

# Matteo Brucato

RESEARCHER · DATA SYSTEMS GROUP · MICROSOFT RESEARCH

Microsoft Research, Redmond WA, USA

✉ mbrucato@microsoft.com | 🌐 <https://matteo-brucato.github.io/> | 📷 MatteoBrucato

## Summary

**Researcher** at Microsoft Research, in the Data Systems Group.

Ph.D. graduate from the University of Massachusetts Amherst with a focus on prescriptive analytics and database systems. I designed and implemented a new system for enabling prescriptive analytics in relational databases through declarative queries, as demonstrated in my doctoral thesis titled "Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data." This research contributes to simplifying the complexity of analytics applications and enhancing scalability. With a robust background in the theoretical and practical aspects of database management and analytics, I am equipped to deliver valuable insights and innovative solutions in technology and analytics environments.

## Research Interests

**Core areas:**

**Data Management Systems:** Enhancing capabilities to support predictive and prescriptive analytics. **Scalable Machine Learning:** Developing methods and systems that improve fairness and efficiency in machine learning applications. **Decision Support Systems:** Building advanced tools to facilitate complex decision-making processes in uncertain environments.

**Selected topics:**

**Prescriptive Analytics:** Extending query engines to support in-database prescriptive analytics. **Markov Decision Processes:** Creating scalable solutions for sequential decision-making problems. **Fair Machine Learning:** Integrating fairness into existing and new machine learning frameworks. **Personalized Document Summarization:** Designing example-driven summarization techniques. **Temporal Information Retrieval:** Advancing the retrieval of time-based data for enhanced contextual understanding.

## Professional Experience

### INDUSTRY

#### Microsoft Research

RESEARCHER

*Redmond, WA, USA*

*Oct 2021 - present*

#### Microsoft Research

RESEARCH INTERN

*Redmond, WA, USA*

*May - Aug 2017*

Worked under the supervision of **Kaushik Chakrabarti** on natural language to SQL using neural networks.

#### IBM Watson

RESEARCH INTERN, WATSON HEALTH

*T. J. Watson Research Center, Yorktown Heights, NY, USA*

*Jun - Aug 2016*

Worked under the supervision of **Shilpa Mahatma** and **Yajuan Wang** on "conceptual queries" for healthcare applications.

#### CINECA

SOFTWARE ENGINEERING INTERN

*Casalecchio di Reno, Bologna, Italy*

*Jun - Oct 2012*

- Responsible for developing and managing a data warehouse and analytics system for a nationwide videoconferencing system.
- Duties included generating aggregated data and analytics, deployed through usable interfaces to let decision makers understand usage and performance of the videoconferencing system.

#### Maggiore Hospital

SOFTWARE DEVELOPER (VOLUNTEER)

*Bologna, Italy*

*2011 - 2013*

- Co-developed software to help the Emergency Room visit their patients using portable ultrasound technology.
- Software deployed on medical tablets to help doctors collect patient's information on the field.
- The collected data was then used to make better informed decisions in the Emergency Room.



### ACADEMIA

## DREAM Lab, University of Massachusetts

Amherst, MA, USA

### RESEARCH ASSISTANT

Sep 2013 - May 2021

- Worked with **Prof Peter J. Haas**, **Prof Alexandra Meliou**, and **Prof Azza Abouzied** (NYU Abu Dhabi) to create  **Package Builder**, a system that extends query engines to support the generation of *packages*, i.e., collections of tuples with global properties.
- Technical paper at VLDB 2024
- Technical paper at SIGMOD 2020
- Demo papers at VLDB 2020
- Invited journal paper at CACM Research Highlights 2019.
- Invited journal paper at the VLDB Journal 2018.
- Invited journal paper at SIGMOD Records 2017.
- Technical paper at VLDB 2016; workshop paper and demo paper at VLDB 2014; undergraduate research poster at SIGMOD 2014.
- Project website: [packagebuilder.cs.umass.edu](http://packagebuilder.cs.umass.edu).
- DREAM Lab website: [dbgroupp.cs.umass.edu](http://dbgroupp.cs.umass.edu).
- Code: Scalable-PaQL-Queries on  GitHub

## Design Technology Lab, NYU Abu Dhabi

Abu Dhabi, UAE

### VISITING PH.D. CANDIDATE

Mar - May 2016

- Paper: Redistributing Funds across Charitable Crowdfunding Campaigns (arXiv:1706.00070).

