

# hybrid-search

November 21, 2024

```
[6]: # Import client library
from qdrant_client import QdrantClient
from sentence_transformers import SentenceTransformer, util
from PIL import Image

import numpy as np
import matplotlib.pyplot as plt
```

```
[7]: client = QdrantClient(url="http://qdrant.mavcode.io:6333", api_key="ADD KEY" )
```

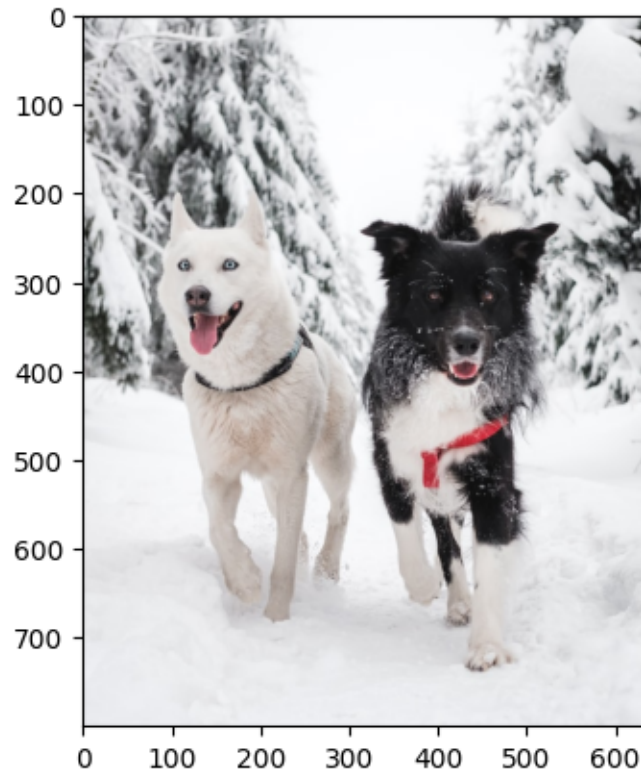
```
/home/mvasquez/opt/miniconda3/envs/torch-2.5.1/lib/python3.12/site-
packages/qdrant_client/qdrant_remote.py:130: UserWarning: Api key is used with
an insecure connection.
```

```
warnings.warn("Api key is used with an insecure connection.")
```

```
[10]: image = Image.open('two_dogs_in_snow.jpg')
```

```
[11]: plt.imshow(image)
```

```
[11]: <matplotlib.image.AxesImage at 0x7f8c501fd790>
```



```
[15]: client.set_model("sentence-transformers/all-MiniLM-L6-v2")
```

```
Fetching 5 files: 0%|          | 0/5 [00:00<?, ?it/s]
```

```
[16]: # comment this line to use dense vectors only
      client.set_sparse_model("prithivida/Splade_PP_en_v1")
```

```
Fetching 5 files: 0%|          | 0/5 [00:00<?, ?it/s]
```

```
[20]: if not client.collection_exists("startups"):
      client.create_collection(
          collection_name="startups",
          vectors_config=client.get_fastembed_vector_params(),
          # comment this line to use dense vectors only
          sparse_vectors_config=client.get_fastembed_sparse_vector_params(),
      )
```

```
[17]: import json

      payload_path = "startups_demo.json"
      metadata = []
      documents = []
```

```

with open(payload_path) as fd:
    for line in fd:
        obj = json.loads(line)
        documents.append(obj.pop("description"))
        metadata.append(obj)

```

```

[ ]: from tqdm import tqdm
client.add(
    collection_name="startups",
    documents=documents[1000:2000],
    metadata=metadata[1000:2000],
    parallel=16,
    ids=tqdm(range(len(documents))),
)

