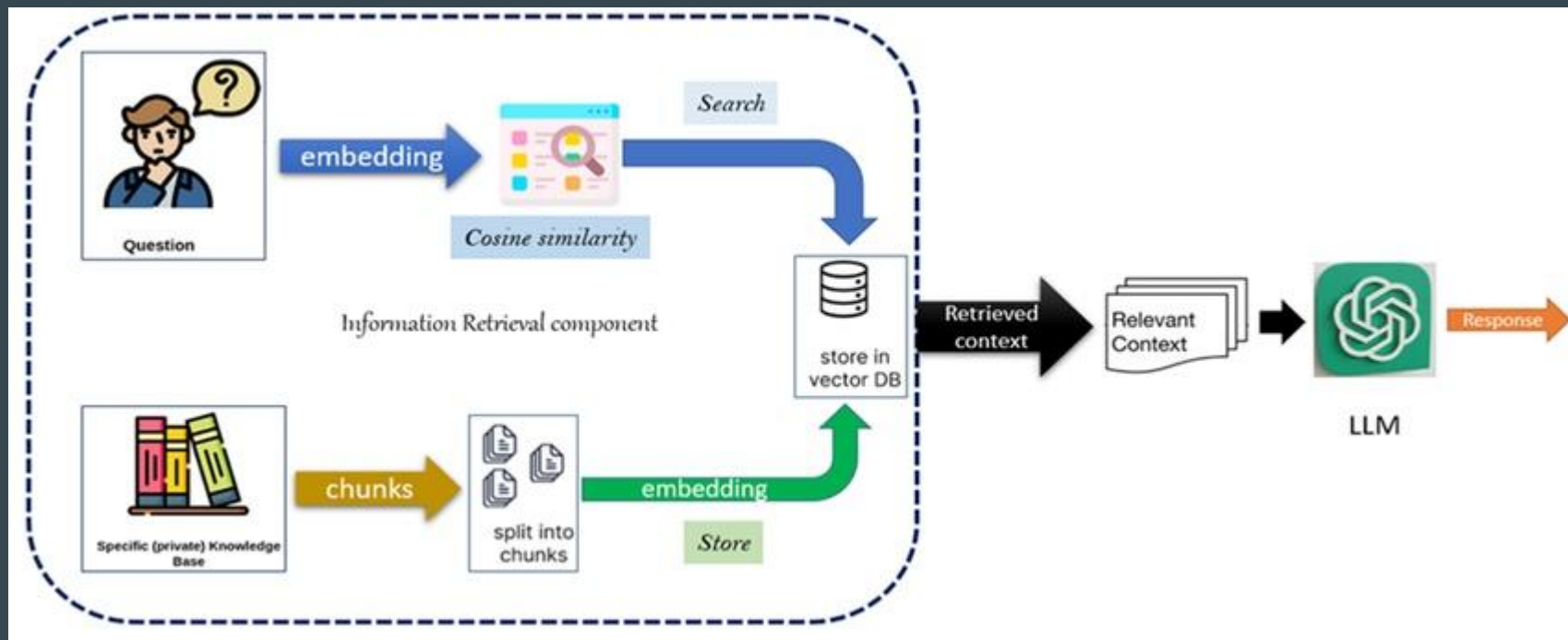


Enhancing RAG with Hybrid Search

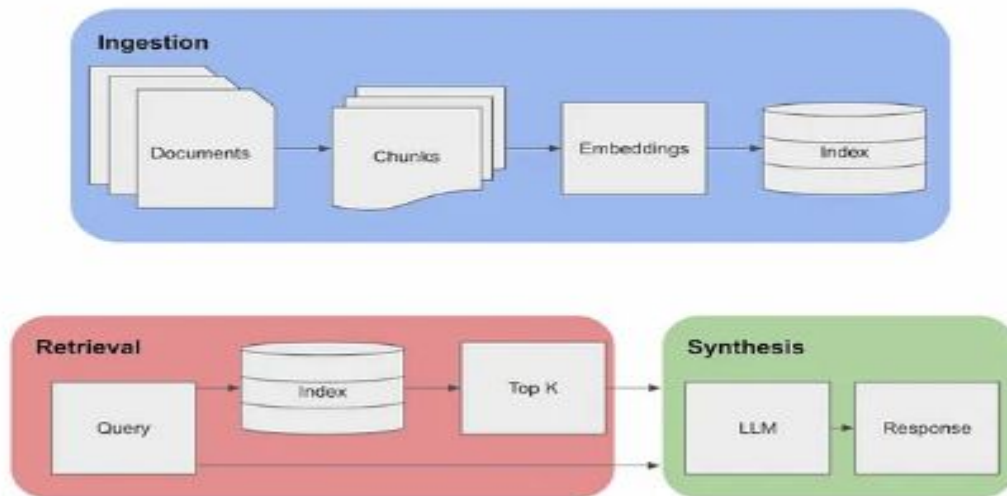
- V G Samvardhan

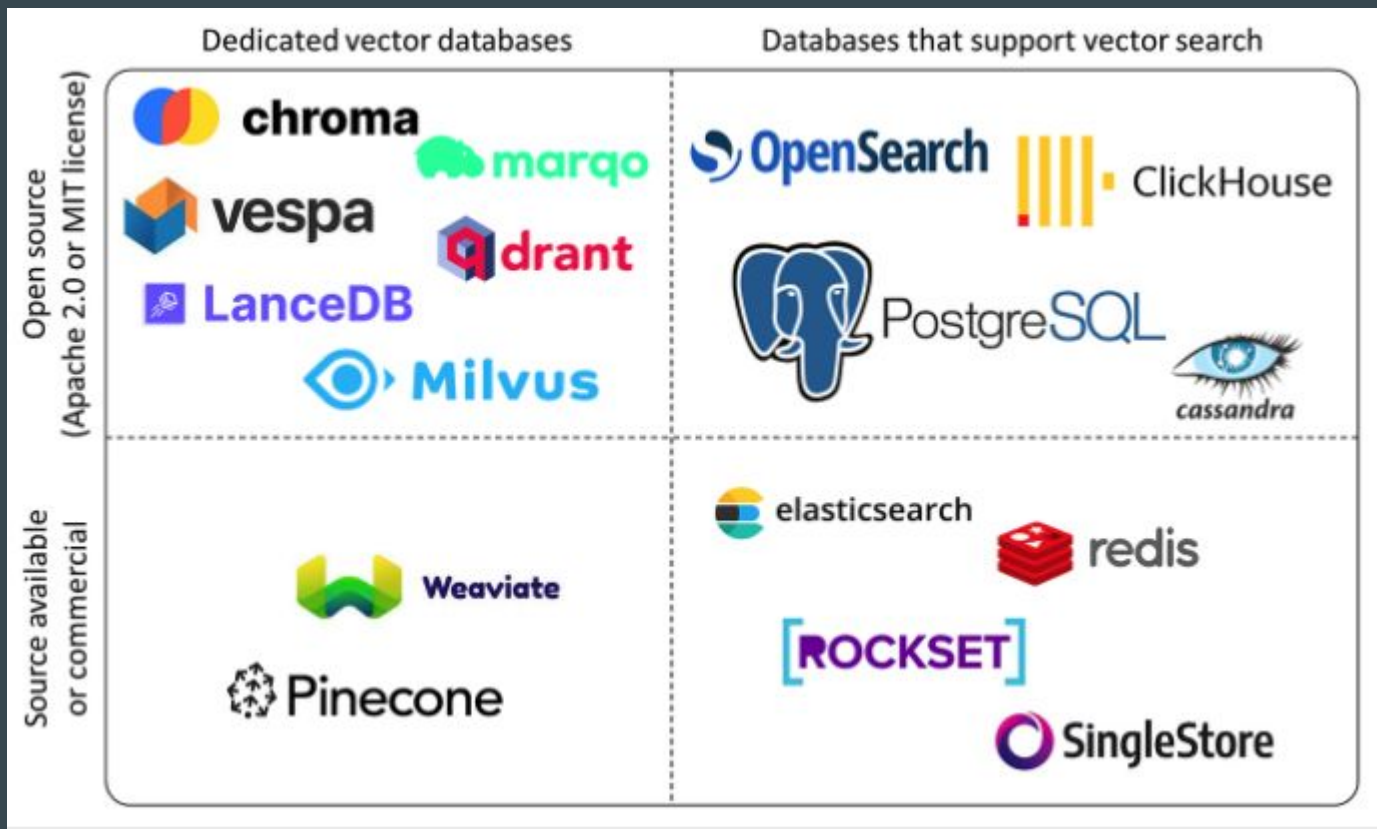
Retrieval-Augmented Generation (RAG)



RAG Pipeline

Basic RAG Pipeline





Types of Search

Keyword-based search

The screenshot shows a web browser at femi.case.edu:3002 displaying the NetWellness website. A search bar contains the text 'stress stroke' and a 'Search Questions' button. Below the search bar, it says '14 questions found'. The first result is '23918 Health Heart Diet' with a question about a diet for arterial deposits. The second result, '83614 Prednisone Decrease Causing Severe Chills', is highlighted with a red box. A red arrow points from the text 'Click to go to the actual question page' to this result. A red label 'Question ID' with an arrow points to the ID '83614'.

NetWellness Home

stress stroke Search Questions Reset

14 questions found

23918 Health Heart Diet Answer

The question: I understand that there is a diet that will remove arterial deposits. Is this correct? If correct, where can I get a copy of the diet? Background: I just...

83614 Prednisone Decrease Causing Severe Chills Answer

I have taken prednisone for 27 years. Usually around 10-15 mg per day but up to 35mg as needed when stressed, as my Crohn's gastro doctor is aware. I began remicade three years...

15560 Medication and fluctuating bp Answer

I'm a female, 47 and overweight taking small doses of prozac and elavil for fibromyalgia and am on HRT. My dr is watching my bp- my diastolic reading and pulse rise to over 100...

