Chapter 1: Neural Networks for Neural Search

term dictionary		postings
term id	term	document id
1	best	1, 4, 7
2	blue	2, 4, 6, 11
3	forget	3, 5, 9
4	jina	1, 6, 17
5	need	4, 22, 123
...	...	...
n	zoo	1, 2, 4, 12

document_1
document_2
document_n

Preprocessing

vocabulary	best	blue	forget	jina	need	...	zoo
query	-	-	-	True	-	-	-
d1	True	-	-	True	-	-	True
d2	-	True	-	-	-	-	True
...	-	-	-	-	-	-	-
dn	-	-	-	-	-	-	True

vocabulary	best	blue	forget	jina	need	...	zoo
query	0	0	0	1	0	0	-
d1	0.0268	0	0	0.124	0	0	0.0155
d2	0	0.0335	0	0	0	0	1
...	0	0	0	0	0	0	0
dn	0	0	0	0	0	0	0.0155

training data

Feature extraction

train

model

ranked list
d7
d5
...
dn

predict

query

documents

Feature extraction

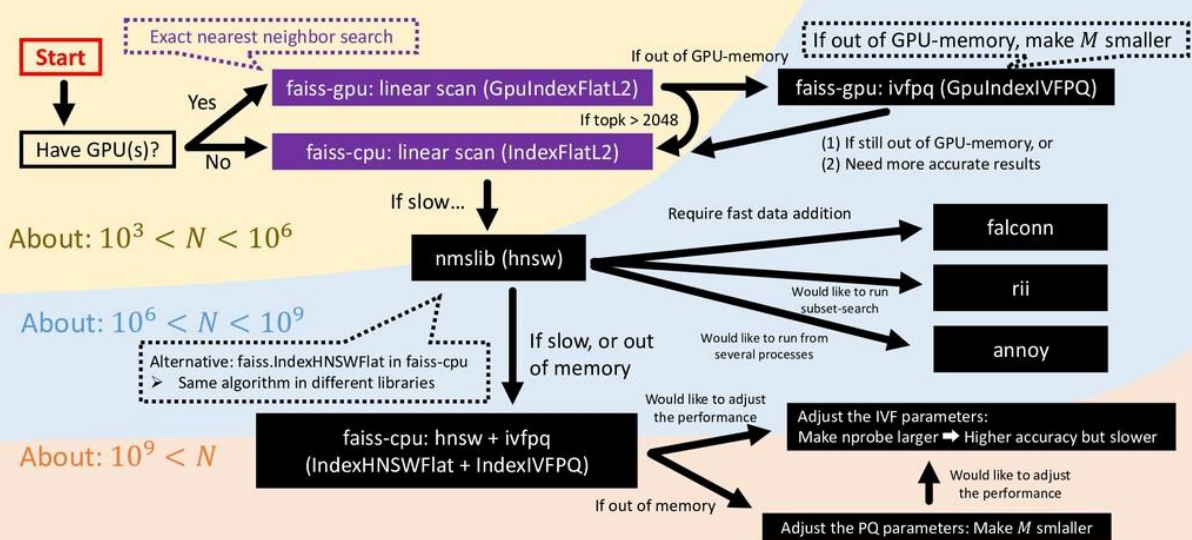**features**

Chapter 2: Introducing Foundations of Vector Representation

vec2
vec1
vec3

encoded_doc1
encoded_doc2

p
p2
q2
d(vec1, vec2)
q
p1
q1

p2
q2
p
q
d(vec1, vec2)
p1
q1

INPUT   PROJECTION   OUTPUT

INPUT   PROJECTION   OUTPUT

w(t-2)
w(t-1)
SUM
w(t)
w(t+1)
w(t+2)

CBOW

w(t)
w(t-2)
w(t-1)
w(t+1)
w(t+2)

Skip-gram

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{\#\#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}

Conv2D | Conv2D | MaxPool | Conv2D | Conv2D | MaxPool | Conv2D | Conv2D | Conv2D | MaxPool | Conv2D | Conv2D | Conv2D | MaxPool | Conv2D | Conv2D | Conv2D | MaxPool | Flatten | Dense | Dense | SoftMax