## AMPLab, UC Berkeley

Berkeley, CA, USA

### VISITING PH.D. STUDENT

May - Aug 2015

- Worked on extensions to Package Builder to identify ideal packages by example, under NYU Summer internship.

## Data-Intensive Systems Lab, Århus University

Århus, Denmark

### VISITING M.S. STUDENT

Jan - Jun 2013

- Worked on M.S. thesis on *Temporal Information Retrieval* under the supervision of **Prof Christian S. Jensen**.
- Actively collaborated with researchers within the group.
- Published two papers in main international conferences in the fields of NLP and IR.

## Database Lab, UC Riverside

Riverside, CA, USA

### UNDERGRADUATE RESEARCH INTERN (VOLUNTEER)

Jan - Jun 2013

- Worked closely with a Ph.D. candidate and **Prof Vassilis J. Tsotras** on query optimization techniques for query result diversification.
- Developed a query parser for a SQL-like diversification query language.
- Devised and developed a query optimizer to select optimal diversification query plans.
- Developed client/server technology to issue and evaluate diversification queries.








## Honors & Awards

2020	<b>Best Demo Award</b> , VLDB 2020	Tokyo, Japan
2020	<b>Best Demo Runner-up Award</b> , VLDB 2020	Tokyo, Japan
2019	<b>Krithi Ramamritham Computer Science Scholarship</b>	Amherst, MA, USA
2018	<b>CACM Research Highlight Award (Invited journal paper in CACM Research Highlights)</b>	Amherst, MA, USA
2017	<b>ACM SIGMOD Research Highlight Award (invited journal paper in SIGMOD Records)</b>	Amherst, MA, USA
2017	<b>Invited journal paper in the VLDB Journal, Special Issue on Best Papers of VLDB 2016</b>	Amherst, MA, USA
2016	<b>Best Papers of</b> , VLDB 2016	New Delhi, India
2013	<b>M.S. with Honors</b> , University of Bologna	Bologna, Italy
2013	<b>Best Young Researcher Award Nominee</b> , RANLP 2013	Hissar, Bulgaria
2013	<b>Merit Scholarship</b> , awarded to 0.2% of all students of the University of Bologna	Bologna, Italy
2012	<b>Research Scholarship</b> , University of Bologna	Bologna, Italy
2012	<b>Merit Certificate</b> , University of Bologna	Bologna, Italy
2012	<b>Complimentary AAAI Membership</b> , Stanford Online AI Course	Stanford, CA, USA
2011	<b>B.S. with Honors</b> , University of Bologna	Bologna, Italy
2011	<b>Merit Certificate</b> , University of Bologna	Bologna, Italy
2010	<b>Merit Certificate</b> , University of Bologna	Bologna, Italy
2010	<b>Exchange Scholarship</b> , University of Bologna	Bologna, Italy

## Publications

1. **Matteo Brucato**, Tarique Siddiqui, Wentao Wu, Vivek Narasayya, Surajit Chaudhuri  
Wred: Workload Reduction for Scalable Index Tuning  
*SIGMOD 2024 (conference) pp.1-26*
2. Anh L. Mai, Pengyu Wang, Azza Abouzied, **Matteo Brucato**, Peter J. Haas, Alexandra Meliou  
Scaling Package Queries to a Billion Tuples via Hierarchical Partitioning and Customized Optimization  
*VLDB 2024 (conference) pp.1146-1158*
3. S. G. Rizzo, **M. Brucato**, D. Montesi  
Ranking Models for the Temporal Dimension of Text