Query	
What color is the grass?	
Responses	
Tomorrow is Saturday	1
The grass is green	3
The capital of Canada is Ottawa	2
The sky is blue	2
A whale is a mammal	1

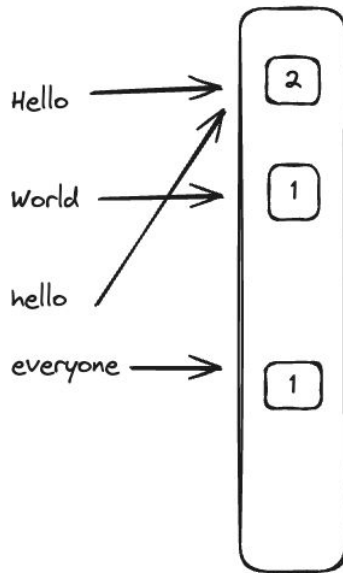
Number of words in common

Sparse Vectors

Keyword-based search in the context of hybrid search often uses a representation called sparse embeddings, which is why it is also referred to as sparse vector search. Sparse embeddings are vectors with mostly zero values with only a few non-zero values, as shown below.

Sparse Vectors

Hello world, hello everyone



let's say hello to the world with our new project

let's

say

hello

to

the

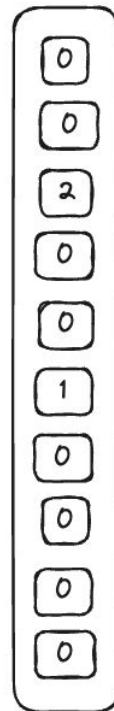
world

with

our

new

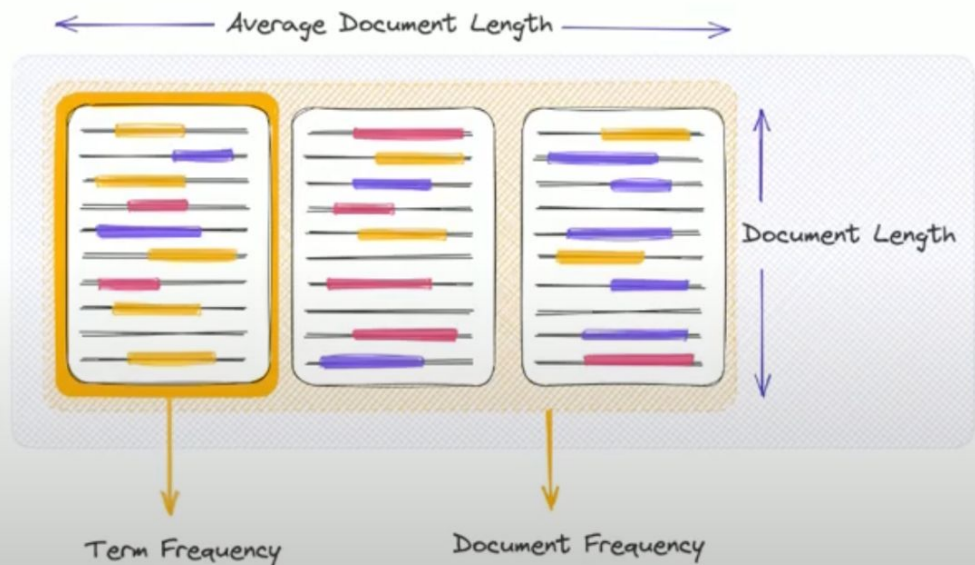
project



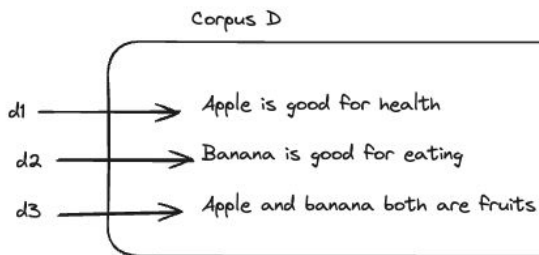
TF-IDF



BM-25



TF-IDF



Search: Apple

TF Calculation:

$$TF(\text{"Apple"}, d1) = 1/5 = 0.2$$

$$TF(\text{"Apple"}, d2) = 0/5 = 0$$

$$TF(\text{"Apple"}, d3) = 1/6 = 0.167$$

TF-IDF

Term Frequency - Inverse Document Frequency

IDF Calculation:

$$IDF(\text{"Apple"}) = \log(3/2) = 0.4054651081081644$$

TF-IDF Calculation

$$TF-IDF(\text{"Apple"}, d1) = 0.405 * 0.2 = 0.081 \text{ (approx)}$$

$$TF-IDF(\text{"Apple"}, d2) = 0.405 * 0 = 0$$

$$TF-IDF(\text{"Apple"}, d3) = 0.405 * 0.167 = 0.068$$

Result: The word apple has more relevant to Document 1

BM-25

BM-25 (Best Match)

Document Length Normalization and non-linear term frequency scaling (saturation)

Given a query Q , containing keywords q_1, \dots, q_n , the BM25 score of a document D is:

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

where $f(q_i, D)$ is q_i 's **term frequency** in the document D , $|D|$ is the length of the document D in words, and avgdl is the average document length in the text collection from which documents are drawn. k_1 and b are free parameters, usually chosen, in absence of an advanced optimization, as $k_1 \in [1.2, 2.0]$ and $b = 0.75$.^[1] $\text{IDF}(q_i)$ is the IDF (**inverse document frequency**) weight of the query term q_i . It is usually computed as:

$$\text{IDF}(q_i) = \ln \left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} + 1 \right)$$

where N is the total number of documents in the collection, and $n(q_i)$ is the number of documents containing q_i .

