

FASHION IMAGE RETRIEVAL SYSTEM

NGUYỄN CHÁNH NGHĨA : 21520362

NGUYỄN QUỐC HƯNG : 21520523

LECTURE: MAI TIẾN DŨNG

Agenda

INTRODUCTION

METHOD

EVALUATION

DEMO

Motivation



- The surge in **online shopping** is increasing steadily
- E-commerce businesses are looking to enhance their search engine systems
- Traditionally, search based on only **Image or Text**

Improvement



- Combining both **text** and **images** based on user descriptions (**images + description**)
- Hybrid search combines the strengths of **traditional keyword-based** and **semantic searches**

Provides results that are the more accurate and relevant -> Enhances user experience

PROBLEM DEFINITION

INPUT

**A database contains fashion products and
corrsponding metadata**

An image of fashion or A description of that fashion

OUTPUT

The most relevant fashion images in the database

METHOD

Dense Vector With CLIP

Contrastive Language - Image Pretraining

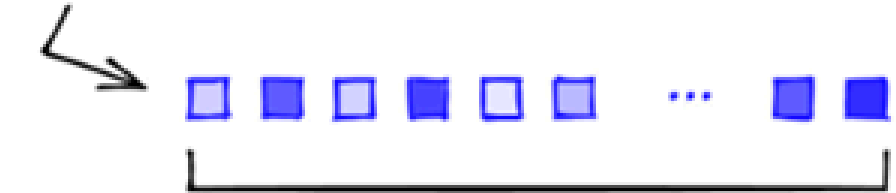
Dense Vector - Semantic Search

Definition

A dense vector is a compact list of mostly non-zero numbers, providing a detailed and nuanced description for accurate searches and comprehensive understanding.

dense

$[0.2, 0.7, 0.1, 0.8, 0.1, \dots 0.9]$



784

Advantages

- Semantic Understanding
- Contextual Analysis
- Multi-modality (text, images, audio, etc.) and cross-modal search (e.g., text-to-image)
- Handling Ambiguous Queries

Disadvantages

- High Dimensionality and Resource Intensiveness
- Need for Vast Amounts of Data for training
- Scalability Issues
- Lack of Explainability - No exact match

Dense Vector With CLIP

Contrastive

is a pretrained model for telling you how well a given image and a given text fit together, introduced by OpenAI in 2021

Language

builds on a large body of work on zero-shot transfer, contrastive pre-training, and multimodal learning

Image

is trained on approximately 400 million text and image pairs which are scrapped on the Internet.

