

# Besat Kassaie

PHD CANDIDATE • COMPUTER SCIENCE

---

The Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1

Phone: +1(519)722-5963 | Email: bkassaie@uwaterloo.ca | Nationality: Canadian

---

**Research Interests: Large Database Systems – Large Information Extraction Systems**

I am a PhD candidate, supervised by Professor Frank Tompa, and member of the Data Systems Group in the Cheriton School of Computer Science at the University of Waterloo. Our group is one of the most influential database and information retrieval groups in North America.

## Research and Industrial Experience

---

2019-2022 **Graduate Research Assistant.** WATERLOO-HUAWEI Joint Innovation Lab.

I am a member of the research team developing a platform that searches documents with text and mathematical content using a pen-based interface. The WATERLOO-HUAWEI Joint Innovation Lab has more than 20 faculty members from the computer science department with various areas of expertise such as data management, programming languages, cloud computing, and mobile trust. Huawei has invested 3 million dollars in research projects at this lab.

2015-2019 **Graduate Teaching Assistant.** The Cheriton School of Computer Science, University of Waterloo.

During my PhD studies, I was a teaching assistant for multiple undergraduate courses in Data Structures & Algorithms, Database Systems, and Introduction to Computer Science. My responsibilities included assisting in recitations, performance assessments, and design of course assignments and projects. For each course, multiple sessions are offered with hundreds of students in each session.

2014 **Senior Research Engineer.** Agency for Science, Technology, and Research (A\*STAR). Singapore.

I was a senior research engineer in a research team working on a project for the McLaren company. The team was working on developing a machine-learning-based approach to design a sleep quality assessment tool. My responsibilities included: model deployment and evaluation as well as participating in conducting and analyzing sleep quality experiments. A\*STAR is a catalyst and enabler of significant research initiatives among the research community in Singapore and beyond.

2013-2014 **Information Technology Consultant.** ITFORCE. Singapore.

I was a consultant on software engineering projects.

## Publications

---

### *JOURNAL PAPERS*

P1- **Besat Kassaie**, Frank Wm. Tompa. 2023. Autonomously Computable Information Extraction. Accepted to appear in Proceedings of the VLDB Endowment. 13 pages.

PVLDB is a major journal for high-impact computer science research on data management with an acceptance rate of 18.6%.

P2- **Besat Kassaie**, Elizabeth L. Irving, and Frank Wm. Tompa. 2021. Computer-Assisted Cohort Identification in Practice. In ACM Transactions on Computing for Healthcare (HEALTH) 3, 2, Article 17. 21 pages.

HEALTH is a multi-disciplinary journal for the publication of high-quality original research papers that have scientific and technological results pertaining to how computing is improving healthcare.

### *PEER-REVIEWED CONFERENCE PAPERS*

P3- Yin Ki Ng, Dallas J Fraser, **Besat Kassaie**, Frank Wm. Tompa. 2021. Dowsing for Math Answers. In Proceedings of CLEF 2021. 12 pages.

The work was chosen as a "best of labs" paper.

P4- **Besat Kassaie**, and Frank Wm. Tompa. 2020. A Framework for Extracted View Maintenance. In Proceedings of the ACM Symposium on Document Engineering 2020. Association for Computing Machinery. New York, USA, Article 14. 4 pages.

The ACM Symposium on Document Engineering publishes original research papers that focus on the design, implementation, development, management, use and evaluation of advanced systems where documents and document collections play a key role with an acceptance rate of 35%.

P5- **Besat Kassaie** and Frank Wm. Tompa. 2019. Predictable and Consistent Information Extraction. In Proceedings of the ACM Symposium on Document Engineering 2019. Association for Computing Machinery, New York, USA, Article 1414, 1-10. 10 pages.

### *WORKSHOP PAPERS*

P6- Yin Ki Ng, Dallas J Fraser, **Besat Kassaie**, Frank Wm. Tompa. 2021. Dowsing for Answers to Math Questions: Ongoing Viability of Traditional MathIR. In CLEF 2021 Working Notes. <http://ceur-ws.org/Vol-2936/>. 19 pages.

P7- Yin Ki NG, Dallas J. Fraser, **Besat Kassaie**, George Labahn, Mirette S. Marzouk, Frank Wm. Tompa, and Kevin Wang. 2020. Dowsing for Answers with Tangent-L. In CLEF 2020 Working Notes. <http://ceur-ws.org/Vol-2696/>. 39 pages.

### *TECHNICAL REPORTS*

P8- **Besat Kassaie** and Frank Wm. Tompa. 2020. Detecting Opportunities for Differential Maintenance of Extracted Views. arXiv:2007.01973. 19 pages.

