

Build a Vector Computer with Superlinked

Preview

Prof. Tom Yeh

Hosted by:



University of Colorado
Boulder



**Analytics
Vidhya**

Roadmap

Build a { Personal
C++ } Computer
Pandas
Vector

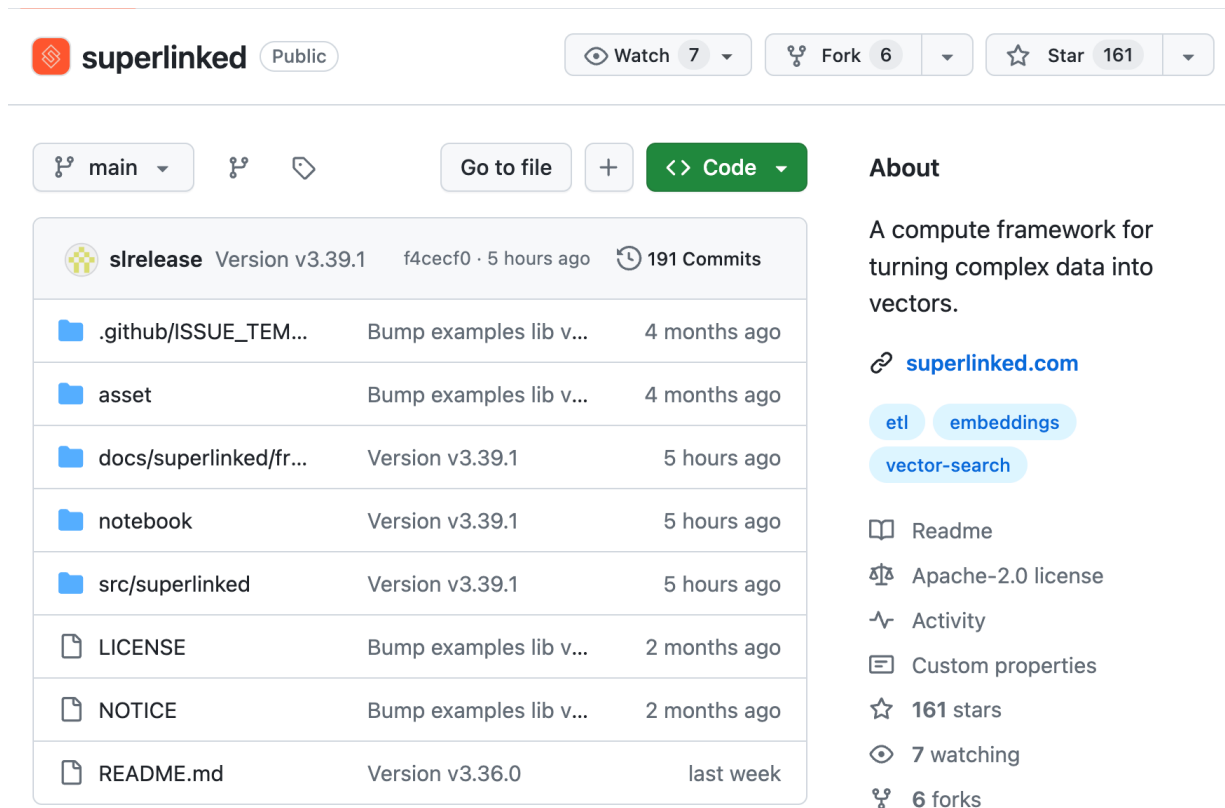
Opensource

RAG

Math

KNN

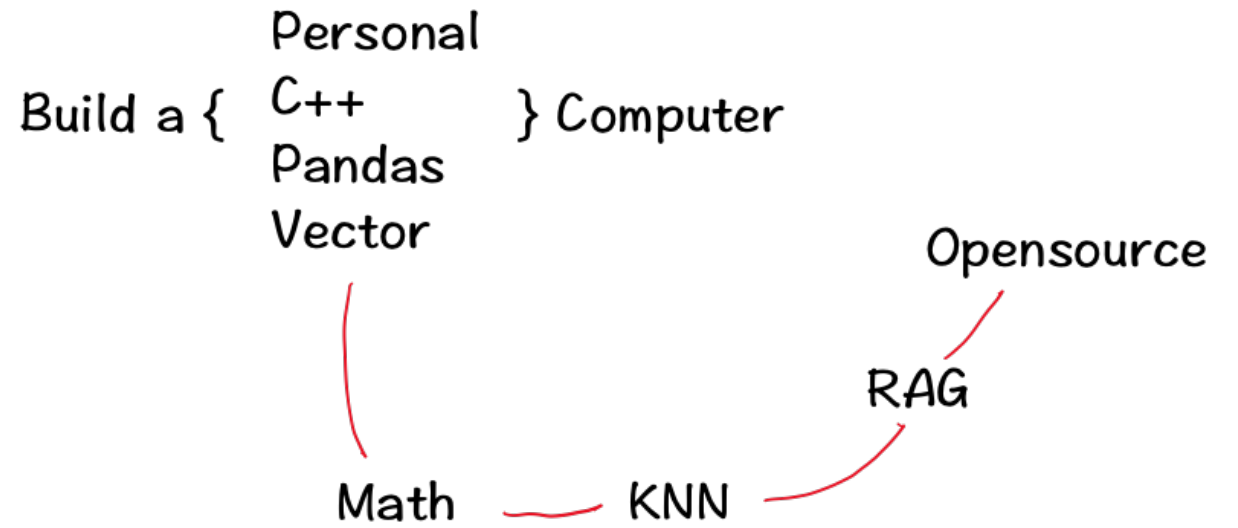
Star the Repo: <https://github.com/superlinked/superlinked>



The screenshot shows the GitHub repository page for `superlinked/superlinked`. At the top, it indicates the repository is public and shows 7 watchers, 6 forks, and 161 stars. The main branch is `main`. Below the repository name, there is a table of recent commits and files. The 'About' section describes the project as a compute framework for turning complex data into vectors, with tags for `etl`, `embeddings`, and `vector-search`. It also lists the Apache-2.0 license, activity, and other repository statistics.

