Bachelor-/Master-Forum 2022

Information Systems and Quantum Computing (ISQC)

Institute of Information Systems (IFIS)

4.11.2022

Professor Dr. rer. nat. habil. Sven Groppe

https://www.ifis.uni-luebeck.de/index.php?id=groppe
Institute of Information Systems (IFIS)

- 5 Professors/PDs
- currently \(\approx\) 15 PhD students, 4 Postdocs, 3 External PhD students
- Labs
  - Information Systems and Quantum Computing (ISQC)
    - Prof. Dr. Sven Groppe
  - Cyber-Physical Medical Systems (CPMS)
    - Prof. Dr.-Ing. Jörg-Uwe Meyer
  - Foundations of AI (FAI)
    - Prof. Dr. Ralf Möller
  - Intellectics
    - PD Dr. Özugür Özçep
  - Human-Aware AI (HAI)
    - Prof. Dr. Nele Russwinkel
Information Systems and Quantum Computing (ISQC)

- **Head:** Prof. Dr. rer. nat. habil. Sven Groppe
- **Projects and Research Assistants**
  - **QC4DB:** Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - N.N.
  - **QualityOnt:** High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **BigSloT:** Big Data Management for the Semantic Internet of Things (DFG)
    - Benjamin Warnke
  - **Hybrid²**-Index Structures for Main Memory Databases (DFG)
    - Tobias Groth
Supervision of Bachelor/Master Thesis & Result

- Often co-supervision of Prof. Groppe together with PhD student
  - meetings regularly and on request, typical:
    - weekly meetings with PhD student
    - monthly meetings with Prof. Groppe

Experience
- 91 supervised thesis (bachelor/master/student research project/Diploma/PhD)

Publications based on bachelor/master thesis
- improves visibility of student's contribution
- improves chances for good job (in academia and industry)
- 43 (out of 141) publications (of ISQC lab) are co-authored by a bachelor/master student (being a student at time of writing)
  - 30% of the publications
Typical Outline of Bachelor/Master Thesis

1. Introduction/Einführung
   1.1. Motivation
   1.2. Tasks of the Thesis/Aufgabenstellung
   1.3. Organization/Organisation der Arbeit

2. Basics/Grundlagen
   2.1. ...
   2.2. Further Related Work/Weitere wissenschaftliche Literatur

3. Concept/Konzept

4. Realization/Realisierung

5. Evaluation

6. Summary and Conclusions/Zusammenfassung und Ausblick

- Latex Template available, e.g.: In Moodle
- FAQs on Bachelor’s and Master’s Theses (from examination board for MINT): in English /German
DVD in addition to Thesis & for IFIS Archive

Please do not forget to burn DVDs for each of the thesis, content:

- **Readme-file** with installation instructions
- **source code** with documentation
  - additionally push to thesis repository in IFIS-Gitlab
- all necessary **third-party-libraries**
- all **data sets for reproducing evaluation** in thesis
- **PDF of the thesis**
- **source files of the thesis** (Word-file/latex-folder)

On the day of defense please deliver **DVD for the IFIS-archive**:

- Content as above
- **PDF of the presentation** for the defense
- **source files of the presentation** for the defense (Powerpoint file/latex-folder etc.)
Typical **Defense of Thesis**

- **20-25 minutes presentation**
  - often similar structure like thesis, but without 4. Realization
- **maybe with succeeding short demonstration (∼ 5 minutes) of developed software (dependent on thesis)**
- **Afterwards questions of reviewers and listeners**
- Reviewers discuss alone in room about result
- **Reviewers announce score** to student and explain the reasons for the score
- **In total: up to 1 hour**
Hybrid Multi-Model Multi-Platform (HM3P) Database

- full and uniform data integration at database level
- performance: fully optimized across different data models
- transparent fault-tolerance
- SQL standards: relational ('87), XML ('03), temporal ('11), JSON ('16), Multi-dimensional Arrays ('19), schemaless ('19), streams ('20?), property graphs ('21?)
- features of different types of databases running on different platforms can be used
Variant: **Semantic HM3P (SHM3P) DB**

Single instance of **SHM3P Database** offers (fully cross-platform optimized) functionality of & replaces...

