Final Project

CS471000

NTHU

Outline

- Goal
- Timeline
- Final Presentation
- Submission

Outline

- Goal
- Timeline
- Final Presentation
- Submission

Goal

- Can you extend a relational database system to support storing and querying over vectors?

TA60

- Build any index for approximate nearest neighbor search.

- Options:
  - Inverted File Index (IVF)
  - Scalar Quantization (SQ)
  - Product Quantization (PQ)
  - Locality Sensitive Hashing (LSH)
 - Hierarchical Navigable Small-World (HNSW)

TA70

- Can you make distance calculation faster?

- Hint: Single Instruction Multiple Data (SIMD)
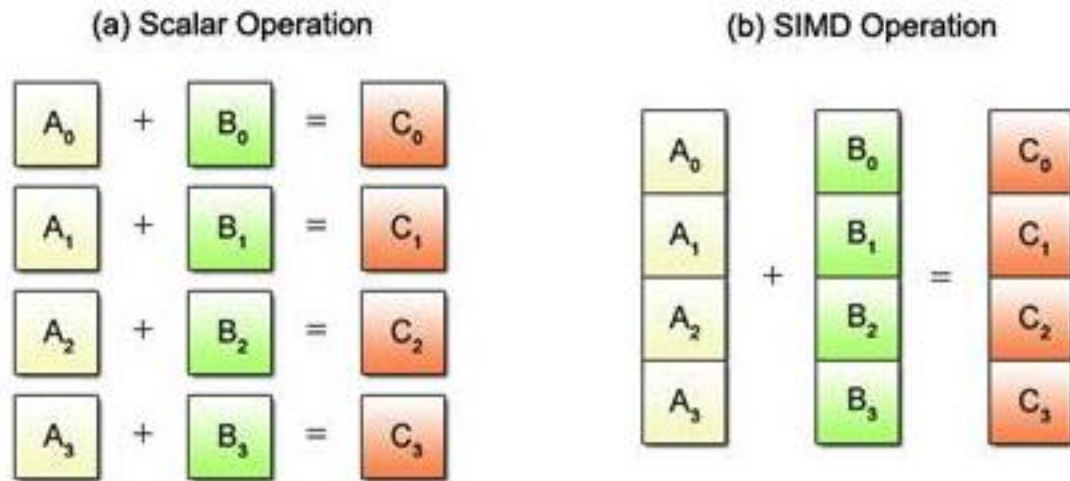
Single Instruction Multiple Data (SIMD) in Java

- Import `jdk.incubator.vector` module

- Write your SIMD version of distance calculation

- Use `jdk17 with `jdk.incubator.vector` package (default jdk17 in VScode is not contain this package)` when running
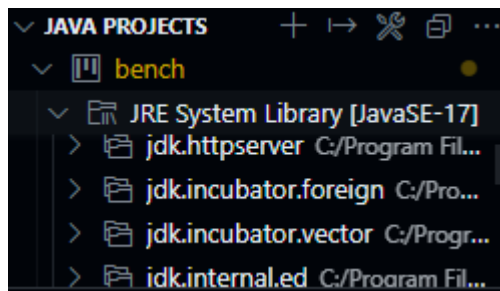
TA70up Evaluation

- Throughput
  - How many nearest neighbor search can you finish over a period of time

- Recall
  - A : { Your nearest neighbor result }
  - T : { True nearest neighbors }
  - Recall: $|(A \cap T)| / |T|$
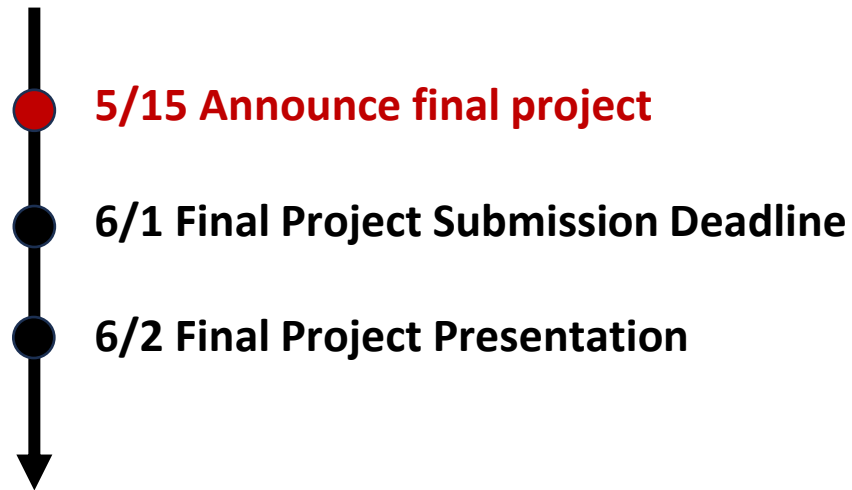
  Each team compete with Throughput * Recall
  The lowest get 70 the highest get 100

Outline

- Goal
- **Timeline**
- Final Presentation
- Submission

Timeline

5/15 Announce final project

6/1 Final Project Submission Deadline

6/2 Final Project Presentation

Outline

- Goal
- Timeline
- **Final Presentation**
- Submission

Final Presentation

- 6/2 (Mon) 15:30 - 18:00 Delta 103

- 4 mins presentation + 2 mins QA for each group

Final Presentation

- Your presentation should cover:
    - Implementations
    - Experiments (Show your throughput and recall)
    - Conclusion

Evaluation

- We will evaluate each group based on:
    - Insight
    - Experiment
    - Presentation

Outline

- Goal
- Timeline
- Final Presentation
- Submission

Submission

- Requirements
  - You have to write a report as usual

- The details of submission will be on GitLab

- Deadline: 2025/6/2(Sun) 23:59

50% Code

30% Presentation

20% Report

References

- https://dl.acm.org/doi/pdf/10.1145/3318464.3386131
- https://vksegfault.github.io/posts/java-simd/