P9- **Besat Kassaie**. 2017. SPARQL over GraphX. arXiv:1701.0309. 11 pages.

## Selected Projects

---

### I. Differential Maintenance Engine: DME

DME identifies documents' updates that can be autonomously applied on extracted relations. DME statically analyzes an update expression and an extraction program, expressed as a core AQL query, to test sufficient conditions for being a pseudo-irrelevant update. If the input program passes the test, the extracted view content is updated by running a shift algorithm. If it does not pass the test the extractor needs to be executed from scratch.

#### *OBJECTIVES*

Developed to demonstrate the practicality of the differential maintenance approach proposed in my PhD thesis.

#### *RESPONSIBILITIES*

I was the sole designer and developer of the engine. Also, I have designed realistic extractors for the DBLP dataset to show the applicability of the proposed optimization strategy in practice.

#### *UTILIZED TECHNOLOGIES*

Java/Scala Programming Languages. AQL/Jape Information Extraction Languages.

#### *CODE REPOSITORY*

[github.com/Besatkassaie/Differential-Maintenance-Engine](https://github.com/Besatkassaie/Differential-Maintenance-Engine)

#### *FUNDING*

NSERC, the Natural Sciences and Engineering Research Council of Canada.

### II. Iterative Query Refinement System: IQREF

IQRef is an alternative approach to active learning for expert-in-the-loop cohort identification. IQRef uses a hold-out sample for evaluating cohort selection criteria and deploys an adaptive data analysis technique to prevent overfitting to the hold-out data. IQRef operates in two modes: Exploration in which the learner builds queries with high precision; Integration in which the learner combines multiple queries to achieve high recall.

#### *OBJECTIVES*

Developed to assist the University of Waterloo's optometry researchers to conduct medical studies using an Electronic Health Records System.

#### *RESPONSIBILITIES*

I was the sole designer and developer of the system.

#### *UTILIZED TECHNOLOGIES*

Java/Python Programming Languages. Apache Lucene Search Engine.

#### *FUNDING*

NSERC, the Natural Sciences and Engineering Research Council of Canada.

### III. Math-Aware Search Engine

A Math Retrieval system that is developed by the University of Waterloo's researchers. The engine is part of BrushSearch, a platform for searching documents with text and mathematical content using a pen-based interface: <https://www.scg.uwaterloo.ca/brushsearch/>.

#### *OBJECTIVES*

Developed to search and retrieve documents that include mathematical content for given math formulas and terms.

## Selected Projects (Cont.)

---

### *RESPONSIBILITIES*

I implemented many features of the search engine, including SLT tree conversion to math tokens, extraction of repetition tokens, and formula normalizations. I conducted the study of proximity signals and created the proximity re-ranking run for the 2021 ARQMath lab.

### *UTILIZED TECHNOLOGIES*

Java/Python Programming Languages. Apache Lucene Search Engine.

### *ACHIEVEMENTS*

The best participant run of the Answer Retrieval task in the ARQMath Lab in years 2020 and 2021; also, the best automatic run of the Formula Retrieval task in the ARQMath Lab in year 2021.

### *CODE REPOSITORY*

[github.com/kikingo501/MathDowsers-ARQMath](https://github.com/kikingo501/MathDowsers-ARQMath)

### *FUNDING*

NSERC, the Natural Sciences and Engineering Research Council of Canada. **Huawei**, WATERLOO-HUAWEI Joint Innovation Lab.

## Teaching Experience

---

**CS338: Computer Applications in Business: Databases.** Teaching Assistant. University Of Waterloo. Winter 2017, Winter 2019. Multiple sessions with ~200 students and several TAs each.  
Main responsibilities: Design of course assignments and projects, recitation, and performance assessment.

**CS348: Introduction to Database Management.** Teaching Assistant. University of Waterloo. Spring 2017, Fall 2017. Multiple sessions with ~250 students and several TAs each.  
Main responsibilities: Recitation and performance assessment.

**CS240: Data Structures and Data Management.** Teaching Assistant. University of Waterloo. Winter 2016. Multiple sessions with ~300 students and several TAs each.  
Main responsibilities: Recitation and performance assessment.

**CS115: Introduction to Computer Science.** Teaching Assistant. University of Waterloo. Spring 2015. Multiple sessions with ~300 students and several TAs each.  
Main responsibilities: Recitation and performance assessment.

**CS116x: Introduction to Computer Science.** Teaching Assistant. University of Waterloo. Winter 2015. Multiple sessions with ~250 students and several TAs each.  
Main responsibilities: Recitation and performance assessment.

## Education

---

**The Cheriton School of Computer Science. University of Waterloo. Canada.**  
PhD in Computer Science

Thesis Advisor: Frank William Tompa

PhD Thesis: Update-Aware Information Extraction. 2023 (expected).