File	Commit	Time
<code>.github/ISSUE_TEM...</code>	Bump examples lib v...	4 months ago
<code>asset</code>	Bump examples lib v...	4 months ago
<code>docs/superlinked/fr...</code>	Version v3.39.1	5 hours ago
<code>notebook</code>	Version v3.39.1	5 hours ago
<code>src/superlinked</code>	Version v3.39.1	5 hours ago
<code>LICENSE</code>	Bump examples lib v...	2 months ago
<code>NOTICE</code>	Bump examples lib v...	2 months ago
<code>README.md</code>	Version v3.36.0	last week





PC

Build a Vector Computer with Superlinked 

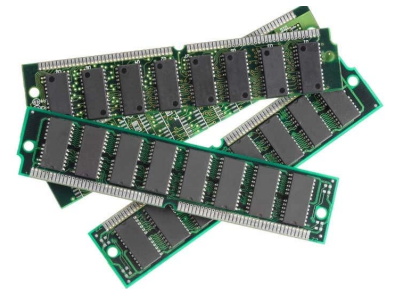


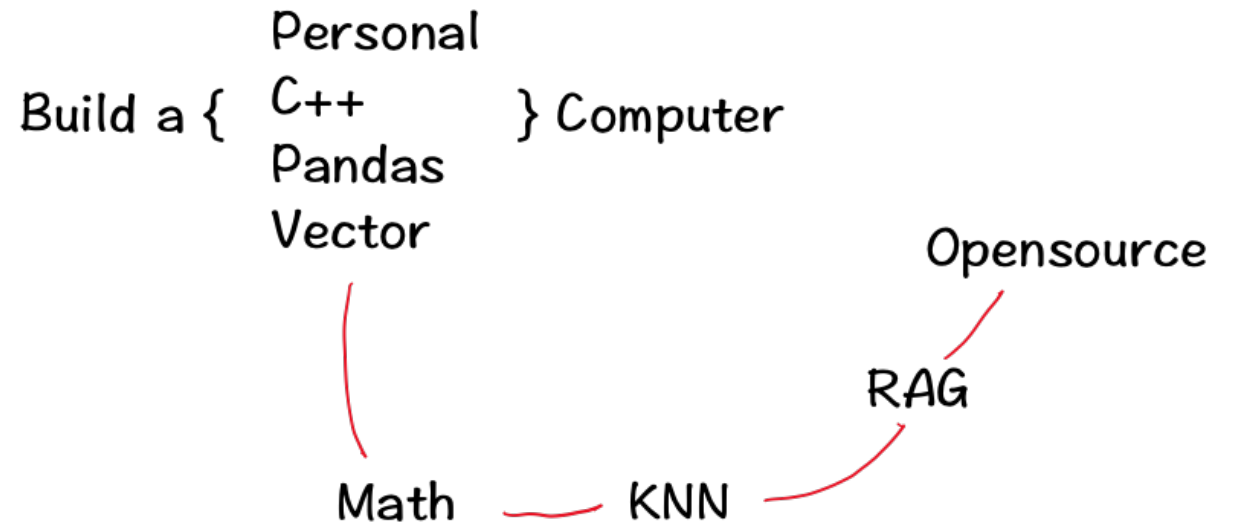
Memory

Output

Processor

Input





C++

Build a Vector Computer with Superlinked 



University of Colorado
Boulder

Input

```
cin    c;
```

Processor

```
#include < >  
#include < >  
int a = 5;  
int b = 9;  
int c;
```

Output

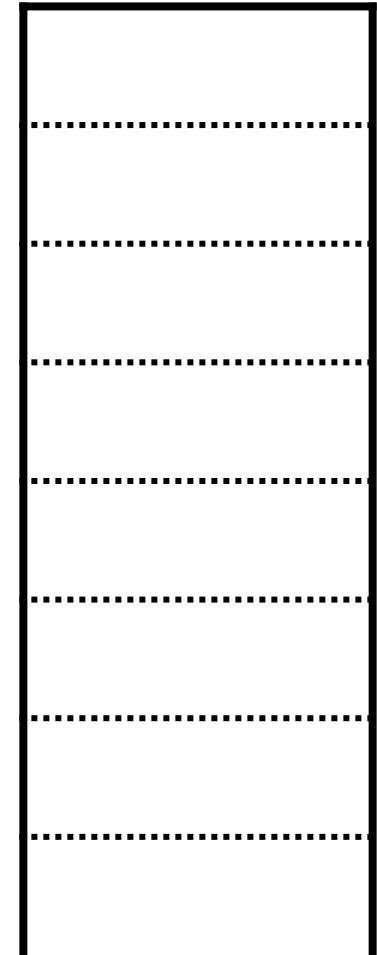
```
cout   a;
```

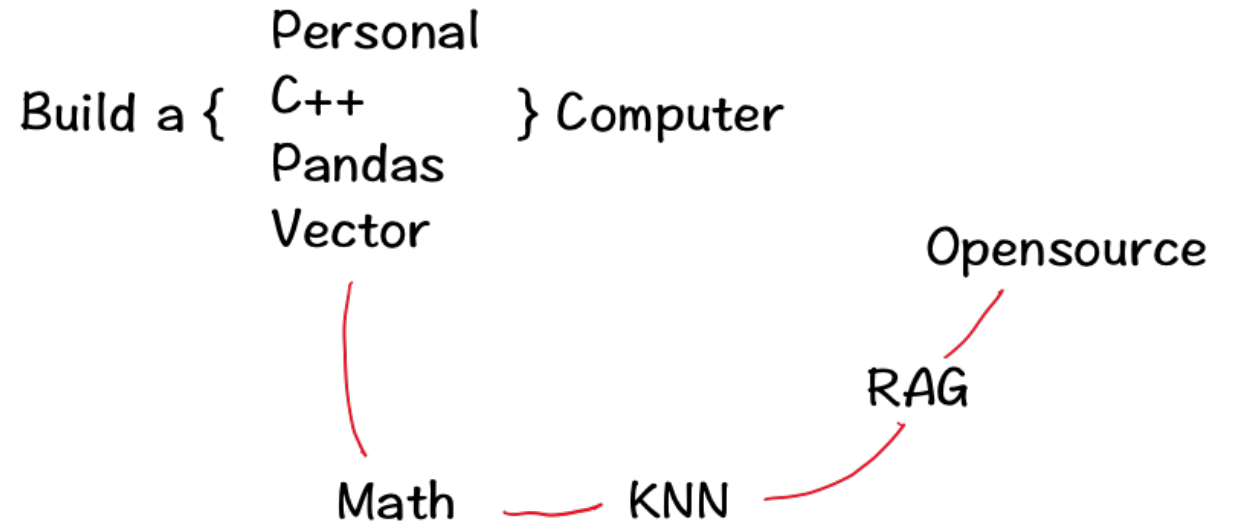
```
cout   b;
```

```
// find nearest to c  
if ( > )
```

```
else
```

Memory





Pandas

Build a Vector Computer with Superlinked 



data.csv

```
name, age, city  
John, 30, NYC  
Alice, 22, LA  
Bob, 35, Chicago
```

Input

```
df = pd. [REDACTED] ('data.csv')
```

name, age, city
John, 30, NYC
Alice, 22, LA
Bob, 35, Chicago

Memory

Processor

```
# select users older than 25
```

```
s = df[ [REDACTED] [REDACTED] ]
```

```
df1 = df[ [REDACTED] ]
```

ID	name	age	city
1			
2			
3			

Output

```
print(df1)
```

Input

```
df = pd.read_csv('data.csv')
```

Processor

```
# select users older than 25 in DC
```

```
s = (df[ ] ) & (df[ ] )
```

```
df1 = df[ ]
```