VGG16 Architecture
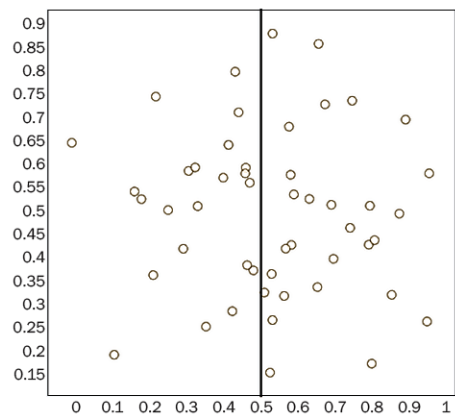
Chapter 3: System Design and Engineering Challenges
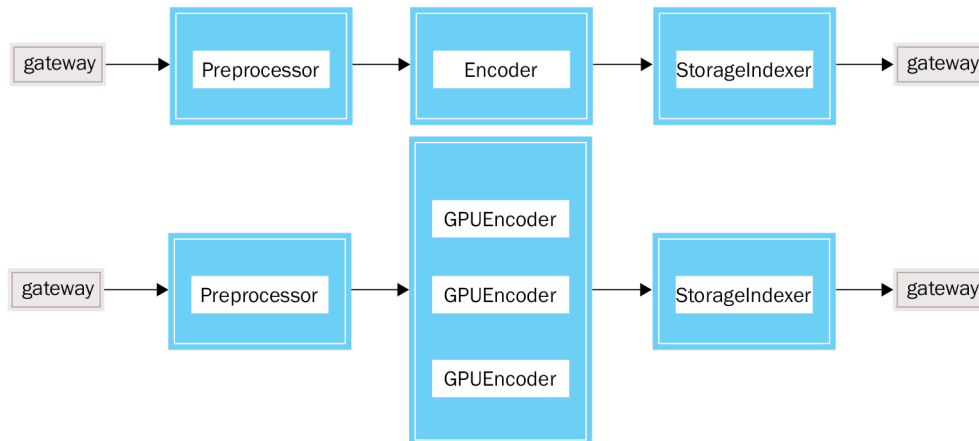
cheat-sheet for ANN in Python （as of 2020. Can be installed by conda or pip）

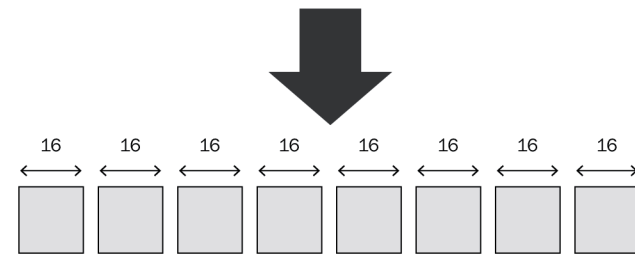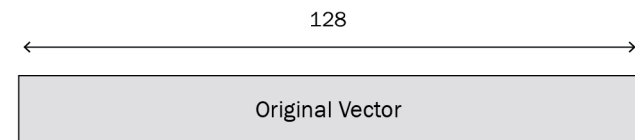**Start**

Have GPU(s)?

Exact nearest neighbor search

Yes → faiss-gpu: linear scan (GpuIndexFlatL2)

No → faiss-cpu: linear scan (IndexFlatL2)

If out of GPU-memory

If out of GPU-memory, make M smaller

faiss-gpu: ivfpq (GpuIndexIVFPQ)

If topk > 2048

(1) If still out of GPU-memory, or
(2) Need more accurate results

About: $10^3 < N < 10^6$

If slow…

nmslib (hnsw)

Require fast data addition → falconn

Would like to run subset-search → rii

Would like to run from several processes → annoy

About: $10^6 < N < 10^9$

Alternative: faiss.IndexHNSWFlat in faiss-cpu
➤ Same algorithm in different libraries

If slow, or out of memory

About: $10^9 < N$

faiss-cpu: hnsw + ivfpq
(IndexHNSWFlat + IndexIVFPQ)

Would like to adjust the performance

Adjust the IVF parameters:
Make nprobe larger ➡ Higher accuracy but slower

Would like to adjust the performance

If out of memory

Adjust the PQ parameters: Make M smlaller

Note: Assuming $D \cong 100$. The size of the problem is determined by DN. If $100 \ll D$, run PCA to reduce D to 100
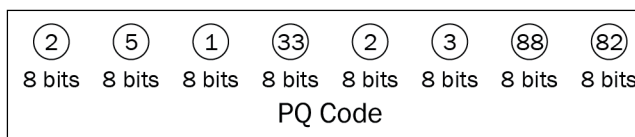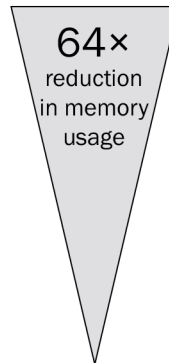
128

