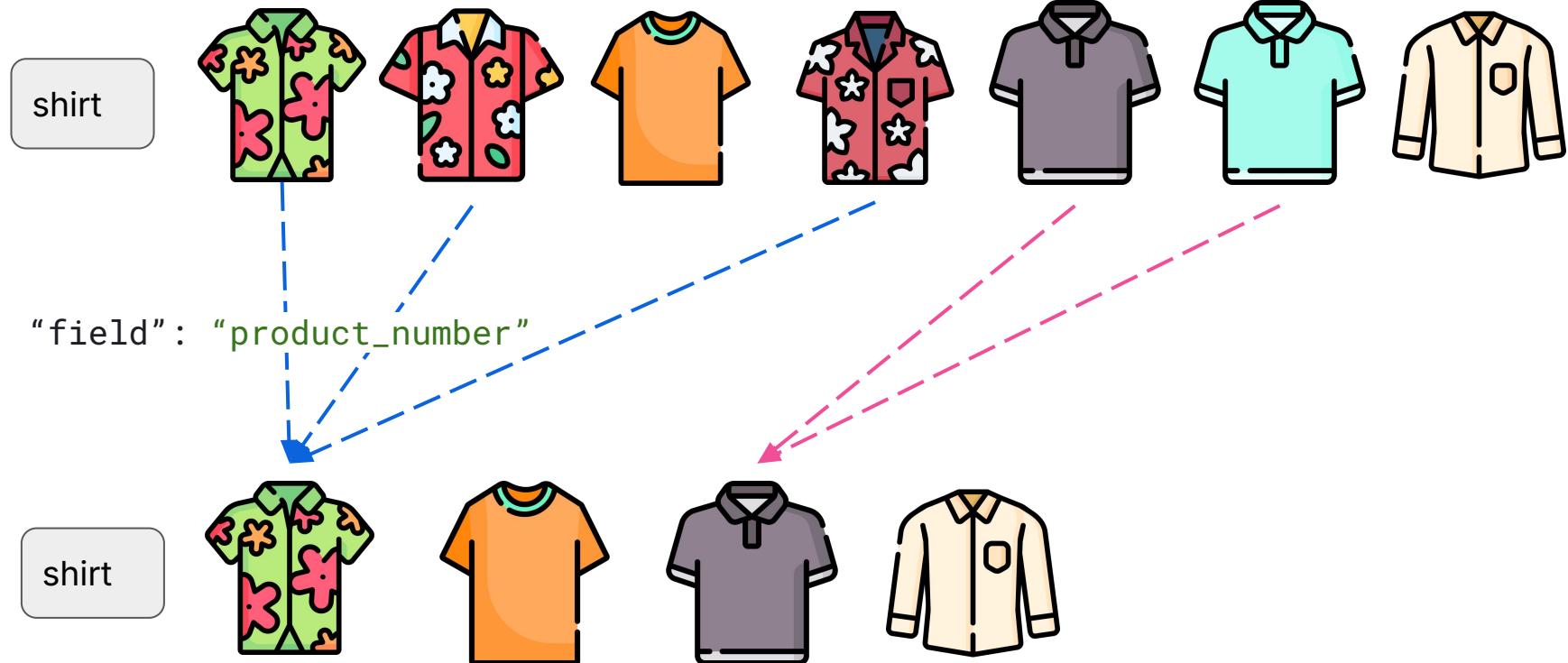
 Designer Polo Ralph Lauren MASTERS CRT TOP UNISEX - Trainers - white/navy €105.00 Regular price: €150.00 -30% Plus Premium delivery	 adidas Performance GALAXY - Road running shoes - core black €54.95 Plus Premium delivery	 Jack & Jones JFWJORDAN - Trainers - navy/bright white/cognac €20.99 Regular price: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