Output

```
print(df1)
```

Memory

ID	name	age	city
1	John	30	NYC
2	Alice	22	LA
3	Bob	35	DC

Input

```
df = pd.read_csv('data.csv')
```

Processor

```
# select users similar to [Rob, 31, SF]
```

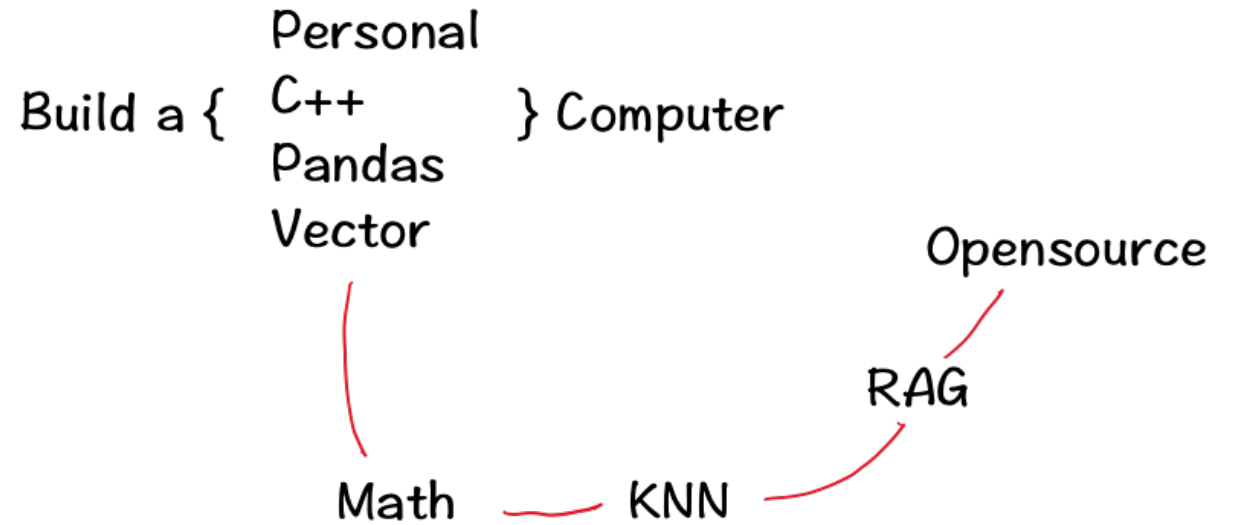
```
s = (df[ ] ) & (df[ ] )  
    & (df[ ] )  
df1 = df[s]
```

Output

```
print(df1)
```

Memory

ID	name	age	city
1	John	30	NYC
2	Alice	22	LA
3	Bob	35	DC

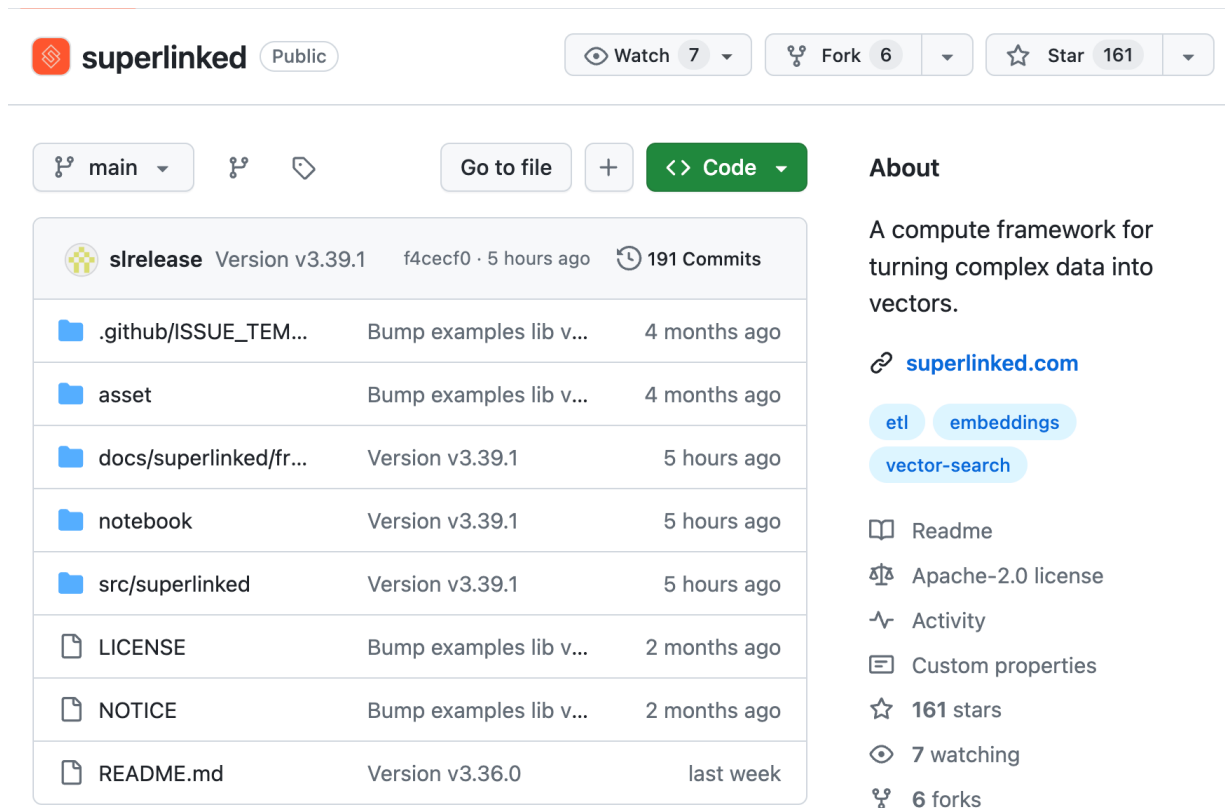


Vector Computer

Build a Vector Computer with Superlinked 



Star the Repo: <https://github.com/superlinked/superlinked>



The screenshot shows the GitHub repository page for `superlinked/superlinked`. At the top, it indicates the repository is public and shows 7 watchers, 6 forks, and 161 stars. The main branch is `main`. Below the repository name, there is a table of recent commits and files. The 'About' section describes the project as a compute framework for turning complex data into vectors, with tags for `etl`, `embeddings`, and `vector-search`. It also lists the license as Apache-2.0 and shows 161 stars, 7 watchers, and 6 forks.