**Iran University of Science and Technology, Azad University Science and Research Branch. Iran.**  
BENG and MENG in Software Engineering

Undergrad Research Advisor: Mohsen Sharifi, Thesis Advisor: Amir Masoud Rahmani

Master's Thesis: Application of Textual Corpus in Ontology Matching. 2010.

## Talks and Presentations

---

T1- **Besat Kassaie**. March 2023. Autonomously Computable Information Extraction. Departmental seminar: The Cheriton School of Computer Science, University of Waterloo.

T2- **Besat Kassaie**. October 2020. Introduction to MATH Information Retrieval. The Cheriton School of Computer Science, University of Waterloo.

I presented our research on mathematical information retrieval to members of WATERLOO-HUAWEI Joint Lab.

T3- **Besat Kassaie**. Frank Wm Tompa. September 2020. A Framework for Extracted View Maintenance.

Oral presentation: DocEng20 San Jose, CA, USA.

T4- **Besat Kassaie**. May 2019. Predictable and Consistent Information Extraction. Departmental seminar: The Cheriton School of Computer Science, University of Waterloo.

T5- **Besat Kassaie**. Frank Wm Tompa. September 2019. Predictable and Consistent Information Extraction.

Oral presentation: DocEng19 Berlin, Germany.

T6- **Besat Kassaie**. June 2019. Poster: Predictable and Consistent Information Extraction. The 1st Annual CS-Can Student Symposium, McGill University, Montreal.

I was invited to present my research in a poster session showcasing emerging researchers from every field in computer science and engineering.

T7- **Besat Kassaie**. September 2018. Poster: Applying Differential Privacy to Text. CPI Launch Event. Waterloo CyberSecurity and Privacy Institute.

I was invited to present my research in a poster session as part of the institute's launch event. The University of Waterloo's Cybersecurity and Privacy Institute (CPI) is tackling privacy and security challenges by building on Waterloo's expertise in computer science, engineering, mathematics, cryptography and quantum computing to create world-leading cybersecurity research and technologies and increasing interdisciplinary collaboration across all faculties.

T8- **Besat Kassaie**. December 2017. Applying Local Differential Privacy to Text. Departmental seminar: The Cheriton School of Computer Science, University of Waterloo.

## Awards and Grants

---

A1- **The Doctoral Thesis Completion Award**. University of Waterloo. The Faculty of Mathematics 2023. Competitive.

The intention is to assist highly qualified, full-time doctoral students to complete their thesis writing and defense.

A2- **Student Travel Grants**. The ACM Special Interest Group on Hypertext and the Web. 2019. Competitive.

The grant was for students traveling to attend any SIGWEB Sponsored Conference.

A3- **Math Domestic Graduate Student Award**. University of Waterloo. 2015, 2016, 2017, 2018. Competitive.

The goal of this award is to support domestic graduate students in the Faculty of Mathematics, who are engaged in research-based (thesis) programs.

A4- **Provost Doctoral Entrance Award for Women**. University of Waterloo. 2015. Competitive.

The main purpose of this award is to provide outstanding full-time female doctoral students with an entrance scholarship.

A5- **Mathematics Graduate Experience Award**. University of Waterloo. 2015. Competitive.

This award is intended to provide financial support for full-time graduate students who acquire experience as a Teaching Assistant during their graduate degree program in the University of Waterloo.

## Professional Development and Services

---

### *PROFESSIONAL SERVICES*

**S1- Conference Session Chair.** The ACM Symposium on Document Engineering. 2021.

“DocEng is the leading international ACM Symposium for researchers, practitioners, developers, and users to explore cutting-edge ideas and to exchange techniques, tools, and experiences in the domain of document engineering. It is sponsored by ACM by means of the ACM SIGWEB Special Interest Group and in cooperation with ACM SIGDOC. All DocEng Proceedings are available through the ACM Digital Library.”

**S2- Conference Reviewer.** The ACM Symposium on Document Engineering. 2017, 2018, 2019, 2020, and 2021.

**S3- Big Data Challenge Reviewer.** STEM Fellowship Charity. 2021, and 2023.

“STEM Fellowship is a Canadian registered charity that uses mentorship and experiential learning to equip the next generation of change-makers with indispensable skills in data science and scholarly writing. We have 20 university branches in over 8 provinces across Canada.”

**S4- Journal Reviewer.** Springer Applied Sciences Journal. 2020.

“SN Applied Sciences is a fully open access journal and indexed in Web of Science’s Emerging Sources Citation Index (ESCI), SCOPUS, Ei Compendex and DOAJ.”

