
SE Wissenschaftliches Arbeiten,

Gruppen 192/2 abcd

Katja Hose, Matthias Lanzinger, Reinhard Pichler und
Stefan Woltran

<https://www.dbai.tuwien.ac.at/education/gwa/>

Wintersemester 2024/25

Ziele

Kennenlernen von Grundprinzipien des wissenschaftlichen Arbeitens und des Wissenschaftsbetriebs

- Erstellen einer wissenschaftlichen Arbeit
- Begutachtung (Review) wissenschaftlicher Arbeiten
- Einarbeiten der Review-Kommentare
- Präsentation einer wissenschaftlichen Arbeit

Forschungsbereich DBAI

Die Themen kommen aus dem Forschungsbereich 192/2 "DBAI"
(= Datenbanken und Artificial Intelligence)



Gruppe 192/2a: Knowledge-Engineering-Themen (1)

■ ML for DB / DB for ML

- Die Grenze zwischen Machine Learning und Datenbanken verschwimmt immer mehr.
- Datenbanksysteme (Indexe, Anfrageoptimierung, Materialized View Selection, etc.) können durch ML-Ansätze verbessert werden.
- Machine Learning Pipelines können durch effizientes Datenmanagement und Data-Processing effektiv unterstützt werden.
- Beispielstichworte: DB tuning, learned indexes, cardinality estimation, data science pipelines, machine learning frameworks, ML operators in DB, etc.

Gruppe 192/2a: Knowledge-Engineering-Themen (2)

■ Graph Data Management and Querying

- Graphen sind sehr flexible Strukturen, um Informationen darzustellen, insbesondere soziale Netzwerke, biologische Interaktionen, Knowledge Graphs, etc.
- Graphen erfordern nicht nur besondere Ansätze zum Datenmanagement sondern auch zur Anfragebearbeitung und -optimierung in verschiedenen Systemen.
- Beispielstichworte: RDF, property graphs, triple stores, SPARQL, Cypher, graph analytics, path queries, summarization, clustering, graph pattern matching, indexing, partitioning, temporal/dynamic/evolving knowledge graphs, Linked (Open) Data, versioning, archiving, reification, multi-model data management, graph processing frameworks, ...

Gruppe 192/2a: Knowledge-Engineering-Themen (3)

■ Knowledge Graphs

- Knowledge Graphs sind eine flexible Struktur zur Wissensrepräsentation, in der nicht nur Informationen erfasst werden sondern auch deren Bedeutung (Semantics).
- Knowledge Graphs sind die Grundstruktur für Informationen im Semantic Web (Web of Data bzw. Linked (Open) Data), werden aber auch zur Datenintegration und für Machine Learning verwendet.
- Beispielstichworte: neurosymbolic AI, knowlege extraction, entity recognition/resolution/alignment, heterogeneous data integration, data lakes, enterprise knowledge graphs, multi-modal knowledge graphs, semantics, ontologies (OWL, RDFS, RDF), reasoning, machine learning on graphs, graph embeddings, graph neural networks, rule mining, ...

Gruppe 192/2a: Knowledge-Engineering-Themen (4)

- Datenqualität, Explainability und Data Governance
 - Anwendungen inkl. Machine Learning können nur so gut sein, wie die Daten, auf denen sie basieren. Datenqualität spielt daher eine entscheidende Rolle.
 - Fragestellungen beinhalten: Kontrolle des Zugriffs auf Daten, wer hat die Daten bearbeitet, wie wurden sie integriert, wie wurde ein Ergebnis einer Anfrage generiert, kann man den Daten vertrauen, wie korrekt sind die Daten, etc.
 - Beispielstichworte: data profiling, data lakes, data warehouses, trust, data stewardship, data clean(s)ing, SHACL shapes, validation, metadata management, provenance/lineage, dataset search, FAIR¹ data principles, policies, etc.

