Christina Christodoulakis, PhD

(647)909-1407 christina@cs.toronto.edu linkedin.com/in/christina-christodoulakis github.com/cchristodoulaki

Data Management & AI

I am looking for Senior Applied Scientist or Research Scientist opportunities where I can combine my expertise on data with my passion for user-friendly access to data. My expertise lies in understanding and modeling data, crafting actionable metadata that facilitates data search, data mining, data exploration and integration using machine learning approaches. I'm passionate about providing effective and user-friendly access to complex data with end-to-end development of processes for data collection, curation, and enhancement. I am eager to collaborate with forward-thinking teams to produce innovative and impactful data solutions.

Core Skills

- Data Modeling
- Requirements Engineering
- Full-stack Development
- Data Management
- Applied Machine Learning
- Natural Language Processing
- Data Analysis & Visualization
- Communication
- Collaboration & Leadership

Technical Skills

Languages: SQL, Python, Java, HTML/CSS, JavaScript, Matlab, C
Libraries and Tools: Pandas, scikit-learn, Seaborn, Matplotlib, NLTK, spaCy, GloVe, Django, jQuery, Bootstrap
Technologies/Frameworks: JSP (JavaServlet Pages), RESTful API, BCED layers (Boundary, Control, Entity, Database Interface), MVC (Model View Controller), DAO (Data Access Objects)
Developer Tools: Git/GitHub, GitHub Copilot, Jupyter Notebooks, Visual Studio Code, Eclipse, Netbeans
Requirements Modeling: Use Cases (Cockburn Methodologies), UML (class, sequence, activity, robustness, state, deployment diagrams)
UI Design: Low Fidelity Prototyping (pen & paper wireframing, storyboarding), UI Walkthrough Evaluation, Think Aloud Evaluation, Heuristic Evaluations (Walkthrough, Think Aloud, Heuristic)
Research and Development: Experimentation and Performance Evaluation, Design and Implementation of Research Prototypes, Applied Machine Learning

Education

University of Toronto **Jan. 2015 – April 2024**
Doctor of Philosophy in Computer Science *Toronto, Ontario, Canada*

- **Advisors**: Professor Angela Demke Brown & Professor Moshe Gabel
- **Doctor of Philosophy Thesis**: Facilitating Discovery in Open Data Repositories, Department of Computer Science, University of Toronto, Canada, Apr. 2024

University of Toronto **Sep. 2012 – Dec 2015**
Master of Science in Computer Science *Toronto, Ontario, Canada*

- **Advisor**: Professor Renée J Miler
- **Master of Science Thesis:** A Survey of Recommendation Approaches and New Applications of Recommendations to Data Science, Department of Computer Science, University of Toronto, Canada, Jan. 2015

Technical University of Crete **Sep. 2006 – May 2012**
Bachelor of Engineering in Electronic and Computer Engineering *Chania, Crete, Greece*

- **Advisor**: Professor Antonis Deligiannakis
- **Honors Engineering Degree Thesis:** A Rich Media Mobile Web Application for Visitors & the Community of the Technical University of Crete, School of Electronic and Computer Engineering, Technical University of Crete, Greece, Aug. 2012

Work Experience

Research Assistant, Computer Science, University of Toronto **Jan 2012 – April 2024**
Primary Research Project: Facilitating Discovery in Open Data Repositories

- Led research teams of up to 4 people aimed at enhancing the utility and accessibility of Open Data repositories. Focused on developing methodologies for automatically identifying and standardizing structural and semantic metadata for Open CSV files, enabling more efficient search and exploration.

- Designed/developed end-to-end data ingestion and data annotation pipelines and infrastructure, including UI for Ground Truth annotations in 4500 CSV files.
- Designed/developed the Pytheas Python library, a fuzzy-rule based approach to identify table boundaries and classify lines in such files.
- Designed and implemented a comparative evaluation to state-of-the-art approaches, demonstrating F1-score improvement of almost 13%. Proposed a confidence measure for Pytheas table annotations, reducing labeling effort by 75%.
- Designed and implemented Gnomon, an approach using stacked machine learning and weighted bipartite graph matching for metadata unification from Open Data documentation files.
- Performed comparative evaluation against popular schema matching methods, improving accuracy by 68%.
- Led authorship of academic publications at prestigious data management conferences (e.g., VLDB, ICDE).

Research Intern, Accelerated Discovery Lab, IBM Research, Almaden, California

Asset Discovery for Collaborative Analytics in LabBook *May 2014 – Aug 2014*

- Improved asset recommendation in the LabBook project, a collaborative analysis platform proposed by IBM, by mining the knowledge graph of users, data, software, data science pipelines and digital notebooks.
- Used pen & paper wireframing and storyboarding to propose and iterate on UI design.
- Contributed to manuscript published in IEEE BigData.

Query Recommendations for Watson Analytics *May 2015 – Dec 2015*

- Analyzed Watson Analytics data and query logs, generating abstractions using semantic annotations with WordNet.
- Designed and implemented a GUI for recommendation and visualization of queries.
- Led and coordinated authorship of manuscript published at IUI ESIDA.

Teaching Assistant, Computer Science, University of Toronto **Sep 2012 – Apr 2024**

Courses: Introduction to Databases (9 semesters), Design of Interactive Computational Media (3 semesters), Technologies of Knowledge Media, Software Design

- Created new course materials, including programming assignments, automated tests, and exam questions, tailored to the curriculum and learning outcomes of each course.
- Led tutorials, evaluated and graded student assignments, exams, and projects, while also providing constructive feedback to support student learning and improvement.

Publications

Peer Reviewed Publications

- **C. Christodoulakis**, M. Gabel, A. D. Brown, Metadata Unification in Open Data with Gnomon, 28th International Conference on Extending Database Technology (EDBT) 2025
- **C. Christodoulakis**, M. Gabel, A. D. Brown, Leveraging Guides to Empower Open Data Research , Document Analysis Systems (DAS) 2022
- **C. Christodoulakis**, E. Munson, M. Gabel, A. D. Brown, R. J. Miller, Pytheas: Pattern-based Table Discovery in CSV Files, Very Large Data Bases (Proc. VLDB Endowment) 2020
- R. J. Miller, F. Nargesian, E. Zhu, **C. Christodoulakis**, K. Q. Pu, & P. Andritsos, Making Open Data Transparent: Data Discovery on Open Data, IEEE Data Eng. Bull. 2018
- **C. Christodoulakis**, A. Asgarian, S. Easterbrook, Barriers to Adoption of Information Technology in Healthcare, IBM Conference on Collaborative Advances in Software and COmputiNg (CASCON) 2017
- **C. Christodoulakis**, E. Kandogan, I. Terrizzano, R. J. Miller, VIQS: Visual Interactive Exploration of Query Semantics, ACM IUI ESIDA 2017
- E. Kandogan, M. Roth, P. Schwarz, J. Hui, I. Terrizzano, **C. Christodoulakis**, R. J. Miller, LabBook: Metadata-driven Social Collaborative Data Analysis, IEEE BigData 2015
- B. Ghadiri, **C. Christodoulakis**, S. H. Yeganeh, O. Hassanzadeh, R. J. Miller, K. Lyons, VizCurator: Curating Five-Star Open Data, 24th International World Wide Web Conference (WWW), 2015
- **C. Christodoulakis**, C. Faloutsos, R. J. Miller, VoidWiz: Resolving Incompleteness Using Network Effects, 30th IEEE International Conference on Data Engineering (ICDE), Chicago, 2014

Invited Talks

- Technical University of Crete ECE Summer School on Data Analysis, July 2023
- Cornell University Database Seminar, November 2020

Awards

Bell Graduate Scholarship (2021–2022, 2022–2023)
Wolfond Scholarship in Wireless Information Technology(2020–2021)
VLDB Registration Support (2020)
Bell Graduate Scholarship (2020–2021)
General Motors Women in Science and Mathematics Award (2019, 2020)
Doctoral Completion Award (2019)
Alexander Graham Bell Canada NSERC CGS D Scholarship (2016–2018)
Ontario Graduate Scholarship (OGS) (2014–2015)
Service Distinction Award, Computer Science, University of Toronto (2016)

Activities

Reviewer or Program Committee Member

- Reviewer, The 9th International Conference on Advanced Collaborative Networks, Systems & Applications, COLLA 2019
- The 34th IEEE International Conference on Data Engineering, ICDE 2018
- Reviewer, The 7th International Conference on Social Media Technologies, Communication & Informatics, SOTICS 2017
- Reviewer, The 7th International Conference on Advanced Collaborative Networks, Systems & Applications, COLLA 2017
- 6th ACM Symposium on Cloud Computing, SoCC 2015
- 41st International Conference on Very Large Data Bases, VLDB 2015
- 23rd European Conference on Information Systems, ECIS 2015

Academic Service

- PC meeting facilitator, The 15th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2021
- UofT DCS faculty recruitment, round table host for System and Networks candidates (2021, 2022)
- Graduate application screening, UofT DCS (2019 – 2022)
- Grad Student Recruitment, Lead volunteer, UofT DCS Grad Visit Day (2016, 2017)
- Student panelist, UofT DCS graduate skills seminar 2015
- Student panelist, UofT DCS incoming graduate students 2015