George Mathew

bigfatnoob.us | george.megg1@gmail.com | 614.535.8678

EDUCATION

NORTH CAROLINA STATE
PhD in Computer Science
Expected Dec 2021| Raleigh, NC
Current. GPA: 3.92 / 4.0

MS in Computer Science
Aug 2014 - May 2016 | Raleigh, NC
Current. GPA: 4.0 / 4.0

AMRITA UNIVERSITY
B.Tech in Electronics Engg
Jul 2008 - May 2012 | CBE, India
Cum. GPA: 3.92 / 4.0
(Gold Medalist)

SKILLS

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

Semi-Pro: C++ • Shell • LaTeX • 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

OS
Linux • OSX • Windows

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

- Developed policies to **automatically identify and ban bad actors** on WhatsApp based on existing and new signals monitored from the platform.
- Created a dashboard **highlighting periodically the bad actors and their actions** based on different signals.
- **Deployed services** across different data centers and monitor their progress.

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

- Optimized **selection of projects for builds** based on usage and history. This resulted in faster setup prior to build and efficient utilization of space.
- **Implemented logging framework** to profile different modules of the selection and build framework.
- Developed **telemetry dashboards** of the build process in **PowerBI** .

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

- Implemented **Gradient Boosting Trees** on Enterprise Computing Language (ECL).
- **Parallelized Gradient Boosting algorithms** on LexisNexis' HPCC platform to benchmark 500GB of 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** that helps you connect with people and subscribe to their activities on twitter. Built on the Java-MySQL stack and scaled for over 250 GB of analyzed 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 a device** 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

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 in **Java8** based on **Object Oriented Paradigms** & **Test Driven Development** with threading and parallel processing. →**git.io/vFrK7**

CERTIFICATIONS

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

ACHIEVEMENTS

- **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.
- **NCSU Graduate Merit Award** for Cross-Language Source Code Similarity.

RESEARCH

SEMANTIC CODE SEARCH guided by Dr. Katie Stolee

Feb 2018 – Present | Raleigh, NC

- Designed and maintain **tools to identify similar code** based on dynamic and static similarity measures.
- Developed the **first cross-language 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

- Studying **trends in SE venues** using topic modeling. We identify 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.
- Part of obtaining **funding from NASA-JPL, LexisNexis Risk and SEI-CMU** .

SOFTWARE ENGINEERING INSTITUTE - CMU (SEI-CMU)

Aug 2015 – Feb 2016

- **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

- **S/W Effort Estimation** on industrial S/W projects and space programs.
- Published in NASA Cost Symposium 2015 and EMSE Journal 2016.

TOP PUBLICATIONS

Preprints

- **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***