**S5- Journal Reviewer.** The International Journal on Semantic Web and Information Systems (IJSWIS). 2017.

“IJSWIS is an archival journal that publishes high quality original manuscripts in all aspects of Semantic Web that are relevant to computer science and information systems communities.”

**S6- Conference Reviewer.** The International Conference on Data Management Technologies and Applications (DATA). 2017.

“The purpose of DATA is to bring together researchers, engineers and practitioners interested on databases, big data, data mining, data management, data security and other aspects of information systems and technology involving advanced applications of data.”

### *DEVELOPMENTS*

**D1- Waterloo Woman in Computer Science Initiative.** Member. 2015-2022.

“Women in Computer Science (WiCS) is dedicated to promoting gender equity in computing by advocating for and supporting women, trans, gender-fluid, gender-queer, and non-binary students enrolled in computer science and related computing programs at the University of Waterloo.”

**D2- Grad Cohort for Women.** Participant. 2022.

“At the Grad Cohort for Women, you will spend two days interacting with about 20 senior female computing-related researchers and professionals, who will share pertinent information on graduate school survival skills, as well as more personal information and insights about their experiences. The workshop will include a mix of formal presentations and informal discussions and social events. By attending Grad Cohort for Women, you will be able to build mentoring relationships and develop peer networks that will form the basis for ongoing activities during your graduate career and beyond.”

**D3- Deep Learning and Reinforcement Learning Summer School.** Participant. 2018.

“DLRL is a keystone next-generation offering of the CIFAR Learning in Machines & Brains program and the CIFAR Pan-Canadian AI Strategy’s AI4Good Training Program, hosted each year in partnership with Canada’s three national AI institutes: Amii in Edmonton, Mila in Montreal and the Vector Institute in Toronto.”

**D4- A/B Testing at Scale: Accelerating Software Innovation.** Tutorial Participant. 2017.

“The goal of the tutorial is to teach attendees how to scale experimentation for their teams, products, and companies, leading to better data-driven decisions. We also want to inspire more academic research in the relatively new and rapidly evolving field of online controlled experimentation.”

## References

---

### **Frank William Tompa** (*PhD Supervisor*)

Distinguished Professor Emeritus and Adjunct Professor. Data Systems Research Group. The David R. Cheriton School of Computer Science. University Of Waterloo. Canada. [fwtompa@uwaterloo.ca](mailto:fwtompa@uwaterloo.ca).

- Founder of OpenText Corporation, Canada's fourth-largest software company as of 2022.
- Recipient of Queen Elizabeth II Diamond Jubilee Medal for his significant contributions to text data and design systems for maintaining large reference texts.
- ACM Fellow for his contributions to text-dominated and semi-structured data management.
- Recognized by the University of Waterloo and the City of Waterloo for epitomizing the energy and enterprise that characterize the University of Waterloo through the naming of the road Frank Tompa Drive.

### **M. Tamer Ozsü** (*Academia*)

University Professor. Data Systems Research Group. David R. Cheriton School of Computer Science. University Of Waterloo. Canada. [tamer.ozsu@uwaterloo.ca](mailto:tamer.ozsu@uwaterloo.ca).

- University Professor, a position the University of Waterloo established to recognize exceptional scholarly achievement and international pre-eminence among its faculty.
- Recipient of the IEEE Innovation in Societal Infrastructure Award for his contributions to data science infrastructure and distributed data management.
- Royal Society of Canada Fellow.
- ACM, IEEE, and AAAS Fellow.

### **Elizabeth L Irving** (*Academia*)

Professor. School Of Optometry and Vision Science. University Of Waterloo. Canada. [elirving@uwaterloo.ca](mailto:elirving@uwaterloo.ca).

- Canada Research Chair in Vision Science.
- Affiliated with the Eye Research Institute of Canada in Toronto.
- Affiliated The Department of Ophthalmology at The University of Toronto.
- Recipient of Royal Society of Canada's Alice Wilson Award.

### **Steve Simske** (*Academia*)

Professor. Department of Systems Engineering. Colorado State University. USA. [steve.simske@colostate.edu](mailto:steve.simske@colostate.edu)

- National Academy of Inventors Fellow.
- IEEE Fellow, elevated for contributions to anti-counterfeiting and cyber-physical security.
- Honorary Professor at the University of Nottingham.
- Steering Committee Chair for the ACM DocEng Symposium.

### **Shonali Krishnaswamy** (*Industry*)

Chief Technology Officer. AIDA Technologies Pte Ltd. Singapore. [shonali@aidatech.io](mailto:shonali@aidatech.io)

- Professor and Deputy Director of Data Science Research Institute. Swinburne University of Technology. Australia. (Past)
- Head, Data Analytics Department, Institute for Infocomm Research, Singapore. (Past)