



IfI Advisory Board Meeting 2021

Prof. Dr. Renato Pajarola
Department Chair

Prof. Dr. Burkhard Stiller
Deputy Department Chair



Agenda

Start 15:00

1. Welcome and Agenda
2. Introduction of Board Members
3. Achievements 2019-2021
4. Overview on Studies@IfI
 - Student numbers
 - Examination numbers
 - New study program overview in BSc and MSc
 - Plans for AI Major
5. Professorship Planning
6. Past CASEs
 - Data Science and Machine Learning
 - Blockchains
7. Technical Research Presentation incl. Q&A
8. Closing and Farewell

R. Pajarola / B. Stiller

All

R. Pajarola

T. Fritz

D. Scaramuzza

R. Pajarola

B. Stiller

M. Volk

B. Stiller

D. Olteanu

R. Pajarola / B. Stiller

End 18:00



Material Download Information

- Full slide set available as a PDF file at
 - <https://files.ifi.uzh.ch/stiller/IfI-AdvisoryBoardMeeting2021.pdf>





**Universität
Zürich** ^{UZH}

Institut für Informatik



Welcome



Introduction of Board Members



IfI Professors

Renato Pajarola
Department Chair and
Visualization and
MultiMedia Lab



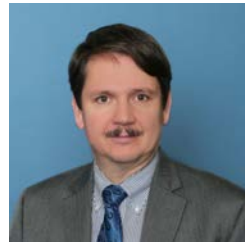
Abraham Bernstein
Dynamic + Distributed
Information Systems
and DSI



Manuel Günther
Artificial Intelligence
and Machine Learning



Burkhard Stiller
Deputy Department
Chair and
Communication Systems



Michael Böhlen
Database Technology



Anikó Hannák
Social Computing



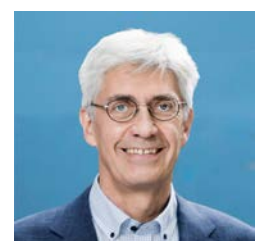
Alberto Bacchelli
Empirical Software
Engineering



Thomas Fritz
Software Evolution and
Architecture



Lorenz Hilty
Informatics and
Sustainability



Jürgen Bernard
Interacting with Data



Harald Gall
Dean WWF and
Software Engineering



Elaine Huang
Human-Computer
Interaction





**Universität
Zürich** ^{UZH}

Institut für Informatik

IfI Professors

Dan Olteanu
Data Systems and Theory



Sven Seuken
Computation and
Economics



Davide Scaramuzza
Robotics and Perception



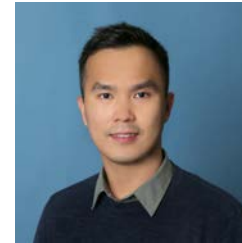
Martin Volk
Computational Linguistics



Ingo Scholtes
Data Analytics



Chat Wacharamanotham
Interaction Design



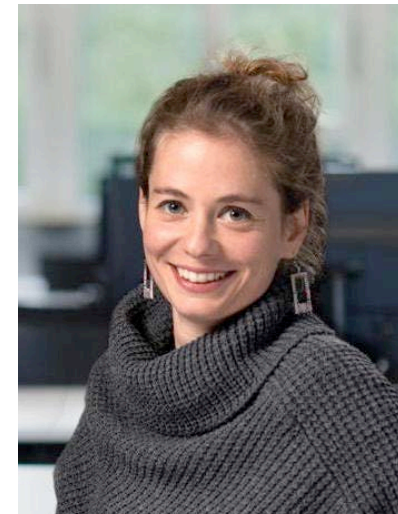
Gerhard Schwabe
Information Management





New Professors at IfI

- **Prof. Dr. Aniko Hannak**
 - Assistant Professor
 - Social Computing
 - Appointed 1 October 2019
 - Start at IfI 1 January 2020





New Professors at IfI

- **Prof. Dr. Dan Olteanu**
 - Full Professor
 - Data Systems and Theory
 - Appointed 1 May 2020
 - Start at IfI 1 August 2020





New Professors at IfI

- **Prof. Dr. Manuel Günther**
 - Assistant Professor
 - Artificial Intelligence and Machine Learning
 - Appointed 1 July 2020
 - Start at IfI 1 July 2020





New Professors at Ifl

- **Prof. Dr. Jürgen Bernard**
 - Assistant Professor
 - Interacting with Data
 - Appointed 1 September 2020
 - Start at Ifl 1 November 2020





Organization of IfI

- **Department Chair**
- **Deputy Chair / Lead Advisory Board**

- **Program Director BSc & MSc**
- **Program Director PhD**

- **Study Advisor**

- **Lead Continued Education**

- **CTO / Colloquium Coordinator**
- **Library Liaison**

Dean WWF – Prof. Dr. Harald Gall

Prof. Dr. Renato Pajarola

Prof. Dr. Burkhard Stiller

Prof. Dr. Gerhard Schwabe

Prof. Dr. Thomas Fritz

Prof. Dr. Sven Seuken

Prof. Dr. Martin Volk

Prof. Dr. Burkhard Stiller

Prof. Dr. Lorenz Hilty



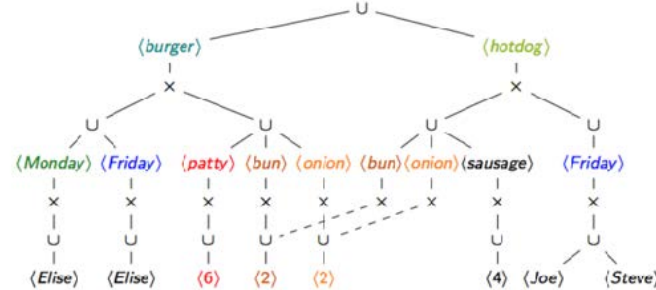
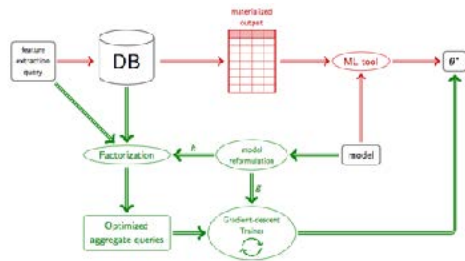
**Universität
Zürich** ^{UZH}

Institut für Informatik



Achievements 2019-2021

Major Grants (Examples)



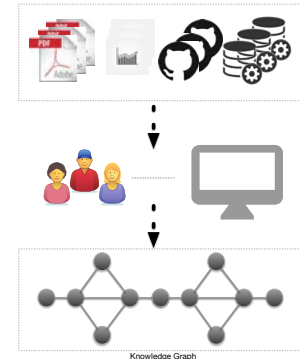
ERC Consolidator Grant (2M CHF) Prof. Dan Olteanu
Foundations of Factorised Data Management Systems



SNF Project Grant 2020 (800K CHF)

Prof. Alberto Bacchelli

Enhanced Code Review:
Using Context and Learning
from Review Experience



SNF Project Grant 2020 (700K CHF)

Prof. Abraham Bernstein

Large-Scale Human-Machine Systems for
Data Science



ERC Consolidator Grant (2M CHF)