*ACM Transactions on Information Systems (TOIS) 2023 (journal) pp.1-34*

4. N. Yadav, **M. Brucato**, A. Fariha, O. Youngquist, J. Killingback, A. Meliou, P. J. Haas  
SUBSUME: A Dataset for Subjective Summary Extraction from Wikipedia Documents  
*NewSum @ EMNLP 2021 (conference)*
5. S. B. Nashed, J. Svegliato, **M. Brucato**, C. Basich, R. Grupen, S. Zilberstein  
Solving Markov Decision Processes with Partial State Abstractions  
*ICRA 2021 (conference) pp.813-819*
6. **M. Brucato**, M. Mannino, A. Abouzied, P. J. Haas, A. Meliou  
sPaQLTools: A Stochastic Package Query Interface for Scalable Constrained Optimization  
*VLDB 2020 (conference) pp.2881-2884*  
 **Best Demo Award. Video: <https://www.youtube.com/watch?v=1v5KevGKbfE>**
7. A. Fariha, **M. Brucato**, P. J. Haas, A. Meliou  
SuDocu: Summarizing Documents by Example  
*VLDB 2020 (conference) pp.2861-2864*  
 **Best Demo Runner-Up Award. Video: <https://www.youtube.com/watch?v=ZGzF6cBSYjo>**
8. **M. Brucato**, N. Yadav, A. Abouzied, P. J. Haas, A. Meliou  
Stochastic Package Queries in Probabilistic Databases  
*SIGMOD 2020 (conference) pp.269-283*
9. **M. Brucato**, A. Abouzied, A. Meliou  
Scalable Computation of High-Order Optimization Queries  
*Communications of the ACM (CACM) 2019 (journal) pp.108-116*
10. **M. Brucato**, A. Abouzied, A. Meliou  
Package queries: Efficient and scalable computation of high-order constraints  
*The VLDB Journal 2018 (Special Issue on Best Papers of VLDB 2016) (journal) pp.693-718*
11. **M. Brucato**, A. Abouzied, A. Meliou  
A Scalable Execution Engine for Package Queries  
*SIGMOD Record 2017 (ACM SIGMOD Research Highlight Award 2017) (journal) pp.24-31*
12. **M. Brucato**, A. Abouzied, C. Blauvelt  
Redistributing Funds across Charitable Crowdfunding Campaigns  
(tech report) arXiv:1706.00070 2017
13. **M. Brucato**, J. F. Beltran, A. Abouzied, A. Meliou  
Scalable Package Queries in Relational Database Systems  
*VLDB 2016 (conference) pp.576-587*  
 **Best papers of VLDB 2016**  
 **Invited journal paper in the VLDB Journal (Special Issue on Best Papers of VLDB 2016)**  
 **ACM SIGMOD Research Highlight Award (Invited journal paper in SIGMOD Records 2017)**  
 **CACM Research Highlight Award (Invited journal paper in CACM Research Highlight 2018)**
14. **M. Brucato**, A. Abouzied, and A. Meliou  
Improving Package Recommendations through Query Relaxation  
*Data4U @ VLDB 2014 (conference) pp.13-18*
15. **M. Brucato**, R. Ramakrishna, A. Abouzied, A. Meliou  
PackageBuilder: From Tuples to Packages  
*VLDB 2014 (conference) pp.1593-1596*
16. K. Fernandes, **M. Brucato**, R. Ramakrishna, A. Abouzied, A. Meliou  
PackageBuilder: Querying for packages of Tuples  
*SIGMOD 2014 (conference) pp.1613-1614*
17. **M. Brucato**, D. Montes  
Metric Spaces for Temporal Information Retrieval  
*ECIR 2014 (conference) pp.385-397*
18. **M. Brucato**, L. Derczynski, H. Llorens, K. Bontcheva, C. S. Jensen  
Recognising and Interpreting Named Temporal Expressions  
*RANLP 2013 (conference) pp.113-121*  
 **Best Young Researcher Award Nominee**

## Education

### UMASS AMHERST

PH.D. IN COMPUTER SCIENCE

- Co-advisors: Peter J. Haas, Alexandra Meliou, and Azza Abouzied (NYU)
- Thesis: *Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data*

Amherst, MA, USA

2013 - 2021

## UNIVERSITY OF BOLOGNA

M.S. IN COMPUTER SCIENCE

- Summa cum laude
- Thesis Advisor: Danilo Montesi
- Thesis: *Temporal Information Retrieval*

*Bologna, Italy*

2011 - 2013

## UNIVERSITY OF BOLOGNA

B.S. IN COMPUTER SCIENCE

- Summa cum laude
- Thesis Advisor: Danilo Montesi
- Thesis: *Mobile systems for patient-reported outcomes*

*Bologna, Italy*

2008 - 2011

## EXCHANGE PROGRAMS

---

### ÅRHUS UNIVERSITY

EXCHANGE PROGRAM DURING M.S., COMPUTER SCIENCE

- Advisor: Christian S. Jensen

*Århus, Denmark*

Jan - Jun 2013

### UC RIVERSIDE

EXCHANGE PROGRAM DURING B.S., COMPUTER SCIENCE

- Advisor: Vassilis J. Tsotras

*Riverside, CA, USA*

Sep - Jun 2011

## Advising & Mentoring

---

### Research Advisor for Ph.D. students Riddho Ridwanul and Anh Mai

UMASS AMHERST

Riddho (UMass Amherst) and Anh (NYU Abu Dhabi) are first-year Ph.D. students and working on package queries.