- **IoT DB On the Edge**
- **Main-Memory DB GPU-accelerated Parallel Server**
- **Quantum DB Quantum Computer**
- **Cloud DB Cloud**

**Reasoning:**
- Lightweight reasoning on large data sizes of IoT devices
- Heavyweight reasoning on moderate data sizes
- Heavyweight reasoning on large data sizes
- Reasoning on small data sizes of mobile devices

How to integrate the different reasoning capabilities and requirements into one transparent global reasoner?

- **Semantic Layer as glue** between other models and platforms
- **new challenges** like integrating different types of reasoners in a transparent global reasoner

**Features of HM3P databases**
- **Easier data integration**
- **Performance issues** may occur due to semantic layer
Types of DBMS

- **state-of-the-art**
- **partly/rudimentarily addressed**
- **visionary/single attempts**
  - Example: hybrid cloud

The Power of Multi-Platform: LUPOSDATE3000

- ultra-fast in jvm...

- ...but also enabling web demos running completely in the browser!

B. Warnke, M.W. Rehan, S. Fischer, S. Groppe: Flexible data partitioning schemes for parallel merge joins in semantic web queries in: BTW’21

Thesis supervised by Prof. Groppe

(S)IoT Database LUPOSDATE3000

- Project BigSloT (DFG, together with ITM) (Benjamin Warnke)
- External PhD student at Bosch (Simon Paasche)
- Combining (e.g., geo) routing protocols with query processing
- Compress data (in storage/messages) according to SHACL definitions, reformulate queries to directly work on compressed data
- Continuous Queries (e.g., redundant processing, efficient recovery after crashes, new types of windows for consistency checks)
- Distributed RDFS Inference
- Digital twins of machines in production
- Green Computing for digital twins/monitoring
- Automatic Testing of Databases: randomized testing, generation of test queries, automatic minimization of test cases, ...
- Semantic Web Layer for NebulaStream (co-supervision with DIMA/Berlin)
Architectures of Emergent Hardware

<table>
<thead>
<tr>
<th>Multi-Core CPU</th>
<th>Many-Core CPU</th>
<th>Graphics Processing Unit (GPU)</th>
<th>Field Programmable Gate Arrays (FPGA)</th>
<th>Quantum Computer/Annealer</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Multi-Core CPU Diagram]</td>
<td>![Many-Core CPU Diagram]</td>
<td>![Graphics Processing Unit Diagram]</td>
<td>![Field Programmable Gate Arrays Diagram]</td>
<td>~100/~5000 qubits</td>
</tr>
</tbody>
</table>

- **Cores**: ~10 ~100 ~1000 ~100 000
- **Core Complexity**: Complex (optimized for single thread performance) Simple
- **Computational Model**: MIMD + SIMD SIMD Data-Flow Universal/Adiabatic Quantum computing
- **Parallelism**: Thread and Data Parallel Data Parallel Arbitrary Quantum Logic Gates/Fluctuation (glob. Min.)
- **Memory Model**: Shared Distributed Quantum Superposition
- **Power**: 150 W 200 W 250 W 50 W 25 KW
- **Database Op.**: Query Optimization (Enumeration of Plans), Concurrency Control

**Legend:**
- Purple: Computational Unit
- Green: Execution Controller
- Yellow: Interconnection Network
- Pink: On-Chip Memory
Timeline of Quantum Computers

- IBM
- IBM Roadmap’20/
  Think’22
- Google
- Google Roadmap’21
- Intel
- Rigetti
- QuTech
- USTC
- Xanadu Quantum Technologies
- Quantum Brilliance
  Roadmap’21
  (Room Temperature)
- D-Wave
  (Quantum Annealing)
- D-Wave Roadmap’21

Main Data Source Roadmaps IBM IBM’25 Google Quantum Brilliance D-Wave #Atoms on Earth #Particles in Universe
Using **Hardware Accelerator** for optimizing Queries / Transaction Schedules
Approaches for Query/Transaction Schedule Optimization

