



# ENI TECH FEST



Et si nous cherchions



**des morceaux de musique ?**



GRUPE ENI

Présenté par

**David Pilato**

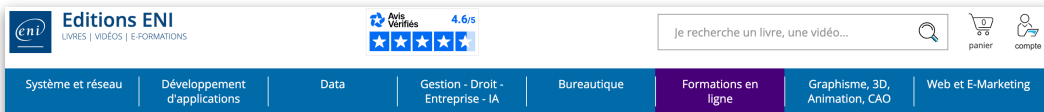


@pilato.fr



@dadoonet

# Découvrez Elasticsearch en 2H30 !



eni Editions ENI LIVRES | VIDÉOS | E-FORMATIONS

4.6/5 Avis Vérifiés

Je recherche un livre, une vidéo...

panier compte

Système et réseau | Développement d'applications | Data | Gestion - Droit - Entreprise - IA | Bureautique | Formations en ligne | Graphisme, 3D, Animation, CAO | Web et E-Marketing

Accès illimité à tous les livres & vidéos, sur l'IA, le dev, les réseaux, la data... Cliquez ici

## Elasticsearch

Indexez vos données pour une recherche efficace

David PILATO



★★★★★ 1 avis

Dans nos systèmes d'information actuels, les utilisateurs sont amenés à effectuer des recherches en temps réel dans des volumes de données souvent importants. L'objectif de cette vidéo est de vous apprendre à utiliser Elasticsearch pour indexer des données de façon à proposer ensuite des résultats pertinents. Pour cela, après une présentation générale d'Elasticsearch et de son écosystème, vous verrez comment démarrer un cluster puis comment structurer des documents JSON pour qu'ils...



Niveau Confirmé



Durée 2h29



Parution mai 2022

Présentation

Table des matières

Avis clients

Auteur

Dans nos systèmes d'information actuels, les utilisateurs sont amenés à effectuer des **recherches en temps réel** dans des **volumes de données** souvent importants. L'objectif de cette vidéo est de vous apprendre à utiliser **Elasticsearch** pour **indexer des données** de façon à proposer ensuite des **résultats pertinents**.

Pour cela, après une présentation générale d'Elasticsearch et de son écosystème, vous verrez comment **démarrer un cluster** puis comment **structurer des documents JSON** pour qu'ils correspondent à un besoin métier. Vous étudierez ensuite comment les indexer et comment les rechercher, que ce soit en mode « Full text » (texte intégral) ou à l'aide de valeurs précises ou numériques.

Ajouter au panier 34,99 €



Accédez en illimité à tous les livres et vidéos du catalogue Editions ENI, avec la Bibliothèque Numérique ENI

Je découvre

### Informations

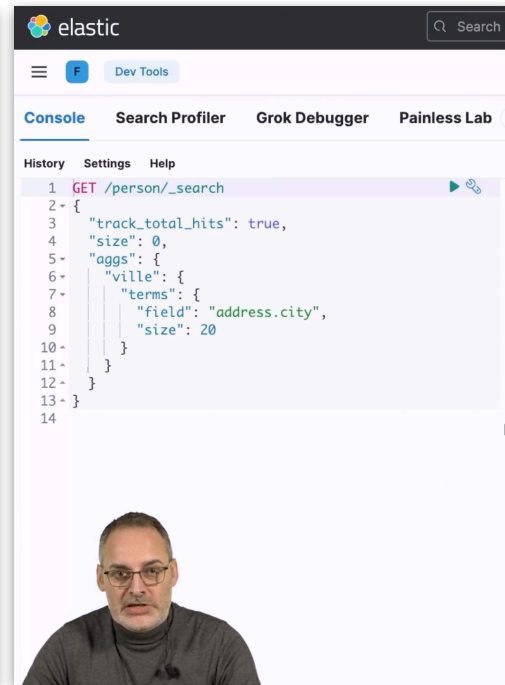
- Consultable en ligne immédiatement après validation du paiement et pour une durée de 10 ans.