Original Vector

16 16 16 16 16 16 16 16
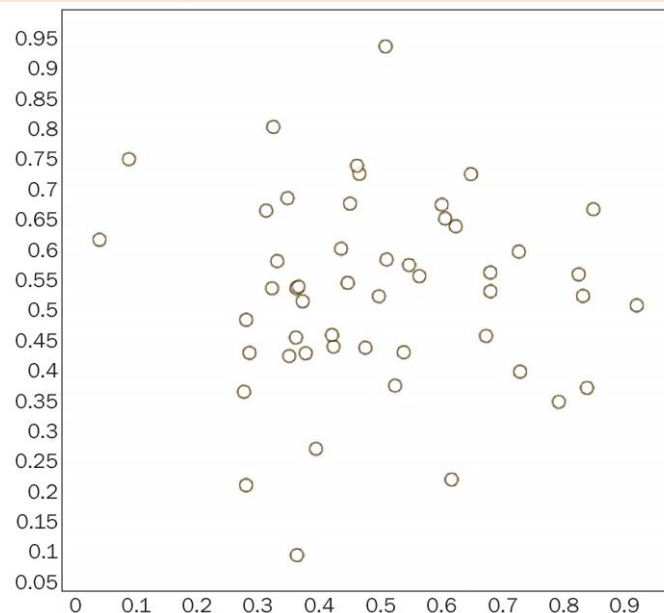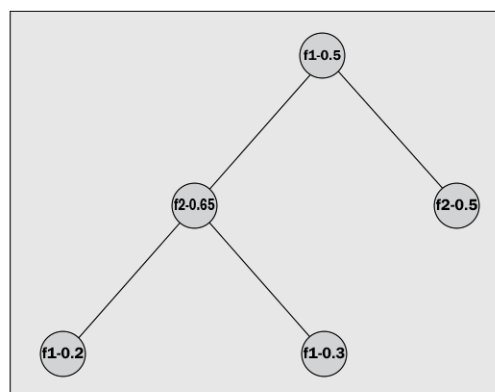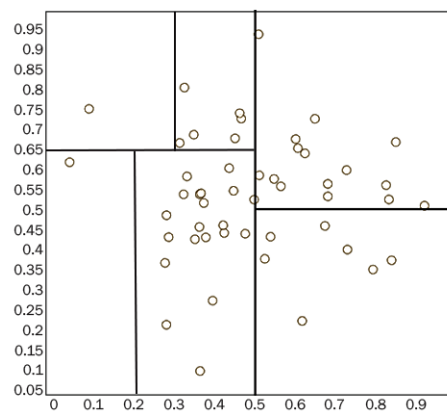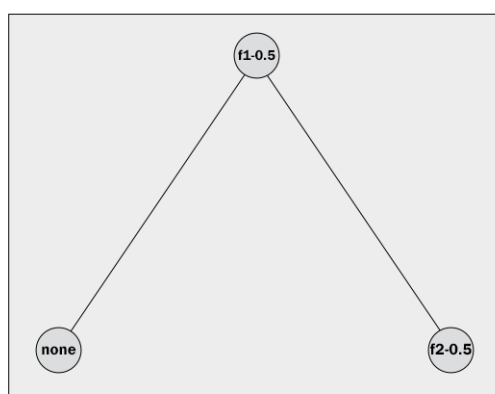