

April 11th Milestone Report

Sunny Nahar

Major Changes:

No major changes, project moving as expected.

What You Have Accomplished Since Your Last Meeting:

Parallelized all seed selector bench framework. I had to restructure the code so that it would utilize per thread data structures to prevent cache coherence overhead and false sharing. I also padded up to a cacheline. The seed selectors share access to the predictor and HashTree. Since there are only reads, multiple threads can easily access the data structures. However, there is still some overhead due coherence traffic. I was able to achieve 40-70x over the various seed selectors.

I implemented two compressed versions of the HashTree data structure. One simply uses the original C++ class based implementation and removes some excess baggage not required for search queries. This reduces the in-memory representation from 296GB to 186GB. I then implemented a C struct based implementation. This removes the memory overhead from virtual function pointers, but the code is messy. This reduces the memory usage further to 147GB.

This had some effect on the runtime. The compressed HashTree was 10% faster, and the struct based compressed HashTree was 20% faster.

I see more room for improvement. I have devised a considerably more complex but efficient representation which can fit in about 59GB. However, I'll put this to future work. There is also bitwise compression that can be done, since I use 32 bit integers for quantities which are at most 26 bits.

Using malloc through new proved again to be a massive cause of contention, so I allocated all local memory to the stack.

Meeting Your Milestone:

I have met my milestone for the past two weeks.

Surprises:

No particular surprises.

Looking Ahead:

The final steps are to gather results on multiple read datasets. The coding portion of this research is complete. I also need to gather data and statistics over the various performance improvements.

Revisions to Your Future Milestones:

No current revisions are needed. The schedule is continuing as planned in the revised milestones.

Resources Needed:

I don't need any particular resources.