Query Optimization:
\[
\bigotimes_{i=1}^{n} R_i = (R_1 \bowtie R_2) \cdots \bowtie R_n \\
(R_1 \bowtie R_n) \bowtie (\cdots)
\]

Transaction Schedule Optimization:
\[
\{T_1, \ldots, T_m\}
\]

Open Source Relational Database Management System (RDBMS),
e.g. PostgreSQL, MySQL

Dynamic Programming
Random Walk Programming
Simulated Annealing
Linear Programming
Machine Learning
Genetic Algorithm
# Algorithms (used e.g. in Query Optimization) and their Quantum Counterparts

<table>
<thead>
<tr>
<th>Query Optimization Approach</th>
<th>Basic Algorithm</th>
<th>Quantum Computing Counterpart</th>
</tr>
</thead>
<tbody>
<tr>
<td>[S+79]</td>
<td>Dynamic Programming</td>
<td>[R19] [A+19]</td>
</tr>
<tr>
<td>[MP18] [Y+20] [W+19] [O+19]</td>
<td>Reinforcement Learning</td>
<td>[S+21] [DCC05]</td>
</tr>
<tr>
<td>[GPK94]</td>
<td>Random Walk</td>
<td>[ADZ93] [A+01]</td>
</tr>
<tr>
<td>[TC19]</td>
<td>Ant Colony Optimization</td>
<td>[WNF07] [G+20]</td>
</tr>
<tr>
<td>[TK17]</td>
<td>Mixed Integer Linear Programming</td>
<td>[HHL09] [A12] [CKS17] [SSO19] [AL22] [AL22]</td>
</tr>
</tbody>
</table>

This list is not complete...

- Please check my lecture about quantum computing: [https://www.ifis.uni-luebeck.de/~groppe/lectures/qc](https://www.ifis.uni-luebeck.de/~groppe/lectures/qc)
Quantum Machine Learning - Data encoding and Quantum Model

\[ \text{data } x \xrightarrow{\text{data encoding}} f(x, \Theta) \xrightarrow{\text{measurement}} \hat{y} \]

Relations to Join

Join Order
Variational quantum circuits (VQCs) beat classical neural networks for join order optimization.

Extensions to this work ⇒ bachelor/master thesis
Optimizing Transaction Schedules via Quantum Annealing

- Experiments on real Quantum Annealer (D-Wave 2000Q cloud service)
  - first minute free
  (afterwards too much for our budget)
- Versus Simulated Annealing on CPU
- Preprocessing time/Number of QuBits: \(O((n \cdot k \cdot R)^2)\)

<table>
<thead>
<tr>
<th>Fig.</th>
<th>(k)</th>
<th>(n)</th>
<th>(R)</th>
<th>(O)</th>
<th>(l_1, \ldots, l_n)</th>
<th>(r_1, \ldots, r_n)</th>
<th>req. var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>{}</td>
<td>8, 4</td>
<td>0, 4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>{(t_1, t_3)}</td>
<td>4, 5, 1</td>
<td>1, 0, 4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>{(t_2, t_4)}</td>
<td>3, 2, 1, 2</td>
<td>1, 2, 3, 2</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>{(t_1, t_2), (t_4, t_5)}</td>
<td>1, 1, 1, 1, 1</td>
<td>1, 1, 1, 1, 1</td>
<td>10</td>
</tr>
</tbody>
</table>
Open Challenges for QC for Databases/Topics for Thesis

- Replacing basic algorithms with their QC counterparts in query optimizations for speeding up databases
  - Query Optimization: Tobias Winker
  - Transaction Schedule Optimization: Umut Çalıkylimaz

- What should be the properties of a quantum computer (e.g. #qubits, latencies of gates) to achieve certain speedups?

- How to combine classical and quantum computing algorithms to achieve good speedups with few qubits? (...for running database optimizations on current available quantum computers...)

