Tyler J. Skluzacek

skluzacek@uchicago.edu ⋄ (952) 220-1684

RESEARCH INTERESTS

HPC, information extraction, workflow automation, statistical learning, search, user experience

EDUCATION

University of Chicago *2016 - Present*

Ph.D. in Computer Science (expected June 2022; GPA: 3.87/4)
Thesis: "Automated metadata extraction can make data swamps more navigable"

My dissertation enhances scientists' collective ability to navigate large, heterogeneous scientific data collections by providing a scalable system (and programming model) for extracting metadata from files of indeterminate schema, and loading the information into a rich, searchable index. Xtract enables users to leverage both HPC and cloud resources to create and run custom file extraction workflows over large repositories, subject to constraints on time and desired allocation usage, by leveraging statistical learning models that maximize metadata *quality* per unit of time. I am currently studying the extent to which Xtract's workflows improve the navigability of science collections at Argonne and Sandia National Labs.

M.S. in Computer Science (received Nov. 2018)
Thesis: "Automated Workflows for Deriving and Extracting Metadata from Disorganized Data Swamps"

Macalester College *2012 - 2016*

B.A. in Applied Mathematics & Statistics
Minors in Computer Science in Economics

RESEARCH EXPERIENCE

University of Chicago Computer Science Jun. 2016 - Present

Ph.D. Candidate; Research Assistant *(advised by Kyle Chard and Ian Foster)*

- Develop and benchmark a scalable metadata extraction service (Xtract) that executes extraction workflows on large science data repositories across disparate distributed computing systems.
- Parameterize and train file type identification models to increase the quality of metadata outputs.
- Enable container scheduling strategies and pipelines in the funcX function-as-a-service (FaaS) platform to enable remote execution of extraction functions.
- Execute and evaluate extraction workflows on diverse scientific data sets (e.g., batteries, correlation spectroscopy, climate, materials, personal cloud).
- Collaborate with users at UChicago, Argonne National Lab, Sandia National Lab, and various university groups to increase the navigability and FAIR-ness of data collections. Update Xtract, and its corresponding CLI and SDK, to address user needs.

IBM Research Jun. 2018 - Oct. 2018

Graduate Reseach Intern *(advised by Mary Tork Roth and Ed Seabolt)*

- Create and implement microservice-based genome assembly workflows in IBM's Functional Genomics Platform.
- Exercise working knowledge of genome assembly tools: Trimmomatic, Bowtie2, FLASH, SPAdes, QUAST.

- Construct full user interface for genome file uploads and pipeline monitoring.
- Communicate with a number of biologists and bioinformaticians regarding system design.
- Program and implement integration tests to ensure continued quality assurance for IBM's clients.

University of Minnesota Summer 2015
NSF REU Research Assistant *(advised by Jon Weissman and Abhishek Chandra)*

- Edited and built MySQL and HBase layers of a multi-tiered cloud storage middleware for HDFS/Hadoop called Tiera.
- Simulated commercial system usage on batched, tagged Twitter data between EBS, S3, and Glacier cloud storage tiers.

TEACHING AND MENTORSHIP EXPERIENCE

Center for Data & Applied Computing (UChicago, Argonne, Fermi) Jun. 2018 - Oct. 2018
Data4All Workshop Lecturer and Curriculum Development Team

- Designed and delivered a data science curriculum based on COVID-19 and 19th-century London cholera epidemic.
- Taught south side Chicago high school students via Python, Jupyter notebooks, and Kepler.

Department of Computer Science (UChicago) Summer 2020
Principal Lecturer; Introduction to Computer Science 2

- Designed and delivered a full academic quarter of lectures and assignments to a class of 40 students.
- Taught introductory computer science concepts (algorithms, data structures, and memory) via the C programming language.

Master's Program in Computer Science (UChicago) Mar. 2018 - Jun. 2021
Teaching Assistant; Distributed Computing (Kyle Chard) and Cloud Computing (Vas Vasiliadis)

- Lectured classes on microservices, big data frameworks (e.g., Spark, Kafka, Hadoop), and building REST services.
- Curriculum development: wrote assignments and testing suites for graders.
- Held weekly office hours and provided Piazza forum support to 48 master's students.
- Graded assignments, exams, and capstone projects.

Center for Data & Applied Computing (UChicago) Jan. 2019 - Mar. 2020
Lead Summer Research Lab Coordinator

- Provide technical and research help to 27 high school through master's student on a plethora of computing projects.
- Serve as a liaison between students and program/mentors and staff.
- Recruit speakers and host weekly educational lunch-talks for multiple summer internship programs (up to 60 people).

