Tomer Kaftan

Education

July 2016 to Present — **Ph.D. Student**, *University of Washington*.
Computer Science & Engineering

Dec 2013 — **Bachelor of Science**, *University of California, Berkeley*, *GPA 3.97*.
Highest Honors in Electrical Engineering & Computer Sciences

Honors and Awards

2016 National Science Foundation Fellowship

2016 UW CSE Weil Family Endowed Fellowship

Publications

Conference Publications

1. **T. Kaftan**, A. Cheung, M. Balazinska, J. Gehrke. Cuttlefish: A Lightweight Primitive for Online Tuning. *Under Submission*.
2. P. Mehta, S. Dorkenwald, **T. Kaftan**, A. Cheung, M. Balazinska, A. Rokem, A. Connolly, J. Vanderplas, Y. AlSayyad. Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads. In *VLDB 2017*
3. E. R. Sparks, S. Venkataraman, **T. Kaftan**, M. Franklin, and B. Recht. KeystoneML: Optimizing Pipelines for Large-Scale Advanced Analytics. In *ICDE 2017*.
4. M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, **T. Kaftan**, M. J. Franklin, A. Ghodsi, and M. Zaharia. Spark SQL: Relational data processing in Spark. In *SIGMOD 2015*.

Talks

1. Cuttlefish: A Lightweight Primitive for Online Tuning. Accepted Speaker at *Strata Data Conference, 2018*.
2. Optimizing Large-Scale Machine Learning Pipelines With KeystoneML. **Invited Speaker** at *ML Systems Workshop, NIPS 2016*.

Posters & Extended Abstracts

1. **T. Kaftan**. Talpidae: Tuning Library Calls in the Dark. *SIGMOD 2017 Student Research Competition*. **Selected for Presentation**

Patents

1. **T. Kaftan**, M. Avrukin, J. D. Santi. Query Categorizer (2015). Publication no. US 20150324868 A1.

3801 Brooklyn Ave NE, Stevens Court K305C – Seattle, WA 98105
United States
+1 (408) 425 7942 • *tomer.kaftan@gmail.com*
tomerk.github.io

Professional Experience

Oct 2014 to Jun 2016 — **Staff Engineer**, *AMPLab, UC Berkeley*.

As a staff engineer, my role was to support multiple research projects going on in the AMPLab.

- Helped develop KeystoneML, a system for constructing machine learning pipelines
- Designed KeystoneML training exercises used by hundreds of AMP Camp 6 attendees
- Created KeystoneML pipelines to support solar flare forecasting research
- Built model prediction and training modules for Velox, a system to manage model lifecycles
- Contributed to published evaluation efforts of SparkSQL

Jan 2014 to Jun 2014 — **Software Engineer**, *Quixey*.

I was a member of the core search architecture team at Quixey, a company building a search engine for mobile apps.

- Successfully led an effort to move the search engine to a shared-nothing distributed architecture, scale tested to 80+ million documents

Sep 2011 to Dec 2013 — **Software Engineering Intern**, *Quixey*.

During my long-running internship, I worked on a variety of information retrieval projects.

- Improved search latencies by 1500% overall via systems optimizations
- Supported efforts to quantifiably improve search result relevance
- Designed a multilingual, context-sensitive noisy channel spell checker
- Built a probabilistic, hierarchical query classifier trained on the document corpus
- Developed an app recommendation system using low-rank matrix factorization techniques

Technical Skills

Tools — Apache Spark, AWS, Hadoop, Lucene, Elasticsearch, MySQL, MongoDB, Cassandra

Languages — Scala, Python, Java, C, Bash, JavaScript, MATLAB, SQL

Activities & Interests

Societies — Student Member of the ACM

Hobbies — Cooking, Hiking, Traveling, Literature

3801 Brooklyn Ave NE, Stevens Court K305C – Seattle, WA 98105
United States
+1 (408) 425 7942 • *tomer.kaftan@gmail.com*
tomerk.github.io