Vidéos enregistrées dans notre studio nantais

### Caractéristiques

Vidéo en ligne  
Ref. ENI : VTELASTIC



elastic Search Elastic

Dev Tools

Console Search Profiler Grok Debugger Painless Lab

History Settings Help

```
1 GET /person/_search
2 - {
3   "track_total_hits": true,
4   "size": 0,
5   "aggs": {
6     "ville": {
7       "terms": {
8         "field": "address.city",
9         "size": 20
10      }
11    }
12  }
13 - }
14
```



<https://www.editions-eni.fr/video/elasticsearch-indexez-vos-donnees-pour-une-recherche-efficace-vtlastic>



GROUPE ENI

1<sup>re</sup> édition  
25.03.2025  
LA CARRIÈRE, SAINT-HERBLAIN

# Elasticsearch

You Know, for Search



GROUPE ENI



**ENI**  
TECH FEST



# Elasticsearch

*APACHE*

**FLUCENE™**



GRUPE ENI



**ENI**  
TECH FEST





66

These are not the droids  
you are looking for.



GROUPE ENI

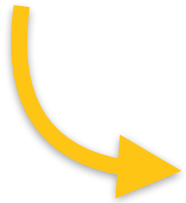


1<sup>re</sup> édition  
25.03.2025  
LA CARRIÈRE, SAINT-HERBLAIN

```
GET /_analyze
{
  "char_filter": [ "html_strip" ],
  "tokenizer": "standard",
  "filter": [ "lowercase", "stop", "snowball" ],
  "text": "These are <em>not</em> the droids
          you are looking for."
}
```

```
"char_filter": "html_strip"
```

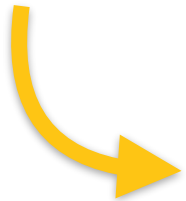
These are `<em>not</em>` the droids you are looking for.



These are not the droids you are looking for.

"tokenizer": "standard"

These are not the droids you are looking for.



These  
are  
not  
the  
droids  
you  
are  
looking  
for

"filter": "lowercase"

<b>T</b> hese	→	<b>t</b> hese
are		are
not		not
the		the
droids		droids
you		you
are		are
looking		looking
for		for

"filter": "stop"

<b>T</b> hese	→	<b>t</b> hese	→	
are		are		
not		not		
the		the		
droids	→	droids	→	droids
you		you		you
are		are		
looking		looking		looking
for		for		

"filter": "snowball"

<b>T</b> hese	→	<b>t</b> hese	→	droids <b>s</b>	→	droid
are		are		you		you
not		not		look <b>ing</b>		look
the		the				
droids		droids				
you		you				
are		are				
looking		looking				
for		for				



These are `<em>not</em>` the **droids** you are **looking** for.

```
{ "tokens": [{
  "token": "droid",
  "start_offset": 27, "end_offset": 33,
  "type": "<ALPHANUM>", "position": 4
}, {
  "token": "you",
  "start_offset": 34, "end_offset": 37,
  "type": "<ALPHANUM>", "position": 5
}, {
  "token": "look",
  "start_offset": 42, "end_offset": 49,
  "type": "<ALPHANUM>", "position": 7
}]}
```