BM-25

BM-25 (Best Match)

$$IDF = \log \left[\frac{N - n(t) + 0.5}{n(t) + 0.5} + 1 \right]$$

IDF calculation

Total Number of documents in the corpus = 3,

$n(\text{Apple}) = 2$

$$IDF(\text{Apple}) = \log \left(\frac{(3-2+0.5)}{(2+0.5)} + 1 \right) = 0.47$$

TF calculation

$$TF(\text{Apple}, d1) = 1/5 = 0.2$$

$$TF(\text{Apple}, d2) = 0/5 = 0$$

$$TF(\text{Apple}, d3) = 1/6 = 0.167$$

BM-25 Calculation

k_1 and b are parameters that control the influence of TF and document length.

Common values are
 $k_1 = 1.5$ and $b = 0.75$.

$$\text{score}(D, t) = IDF(t) * \frac{TF(t, D)^{k_1+1}}{TF(t, D)^{k_1} + k_1 * \left[\frac{1 - b + b * \frac{|D|}{\text{avgdl}}}{1} \right]}$$

$$\text{score}(d1, \text{Apple}) = 0.470$$

$$\text{score}(d2, \text{Apple}) = 0$$

$$\text{score}(d3, \text{Apple}) = 0.470$$

The formula effectively balances the frequency of the term in the document against the frequency of the term in the entire corpus, adjusted for document length,

Key Takeaway

Memory Efficiency: They use less memory due to high proportion of zeros.

Interpretability: Dimensions often correspond directly to specific features, enhancing transparency.

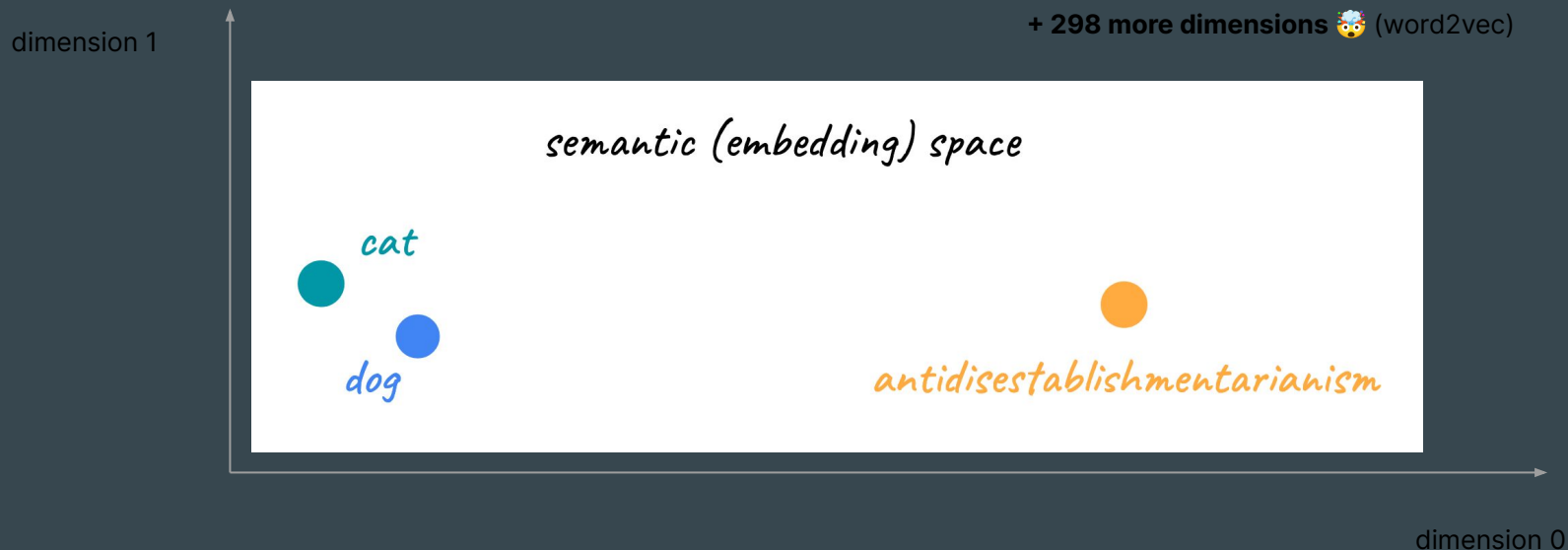
High Dimensionality: May result in computational challenges and inefficiencies.

Difficulty in Capturing Complexity: Sparse nature can make it hard to represent complex patterns effectively.