```

```

[23]: from qdrant_client import QdrantClient

class HybridSearcher:
    DENSE_MODEL = "sentence-transformers/all-MiniLM-L6-v2"
    SPARSE_MODEL = "prithivida/Splade_PP_en_v1"
    def __init__(self, collection_name):
        self.collection_name = collection_name
        # initialize Qdrant client
        self.qdrant_client = QdrantClient(url="http://qdrant.mavcode.io:6333",
        ↪api_key="ADD KEY" )
        self.qdrant_client.set_model(self.DENSE_MODEL)
        # comment this line to use dense vectors only
        self.qdrant_client.set_sparse_model(self.SPARSE_MODEL)

    def search(self, text: str):
        search_result = self.qdrant_client.query(
            collection_name=self.collection_name,
            query_text=text,
            query_filter=None, # If you don't want any filters for now
            limit=5, # 5 the closest results)
            # `search_result` contains found vector ids with similarity scores
            # along with the stored payload

            # Select and return metadata
            metadata = [hit.metadata for hit in search_result]
            return metadata

```

```

[24]: hybrid_searcher = HybridSearcher(collection_name="startups")

```

```

[25]: hybrid_searcher.search("Two dogs in the snow")

```

```
[25]: [{ 'link': 'http://www.kumfytailz.com',
  'document': 'Revolutionary Warming/Cooling Dog Apparel\nKumfyTailz combines
well-known gel-pack technology (think human therapy applications, microwave or
freeze it and put it on your knee or shoulder) with premium-quality dog
harnesses and coats. We keep pets comfortable in any weather, for more outside
play/walk ...',
  'images': 'https://d1qb2nb5cznatu.cloudfront.net/startups/i/257434-
d3fc861a4401e34169f3da806ef2192b-thumb_jpg.jpg?buster=1377799352',
  'city': 'Chicago',
  'name': 'Kumfy Tailz',
  'alt': 'Kumfy Tailz - pets'},
{'document': 'Top Dog in Online Pet Product Discovery, acquired by FamilyPet in
2015\ndoggyloot is a super-simple eCommerce platform that sells curated pet
products to a highly engaged customer base. We help dog-loving customers
discover top-quality products every day at the most competitive prices with our
daily emails, and make it convenient for ...',
  'link': 'http://www.doggyloot.com',
  'city': 'Chicago',
  'alt': 'Doggyloot - e-commerce consumer goods retail pets',
  'images': 'https://d1qb2nb5cznatu.cloudfront.net/startups/i/28002-
cbe0f9a8c2bce0f11a83575cd002c305-thumb_jpg.jpg?buster=1421987208',
  'name': 'Doggyloot'},
{'name': 'Betel Nut Games',
  'city': 'Chicago',
  'document': 'We make mildly addicting and highly stimulating games you should
play\nDog Park is an original arcade puzzler that puts you in the role of a
variety of dogs attempting to become the top dog of a dog park. With each park
you dominate, you add more dogs to your pack, each with their own style of
play.\nMore original games on the way ...',
  'alt': 'Betel Nut Games - ',
  'link': 'http://www.betelnutgames.com/',
  'images': 'https://d1qb2nb5cznatu.cloudfront.net/startups/i/85473-
5d21b6faf1f45067c1a3a45e7c75f365-thumb_jpg.jpg?buster=1338998117'},
{'images': 'https://d1qb2nb5cznatu.cloudfront.net/startups/i/541010-
fd7728034becdea19bef742b586def10-thumb_jpg.jpg?buster=1416685118',
  'link': 'http://www.pupjoy.com',
  'city': 'Chicago',
  'name': 'PupJoy',
  'alt': 'PupJoy - e-commerce pets subscription businesses Pet Care',
  'document': "Helping Dog Lovers Treat Better. Raising the Bar on the
Subscription Box.\nFor your dogs' treat and toy needs, we are a better
subscription box, focused on customization and outstanding service. When you
want to outfit them with cool, quality accessories, visit our online boutique.
We provide you with better products (like healthy, clear ..."}],
{'images': 'https://d1qb2nb5cznatu.cloudfront.net/startups/i/160681-
b1bc028083b8517223ff6324e883a4ed-thumb_jpg.jpg?buster=1359242185',
  'document': 'Snowboard/Ski Contest Events\nsnowboarding/skiing competitions 12
```

```
months a year. looking to take snow sports to the next level',  
  'name': 'World Snow Tour',  
  'city': 'Chicago',  
  'alt': 'World Snow Tour - television events',  
  'link': 'http://None'}]
```

[ ]: