George Mathew

bigfatnoob.us | george.meg91@gmail.com | 614.535.8678

EDUCATION

NORTH CAROLINA STATE
PhD in Computer Science
Expected Dec 2021| Raleigh, NC

MS in Computer Science
Aug 2014 - May 2016 | Raleigh, NC

AMRITA UNIVERSITY
B.Tech in Electronics Engg
Jul 2008 - May 2012 | CBE, India
(*Gold Medalist*)

SKILLS

PROGRAMMING
Pro: Java • JavaScript• Haskell • Matlab • Python • R • Git • LATEX

Semi-Pro: C++ • Shell • Scala • Clojure • RubyOnRails

Familiar: PHP • Android • Perl • C • Julia • C# • Bison

DATA PROCESSING
Hadoop • Spark • Trident • Storm • Hive

MACHINE LEARNING
Scikit-learn • Pandas • Keras • Tensorflow

DATABASES
MySQL • MongoDB • Redis • Cassandra • Neo4j

BUILD
Make • Maven • Ant • Gradle

DEVOPS
Docker • AWS • Google Cloud • Vagrant

SOFTWARE ENGINEERING
Test Driven Development • Scrum • Continuous Integration • Extreme Programming

FOOTPRINTS

Google Scholar: **tiny.cc/dbnvtz**
Github: **bigfatnoob**
LinkedIn: **georgevmathew**
Twitter: **@ThatBigFatNoob**
Facebook: **george.mathew.1690**

INDUSTRIAL EXPERIENCE

WHATSAPP | S.E. Intern
May'19 - Aug'19 | Menlo Park, CA

- Designed and **introduced custom signals** to track bad actors.
- Developed policies using the custom signals to **automatically identify and ban bad actors** .
- Created a **signal-based dashboard** to monitor bad actors and their actions.
- **Deployed and monitored policies** across different data centers.

MICROSOFT | S.E. Intern - OFFICE
May'18 - Aug'18 | Seattle, WA

- Optimized **selection of projects for packaging builds** based on usage and history, resulting in faster setup and efficient space utilization.
- Implemented **logging framework** to profile different modules of the selection and build framework.
- Developed **telemetry dashboards** of the build process in PowerBI.
- Developed **prototype for improving build** using shared resources.

LEXISNEXIS | S.E. Intern
May'17 – Aug'17 | Raleigh, NC

- Implemented and parallelized **Gradient Boosting Trees** in Enterprise Computing Language on LexisNexis' HPCC platform for ∼500GB data.
- Developed an **integrated search platform** for both legal and academic documents using TensorFlow.
- Research article published in Elsevier Connect (tiny.cc/bfn_el)

FACEBOOK | S.E. Intern
May'15 – Aug'15 | Menlo Park, CA

- Developed **Java based parsing module** for Hive and Presto identifying components of Query.
- Optimized **data migration** between datacenters based on usage statistics.
- Created **Apache Thrift based API** to expose service across different languages.
- Developed **React.js and Java based client** for the Thrift API.

CROWDCHAT | S.E.
Oct'13 – Jun'14 | Hyderabad, India

- Developed crowdchat.net, a **hash-tag based chat platform** using NodeJS-Redis stack, Bootstrap, jQuery, and jade.
- Implemented cross platform chat on platform.crowdchat.net, a **data analytics platform** to connect with people and subscribe to their activities on twitter.
- Built **analytics engine** on a Java-MySQL stack and **scaled for over 250 GB** of unprocessed data.
- Designed **notification engine** and **customized ticker linking posts** from different social platforms.
- Developed **database management** and automation scripts supporting periodic maintenance.

PAYODA TECHNOLOGIES | S.E.
Jun'12 – Sep'13 | Coimbatore, India

- Created a **REST based module to register devices** into AppViewX, a software load balancer.
- Developed an **aggregation module** on mongoDB using map-reduce to aggregate statistics periodically.
- Designed **Prototype Adaptor for configuring switches** into the AppViewX environment.

COURSEWORK

- Advanced Machine Learning
- Automated Software Engineering
- Automated Program Repair
- Advanced Algorithms;
- Compilers; DevOps
- Data Guided Business Intelligence
- Advanced Database Management
- Object Oriented Design Paradigm
- Spatial & Temporal Data Mining

CERTIFICATIONS

Stanford
  - Machine Learning
Princeton
  - Algorithms
Deeplearning.ai
  - Neural Networks & Deep Learning
  - Improving Deep Neural Networks
  - Structuring Machine Learning Projects
  - Convolutional Neural Networks

PERSONAL PROJECTS

Region.io : A bookmark manager via a **website** or **chrome extension** . Pages indexed using **Elasticsearch** , with a **Node.js** based server and **MongoDB** based database. → **region.io**

Octorater : Distributed application for text analysis on movie reviews to predict movie ratings. Powered by **R** and parallelly deployed via **Apache Trident** . → **tiny.cc/bfn_pp0**

Optima : Multi-objective optimization framework containing the latest state of the art optimization algorithms. Implemented on **Python** , front-end using **Flask** and graphic rendering via **Seaborn** & **MatPlotLib** . Adopted and utilized by two research labs. → **git.io/vFrK1**

Collections : Prominent datastructures efficiently implemented in **Java8** based on **object oriented paradigms** & **test driven development** with threading and parallel processing. →**git.io/vFrK7**

ACHIEVEMENTS

- **NCSU Graduate Merit Award** for Cross-Language Source Code Similarity
- **Top-3 best poster award** in HPCC summit for summer internship at LexisNexis
- **Best Fresher** at Payoda Technologies
- **University Silver Medallist** in undergraduation at Amrita University

RESEARCH

SEMANTIC CODE SEARCH GUIDED BY DR. KATIE STOLEE
Feb 2018 – Present | Raleigh, NC
- Designed and maintained **tools to identify similar code** based on dynamic and static similarity measures.
- Developed the **first cross-language dynamic code clone detection** (published in ICSE-2020) and **code-to-code search** (published in FSE-2021) tool on Java, Python and R using multi-objective search of code-similarity measures.
- Explored **expansions for functional programming languages** like Haskell and **scalability to large open-source repositories** .

REAL AI IN SOFTWARE ENG. (RAISE) GUIDED BY DR. TIM MENZIES
Oct 2014 – Jan 2018 | Raleigh, NC
- Analyzed **trends in Software Engineering (SE) venues** using topic modeling and identified that SE research can be summarized as 11 topics.
- Published in International Conference on SE 2017, IEEE Transactions on SE 2018 and Automated SE conference 2019.
- Funded by NASA-JPL, LexisNexis Risk and SEI-CMU .

SOFTWARE ENGINEERING INSTITUTE - CMU (SEI-CMU) Aug 2015 – Feb 2016
- Developed GUI based tool for **optimizing requirements engineering models** on i* frameworks using multi objective optimization techniques.
- Published in IEEE Requirements Eng. Conference 2017 and IEEE Software Architecture Workshop 2017.

NASA JET PROPULSION LABORATORY
Jan 2015 – May 2015
- Estimated **Software effort** for industrial projects and space programs.
- Developed PEEKING2, a **non-parametric projection based model** for skewed datasets with insufficient observations.
- Published in NASA Cost Symposium 2015 and EMSE Journal 2016.

TOP PUBLICATIONS

- **FSE-2021** : George Mathew , Katie Stolee. *Cross-language Code Search Using Static and Dynamic Analyses*
- **ICSE-2020** : George Mathew , Chris Parnin, Katie Stolee. *SLACC: Simion-based Language Agnostic Code Clones*
- **IEEE Software** : George Mathew , Tim Menzies. *Software Engineering's Top Topics, Trends, and Researchers*
- **MSR-2018** : Vivek Nair, Amritanshu Agrawal, Jianfeng Chen, Wei Fu, **George Mathew** , Tim Menzies, Leandro Minku, Markus Wagner, Zhe Yu. *Data-driven Search-based Software Engineering*
- **TSE-2018** : George Mathew , Tim Menzies. *Finding Trends in Software Engineering* (Journal First ASE-2019)
- **RE-2017** : George Mathew , Tim Menzies, Neil A. Ernst, John Klein. *"SHORTer" Reasoning About Larger Requirements Models*
- **ICSE-2017** : George Mathew , Amritanshu Agrawal, Tim Menzies. *Trends in Topics at SE Conferences: Preliminary Version*
- **ISAW-2017** : Neil A. Ernst, John Klein, **George Mathew** , Tim Menzies. *Using Stakeholder Preferences to Make Better Architecture Decisions*
- **ESE-2017** : Tim Menzies, Ye Yang, **George Mathew** , Barry Boehm, Jairus Hihn. *Negative Results for Software Effort Estimation*
- **AAAI-2016** : Jairus Hihn, Leora Juster, James Johnson, Tim Menzies, **George Mathew** . *Improving and Expanding NASA Software Cost Estimation Methods*