Prof. Davide Scaramuzza

Low-latency Perception and Action for Agile Vision-based Flight



Honors

- **Digital Shaper 2019 (Bilanz)**
 - Prof. Abraham Bernstein, Ph.D.
- **FOSS Impact Award 2019 at Software Engineering Conference MSR:
git2net: Mining Time-Stamped Co-Editing Networks from Large git Repositories**
 - Prof. Dr. Ingo Scholtes joint work with Christoph Gote (ETH Zürich)
- **UZH-Lehrpreis 2019**
 - Prof. Dr. Chat Wacharamanotham
- **UZH-Alumni Award 2019**
 - Dr. Alain Gut



Honors

- **Member of the Swiss Academy of Engineering Sciences (SATW) 2020**
 - Prof. Abraham Bernstein, Ph.D.
- **Ric Holt Early Career Achievement Award 2020**
 - Prof. Dr. Alberto Bacchelli for contributions on modern code reviewing by the Mining Software Repositories community
- **Keynote Speaker at Very Large Data Bases 2020 Conference**
 - Prof. Dr. Dan Olteanu



ESEC/FSE 2019: Test of Time Awards

The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering

– ESEC/FSE 2009 Proceedings

- Thomas Zimmermann, Nachiappan Nagappan, Harald Gall, Emanuel Giger, Brendan Murphy: *Cross-project Defect Prediction: A Large Scale Experiment on Data vs. Domain vs. Process* (Prof. Gall)
- Christian Bird, Adrian Bachmann, Eirik Aune, John Duffy, Abraham Bernstein, Vladimir Filkov, Premkumar Devanbu: *Fair and Balanced? Bias in Bug-fix Datasets* (Prof. Bernstein)



Best Paper Awards

- **Best Systems Paper Award at DESRIST 2019**
 - Liudmila Zavolokina, Noah Zani, Gerhard Schwabe: *Why Should I Trust a Blockchain Platform? Designing for Trust in the Digital Car Dossier* (Prof. Schwabe)
- **Best Systems Paper Award at CHI 2019**
 - Alexander Eiselmayer, Chat Wacharamanotham, Michel Beaudouin-Lafon, Wendy E. Mackay: *Touchstone2: An Interactive Environment for Exploring Trade-offs in HCI Experiment Design* (Prof. Wacharamanotham)
- **Best Systems Paper Award at HICSS 2019**
 - Rafael Ziolkowski, Geetha Parangi, Gianluca Miscione, Gerhard Schwabe: *Examining Gentle Rivalry: Decision-Making in Blockchain Systems* (Prof. Schwabe)
- **2nd Best Visual Computer Paper at CGI 2019**
 - Georgios-Tsampikos Michailidis and Renato Pajarola: *ASPIRE: Automatic Scanner Position Reconstruction* (Prof. Pajarola)



Best Paper Awards

- **Honorable Mention Award at EuroVis 2020**
 - Alireza Amiraghdam, Alexandra Diehl and Renato Pajarola: *Locally-adaptive Line Simplification for GPU-based Geographic Vector Data Visualization* (Prof. Pajarola)
- **Best Systems Paper Award at RSS and Best Paper Award Finalist**
 - Philipp Foehn, Dario Brescianini, Elia Kaufmann, Titus Cieslewski, Mathias Gehrig, Manasi Muglikar and Davide Scaramuzza: *AlphaPilot: Autonomous Drone Racing* (Prof. Scaramuzza)
- **Best Presentation Award at GECON2020**
 - Muriel Franco: *ProtectDDoS: A Platform for Trustworthy Offering and Recommendation of Protections* (Prof. Stiller)
- **Best Paper Award at CHI 2020**
 - Chatchavan Wacharamanotham, Lukas Eisenring et al.: *Transparency of CHI Research Artifacts: Results of a Self-Reported Survey* (Prof. Wacharamanotham)
- **Best Paper Award at IEEE ICBC 2021**
 - Christian Killer, Markus Knecht, Claude Müller, Bruno Rodrigues, Eder Scheid et al.: *Æternum: a Decentralized Voting System with Unconditional Privacy* (Prof. Stiller)



Junior Researcher Achievements

- **SNF Ambizione Grant**
 - Dr. Fabio Palomba: TEst CoDe (TED) Project (Prof. Bacchelli)
- **Winner Drone Hero Europe Contest 2019**
 - Davide Falanga, Kevin Kleber: The Foldable Drone (Prof. Scaramuzza)
- **SNF Early PostDoc-Mobility-Stipendium**
 - Dr. Steffen Schuldenzucker: Bridging the Gap Between Fast and Slow Stock Trading Through Algorithmic Market Making (Prof. Seuken)
 - Dr. Gianluca Brero: Towards a General Theory of Mechanism Design für Economies of Automated Agents (Prof. Seuken)
- **The Grace Hopper Celebration Scholarship 2020**
 - Lizeth Fuentes (Prof. Pajarola)



Professorship Offers

- **Daniele Dell’Aglio, Assistant Professor at the Aalborg University in Denmark**
 - Prof. Abraham Bernstein
- **Guillermo Gallego, Associate Professor at TU Berlin**
 - Prof. Davide Scaramuzza
- **Jürgen Hackl, Assistant Professor at University of Liverpool**
 - Prof. Ingo Scholtes
- **Sebastian Proksch, Assistant Professor at the Technical University of Delft**
 - Prof. Harald Gall
- **Rafael Ballester-Ripoll, Assistant Professor at the IE School of Human Sciences and Technology, Madrid**
 - Prof. Renato Pajarola

Visibility of IfI Research and Activities 2019 (for more, see IfI News: <http://www.ifi.uzh.ch/en/news.html>)

6 Mar 2019
In den Medien: Blockchain verstehen



Gerhard Schwabe stellt in der Sendung «Einstein» das Cardossier-Projekt vor.

Glaubt man den Machern, soll die Blockchain zur Wegbereiterin eines neuen werden. Was ist Hype, was Realität?

Am 28. Februar 2019 ist «Einstein» der Frage nachgegangen, wie die B funktioniert.

↓ Gerhard Schwabe über Blockchain-Projekt Cardossier – SRF-Einstein v 28.2.2019 (MP4, 21 MB)

Erica Maurer

29 Mar 2019
In the media: Are faster drones possible?



The camera-based perception the greatest challenge of autonomous drone flight.

Davide Scaramuzza and his team are world leaders in the development can fly in demanding areas without GPS.

→ The New York Times 26.3.2019

Erica Maurer

26 Sep 2019
Abraham Bernstein: Digital Shaper 2019



Abraham Bernstein, Professor am Institut für Informatik und Gründer der Digital Society Initiative der UZH, wird zu einem der 100 Digital Shapers 2019 gekürt. Herzliche Gratulation!

In Zusammenarbeit mit Bilanz, Le Temps und einer Jury hat DigitalSchweizland die 10 wichtigsten digitalen Pioniere der Schweiz aus 10 Kategorien gewählt.

Abraham Bernstein wurde in der Kategorie «Vordenker» ausgezeichnet.

→ Digital Shapers 2019

→ Abraham Bernstein Digital Shaper 2019

Erica Maurer

30 Apr 2019
UZH-Lehrpreis für Chat Wacharamanatham



Chat Wacharamanatham ist Juni 2016 UZH-Professor für Interaction Design bzw. Human Computer Interaction (HCI), interdisziplinäres Fachgebiet zwischen Informatik, Psychologie und Design.

Für sein Engagement erhält er den Lehrpreis 2019 (Credit Suisse Award for Best Teaching).

Herzliche Gratulation!

An der Abstimmung für den Lehrpreis 2019 haben über 2000 Studierende teilgenommen. Massgebend für seine Wahl waren insbesondere die Begründungen der Studierenden. Hier eine Auswahl:

«Chats Lehre ist der Inbegriff von Human Computer Interaction – einfach und intuitiv verständlich.»