1: <https://www.go-fair.org/fair-principles/>

Gruppe 192/2b: Datenabfrage-Themen (1)

■ Abfragesprachen:

- Verschiedene Sprachen je nachdem wie die Daten gespeichert sind: Tabellen, Freitext, XML, Knowledge Graphs,
- Wie mächtig sind die jeweiligen Abgabesprachen? Welche Konzepte können damit ausgedrückt werden und welche nicht?
- Die Mächtigkeit von Abfragesprachen steht in der Balance mit der algorithmischen Komplexität der Beantwortung der Abfrage. Wann machen extra Features in der Abfragesprache die Abarbeitung schwieriger?

Gruppe 192/2b: Datenabfrage-Themen (2)

- Graph Neural Networks zur Datenabfrage:
 - Graph Neural Networks (GNNs) sind Deep Learning Modelle die auf Graph Daten (z.B. Knowledge Graphs oder Moleküle) angewandt werden.
 - Ein GNN kann als eine Abfrage an die Daten des Graphs gesehen werden. Diese lassen sich theoretisch präzise in ihrer Aussagekraft analysieren.
 - Durch die Analyse von GNNs als Abfragesprachen können wir die Limits davon erforschen, was GNNs lernen können.

Gruppe 192/2c: Datenbank-Themen (1)

- NoSQL:
 - nicht-relationale Datenformate wie Key/Value Stores, Document Stores oder Column-oriented databases.
 - Einschränkung der "üblichen" DB-Funktionalität von RDBMSs (wie ACID-Anforderung, Anfragesprache) zwecks Performance
- Streaming Daten:
 - Zahlreiche Anwendungen liefern einen (möglicherweise "unendlichen") Datenstrom, z.B.: Sensor Networks, Börsenkurse, Sport Live Ticker, etc.
 - Die Anfragebearbeitung über Streaming Data benötigt spezielle Methoden, da nur sequentieller Zugriff auf einen kleinen Ausschnitt des Datenstroms besteht.

Gruppe 192/2c: Datenbank-Themen (2)

- Security im Datenbankenbereich:
 - Der Schutz von Systemen vor unerlaubtem Zugriff ist ein zentrales Thema in vielen Bereichen der Informatik.
 - Eigene Mechanismen wurden entwickelt, um Daten(banken) vor Beschädigung, Missbrauch, etc. zu schützen.
- Data Integration und Data Exchange:
 - Management von heterogenen Daten
 - Schema Mappings: Austausch/Kombination von unterschiedlich strukturierten Daten

Gruppe 192/2d: Artificial Intelligence Themen (1)

■ Non-Classical Reasoning:

- Abduktion: bezeichnet ein auf Annahmen basiertes Schließen um Beobachtungen zu erklären. Verwandt dazu ist das Problem der logik-basierten Diagnose.
- Closed-World Assumption (CWA): bezeichnet die Annahme, dass alles, was nicht ausdrücklich wahr ist, als falsch angenommen werden kann. Eine Ad-Hoc Anwendung führt schnell zu Inkonsistenzen. Zahlreiche Varianten von CWA wurden daher entwickelt.
- Belief Change: Aktualisieren bzw. Kombinieren von Wissensbasen ist in vielen Applikationen eine zentrale Operation. Ziel ist einerseits Inkonsistenz zu vermeiden andererseits das bereits vorhandene Wissen möglichst unverändert zu lassen.

Gruppe 192/2d: Artificial Intelligence Themen (2)

■ Strategy Games:

- Computerprogramme für "schwere" Strategie-Spiele wie Schach, Go, etc. sind ein klassisches AI-Thema.
- Viele allgemein-verwendbare Ergebnisse (wie z.B. effiziente Suchverfahren) wurden in dieser Forschung entwickelt.

Gruppe 192/2d: Artificial Intelligence Themen (3)

- Computational Argumentation:
 - Beschäftigt sich mit dem Modellieren und Auflösen von Situationen, wo widersprüchliche Meinungen aufeinandertreffen.
 - Methoden reichen von Argument-Mining (erkennen von Argumenten aus Fließtext), über Structured Argumentation (logischer Aufbau eines Arguments), bis zu Abstract Argumentation (Beziehung zwischen Argumenten steht im Vordergrund).

