Matt Crane

mcrane@snapbug.geek.nz
https://snapbug.geek.nz
+1 707 764 6288

San Francisco
CA 94107, USA

Work Experience:

Machine Learning Infrastructure Software Engineer, Instacart, CA, USA **November 2021–present**

Building out the infrastructure that powers machine learning at Instacart.

Research Scientist, Meta *(fka Facebook Inc.)*, CA, USA **July 2018–November 2021**

Focused on solving advertiser facing issues within the delivery system, including diagnosing systemic inefficiencies in ranking and delivery systems, large scale back-end migrations to unblock scaling of products, and development of new products. Served as an internal hiring point of contact for the team, as well as managing and mentoring interns and being a ramp-up buddy for new engineers, both senior and junior, to both the team and org.

—Led a team of seven engineers to migrate ads delivery to a new data model for a single ads product. This project spanned the whole ads delivery system and multiple XFN teams. Defined the project roadmap, timeline, and goals alongside XFN partner teams. Designed key infrastructure components, delegating effectively across this team.

This data model serves % of Facebook ads revenue, and unblocked % additional revenue through scaling and reduced serving infrastructure requirements by %. The project serves as the basis for the entirety of ads delivery to migrate data models, unlocking a further % additional revenue, and further reducing serving infrastructure by %.

—Developed a new rule-based delivery product that allows advertisers to express demographic constraints and hints for each asset within an ad. Led a small team of senior backend engineers, communicating engineering updates and issues encountered to both XFN partner teams, and area leadership. Alphas showed majority improved performance (%), and overwhelming positive sentiment (%).

Identified a key issue that affected % of Facebook ads revenue, and was existential for multiple projects across multiple teams.

—Designed and implemented an experimentation platform for arbitrary demand segmented experiments, that did not need budget controls. This platform supplemented, and bought advertiser experience metrics to, the standardised pre-existing supply segmented experimentation platform. It has since been used to support experiments that have shipped % gains cumulatively.

–Improved the sustainability of an ad campaign performance by investigating all parts of delivery. Analyzed the effect of ranking model state/history, detected when campaigns had entered a sub-optimal state, and tested methods to break that state, and sustain performance.

—Identified the source of large variance observed by advertisers on duplicated campaigns. After root-causing to ranking, and conducting a large-scale feature and model migration, an advertiser facing experiment showed revenue gains of %, alongside other key business-sensitive advertiser experience metrics.

Education:

PhD — Computer Science, University of Otago, New Zealand **February 2012–March 2016**

Thesis: Improved Indexing & Searching Throughput `http://ourarchive.otago.ac.nz/handle/10523/6223`

My doctoral research concentrated on investigating ways to make the processes of indexing and searching of web-scale collections more efficient without impacting on the effectiveness of the system.

For instance, we can choose to *not* index certain documents on the basis of quality indicators, but if the wrong decision is made then this could have a significant impact on search quality.

The research methodology involved the continuing support and ongoing development of the ATIRE search engine (an open source, academically developed project available from `http://atire.org`), with which I have been an active contributor since its initial development. My work on ATIRE has enabled it to scale to larger collections.

During my candidature I was, and continue to be, an active member of the information retrieval community, and have presented at multiple SIGIR, ICTIR, and ADCS conferences and workshops, as well as CIKM.

MSc (Thesis Only) with Distinction — Computer Science, University of Otago **October 2009–May 2011**

Thesis: *The New User Problem in Collaborative Filtering* http://otago.ourarchive.ac.nz/handle/10523/1938

BSc(Hons) First Class — Computer Science, University of Otago **2005–2008**

Publications:

Andrew Trotman and **Matt Crane**. "Micro-and Macro-optimizations of SaaT Search". In: *Software: Practice and Experience* 49.5 (2019), pp. 942–950.

Matt Crane. "Questionable Answers in Question Answering Research: Reproducibility and Variability of Published Results". In: *Transactions of the Association for Computational Linguistics* 6 (2018), pp. 241–252. URL: https://transacl.org/ojs/index.php/tacl/article/view/1299.

Joel Mackenzie, J. Shane Culpepper, Roi Blanco, **Matt Crane**, Charles L. A. Clarke, and Jimmy Lin. "Query Driven Algorithm Selection in Early Stage Retrieval". In: *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. WSDM '18. 2018, pp. 396–404.

Leif Azzopardi, **Matt Crane**, Hui Fang, Grant Ingersoll, Jimmy Lin, Yashar Moshfeghi, Harrisen Scells, Peilin Yang, and Guido Zuccon. "The Lucene for Information Access and Retrieval Research (LIARR) Workshop at SIGIR 2017". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17. 2017, pp. 1429–1430.

Matt Crane, J. Shane Culpepper, Jimmy Lin, Joel Mackenzie, and Andrew Trotman. "A Comparison of Document-at-a-Time and Score-at-a-Time Processing". In: *Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. WSDM '17. 2017, pp. 201–210.

Matt Crane and Jimmy Lin. "An Exploration of Serverless Architectures for Information Retrieval". In: *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. ICTIR '17. 2017, pp. 241–244.

Joel Mackenzie, J. Shane Culpepper, Roi Blanco, **Matt Crane**, Charles L. A. Clarke, and Jimmy Lin. "Efficient and Effective Tail Latency Minimization in Multi-Stage Retrieval Systems". In: *CoRR* abs/1704.03970 (2017). arXiv: 1704.03970.