CONSULTING AND STARTUP EXPERIENCE

MarketRebellion, LLC (Chicago, IL) Jul. 2020 - Jan. 2022
Data Science and Cloud Consultant

- Construct, research, and maintain financial sentiment workflows for a top Chicago-based financial education and trading firm.
- Architect massive-scale cloud systems capable of quickly responding to tens of millions of client requests per day.
- Scrape live and historical market and sentiment data from many APIs, including Reddit, Twitter, Binance, Voyager, and NASDAQ.
- Generate and maintain low-latency dashboards for CEO, analysts, and social media representatives.

NightWare, Inc. Sep. 2015 - May 2017
Founder and Inventor

- Invented cross-platform mobile app to track and treat PTSD survivors' traumatic nightmares.
- Raised $26k via Kickstarter and over $25k by winning MentorMate's MobCon pitch competition.
- Educated people about myBivy through the media, with sources including BBC, PBS, and NPR.
- Successfully exited to NightWare, Inc. (May 2017); still provide algorithmic consulting.

Macalester Consulting Group Sep. 2013 - May 2016
President and Board Member (Elected)

- Led a weekly group of 20+ aspiring consultants; discussed industries, strategies, and case studies.
- Interfaced with speakers for weekly career-readiness seminars and consulting case competitions.

AWARDS

UChicago Data Science Institute Rising Star Award	*Nov 2021*
Illinois Technology Foundation – Fifty for the Future Award	*Mar 2018*
CERES Center for Unstoppable Computing Outstanding Graduate Research Award	*Sep 2017*
NSF Graduate Research Fellowship Program Honorable Mention	*Apr 2017*
Nominet Trust 100 Award (myBivy)	*Jan 2017*
American Psychological Association Innovation Leader	*Oct 2016*
Tech{Dot}MN – Minnesota Rising Tech Star of the Year	*Dec 2015*

SELECTED TALKS AND POSTERS

Sandia National Labs Invited Talk *Feb 2022*
"Metadata Extraction for the Battery Modeling Community"

UChicago Data Science Institute Rising Stars Summit (Chicago, IL) *Nov 2021*
"Enabling Data Utility Across the Sciences"

Parsl and funcX Fest (Virtual) *Oct 2021*
"A Serverless Framework for Distributed Bulk Metadata Extraction"

HPDC: ACM Symp. on High-Performance Parallel and Distributed Computing (Virtual) *Jun 2021*
"A Serverless Framework for Distributed Bulk Metadata Extraction"

CERES Center for Unstoppable Computing Annual Summit (Virtual and Chicago, IL) *2017–2020*
"Draining the Data Swamp Funded Project Updates"

Center for Data and Computing Summer Lab Invited Talk *2018–2020*
"Making the most of your research summer"

Middleware 2020 Doctoral Symposium (Davis, CA) *Dec 2019*
"Dredging a data lake: decentralized metadata extraction"

WoSC: Workshop on Serverless Computing (Davis, CA) *Dec 2019*
"Serverless workflows for indexing large scientific data"

SC21 Lightning Talk Series (Dallas, TX; Denver, CO) *Nov 2018, Nov 2019*
"Container Orchestration for Genome Sequencing Pipelines" (2018)
"Serverless Metadata Extraction Workflows for Better Science" (2019)

IOTFuse *Nov 2017*
"NightWare: Transformative Research on the Human Body"

SSDBM: Conference on Scientific and Statistical Database Management *Jun 2017*
"Skluma: A statistical learning pipeline for taming unkempt data repositories"

PDSW-DISCS: Parallel Data Storage and Data Intensive Scalable Computing Systems *Jun 2017*
"Klimatic: a virtual data lake for harvesting and distribution of geospatial data"

SERVICE

Program Committee	TODO, TODO
Peer Reviewer	Cloud ('18), Cluster ('17), eScience ('17), EuroPar ('17), FGCS ('19), HPCC ('17), HPDC ('17, '21), IPDPS ('17, '20, '21, '22), SC ('18, '19, '20, '21), SciPy ('21)
Student Volunteer	SuperComputing ('17, '18 [lead], '19 [lead], '21 [cloud papers committee])
Hackathon Mentor	assisted teams in designing mobile and cloud apps at Uncommon Hacks 2019
Judge	Department of Energy Junior High Science Bowl (2019)

PUBLICATIONS

T. Skluzacek, E. Hsu, M. Chen, K. Chard, I. Foster. "Models and Metrics for Mining Meaningful Metadata". Proceedings of the International Conference on Computational Science (ICCS). 2022.