Embeddings

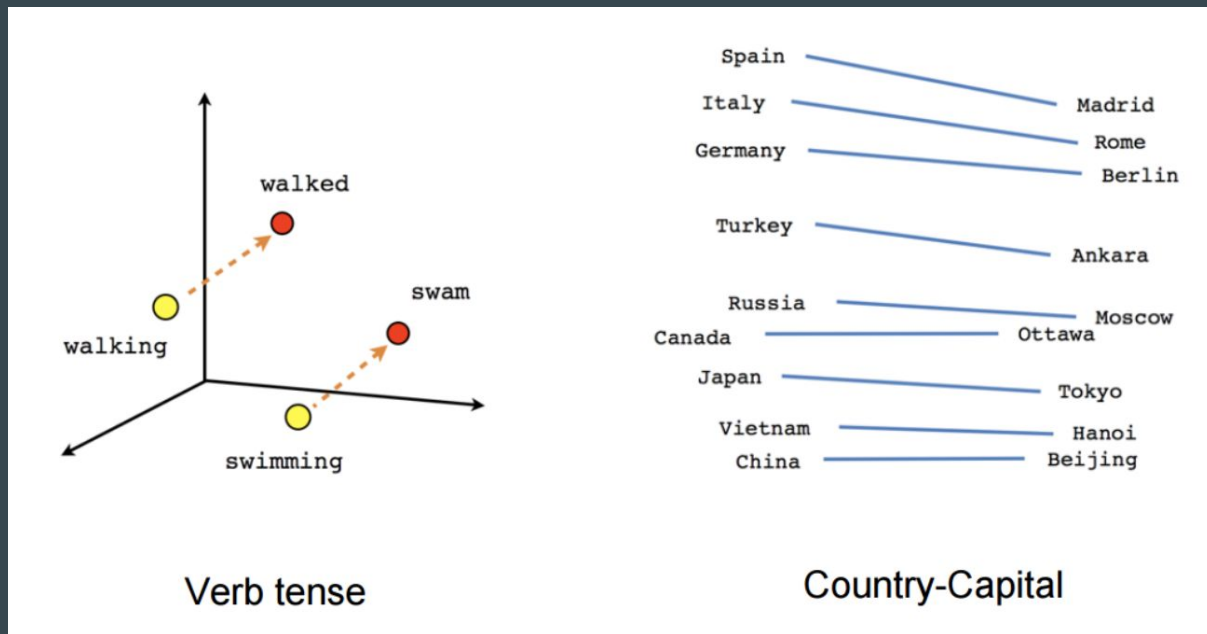
A list (vector) of numbers [0.5, 0.798, 0.03, ...]

