BANG Library

BANG Search can be consumed via this library. The graph index to be generated using DiskANN (https://github.com/microsoft/DiskANN/blob/main/workflows/SSD_index.md)

1. Dataset Generation

Download the dataset file in .bin format from big-ann-benchmarks (https://github.com/harsha-simhadri/big-ann-benchmarks/blob/main/neurips21/t3/README.md)

```
python create_dataset.py --dataset bigann-10M
```

2. Graph Construction

Construct the graph index and compressed vectors using DiskANN/Vamana. The compression factor can be controlled by the '-B' parameter. The higher the value lower is the compression. Set this to the memory on the GPU that can be used to store compressed vectors.

```
./build_disk_index  --data_type  uint8 --dist_fn l2 --data_path
/mnt/ssd_volume/big-ann-
benchmarks/data/bigann/base.1B.u8bin.crop_nb_10000000 --
index_path_prefix sift10m_index -R 64 -L 200 -B 70 -M 48
```

Run a python script to extract required metadata about the dataset using a script provided in BANG repo.

```
python bang_preprocess.py /mnt/ssd_volume/diskANN-
working/build/tests/sift10m_index_disk.index
/mnt/ssd_volume/diskANN-working/build/tests/sift10m_index_disk.bin
128 1 64
```

We construct the groundtruth in the required format by BANG using DiskANN

```
/compute_groundtruth --data_type uint8 --dist_fn l2 --base_file
/mnt/ssd_volume/big-ann-
benchmarks/data/bigann/base.1B.u8bin.crop_nb_10000000 --query_file
/mnt/ssd_volume/big-ann-benchmarks/data/bigann/bigann-10M --K 10 --
gt_file /mnt/ssd_volume/diskANN-
working/build/tests/sift10m_groundtruth.bin
```

Now, we are ready to start the BANG search.

3. BANG Search.

Download the code from BANG Repo : https://github.com/karthik86248/BANG-Billion-Scale-ANN

Navigate to BANG_Base directory.

Build the code using

```
make bang driver
```

Run the search using. Provide various values for worklist length when prompted via the console. The

```
./driver /mnt/ssd_volume/diskANN-working/build/tests/sift10m_index
/mnt/ssd_volume/big-ann-
benchmarks/data/bigann/query.public.10K.u8bin
/mnt/ssd_volume/diskANN-working/build/tests/sift10m_groundtruth.bin
10000 10
```

values could be in the range 10 to 152 (assuming recall parameter used is 10)