**Semantic**  
search  
≠  
**Literal**  
matches



GROUPE ENI

similarweb

**YOU'RE COMPARING  
APPLES TO NECTARINES**



# Elasticsearch

You Know, for Search



GROUPE ENI



**ENI**  
TECH FEST

# Elasticsearch

You Know, for **Vector** Search



GROUPE ENI



**ENI**  
TECH FEST



GROUPE ENI





1<sup>re</sup> édition  
25.03.2025  
LA CARRIÈRE, SAINT-HERBLAIN

# Embeddings represent your data

Example: 1-dimensional vector



Character	Vector
	$[-1]$
	$[1]$



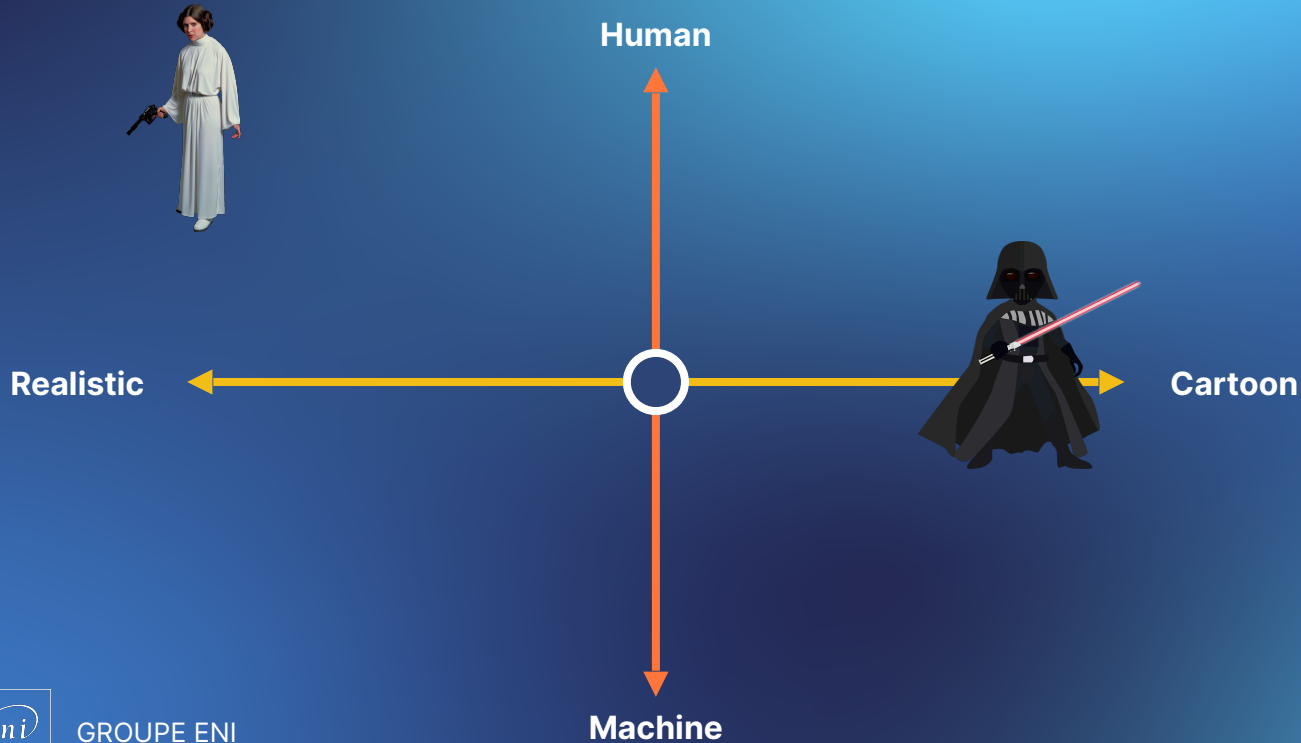
GROUPE ENI







# Multiple dimensions

represent different data aspects

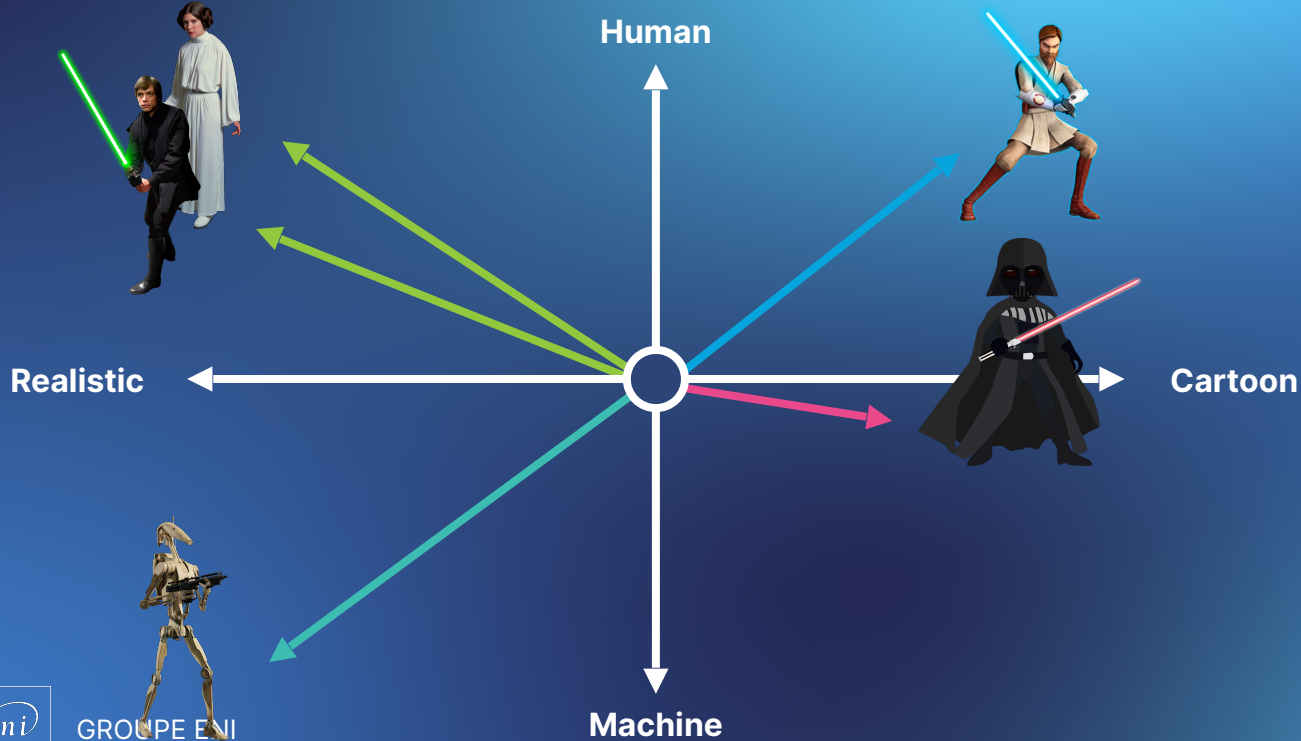



Character	Vector
	$[-1, 1]$
	$[1, 0]$



GRUPE ENI

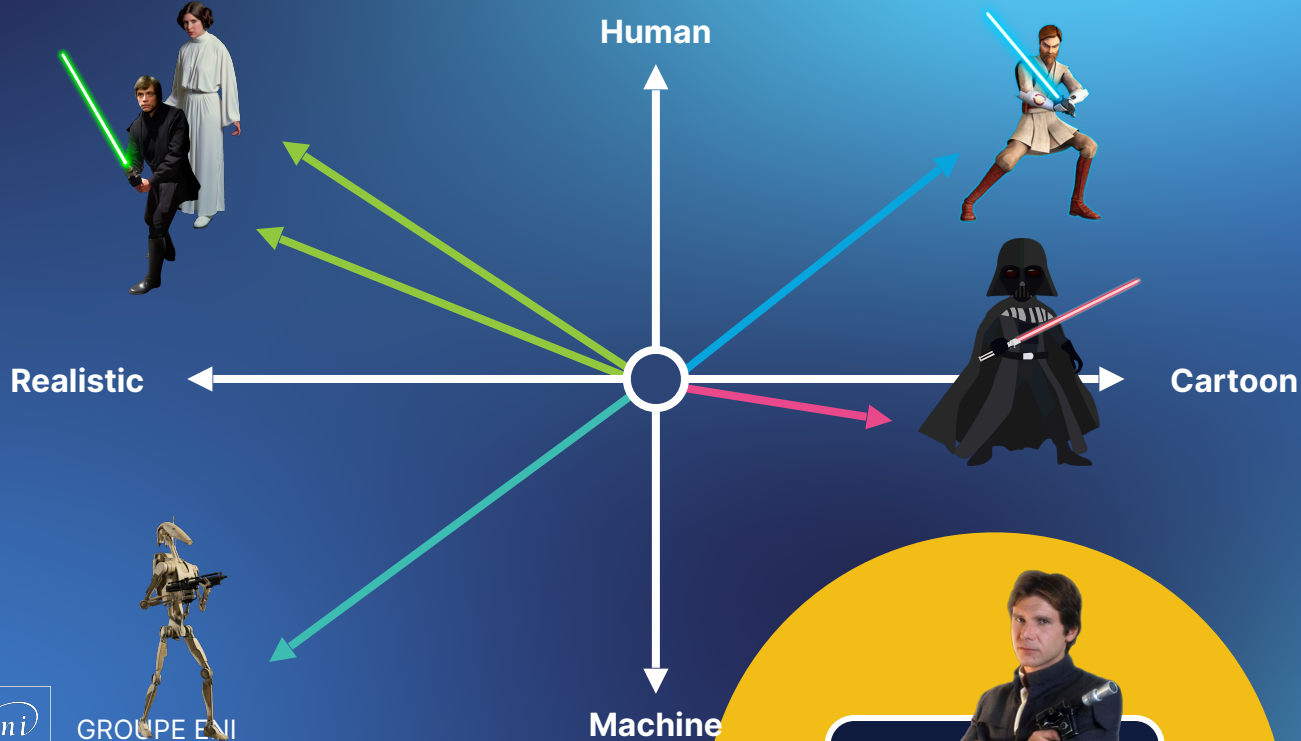
# Similar data is grouped together



Character	Vector
	$[-1.0, 1.0]$
	$[1.0, 0.0]$
	$[-1.0, 0.8]$
	$[1.0, 1.0]$
	$[-1.0, -1.0]$