- What other (database) domains besides query and transaction schedule optimizations benefit from quantum computers? (In short: those based on mathematical optimization problems, but also other...?)
## QC4DB: Accelerating Relational Database Management Systems via Quantum Computing

<table>
<thead>
<tr>
<th><strong>Name:</strong></th>
<th>QC4DB: Accelerating Relational Database Management Systems via Quantum Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proj. Web:</strong></td>
<td>[Project Website](Project Website@Quantentechnologien)</td>
</tr>
<tr>
<td><strong>Funded by:</strong></td>
<td>BMBF, Fördermaßnahme <a href="http://www.fordermaessnahmen.de">Anwendungsnetzwerk</a> für das Quantencomputing</td>
</tr>
<tr>
<td><strong>Duration:</strong></td>
<td>3 years</td>
</tr>
<tr>
<td><strong>Volume:</strong></td>
<td>1.8M Euros</td>
</tr>
<tr>
<td><strong>Topics:</strong></td>
<td>Optimizing an open source relational database management system</td>
</tr>
<tr>
<td></td>
<td>- Queries</td>
</tr>
<tr>
<td></td>
<td>- Transaction Schedules</td>
</tr>
<tr>
<td><strong>Partners:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Coord.)</td>
</tr>
<tr>
<td><strong>Expertises:</strong></td>
<td>Hardware-Acceleration of Databases</td>
</tr>
<tr>
<td></td>
<td>Room Temperature Diamond Quantum Accelerators/qbOS</td>
</tr>
<tr>
<td><strong>Website:</strong></td>
<td><a href="https://www.ifis.uni-luebeck.de/~groppe/">https://www.ifis.uni-luebeck.de/~groppe/</a></td>
</tr>
<tr>
<td></td>
<td><a href="https://quantumbrilliance.com/">https://quantumbrilliance.com/</a></td>
</tr>
</tbody>
</table>
COVID-19 Knowledge Graph (KG)

Some issues of knowledge graph quality:
- contradictions in (evolving) facts
- checking vague formulations and compare them with other given information
- errors with ambiguity: E484K mutation of B.1.1.7 is a SARS-CoV-2 virus, but reinfection with E484K is only possible for patients recovered from SARS-CoV-2 virus not mutating E484K (→ introduce class “SARS-CoV-2 without E484K mutation”)

[1] https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774102
[4] https://www.bmj.com/content/372/bmj.n359
High Quality KGs from recent English, French & German Emergent Trends with the example of COVID-19

- Project with partners in Paris & Toulouse (Hanieh Khorashadizadeh)
- Thesis in the areas of
  - data capturing,
  - visualization & analysis,
  - detection of contradictions in KG, ...

![Diagram showing the process of data capturing and quality assurance in KGs]

Legend:
- Data
- Processing Modules
- Sub-Task/Module

- (Facts & Ontology) Quality Rules
- Reasoner
- Human-in-the-loop
- External Ontologies & Knowledge Graphs

- Facts, Ontological Entities/Expressions/Axioms to be checked
- Extracted Information (Facts, Ontological Entities/Expressions/Axioms)

- Knowledge Graph
- Data Analysis & Visualization (WP5)
- Knowledge Injection (WP3)
- Quality Module (WP4)
- Data Capturing & Checker (WP2)
- News Article
- Social Media
- Scientific Publication

Spatial Time Machine: Knowledge at some point of Time (global/national: Fr/Ger/GB/...)
Statistics about Revised & Evolving Knowledge over Time
Effects of COVID-19 Pandemic and Confinements for: Incidence Rates, Economy, Society, Climate Change, ...
Spread of Information
Analytical Comparisons between different Nations
Information Systems and Quantum Computing (ISQC)

- **Head:** Prof. Dr. rer. nat. habil. Sven Groppe
- **Projects and Research Assistants**
  - **QC4DB:** Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - N.N.
  - **QualityOnt:** High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **BigSloT:** Big Data Management for the Semantic Internet of Things (DFG)
    - Benjamin Warnke
  - **Hybrid^2**-Index Structures for Main Memory Databases (DFG)
    - Tobias Groth