Similar things are close together

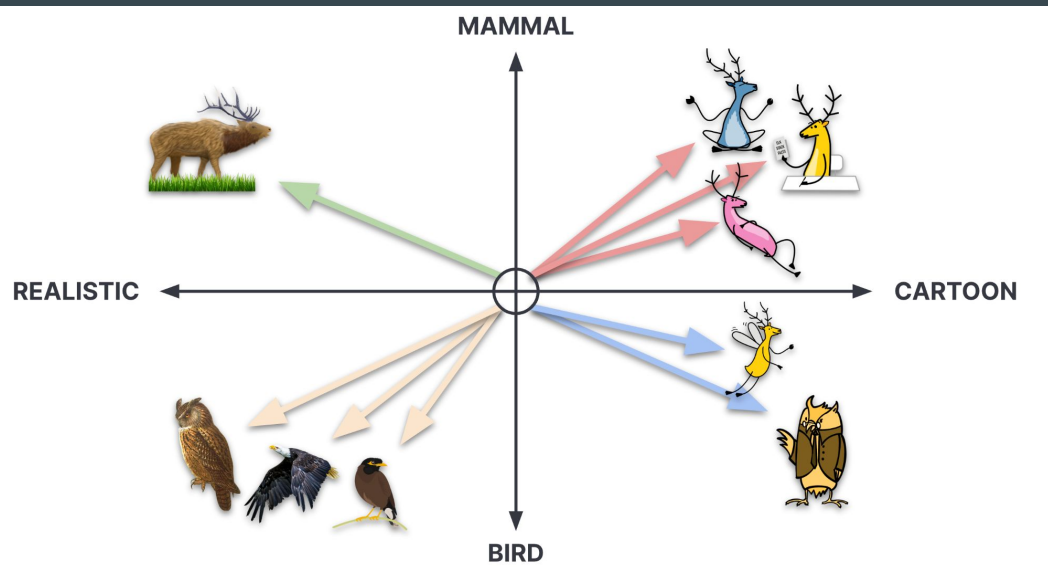


Semantic algebra with embeddings

Kinda handy, like we might do with vectors

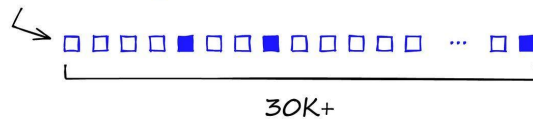


Semantic search



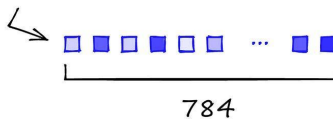
sparse

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



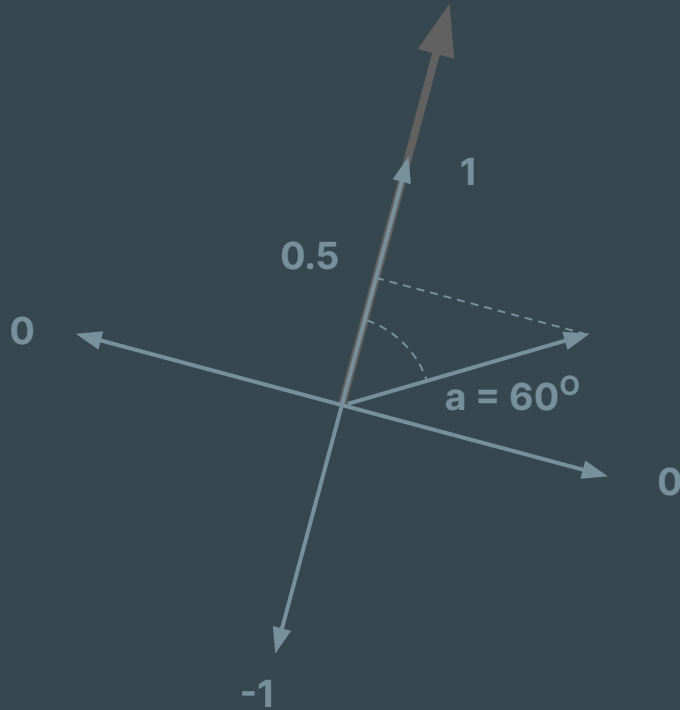
dense

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



Dense Vectors


Measuring similarity with cosine





Cosine similarity puns are sticky


Semantle - a game of semantic

Semantle

Team

Stats

Rules

Settings

Today is puzzle number **494**. The nearest word has a similarity of **83.89**, the tenth-nearest has a similarity of 45.49 and the one thousandth nearest word has a similarity of 19.85.

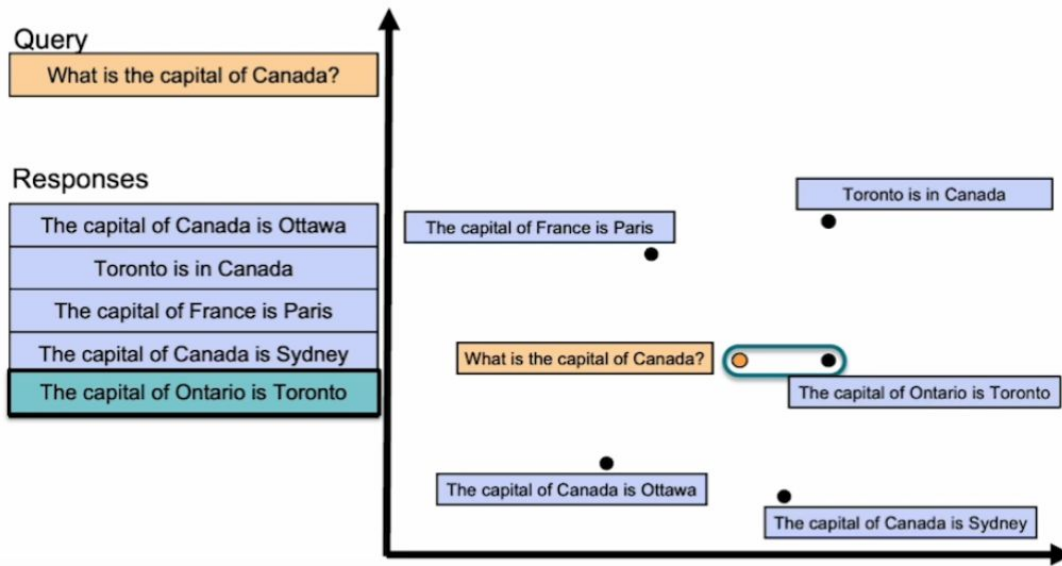
Guess

#	Guess	Similarity	Getting close?
7	vegetable	0.49	(cold)
1	carrot	8.81	(cold)
3	ethics	8.78	(cold)
6	embedding	5.83	(cold)
4	running	3.66	(cold)
2	horse	1.47	(cold)
5	safety	0.07	(cold)