«Er beherrscht die Thematik und gibt sie auf interaktive Weise weiter.»

«Er ist ein überaus motivierter und motivierender Dozent, der die Thematik seines Themengebiets beherrscht und dies auf eine interaktive Weise seinen Studierenden weitergibt. In seiner Vorlesung wird über das ganze Semester an einem Gruppenprojekt gearbeitet, das stets wieder evaluiert wird. Dies fördert sowohl die Interaktion der Studierenden untereinander als auch den Austausch zwischen Dozierenden und Studierenden.»

→ UZH-News 29.4.2019

→ Lehrpreis 2019

Erica Maurer

- Gerd Schwabe on Cardossier in **SRF**
- Davide Scaramuzza mentioned in **New York Times**
- Avi Bernstein Digital Shaper 2019
- Teaching Award for Chat Wacharamanatham

Visibility of IfI Research and Activities 2020 (for more, see IfI News: <http://www.ifi.uzh.ch/en/news.html>)



- Aniko, Avi, and Sven portrayed in **UZH Magazin**
- Davide in **Science Robotics**
- Jan Bieser member of **WEF** Global Future Council of Cities of Tomorrow
- Michael, Renato, and Sven UZH Teachers of the Hour
- Avi and Lorenz at **Swiss Digital Days**





New Open Space @ IfI

- Large open space close to main lecture halls BIN 2.A.01 and BIN 2.A.10
- Renovated and refurbished space for lounge like social meeting and work space
- Upgraded with functional and modern furniture as well as adjustable lighting



New Open Space @ IfI

- For all students and employees to use
- Height-adjustable side tables and electrical outlets built into the furniture for comfortable working
- Dedicated sound-proof phone booth
- Accessible during regular opening hours





Overview on Studies@IfI



BSc, MSc, PhD Program

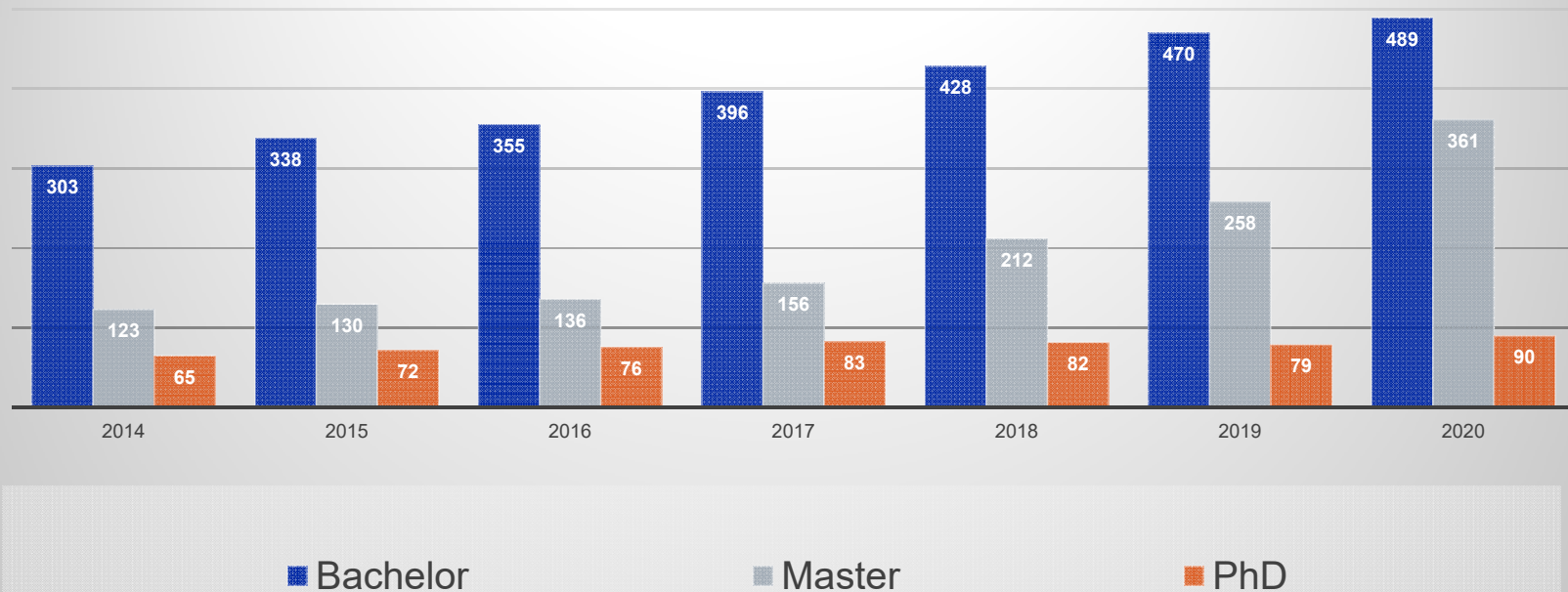
Prof. Dr. Gerhard Schwabe, Prof. Dr. Thomas Fritz, Prof. Dr. Davide Scaramuzza

Table of Content

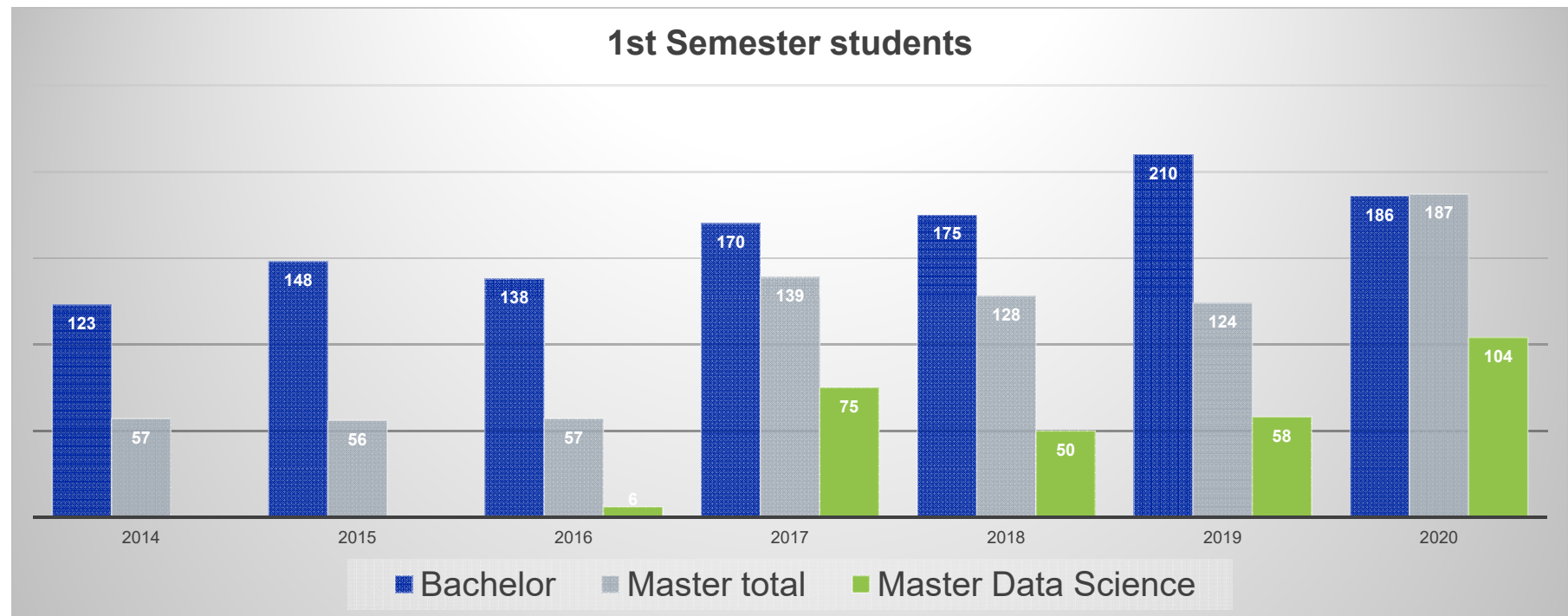
- Student numbers
- Examination numbers
- Overview new study program BSc and MSc
- Plans for AI Master
- Overview planned improvements PhD program