* Slightly modified to fit on screen

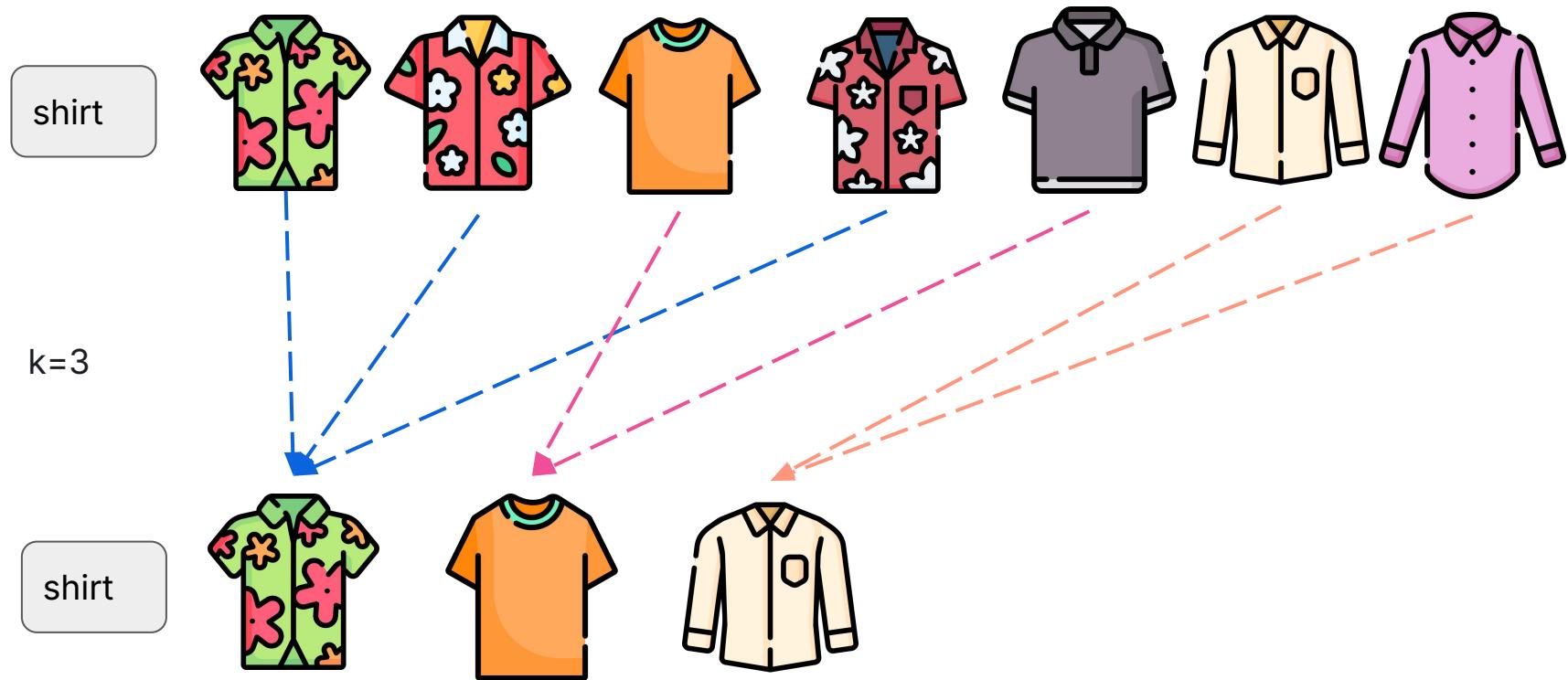


How can we achieve it?

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

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

https://www.cs.cmu.edu/~jgc/publication/The_Use_MMR_Diversity_Based_LTMIR_1998.pdf

Maximum Marginal Relevance

shirt



0.91



0.88



0.85



0.83



0.82



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



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



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

Maximum Marginal Relevance

shirt



0.91



0.88



0.85



0.83



0.82

Results (R)

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

Maximum Marginal Relevance: Similarity Function

shirt



S



0.91



0.88



0.85



0.83



0.82

R\|S

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



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) =



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



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



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



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

Maximum Marginal Relevance: Lambda Parameter

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



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



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



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

Maximum Marginal Relevance: Selected Result

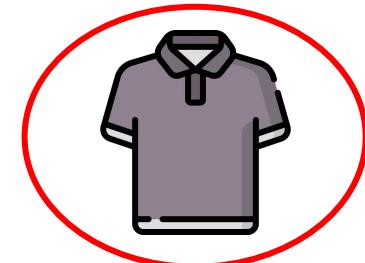
$$= \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot \text{Sim}_1(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} (\text{Sim}_2(D_i, D_j)) \right]$$



