

## **Semantic Search Over Text**

Yujian Tang

## **Speaker**

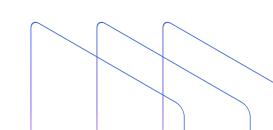


**Yujian Tang** 

Developer Advocate, Zilliz

yujian@zilliz.com
https://www.linkedin.com/in/yujiantang
https://www.twitter.com/yujian\_tang





## Company





@Zilliz\_Universe



linkedin.com/in/zilliz

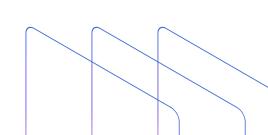


milvusio.slack.com



github.com/milvus-io/milvus

zilliz.com



**01 What is Semantic Search?** 

**What is a Vector Embedding?** 

03 How Can You Do Semantic Text Search?

## 01

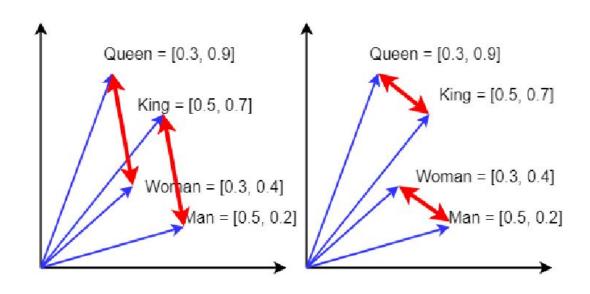
## What is Semantic Search?





## **Semantic Similarity**

Some words are similar



#### **Semantic Search**

```
rangle acids protein

gwh gel nm insulin reactive molecules
isotopes femperature

isotopes femperature

isotopes femperature

clatp indo dialects fuel

sp est pi alpha languages surface

energy

pmid fried genus | m km | y disk energy

pmid fried genus | m km | y java population functions

bww mi chisland affica language are data

ko puerto sweden | n total states can process

violin | du | ice C | othy than is use

hockey | san d th b | war | or that

footballer | san d th b | war | or that

footballer | von | w sing t us an of the approach

earl | born | w sing t us an of the approach

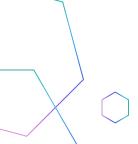
gibson | carl cbs st | father late he in actress | ed | york | nine | his

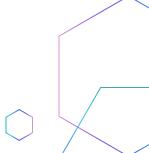
lisa actor | scott james | big | s show

phil gary | tom | paul john | smith | chris | frank | or footballer | smith | movie | smith | chris | frank | or footballer | smith | movie | smith | chris | frank | or footballer | smith | movie | smith | chris | frank | or footballer | smith | chris | frank | or footballer | smith | movie | smith | movie | smith | chris | footballer | smith | movie | smith | movie | smith | chris | footballer | smith | movie | smith | movie | smith | movie | smith | chris | footballer | smith | movie | smith 
                                                                                                                                                   isotopes temperature
                                                                                                                                                                                                                                                                                                                                                                                                                                     featuring
```

# 02

## What is a Vector Embedding?





## **Vector Embedding**

4218

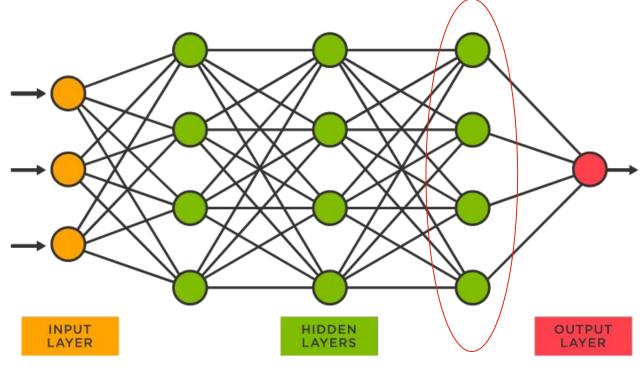
Default project > asdf > medium\_articles ■ medium\_articles LOADED ☐ Connection Guide Actions V Collection Details Schema **Data Preview** Vector Search id title vector title link 125 [0.014838364, -0.017620698, 0.03955149... The dawn of Dark Mode https://uxdes https://media 195 [-0.008266705, -0.009836753,... Scrollsuming - The New Zombie Interaction 1464 [0.036901046, 0.00553058, 0.011379256,... Make Issue Types Great Again https://medic 1538 [0.013065741, 0.011942832, -0.00739964... Creating an exam archive system with a Data... https://towar 1623 [-0.0024157332, -0.007975485,... Using a Golang pattern to write better... https://media 2488 [0.067513265, 0.015994877,... Make Your Data Models Into Websites https://towar 2926 [-0.009442576, 0.0028793914,... Create A Machine Learning Model With Goog... https://towar 3480 [0.048443407, -0.013659755,... What If Only Batch Normalization Layers Wer... https://towar What Zynn's entrance mean for the North... 3573 [0.02339675, -0.0032612802, 0.01034519... https://medic

[0.054374292, 0.011657661, 0.04338219,...

Vibing Out TensorFlow

https://towar

How are these generated?



## How are these generated?

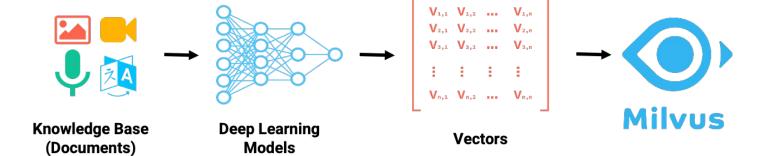


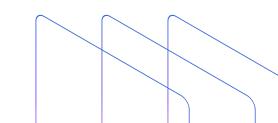




# How Can You Do Semantic Text Search?

#### **Put Your Data into a Vector Database**





## **Steps**

- 1. Download and Import the Data
- 2. Explore and Clean the Data
- 3. Spin up the Vector Database
- 4. Get Embeddings and Populate the Database
- 5. Query the Vector Database



### **THANK YOU FOR LISTENING**