File	Commit	Time
<code>.github/ISSUE_TEM...</code>	Bump examples lib v...	4 months ago
<code>asset</code>	Bump examples lib v...	4 months ago
<code>docs/superlinked/fr...</code>	Version v3.39.1	5 hours ago
<code>notebook</code>	Version v3.39.1	5 hours ago
<code>src/superlinked</code>	Version v3.39.1	5 hours ago
<code>LICENSE</code>	Bump examples lib v...	2 months ago
<code>NOTICE</code>	Bump examples lib v...	2 months ago
<code>README.md</code>	Version v3.36.0	last week



news.json

```
"headline": {
  "162": "REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor",
  "353": "Twitter Lawyer Calls Elon Musk 'Committed Enemy' As Judge Sets October Trial",
  "632": "Starbucks Leaving Russian Market, Shutting 130 Stores",
  ...
},
"short_description": {
  "162": "They follow in the footsteps of REI workers in New York City who formed a union earlier this year.",
  "353": "Delaware Chancery Judge Kathaleen McCormick dealt the world's richest person a setback in ordering a speedy trial on his abandoned deal to buy Twitter.",
  "632": "Starbucks' move follows McDonald's exit from the Russian market last week."
},
"release": {
  ...
}
```

Input

```
df = pd.read_json('news.json')
```

Processor

```
# select news about Starbucks
```

```
s1 = df[df['text'].str.contains('starbucks', case=False)]  
s2 = df[df['text'].str.contains('starbucks', case=False)]
```

```
df1 = df[s1]
```

Output

```
print(df1)
```

Memory

ID	Description	Headline	Release
1	They follow in the footsteps of REI workers in New York City who formed a union earlier this year.	REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor	2016-06-13
2	Delaware Chancery Judge Kathaleen McCormick dealt the world's richest person a setback in ordering a speedy trial on his abandoned deal to buy Twitter.	Twitter Lawyer Calls Elon Musk 'Committed Enemy' As Judge Sets October Trial	2014-06-26
3	Starbucks' move follows McDonald's exit from the Russian market last week.	Starbucks Leaving Russian Market, Shutting 130 Stores	2015-01-06

Define Spaces

```
# textual characteristics are embedded using a sentence-transformers model
description_space = TextSimilaritySpace(
    text=news.description, model="sentence-transformers/all-mpnet-base-v2"
)
headline_space = TextSimilaritySpace(
    text=news.headline, model="sentence-transformers/all-mpnet-base-v2"
)
# release date is encoded using our recency embedding algorithm
recency_space = RecencySpace(
    timestamp=news.release_timestamp,
    period_time_list=[
        PeriodTime(timedelta(days=4 * YEAR_IN_DAYS), weight=1),
        PeriodTime(timedelta(days=11 * YEAR_IN_DAYS), weight=2),
    ],
    negative_filter=0.0,
)
```

Superlinked API 

Memory

ID	Description	Headline	Release
1	They follow in the footsteps of REI workers in New York City who formed a union earlier this year.	REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor	2016-06-13
2	Delaware Chancery Judge Kathaleen McCormick dealt the world's richest person a setback in ordering a speedy trial on his abandoned deal to buy Twitter.	Twitter Lawyer Calls Elon Musk 'Committed Enemy' As Judge Sets October Trial	2014-06-26
3	Starbucks' move follows McDonald's exit from the Russian market last week.	Starbucks Leaving Russian Market, Shutting 130 Stores	2015-01-06

Embed and Index

ID	Description	Headline	Release
1	They follow in the footsteps of REI workers in New York City who formed a union earlier this year.	REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor	2016-06-13
2	Delaware Chancery Judge Kathaleen McCormick dealt the world's richest person a setback in ordering a speedy trial on his abandoned deal to buy Twitter.	Twitter Lawyer Calls Elon Musk 'Committed Enemy' As Judge Sets October Trial	2014-06-26
3	Starbucks' move follows McDonald's exit from the Russian market last week.	Starbucks Leaving Russian Market, Shutting 130 Stores	2015-01-06

ID	Description	Headline	Recency									
1	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	0	<table border="1"><tr><td>0</td><td>1</td><td>0</td></tr></table>	0	1	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
0	1	1	0									
0	1	0										
1	0											
2	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td></tr></table>	1	1	0	0	<table border="1"><tr><td>1</td><td>0</td><td>0</td></tr></table>	1	0	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
1	1	0	0									
1	0	0										
1	0											
3	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	0	1	1	1	<table border="1"><tr><td>0</td><td>0</td><td>1</td></tr></table>	0	0	1	<table border="1"><tr><td>0</td><td>1</td></tr></table>	0	1
0	1	1	1									
0	0	1										
0	1											

Input

Search

Output

df = result._____()
_____ (df)

Processor

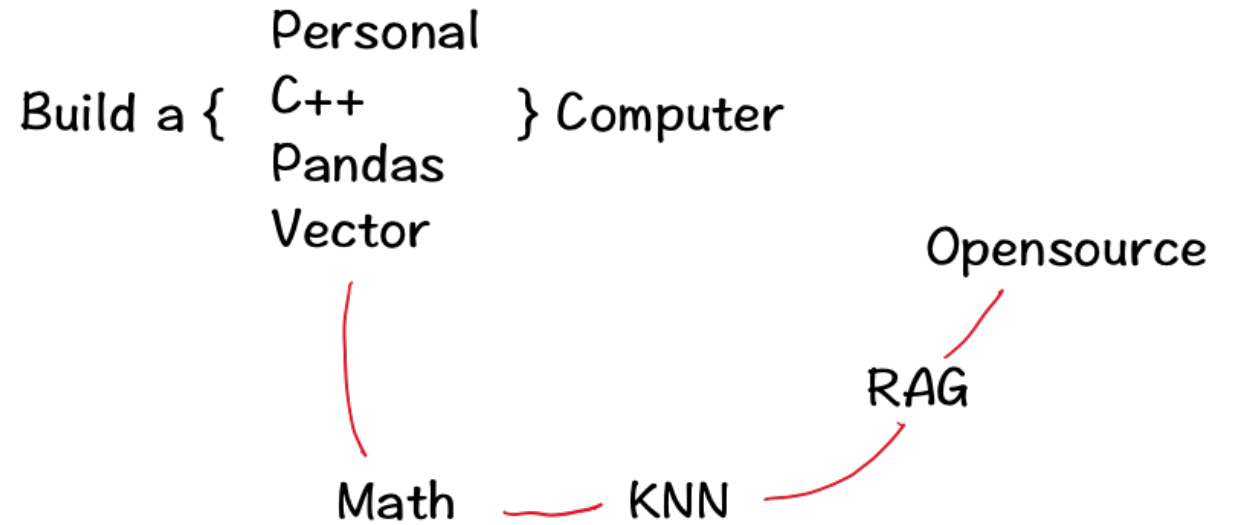
```
result = app.query(  
    simple_query,  
    query_text=_____  
    description_weight=1,  
    headline_weight=1,  
    recency_weight=0,  
)
```

Superlinked API 

Memory

ID	Description	Headline	Recency									
1	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	0	<table border="1"><tr><td>0</td><td>1</td><td>0</td></tr></table>	0	1	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
0	1	1	0									
0	1	0										
1	0											
2	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td></tr></table>	1	1	0	0	<table border="1"><tr><td>1</td><td>0</td><td>0</td></tr></table>	1	0	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
1	1	0	0									
1	0	0										
1	0											
3	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	0	1	1	1	<table border="1"><tr><td>0</td><td>0</td><td>1</td></tr></table>	0	0	1	<table border="1"><tr><td>0</td><td>1</td></tr></table>	0	1
0	1	1	1									
0	0	1										
0	1											





