FEROSH JACOB

https://feroshjacob.github.io

feroshjacob@gmail.com

Research Interests

Software engineering, machine learning, data mining,natural language processing, high-performance computing, domain-specific languages, software modeling, static code analysis, programming languages.

Professional Experience

Senior Software Engineer *Sep 2015 - Present*

The Home Depot, Smyrna, GA

Search Evaluation: One of my main search evaluation goals is to estimate the overall impact of a new feature to the HD.COM search without actually deploying the feature into production. Our current approach is to estimate the overall impact by comparing the search results (products) for the past searches (last month, last year) with and without the feature. My future goal is to develop and deploy an absolute search evaluation system.

HD Ontology: We developed an ontology of products that are relevant to the home improvement store. The ontology was designed to improve the precision and recall of the HD.com search. My focus involved determining the synonyms and the hypernym/hyponym relationships between the products. I have used The HomeDepot search log and customer behavior data along with the publicly available data sets (e.g., Wikipedia, Wordnet) for the data extraction.

Senior Software Engineer *Sep 2013 - Sep 2015*

Careerbuilder, Norcross, GA

Entity Normalization and Instance Matching: I worked on a system that can recognize and normalize company names from text documents (e.g., resumes, job postings). We experimented several machine learning and data mining techniques for the entity recognition and normalization. I also worked on a project that involved removing duplicates in a big data scale using the state of the art instance matching methods.

Graduate Research Assistant *June 2010 - May 2013*

Department of Computer Science, University of Alabama, Tuscaloosa, AL

Software Modeling in HPC: I identified, analyzed, and developed three modeling approaches (Code-level, Program-level, and Domain-level) for computation intensive problems to leverage portability, improved source code maintenance, and heterogeneous computing. I also collaborated with researchers working on cloud computing to understand and evaluate the execution of such problems in a cloud. I am currently supported by a grant from Air Force/AFRL STTR, which I helped Dr. Jeff Gray in preparing and submitting.

Ph.D. Intern *May 2012 - Aug 2012*

Pacific Northwest National Laboratory, Richland, WA

HPC Pipeline DSLs: I developed two Domain-Specific Languages (DSLs) for scientists to describe and deploy signature discovery workflows. The Service Description Language (SDL) describes key elements of a signature discovery service and automatically generates its implementation code. The Workflow Description Language (WDL) describes the pipeline of services and generates deployable artifacts for the Taverna workflow management system.

Software Engineer *Jun 2004 - Jul 2007*
Stabilix Private Solutions Ltd, Trivandrum, India

I worked in the security module of FELIX project, a generic application/claim processing tool for insurance companies.

Education

Doctor of Philosophy *2009-2013*
Department of Computer Science, University of Alabama, Tuscaloosa, AL
Advisor: Dr. Jeff Gray
Thesis: *Modeling Computation Intensive Problems for Heterogeneous Computing and Improved Source Code Maintenance*

Master of Science *2007-2009*
Department of Computer Science, Clarkson University, Potsdam, NY
Advisor: Dr. Daqing Hou
Thesis: *CSeR - A Code Editor for Tracking & Visualizing Detailed Clone Differences*

Bachelor of Technology *2000-2004*
Department of Electronics and Communication, National Institute of Technology, Calicut, India

Technical Skills

Programming Languages
Java, Scala, Python

Big Data tools
Spark, Hadoop, Google Dataproc

Machine Learning tools
Weka, MLlib

Other tools
Jena (Semantic Web), Liftweb (Scala web), SBT (build tool), Maven (build tool)

Refereed Book Chapters

Faizan Javed and **Ferosh Jacob**, Data Science and Big Data Analytics at Career Builder, *Big-Data Analytics and Cloud Computing: Theory, Algorithms and Applications(Marcello Trovati, Richard Hill, Ashiq Anjum, Ying Shao Zhu, and Lu Liueds.)*, Springer International Publishing, ISBN: 978-3-319-25313-8, 2016, pp. 83-96.

Refereed Journal Articles

Ferosh Jacob, Adam Wynne, Yan Liu, Nathan Baker, and Jeff Gray, Domain-Specific Languages For Developing and Deploying Signature Discovery Workflows, *IEEE Computing in Science and Engineering*, vol. 16, no. 1, 2014, pp. 52-64.

Weihua Geng and **Ferosh Jacob**, A GPU Accelerated Direct-sum Boundary Integral Poisson Boltzmann Solver, *Computer Physics Communications*, vol. 184, no. 6, 2013, pp. 1490-1496.

