Vector Search for Data Scientists

A Case Study with **Twitter Analytics**

Weaviate

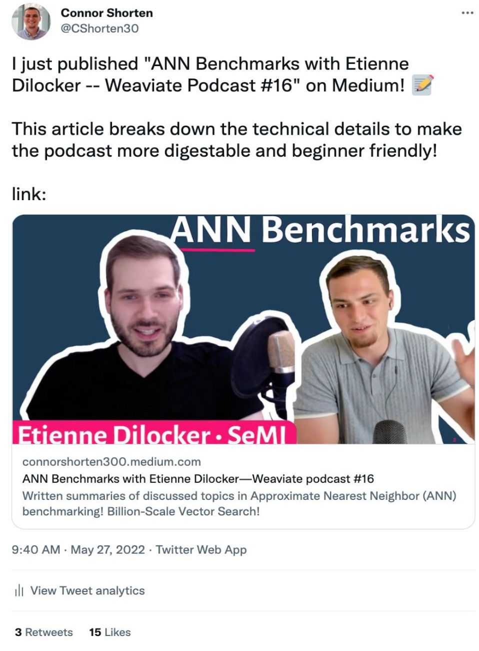
Measuring Performance

How is my data distributed?

Are there any outliers in my data?

Are my variables correlated with each other?

Twitter Analytics



Connor Shorten
@CShorten30

I just published "ANN Benchmarks with Etienne Dilocker -- Weaviate Podcast #16" on Medium! 📝

This article breaks down the technical details to make the podcast more digestable and beginner friendly!

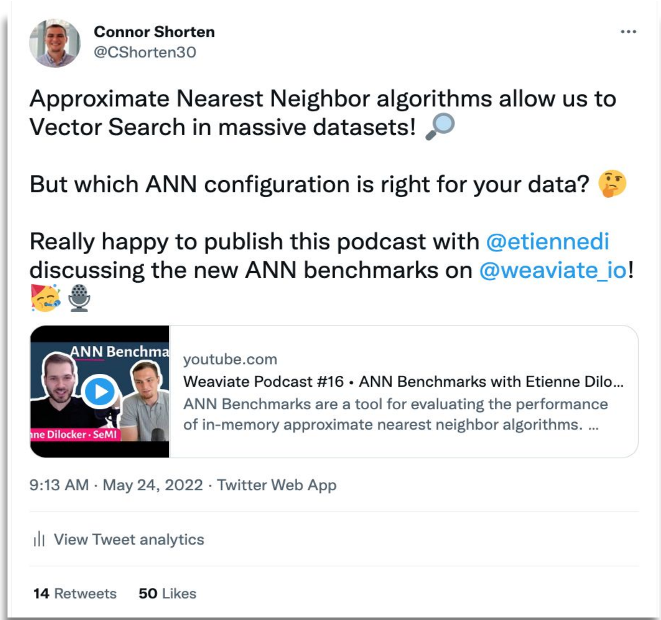
link:

connorshorten300.medium.com
ANN Benchmarks with Etienne Dilocker—Weaviate podcast #16
Written summaries of discussed topics in Approximate Nearest Neighbor (ANN) benchmarking! Billion-Scale Vector Search!

9:40 AM · May 27, 2022 · Twitter Web App

View Tweet analytics

3 Retweets **15** Likes



Connor Shorten
@CShorten30

Approximate Nearest Neighbor algorithms allow us to Vector Search in massive datasets! 🔎

But which ANN configuration is right for your data? 🤔

Really happy to publish this podcast with @etiennedi discussing the new ANN benchmarks on @weaviate_io!
🥳🎙️

youtube.com
Weaviate Podcast #16 • ANN Benchmarks with Etienne Dilo...
ANN Benchmarks are a tool for evaluating the performance of in-memory approximate nearest neighbor algorithms. ...

9:13 AM · May 24, 2022 · Twitter Web App

View Tweet analytics

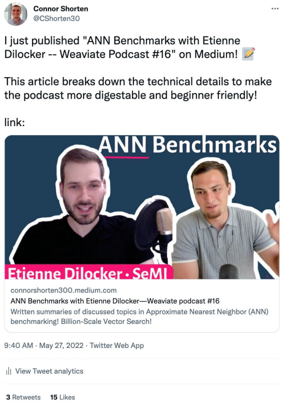
14 Retweets **50** Likes

Twitter Analytics

This presentation will utilize Twitter Analytics data to illustrate Vector Search for Data Scientists

Tweet Text	Time	Impressions	Engagements	Engagement Rate	Retweets	Replies	Likes	User Profile Clicks	Url Clicks
I just published "ANN Benchmarks with Etienne Dilcoker -- Weaviate Podcast #16 on Medium..	May 27th, 1:34pm	1905	50	2.6%	3	1	15	2	18
Approximate Nearest Neighbor algorithms allow us to Vector Search in massive datasets! ...	May 24th, 1:13pm	7182	252	3.5%	14	1	50	27	36

Feature Engineering:
Contains Emoji?
Character Count?
Word Count?
Contains "Weaviate"?



Connor Shorten
@CShorten30

I just published "ANN Benchmarks with Etienne Dilocker -- Weaviate Podcast #16" on Medium! 📝

This article breaks down the technical details to make the podcast more digestable and beginner friendly!

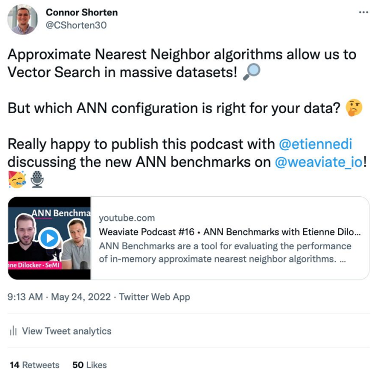
link:

connorshorten300.medium.com
ANN Benchmarks with Etienne Dilocker—Weaviate podcast #16
Written summaries of discussed topics in Approximate Nearest Neighbor (ANN) benchmarking! Billion-Scale Vector Search!

9:40 AM · May 27, 2022 · Twitter Web App

View Tweet analytics

3 Retweets **15** Likes



Connor Shorten
@CShorten30

Approximate Nearest Neighbor algorithms allow us to Vector Search in massive datasets! 🔎

But which ANN configuration is right for your data? 🤔

Really happy to publish this podcast with @etiennedi discussing the new ANN benchmarks on @weaviate_io! 🥳 🎙️

youtube.com
Weaviate Podcast #16 • ANN Benchmarks with Etienne Dilo...
ANN Benchmarks are a tool for evaluating the performance of in-memory approximate nearest neighbor algorithms. ...

9:13 AM · May 24, 2022 · Twitter Web App

View Tweet analytics

14 Retweets **50** Likes

Key Takeaways: “Vector Search for Data Scientists”

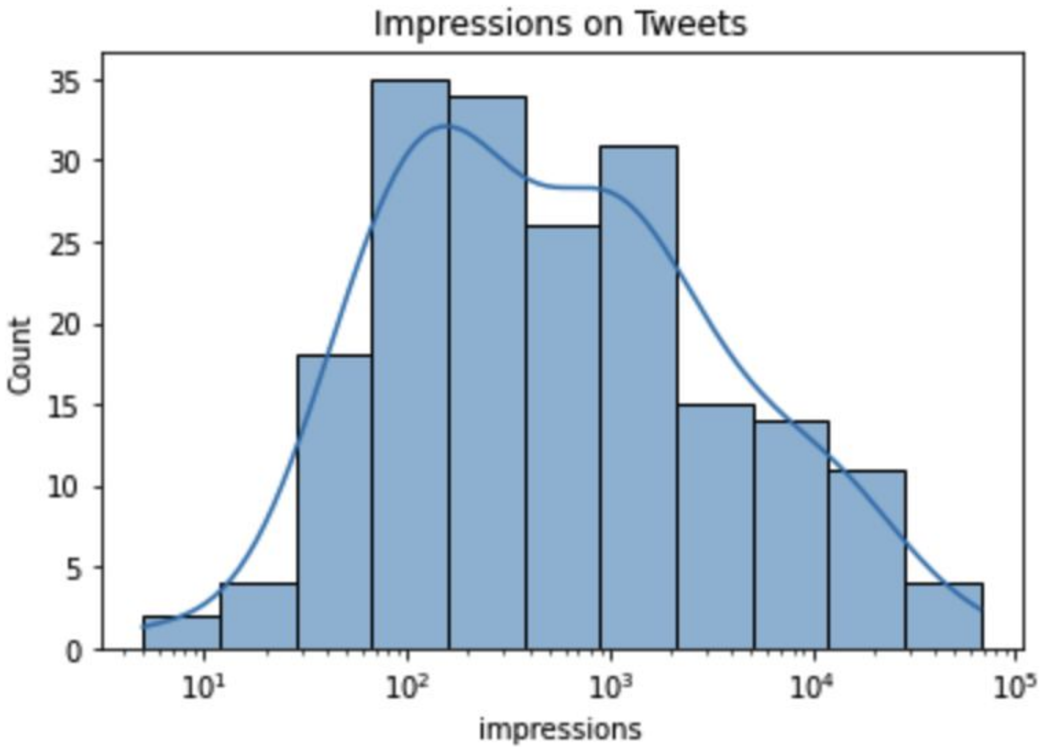
1. Segmentation in Data Science
2. Vector Representations of Unstructured Data
3. Applying Vector Search to Semantic Segmentation
4. Weaviate Example for Twitter Analytics
5. Research Questions and Discussion

Slides, Colab Notebook, Video Presentation available on:

github.com/CShorten/Vector-Search-for-Data-Scientists

Key Takeaway #1 - Segmentation in Data Science

Visualizing Distributions of Values



Key Takeaway #1 - Segmentation in Data Science

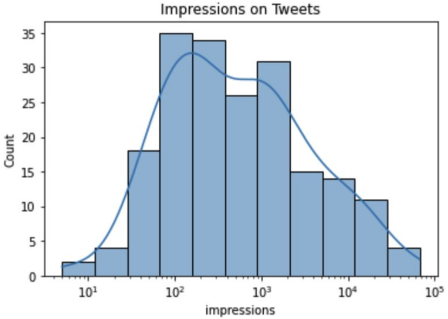
What **Time** was the Tweet sent?