Salman Mohammed, **Matt Crane**, and Jimmy Lin. "Quantization in Append-Only Collections". In: *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. ICTIR '17. 2017, pp. 265–268.

Zhucheng Tu, **Matt Crane**, Royal Sequiera, Junchen Zhang, and Jimmy Lin. "An Exploration of Approaches to Integrating Neural Reranking Models in Multi-Stage Ranking Architectures". In: *NeuIR Workshop at SIGIR2017* abs/1707.08275 (2017). arXiv: 1707.08275.

Jaime Arguello, **Matt Crane**, Fernando Diaz, Jimmy Lin, and Andrew Trotman. "Report on the SIGIR 2015 Workshop on Reproducibility, Inexplicability, and Generalizability of Results (RIGOR)". In: *SIGIR Forum* 49.2 (2016), pp. 107–116.

Matt Crane. "Improved Indexing & Searching Throughput". PhD thesis. 2016.

Ahmed Elbagoury, **Matt Crane**, and Jimmy Lin. "Rank-at-a-Time Query Processing". In: *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*. ICTIR '16. 2016, pp. 229–232.

Jimmy Lin, **Matt Crane**, Andrew Trotman, Jaime Callan, Ishan Chattopadhyaya, John Foley, Grant Ingersoll, Craig Macdonald, and Sebastiano Vigna. "Toward Reproducible Baselines: The Open-Source IR Reproducibility Challenge". In: *Proceedings of the 38th European Conference on Information Retrieval*. ECIR '16. 2016, pp. 408–420.

Matt Crane and Andrew Trotman. "Collision Resolution in Hash Tables for Vocabulary Accumulation During Parallel Indexing". In: *Proceedings of the 20th Australasian Document Computing Symposium*. ADCS '15. 2015, pp. 1–4.

Matt Crane, Andrew Trotman, and David Eyers. "Improving Throughput of a Pipeline Model Indexer". In: *Proceedings of the 20th Australasian Document Computing Symposium*. ADCS '15. 2015, pp. 5–8.

Matt Crane. "Diversified Relevance Feedback". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '13. [Doctoral Consortium Paper]. 2013, p. 1142.

Matt Crane, Andrew Trotman, and Richard O'Keefe. "Maintaining Discriminatory Power in Quantized Indexes". In: *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*. CIKM '13. 2013, pp. 1221–1224.

Matt Crane, Andrew Trotman, and Richard O'Keefe. "Malformed UTF-8 and Spam". In: *Proceedings of the 18th Australasian Document Computing Symposium*. ADCS '13. 2013, pp. 101–104.

Andrew Trotman, Xiang-Fei Jia, and **Matt Crane**. "Managing Short Postings Lists". In: *Proceedings of the 18th Australasian Document Computing Symposium*. ADCS '13. 2013, pp. 113–116.

Matt Crane and Andrew Trotman. "Effects of Spam Removal on Search Engine Efficiency and Effectiveness". In: *Proceedings of the 17th Australasian Document Computing Symposium*. ADCS '12. [Best Paper Honorable Mention]. 2012, pp. 1–8.

Andrew Trotman and **Matt Crane**. "Snip!" In: *Focused Retrieval of Content and Structure*. 2012, pp. 278–282.

Andrew Trotman, Xiang-Fei Jia, and **Matt Crane**. "Towards an Efficient and Effective Search Engine". In: *Open Source Information Retrieval Workshop at SIGIR2012*. 2012, pp. 40–47.

Matt Crane. "The New User Problem in Collaborative Filtering". MSc thesis. 2011.

Active Involvement:

Journal Reviewing:

- Information Processing and Management (Elsevier)
- Information Processing Letters (Elsevier)
- Information Retrieval Journal (Springer)
- Transactions on Information Systems (ACM)
- Transactions on Knowledge and Date Engineering (IEEE)

Conference Reviewing:

- SIGIR2021—44th International ACM SIGIR Conference on Research and Development in Information Retrieval
- SIGIR2020—43rd International ACM SIGIR Conference on Research and Development in Information Retrieval
- ADCS2019—24th Australasian Document Computing Symposium
- CIKM2019—28th ACM Conference on Information and Knowledge Management
- SIGIR2019—42nd International ACM SIGIR Conference on Research and Development in Information Retrieval
- ADCS2018—23rd Australasian Document Computing Symposium
- SIGIR2018—41st International ACM SIGIR Conference on Research and Development in Information Retrieval
- ICTIR2017—3rd ACM International Conference on the Theory of Information Retrieval
- ADCS2017—22nd Australasian Document Computing Symposium
- ADCS2016—21st Australasian Document Computing Symposium
- ADCS2015—20th Australasian Document Computing Symposium
- ADCS2014—19th Australasian Document Computing Symposium

Workshop Involvement:

- LIARR Workshop at SIGIR2017—Lucene for Information Access and Retrieval Research
- RIGOR Workshop at SIGIR2016—Reproducibility, Inexplicability, and Generalizability of Results

Skills:

Programming languages; strong: `C++`, `Python`; moderate: many others, including `SQL`, `Go`, `C`, `Java`.

Troubleshooting and debugging skills of large, complex systems.

Strong written and verbal communication.

Referees:

Available on Request