Explorer

October 10, 2023

```
[1]: from testbed import load_query, K, D, evaluate, NUM_THREADS
     from testbed.test_diskann import build_diskann
     from testbed.test_hnswlib import build_hnswlib
     from testbed.test_faiss import build_faiss

     import hnswlib as hnsw
     import faiss
     import diskannpy as diskann
     import numpy as np
```

```
[2]: queries, gts = load_query()
```

1 HNSW

```
[3]: %%time
     build_hnswlib()
```

```
== HNSW ==
Time taken: 4.057326900233333 minutes
Memory difference: 1.619890176 GB
Memory at end: 2.43755008 GB
CPU times: user 32min 21s, sys: 2.62 s, total: 32min 24s
Wall time: 4min 5s
```

```
[4]: hnswlib_idx = hnsw.Index("l2", D)
     hnswlib_idx.load_index("/home/nawat/muic/senior/anns-war/indices/hnswlib/index.
       ↪bin")
```

```
[5]: %%time
     hnsw_results, _ = hnswlib_idx.knn_query(queries, K)
```

```
CPU times: user 17.8 s, sys: 251 ms, total: 18 s
Wall time: 854 ms
```

```
[6]: hnsw_results
```

```
[6]: array([[932085, 934876, 561813, …, 656997, 871600, 390777],
            [413247, 413071, 706838, …, 538688, 815172, 690476],
```

```
      [669835, 408764, 408462, …, 310475, 971815, 957355],
      …,
      [123855, 123351, 534149, …, 882928,  61007,  99059],
      [755327, 755323, 840765, …, 595134, 601257, 172180],
      [874343, 464509, 413340, …, 419949, 735660, 265726]],
     dtype=uint64)
```

[7]: `evaluate(hnsw_results, gts)`

[7]: 0.9776630000000001

2 FAISS

[8]:
```
%%time
build_faiss()
```

```
== FAISS-based HNSW ==
Time taken: 4.395543714483334 minutes
Memory difference: 1.54814464 GB
Memory at end: 4.103929856 GB
CPU times: user 34min 2s, sys: 1.67 s, total: 34min 4s
Wall time: 4min 25s
```

[9]:
```
faiss_idx = faiss.read_index("/home/nawat/muic/senior/anns-war/indices/faiss/
  ↪index.bin")
```

[10]:
```
%%time
_, faiss_results = faiss_idx.search(queries, K)
```

```
CPU times: user 1.73 s, sys: 11 µs, total: 1.73 s
Wall time: 234 ms
```

[11]: `faiss_results`

[11]:
```
array([[932085, 934876, 561813, …, 565750, 988993, 522282],
       [413247, 413071, 706838, …, 214320, 502137, 134091],
       [669835, 408764, 408462, …, 233819, 250457, 851856],
       …,
       [123855, 123351, 534149, …, 867625,  20062, 575366],
       [755327, 755323, 840765, …, 397414, 860426, 626711],
       [874343, 464509, 413340, …, 768793, 414057, 469001]])
```

[12]: `evaluate(faiss_results, gts)`

[12]: 0.683315

3 DiskANN

```
[13]: %%time
      build_diskann()
```

```
L2: Using AVX2 distance computation DistanceL2Float
Using only first 1000000 from file..
Starting index build with 1000000 points…
90% of index build completed.Starting final cleanup..done. Link time: 137.509s
Index built with degree: max:70  avg:68.7787  min:7  count(deg<2):0
Not saving tags as they are not enabled.
Time taken for save: 0.825731s.
L2: Using AVX2 distance computation DistanceL2Float
== DiskANN ==
Time taken: 2.32362170995 minutes
Memory difference: 0.511905792 GB
Memory at end: 5.037744128 GB
Passed, empty build_params while creating index config
From graph header, expected_file_size: 279114680, _max_observed_degree: 70,
_start: 123742, file_frozen_pts: 0
Loading vamana graph /home/nawat/muic/senior/anns-
war/indices/diskann/ann…done. Index has 1000000 nodes and 68778664 out-edges,
_start is set to 123742
Num frozen points:0 _nd: 1000000 _start: 123742 size(_location_to_tag): 0
size(_tag_to_location):0 Max points: 1000000
CPU times: user 18min 15s, sys: 2.14 s, total: 18min 17s
Wall time: 2min 19s
```

```
/home/nawat/.cache/pypoetry/virtualenvs/anns-war-4cghLm-Y-
py3.10/lib/python3.10/site-packages/diskannpy/_common.py:230: UserWarning: The
number of vectors in the saved index exceeds the max_vectors parameter.
max_vectors is being adjusted to accommodate the dataset, but any insertions
will fail.
  warnings.warn(
/home/nawat/.cache/pypoetry/virtualenvs/anns-war-4cghLm-Y-
py3.10/lib/python3.10/site-packages/diskannpy/_common.py:236: UserWarning: The
number of vectors in the saved index equals max_vectors parameter. Any
insertions will fail.
  warnings.warn(
```

```
[14]: diskann_idx = diskann.StaticMemoryIndex("/home/nawat/muic/senior/anns-war/
      ↪indices/diskann",
                                              distance_metric="l2",
                                              num_threads=NUM_THREADS,
                                              initial_search_complexity=125,
                                              vector_dtype=np.float32
                                              )
```

```
L2: Using AVX2 distance computation DistanceL2Float
```

```
Passed, empty build_params while creating index config
From graph header, expected_file_size: 279114680, _max_observed_degree: 70,
_start: 123742, file_frozen_pts: 0
Loading vamana graph /home/nawat/muic/senior/anns-
war/indices/diskann/ann…done. Index has 1000000 nodes and 68778664 out-edges,
_start is set to 123742
Num frozen points:0 _nd: 1000000 _start: 123742 size(_location_to_tag): 0
size(_tag_to_location):0 Max points: 1000000
```

[15]:
```python
%%time
diskann_result = diskann_idx.batch_search(queries, K, 125,␣
 ↪num_threads=NUM_THREADS).identifiers
```

```
CPU times: user 9.43 s, sys: 13 µs, total: 9.43 s
Wall time: 1.18 s
```

[16]:
```python
diskann_result
```

[16]:
```
array([[932085, 934876, 561813, …, 931721, 989762, 929750],
       [413247, 413071, 706838, …, 846198, 987074, 538688],
       [669835, 408764, 408462, …, 310475, 971815, 937903],
       …,
       [123855, 123351, 534149, …,  90175, 685486, 416474],
       [755327, 755323, 840765, …, 595134, 601257, 172180],
       [874343, 464509, 413340, …, 360985, 419949, 223427]],
      dtype=uint32)
```

[17]:
```python
evaluate(diskann_result, gts)
```

[17]: 0.993302