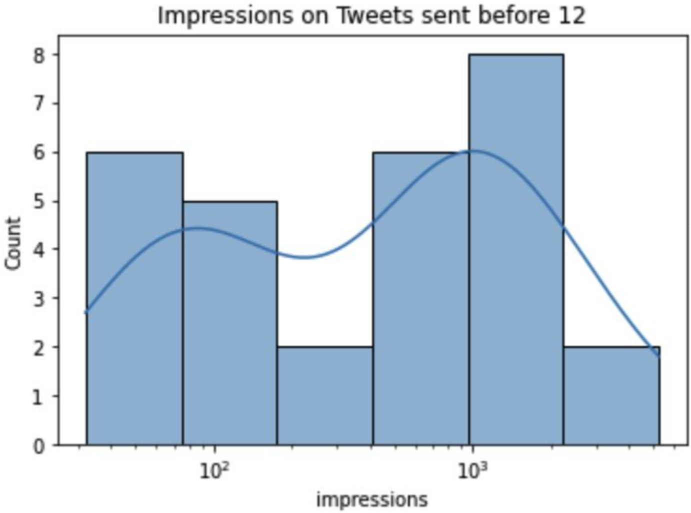
Is there a **URL Link** in the Tweet?

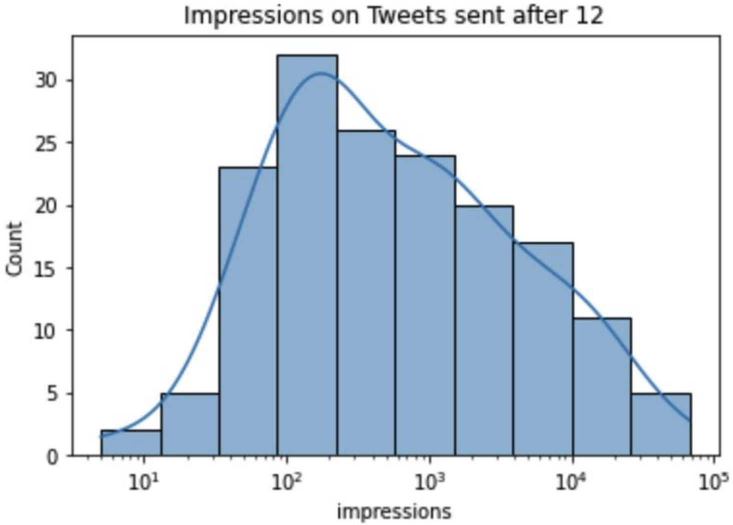
Symbolic Segmentation

Vector Segmentation

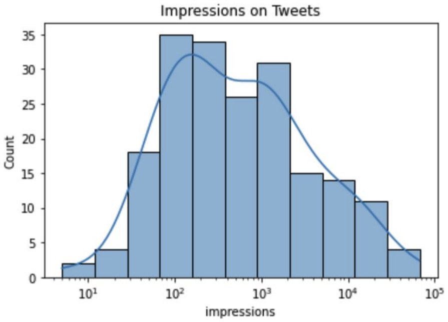
What **Time** was the Tweet sent?

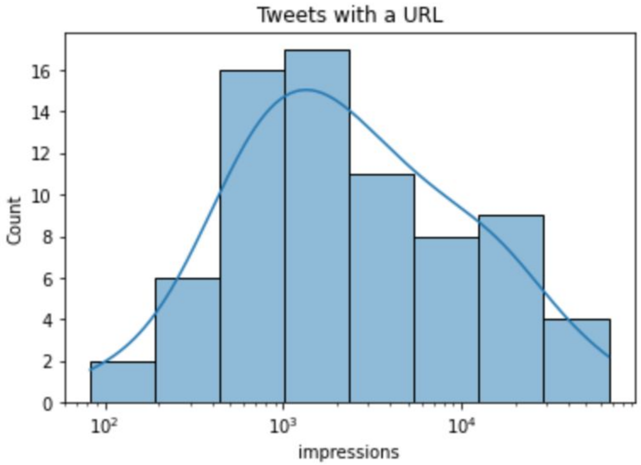


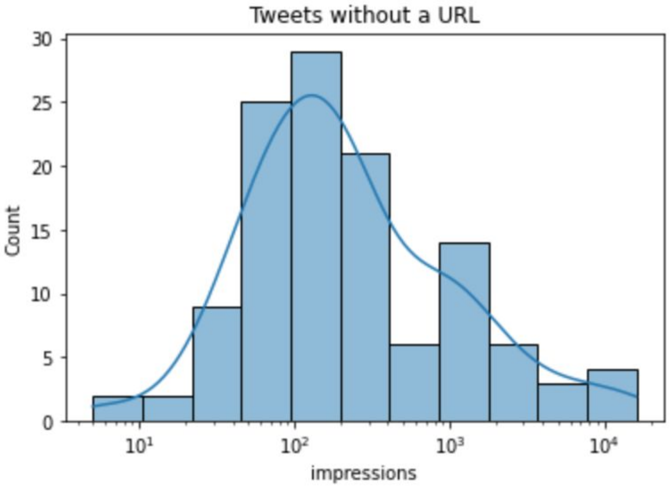




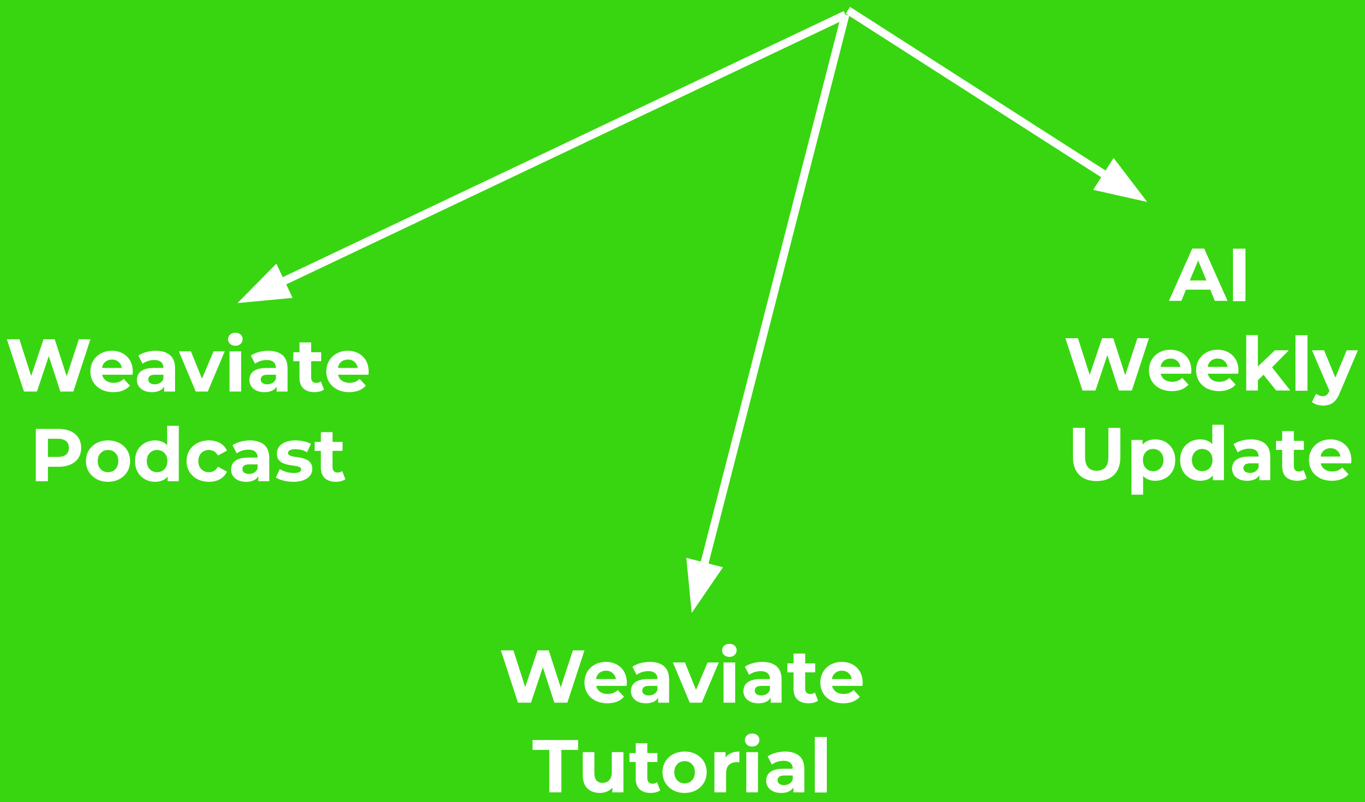
Is there a **URL Link** in the Tweet?







Can we split **Impressions** based on the **Semantics** of the content?



Weaviate Podcast

Weaviate Tutorial

AI Weekly Update

How can we segment analytics based on the semantics of…

- Text
- Images
- Code
- Audio
- Video
- Graph-Structure
- Biological Sequences
- … !

Unstructured Data

Summary of Takeaway #1
Segmentation in Data Science

We visualize the **Distribution** of our data to get a sense of it, for example we see that **Impressions** are somewhat **Normally Distributed**.

Key Takeaway #2 - Vector Representations of Unstructured Data

Vector Representations of Data

Photo by Shayna Douglas on Unsplash

0.83

0.35

..

0.02

0.74

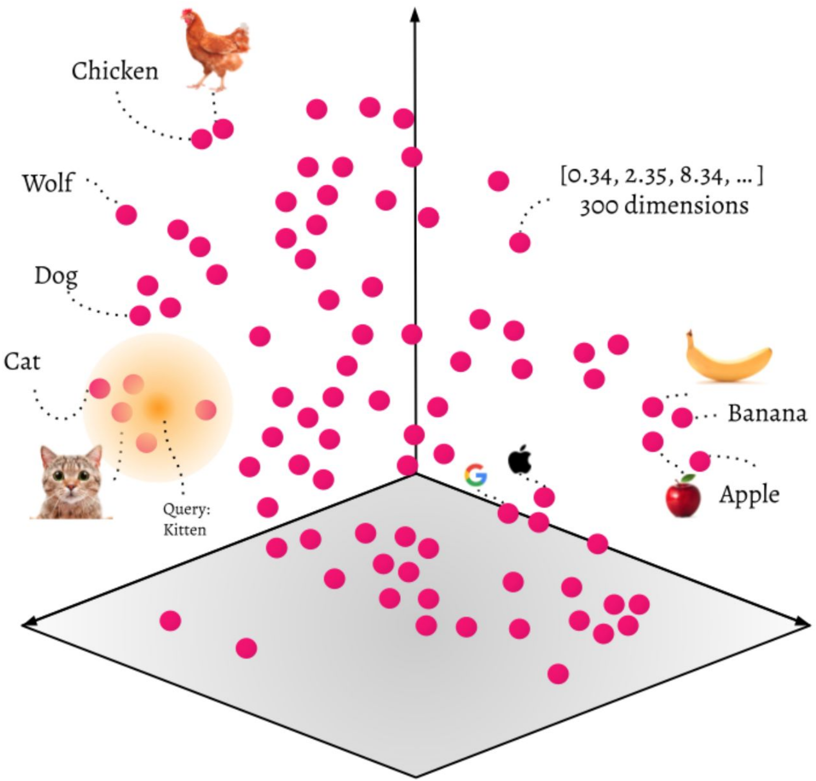
0.01

..

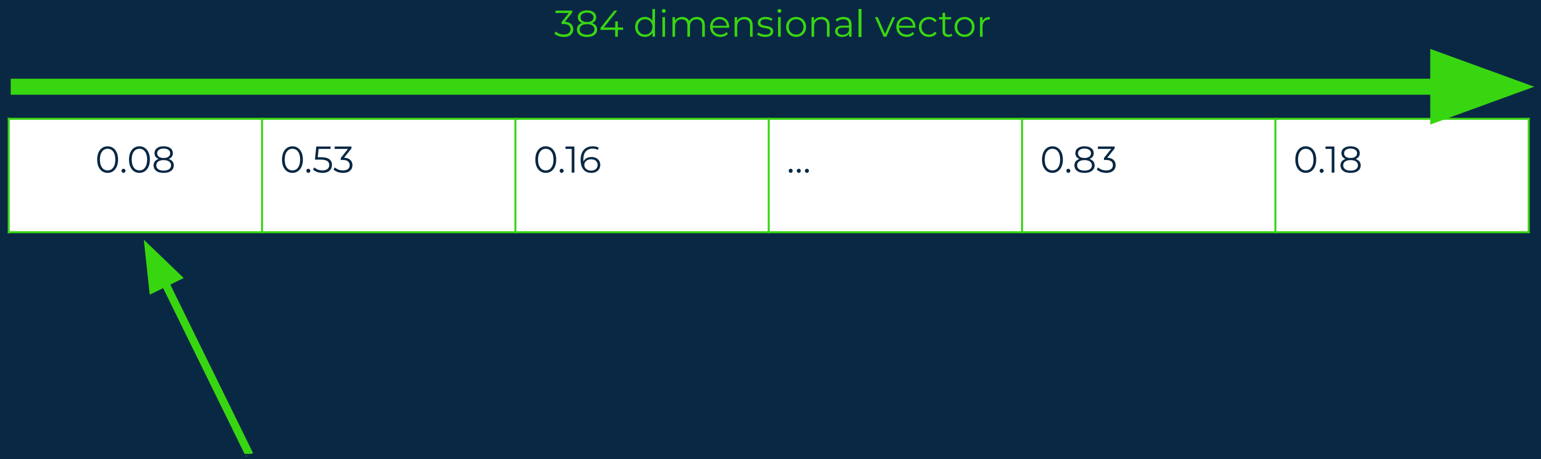
0.95

Photo by Bill Stephan on Unsplash

Capturing Semantics in Vector Representations



How do Vectors represent real-world objects?



384 dimensional vector

0.08	0.53	0.16	...	0.83	0.18

Does this represent how much of a “brand” this is?

We aren’t sure! But there are research fields such as “Multimodal Neurons” from OpenAI, and the general field of **Disentangled Representation Learning** that are making great strides in understanding this.

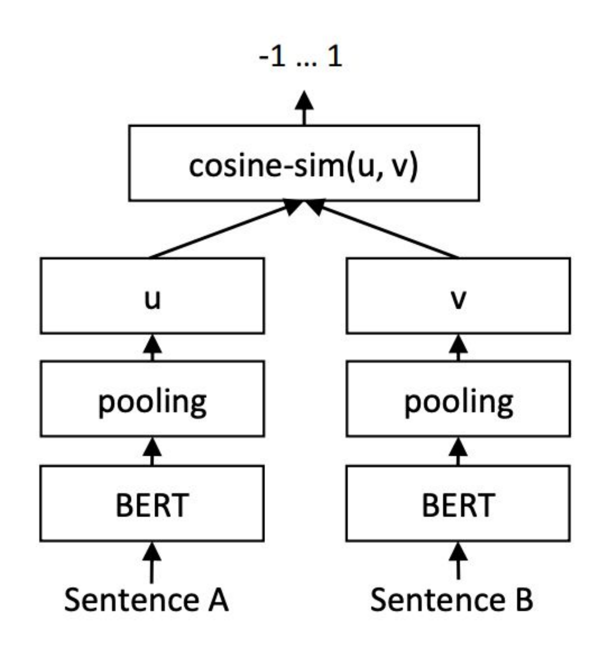
Semantic Similarity with Vector Representations

Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Authored by

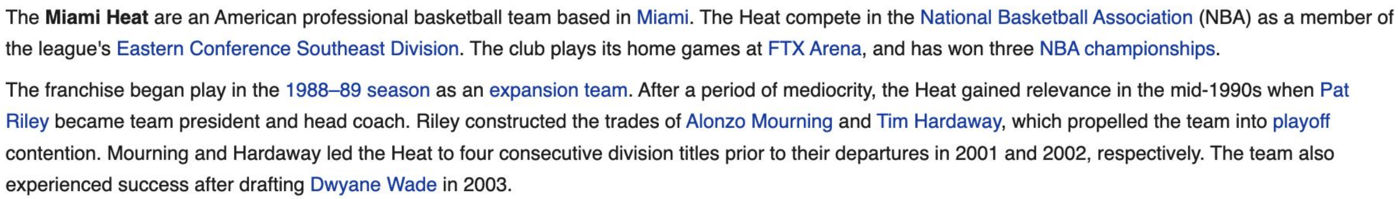
Nils Reimers and Iryna Gurevych

Published 2019



Positive and Negative Pair Sampling

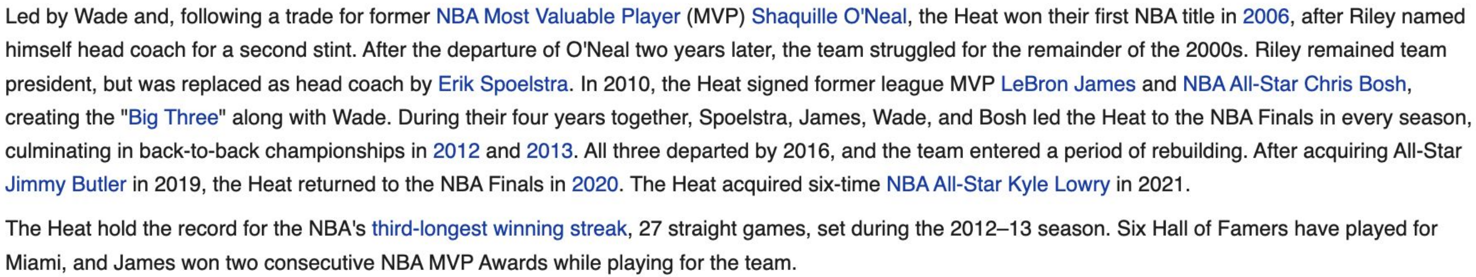
Query Point



The **Miami Heat** are an American professional basketball team based in Miami. The Heat compete in the National Basketball Association (NBA) as a member of the league's Eastern Conference Southeast Division. The club plays its home games at FTX Arena, and has won three NBA championships.

The franchise began play in the 1988–89 season as an expansion team. After a period of mediocrity, the Heat gained relevance in the mid-1990s when Pat Riley became team president and head coach. Riley constructed the trades of Alonzo Mourning and Tim Hardaway, which propelled the team into playoff contention. Mourning and Hardaway led the Heat to four consecutive division titles prior to their departures in 2001 and 2002, respectively. The team also experienced success after drafting Dwyane Wade in 2003.

Positive (Semantically Similar)

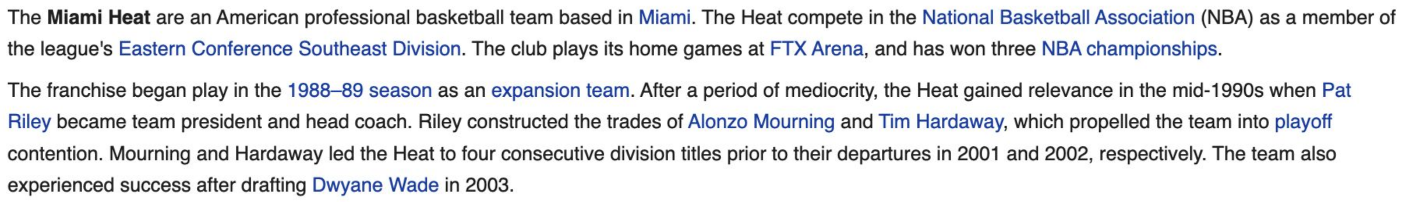


Led by Wade and, following a trade for former NBA Most Valuable Player (MVP) Shaquille O'Neal, the Heat won their first NBA title in 2006, after Riley named himself head coach for a second stint. After the departure of O'Neal two years later, the team struggled for the remainder of the 2000s. Riley remained team president, but was replaced as head coach by Erik Spoelstra. In 2010, the Heat signed former league MVP LeBron James and NBA All-Star Chris Bosh, creating the "Big Three" along with Wade. During their four years together, Spoelstra, James, Wade, and Bosh led the Heat to the NBA Finals in every season, culminating in back-to-back championships in 2012 and 2013. All three departed by 2016, and the team entered a period of rebuilding. After acquiring All-Star Jimmy Butler in 2019, the Heat returned to the NBA Finals in 2020. The Heat acquired six-time NBA All-Star Kyle Lowry in 2021.

The Heat hold the record for the NBA's third-longest winning streak, 27 straight games, set during the 2012–13 season. Six Hall of Famers have played for Miami, and James won two consecutive NBA MVP Awards while playing for the team.

Positive and Negative Pair Sampling

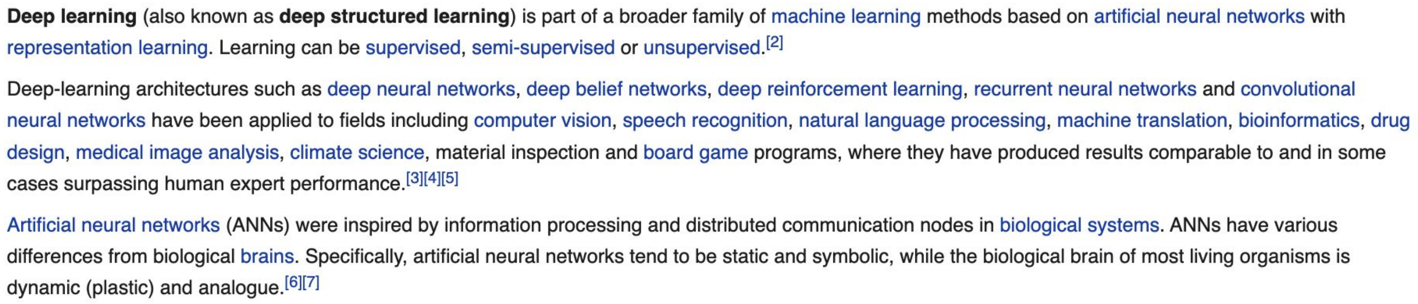
Query Point



The **Miami Heat** are an American professional basketball team based in Miami. The Heat compete in the National Basketball Association (NBA) as a member of the league's Eastern Conference Southeast Division. The club plays its home games at FTX Arena, and has won three NBA championships.

The franchise began play in the 1988–89 season as an expansion team. After a period of mediocrity, the Heat gained relevance in the mid-1990s when Pat Riley became team president and head coach. Riley constructed the trades of Alonzo Mourning and Tim Hardaway, which propelled the team into playoff contention. Mourning and Hardaway led the Heat to four consecutive division titles prior to their departures in 2001 and 2002, respectively. The team also experienced success after drafting Dwyane Wade in 2003.

Negative (Semantically Different)



Deep learning (also known as **deep structured learning**) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.[2]

Deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.[3][4][5]

Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, artificial neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analogue.[6][7]

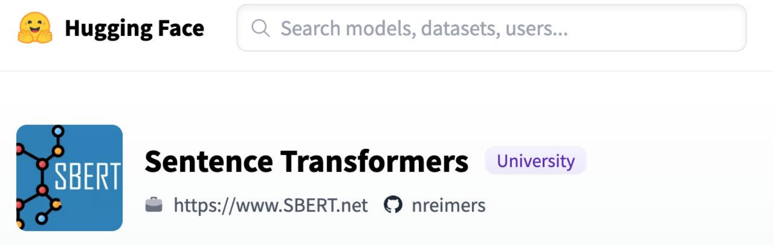
Do we need to **train** our own models?

No! There are many pre-trained models that work very well for a broad range of data!

Do we need to **train** our own models?

No! There are many pre-trained models that work very well for a broad range of data!

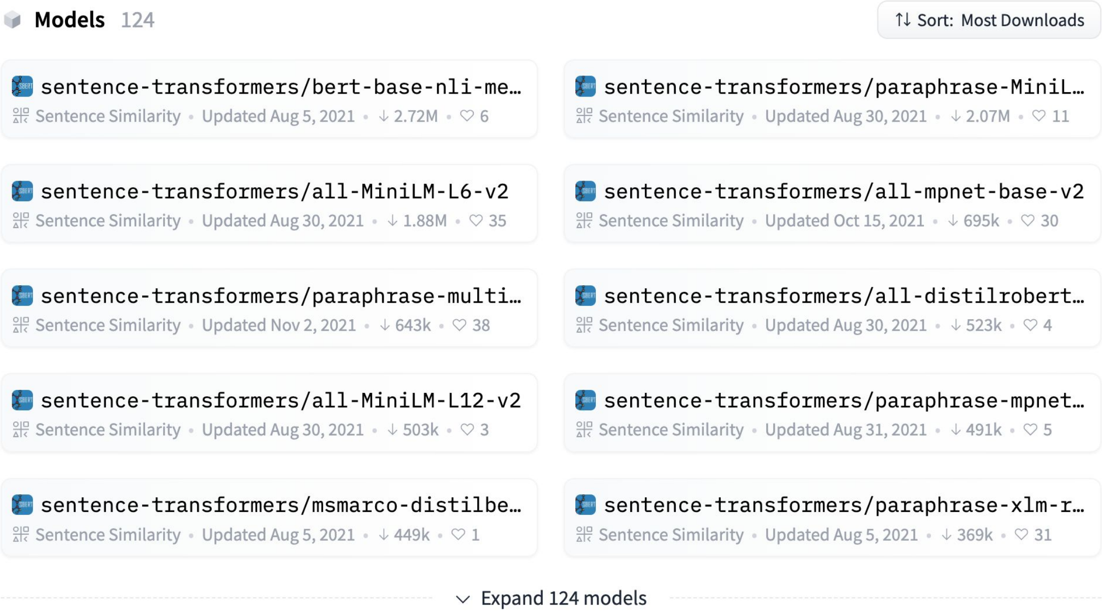
Great place to get started: **Sentence Transformers**



🤗 **Hugging Face** Search models, datasets, users...

Sentence Transformers University

https://www.SBERT.net nreimers



Models 124

Sort: Most Downloads

sentence-transformers/bert-base-nli-me…
Sentence Similarity • Updated Aug 5, 2021 • ↓ 2.72M • ♡ 6

sentence-transformers/paraphrase-MiniL…
Sentence Similarity • Updated Aug 30, 2021 • ↓ 2.07M • ♡ 11

sentence-transformers/all-MiniLM-L6-v2
Sentence Similarity • Updated Aug 30, 2021 • ↓ 1.88M • ♡ 35

sentence-transformers/all-mpnet-base-v2
Sentence Similarity • Updated Oct 15, 2021 • ↓ 695k • ♡ 30

sentence-transformers/paraphrase-multi…
Sentence Similarity • Updated Nov 2, 2021 • ↓ 643k • ♡ 38

sentence-transformers/all-distilrobert…
Sentence Similarity • Updated Aug 30, 2021 • ↓ 523k • ♡ 4

sentence-transformers/all-MiniLM-L12-v2
Sentence Similarity • Updated Aug 30, 2021 • ↓ 503k • ♡ 3

sentence-transformers/paraphrase-mpnet…
Sentence Similarity • Updated Aug 31, 2021 • ↓ 491k • ♡ 5

sentence-transformers/msmarco-distilbe…
Sentence Similarity • Updated Aug 5, 2021 • ↓ 449k • ♡ 1

sentence-transformers/paraphrase-xlm-r…
Sentence Similarity • Updated Aug 5, 2021 • ↓ 369k • ♡ 31

Expand 124 models

Summary of Takeaway #2
Vector Representations of Unstructured Data

Unstructured Data such as Images, Text, Code, … can be represented as **Vectors** with Deep Learning models.

These models are trained to **maximize semantic similarity** with **massive** collections of data.

We often do not need to train the models ourselves for **particular data domains** to reach reasonable performance.

Key Takeaway #3 - Applying Vector Search to Semantic Segmentation

Summary of Takeaway #3
Applying Vector Search for Semantic Segmentation

Vector embeddings enable an **Interface** to split analytics based on the

Semantics of the content.

Key Takeaway #4 - Weaviate for Twitter Analytics

Twitter Analytics

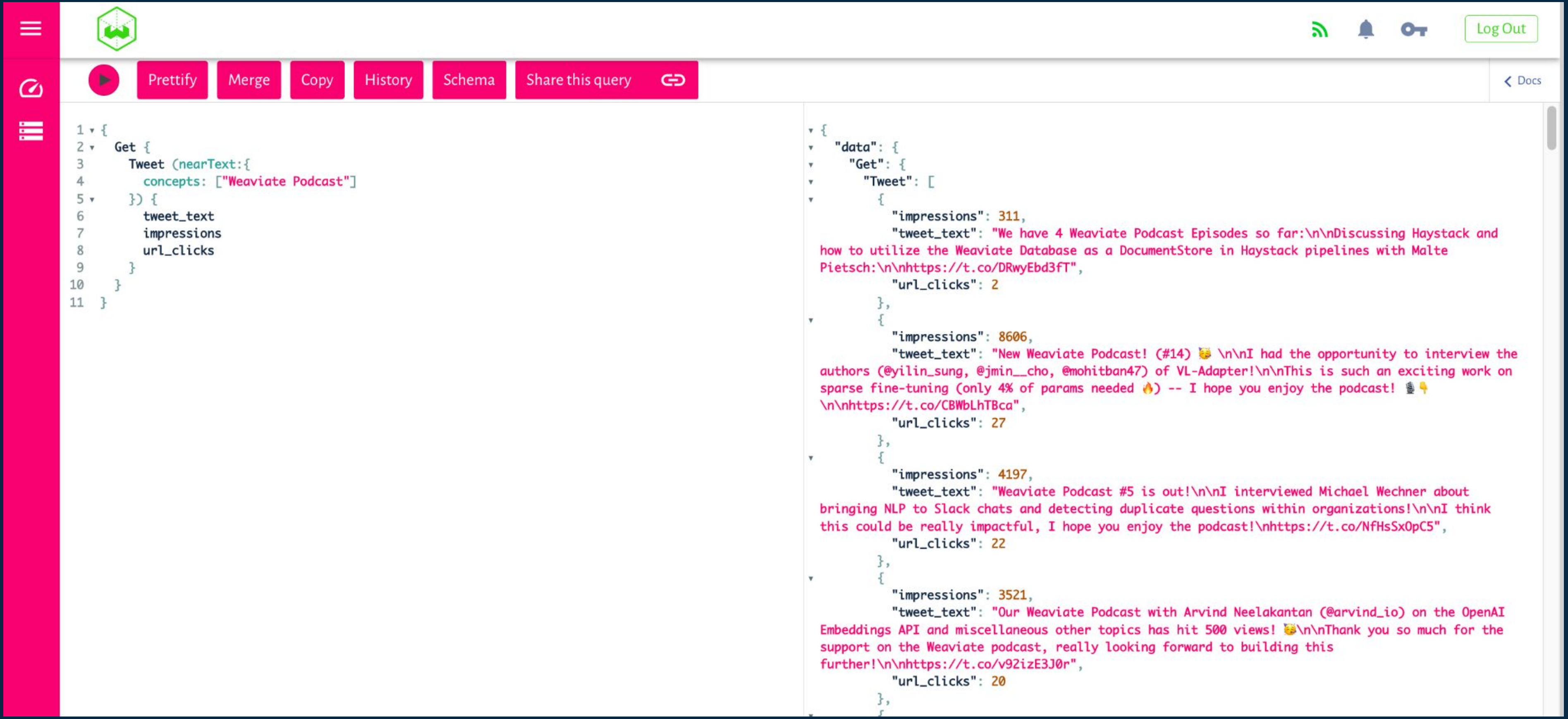
Tweet Text	Time	Impressions	Engagements	Engagement Rate	Retweets	Replies	Likes	User Profile Clicks	Url Clicks
I just published “ANN Benchmarks with Etienne Dilcoker -- Weaviate Podcast #16 on Medium..	May 27th, 1:34pm	1905	50	2.6%	3	1	15	2	18
Approximate Nearest Neighbor algorithms allow us to Vector Search in massive datasets! ...	May 24th, 1:13pm	7182	252	3.5%	14	1	50	27	36

Twitter Analytics CSV

Tweet text	time	impressions	engagements	Engagement rate	retweets	replies	likes	User profile clicks	Url clicks

Column to be vectorized with a pre-trained sentence transformer

Weaviate

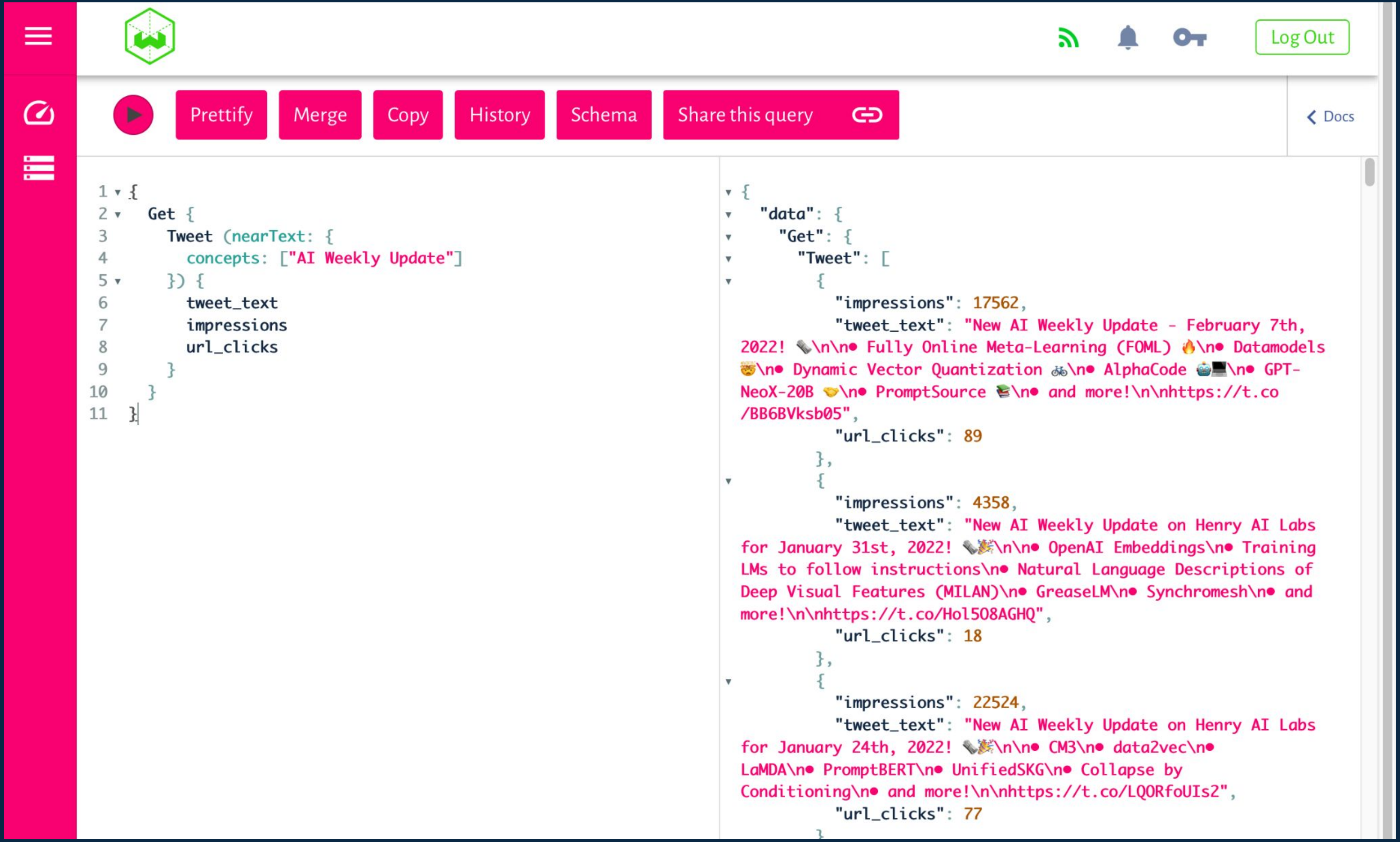


```
{
  Get {
    Tweet (nearText:{
      concepts: ["Weaviate Podcast"]
    }) {
      tweet_text
      impressions
      url_clicks
    }
  }
}
```

```
{
  "data": {
    "Get": {
      "Tweet": [
        {
          "impressions": 311,
          "tweet_text": "We have 4 Weaviate Podcast Episodes so far:\n\nDiscussing Haystack and
how to utilize the Weaviate Database as a DocumentStore in Haystack pipelines with Malte
Pietsch:\n\nhttps://t.co/DRwyEbd3fT",
          "url_clicks": 2
        },
        {
          "impressions": 8606,
          "tweet_text": "New Weaviate Podcast! (#14) 🥳 \n\nI had the opportunity to interview the
authors (@yilin_sung, @jmin__cho, @mohitban47) of VL-Adapter!\n\nThis is such an exciting work on
sparse fine-tuning (only 4% of params needed 🔥) -- I hope you enjoy the podcast! 🎙👇
\n\nhttps://t.co/CBWbLhTBca",
          "url_clicks": 27
        },
        {
          "impressions": 4197,
          "tweet_text": "Weaviate Podcast #5 is out!\n\nI interviewed Michael Wechner about
bringing NLP to Slack chats and detecting duplicate questions within organizations!\n\nI think
this could be really impactful, I hope you enjoy the podcast!\nhttps://t.co/NfHsSxOpC5",
          "url_clicks": 22
        },
        {
          "impressions": 3521,
          "tweet_text": "Our Weaviate Podcast with Arvind Neelakantan (@arvind_io) on the OpenAI
Embeddings API and miscellaneous other topics has hit 500 views! 🥳\n\nThank you so much for the
support on the Weaviate podcast, really looking forward to building this
further!\n\nhttps://t.co/v92izE3J0r",
          "url_clicks": 20
        },
        {
```