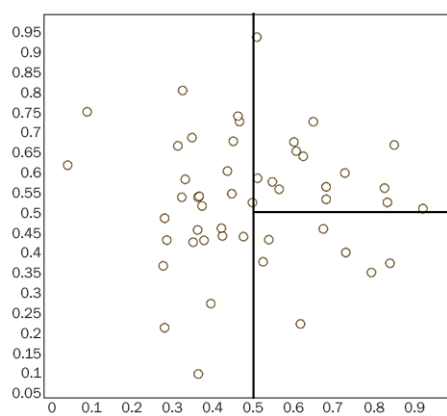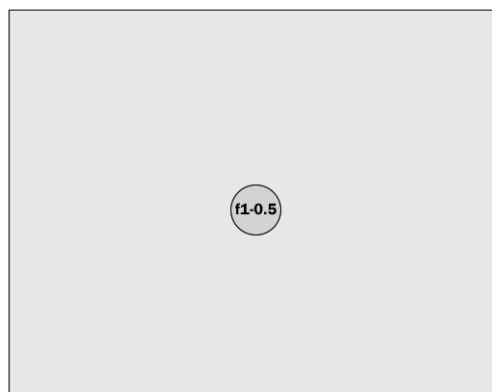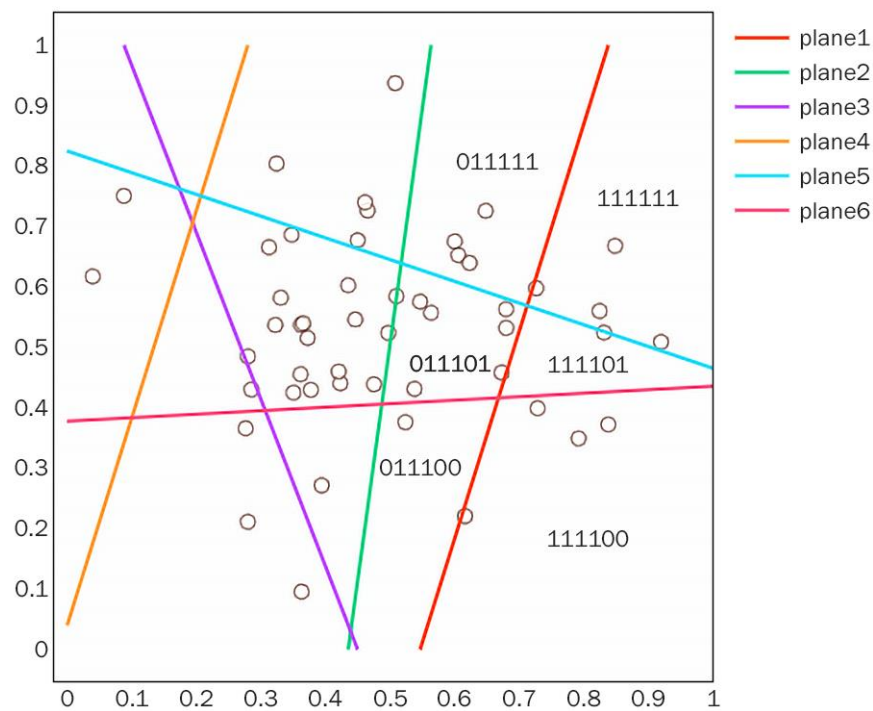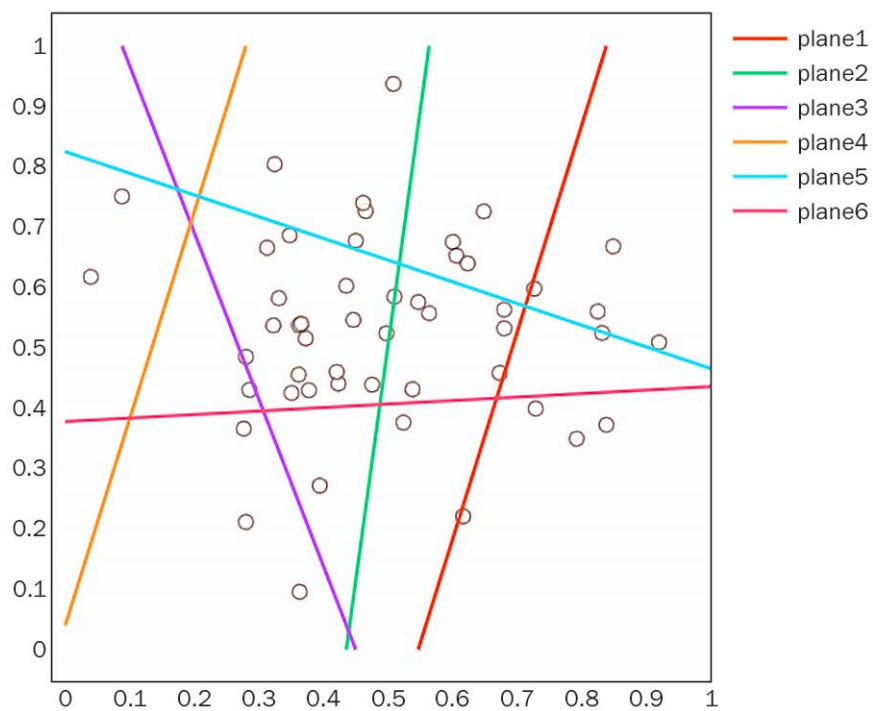
After Product Quantization