*Amherst, MA*

2021

### Research Advisor for M.S. students Oscar Youngquist and Julian killingback

UMASS AMHERST

Oscar is working on enhancing SuDocu, our system for extractive document summarization by example. Julian helped us collecting user-annotated data from Mechanical Turk.

*Amherst, MA*

2021

### Research Advisor for undergraduate student Agam Sandhu

UMASS AMHERST

Agam Sandhu worked on scalable stochastic constrained optimization under costly sampling methods.

*Amherst, MA*

2020

### Research Advisor for PhD student Ahmad Chatha (NYU)

UMASS AMHERST

Ahmad Chatha worked on combining scalable constrained optimization methods with scenario-based methods for stochastic optimization.

*Amherst, MA*

2020

### Project Mentor (in collaboration with Microsoft Research)

UMASS AMHERST

Mentored MS students (Abdul Hannan Kanji, Parul Gupta, Ge Gao, Xinlu Liu) for their Independent Class project (CS 696DS), in collaboration with Microsoft Research, to improve the Fairlearn library (<https://fairlearn.github.io>).

*Amherst, MA*

2020

### Research Advisor for MS student Kanchi Masalia

UMASS AMHERST

Kanchi Masalia worked on parts of SuDocu, our system for personalized efficient document summarization by example.

*Amherst, MA*

2020

### Research Advisor for undergrad Genglin Liu

UMASS AMHERST

Genglin Liu worked on the frontend of SuDocu, our system for personalized efficient document summarization by example.

*Amherst, MA*

2020

### Research Mentor for undergraduate student Kyaw Htet

UMASS AMHERST

Kyaw worked on analyzing scenario reductions methods for stochastic optimization.

*Amherst, MA*

Fall 2019

### Project Mentor (in collaboration with Microsoft)

UMASS AMHERST

Mentored a group of four MS students (Katie House, Shanu Vashishtha, Deeksha Razdan, Cheng-Yin Eng) for their Independent Class project (CS 696DS), in collaboration with Microsoft to produce notebooks that showcase their Machine-as-a-service platform Azure ML.

*Amherst, MA*

Spring 2019

### Research Mentor for undergraduate student Nishad Ranade

UMASS AMHERST

Nishad worked on Monte Carlo scenario reductions methods.

*Amherst, MA*

2018-2019

### Research Mentor for 1st-year PhD student Nishant Yadav

UMASS AMHERST

Nishant and I worked together on stochastic package queries.

*Amherst, MA*

2018

## Research Mentor for graduate student Sandhya Sankaranarayanan

UMASS AMHERST

Sandhya and I worked together on extending the package query language PaQL.

Amherst, MA

Fall 2015

## Research Mentor for undergraduate student Kevin Fernandez

UMASS AMHERST

Kevin and I worked together on evaluation techniques for package queries. This work was invited to present at the *ACM SIGMOD 2014 Undergraduate Research Poster Competition*.

Amherst, MA

Spring 2014

## Research Mentor for graduate student Rahul Ramakrishna

UMASS AMHERST

Rahul and I worked together on interface abstractions for package queries. Our work was demonstrated at *VLDB 2014*.

Amherst, MA

2013 - 2014

## Teaching

---

### Teaching Assistant for Graduate Database Systems (CS 645)

UMASS AMHERST

Office hours, question answering, grading, projects.

Amherst, MA

Spring 2021

### Teaching Assistant for Undergraduate Database Systems (CS 345)

UMASS AMHERST

Office hours, question answering, grading and labs.

Amherst, MA

Fall 2020

### Teaching Assistant for Undergraduate Class “A Networked World” (INFO 203)

UMASS AMHERST

Homework grading.

Amherst, MA

Spring 2018

### Teaching Assistant for Undergraduate Introduction To Computation (CS 250)

UMASS AMHERST

Responsible for discussion sessions, occasional lectures, office hours, homework and discussion grading.

Amherst, MA

Fall 2017

### Teaching Assistant for Undergraduate Database Systems (CS 345)

UMASS AMHERST

Responsible for occasional lectures, project advising, office hours, homework and project grading.

