

March 28th Milestone Report

Sunny Nahar

Major Changes:

No major changes, project moving as expected.

What You Have Accomplished Since Your Last Meeting:

I thought about other ways to improve the frequency predictor construction time. I don't see further direct improvements I can make at this time. I reduced the synchronization to simply atomic writes.

I was able to improve HashTree construction by about a factor of 3. The key to significant improvement is using memory mapped files. Instead of copying files into a memory buffer, they are mapped to virtual memory, which allows for faster reads. This cuts down the sequential read time to almost 1 second, about a 50x improvement over the disk reading time. Reading from virtual memory will cause page table misses, but this is not as costly as doing a full byte-by-byte copy.

In addition, the program can advise the kernel its page access policy; in this case it was sequential. I was able to squeeze some performance through this.

I implemented a greedy schedule where threads take the largest pieces of work first. This is done by picking the files with the largest size. This allows for a more even distribution of work, and ensures that a later thread does not get a large piece of work, bottlenecking the entire runtime. This improved the runtime by 10%.

I tried different number of file splits (64, 128, 256, 512) to see if I could further reduce the runtime, but this did not add any benefit. The original split I had was the best.

I found an optimization to reduce the space usage by almost 40%. This is huge and could potentially have a large effect on runtime. However, implementing this would require a write of the entire infrastructure, so this is something best left to do in the future.

I have started work on parallelizing the heuristic itself. This is more challenging since most of the code is fairly sequential, since it is a dynamic programming algorithm. I think that the parallelism will be mostly over data.

Meeting Your Milestone:

I have met my milestone for the past two weeks. The goal was to continue existing work and optimize as much as possible.

Surprises:

No particular surprises.

Looking Ahead:

It seems like the work on the HashTree and predictor is done. The next steps are to finish the optimization for the heuristic, and start benchmarking.

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.