**Approximate Nearest Neighbour Search**

1. Cross-Platform Compilation

Can be compiled on Windows, Linux and MacOS.

1. High Performance Implementation on CPU

Parallel loop which iterates through the layers of Graph.

Enhance the efficiency of certain operations that involve searching for neighbors and modifying the graph structure.

Neighbour Search and Graph Modification: Within each layer, several tasks are performed simultaneously for different nodes in the graph. The “for” loop within the parallel section conducts tasks such as searching for neighbours, adding edges, and adjusting the graph structure.

1. High Performance Implementation on GPU

Cosine Similarity with GPU for benchmarking using CUDA

1. Illegal Input Handling

Should be able to open the Input file.

The first line of the input file must have Dimensions, Base Vectors and Number of Query Vectors

Actual Base and Query Vectors should match the sizes defined on first line.

K should be a positive integer, and should be no greater than the total number of base vectors

1. Language Support – C++
2. Non – Trivial Optimization Techniques
3. Implemented HNSW (Hierarchical Navigable Small World), which refers to a data structure and algorithm used for approximate nearest neighbor search in high-dimensional spaces. It's a method designed to efficiently search for approximate nearest neighbors in large collections of high-dimensional data.

Referred the following paper for algorithm:

<https://arxiv.org/ftp/arxiv/papers/1603/1603.09320.pdf>

1. Normalizing the vectors before calculating cosine similarity. After normalizing the vectors, just computing dot product is enough, as we don’t have to worry about magnitude.

Benchmarking:

For input test – Dimensions 10, Base Vectors 10000, Query Vectors 1000

Total euclidean time: 3.901 sec

Total HNSW time: 0.144 sec

Total cosine similarity time: 1.398 sec

Total cosine similarity with normalization time: 0.667 sec

Total cosine similarity with GPU time:

For input test – Dimensions 100, Base Vectors 10000, Query Vectors 1000

Total euclidean time: 11.486

Total HNSW time: 0.417

Total cosine similarity time: 11.966

Total cosine similarity with normalization time: 5.251