Amherst, MA

Fall 2015

## Professional Service & Leadership

---

### Program Committee

- *SIGMOD 2025*
- *VLDB 2025*
- *DataPlat workshop @ ICDE 2024*
- *IEEE Transactions of Knowledge and Data Engineering (TKDE) 2020*

### Guest Editorial Board

- *Information Processing and Management*, special issue on *Time and Information Retrieval*, Elsevier, 2014/2015.

### External Review Committee

- *VLDB 2017*
- *SIGMOD 2015*
- *VLDB 2015*
- *RANLP 2013*

### Lab Seminar Organizer

UMASS AMHERST DATABASE LAB

- Coordinated weekly discussion seminar, including invited talks and paper readings, for the UMass Amherst Database Research group (~10 Ph.D. students and faculty), and webmaster duties.

Amherst, MA, USA

2014 - 2015

### Student Seminar Leader

UMASS AMHERST DATABASE LAB

- Organized and coordinated a weekly discussion seminar among Ph.D. and M.S. students in the UMass Amherst Database Research group to foster research collaborations, exchange of ideas, group discussions, and group cohesion.

Amherst, MA, USA

Fall 2015

## Talks

---

### INDUSTRY

---

#### Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data

MEGAGON LABS

Virtually

Jan 2021

#### Stochastic Package Queries for In-Database Prescriptive Analytics

SISU DATA

Virtually

Sep 2021

## Scalable Package Queries in Relational Database Systems

MICROSOFT RESEARCH

Redmond, WA, USA

Jun 2017

## Conceptual Queries for Multiple Data Sources

IBM WATSON

Austin, TX, USA

Aug 2016

## Package Queries: From Tuples to Packages

BERKELEY LABS

Berkeley, CA, USA

Jul 2015

## ACADEMIA

---

## Package Queries: Scalable Prescriptive Analytics Close to the Data

GEORGIA TECH

Virtually

Mar 2021

## Package Queries: Scalable Prescriptive Analytics Close to the Data

CARNEGIE MELLON UNIVERSITY

Virtually

Mar 2021

## Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data

HARVARD UNIVERSITY, DASLAB

Virtually

Jan 2021

## Stochastic Package Queries For In-database Optimization Under Uncertainty

INFORMS 2020

Virtually at INFORMS 2020

Nov 2020

## sPaQLToolS: A Stochastic Package Query Interface for Scalable Constrained Optimization

VLDB 2020

Tokyo, Japan (virtually)

Sep 2020

- Video: <https://www.youtube.com/watch?v=1v5KevGKbfE>

## Stochastic Package Queries in Probabilistic Databases

SIGMOD 2020

Portland, OR, USA (virtually)

Jun 2020

- Video: <https://www.youtube.com/watch?v=UrEOEplK06I>

## Package Queries: Enabling Declarative and Scalable Prescriptive Analytics in Relational Data

MIT CSAIL, DATA SYSTEMS GROUP

Cambridge, MA, USA

Feb 2020

## Stochastic Package Queries for In-Database Constrained Optimization Under Uncertainty

MIT, NORTH EAST DATABASE DAY, SAMBERG CONFERENCE CENTER

Cambridge, MA, USA

Jan 2020

## Scalable Package Queries in Relational Database Systems

VLDB 2016

New Delhi, India

Sep 2016

- Slides: [packagebuilder.cs.umass.edu/papers/scalable-paql-presentation-share.pdf](http://packagebuilder.cs.umass.edu/papers/scalable-paql-presentation-share.pdf)

## Package Queries: From Tuples to Packages

AMPLAB, UC BERKELEY

Berkeley, CA, USA

Jun 2015

## Improving Package Recommendations through Query Relaxation

VLDB 2014

Hangzhou, China

Sep 2014

- Slides: [packagebuilder.cs.umass.edu/papers/package-recomm-slides-web.pdf](http://packagebuilder.cs.umass.edu/papers/package-recomm-slides-web.pdf)

## Metric Spaces for Temporal Information Retrieval

ECIR 2014

Amsterdam, the Netherlands

Apr 2014

## Metric Spaces for Temporal Information Retrieval

UNIVERSITY OF BOLOGNA

Bologna, Italy

Dec 2013

## Metric Spaces for Temporal Information Retrieval

ÅRHUS UNIVERSITY

Århus, Denmark

Jan 2013