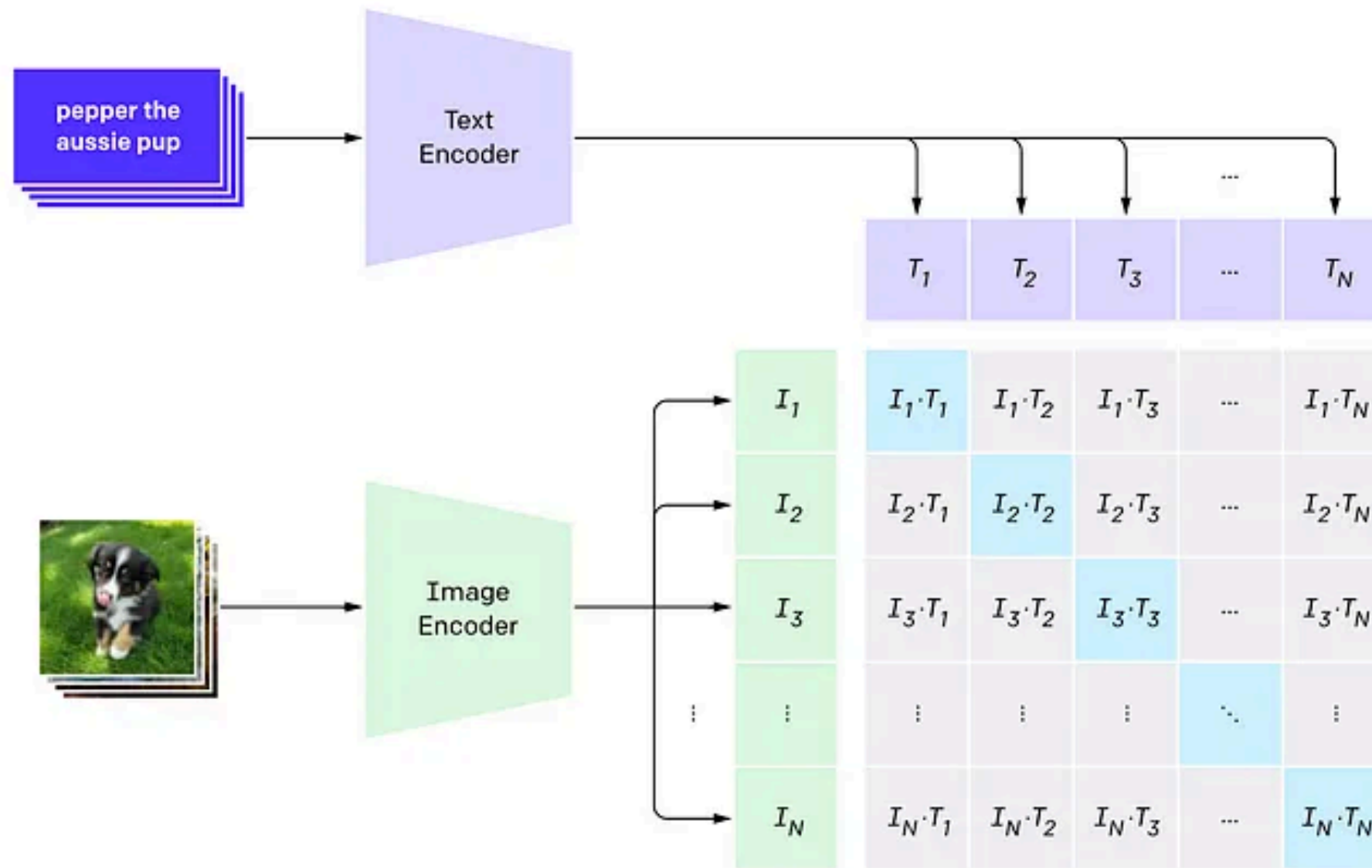
Pre-training

CLIP can find whether a given image and textual description match without being trained for a specific domain.

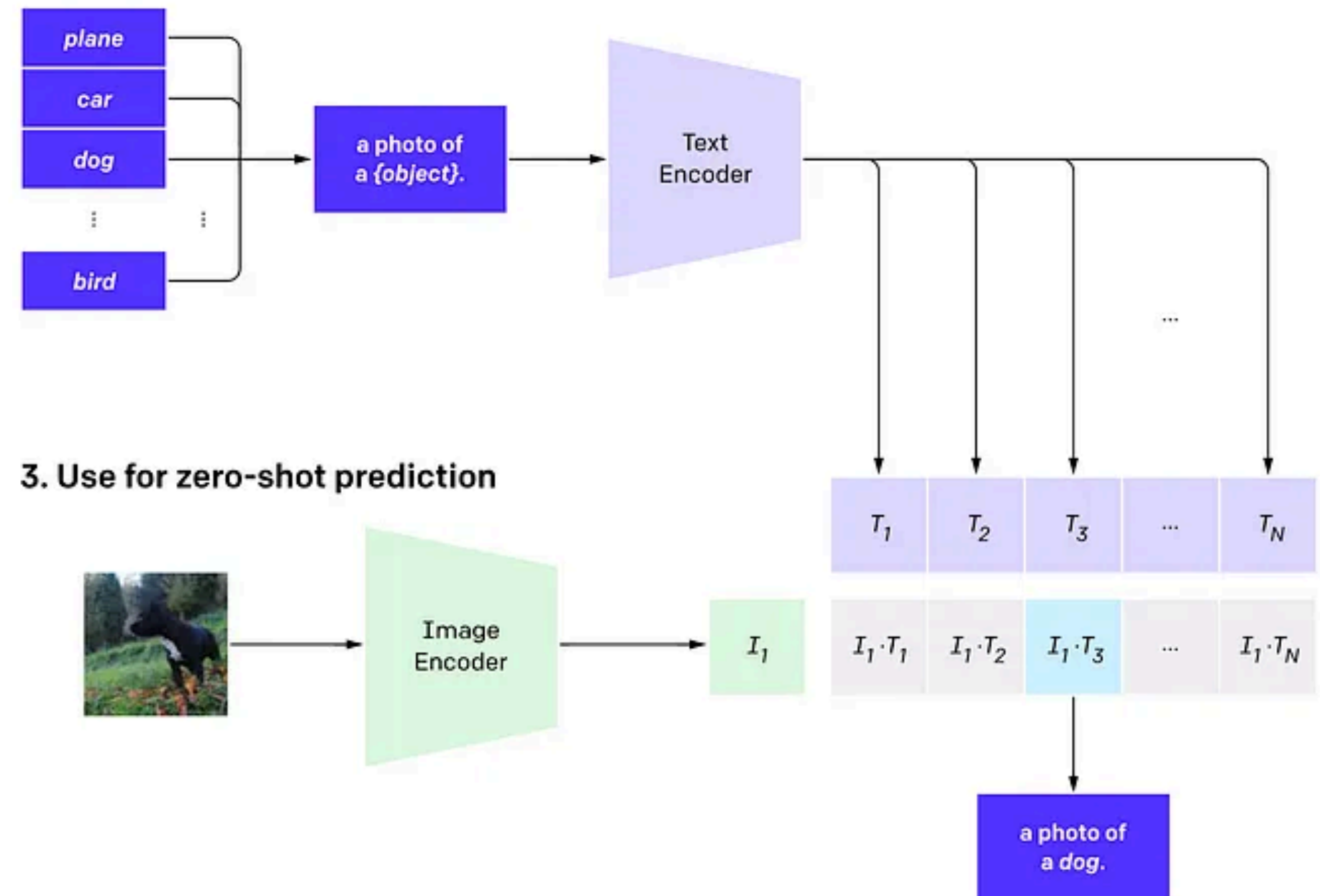


CLIP is suitable for the increasing amount of new data problem

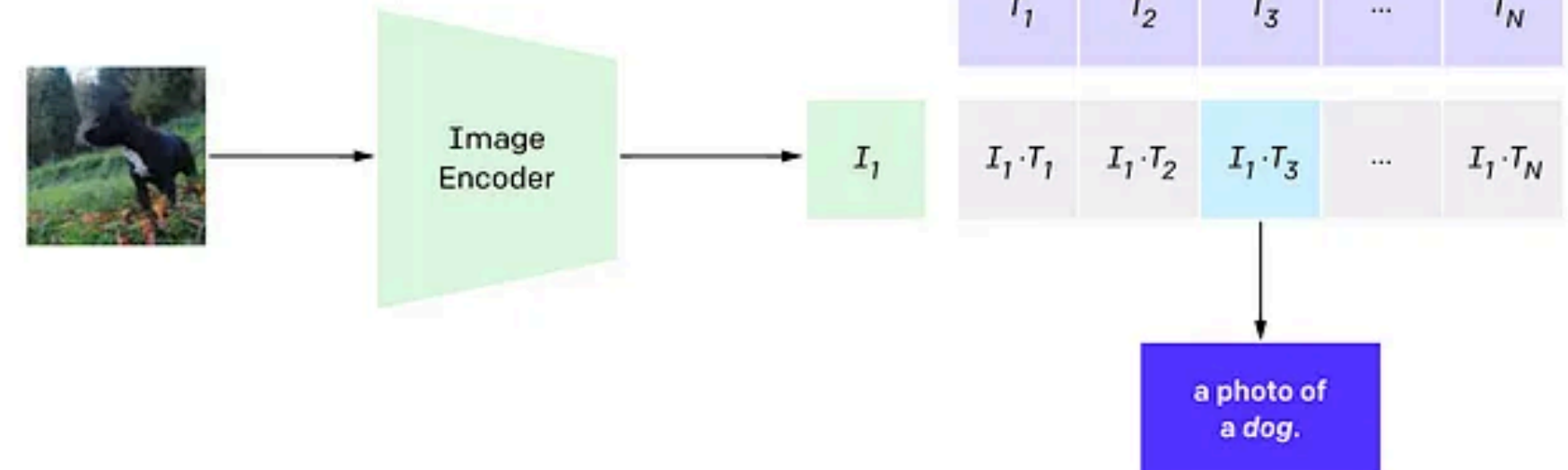
1. Contrastive pre-training



2. Create dataset classifier from label text



3. Use for zero-shot prediction



CLIP pre-trains an image encoder and a text encoder to predict which images were paired with which texts in our dataset. We then use this behavior to turn CLIP into a zero-shot classifier. We convert all of a dataset's classes into captions such as "a photo of a dog" and predict the class of the caption CLIP estimates best pairs with a given image.

Sparse Vector With SPLADE

Spars**e** **L**exical **a**nd **E**xpans**i**on

Sparse Vector - Keyword Search

Definition

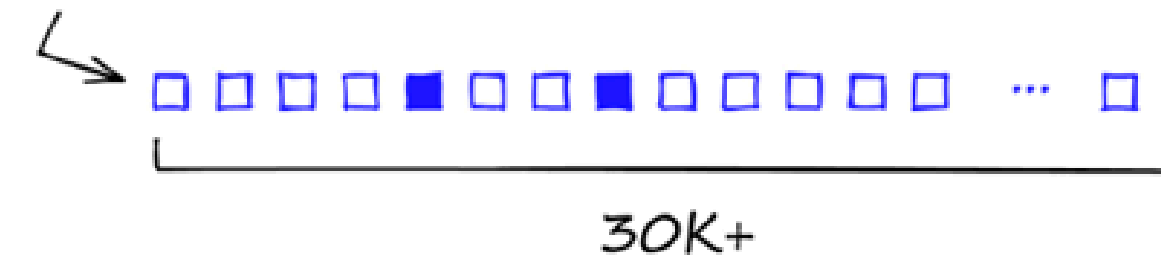
Sparse vectors are representations where the majority of elements are zero, and only non-zero elements and their positions are explicitly stored

Advantages

- Memory Efficient
- Interpretability
- Literal Matching
- Explicit Query

sparse

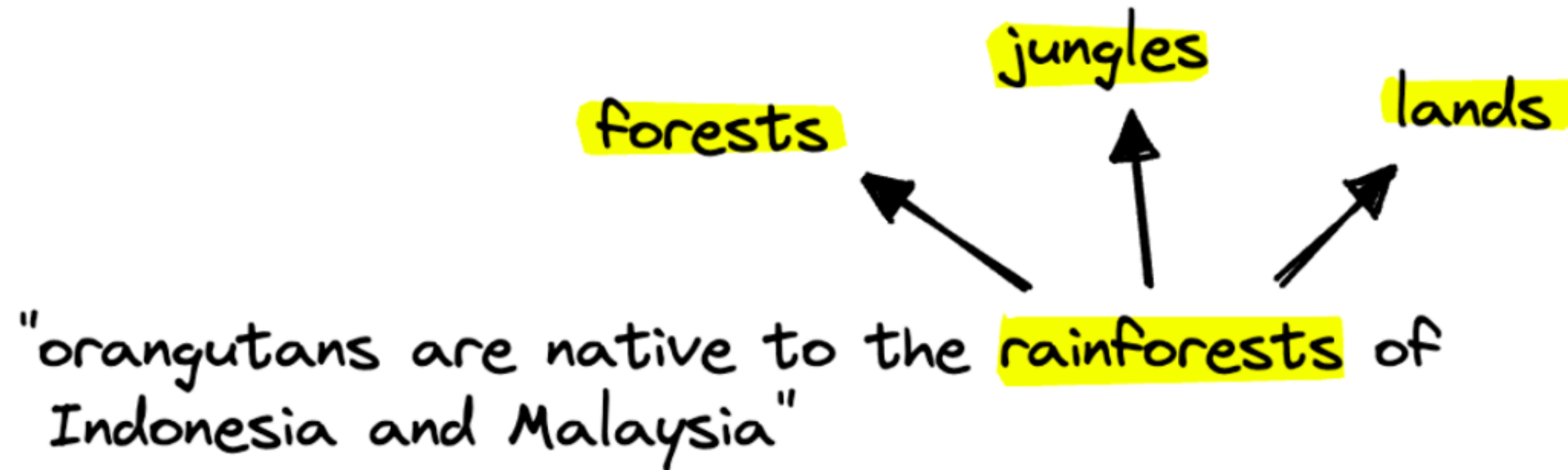
$[0, 0, 0, 1, 0, \dots 0]$



Disadvantages

- Difficulty with Synonyms and Variations:
 - Slipper - flip flop, thắt lưng - dây nịt
- Overly Broad or Narrow Results
- Query Ambiguity: User seeks "sneakers," query is "Comfortable casual shoes for everyday activities."