T. Skluzacek, R. Wong, Z. Li, R. Chard, K. Chard, I. Foster. "A Serverless Framework for Distributed Bulk Metadata Extraction". Proceedings of High-Performance Parallel and Distributed Computing (HPDC). 2021.

Z. Li, R. Chard, L. Ward, K. Chard, **T. Skluzacek**, ..., M. Franklin, I. Foster. "DLHub: Simplifying publication, discovery, and use of machine learning models in science". Journal of Parallel and Distributed Computing (JPDC). 2020.

R. Chard, Y. Babuji, Z. Li, **T. Skluzacek**, ..., K. Chard, I. Foster. "funcX: a Federated Function Serving Fabric for Science". Proceedings of High-Performance Parallel and Distributed Computing (HPDC). 2020.

T. Skluzacek, R. Chard, R. Wong, Z. Li, Y. Babuji, L. Ward, B. Blaiszik, K. Chard, I. Foster. "Serverless Workflows for Indexing Large Scientific Data". 5th International Workshop on Serverless Computing (WoSC). 2019.

T. Skluzacek. "Dredging a Data Lake: Decentralized Metadata Extraction". International Middleware Conference Doctoral Symposium. 2019.

R. Chard*, **T. Skluzacek***, et al., K. Chard, I. Foster. "Serverless Supercomputing: High Performance Function as a Service for Science". ArXiv Preprint: https://arxiv.org/abs/1908.04907. 2019. (*: co-first authors)

T. Skluzacek, R. Kumar, R. Chard, G. Harrison, P. Beckman, K. Chard, I. Foster. "Skluma: An extensible metadata extraction pipeline for disorganized data". IEEE 14th International Conference on e-Science (e-Science). 2018.

P. Beckman, **T. Skluzacek**, K. Chard, I. Foster "Skluma: A Statistical Learning Pipeline for Taming Unkempt Data Repositories". International Conference on Scientific and Statistical Database Management (SSDBM). 2017.

T. Skluzacek, K. Chard, I. Foster. "Klimatic: A Virtual Data Lake for Harvesting and Distribution of Geospatial Data". In 1st Joint International Workshop on Parallel Data Storage and data Intensive Scalable Computing Systems (PDSW-DISCS). 2016.

SELECTED MEDIA

CBS Henry Ford's Innovation Nation – "From idea to software"	Airs January 2022
CNN Online – "[Tyler] designed an app to stop his dad's PTSD nightmares"	Dec. 2020
NPR Online – "[Tyler] Designed a Smartwatch App to Help Stop His Dad's Nightmares"	Dec. 2020
Wired Magazine – "How a Vibrating Smartwatch Could Be Used To Stop Nightmares"	Nov. 2020

PBS Almanac – "An App that Combats PTSD" May 2016

RESEARCH FUNDING PROPOSALS

XSEDE Jetstream Compute Allocations and Supplemental Allocations $ 22,256
(PIs: Kyle Chard, Ian Foster)
Co-led the writing of a proposal to receive funding for hosting Xtract on Jetstream resources at Indiana University and TACC, in addition to other Globus Labs research initiatives. Renewable annually (2017–2022).

Midway2 Research Computing Center Allocation (UChicago) 500K SUs
(PI: Ian Foster)
Provided significant components of a proposal to obtain research computing service units to support large-scale cluster computing in academic years 2020–2021 and 2021–2022.

Petrel and Eagle Storage Systems (Argonne) 200TB storage
(PI: **Tyler J. Skluzacek**)
Led a proposal to obtain persistent research storage at the Argonne Leadership Computing Facility (ALCF). These resources are critical to the development and evaluation of Xtract.

funcX: A Function Execution Service for Portability and Performance $2,658,096
(PI: Ian Foster)
Wrote components of National Science Foundation Award 2004894
to promote the application of funcX—our federated FaaS platform—to myriad science projects, including as part of metadata extraction workflows.

RELEVANT SKILLS

Programming Languages	Python, C, Java, Parsl, R, Swift
Computing Tools	APIs, AWS, cloud, containers, data lakes, databases, distributed systems, HPC, job scheduling, message queues (RabbitMQ, ZMQ), serverless, web development, science workflows, search systems (ElasticSearch, Globus Search)
Statistical Tools	machine learning, regression analysis, survival analysis, time series
Engineering practices	FAIR, open source, Agile, "fail fast"
Science	batteries and battery modeling, materials science data, correlation spectroscopy data, climate science modeling and data