# Vector search ranks objects by similarity (~relevance) to the query



Rank	Result
Query	
1	
2	
3	
4	
5	



GRUPE ENI



ENI  
TECH FEST

# How do you

index **vectors**?

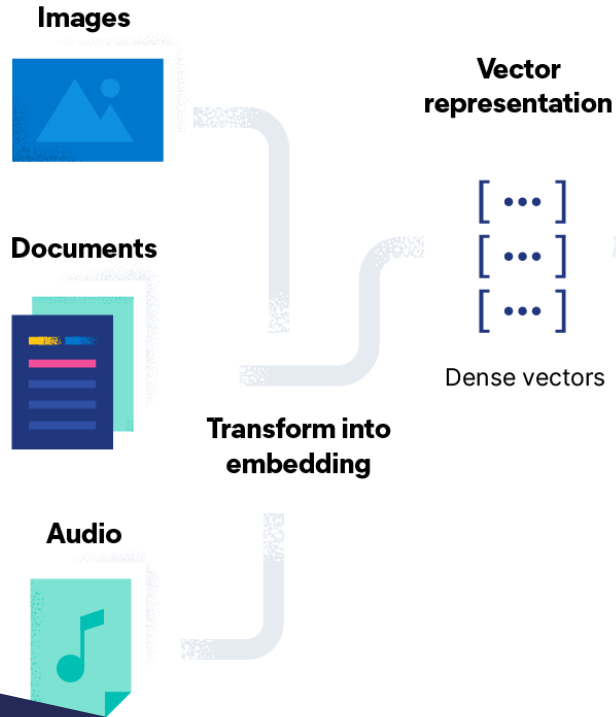


GROUPE ENI



**ENI**  
TECH FEST

# Architecture of Vector Search



# dense\_vector field type

```
PUT ecommerce
{
  "mappings": {
    "properties": {
      "description": {
        "type": "text"
      }
      "desc_embedding": {
        "type": "dense_vector"
      }
    }
  }
}
```

# Data Ingestion and Embedding Generation

POST /ecommerce/\_doc

```
{
  "_id": "product-1234",
  "product_name": "Summer Dress",
  "description": "Our best-selling...",
  "Price": 118,
  "color": "blue",
  "fabric": "cotton",
}
{ "desc_embedding": [0.452, 0.3242, ...],
}
{ "img_embedding": [0.012, 0.0, ...]
}
```

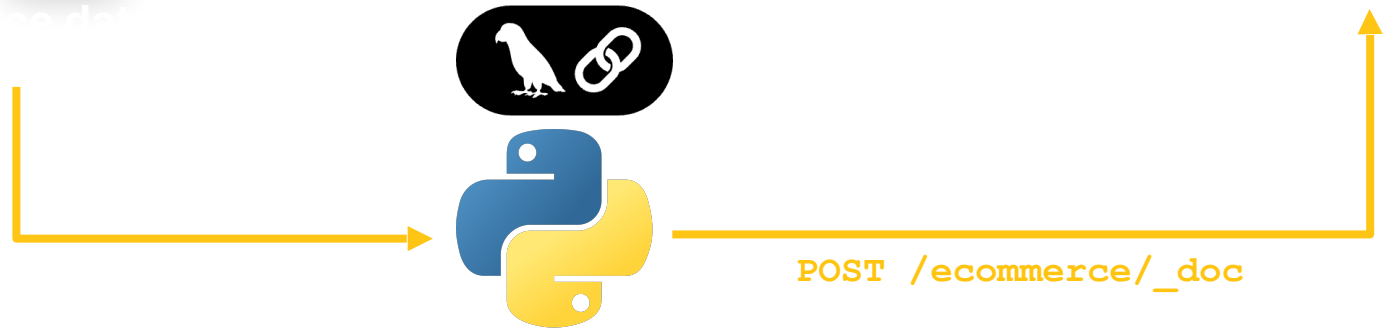

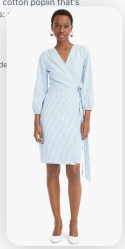
You asked, we answered: Our best-selling classic wrap dress now comes in a cotton poplin that's wear-all-day perfect. Bonus: stripes (our favorite).