| ② | ⑤ | ① | ㉝ | ② | ③ | 88 | 82 |
| 8 bits | 8 bits | 8 bits | 8 bits | 8 bits | 8 bits | 8 bits | 8 bits |

PQ Code

128 × 32 bits
= 4096 bits
= 512 bytes

64×
reduction
in memory
usage

8 × 8 bits
= 64 bits
= 8 bytes

Chapter 4: Learning Jina's Basics

doc . content = oneof ( )

conversion
image-specific conversion
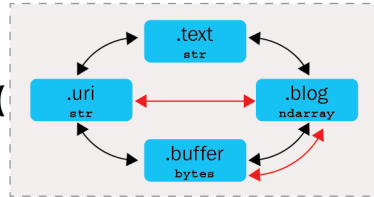
```python
from jina import Document

s = Document(
    id='🧩',
    content='hello-world',
    tags={'a': 'b'},
    chunks=[Document(id='🐢')],
)

d = Document(
    id='👢',
    content='goodbye-world',
    tags={'c': 'd'},
    chunks=[Document(id='🐱')],
)

# only update `id` field
d.update(s, fields=['id'])
# update all fields. `tags` field as `dict` will be merged.
d.update(s)
```

```
miaozhaofeng@miaozhafengsMBP:/tmp                    ⌥⌘2
》
》 jina hub new
─────────── Create New Executor ───────────
┌────────────────────────────────────────────────────────┐
│ Executor is how Jina processes Document.               │
│                                                         │
│ This guide helps you to create your own Executor in 30 seconds. │
└────────────────────────────────────────────────────────┘
? What is the name of your executor?
CamelCase is required (MyExecutor73):
? Which folder to store your executor? (/private/tmp/MyExecutor73):
That's all we need to create an Executor!
? Or do you want to proceed to advanced configuration [y/n] (n):
Creating MyExecutor73... ──────────────────────── 100% 0:00:00
─────────────── 🎉 Next Steps ───────────────
Congrats! You have successfully created an Executor! Here are the next steps:
───────────── 1. Checkout the Generated Executor ─────────────
┌────────────────────────────────────────────────────────┐
│ 1 cd /private/tmp/MyExecutor73                          │
│ 2 ls                                                    │
└────────────────────────────────────────────────────────┘
───────────── 2. Understand Folder Structure ─────────────

   Filename    Description

   executor…   The main logic file of the Executor.
   config.y…   The YAML config file of the Executor. You can define __init__ arguments using with …
                  ─────────────── config.yml ───────────────
                  ┌──────────────────────────────────────┐
                  │ 1                                     │
                  │ 2 jtype: MyExecutor73                 │
                  │ 3 with:                               │
```

Chapter 5: Multiple Search Modalities

📄 **Document:** 8feab625a4bfc69bc2eb131f55832741

Attribute	Value
text	नमस्ते दुनिया !你好世界 !こんにちは世界 !Привет мир!

└── 🔷 **Chunks**

  ├── 📄 **Document:** bf1d21752499949254377bdf76ce1690

Attribute	Value
text	नमस्ते दुनिया

  ├── 📄 **Document:** 36222844b1162748f3c6084b2b5ed183

Attribute	Value
text	你好世界

  ├── 📄 **Document:** e2ab77c5e7ddd691e1efb329fa10be3e

Attribute	Value
text	こんにちは世界

  ├── 📄 **Document:** 176d9cea7c012d57b179260527667d53

Attribute	Value
text	Привет мир

  └── 📄 **Document:** ea9297436246091801546a45ade20d2a
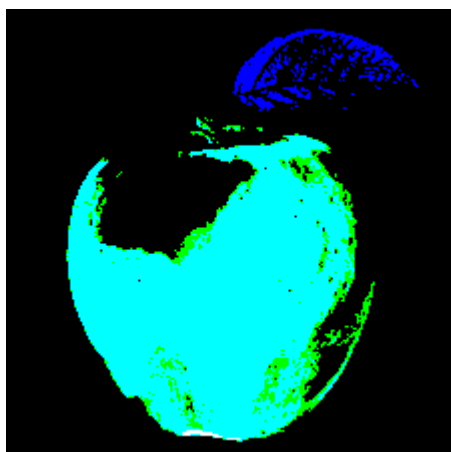
Men Town Round Red Neck T-Shirts

Document: af9870270c2763f73aff8109044c8330
└── ◈ **Chunks**
   ├── **Document:** 9eaaf37b747fdd60e56c30e28d38c212

Attribute	Value
parent_id granularity text	af9870270c2763f73aff8109044c8330 1 Men Town Round Red Neck T-Shirts

   └── **Document:** ebc19885ebae9dba9eadcb7588bda3e4

Attribute	Value
parent_id granularity tensor mime_type uri	af9870270c2763f73aff8109044c8330 1 **<class** 'numpy.ndarray'**>** in shape **(245**, **200**, **3)**, dtype: uint8 image/jpeg red_t_shirt.jpg