Sparse Vector With **SPLADE**



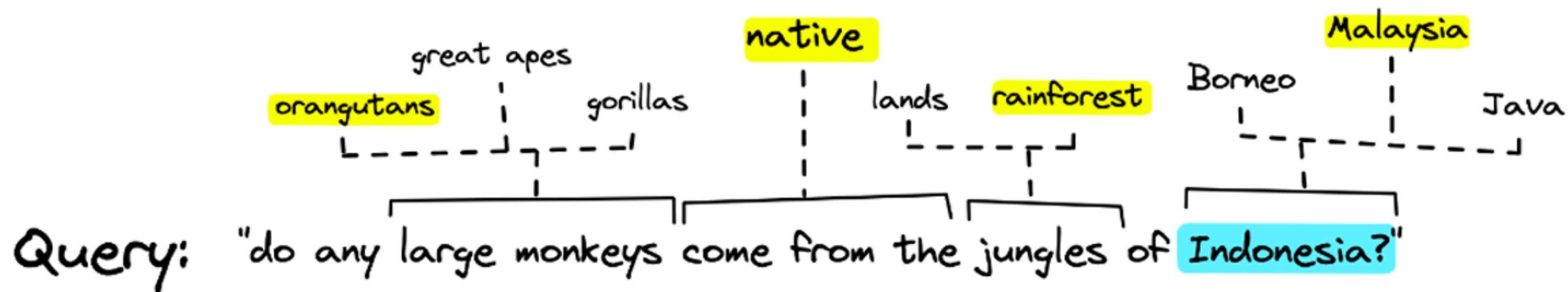
Traditional Sparse

Query: "do any large monkeys come from the jungles of **Indonesia?**"

term overlap

Doc: "Orangutans are native to the rainforests of **Indonesia** and Malaysia"