Ferosh Jacob, Jeff Gray, Jeffrey C. Carver, Marjan Mernik, and Purushotham Bangalore, "PPModel: A Modeling Tool for Source Code Maintenance and Optimization of Parallel Programs," *The Journal of Supercomputing*, vol. 62, no 3, 2012, pp. 1560-1582.

Ferosh Jacob, Sonqging Yue, Jeff Gray, and Nicholas Kraft, "SRLF: A Simple Refactoring Language for High Performance FORTRAN," *Journal of Convergence Information Technology*, vol. 7, no. 12, 2012, pp. 256-263.

Ferosh Jacob, Amber Wagner, Prateek Bahri, Susan Vrbsky, and Jeff Gray, "Simplifying the Development and Deployment of MapReduce Algorithms," *International Journal of Next-Generation Computing* (Special Issue on Cloud Computing Yugyung Lee and Praveen Rao, eds.), vol. 2, no. 2, 2011, pp. 123-142.

Ferosh Jacob, Ashfakul Islam, Weihua Geng, Jeff Gray, Brandon Dixon, Susan Vrbsky, and Purushotham Bangalore, "PNBsolver: A Case Study on Modeling Parallel N-body Programs," *Journal of Parallel and Distributed Computing*, 26 pages (in submission).

Refereed Conference Publications (Last three years)

Ferosh Jacob, Ilamgumaran Karunanithi, Pramod Salian, and Ravi Sambhu, BBC: A DSL for Designing Cloud-based Heterogeneous Bigdata Pipelines *Proceedings of the International Conference on Big Data*, Boston, MA, December 2017.

Mayank Kejriwal, Qiaoling Liu, **Ferosh Jacob**, and Faizan Javed, A Pipeline for Extracting and Deduplicating Domain-Specific Knowledge Bases in *Proceedings of the International Conference on Big Data*, Santa Clara, CA, November 2015, pp. 1144-1153.

Qinlong Luo, Meng Zhao, Faizan Javed, and **Ferosh Jacob**, Macau: Large-Scale Skill Sense Disambiguation in the Online Recruitment Domain in *Proceedings of the International Conference on Big Data*, Santa Clara, CA, November 2015, pp. 1324-1329.

Ferosh Jacob, Aaron Johnson, Faizan Javed, Meng Zhao, and Matt McNair, WebScalding: A Framework for Big Data Web Services in *Proceedings of the International Conference of Big Data Computing Service and Applications*, San Francisco Bay, CA, April 2015, pp. 493-498.

Faizan Javed, Qinlong Luo, Matt McNair, **Ferosh Jacob**, Meng Zhao, Tae Seung Kang, Carotene: A Job Title Classification System for the Online Recruitment Domain in *Proceedings of the International Conference of Big Data Computing Service and Applications*, San Francisco Bay, CA, April 2015, pp. 286-293.

Meng Zhao, Faizan Javed, **Ferosh Jacob**, and Matt Mcnair, SKILL: A System for Skill Identification and Normalization in *Proceedings of the Association for the Advancement of Artificial Intelligence (Industry track)*, Austin, TX, January 2015, pp. 4012-4017.

Ferosh Jacob, Faizan Javed, Meng Zhao, and Matt Mcnair, sCooL: A System for Academic Institution Name Normalization in *Proceedings of the Second International Symposium on Big Data and Data Analytics*, Minneapolis, MN, May 2014, pp. 86-93.

Presentations

"Developing an Ontology for the Home Improvement/Retail Domain" *ACM Southeast conference*, Kennesaw, GA, April 2017.

"Machine learning techniques in Java" *Atlanta Java Users Group meetup*, Atlanta, GA, May 2016.

"Modulo-X: A Simple Transformation Language for HPC Programs" *Southeast Regional Conference*, Tuscaloosa, AL, March 2012 (Poster).

"CUDACL+: A Framework for GPU Programs," *International Conference on Object-Oriented Programming, Systems, Languages, and Applications (SPLASH/OOPSLA)*, Portland, OR, October 2011 (Doctoral Symposium).

"Refining high performance FORTRAN code from programming model dependencies" *International Conference on High Performance Computing Student Research Symposium*, Goa, India, December 2010 (Poster). **Best Presentation Award**

"Extending abstract APIs to shared memory," *International Conference on Object-Oriented Programming, Systems, Languages, and Applications (SPLASH/OOPSLA)*, Reno, NV, October 2010 (Student Research Competition). **Bronze Medal**