Overview Student Numbers

Students majoring in Informatics total



Overview Student Numbers, 1st Semester



→ High increase on Master level mainly due to Data Science

Changes on Bachelor Level



Universität
Zürich^{UZH}

Department of Informatics

- Change from 4 to 2 specializations in the major (SOSY and WI).
- Computing and Economics will not be offered anymore
- People-Oriented Computing will only be offered as a minor (along with SOSY and WI)
- Informatics and Natural Sciences (INW) remains
- Minor programs remain unchanged, certain changes in modules

New Structure of Bachelor Level



Universität
Zürich^{UZH}

Department of Informatics

1st year (Assessment)

Information Systems	Software Systems	Informatics with Natural Sciences
WI	SOSY	INW
Assessment 60 ECTS		

2nd and 3rd year (Vertiefungsstufe)

Information Systems	Software Systems	Informatics with Natural Sciences
WI	SOSY	INW
Bachelor's thesis 18 ECTS		
15 ECTS		
Common compulsory program 27 ECTS		

1st & 2nd sem

Changes on Master Level

- New major “Artificial Intelligence”
- Computing and Economics will not be offered anymore
- Minor “Information Systems” will be offered for Informatics students
- More structure through the introduction of core elective areas
- Master’s Basic module has been dropped
- Master’s project has been reduced to 15 ECTS



New Structure of Master Level

Five Major study programs 90 ECTScredits

Information Systems	Software Systems	People-Oriented Computing	Artificial Intelligence	Data Science
IS	SOSY	POC	AI	DS
Compulsory module 6 ECTS	Compulsory module 6 ECTS	Compulsory module 6 ECTS	Compulsory module 6 ECTS	Compulsory module 6 ECTS

Master's project
15 ECTS

Core elective area 18 ECTS	Core elective area 18 ECTS	Core elective area 18 ECTS	Core elective area 18 ECTS	Core elective area 18 ECTS
-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------------------------------

INF elective area
15 ECTS

WWF elective area
6 ECTS

Master's thesis
30 ECTS

Minor study program of choice 30 ECTScredits

Informatics
Information Systems
Data Science
Economics
Business Administration
Banking and Finance

Core elective area AI

- Advanced Machine Learning
- Advanced Techniques of Machine Translation
- Algorithmic Game Theory and Mechanism Design
- Big-Data Analytics
- Combinatorial Algorithms
- Computer Graphics
- Deep Learning
- Essentials in Text and Speech Processing
- Machine Learning for Natural Language Processing 1
- Mathematical and Computational Statistics with a View Towards Finance and Risk Management
- Network Science for Business, Economics, Informatics and Social Sciences
- Randomized Algorithms
- Vision Algorithms for Mobile Robotics

Objectives for PhD Program Revisions

- Improve and simplify overall structure as well as workflows
- Create more of a cohort like feeling
- Provide more continuous feedback throughout the PhD
- Facilitate the hiring of highly qualified students into the PhD program

New Curriculum Overview

Core area (*within first 3 semesters*)

- 3 courses (out of core courses offered)

Elective area (*within first 8 semesters*)

- IFI Summer School courses
- External Summer Schools
- Graduate Campus courses (UZH)
- Doctoral level courses announced in the UZH VVZ

Research Colloquia (*within first 8 semesters*)

- 0.17 credits per talk → 18 talks

Teaching Practice and Research Practice (*before registering for the defense*)

- Advisor binary

Core	6 ECTS
Scientific Writing for PhD Students	2 ECTS
Basics in University Teaching	2 ECTS
Research Methodology	2 ECTS
Core elective area	6 ECTS
Research Colloquia	3 ECTS
Teaching Practice	6 ECTS
Research Practice	9 ECTS
Total	30 ECTS



Committees for PhD Milestones

PhD proposal defense – Committee

- 1st and 2nd Reader: external (or internal) with expertise in the area (conflict of interest ok)
- Meta-Reviewer: Member of the PhD committee
- statement from supervisor endorsing (or commenting on) the direction of the proposal

PhD defense – Committee

- Chair: PhD Committee Member
- 1st Reviewer: Supervisor
- 2nd Reviewer: external, *without* conflict of interest
- 3rd Reviewer: external or internal (ok if there is a conflict of interest)
- Potential additional reviewers

Master Fast Track

90 ECTS (instead of 120)

- 60 credits courses
- 30 credits thesis (including a 1-2 page PhD proposa)
- No minor

Admission

- For students that advisor wants to keep as PhD student;
- Prof only admit students into Master FastTrack if they can provide a position when the student successfully completes the program (can be industry funded)
- Students can be switched from MSc to MSc fast track program until start of thesis (makes sense to wait in case students are not known very well yet)
- Require 5.0 (normal MSc admission)



Professorship Planning



Status of Recent Cases

- AOPaP in Empirical Software Engineering – Promotion Alberto Bacchelli
 - Completed in 2021, started on October 1.
- OPaP in Informatics and Sustainability – Promotion Lorenz Hilty
 - Ongoing hiring negotiations
- AOPaP in Blockchain Technologies – Promotion Claudio Tessone
 - Ongoing hiring negotiations
- AOP in Robotics and Perception – Direct Appointment Davide Scaramuzza
 - Starting hiring negotiations

- **Social Informatics**
- Technology, Democracy and Government
- LegalTech
- IT Architecture

- **Electronic Markets and Business**

Computing and Economics
(part of Governing Markets and Institutions)

People-Oriented Computing
(part of Managing People and Organizations)

Big Data Analytics
(part of Methodological Foundations)

- **Cyber-Physical Systems**
- Software Analytics
- Computational Methods in Sustainable Systems Design

- **Information and Open Science**
- **Analytical Computing**
- Computational Thinking

- **Proposed as UZH financed**
- **Proposed as 3rd party sponsoring**



Information and Open Science (Full or Associate Professorship)

- Data is the currency of the digital transformation age
- Libraries as infrastructure for innovative data-driven projects (*Library and Information Science* in the US)
- Open Science increasingly important
- Research foci of professorship in **Information and Open Science**:
 - Information Retrieval
 - Data/Knowledge Management
 - Visual Analytics
 - Data Curation
 - (Linked) Open/Scientific Data
- Complementing professorship in *Digital Libraries* at the Faculty of Arts and Social Sciences
- Collaborating with Digital Society Initiative, Citizen Science, university libraries



Past CASEs



CAS

= Certificate of Advanced Studies

Data Science and Machine Learning

Prof. Dr. Martin Volk, Prof. Dr. Renato Pajarola, Tamar Tolcachier



Central Goal of our CAS

Presentation of the latest developments

1. in Big Data Processing and Deep Learning
2. for Text and Image processing
3. including the most important applications



Target Audience

IT professionals who

- graduated some years ago in Computer Science and
- have some experience on the job

We accept applicants based on their CV.



