

PARTH PARIKH

pmparikh@ncsu.edu • <https://github.com/pncnmnp> • (+1) (984) 810-9749

SUMMARY

I am an aspiring software engineer with a passion for solving unique problems in Computer Science. I have experience working in multidisciplinary, multicultural, and remote teams, and I am an active learner. I have an established interest in research, which is reflected in my efforts to **maintain a blog** that allows me to express my ideas and perspectives to others. I am currently pursuing a Master's degree (Thesis) in Computer Science to support my career development.

EDUCATION

North Carolina State University - Masters in Computer Science (GPA: 4/4)

Aug 2021 - May 2023

University of Mumbai - Bachelor in Computer Engineering (GPA: 9.19/10)

Aug 2017 - June 2021

PUBLICATIONS

Proximity Search in the Greedy Tree

SIAM Symposium on Simplicity in Algorithms, 2023

Linear-time Approximate Hausdorff DistanceThe 30th Fall Workshop on Computational Geometry, 2022**Spectral Bloom Filters for Client Side Search**The 11th IEEE Annual IEMCON, 2020

TECHNICAL EXPERIENCE

Production Engineering Intern - Meta (formerly Facebook)

May 2022 - Aug 2022

- Build performance measuring and stress testing tools at Meta's *Warm Storage team*.
- Used Fio to benchmark Meta's clusters with different kinds of workloads to test performance. The benchmarks involved designing the IO behavior of the file systems - like tuning the block size, file size, total IO, bandwidth rate limit, the ratio of reads/writes, number of threads, etc.
- Architected a tool to perform A/B testing on OS metrics such as IOPS, memory bandwidth, CPU bandwidth, and network latency. Integrated this tool into Meta's CI/CD pipeline to gain routine information on such metrics and catch regressions.

Research Assistant - NC State's Theoretical Computer Science Lab

Aug 2021 - Present

Advisor: Prof. Donald Sheehy

- Analyzing a new metric tree data structure called *Greedy Tree* that utilizes the farthest point sampling (or greedy permutations) for its incremental construction. This research focuses on the intersection of *Computational Geometry* and *Data Structures*.
- Greedy Trees can be used to approximate the bichromatic all nearest neighbors algorithm and Hausdorff distance in $O(n \log n)$ time.
- Surveying the landscape of heuristics and algorithmic ideas that have been employed in Hausdorff distance algorithms.

MLH Fellow - Major League Hacking

Sept 2020 - Dec 2020

- Worked on BentoML (framework for managing and deploying machine learning models) and added support for URL prefix to allow users to run BentoML's storage and deployment model behind a reverse proxy server.

Research Intern - Indian Institute of Information Technology, Allahabad

May 2020 - July 2020

Affective Analysis of Project Gutenberg's corpus

Advisor: Prof. Uma Shanker Tiwary and Mr. Punit Singh

- Designed models to predict and classify emotions of all the passages in popular books from Project Gutenberg's collection.
- Observed and documented issues affecting the emotion analysis domain - such as skewed datasets, difficulty observing the neutral space, and lack of semantic understanding of Modern English in pre-trained transformer models.

PROJECTS AND PERSONAL RESEARCH

Shimmey

Oct 2022

- Designed an end-to-end privacy model for *link shimming* using *Private Information Retrieval (PIR)*.
- Link shimming allows URLs of a website to be redirected to an intermediary endpoint before redirecting it back to the original destination. Facebook, Twitter, Microsoft, and Google have widely adopted this technique.
- Engineered a 2-server PIR architecture with $O(n^{0.5})$ communication cost and observed a **96.2% decrease in runtime cost** over the state-of-the-art technique that relied upon Homomorphic Encryption.

Reversing the 20 Questions Game

Sept 2021 - Nov 2021

- Engineered a transformer-based boolean question-answering model wherein the model chooses an entity at random and the human aims to guess this entity by asking natural language queries with an **accuracy of 78.7%**.

LuaNLP - Natural Language Processing Library for Lua

Feb 2021 - April 2021

- Presently, it is **one of the largest native libraries for statistical NLP in Lua**.
- Implemented 14 modules: tokenizers, lemmatization, stemming, parts-of-speech tagger, sentiment analysis, keyword extraction, named-entity recognition, and text summarization.

Sthir - Spectral Bloom Filters for Client-Side Search

June 2020 - Oct 2020

- Pioneered a memory-efficient library for client-side searching using a probabilistic data structure *Spectral Bloom Filters*.
- This library produces rankings comparable to Lunr.js but **with an 85% decrease in memory footprint**.

Detecting air pollution hotspots and identifying their source trajectories

Jan 2020 - Feb 2020

- Architected two models using satellite data from ERA5 and Sentinel-5P and submitted them to the *Indian Space Research Organization* for predicting ground pollutant concentration using a geographically weighted regression model.

Crossword Solver to solve mini New York Times crosswords

Dec 2019

- Probabilistically solved mini New York Times crosswords (**in under 2 minutes**) by guessing clues and positioning them on the grid (an NP-complete problem) using the Z3 Theorem Prover.
- This research was used by Rumshisky et al. (UMass Lowell) to introduce crossword-solving as a new NLP benchmark.

Indian Movie Recommendation System

Sept 2019 - Nov 2019

- Curated *The Indian Movie Database*, **currently the largest dataset available for Indian movies**, with over 4500 titles released between 1950 and 2019.
- Crafted content-based, collaborative filtering, and hybrid models for the dataset.

Popup Encyclopedia

May 2019

- Developed a Firefox extension that allows users to quickly access the meanings of words by double-clicking on them.
- Optimized to perform faster than state-of-the-art software like Google Dictionary using offline indexing techniques.

Branch Target Buffer

April 2019

- Designed and implemented a *branch prediction buffer* and a *branch target buffer* using C, incorporating a 2-bit saturating counter and an LRU replacement policy to improve performance.

TECHNICAL SKILLS**Programming Languages
Technologies**

Proficient in Python; Prior experience in C/C++, Lua, Javascript, Java, Bash, HTML/CSS
Linux Kernel, AWS Lambda, CI/CD, CUDA, OpenMP/MPI, MySQL, SQLite, Hadoop,
MapReduce, Boost, Flask, Django, L^AT_EX, Docker, PyTorch, HuggingFace, NLTK,
Scikit-learn, Pandas, Numpy, Scrapy, Tkinter, REST, Three.js, WebGL, Git, AutoCAD
Parallel Systems, Compiler Construction, Operating Systems, Databases, Graph Theory,
Natural Language Processing, Privacy in the Digital Age, Software Engineering,
Computer Networks, Computer Graphics

Relevant Courses**Operating Systems**

Ubuntu, MacOS, Debian, Tiny Core Linux, CentOS

TEACHING

- To summarize the learning of my *Interactive Narrative* course, I engineered **Phoenix10.1** - a software to generate personalized pre-recorded internet radios with a text-to-speech based radio jockey.
- Authored technical blogs on topics such as **Approximate Distance Oracles** (scalable pathfinding data structure), **Pseudocode to Code Generation** (analyzing publications in this domain), **Exploring Large Open-Source Codebases**, and **Moderating an Online Discourse**.
- Teaching Assistant for NC State's CSC 442 - *Introduction to Data Science* course during Fall '22.
- Guest lectured on the topic of Support Vector Machines in my undergraduate Machine Learning class.