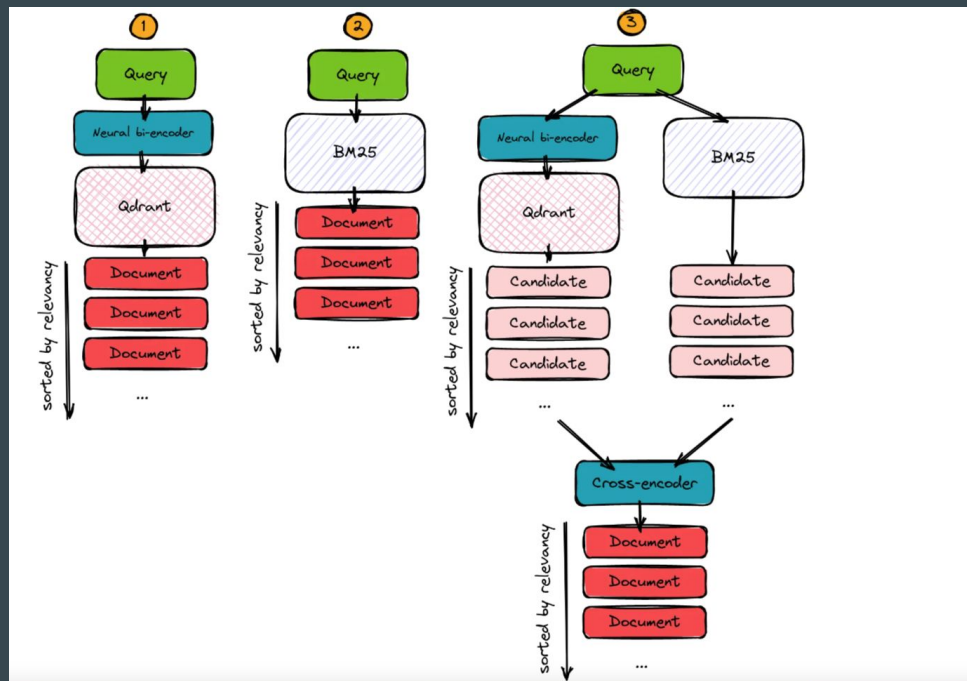
Hint

Give up

Dense Retrieval is also not perfect

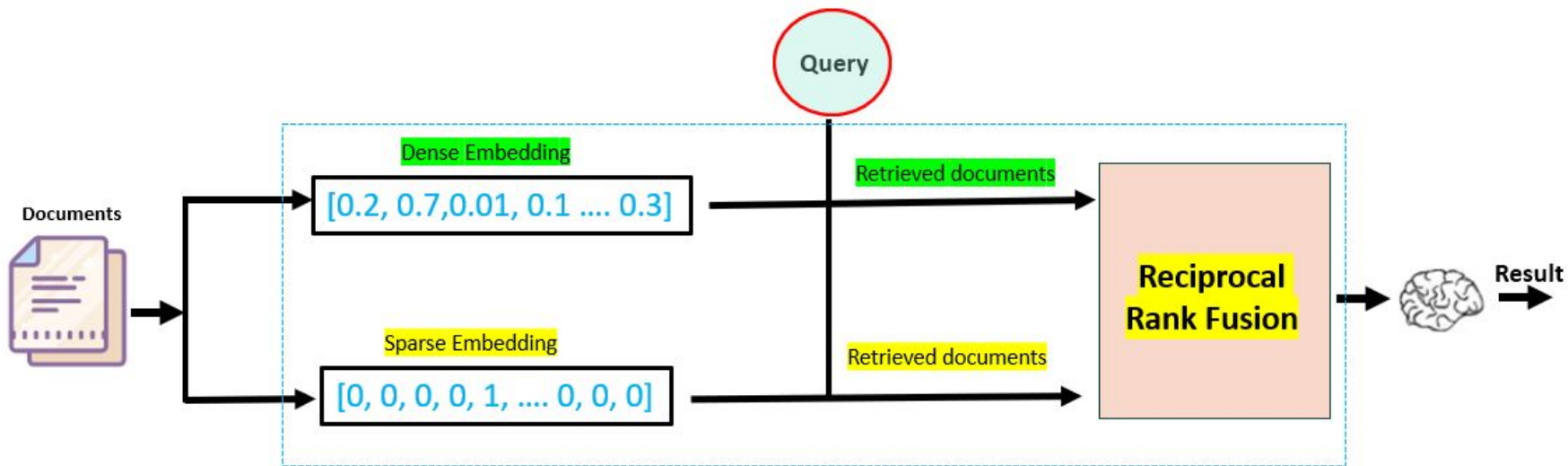


Hybrid Search



Hybrid search operates by first performing individual searches across both sparse and dense vectors simultaneously. It then utilizes a method known as Reciprocal Rank Fusion (RRF)