CAS “Data Science & ML”

- took place 2018, 2019, 2020 (partly online); *canceled in 2021*
- will resume in the spring 2022
- 10 course days, mostly Ifl professors
- participants: 2018: **30** / 2019: **23** / 2020: **23**
- course fee: 8500 CHF



- Probability: A refresher course

- Why do we need mining or learning?
- Learning as induction

- Classification - Learning Decision Trees
- Class Probability Estimation – Naïve Bayes
- Regression – Linear Regression & Neural Networks
- Clustering – K-Means & Hierarchical Clustering
- Association Rules – Apriori

A Data Mining Project A $\rightarrow$ Z

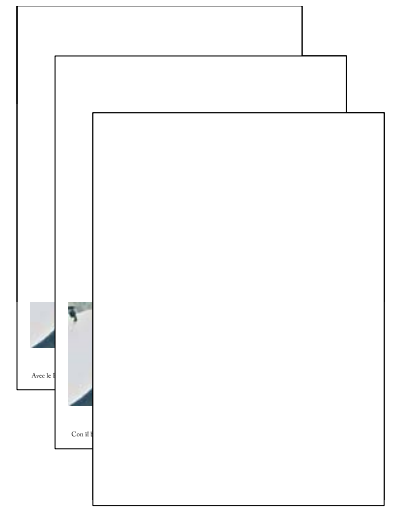
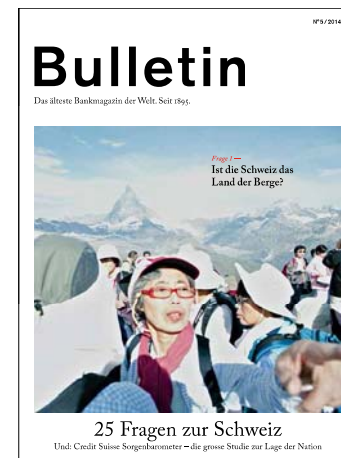




Introduction to Text Analytics and Text Mining

Course Topics

- Introduction to Language Technology
- Multilingual Text Analytics for Big Data
- Named Entity Recognition
- Sentiment Analysis
- Information Extraction from Large Text Collections
- Word Embeddings and Text Classification



Course Structure

- Presentations on Methods and Best Practices
- Tool Overview and Hands-on Exercises
- Group Discussions



Ethical and Legal Aspects

Course Topics

- Introduction to Ethical Aspects
- Ethical Challenges of Big Data
- Learning/Autonomous Systems: Issues of Responsibility
- Introduction to Legal Aspects
- Relevant Fields of Law
- Examples and Case Studies

Course Structure

- Presentations
- Group Discussions
- Groupwork on Case Studies





Examinations

1. Multiple-Choice Test

- 5 questions each to the 5 course days of module 1

2. Course Paper

- 15 pages (DE or EN) about a selected topic under the supervision of one of the lecturers



Lessons Learned – Open Questions

- Recruiting participants for the CAS requires advertising on various channels → suggestions welcome !
- The quality of some course papers is low → how strict shall we be in enforcing minimum quality requirements ? what do comparable CASes require?
- More CAS programs?



Ideas for More CAS Programs?

- 1. AI for Managers**
- 2. Informatics/AI for high school (*Gymnasium*) teachers**
3. AI for Health
4. AI for Law
5. Others ... ???



University of
Zurich^{UZH}

Blockchain Center and Department of Informatics IfI

CAS in Blockchain

A Collaboration of the IfI and the UZH Blockchain Center (BCC)

A Competence Center of the University of Zurich

<https://www.blockchain.uzh.ch/>

Blockchain
Center

BCC Across Disciplines

- Communication Systems Group
- Computation and Economics Research Group
- Criminal Law, Criminal Procedure Law and Criminology
- Dynamic and Distributed Information Systems
- Finance
- Financial Accounting
- Financial Economics
- Informatics and Sustainability Research
- Information and Communication Law
- Information Economics and Contract Theory
- Intellectual Property and Competition Law
- International and Foreign Constitutional Law
- Information Management Research Group
- Network Science
- Private, Economic and European Law
- Public Law
- Quantitative Business Administration
- Quantitative Finance
- Services and Operations Management





BC CAS Administration

- Study program divided into **three modules**:
 - M1: Blockchain Basics (3 ECTS credits)
 - M2: Blockchain Regulation and Law (3 ECTS credits)
 - M3: Blockchain Business and Economics (4 ECTS credits)
- A total of **10 ECTS credits** are required in order to successfully complete the CAS
 - An ECTS credit corresponds to a workload of 30 hours.
 - This comprises attendance time during the course days as well as preparatory and follow-up work.
- An **assessment** must be completed for each module.
 - For M1: multiple-choice test on all topics covered
 - For M2+M3: written paper on a selected topic (on all topics covered during the entire course) must be completed



University of
Zurich ^{UZH}

Blockchain Center and Department of Inf

UZH Blockchain Center





2019 CAS Instance

- 32 attendees
- *Informatics, economics, and law background*





2020 CAS Instance

- 16 attendees
 - Quite many *de-registrations* happened due to the pandemic
 - Potential lack of on-site teaching
 - Change of budgets of companies for “sending” their employees
 - Uncertainty in general
- *Informatics, economics, and law background*
- Started operations finally in hybrid mode
... and had been forced to move to full on-line mode due to the pandemic (legislative orders)



Overview – CAS in Blockchain

Module 1: "Blockchains"

- [Fri 4.9.2020 1.0 Introduction to Blockchains, Prof. Dr. Burkhard Stiller](#)
- [Fri 11.9.2020 1.0 Blockchain Platforms and Architectures, Prof. Dr. Burkhard Stiller, Dr. Thomas Bocek](#)
- [Fri 18.9.2020 1.0 Smart Contracts, Prof. Abraham Bernstein, Ph.D., Dr. Daniele Dell'Aglio](#)
- Thu 24.9.2020 Written Exam (on-line)

IfI Colleagues

Module 2: "Blockchain Business and Economics"

- [Sat 26.9.2020 1.0 Enterprise Blockchains, Prof. Dr. Gerhard Schwabe](#)
- Sat 3.10.2020 1.0 Cryptocurrencies, Prof. Dr. Thorsten Hens
- Fri 9.10.2020 1.0 Applications of Public Blockchains, Prof. Dr. Helmut Dietl
- [Sat 10.10.2020 1.0 Blockchain Cryptoeconomics and Analytics, Prof. Dr. Claudio J. Tessone](#)
- Fri 16.10.2020 Hand-out Written Assignment

“to be” IfI Colleague

Module 3: "Blockchain Regulation and Law"

- Fri 16.10.2020 1.0 Token Regulations and ICOs, Prof. Dr. Rolf H. Weber
- Sat 17.10.2020 1.0 Cryptocurrencies and Smart Contracts, Prof. Dr. Rolf H. Weber, Prof. Dr. Peter Picht
- Sat 24.10.2020 1.0 Data Protection and Property Law, Prof. Dr. Florent Thouvenin, Dr. Alfred Früh
- Thu 29.10.2020 Written Exam
- Thu 10.12.2020 Hand-in Written Assignment
- Thu 17.12.2020 Graduation Party and Handover of Certificates



2021 CAS Instance (Plan and Status)

- Study program **admittance requirements still remain**
 - University degree at Master's level or an equivalent qualification
 - And basic knowledge in the fields of IT, law, *or* economics
- Information event was run as on-line event on Thursday, May 6, 2021
- Operations expected to run in on-site mode
- <https://www.ifi.uzh.ch/en/studies/cas/Blockchain.html>



Key BC CAS Challenge

- Study program **covers three areas of informatics, economics, and law**
 - Excellent combination of areas, which interact in reality
 - The interplay is key for many applications
 - Security concerns on the technical basis impact trust, reputation, and acceptability of application
 - Economic incentive models impact the use of certain algorithms on the technical basis
 - Dedicated angles of laws in countries and regions impact the deployment, use, or validity
- Taught by separate experts in each of these fields
 - University degree at Master's level or an equivalent qualification in ONE area typical
 - Challenging requirements for any of the students and teachers, who cannot, but do assume at least basics in their respective field!



