

# FATEMEH NARGESIAN

University of Toronto, Department of Computer Science  
✉ [fnargesian@cs.toronto.edu](mailto:fnargesian@cs.toronto.edu)  
<http://www.cs.toronto.edu/~fnargesian>

## RESEARCH INTERESTS

---

Data management and Data science, with a focus on data discovery, scalable feature engineering, data augmentation, schema inference, and data cleaning for datasets with missing values.

## EDUCATION

---

### University of Toronto

PhD Candidate, Computer Science, 2018 (expected)  
Advisor: Prof. Renée J. Miller  
Dissertation: “Dataset Discovery in Data Lakes”

### University of Ottawa

MSc, Computer Science, 2010  
Advisors: Profs. Iluju Kiringa and Liam Peyton  
Thesis: “Bridging Decision Applications and Multidimensional Databases”

### Sharif University of Technology

MSc, Artificial Intelligence, 2007  
Thesis on “Learning Method Preconditions of Hierarchical Task Network Planning”

### Shahid Beheshti University

BSc, Computer Engineering, 2004

## INDUSTRY RESEARCH EXPERIENCE

---

### IBM T. J. Watson Research Center

Research Intern at Automated Machine Learning Group, 7/2016-11/2016  
*Project Summary: design and development of a feature recommendation module for end-to-end automated data science.*

### IBM T. J. Watson Research Center

Research Intern at Stream Analytics Group, 8/2014-11/2014  
*Project Summary: design and development of a model selection module for end-to-end automated data science.*

### McGill University, Montreal General Hospital

Researcher at Clinical Informatics Research Group, 1/2011-9/2011  
*Project Summary: automatic integration and cleaning of medical records for computerized drug and disease management.*

## PUBLICATIONS

---

### Optimizing Organizations for Navigating Data Lakes.

F. Nargesian, K. Q. Pu, E. Zhu, B. G. Bashardoost, R. J. Miller. Revise and Resubmit, <https://arxiv.org/abs/1812.07024>, PVLDB, 2019.

### A Set Overlap Search Algorithm for Finding Joinable Tables in Massive Data Lakes.

E. Zhu, D. Deng, [F. Nargesian](#), R. J. Miller. To appear in SIGMOD, 2019.

### Making Open Data Transparent: Data Discovery on Open Data.

R. J. Miller, [F. Nargesian](#), E. Zhu, Christina Christodoulakis, K. Q. Pu, Periklis Andritsos. IEEE Data Engineering Bulletin, 2018.

**Table Union Search on Open Data.**

F. Nargesian, E. Zhu, K. Q. Pu, R. J. Miller. In Proceedings of the VLDB Endowment, (PVLDB) 11(7): 813-825, 2018.

**Dataset Evolver: An Interactive Feature Engineering Notebook.**

F. Nargesian, U. Khurana, H. Samulowitz, D. S. Turaga, T. Pedapati. In Proceedings of the Conference on Artificial Intelligence (AAAI), 2018. Demonstration.

**Interactive Navigation of Open Data Linkages. (*Best Demo Award*)**

E. Zhu, K. Q. Pu, F. Nargesian, R. J. Miller. In Proceedings of the VLDB Endowment (VLDB), 2017. Demonstration.

**Learning Feature Engineering for Classification.**

F. Nargesian, H. Samulowitz, U. Khurana, E. B. Khalil, D. S. Turaga. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2017.

**Automating Feature Engineering.**

U. Khurana, F. Nargesian, H. Samulowitz, D. S. Turaga, E. B. Khalil. Artificial Intelligence for Data Science Workshop, NIPS, 2016.

**LSH Ensemble: Internet-Scale Domain Search.**

E. Zhu, F. Nargesian, K. Q. Pu, R. J. Miller. In Proceedings of the VLDB Endowment, (PVLDB) 9(12): 1185-1196, 2016.

**LinkedCT Live: Platform for Online Curation of Clinical Trials Data.**

O. Hassanzadeh, R. J. Miller, F. Nargesian, E. Zhu. In International Semantic Web Conference ISWC, 2015. Demonstration.

**SOFIA: An Analytics Recommendation System.**

F. Nargesian, A. Biem, P. Jain, S. Parthasarathy, D. S. Turaga. In International Semantic Web Conference (ISWC), 2015. Demonstration.

**Data-driven Recommendations for Exploratory Query Formulation.**

F. Nargesian. SIGMOD PhD Symposium: 31-35, 2014.

**Managing and Mapping Data Lineage for Business Intelligence and Analytics Applications in Health Care.**

M. Azarm, F. Nargesian, L. Peyton, i-Society: 120-126, 2011.

**Tool Support and Data Management for Business Analytics Applications in Healthcare.**

M. Azarm, F. Nargesian, L. Peyton, International Journal for Infonomics (IJI) 4(4): 484-493, 2011.

**LHTNDT: Learn HTN Method Preconditions using Decision Tree.**

F. Nargesian, G. Ghassem-Sani, ICINCO-ICSO: 60-65, 2008.

**Center of Confusion Estimation for Out-of-Focus Images Based on Bispectrum.**

F. Nargesian, A. A. Darabi, M. Jamzad, IEEE ICCIMA (3): 501-506, 2007.

**TEACHING EXPERIENCE***Teaching in Higher Education Certificate by the University of Toronto***Sessional instructor, University of Toronto**

Introduction to Databases, Fall 2015

**Guest lecturer, University of Toronto**

Introduction to Databases, Winter 2018

Operating Systems, Winter 2016, Fall 2017

**Teaching assistant, University of Toronto**

Operating Systems, 2013-2017

Database System Technology, Fall 2012

Introduction to Databases, 2012-2018

---

**INVITED TALKS**

---

**Northeastern University** (Dataset Discovery in Data Lakes, Jan. 2019)  
**School of Information, University of Toronto** (Table Union Search, 2018)  
**IBM T. J. Watson** (Learning Feature Engineering, 2016)  
**IBM T. J. Watson** (Dataset Search and Semantics, 2016)  
**IBM T. J. Watson** (Knowledge Base Driven Model Selection, 2014)  
**Carleton University** (Bridging the Gaps between Multidimensional Databases and Applications, 2011)

---

**MENTORSHIP AND OUTREACH**

---

**Research at University of Toronto**  
Mihai Nicolae (now at IBM), Hugh Bui (now at RBC), Vishrant Vasavada (now at UCLA), Ludi Zhan (now at University of Washington)  
**Vice President of Women in Science and Engineering**  
University of Ottawa, 2009-2011

---

**HONORS AND AWARDS**

---

**WiML-17 Travel Grant** (2017)  
**Best Demo Award** (2017) - 43th International Conference on Very Large Data Bases (VLDB)  
**IJCAI-17 Travel Grant** (2017)  
**BELL Scholarship** (2015-2017) - University of Toronto  
**Best Poster Award** (2010) - CASCON, Toronto, Canada  
**International Francophone scholarship** (2008-2011) - University of Ottawa

---

**PATENTS**

---

**USPTO Application #82017030001:** Methods and Systems for Feature Engineering