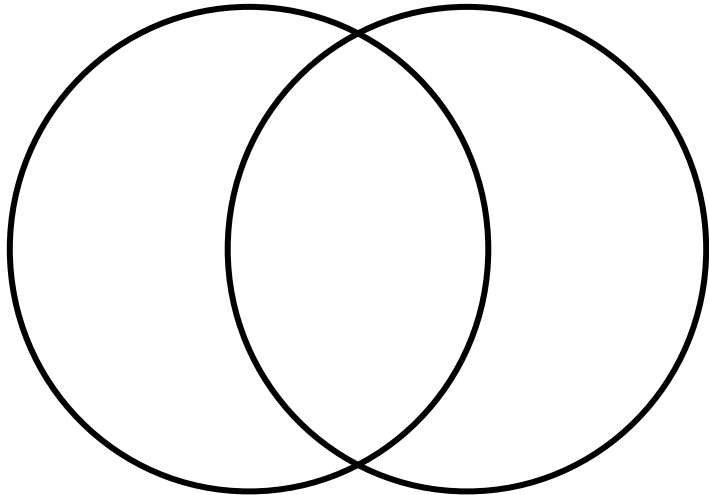
Math

Build a Vector Computer with Superlinked 

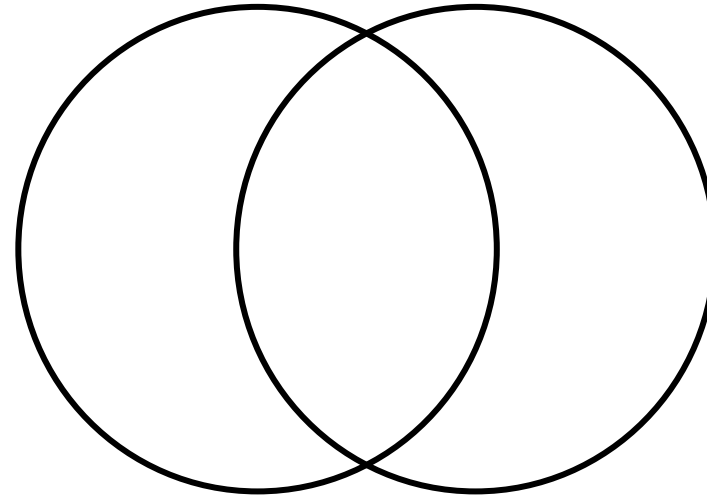


Draw distance vs similarity

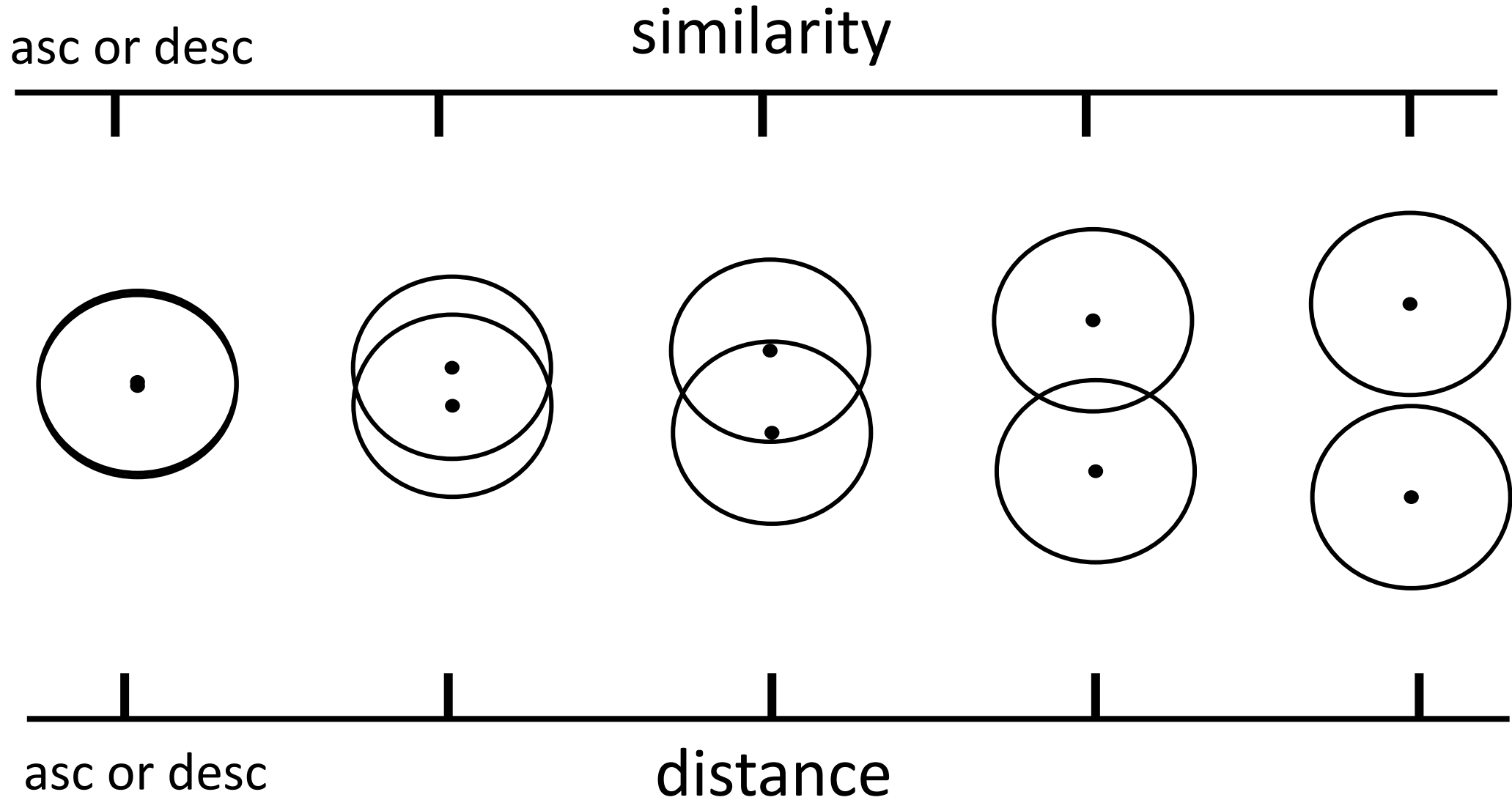
distance



similarity



Distance vs similarity on a scale of 1 to 5



How to compute dot product?

Example:

1	2	3
---	---	---

* * *

2	2	0
---	---	---

= = = Σ

2	4	0
---	---	---

6

Result

dog

2	1	0
---	---	---

* * *

cat

1	2	0
---	---	---

= = = Σ

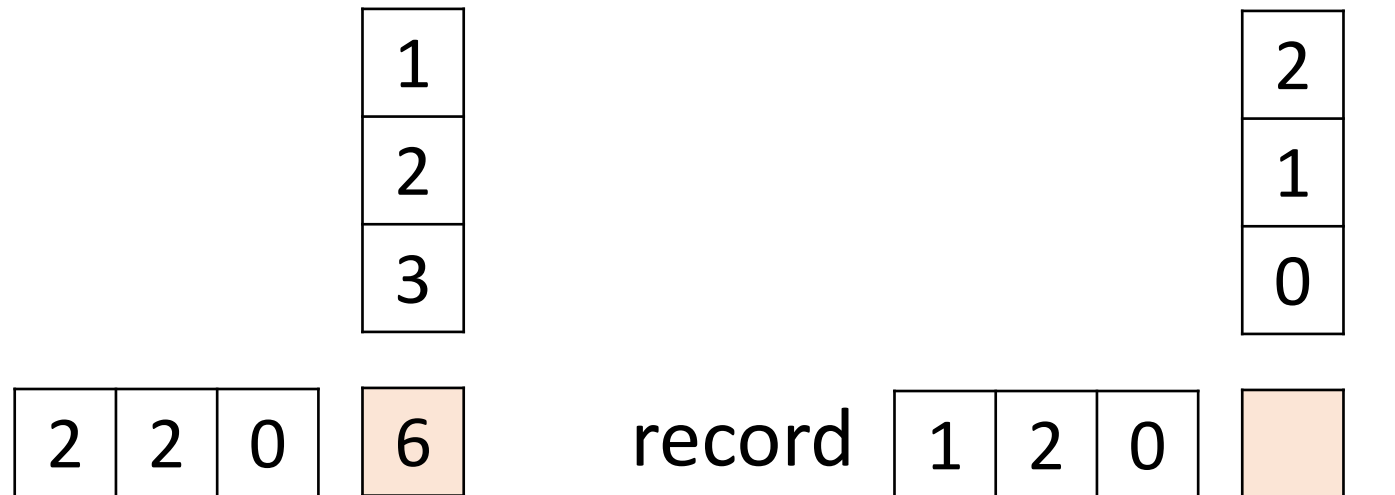
--	--	--

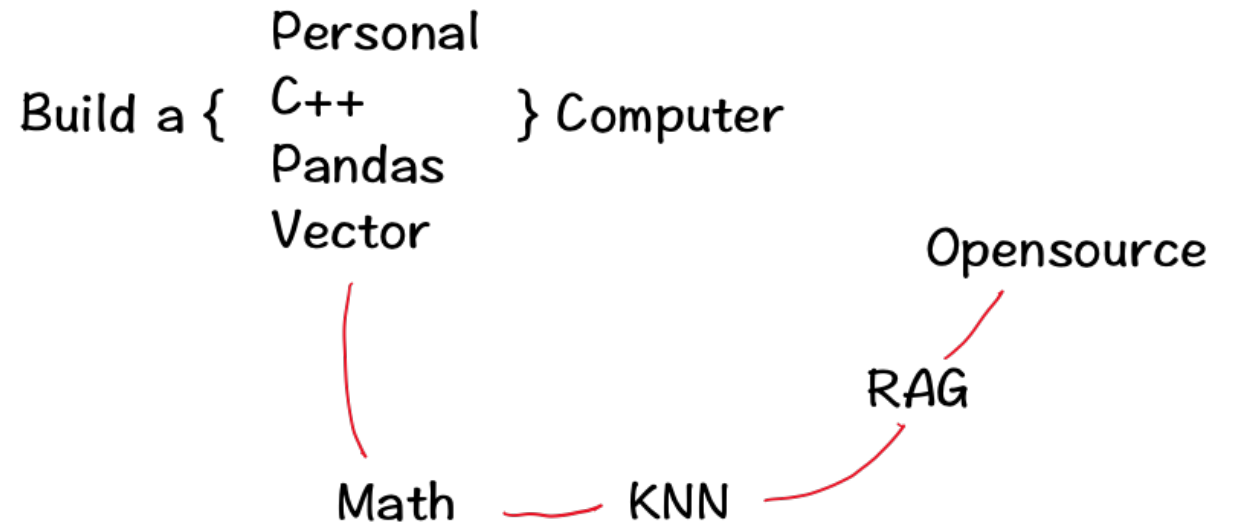
--

Result

How to compute dot product using matrix multiplication?

Example:





KNN

Build a Vector Computer with Superlinked 

Input

Search

Output

df = result._____()

_____ (df)

Processor

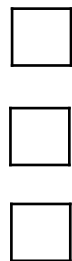
```
result = app.query(  
    simple_query,  
    query_text=██████████,  
    description_weight=1,  
    headline_weight=1,  
    recency_weight=0,  
)
```

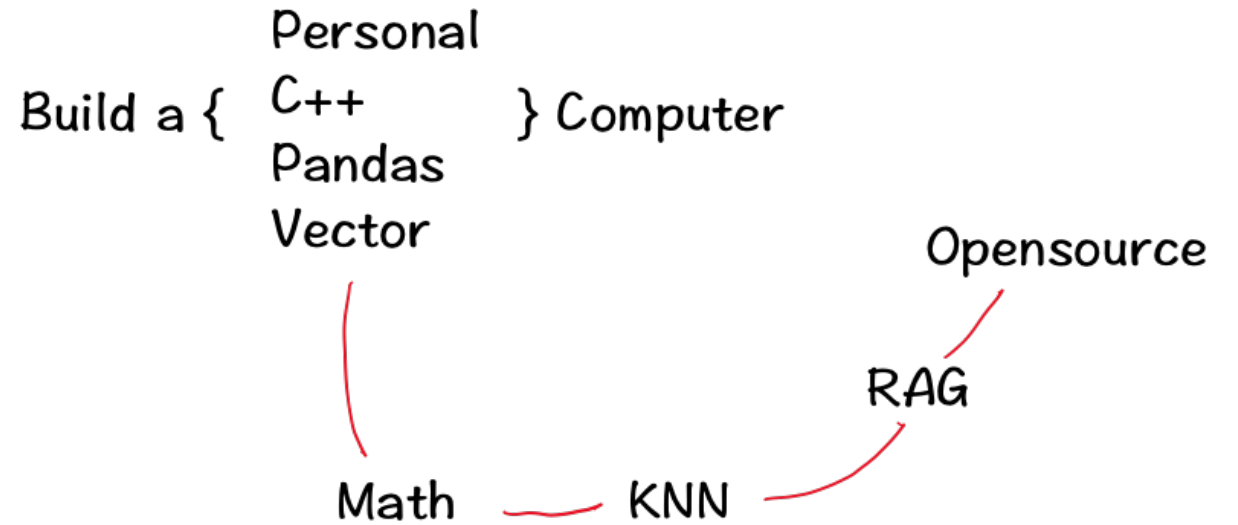
Superlinked API 

Memory



ID	Description	Headline	Recency									
1	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	0	<table border="1"><tr><td>0</td><td>1</td><td>0</td></tr></table>	0	1	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
0	1	1	0									
0	1	0										
1	0											
2	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td></tr></table>	1	1	0	0	<table border="1"><tr><td>1</td><td>0</td><td>0</td></tr></table>	1	0	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
1	1	0	0									
1	0	0										
1	0											
3	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	0	1	1	1	<table border="1"><tr><td>0</td><td>0</td><td>1</td></tr></table>	0	0	1	<table border="1"><tr><td>0</td><td>1</td></tr></table>	0	1
0	1	1	1									
0	0	1										
0	1											





RAG

Build a Vector Computer with Superlinked 



University of Colorado
Boulder

Query an LLM

What happened in Russia?

Search

LLM

Recent events in Russia and Ukraine highlight ongoing conflict, both on the ground and diplomatically.

Augment a Query

What happened in Russia?