General - CAS' Future

- Discussion started at WWF level to integrate all CAS' into the "Executive School"
 - **Different drawbacks and advantages attached to such a decision**
- Other general topics of interest to IfI professors' have not yet been derived
 - Although topics may be of relevance ...
 - ... a key factor seems to be time and (over-)load of colleagues to invest in CASes



Technical Research Presentation

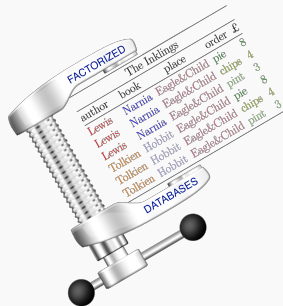
D. Olteanu

Factorised Learning over Relational Data

[fdbresearch.github.io](https://github.com/fdbresearch)

Dan Olteanu
University of Zurich

IfI Advisory Board Meeting
Virtual Zurich, June 9, 2021



Computer Scientist

Since May 2020:

- Full Professor, Department of Informatics, **University of Zurich**
- Head of Data Systems and Theory (DaST) Group

Prior lives (sample):

- Full Professor, Department of Computer Science, **University of Oxford** (previously university lecturer, then associate professor, 13 years)
- Computer Scientist at start-ups **LogicBlox** and **RelationalAI** (7 years)
- Visiting Associate Professor at **University of California, Berkeley** (1 year)
- Research Scholar at **Cornell University** (3 months)
- Postdoc at **University of Saarland** (2.5 years)
- Research and Teaching Assistant at **University of Munich** (4.5 years)
- Research Intern at **INRIA Rocquencourt** (6 months)

Agenda

- Understand computational challenges for data processing
- Design simple and scalable solutions towards these challenges

Theoretical research

- Develop algorithms and languages for data processing
- Analyse how expensive they are and what they can do

Systems research

- Build data systems based on well-understood theory
 - Academic prototypes in the public domain: SPEX, MayBMS, SPROUT, LMFAO, F-IVM, IFAQ, ..
 - Commercial systems: LogicBlox, RelationalAI

Foundations of Data Science course (6 ECTS)

- Mandatory for Master students with minor/major in Data Science
- Also offered to PhD students across UZH
- Mathematical foundations of machine learning
- Strong practical component

Foundations of Computing II course (6 ECTS)

- 3rd semester Bachelor in Informatics course
- Mathematical foundations of computing

Efficient Algorithms course (3 ECTS)

- MSc+PhD level course
- Unifying digest of latest research in efficient computation over relational data
- Strong practical component
- Unique, not offered anywhere else in the world

Modern Data Analytics seminar (3 ECTS)

Acknowledgments for Joint Work on Factorised Learning

FDB team, in particular:



Ahmet



Amir



Haozhe

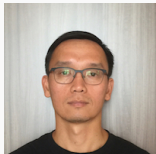


Max

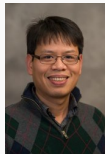


Milos

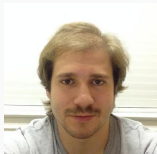
RelationalAI team, in particular:



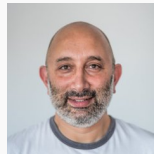
Hung



Long



Mahmoud



Molham

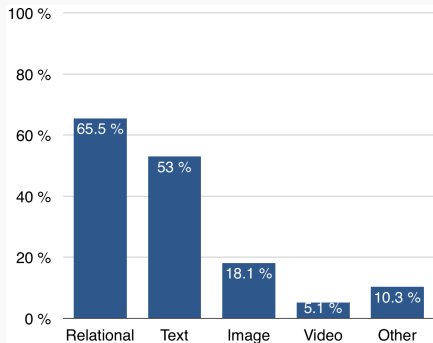
Reasons for DB research community to feel empowered:

1. **Pervasiveness of relational data** in data science
 - Hard fact
2. **Widespread need for efficient data processing**
 - Core to our community's *raison d'être*
3. **New processing challenges** posed by data science workloads
 - Assimilate ideas and applications of related fields
 - to adapt to new requirements and
 - become ever more powerful and versatile

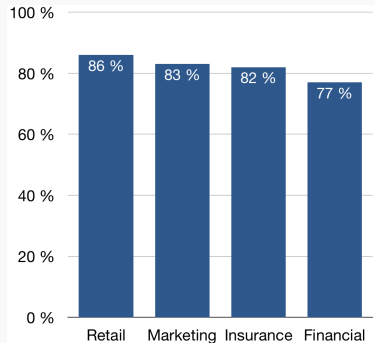
These reasons also serve as motivation for our work.

Relational Data is Ubiquitous

Kaggle Survey: Most Data Scientists use Relational Data at Work!



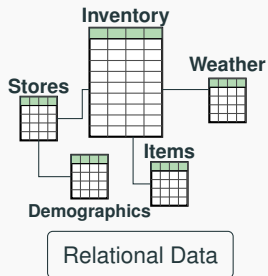
Overall



By Industry

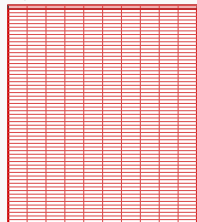
Source: The State of Data Science & Machine Learning 2017, Kaggle, October 2017
(based on 2017 Kaggle survey of 16,000 ML practitioners)

State of Affairs in Learning over Relational Data



Feature Extraction Query
Inventory $\bowtie$ Stores $\bowtie$ Items
 $\bowtie$ Weather $\bowtie$ Demographics

10,000s of Features



Training Dataset

ML Tool

Model

Structure-Agnostic Learning:

