Paul Komarek

Last Updated: 2013-11-04 San Jose, CA

[**komarek.paul@gmail.com**](mailto:komarek.paul@gmail.com)412.983.1284

U.S. Citizen

# Objectives

I am a project-oriented scientist and software engineer that enjoys solving large and complex technical challenges in a team environment.

# Education

**Ph.D** (Algorithms, Combinatorics, and Optimization) **Carnegie Mellon University**, advised by Andrew Moore, May 2004.

**M.S** (Algorithms, Combinatorics, and Optimization) **Carnegie Mellon University**, advised by Andrew Moore, May 1998.

* 1. (Mathematics) **Western Washington University**, Magna Cum Laude, Graduation with Distinction in Mathematics, June 1997

# Awards

* + - **Google OC Award**, for improvements to cluster utilitzation, 2009
    - **Google OC Award**, for Google Custom Search, 2009
    - **NASA Space Grant Scholarship** Pennsylvania Space Grant Consortium, December 2001
    - **Outstanding Participant** CMU Center for Nonlinear Analysis Summer Undergraduate Applied Mathematics Institute, July 1996
    - **Outstanding Mathematics Graduate** Western Washington University, June 1997

# Current Employment

## Software Engineer/Site Reliability Enginer

**Software Engineer/Site Reliability Enginer**

**Software Engineer**

Google, **Site Reliability Engineering, Another Production Automation Project**, September 2013 through present. I was asked to lead (technical leadership, component de- sign and integration, project management) a new production automation project that would replace several existing projects and add significant new capabilities to automation across Google. This project combines Python, C++, Java, and Go. It will soon include some web tech. I am an active enginer, but primarily at the 25% level due to my leadership require- ments (including the other automation project, below). Product launches in Q4 2013.

Google, **Site Reliability Engineering, Production Automation**, December 2007 through present. Joined new internal “production automation” project in Google’s Site Reliabil- ity group. Became Technical Lead (2009-) and soon after took responsibility for entire project. Functioned as software engineer, software architect, project manager, and any- thing else needed to continue growing the project. From 2011 through 2013, served as personel manager as well. Project was assumed dead when I took over, and grew two or- ders of magnitude under my leadership. This project is primarily Python, and I remain an active engineer as needed. We also have various bits of web tech, data analysis (including mapreduce) and integration points with other languages.

Google, **Search Quality, Google Co-op and Custom Search**, April 2006 through De- cember 2007. Worked on small team to develop Google’s public *Custom Search* product. Responsibilities included feature design and development (in conjunction with team) for frontend (C++, HTML, Javascript, ClearSilver) and backend (C++), tools for load testing and exploring public usage (Bash, Python), customer outreach (iGoogle Gadgets, devel- oper days, blog posts), and participate in a pager rotation to keep everything up-to-date and running (emergency fixes, planned roll-outs and upgrades).

## Software Engineer

Google, **Search Quality, Search Result Ranking**, November 2005 through April 2006. Worked with large team to improve quality of Google’s search results through data mining, machine learning, and advanced alorithms. On the side, I developed a novel tool for identi- fying sets of bad ranking results with common cause, from conception through beta testing and changing my manager’s mind with evidence of impact.

# Education Related Activities and Previous Employment

**Postdoctoral Fellow** Carnegie Mellon University, Robotics Institute, **Auton Lab**, May 2004 through September 2005

Carnegie Mellon University, Robotics Institute, **Auton Lab**, January 2003 through present, coordinated development of Auton software products for the Auton-Pfizer collaboration

**Pfizer Collaboration Lead**

**Aethon Inc.**, January 2002 to April 2002, researched wireless communications, developed

## Consultant

**Systems Administrator**

an elevator interface protocol and accompanying software, provided some systems support, and delivered finished functional prototypes for these systems (Aethon develops a mobile robot for hospital use)

Carnegie Mellon University, Robotics Institute, **Auton Lab**, 2000 through present, includ- ing planning, acquisition, deployment, maintenance and security of computing resources and servers; also helped hire, train, and supervise three systems administrators.

Carnegie Mellon University, **Department of Mathematical Sciences**, August 1997 through May 2004, except when teaching (see below)

**Research Assistant**

Western Washington University, **Department of Mathematics**, September 1996 to June

## Undergraduate Researcher

1997, developed Maple-based two- and three-dimensional tomographic reconstruction soft- ware for convex polytopes

Western Washington University, **Department of Mathematics’ Math Center**, September 1994 to June 1997, upper-division mathematics tutor

**Mathematics Fellow**

# Deployed Hardware and Software Systems

**Internal Google Software** *Jan 2006-* I have design, built, and deployed several internal and customer-facing software systems at Google, related to ranking, Google Custom Search, and automation tools.

*May 2005-* First source release for my Logistic Regression with Truncated Regularized It-

## LR-TRIRLS

**AFC Active Learning**

**Auton Fast Classifiers (AFC)**

**Aethon Elevator Controller**

eratively Re-weighted Least Squares software. Licensed under the GNU General Public Li- cense (GPL), available at [http://www.autonlab.org](http://www.autonlab.org/) and [http://komarix.org/ac/lr.](http://komarix.org/ac/lr)

*September 2004-* Active learning software for scheduling roboticized pharmaceutical ex- periments (see AFC, below). I am responsible for the design, implementation, and main- tenance of this software. This software has been delivered to the sponsor, and will be maintained and distributed contingent on future contracts.