Gruppe 192/2d: Artificial Intelligence Themen (4)

■ Computational Social Choice:

- Entscheidungen zwischen mehreren Alternativen basieren häufig auf Präferenzen.
- Wenn Präferenzen mehrerer Menschen, SW-Agenten, etc. berücksichtigt werden sollen, kann es Widersprüche geben.
- Typische Ziele sind daher die Erstellung einer gemeinsamen Präferenzenliste, die Auswahl einer "besten" Alternative, etc.

Beurteilung

- Voraussetzung: Wer bei der Prüfung für den **allgemeinen Teil** mitgemacht hat, bekommt eine Note. Der allgemeine Teil zählt 17% für die Beurteilung.
- Komponenten der Beurteilung (**spezieller Teil**: 83%):
 - Seminararbeit (ca. 10 Seiten) 60/100
 - Review von 2 anderen Arbeiten (je 1-2 Seiten) 10/100
 - Präsentation (20min) 20/100
 - Einarbeitung der Reviews 10/100
- 5 Termine (siehe <https://dbai.tuwien.ac.at/education/gwa/>)
 - Anwesenheitspflicht bei allen 4 Terminen
(Entschuldigungen **vorab** per mail an gwa@list.tuwien.ac.at)

Organisation und wichtige Termine (1)

1. Einteilung der Themen:

- im Anschluss an den Termin am 21. Oktober (in TUWEL)

2. Abgabe von Titel und Literaturliste:

- bis 27. Oktober 2024 in TUWEL hochladen
- Feedback durch die LVA-Leiter in TUWEL

3. Abgabe von Titel, Literaturliste und Gliederung:

- bis 3. November 2024 in TUWEL hochladen
- Durchsprache im Seminar am 4. November 2024

4. Abgabe der Seminararbeit

- bis 9. Dezember 2024 in TUWEL hochladen
- Anforderungen: Englisch verpflichtend; Latex erwünscht

Organisation und wichtige Termine (2)

- 5. Zuteilung der zu begutachtenden Seminararbeiten:
 - am 10./11. Dezember 2024 durch die LVA-Leiter_innen

- 5. Abgabe der Reviews:
 - bis 19. Dezember 2024 in TUWEL hochladen

- 5. Präsentation im Seminar:
 - Getrennt nach den Gruppe 192/2 a,b,c,d (siehe übernächste Folie)
 - Termineinteilung der Vorträge: Benachrichtung im TUWEL durch die LVA-Leiter_innen nach der Abgabe der Seminararbeiten.

- 5. Abgabe der überarbeiteten Seminararbeit:
 - bis 27. Jänner 2025 in TUWEL hochladen

Vorbereitungstermine

21. Oktober 2024, 9:00 – 11:00, Zemanek SEM:

- Vorbesprechung
- Literatursuche

28. Oktober 2024, 9:00 – 11:00, Zemanek SEM:

- Struktur wissenschaftlicher Arbeiten
- Begutachtung im Wissenschaftsbetrieb

4. November 2024, 9:00 – 11:00, Zemanek SEM:

- Präsentationstechnik

Präsentationstermine

Gruppe 192/2a:	09.1.2025: 09:00–16:00 (FAV 01A)
	10.1.2025: 09:00–16:00 (Gödel SEM)
Gruppe 192/2b:	13.1.2025: 13:00–18:00 (AE U1-1)
	14.1.2025: 13:00–18:00 (AE U1-3)
Gruppe 192/2c:	09.1.2025: 13:00–18:00 (FAV Hörsaal 2)
	10.1.2025: 13:00–18:00 (von Neumann SEM)
Gruppe 192/2d:	09.1.2025: 8:00 – 13:00 (von Neumann SEM)
	10.1.2025: 8:00 – 13:00 (von Neumann SEM)