1. **Unnecessary** data matrix materialization

Relational structure carefully crafted by domain experts thrown away

2. **Expensive** data move

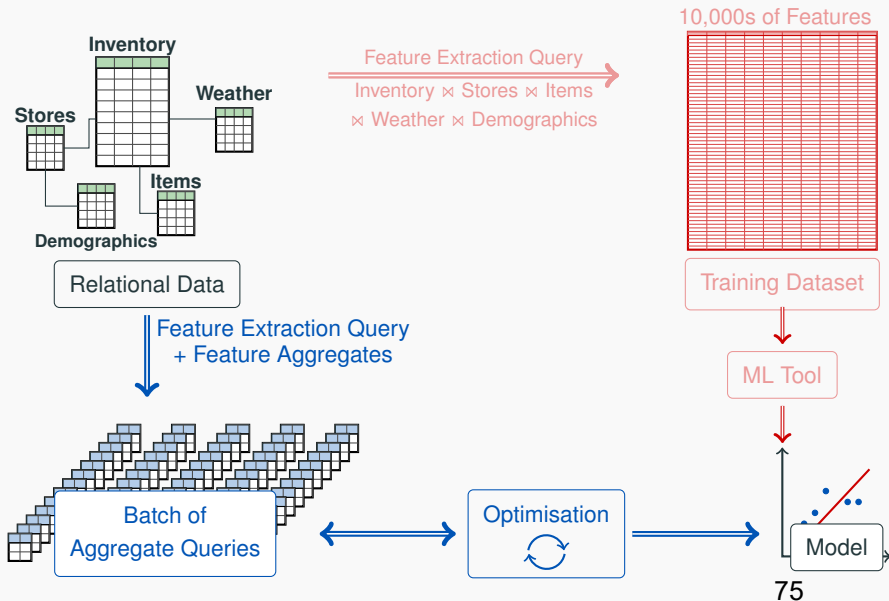
Training dataset can be **order-of-magnitude** larger than the input DB

3. **High** maintenance cost

Recomputation from scratch after updates

4. **Limitations** inherited from both DB and ML tools

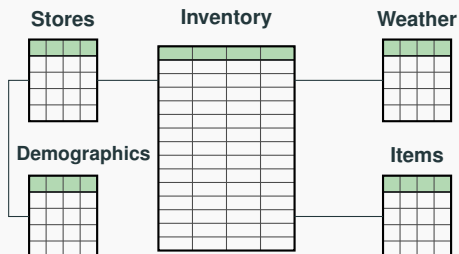
Structure-Aware Learning over Relational Data



Conjecture

The learning time and accuracy of the model can be drastically improved by exploiting the structure and semantics of the underlying multi-relational database.

Structure-aware Learning FASTER than Feature Extraction Query alone



Relation	Cardinality	Arity (Keys+Values)	File Size (CSV)
Inventory	84,055,817	3 + 1	2 GB
Items	5,618	1 + 4	129 KB
Stores	1,317	1 + 14	139 KB
Demographics	1,302	1 + 15	161 KB
Weather	1,159,457	2 + 6	33 MB
Join	84,055,817	3 + 41	23GB

Structure-aware versus Structure-agnostic Learning

Train a linear regression model to predict *inventory* given all features

	PostgreSQL+TensorFlow		LMFAO (SIGMOD'19)	
	Time	Size (CSV)	Time	Size (CSV)
Database	—	2.1 GB	—	2.1 GB
Join	152.06 secs	23 GB	—	—
Export	351.76 secs	23 GB	—	—
Shuffling	5,488.73 secs	23 GB	—	—
Query batch	—	—	6.08 secs	37 KB
Grad Descent	7,249.58 secs	—	0.05 secs	—
Total time	13,242.13 secs		6.13 secs	

2, 160× faster while being more accurate (RMSE on 2% test data)

TensorFlow trains one model. Our approach takes < 0.1 sec for any extra model over a subset of the given feature set.

TensorFlow's Behaviour is the Rule, not the Exception!

Similar behaviour (or outright failure) for more:

- **datasets:** Favorita, TPC-DS, Yelp, Housing
- **systems:**
 - used in industry: R, scikit-learn, Python StatsModels, mlpack, XGBoost, MADlib
 - academic prototypes: Morpheus, libFM
- **models:** decision trees, factorisation machines, *k*-means, ..

This is to be contrasted with the scalability of DB Management Systems!

**How to achieve the performance improvement
of structure-aware learning?**

Idea 1: Turn the ML Problem into a DB Problem



Through DB Glasses, Everything is a Batch of Queries

Workload	Query Batch	# Queries
Linear Regression	$SUM(X_i * X_j)$ [WHERE $\sum_k X_k * w_k < c$]	814
Covariance Matrix	$SUM(X_j)$ GROUP BY X_j [WHERE ...]	
(Non)poly. loss	$SUM(1)$ GROUP BY X_i, X_j [WHERE ...]	
Decision Tree Node	$VARIANCE(Y)$ WHERE $X_j = c_j$	3,141
Mutual Information	$SUM(1)$ GROUP BY X_i, X_j	56
Rk-means	$SUM(1)$ GROUP BY X_j $SUM(1)$ GROUP BY $Center_1, \dots, Center_k$	41

(# Queries shown for Retailer dataset with 39 attributes)

Queries in a batch:

- Same aggregates but over different attributes
- Expressed over the same join of the database relations

AMPLE opportunities for sharing computation in a batch

So far:

- Polynomial regression
- Factorisation machines
- Classification/regression trees
- Mutual information
- Chow Liu trees
- k -means clustering
- k -nearest neighbours
- (robust, ordinal) PCA
- SVM

On-going:

- Boosting regression trees
- AdaBoost
- Sum-product networks
- Random forests
- Logistic regression
- Linear algebra:
 - QR decomposition
 - SVD
 - low-rank matrix factorisation

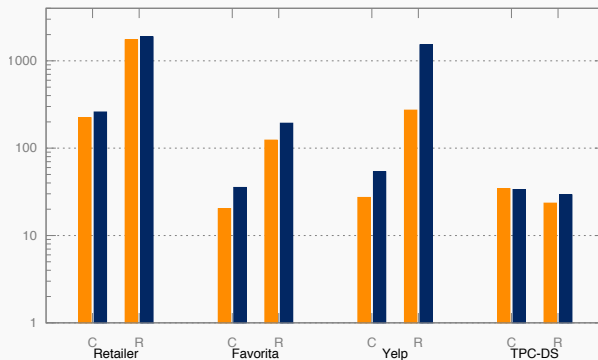
All these cases can benefit from **structure-aware computation**

Natural Attempt:

Use Existing DB Systems to Compute Query Batches

Existing Database Systems are **NOT** Designed for Query Batches

Relative Speedup for **Our Approach** over **DBX** and **MonetDB**

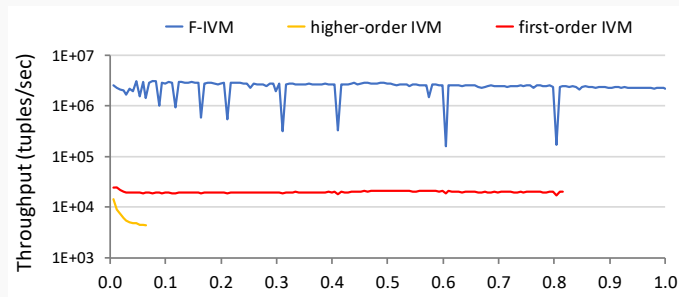


C = Covariance Matrix; R = Regression Tree Node; AWS d2.xlarge (4 vCPUs, 32GB)

Existing Data Stream Systems are **NOT** Designed for Query Batches

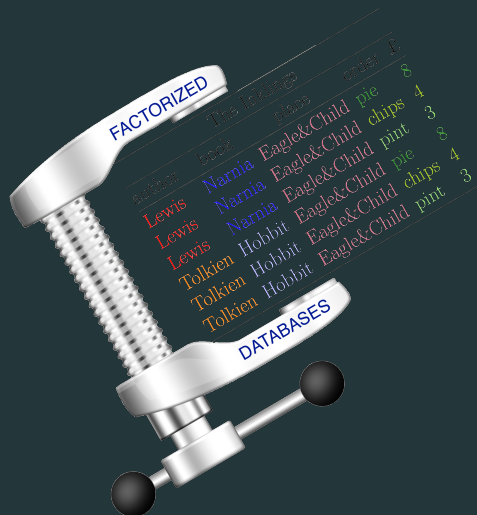
Task: Maintain the covariance matrix over Retailer