]

Text Query from “AI Weekly Update”



```
{
  Get {
    Tweet (nearText: {
      concepts: ["AI Weekly Update"]
    }) {
      tweet_text
      impressions
      url_clicks
    }
  }
}
```

```
{
  "data": {
    "Get": {
      "Tweet": [
        {
          "impressions": 17562,
          "tweet_text": "New AI Weekly Update - February 7th,
2022! 🗞\n\n● Fully Online Meta-Learning (FOML) 🔥\n● Datamodels
🤯\n● Dynamic Vector Quantization 🚲\n● AlphaCode 🤖💻\n● GPT-
NeoX-20B 🤝\n● PromptSource 📚\n● and more!\n\nhttps://t.co
/BB6BVksb05",
          "url_clicks": 89
        },
        {
          "impressions": 4358,
          "tweet_text": "New AI Weekly Update on Henry AI Labs
for January 31st, 2022! 🗞🎉\n\n● OpenAI Embeddings\n● Training
LMs to follow instructions\n● Natural Language Descriptions of
Deep Visual Features (MILAN)\n● GreaseLM\n● Synchromesh\n● and
more!\n\nhttps://t.co/Hol5O8AGHQ",
          "url_clicks": 18
        },
        {
          "impressions": 22524,
          "tweet_text": "New AI Weekly Update on Henry AI Labs
for January 24th, 2022! 🗞🎉\n\n● CM3\n● data2vec\n●
LaMDA\n● PromptBERT\n● UnifiedSKG\n● Collapse by
Conditioning\n● and more!\n\nhttps://t.co/LQORfoUIs2",
          "url_clicks": 77
        }
```

Query in Python

GET FROM CARBON!

5 Nearest Neighbors to → **“Weaviate Podcast”**

Content	Impressions
“New Weaviate Podcast! (#14) \n\nI had the opportunity to interview the authors (@yilin_sung, @jmin__cho, @mohitban47)…”	8606
“We have 4 Weaviate Podcast Episodes so far:\n\nDiscussing Haystack and how to utilize the Weaviate Database as a…”	311
“Weaviate Podcast #5 is out!\n \nI interviewed Michael Wechner about bringing NLP to Slack chats and detecting duplicate…”	4197
“New Weaviate Podcast!! \n \nI spoke with Alex Cannan about @zencastr and applying Weaviate Vector Search to Podcast…”	5104
“New Weaviate Podcast with Maximillian Werk of @JinaAI_! …”	1580

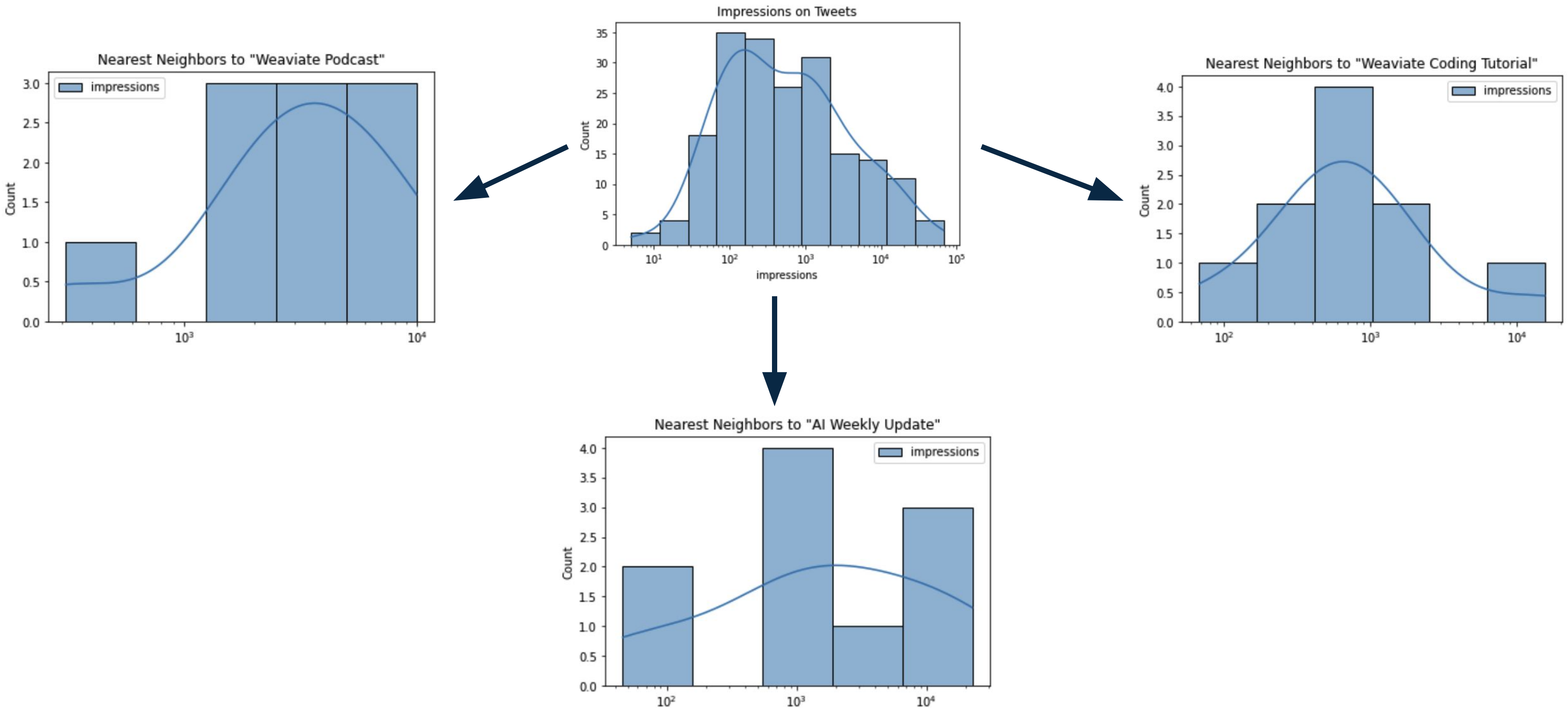
5 Nearest Neighbors to → **“Weaviate Coding Tutorial”**

Content	Impressions
“We have 4 Weaviate Podcast Episodes so far […] **how to utilize the Weaviate Database as a Document Store in Haystack pipelines … ”**	311
“We have 2 new coding tutorials on Weaviate YouTube…”	1144
“@weaviate_io Love the integration of this with the GraphQL API!”	378
“Here are some thoughts on combining Weaviate and Haystack! \n\nTLDR: Weavaite is a great Vector Search database…”	15563
“Weaviate (@weaviate_io) is also announcing a collaboration with Jina AI (@JinaAI_)! …”	586

5 Nearest Neighbors to → **“AI Weekly Update”**

Content	Impressions
“New AI Weekly Update - February 7th, 2022! \n\n Fully Online Meta-Learning (FOML) \n Datamodels \n Dynamic Vector…”	17562
“New AI Weekly Update on Henry AI Labs	4358
	22524
	86
	8491

What was the Tweet **about?**



Impressions on Tweets

Nearest Neighbors to "Weaviate Podcast"

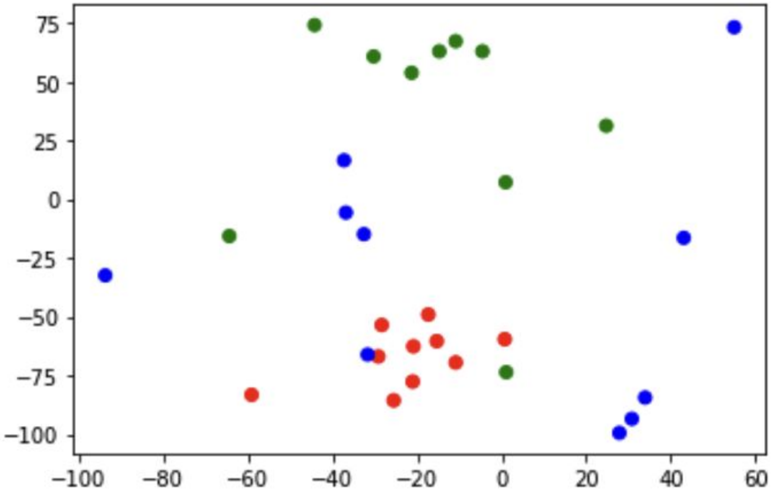
Nearest Neighbors to "Weaviate Coding Tutorial"

Nearest Neighbors to "AI Weekly Update"

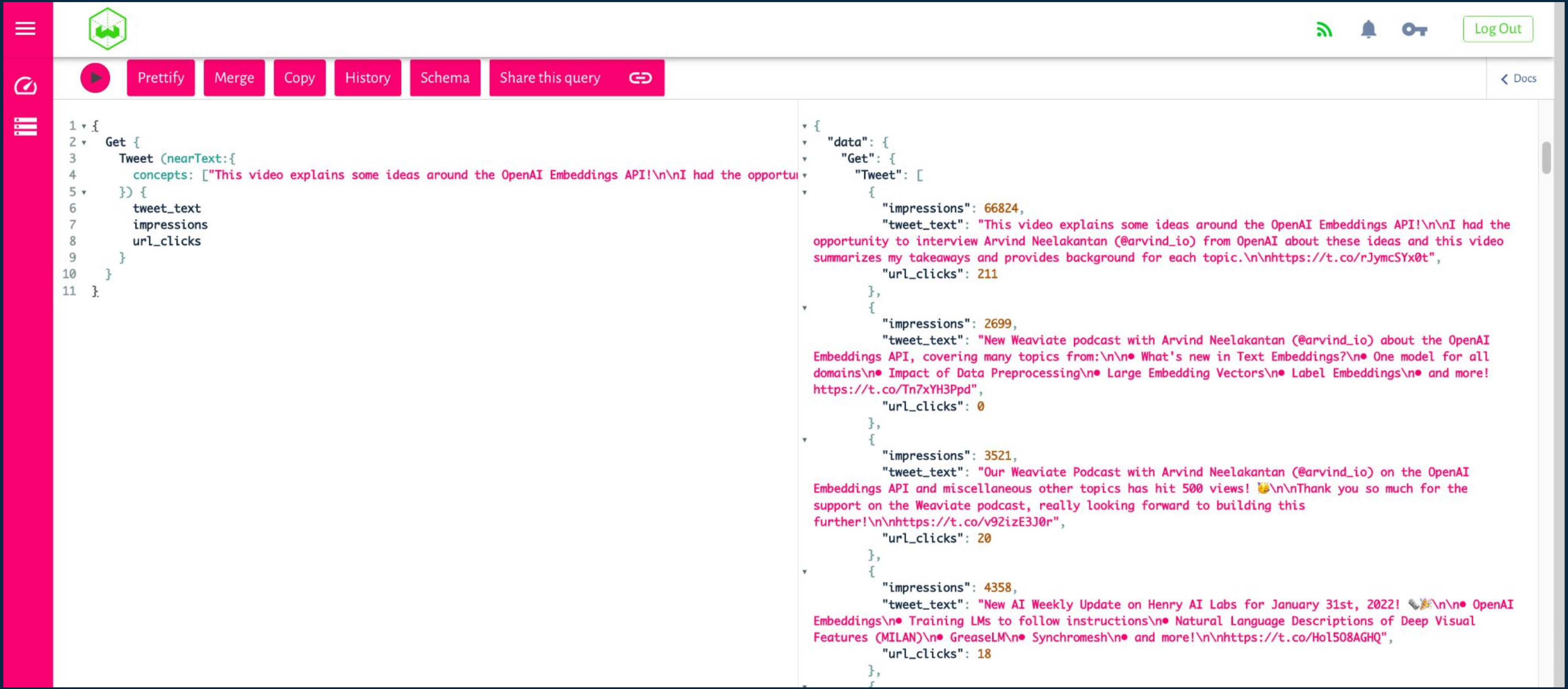
t-SNE Vector Embedding Visualization

Work in Progress - Not sure if it’s worth the trouble

This kind of idea



Have I tweeted something like this before?



```
{
  Get {
    Tweet (nearText:{
      concepts: ["This video explains some ideas around the OpenAI Embeddings API!\n\nI had the opportu
    }) {
      tweet_text
      impressions
      url_clicks
    }
  }
}
```

```
{
  "data": {
    "Get": {
      "Tweet": [
        {
          "impressions": 66824,
          "tweet_text": "This video explains some ideas around the OpenAI Embeddings API!\n\nI had the opportunity to interview Arvind Neelakantan (@arvind_io) from OpenAI about these ideas and this video summarizes my takeaways and provides background for each topic.\n\nhttps://t.co/rJymcSYx0t",
          "url_clicks": 211
        },
        {
          "impressions": 2699,
          "tweet_text": "New Weaviate podcast with Arvind Neelakantan (@arvind_io) about the OpenAI Embeddings API, covering many topics from:\n\n• What's new in Text Embeddings?\n• One model for all domains\n• Impact of Data Preprocessing\n• Large Embedding Vectors\n• Label Embeddings\n• and more! https://t.co/Tn7xYH3Ppd",
          "url_clicks": 0
        },
        {
          "impressions": 3521,
          "tweet_text": "Our Weaviate Podcast with Arvind Neelakantan (@arvind_io) on the OpenAI Embeddings API and miscellaneous other topics has hit 500 views! 🥳\n\nThank you so much for the support on the Weaviate podcast, really looking forward to building this further!\n\nhttps://t.co/v92izE3J0r",
          "url_clicks": 20
        },
        {
          "impressions": 4358,
          "tweet_text": "New AI Weekly Update on Henry AI Labs for January 31st, 2022! 🗞️🎉\n\n• OpenAI Embeddings\n• Training LMs to follow instructions\n• Natural Language Descriptions of Deep Visual Features (MILAN)\n• GreaseLM\n• Synchromesh\n• and more!\n\nhttps://t.co/Hol5O8AGHQ",
          "url_clicks": 18
        },
        {
```

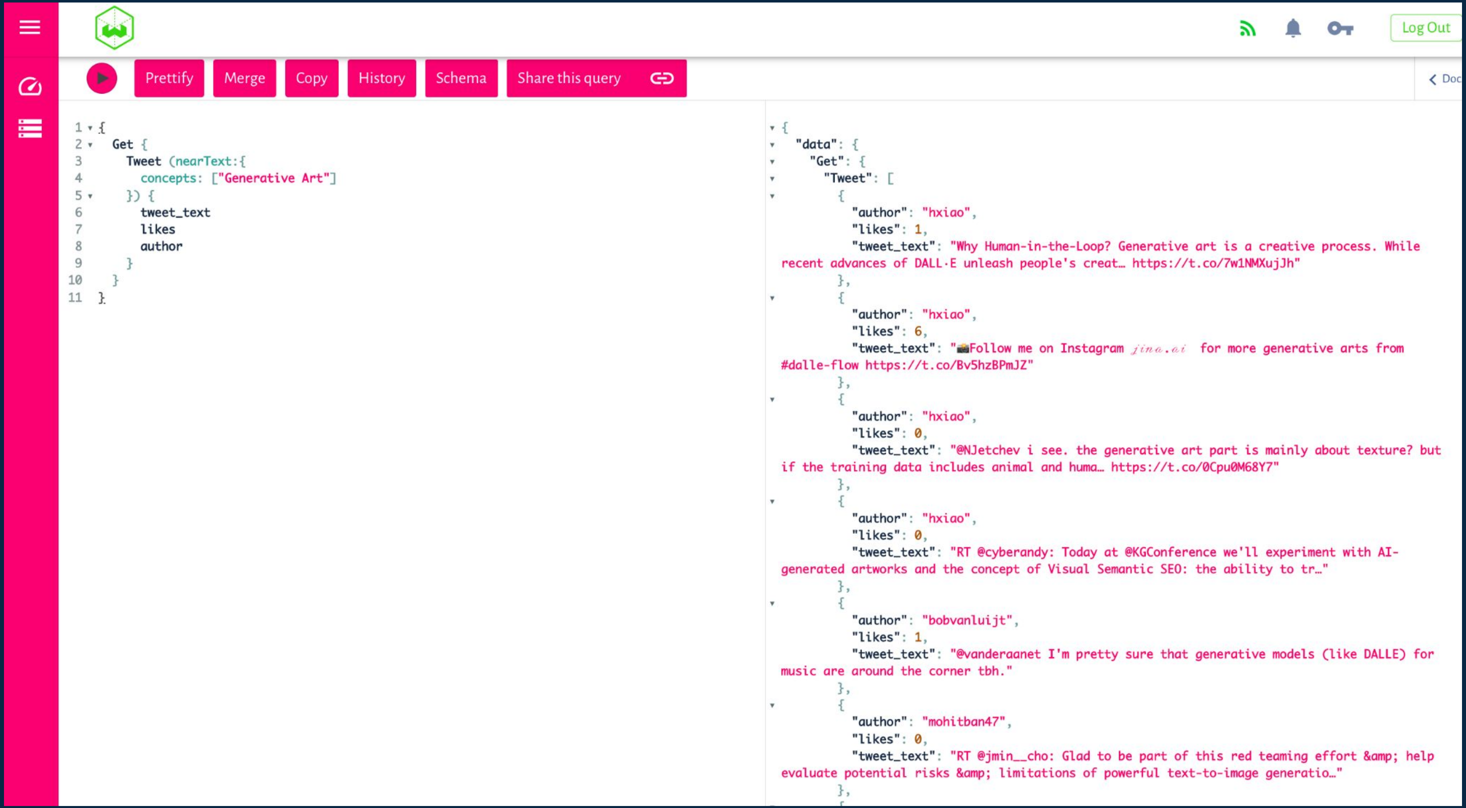
Have any **Weaviate Podcast guests** tweeted something like this recently?

Weaviate
podcast

Twitter API Request

Tweet, Author, Likes → Weaviate

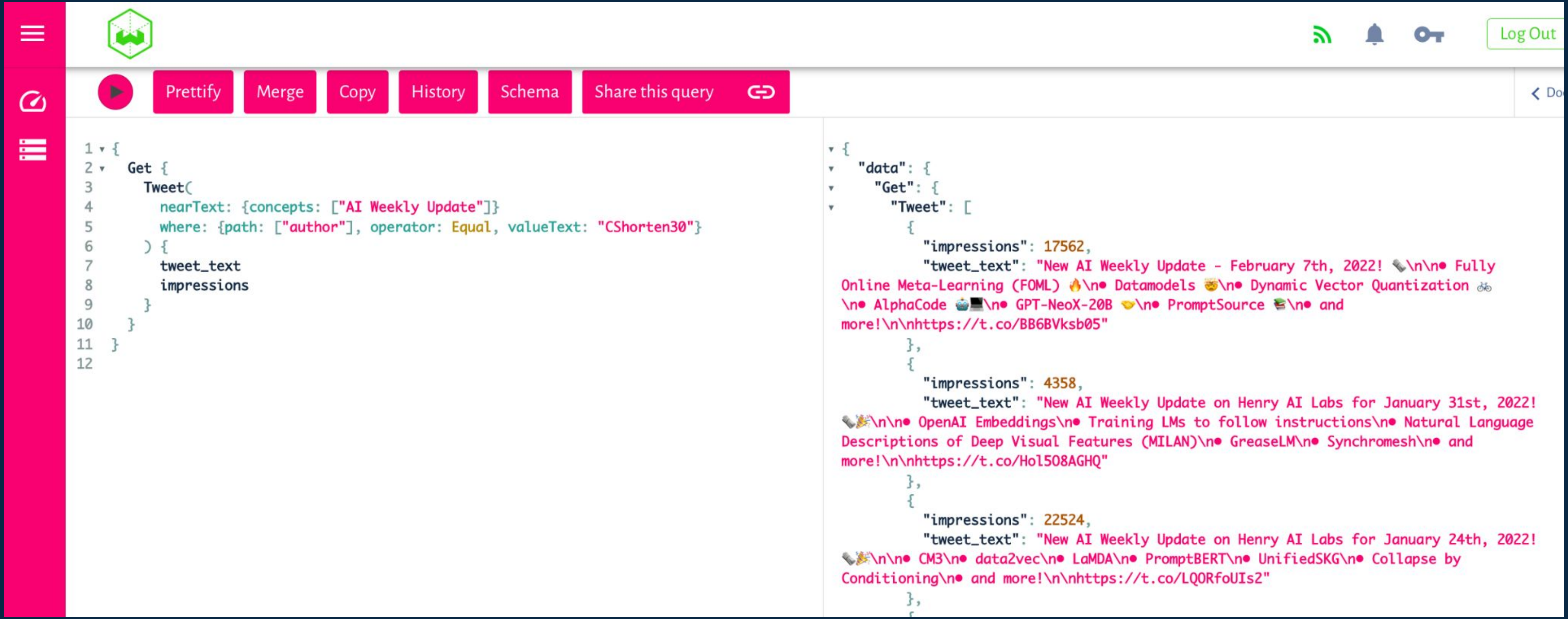
Is anyone tweeting about **Generative Art?**



```
{
  Get {
    Tweet (nearText:{
      concepts: ["Generative Art"]
    }) {
      tweet_text
      likes
      author
    }
  }
}
```

```
{
  "data": {
    "Get": {
      "Tweet": [
        {
          "author": "hxiao",
          "likes": 1,
          "tweet_text": "Why Human-in-the-Loop? Generative art is a creative process. While
recent advances of DALL·E unleash people's creat… https://t.co/7w1NMXujJh"
        },
        {
          "author": "hxiao",
          "likes": 6,
          "tweet_text": "📸Follow me on Instagram 𝑗𝑖𝑛𝑎.𝑎𝑖  for more generative arts from
#dalle-flow https://t.co/Bv5hzBPmJZ"
        },
        {
          "author": "hxiao",
          "likes": 0,
          "tweet_text": "@NJetchev i see. the generative art part is mainly about texture? but
if the training data includes animal and huma… https://t.co/0Cpu0M68Y7"
        },
        {
          "author": "hxiao",
          "likes": 0,
          "tweet_text": "RT @cyberandy: Today at @KGConference we'll experiment with AI-
generated artworks and the concept of Visual Semantic SEO: the ability to tr…"
        },
        {
          "author": "bobvanluijt",
          "likes": 1,
          "tweet_text": "@vanderaanet I'm pretty sure that generative models (like DALLE) for
music are around the corner tbh."
        },
        {
          "author": "mohitban47",
          "likes": 0,
          "tweet_text": "RT @jmin__cho: Glad to be part of this red teaming effort &amp; help
evaluate potential risks &amp; limitations of powerful text-to-image generatio…"
        },
```

Filtering Semantic Searches with Symbolic Attributes



```
{
  Get {
    Tweet(
      nearText: {concepts: ["AI Weekly Update"]}
      where: {path: ["author"], operator: Equal, valueText: "CShorten30"}
    ) {
      tweet_text
      impressions
    }
  }
}
```

```
{
  "data": {
    "Get": {
      "Tweet": [
        {
          "impressions": 17562,
          "tweet_text": "New AI Weekly Update - February 7th, 2022! 🗞️\n\n• Fully
Online Meta-Learning (FOML) 🔥\n• Datamodels 🤯\n• Dynamic Vector Quantization 🚲
\n• AlphaCode 🤖💻\n• GPT-NeoX-20B 🤝\n• PromptSource 📚\n• and
more!\n\nhttps://t.co/BB6BVksb05"
        },
        {
          "impressions": 4358,
          "tweet_text": "New AI Weekly Update on Henry AI Labs for January 31st, 2022!
🗞️🎉\n\n• OpenAI Embeddings\n• Training LMs to follow instructions\n• Natural Language
Descriptions of Deep Visual Features (MILAN)\n• GreaseLM\n• Synchromesh\n• and
more!\n\nhttps://t.co/Hol5O8AGHQ"
        },
        {
          "impressions": 22524,
          "tweet_text": "New AI Weekly Update on Henry AI Labs for January 24th, 2022!
🗞️🎉\n\n• CM3\n• data2vec\n• LaMDA\n• PromptBERT\n• UnifiedSKG\n• Collapse by
Conditioning\n• and more!\n\nhttps://t.co/LQORfoUIs2"
        },
```

Technical Details of how this is setup

Pandas DataFrame → Weaviate

A look under the hood of client.from_pandas

Hosting Weaviate

- Weaviate Cloud Service!
- Localhost / Cloud DIY setup with Docker-Compose

Weaviate Schema Setup

```
Weaviate_schema = {
    “classes”: [{
                “class”: “Tweet”,
                “description”: “Tweet Analytics”,
                “properties”: [{
                        “name”: “tweet_text”,
                        “dataType”: [“text”],
                        “description”: “The text in the Tweet.”,
                        “moduleConfig”: {
                            “Text2vec-transformers”: { “skip”: False”, “vectorizePropertyName”: False }
                        }
                      },
                        …
```

Batch upload

```
Def add_tweet(batch: Batch, data: dict) -> str:
    Tweet_object = {
        “Tweet_text”: data[“tweet_text”],
        “Hour”: data[“hour”],
        …
    }
    batch.add_data_object(
        Data_object = tweet_object,
        Class_name = “Tweet”,
        Uuid = tweet_id
    )
```

Weaviate

- Weaviate is a Vector Search **Database**, rather than a **Library** such as Facebook's FAISS or Spotify's ANNOY

- Weaviate has a **Graph-like Data Model**

Summary of Takeaway #4
Vector Representations of Unstructured Data

Weaviate is a **Vector Search Database** that can be used to store and search through semantic embeddings of unstructured data.

Key Takeaway #5 - Research Questions and Discussion

Research Questions and Discussion

- Should I fine-tune my embedding model?
- Large-Scale Vector Search with Approximate Nearest Neighbor (ANN) Algorithms
- How does Vector Search differ from Classification or Regression models?

What do we want to know about our Tweets?

Should I post this?

When might be a better time to post it?

What might be a better phrasing of this tweet?

Expanding from individuals to teams

- Has anyone on my team tweeted something like this recently?
- Who on our team would be best fit to tell this story?
- What topics should we be tweeting about?

Summary of Takeaway #5

Research Questions and Discussion

How can we improve these systems?

How are systems like this changing our world?

Key Takeaways: “Vector Search for Data Scientists”

1. Segmentation in Data Science
2. Vector Representations of Unstructured Data
3. Applying Vector Search to Semantic Segmentation
4. Weaviate Example for Twitter Analytics
5. Research Questions and Discussion

Slides, Colab Notebook, Video Presentation available on:

github.com/CShorten/Vector-Search-for-Data-Scientists

Thank you for Watching!

Special thanks to **Sebastian Witalec** in advising the development of this presentation