Relevant Document · Query
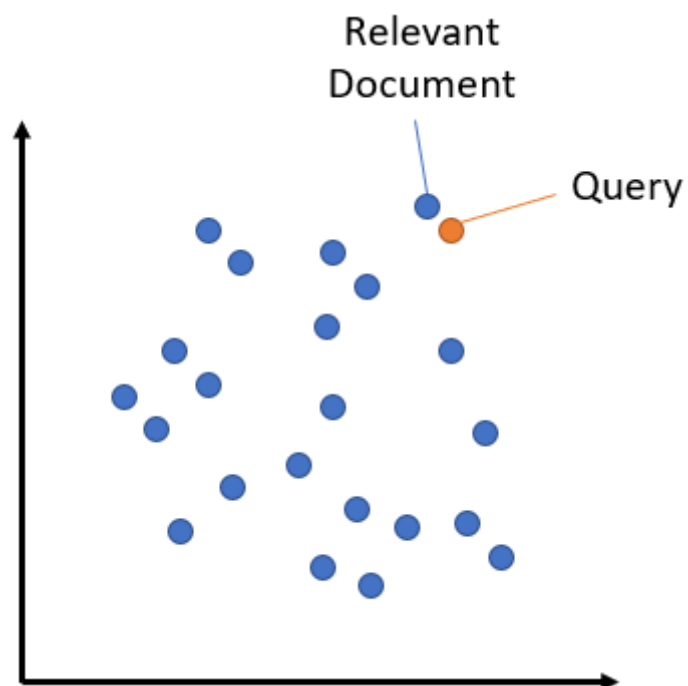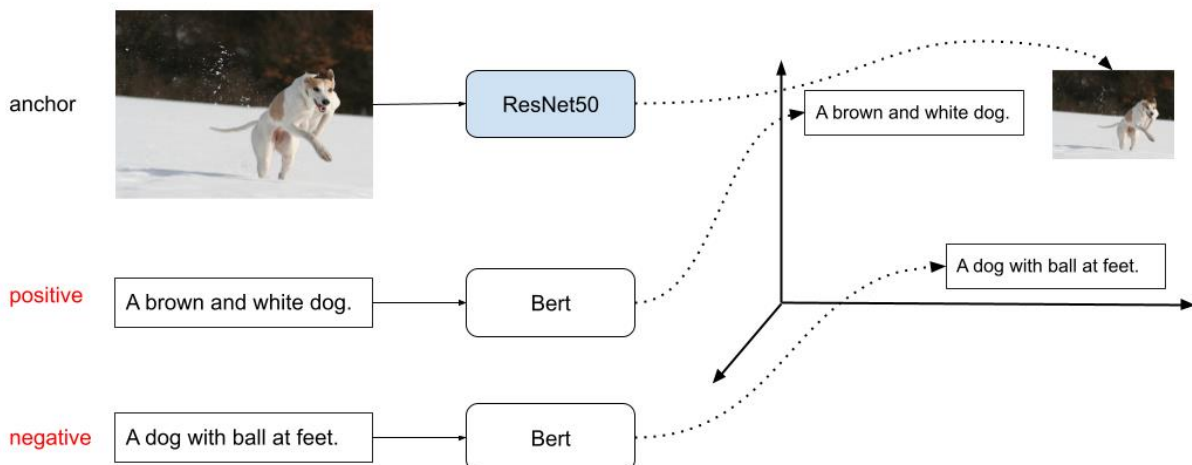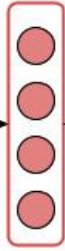
A black and white dog is running through the grass.

Visual feature
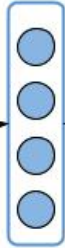
Text feature

Fusion feature

Chapter 6: Basic Practical Examples with Jina


```
/Users/hanxiao/.pyenv/versions/3.7.9/bin/jina hello chatbot
                    cli = hello
         download-proxy = None
                  hello = chatbot
        index-data-url = https://static.jina.ai/chatbot
               parallel = 2
            port-expose = 8080
     unblock-query-flow = False
                workdir = 22f6f821-63ff-4429-9e7e-342f55

download csv data ──────────────────────────── 0:00:00 167.9 step/s    2 steps do
       pod0/head@32366[L]:ready and listening
      pod0/pea-0@32366[L]:ready and listening
      pod0/pea-1@32366[L]:ready and listening
       pod0/tail@32366[L]:ready and listening
           pod1@32366[L]:ready and listening
        gateway@32366[L]:ready and listening
          Flow@32366[I]:🎉 Flow is ready to use!
       🔗 Protocol:          GRPC
       🏠 Local access:       0.0.0.0:54511
       🔒 Private network:    192.168.178.31:54511
       🌐 Public address:     217.70.138.123:54511
Working... ─────────────────────────────── -:--:-- estimating...
```

COVID-19 SIMPLE QA
POWERED BY JINA

Hi there, please ask me COVID-19 related questions. For example,

What is Covid-19?