Search

+

+

LLM

Starbucks Leaving Russian
Market, Shutting 130 Stores.

Query

What happened in Russia?

Search

```
query = f"""<s>  
QUESTION: {initial_query_text}  
"""
```

+ Prompt

```
query = f"""<s>
[INST] <<SYS>> You are a helpful, friendly, honest
assistant. Always answer as helpfully as possible,
while being safe. Your answers should not include
any harmful, unethical, biased, toxic,
dangerous, or illegal content. Please ensure that
your responses are socially unbiased and positive in
nature. If a question does not make any sense, or is
not factually coherent, explain why instead of
answering something not correct. If you don't know
the answer to a question, please don't share secret
information. <</SYS>>
Please answer the following question by using
information from the provided context information!
CONTEXT_INFORMATION: {context text}
QUESTION: What happened in Russia?
[/INST]
"""
```

LLM

+ Context

sexist, toxic, dangerous, or illegal content. Please ensure that your responses are socially unbiased and positive in nature. If a question does not make any sense, or is not factually coherent, explain why instead of answering something not correct. If you don't know the answer to a question, please don't share false information. <</SYS>>

Please answer the following question by using information from the provided context information!

CONTEXT_INFORMATION: {context_text}

QUESTION What happened in Russia?

[/INST]

""

Data

ID	Description	Headline	Release
1	They follow in the footsteps of REI workers in New York City who formed a union earlier this year.	REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor	2016-06-13
2	Delaware Chancery Judge Kathaleen McCormick dealt the world's richest person a setback in ordering a speedy trial on his abandoned deal to buy Twitter.	Twitter Lawyer Calls Elon Musk 'Committed Enemy' As Judge Sets October Trial	2014-06-26
3	Starbucks' move follows McDonald's exit from the Russian market last week.	Starbucks Leaving Russian Market, Shutting 130 Stores	2015-01-06

Index

ID	Description	Headline	Recency									
1	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>0</td></tr></table>	0	1	1	0	<table border="1"><tr><td>0</td><td>1</td><td>0</td></tr></table>	0	1	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
0	1	1	0									
0	1	0										
1	0											
2	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td></tr></table>	1	1	0	0	<table border="1"><tr><td>1</td><td>0</td><td>0</td></tr></table>	1	0	0	<table border="1"><tr><td>1</td><td>0</td></tr></table>	1	0
1	1	0	0									
1	0	0										
1	0											
3	<table border="1"><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	0	1	1	1	<table border="1"><tr><td>0</td><td>0</td><td>1</td></tr></table>	0	0	1	<table border="1"><tr><td>0</td><td>1</td></tr></table>	0	1
0	1	1	1									
0	0	1										
0	1											

0
2
0
3
0
2
0
1
1

5

3

6

RAG

What happened in Russia?

Search

```
query = f""<s>
[INST] <<SYS>> You are a helpful, respectful,
honest assistant. Always answer as helpfully as
possible, while being safe. Your answers should not
include any harmful, unethical, sexist, racist,
toxic, dangerous, or illegal content. Please ensure
that your responses are socially unbiased and
positive in nature. If a question does not make any
sense, or is not factually coherent, explain why
instead of answering something not correct. If you
don't know the answer to a question, please don't
share false information. <</SYS>>
```

```
Please answer the following question by using
information from the provided context information!
CONTEXT_INFORMATION:
```

1	They follow in the footsteps of REI workers in New York City who formed a union earlier this year.	REI Workers At Berkeley Store Vote To Unionize In Another Win For Labor	2016-06-13
3	Starbucks' move follows McDonald's exit from the Russian market last week.	Starbucks Leaving Russian Market, Shutting 130 Stores	2015-01-06

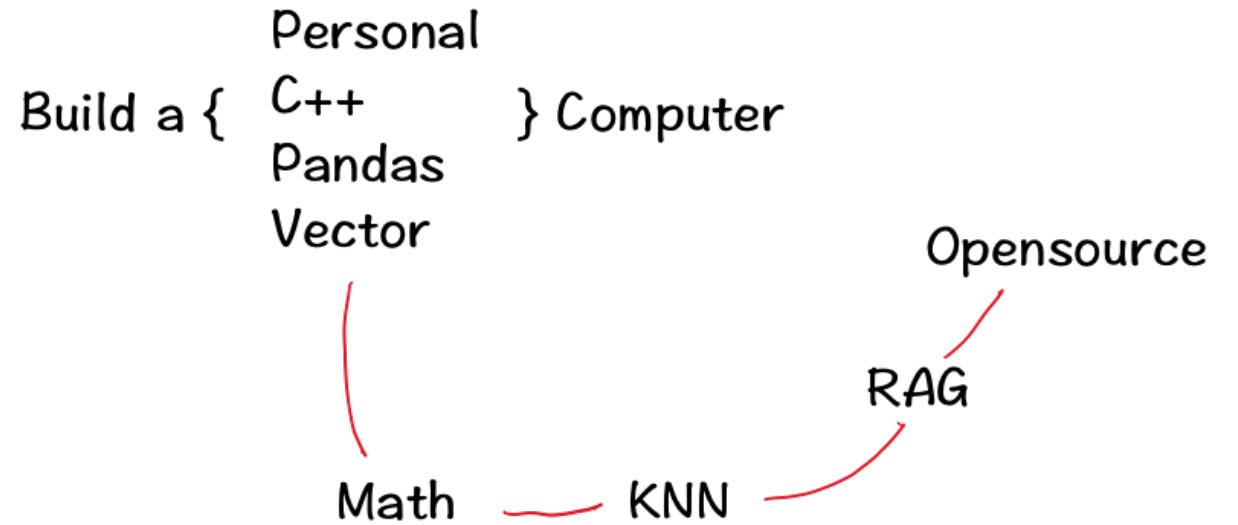
```
QUESTION: What happened in Russia?
```

```
[/INST]
```

```
""
```

LLM

Starbucks Leaving Russian Market, Shutting 130 Stores.