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



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

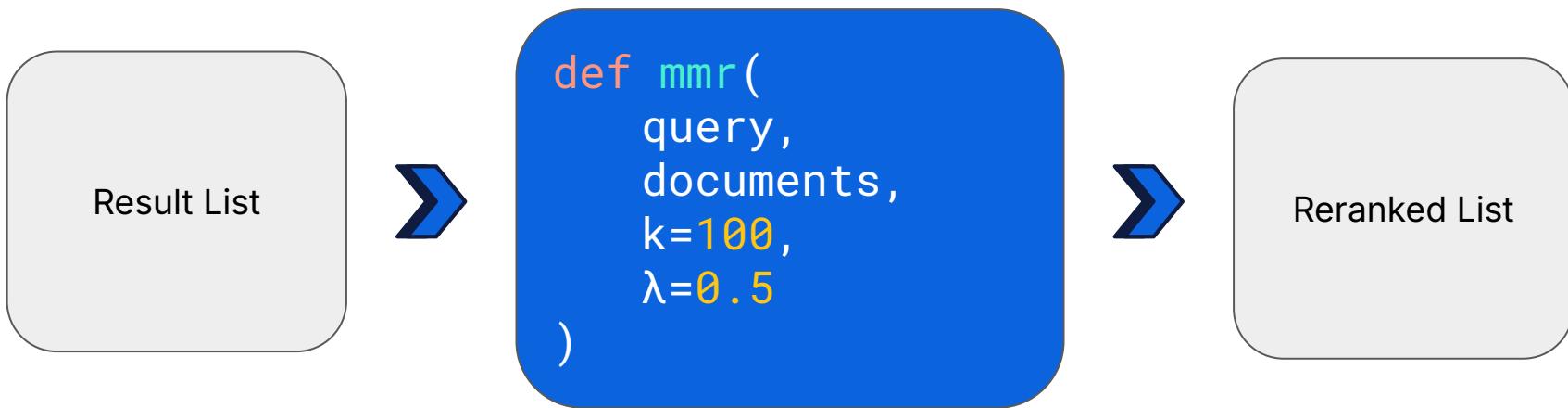


$$\begin{aligned} 0.5 * 0.82 - \\ 0.5 * \max(0.60, 0.70) \\ = 0.06 \end{aligned}$$



Maximum Marginal Relevance

What do I need to know?



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



Pure Relevance



ID: 9785

Urban Yoga Women Summer B...

Track Pants - Navy Blue

Score: 0.862



ID: 7128

Urban Yoga Women Bottom B...

Track Pants - Black

Score: 0.861



ID: 19242

Puma Women Grey Capri Pan...

Capris - Grey

Score: 0.858



ID: 3921

Urban Yoga Men's Bottom B...

Track Pants - Black

Score: 0.857



ID: 52529

Pepe Jeans Men Grey 3/4 L...

Shorts - Grey

Score: 0.856



ID: 4826

ADIDAS Men's Woven Dark N...

Track Pants - Navy Blue

Score: 0.855



ID: 44664

Wills Lifestyle Women Cha...

Trousers - Charcoal

Score: 0.854



ID: 7133

Urban Yoga Men Bottom Gre...

Track Pants - Grey

Score: 0.854



ID: 43522

French Connection Women N...

Trousers - Navy Blue

Score: 0.854



ID: 18869

Puma Women Black Core Tra...

Track Pants - Black

Score: 0.854

MMR($\lambda=0.5$)



ID: 9785

Urban Yoga Women Summer B...

Track Pants - Navy Blue

Score: 0.862



ID: 41163

Allen Solly Woman Khaki T...

Trousers - Khaki

Score: 0.839



ID: 13255

Palm Tree Kids Boys Check...

Shorts - White

Score: 0.835



ID: 4774

ADIDAS Women 3S Pink Trac...

Track Pants - Pink

Score: 0.837



ID: 52529

Pepe Jeans Men Grey 3/4 L...

Shorts - Grey

Score: 0.856



ID: 22466

Mynta Women Cream Patial...

Leggings - Cream

Score: 0.836



ID: 44906

Puma Men White 3/4 Length...

Shorts - White

Score: 0.853



ID: 32406

Arrow Woman Black Trouser...

Trousers - Black

Score: 0.853



ID: 57824

United Colors of Benetton...

Trousers - Green

Score: 0.842



ID: 30919

Fabindia Women Pink Harem...

Trousers - Pink

Score: 0.840



Is this
always a
good idea?