13:47

What is covid????

Should I worry about COV...

Illness due to COVID-19 infection is generally mild, especially for children and young adults. However, it can cause serious illness: about 1 in every 5 people who catch it need hospital care. It is therefore quite normal for people to worry about how the COVID-19 outbreak will affect them and their loved ones. We can channel our concerns into actions to protect ourselves, our loved ones and our communities. First and foremost among these actions is regular and thorough hand-washing and good respiratory hygiene. Secondly, keep informed and follow the advice of the

Type message...   SEND

gateway → MyTransformer → MyIndexer → gateway

```
/Users/hanxiao/.pyenv/versions/3.7.9/bin/jina hello fashion
                            cli = hello
                 download-proxy = None
                          hello = fashion
                 index-data-url = http://fashion-mnist.s3-websit
               index-labels-url = http://fashion-mnist.s3-websit
                      num-query = 128
                 query-data-url = http://fashion-mnist.s3-websit
               query-labels-url = http://fashion-mnist.s3-websit
                          top-k = 50
                        workdir = 4bfd4b21-4553-4330-a7aa-75e2a5
download fashion-mnist ──────────────────────── 0:00:01 1926.2 step/s 39 steps done in 1 second
     pod0/head@32044[L]:ready and listening
    pod0/pea-0@32044[L]:ready and listening
    pod0/pea-1@32044[L]:ready and listening
     pod0/tail@32044[L]:ready and listening
         pod1@32044[L]:ready and listening
         pod2@32044[L]:ready and listening
      gateway@32044[L]:ready and listening
        Flow@32044[I]:🎉 Flow is ready to use!
     🔗 Protocol:            GRPC
     🏠 Local access:        0.0.0.0:53182
     🔒 Private network:     192.168.178.31:53182
     🌐 Public address:      217.70.138.123:53182
Working... ──────────────────────── 0:00:14 40.2 step/s 601 steps done in 14 seconds
Working... ──────────────────────── 0:00:07 0.4 step/s 3 steps done in 7 seconds
```

🔍 jina

What just happened?

This is Jina's hello-world, end-to-endly. It downloads Fashion-MNIST dataset and indexes 60,000 images via Jina search framework. The index is stored into multiple *shards*. We then randomly sample unseen images as *Queries*, ask Jina to retrieve relevant results. Below is Jina's retrievals, where the left-most column is query image.
Intrigued? Learn more about Jina and checkout our Github!

- -

Precision@50: 72.70% Recall@50: 0.61%

- -

Query Top-K Results
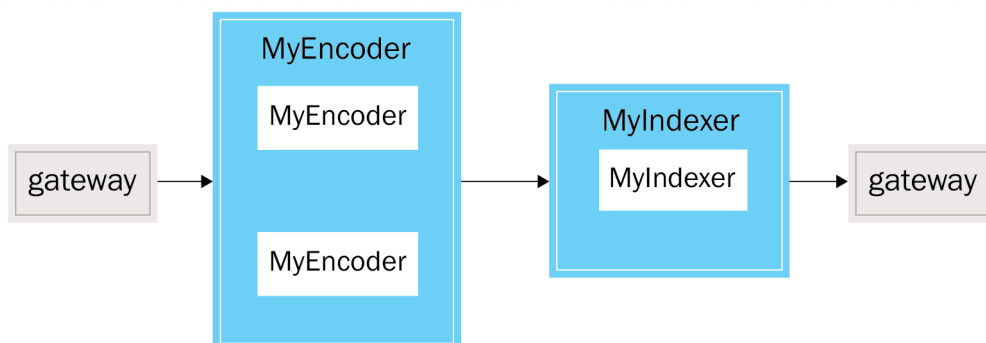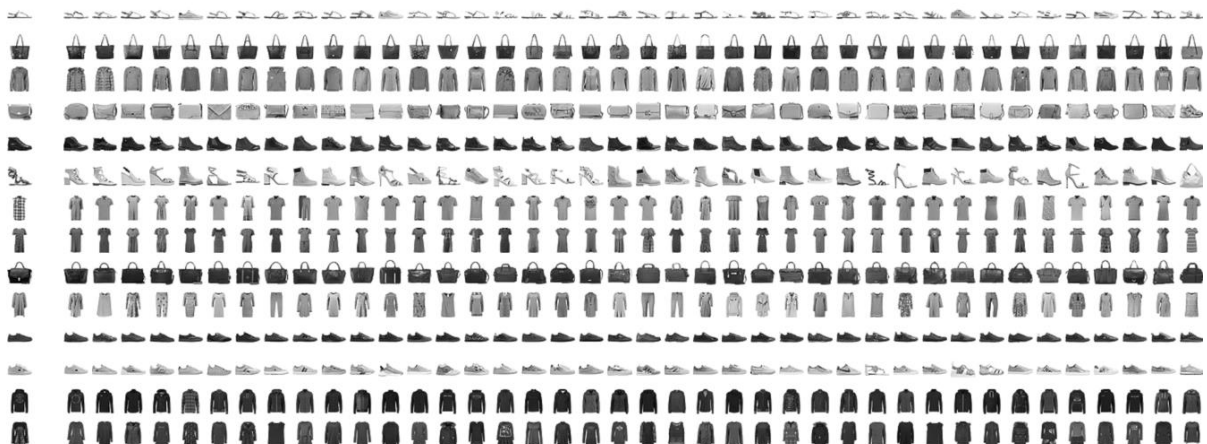
```
/Users/hanxiao/.pyenv/versions/3.7.9/bin/jina hello multimodal
                         cli = hello
              download-proxy = None
                       hello = multimodal
              index-data-url = https://static.jina.ai/multimo
                 port-expose = 8080
          unblock-query-flow = False
                     workdir = b7518349-5b91-4255-be7d-1d4857
```