Hybrid Vector Search



Demo Load Index

```
15     "max_score": 1,
16     "hits": [
17     {
18         "_index": "my-semantic-search-index",
19         "_id": "1",
20         "_score": 1,
21         "_source": {
22             "id": "d1",
23             "event_text": ""Join us for a day of laughter, joy, and endless smiles at our spectacular
                Family and Kids Event!
24 This enchanting celebration is designed to bring families together for a memorable experience filled
25 with excitement and wholesome activities.
26 Highlights:
27 1. Magical Adventure Zone:
28 Step into a world of enchantment where your little ones can embark on a magical journey. From face
    painting
29 to interactive story sessions, the Adventure Zone promises endless wonders for the young at heart.
30 2. Arts and Crafts Wonderland:
31 Unleash your creativity in our Arts and Crafts Wonderland! Join hands with your kids to create beautiful
32 masterpieces and cherished memories together. From DIY crafts to painting stations, let imagination run
    wild.
33 3. Live Entertainment & Performances:
34 Be captivated by live performances that will leave the whole family in awe. Magicians, puppet shows, and
35 lively music will fill the air with laughter and cheer.
36 4. Delicious Delights:
37 Indulge in a variety of mouthwatering treats from our food stalls. From cotton candy to popcorn, there's
38 something to satisfy every craving.
39 5. Family Games & Contests:
40 Engage in friendly competition with our exciting games and contests. Prizes await the winners, but the
    real
41 reward is the bonding experience with your loved ones.
42 6. Petting Zoo & Pony Rides:
43 Connect with furry friends at our delightful petting zoo and give your little ones the joy of pony rides
    .
44 A perfect opportunity for animal lovers of all ages.
45 7. Photo Booth Fun:
46 Capture the magic of the day at our themed photo booths. Dress up in costumes, strike a pose, and take
    home
```

```
1     "_index": "my-semantic-search-index",
2     "_id": "2",
3     "_score": 1,
4     "_source": {
5         "id": "d2",
6         "event_text": ""Unlock the Power of Pen: Dramatic Writing Class
7 Embark on a journey into the captivating world of storytelling with our exhilarating Dramatic Writing
    Class!
8 Whether you're a seasoned wordsmith or a budding storyteller, this course promises to ignite your
    imagination,
9 refine your narrative skills, and unleash the dramatist within you.
10 What to Expect:
11 1. Craft Compelling Narratives:
12 Dive deep into the art of storytelling as we explore the intricacies of plot development, character arcs
    , and
13 creating tension. Uncover the secrets of crafting narratives that leave a lasting impact.
14 2. Character Development Workshop:
15 Bring your characters to life! Learn the techniques to create multidimensional characters with relatable
    motivations
16 and authentic voices. Delve into the psychology of your characters and make them resonate with your
    audience.
17 3. Mastering Dialogue:
18 Discover the power of dialogue in creating riveting scenes. From natural conversations to intense
    confrontations,
19 our class will guide you in crafting dialogues that ring true and enhance the emotional depth of your
    writing.
20 4. Exploring Genres:
21 Whether your heart lies in drama, suspense, or comedy, our class will help you navigate various genres.
    Unleash
22 your creativity by experimenting with different styles, and find your unique voice as a dramatic writer.
```

Demo - Keyword Search

GET /my-semantic-search-index/_search

```
{
  "query": {
    "match": {
      "event_text": "samvardhan"
    }
  }
}
```

```
{
  "took": 36,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 0,
      "relation": "eq"
    },
    "max_score": null,
    "hits": []
  }
}
```

Demo - Semantic Search

```
GET /my-semantic-search-index/_search
{
  "query": {
    "neural": {
      "event_dense_embedding": {
        "query_text": "samvardhan",
        "model_id": "m2lMMI8BfMxv4jg4ugFP",
        "k": 100
      }
    }
  }
}
```

```

"took": 183,
"timed_out": false,
"_shards": {
  "total": 1,
  "successful": 1,
  "skipped": 0,
  "failed": 0
},
"hits": {
  "total": {
    "value": 4,
    "relation": "eq"
  },
  "max_score": 0.013157864,
  "hits": [
    {
      "_index": "my-semantic-search-index",
      "_id": "4",
      "_score": 0.013157864,
      "_source": {
        "id": "d4",
        "event_text": "\"\"Harmony in Elegance: A Night of Classical Splendor
Indulge your senses in an enchanting evening of refined melodies and timeless compositions at \"Harmony
in Elegance.\"
Join us for a classical music event that promises to transport you to a world of sophistication and
emotive beauty.
Program Highlights:
1. Masterful Orchestral Performance:
Immerse yourself in the sublime sounds of a world-class orchestra as they bring to life classical
masterpieces
from renowned composers. Experience the rich textures and emotive nuances of timeless symphonies.
2. Virtuoso Solo Performances:
Witness the breathtaking skill of accomplished soloists as they showcase their artistry on instruments
like the
violin, cello, and piano. Let their virtuosity leave you spellbound in a cascade of musical brilliance.

```

Demo - Hybrid Search

```
GET /my-semantic-search-index/_search?search_pipeline=hybrid-search-pipeline
```

```
{
  "query": {
    "hybrid": {
      "queries": [
        {
          "match": {
            "text": {
              "query": "children"
            }
          }
        },
        {
          "neural": {
            "event_dense_embedding": {
              "query_text": "children",
              "model_id": "m2LMMI8BfMxv4jg4ugFP",
              "k": 100
            }
          }
        }
      ]
    }
  }
}
```

```
"max_score": 1,
"hits": [
  {
    "_index": "my-semantic-search-index",
    "_id": "4",
    "_score": 1,
    "_source": {
      "id": "d4",
      "event_text": ""Harmony in Elegance: A Night of Classical Splendor
```

Indulge your senses in an enchanting evening of refined melodies and timeless compositions at "Harmony in Elegance."

Join us for a classical music event that promises to transport you to a world of sophistication and emotive beauty.

Program Highlights:

1. Masterful Orchestral Performance:

Immerse yourself in the sublime sounds of a world-class orchestra as they bring to life classical masterpieces

from renowned composers. Experience the rich textures and emotive nuances of timeless symphonies.

2. Virtuoso Solo Performances:

Witness the breathtaking skill of accomplished soloists as they showcase their artistry on instruments like the

violin, cello, and piano. Let their virtuosity leave you spellbound in a cascade of musical brilliance.

3. Chamber Music Intimacy:

Delight in the intimacy of chamber music as talented ensembles perform in perfect harmony. Experience the interplay

of instruments in an up-close and personal setting, revealing the intricate beauty of classical compositions.

4. Timeless Compositions Reimagined:

Be captivated by innovative interpretations of classical compositions, breathing new life into familiar pieces



Repo Link