**FIT**

- 39" from high point of shoulder

**DETAILS**

- Cotton
- Lined
- Machine wash
- Import



# How do you

search **vectors**?



GROUPE ENI



**ENI**  
TECH FEST

# Architecture of Vector Search



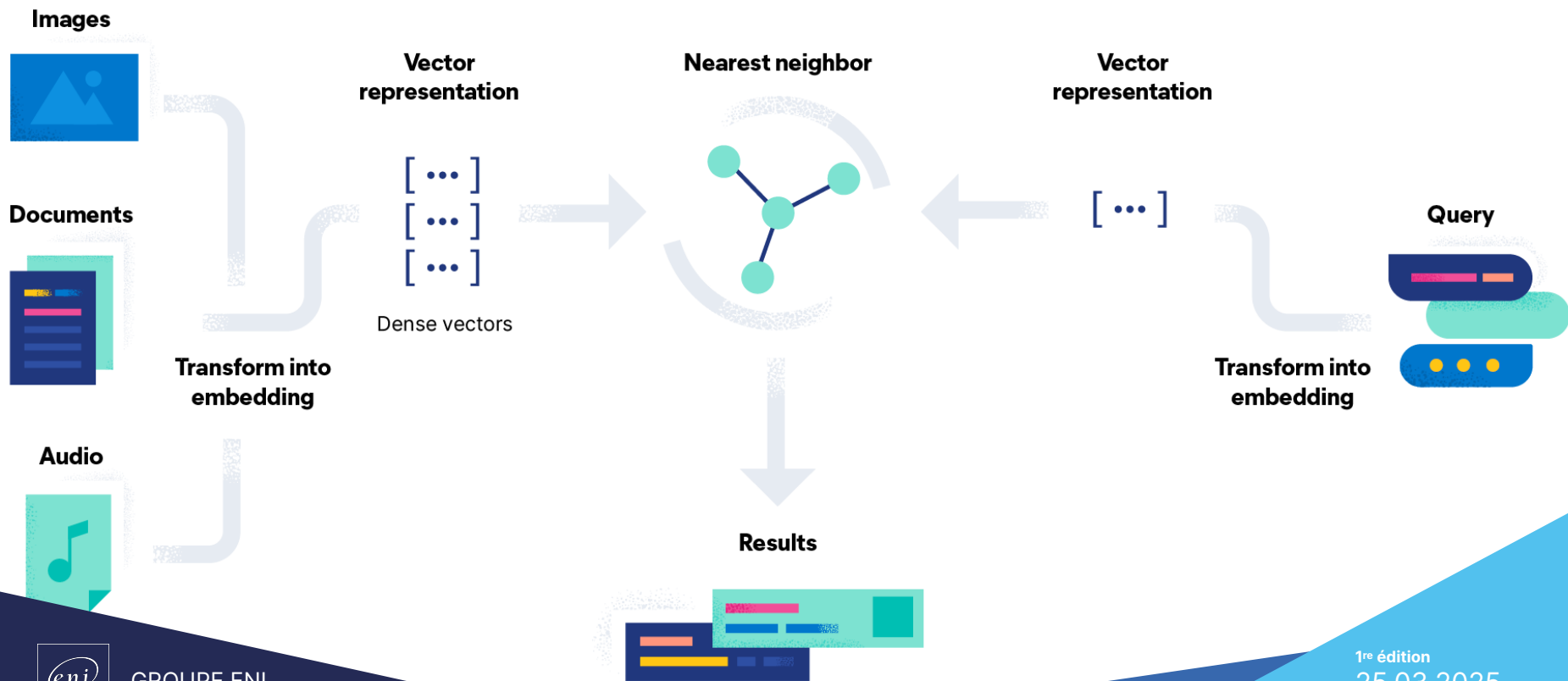
# knn query

```
GET /ecommerce/_search
{
  "query" : {
    "bool": {
      "must": [{
        "knn": {
          "field": "desc_embedding",
          "query_vector": [0.123, 0.244, ...]
        }
      ]
    },
    "filter": {
      "term": {
        "department": "women"
      }
    }
  }
},
"size": 10
}
```