It's about **user intent**

Pixel 10 pro

<KNOWN_ITEM>



Sneakers

<EXPLORATION>

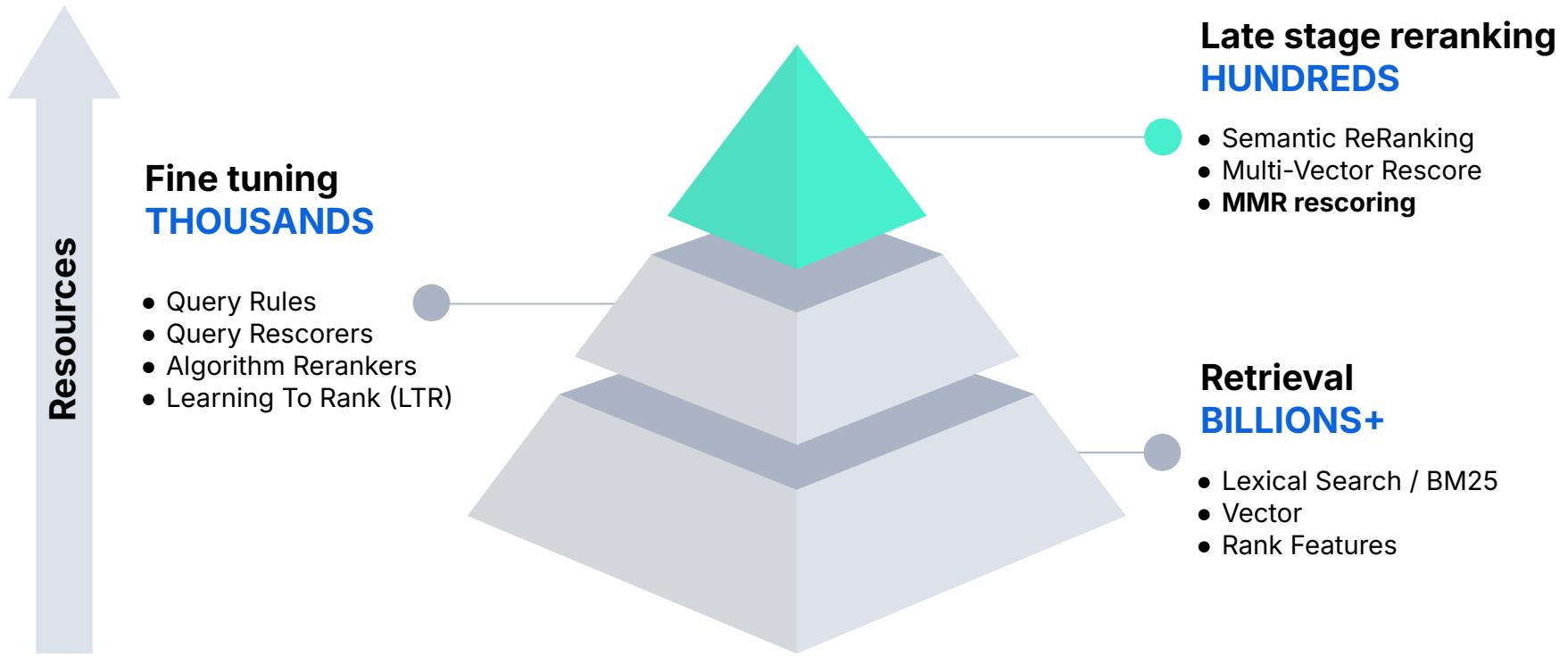


Present for moms birthday?

<INSPIRATIONAL>



It's about performance



Improve answer quality of RAG applications

Where should I eat?



Vacation ideas?



Wedding ceremony?



Want to
learn more?



The image shows a screenshot of a web browser displaying a blog post from elastic search labs. The header includes the elastic logo, search bar, and navigation links for Tutorials, Examples, Integrations, Blogs, and a yellow "Start free trial" button. The main content area has a dark background with a white header titled "BLOG". The main title of the post is "Diversifying search results with Maximum Marginal Relevance". Below the title, a subtitle reads: "Implementing the Maximum Marginal Relevance (MMR) algorithm with Elasticsearch and Python. This blog includes code examples for vector search reranking." A small bio indicates the author is Peter Straßer, posted on July 10, 2025. The post content discusses the limitations of traditional relevance-based search and introduces MMR as a solution. It includes a section on the problem of redundancy in search results and how MMR addresses this. The right sidebar contains a "JUMP TO" menu with links to various sections of the post, such as "The problem: When relevance isn't enough", "Enter Maximum Marginal Relevance", "How MMR works", "Implementing MMR", and "The impact: Before and after MMR". There is also a "+ Show more" link. At the bottom, there are social sharing icons for X, Facebook, and LinkedIn.

<https://www.elastic.co/search-labs/blog/maximum-marginal-relevance-diversify-results>

Thank you!