*April 2002-* Fast classification software for high-dimensional datasets. I provided new al- gorithms and eventually took over the entire software system, including the user interface, learner and and dataset framework, performance evaluation, and documentation. This soft- ware is still in use by the sponsor, is maintained regularly, and has been widely distributed.

*January 2002-* Software for managing a single-board computer and serial interface board connected to an elevator’s control system. I developed a protocol and daemons for bidirec- tional communication between a mobile robot and a passenger elevator. Aethon’s current elevator controller is a derivative of my prototype.

## Aethon Wireless Relays

*January 2002-* Stand-alone devices for ad-hoc relaying of communications between mobile robots and an elevator controller (above). I selected the embedded hardware, created a small GNU/Linux operating system and installation utilities adapted to compact flash, and wrote the message relay software. Aethon’s current version of this device is a derivative of my prototype.

**Auton Build System** *-July 2001-* I maintain the makefile and scripts used for building all Auton software on various compilers, microprocessors, and POSIX-ish environments.

*-January 2000-* Servers, compute machines, storage and services used by Auton lab mem- bers for research. My responsibilities include

* + - * hardware and software selection, procurement, and deployment
      * maximizing performance for niche scientific needs on a limited capital budget in a university environment

## Auton Compute Infrastructure

**Tomographic Reconstruction Software**

* + - maintaining vendor relationships and negotiating affordable prices
    - understanding the current high-performance and consumer computing markets, both for our needs and for occasional advisement of clients and other academics.
    - maintaining software, security, and services

This collection of user and server systems is used daily and maintained constantly. Some responsibilities have been shared with additional admins since Spring 2002.

*Fall 1996* Software package for reconstructing a density function over a convex polytope using only information from (n-1)-dimensional integrals (“x-rays”). I developed this soft- ware for a math professor to use as part of his Geometric Tomography classes.

# Publications and Talks

* + - Paul Komarek and Andrew Moore, *Making Logistic Regression A Core Data Mining Tool with TR-IRLS*, **International Conference on Data Mining**, 2005 (ICDM 2005)
    - Paul Komarek, *Logistic regression for fast, accurate, and parameter free data mining*, Invited talk at **Google Inc.**, July 2005.
    - Paul Komarek and Andrew Moore, *Making Logistic Regression A Core Data Mining Tool: A Practical Investigation of Accuracy, Speed, and Simplicity*, Technical Report TR-05-27 at the **Robotics Institute, Carnegie Mellon University**, May 2005.
    - Paul Komarek, *Autonomous Fast Classifiers for Pharmaceutical Data Sets*, Invited talk at **Applied Biosystems Inc.**, July 2004
    - Paul Komarek, *Autonomous Fast Classifiers for Pharmaceutical Data Sets*, Invited talk at the **Midwest Biopharmaceu- tical Statistics Workshop** 2004 (MBSW 2004)
    - Paul Komarek, *Logistic Regression for Data Mining and High-Dimensional Classification*, **Doctoral Thesis**, 2004
    - Alex Gray, Paul Komarek, Ting Liu, and Andrew Moore, *High-Dimensional Probabilistic Classification for Drug Dis- covery*, **Computational Statistics**, 2004 (CompStat 2004)
    - Anya Goldenberg, Paul Komarek, Jeremy Kubica, Andrew Moore, and Jeff Schneider *A Comparison of Statistical and Machine Learning Algorithms on the Task of Link Completion*, **Knowledge Discovery in Databases**, 2003 (KDD 2003)
    - Paul Komarek and Andrew Moore, *Fast Logistic Regression for Large Sparse Datasets with Binary Outputs*, **Artificial Intelligence and Statistics**, 2003 (AISTAT 2003)
    - Paul Komarek and Andrew Moore, A Dynamic Adaptation of AD-trees for Efficient Machine Learning on Large Data Sets, **International Conference on Machine Learning**, 2000 (ICML 2000)
    - Paul Komarek, Canonical Ramsey Numbers–A New Lower Bound for Off-Diagonal Ramsey Numbers, **Joint Mathe- matics Meetings (AMS/MAA)**, 1997

# Other Professional Activities

* + - Active in all aspects of recruiting at Google.
    - Teaching weekly yoga class to co-workers, 2010 Q2 to present.
    - Mentoring Auton graduate students, Jan 2002 to Dec 2002 and Jan 2005 to present.
    - Advising an undergraduate intern in the Auton lab, June 2005 to present.
    - Supervising part-time system administration employees, March 2002 to June 2005.
    - Research advisement of graduate and undergraduate students in the Auton Lab, at the Robotics Institute, Carnegie Mellon University, May 2004 to present.
    - Refereeing submissions to the *Information Systems* journal and several conferences, including the IEEE *Transactions on Knowledge and Data Engineering* (TKDE), *Uncertainty in Artificial Intelligence* (UAI), *Knowledge Discovery and Data Mining* (KDD), and *Neural Information Processing Systems* (NIPS).
    - Teaching Assistant for the Department of Mathematical Sciences, Carnegie Mellon University, August 1998 to December 1998.
    - Participant in the Center for Nonlinear Analysis’ Summer Undergraduate Applied Mathematics Institute at Carnegie Mellon University, Summer 1996.

# Personal

When outdoors I enjoy soccer, hiking, travel and photography. Indoors, I dabble in electronics and use my embedded computing experience for entertainment. I like to combine software, hardware, woodworking and metalworking to complete “essential” upgrades to our home, including a small home theater.

# References

Available on request.