# Architecture of Vector Search



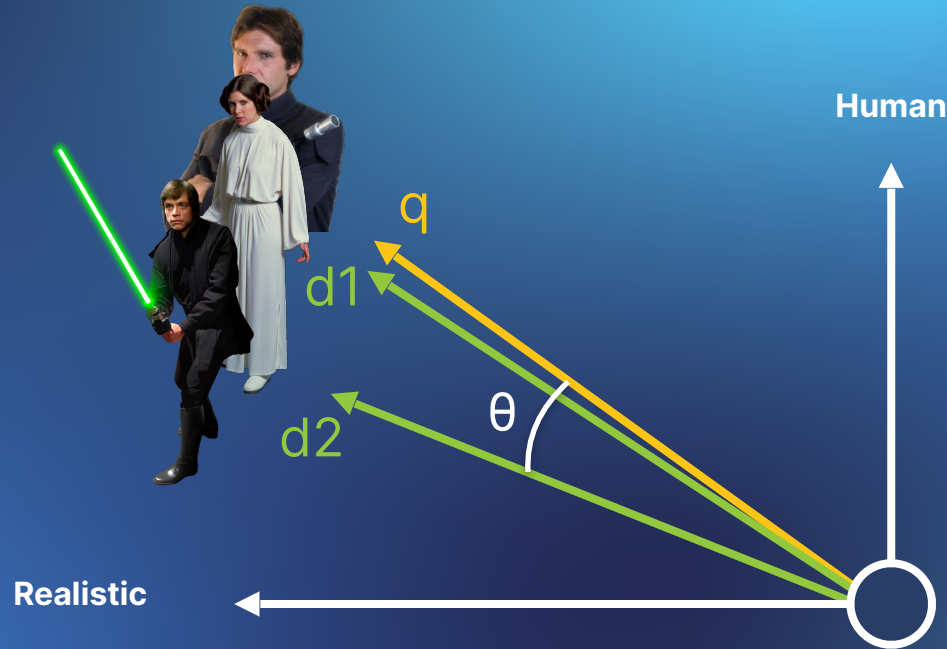
But how does it  
really work?



GROUPE ENI



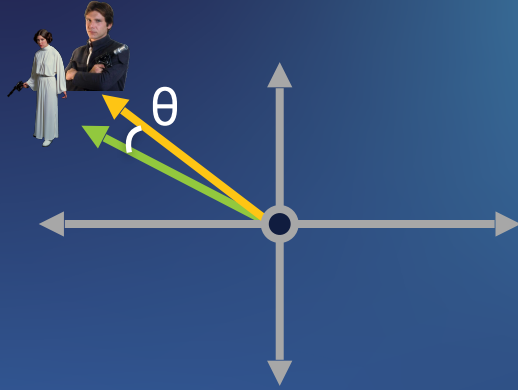
# Similarity



$$\cos(\theta) = \frac{\vec{q} \times \vec{d}}{|\vec{q}| \times |\vec{d}|}$$

$$\text{\_score} = \frac{1 + \cos(\theta)}{2}$$

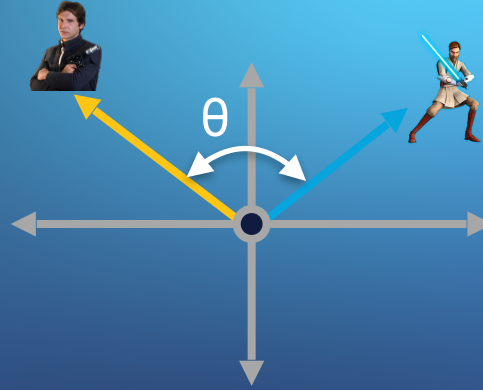
# Similarity: cosine (cosine)



