

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND STATISTIK
INSTITUT FÜR INFORMATIK

LEHRSTUHL FÜR DATENBANKSYSTEME
UND DATA MINING

Seminar

Aktuelle Themen im Bereich Data Science

SoSe19
17.05.2019



DBS

Dozent: Peer Kröger
Betreuung: Anna Beer



Übersicht

- 25.04.19, 12:00 Uhr:
 - Seminarvorstellung
 - Zuteilung Themen Phase 1
- **17.05.19, 10:00- 18:00, Oettingenstr. 67, Raum 169 :**
 - **Vorträge Phase 1**
 - **Jeweils 10-15 Minuten + 5 Minuten Fragen**
 - **Zuteilung Themen Phase 2**
- **14.06.19:**
 - Abgabe der schriftlichen Ausarbeitungen zu Phase 2 (7.000 – 14.000 Zeichen)
 - (Feedback zu Vorversionen nur, wenn sie bis spätestens 05.06. per Mail eingegangen sind)
- **27. + 28.06.19, jeweils 10:00-16:00 Uhr**
 - **27.06.19:** Edmund-Rumpler-Str. 13, B185
 - **28.06.19:** Oettingenstr. 67, C007:
 - Vorträge Phase 2
 - Jeweils 20-25 Minuten + 10 Minuten Fragen

Hinweise

- Alle Themen werden verlost
- **Alle Vortragsfolien** in Uniworx hochladen **bis 16.05. bzw 26.06., spätestens 16:00 Uhr.**
Format: PDF. (Falls PowerPoint gewünscht, bitte sowohl die pptx, als auch ein daraus erstelltes pdf). Wichtig: mit Seitenzahlen.
Template: [http://www.dbs.ifi.lmu.de/cms/studium lehre/bama theses/index.html#BA](http://www.dbs.ifi.lmu.de/cms/studium_lehre/bama_theses/index.html#BA)
Benennung: Vortrag<Nummer des Themas><Thema>.pdf
Beispiel: Vortrag01Präsentationen.pdf
- Die **Ausarbeitung** wird mit LaTeX gemacht, Abgabe als PDF bis spätestens **14.06.19** (4 Wochen Zeit).
Template: <https://www.overleaf.com/latex/templates/association-for-computing-machinery-acm-sig-conference-proceedings-template/bmvfhcdxfty>
Benennung: Ausarbeitung<Nummer des Themas><Thema>.pdf
- Sprecht euch mit den jeweils vor und nach euch Vortragenden ab! Noch wichtiger falls ihr Teil a oder b eines Doppelthemas habt.

Phase 2: Ausarbeitungen- Formal

- Sprache: Englisch (empfohlen) / deutsch
- Abgabe bis 14.06.19, über uniworx (spätere Abgaben werden nicht akzeptiert)
- Abgabe wird nur mit korrekter Benennung akzeptiert
- Umfang: 7 000 – 14 000 Zeichen
- Freiwilliges (aber empfehlenswertes) Zwischenstandstreffen mit mir
- Auf das achten, was ihr gerade gelernt habt, u.a.:
 - Korrektes zitieren, alle Quellen angeben, nicht einfach übersetzen/abschreiben
 - Wissenschaftlicher Schreibstil
 - Ansprechende, ordentliche Form. Gerne eigene Abbildungen erstellen- diese haben im Paper häufig einfach nur keinen Platz
 - Eigene Recherche zum Thema durchführen
 - Aufbau und Sprache klar, einfach zu verstehen, und fehlerfrei
 - ...

Phase 2: Überblick

- Überblick geben und in das Thema einführen
- Grundlagen und bekannteste Ansätze erklären
- Herausfinden, was für das Gebiet wichtig ist
- Welche Problemstellungen gibt es?
- ...



Verlosung



Phase 2: Themen

1. Clustering
2. Curse of Dimensionality
3. Subspace Clustering
4. Correlation Clustering
5. Clustering of Time Series
6. Clustering in Streams
7. Distributed Clustering of Big Data
8. Anytime Clustering
9. Interactive Clustering
10. Clustering of Graphs
11. Spectral Clustering, Graphs, and Linear Algebra
12. Index Structures used for Clustering
13. Process Mining
14. Evaluation Measurements used in Clustering
15. Outlier Detection
16. Data Cleansing

Fragen?