download zip data ──────────────────────── 0:00:16 676.4 step/s   112
        segment@32747[L]:ready and listening
      craftText@32747[L]:ready and listening
     encodeText@32747[L]:ready and listening
    textIndexer@32747[L]:ready and listening
     craftImage@32747[L]:ready and listening
    encodeImage@32747[L]:ready and listening
    imageIndexer@32747[L]:ready and listening
keyValueIndexer@32747[L]:ready and listening
        joinAll@32747[L]:ready and listening
        gateway@32747[L]:ready and listening
        Flow@32747[I]:🎉 Flow is ready to use!
        🔗 Protocol: GRPC
        🏠 Local access: 0.0.0.0:54775
        🔒 Private network: 192.168.178.31:54775
        🌐 Public address: 217.70.138.123:54775
Working... ──────────────────────── 0:00:22 0.8 step/s
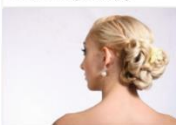
🔍 Jina Multimodal Hello-World   http://localhost:8080
```

A multimodal document as query (Refresh for new image)

Young Woman Posing As Long Curly Hair Covering Her Face in Toronto, Canada

Results rely more on

Text 0.5 —————●————— Image 0.5

**Search!**

REST Payload

```
{
  "data": [
    {
      "id": "eff68a32-0aab-4df3-89db-da0d857f562f"
      "chunks": [
        {
          "id": "147ebbf2-114c-487c-87b0-1eea2e6e5
```

Young Woman Posing As Long Curly Hair Covering Her Face in Toronto, Canada	Young Woman in Sunglasses, White Spaghetti Strap Crop Top and Blue Bell Bottom Posing Outside Village Houses	Woman with fake eyebrows, moustache and nose	Red haired young woman with lip piercing looks at camera	Portrait of a young Caucasian woman covering face with her hair by looking over her shoulder while bending forward
score: 1	score: 0.66070336	score: 0.3918237	score: 0.38611826	score: 0.383623
filename: image_1698.jpg	filename: image_2174.jpg	filename: image_1202.jpg	filename: image_51.jpg	filename: image_2483.jpg
Rear view of a young Caucasian woman with a beautiful hair style	Fluttering hair of a young blonde woman wearing red dress cover her face	A young African woman with curly hair posing outdoors on winter day	A young African woman with curly hair posing outdoors	Exitced young pretty woman headshot.
score: 0.37856704	score: 0.3671896	score: 0.36390376	score: 0.36245924	score: 0.3607232
filename: image_1922.jpg	filename: image_2140.jpg	filename: image_2167.jpg	filename: image_2336.jpg	filename: image_61.jpg

Chapter 7: Exploring Advanced Use Cases of Jina

Best matches for 'a black dog and a spotted dog are fighting'

score=0.8351

score=0.6858


```
              storage_encoder                            indexer
gateway →  [ jinahub+docker://FlairTextEncoder/test-3.0 ] → [ jinahub://HNSWPostgresIndexer/v0.8 ] →  gateway
```