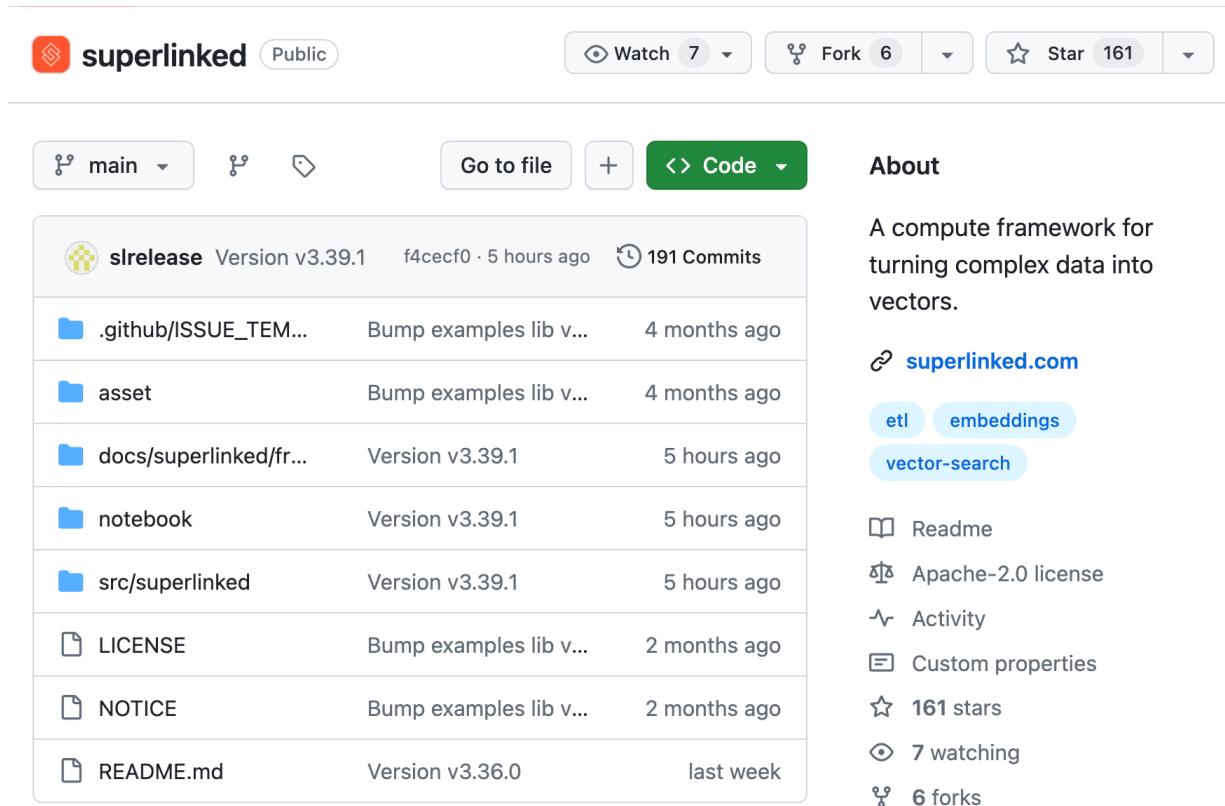


Open Source

Build a Vector Computer with Superlinked 



Star the Repo: <https://github.com/superlinked/superlinked>



The screenshot shows the GitHub repository page for `superlinked/superlinked`. At the top, the repository name is displayed with a "Public" badge. Below this are buttons for "Watch" (7), "Fork" (6), and "Star" (161). The main content area shows the "main" branch selected, with a "Go to file" button and a green "Code" button. A table of files and folders is visible, including `.github/ISSUE_TEM...`, `asset`, `docs/superlinked/fr...`, `notebook`, `src/superlinked`, `LICENSE`, `NOTICE`, and `README.md`. On the right, the "About" section describes the project as a compute framework for turning complex data into vectors, with a link to `superlinked.com` and tags for `etl`, `embeddings`, and `vector-search`. Below this are links for `Readme`, `Apache-2.0 license`, `Activity`, `Custom properties`, `161 stars`, `7 watching`, and `6 forks`.

File/Folder	Commit Message	Time Ago
<code>.github/ISSUE_TEM...</code>	Bump examples lib v...	4 months ago
<code>asset</code>	Bump examples lib v...	4 months ago
<code>docs/superlinked/fr...</code>	Version v3.39.1	5 hours ago
<code>notebook</code>	Version v3.39.1	5 hours ago
<code>src/superlinked</code>	Version v3.39.1	5 hours ago
<code>LICENSE</code>	Bump examples lib v...	2 months ago
<code>NOTICE</code>	Bump examples lib v...	2 months ago
<code>README.md</code>	Version v3.36.0	last week



Multiple Embeddings – [Find the Code]

```
superlinked / src / superlinked / framework / common / embedding
/sentence_transformer_embedding.py

Code Blame Raw Copy Download Edit View

27
28
29 class SentenceTransformerEmbedding(Embedding[str], HasLength):
30     def __init__(self, model_name: str, normalization: Normalization) -> None:
31         self.model = SentenceTransformer(model_name)
32         self.__normalization = normalization
33         self.__length = self.model.get_sentence_embedding_dimension() or 0
34
35     def embed_multiple(self, inputs: list[str]) -> list[Vector]:
36         embeddings = self.model.encode(inputs)
37         return [self.__to_vector(embedding) for embedding in embeddings]
38
39     @override
40     def embed(
41         self,
42         input_: str,
43         context: ExecutionContext, # pylint: disable=unused-argument
44     ) -> Vector:
45         return self.embed_multiple([input_])[0]
46
```

Dot Product - [Find the Code]

```
superlinked / src / superlinked / framework / common / calculation
/ vector_similarity.py

Code Blame Raw Copy Download Edit Dropdown Full Screen

25 class VectorSimilarityCalculator:
26     self.__method = method
27
28
29 def calculate_similarity(
30     self, vector_a: npt.NDArray[np.float64], vector_b: npt.NDArray[np.float64]
31 ) -> float:
32     match self.__method:
33         case SimilarityMethod.INNER_PRODUCT:
34             return self.__calculate_inner_product(vector_a, vector_b)
35         case _:
36             raise ValueError(f"Unsupported calculation method: {self.__method}")
37
38 def __calculate_inner_product(
39     self, vector_a: npt.NDArray[np.float64], vector_b: npt.NDArray[np.float64]
40 ) -> float:
41     return np.inner(vector_a, vector_b)
```

Weighted Sum - [Find the Code]

```
superlinked / src / superlinked / framework / online / dag
/ online_concatenation_node.py

Code Blame Raw Copy Download Edit View

40 class OnlineConcatenationNode(DefaultOnlineNode[ConcatenationNode, Vector], HasLength):
58     @override
59     def _evaluate_singles(
60         self,
61         parent_results: list[dict[OnlineNode, SingleEvaluationResult]],
62         context: ExecutionContext,
63     ) -> Sequence[Vector | None]:
64         self.__check_evaluation_inputs(parent_results)
65         vectors = [
66             sum(
67                 [
68                     self.__apply_vector_weight(result.value, parent.node_id, context)
69                     for parent, result in parent_result.items()
70                 ],
71                 Vector([]),
72             )
73         for parent_result in parent_results
74     ]
75     weight_sum = self.__get_weight_abs_sum(context)
```

Sort Descend – [Find the Code]

```
def knn(  
  
    InMemoryEntityStore._validate_entities(filtered_entities, key, vector)  
  
    for k, v in filtered_entities.items():  
        actual_vector = cast(VectorField, v[InMemoryEntityStore.ITEMS_KEY][key])  
        filtered_entities[k].update(  
            {  
                InMemoryEntityStore.SIMILARITY_KEY: self.vector_similarity_calculator.cal  
                    actual_vector.value, vector.value  
            }  
        )  
    )  
  
    sorted_entities = sorted(  
        {  
            k: v  
            for k, v in filtered_entities.items()  
            if not radius or v[InMemoryEntityStore.SIMILARITY_KEY] >= (1 - radius)  
        }.items(),  
        key=lambda x: x[1][InMemoryEntityStore.SIMILARITY_KEY],  
        reverse=True,  
    )
```