Spalde Sparse Vector

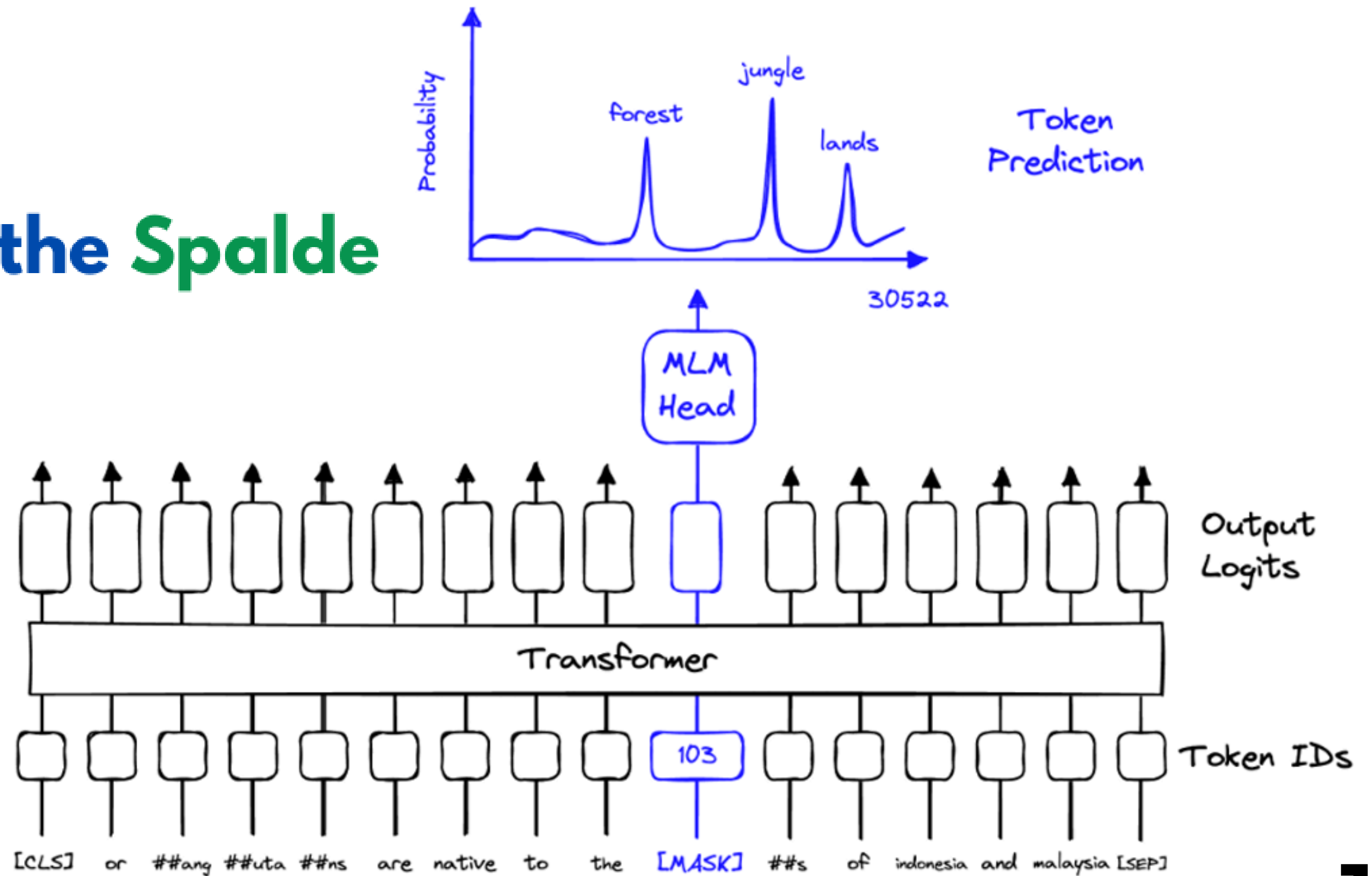


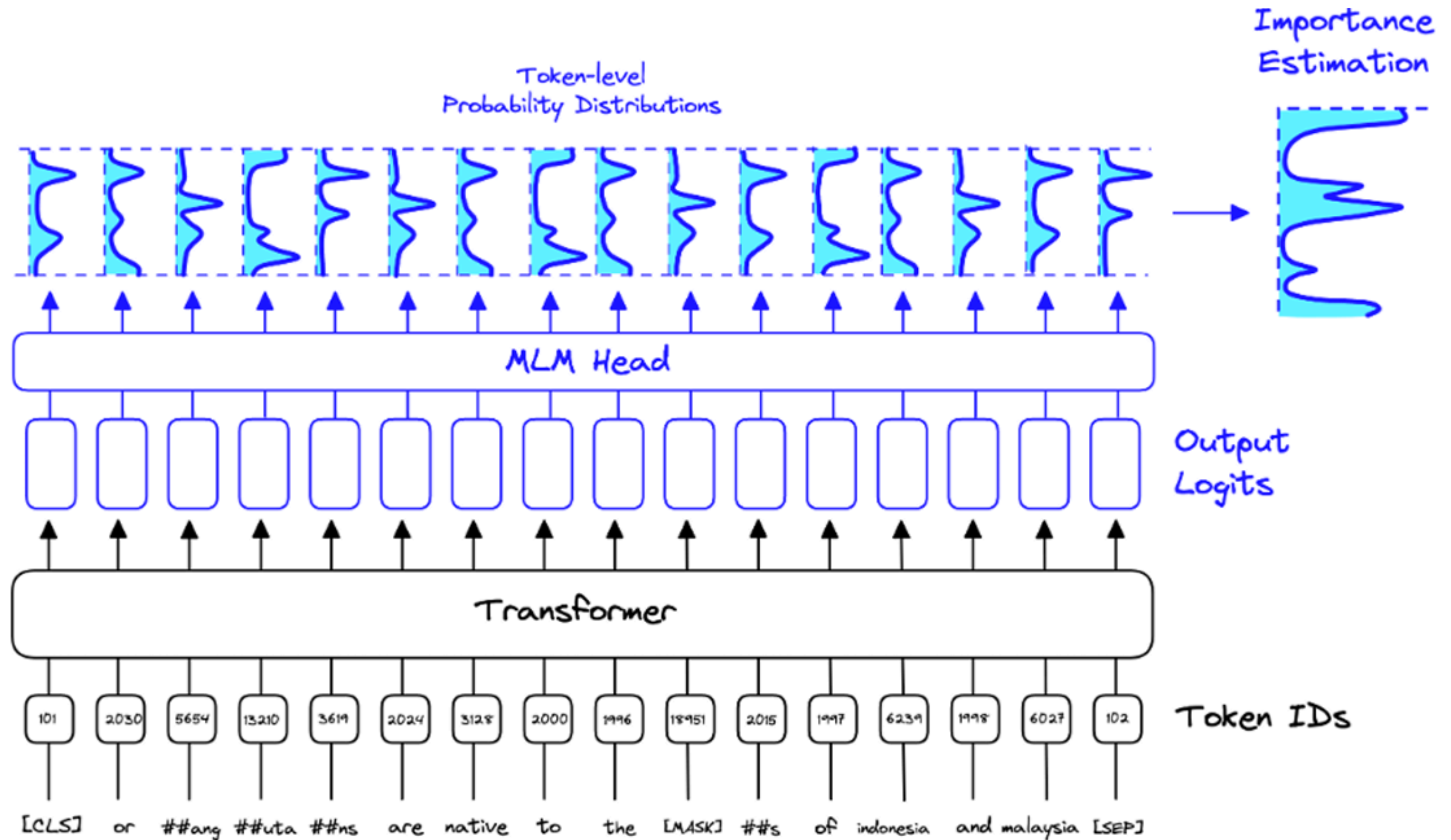
with query expansion

without query expansion

Doc: "Orangutans are native to the rainforests of Indonesia and Malaysia"

How to build the Spalde





Datasets: ashraq/fashion-product-images-small

like 19

Tags: Croissant

DATASET

Dataset card Viewer Files and versions Community

Dataset Viewer








Auto-converted to Parquet

</> API

View in Dataset Viewer

Split (1)
train · 44.1k rows

Search this dataset

baseColour	season	year	usage	productDisplayName	image
string · classes	string · classes	float64	string · classes	string · lengths	image
 46 values	 4 values	 2.01k 2.02k	 8 values	 7 88	53
Navy Blue	Fall	2,011	Casual	Turtle Check Men Navy Blue Shirt	
Blue	Summer	2,012	Casual	Peter England Men Party Blue Jeans	

Downloads last month

</> Use in Datasets library

Edit

⋮

Size of downloaded dataset files:
271 MBSize of the auto-converted Parquet files:
271 MB