- Round-robin insertions in all relations
- All maintenance strategies implemented in DBToaster



Azure DS14, Intel Xeon, 2.40GHz, 112GB, 1 thread; one hour timeout

Idea 2: Exploit Problem Structure to Lower Complexity



	author	book	place	order	F
Lewis	Narnia	Eagle&Child	pie	8	
Lewis	Narnia	Eagle&Child	chips	4	
Lewis	Narnia	Eagle&Child	pint	3	
Tolkien	Hobbit	Eagle&Child	pie	8	
Tolkien	Hobbit	Eagle&Child	chips	4	
Tolkien	Hobbit	Eagle&Child	pint	3	

Algebraic structure: (semi)rings $(\mathcal{R}, +, *, \mathbf{0}, \mathbf{1})$

- Distributivity law $\rightarrow$ Factorisation

Factorised Databases

[VLDB'12+'13, TODS'15, SIGREC'16]

Factorised Machine Learning

[SIGMOD'16+'19, DEEM'18, PODS'18+'19, TODS'20]

- Additive inverse $\rightarrow$ Uniform treatment of updates

Factorised Incremental Maintenance

[SIGMOD'18+'20]

- Sum-Product abstraction $\rightarrow$ Same processing for distinct tasks

DB queries, Covariance matrix, PGM inference, Matrix chain multiplication

[SIGMOD'18+'19]

Combinatorial structure: query width and data degree measures

- Width measure w for FEQ $\rightarrow$ Low complexity $\tilde{O}(N^w)$

factorisation width $\geq$ fractional hypertree width $\geq$ sharp-submodular width
worst-case optimal size and time for factorised joins

[ICDT'12+'18, TODS'15, PODS'19, TODS'20]

- Degree $\rightarrow$ Adaptive processing depending on high/low degrees

worst-case optimal incremental maintenance

[ICDT'19a, PODS'20]

evaluation of queries with negated relations of bounded degree

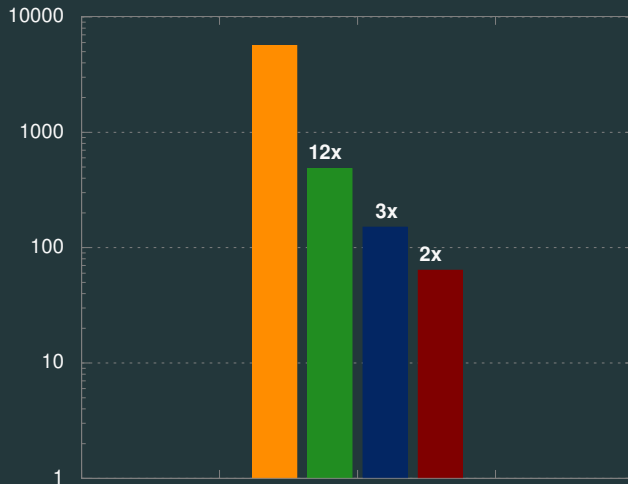
[ICDT'19b]

- Functional dependencies $\rightarrow$ Learn simpler, equivalent models

reparameterisation of polynomial regression models and factorisation machines

[PODS'18, TODS'20]

Idea 3: Lower the Constant Factors



1. **Specialisation** for workload and data

- Generate code specific to the query batch and dataset

- Improve cache locality for hot data path

2. **Sharing low-level data access**

- Aggregates decomposed into views over join tree

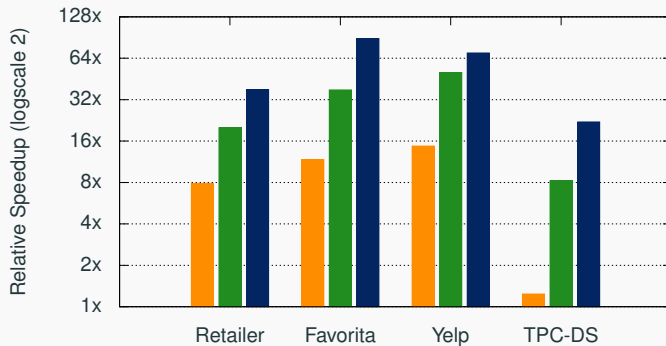
- Share data access across views with different output schemas

3. **Parallelisation**: multi-core (SIMD & distribution to come)

- Task and domain parallelism

[DEEM'18,SIGMOD'19, CGO'20]

Relative Speedup of Code Optimisations



Added optimisations for covariance matrix computation:

specialisation → + sharing → + parallelisation

AWS d2.xlarge (4 vCPUs, 32GB)

Conclusion: Three-step Recipe for Factorised Learning over Relational Data

1. Turn the learning problem into a database problem
2. Exploit the problem structure to lower the complexity
3. Specialise and optimise the code to lower the constant factors

Q.E.D.



**Universität
Zürich** ^{UZH}

Institut für Informatik



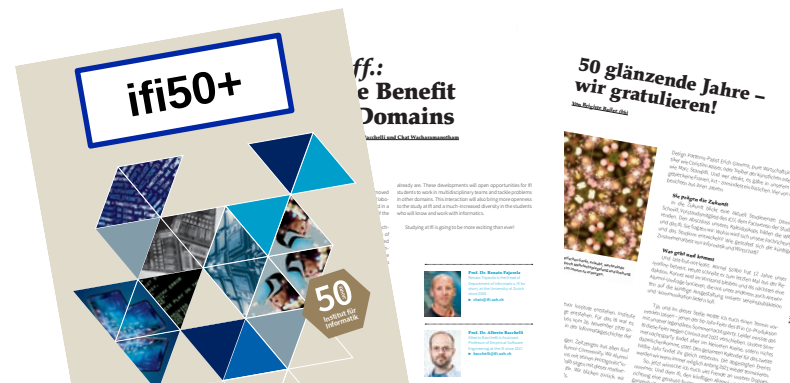
Closing

ifi50 → ifi50+ (or ifi5x or ifi50++ or ifi50 2.0 or ifi50 Ultimate 64 bit Edition or ifi50_(_ツ)_/ or...)

- 1967: Founding of **Institut für Operations Research und Elektronische Datenverarbeitung**
- 1970: Splitting into **Institut für Operations Research** and **Institut für Elektronische Datenverarbeitung**
- 1973: Renamed to **Institut für Informatik**

Founding of IfI ('s predecessor)

- Events planned for 2020 (public symposium with exhibition, department party) postponed
- Publication scheduled for later this year (2021)
- Virtual globe with interactive timeline (persons, achievements, events)
- Perhaps big events in 2023 (50 years “Institut für Informatik”)



IfI Advisory Board Meeting 2021



A Decade of Ifl Advisory Board

- First Ifl Advisory Board meeting was on April 15th, 2011
- But our professors do not seem to be affected by time



Thomas Fritz



Sven Seuken



Davide Scaramuzza

- Weak memories remind us of regularly practiced social proximity





Farewell

For their long service in our Advisory Board, we would like to recognize their efforts and their contributions to our discussions, and wish them all the best in their future activities

- Sabine Wettstein
 - Since 2013
- Bruno Kaiser
 - Since 2014



Next Advisory Board Meeting

- Change to bi-annual meeting
- Planned for June 2023
- Hopefully on-site with more direct and additional information interactions