**Similar vectors**

$\theta$  close to 0  
 $\cos(\theta)$  close to **1**

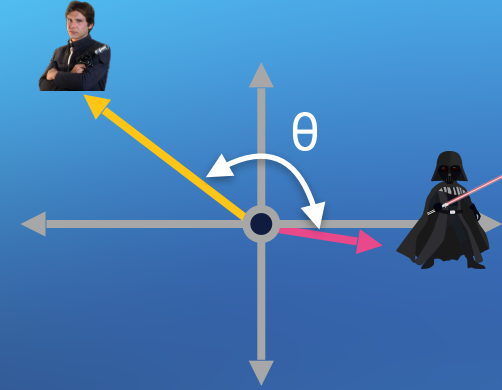
$$\text{\_score} = \frac{1 + 1}{2} = 1$$



**Orthogonal vectors**

$\theta$  close to  $90^\circ$   
 $\cos(\theta)$  close to **0**

$$\text{\_score} = \frac{1 + 0}{2} = 0.5$$



**Opposite vectors**

$\theta$  close to  $180^\circ$   
 $\cos(\theta)$  close to **-1**

$$\text{\_score} = \frac{1 - 1}{2} = 0$$



GRUPE ENI



ENI  
TECH FEST



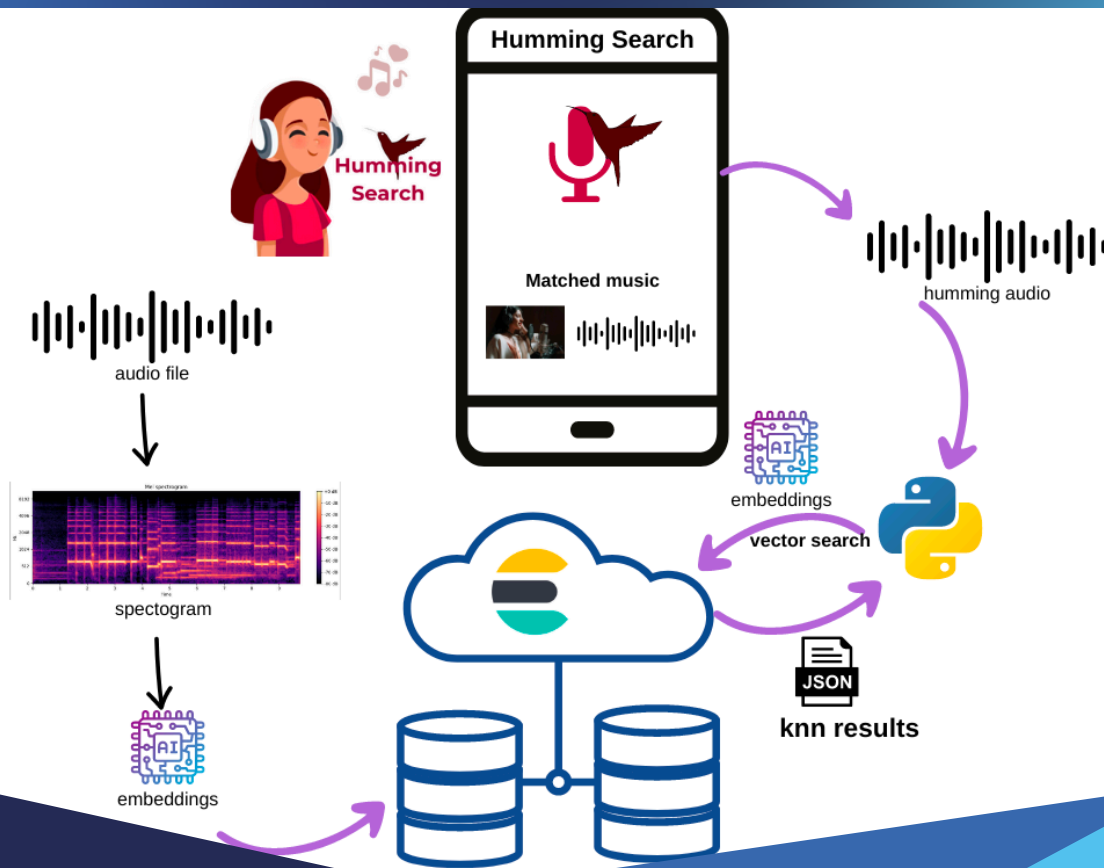
<https://djdadoo.pilato.fr/>



GRUPE ENI



<https://github.com/dadoonet/music-search/>





# ENI TECH FEST



Et si nous cherchions



**des morceaux de musique ?**



GRUPE ENI

Présenté par

**David Pilato**



@pilato.fr



@dadoonet

slides & demo

