# George Mathew

bigfatnoob.us | george.megg91@gmail.com | 614.535.8678

## EDUCATION

### NORTH CAROLINA STATE

PhD in Computer Science
Expected Dec 2021| Raleigh, NC
Current. GPA: 3.95 / 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

## FOOTPRINTS

Google Scholar: tiny.cc/bfn0
Github: bigfatnoob
LinkedIn: georgevmathew
Twitter: @ThatBigFatNoob
Facebook: george.mathew.1690

## SKILLS

### PROGRAMMING

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

**Semi-Pro:** C++ • Shell • LaTeX • Clojure • RubyOnRails

**Familiar:** PHP • Android • Perl • C • Julia • .NET • Bison

### COURSEWORK

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

### CERTIFICATIONS

- Stanford
  - Machine Learning → tiny.cc/bfnC1
- Princeton
  - Algorithms → tiny.cc/bfnC2
- Deeplearning.ai
  - Neural Networks & Deep Learning → tiny.cc/bfnC3
  - Improving Deep Neural Networks → tiny.cc/bfnC4
  - Structuring Machine Learning Projects → tiny.cc/bfnC5
  - Convolutional Neural Networks → tiny.cc/bfnC6

## TOP PUBLICATIONS

Preprints: tiny.cc/gvm

- **ICSE-2020** : George Mathew , Chris Parnin, Katie Stolee *Software Engineering's Top Topics, Trends, and Researchers* → tiny.cc/bfn_p8
- **IEEE Software** : George Mathew , Tim Menzies. *Software Engineering's Top Topics, Trends, and Researchers* → tiny.cc/bfn_p8
- **TSE-2018** : George Mathew , Tim Menzies. *Finding Trends in Software Engineering* (Journal First ASE-2019) → tiny.cc/bfn_p6
- **RE-2017** : George Mathew , Tim Menzies, Neil A. Ernst, John Klein. *"SHORTer" Reasoning About Larger Requirements Models* → tiny.cc/bfn_p3
- **ICSE-2017** : George Mathew , Amritanshu Agrawal, Tim Menzies. *Trends in Topics at SE Conferences: Preliminary Version* → tiny.cc/bfn_p5
- **ISAW-2017** :Neil A. Ernst, John Klein, George Mathew , Tim Menzies. *Using Stakeholder Preferences to Make Better Architecture Decisions* → tiny.cc/bfn_p4
- **ESE-2017** : Tim Menzies, Ye Yang, George Mathew , Barry Boehm, Jairus Hihn *Negative Results for Software Effort Estimation.* → tiny.cc/bfn_p2
- **AAAI-2016** : Jairus Hihn, Leora Juster, James Johnson, Tim Menzies, George Mathew *Improving and Expanding NASA Software Cost Estimation Methods* → goo.gl/QFBffG
- **NASA-2015** : Jairus Hihn, Tim Menzies, George Mathew , James Johnson *NASA Software Cost Estimation Model: An Analogy Based Estimation Method* → tiny.cc/bfn_p1

## RESEARCH

### SEMANTIC CODE SEARCH guided by Dr. Katie Stolee

Feb 2018 – Present | Raleigh, NC

- **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** (under-review in ASE-2021) tool on Java, Python and R using multi objective search of code similarity measures
- Exploring **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(tiny.cc/bfn_p5) and under review in IEEE Transactions on SE(tiny.cc/bfn_p6)
- 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(tiny.cc/bfn_p3) and IEEE Software Architecture Workshop 2017(tiny.cc/bfn_p4).

### 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(tiny.cc/bfn_p1) and EMSE Journal 2016(tiny.cc/bfn_p2).

# EXPERIENCE

**WHATSAPP** | Software Engineering Intern — May 2019 - August 2019 | Menlo Park, CA

- Developed policies that can **automatically identify and ban bad actors** on WhatsApp. This was based on existing signals or new signals monitored from the platform.
- Create a dashboard **highlighting periodically the bad actors and their actions** based on different signals.
- **Deployment of these services** across different data centers and monitor their progress.

**MICROSOFT** | Software Engineering Intern - OFFICE — May 2018 - August 2018 | Seattle, WA

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

**LEXISNEXIS** | Software Engineering Intern — May 2017 – August 2017 | Raleigh, NC

- Implement **Gradient Boosting Trees** on Enterprise Computing Language(ECL).
- **Parallelize Gradient Boosting algorithms** on LexisNexis' HPCC platform to benchmarking close to 500GB of data.
- Develop a **common search platform for both legal and academic documents** using TensorFlow.
- Top-3 best poster award for the internship.(tiny.cc/bfn_a1)
- Research article published in Elsevier Connect(tiny.cc/bfn_el)

**FACEBOOK** | Software Engineering Intern — May 2015 – August 2015 | Menlo Park, CA

- Developed a **Java based parsing module for Hive and Presto** identifying components of Query.
- Module to **migrate between datacenters based on statistics** and **optimize data migration** and time taken.
- Created a **Apache Thrift based API** to expose service across different languages.
- Developed a **React.js and Java based client** to aid front end users to use the API.

**CROWDCHAT** | FullStack Software Engineer — October 2013 – June 2014 | Hyderabad, India

- Developed crowdchat.net, a **hash-tag based chat platform** . It was built using the NodeJS Redis stack. Bootstrap, jQuery and client jade were used to develop UI.
- Worked 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.
- **Notification engine** and **customized ticker linking posts** from different social platforms.
- **Database management** and automation scripts supporting periodic maintainance scripts.

**PAYODA TECHNOLOGIES** | Software Engineer — June 2012 – September 2013 | Coimbatore, India

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

# PET PROJECTS

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

Octorater : Octorater is a distributed application built to perform text analysis on movie reviews streamed from Rotten Tomatoes to predict the rating of a movie as good or bad. The machine learning was powered by **R** and parallelly deployed via **Apache Trident** . The main objective behind this project was to learn the working of various text analysis methods and select the best out of them to build a predictive model based on the dataset and business requirements. → **tiny.cc/bfn_pp0**

Optima : Optima is a Multi-Objective Optimization framework containing the latest state of the art optimization algorithms. The core business logic is implemented on **Python** , front-end using **Flask** and graphic rendering via **Seaborn** & **MatPlotLib** . Package has been adopted and utilized by two research labs. → **git.io/vFrK1**

Collections : A collection of prominent and efficient datastructures in **Java8** based on **Object Oriented Paradigms** & **Test Driven Development** . It also supports threading and parallel processing. →**git.io/vFrK7**

# ACHIEVEMENTS

- **Top-3 best poster award** in HPCC summit for summer internship at LexisNexis.
- **Best Fresher** at Payoda Technologies.
- **University Silver Medallist** & **Department Gold Medallist** in under-graduation at Amrita School Of Engineering.
- **National Gold Medallist** in High School Math Olympiad.