Spaces using ashraq/fashion-1

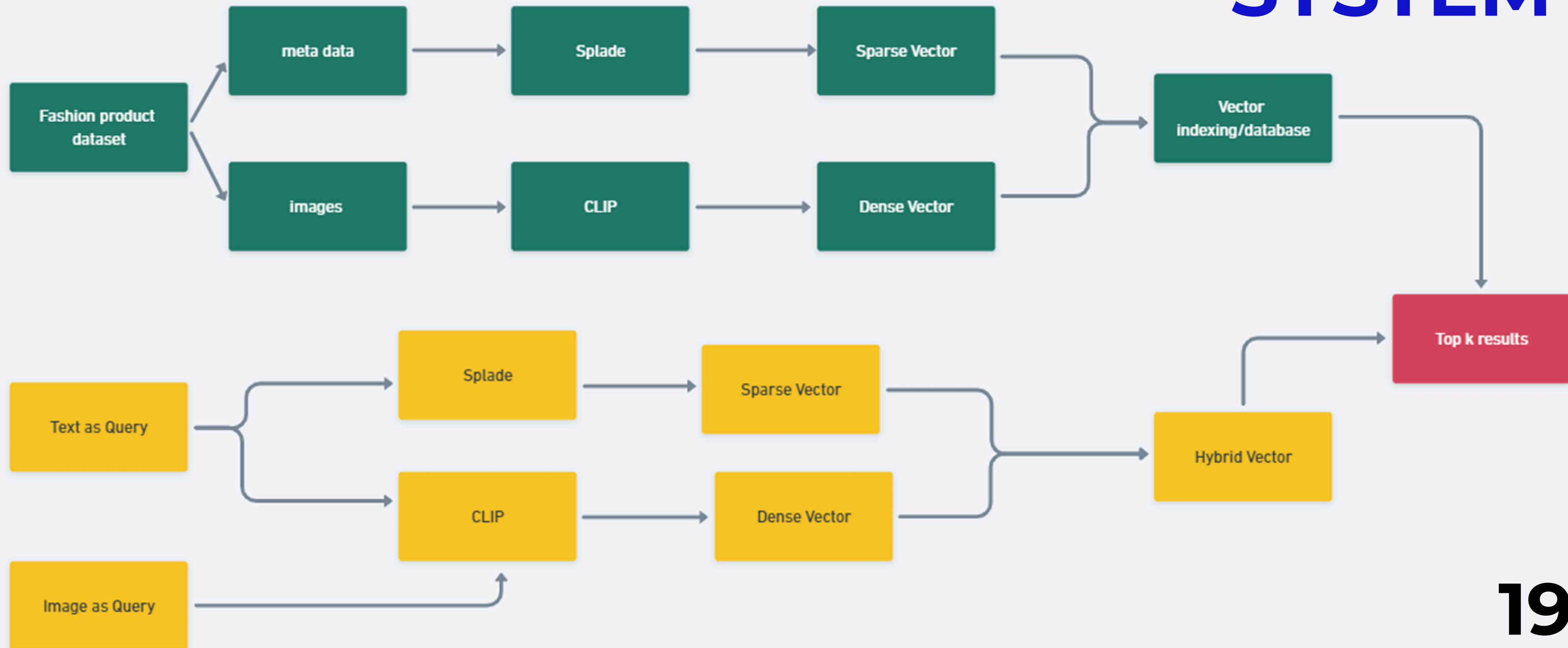
DATASET

Rows: 44,072

Columns : 11

(10 for metadata 1 for images)

SYSTEM



Score Ranking of Pinecone

The score is the sum of the **dot product** of its **dense values** with the **dense part** of the query, together with the **dot product** of its **sparse values** with the **sparse part** of the query.

MAP

EVALUATION WITH MAP

Dense	Sparse	Hybrid
0,58	0,72	0,93

EVALUATION WITH MAP

Multi
0,89

0.5

Text Query

black sock for man











Query image

Drop file here

Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

Browse files

<div>Score: 28.99</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.63</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.48</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.42</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.36</div> <div></div> <div>Raymond Men Black Socks</div>
<div>Score: 28.3</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.21</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 28.05</div> <div></div> <div>Park Avenue Men Black Socks</div>	<div>Score: 27.93</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 27.76</div> <div></div> <div>Raymond Men Black Socks</div>

0.5

Text Query

Query image

Drop file here



Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG

Browse files



Screenshot 2024-05-06 161412.png 22.1KB X



Score: 51.94



Playboy Men White Socks

Score: 51.63



Jockey LCESCBRA Men
White Pack of 3 Socks
2004

Score: 51.28



Playboy Men Black and
White Socks

Score: 51.21



Puma Men Sports Socks
1 Pair White Socks

Score: 50.93



ADIDAS Unisex Pack of 3
Crew Socks

Score: 50.89



Nike Unisex Pack of 3
White Socks

Score: 50.81



Hanes Men Pack Of 3
White Cushion Max Crew

Score: 50.69



Playboy Men Grey Socks

Score: 50.6



Playboy Men White
Striped Socks

Score: 50.59



ADIDAS Originals Men
White Casual Socks


0.5

Text Query


Black


Query image











Drop file here

 Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG

Browse files

 Screenshot 2024-05-06 161412.png 22.1KB X



<div>Score: 55.35</div> <div></div> <div>Playboy Men Black and White Socks</div>	<div>Score: 53.51</div> <div></div> <div>Reid & Taylor Men Solid Black Socks</div>	<div>Score: 53.49</div> <div></div> <div>Playboy Men Grey & Black Socks</div>	<div>Score: 53.36</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 53.31</div> <div></div> <div>Raymond Men Black Socks</div>
<div>Score: 53.31</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 53.27</div> <div></div> <div>Playboy Men White Socks</div>	<div>Score: 53.27</div> <div></div> <div>Raymond Men Black Socks</div>	<div>Score: 52.91</div> <div></div> <div>Playboy Women Black Socks</div>	<div>Score: 52.89</div> <div></div> <div>Playboy Men Black & White Socks</div>

DEMO

REFERENCES

1. Information Retrieval Model System Design
2. Vector Search: What is Vector Search and How Does it Work?
3. Hybrid Search Introduction - Pinecone
4. MetaRank - Pinecone
5. SPLADE - Pinecone
6. Paper: Learning Transferable Visual Models From Natural Language Supervision
7. Sparse-Dense Vectors - Pinecone Documentation
8. Hybrid Search and SPLADE - Medium Article
9. YouTube Video: Hybrid Search and Neural Retrieval Models
10. Paper: Hybrid Search with Sparse and Dense Vectors
11. OpenAI's CLIP - What is CLIP and How to Use It?

*Thank
You*

