Get-Together mit den Erstsemestern des Master Wirtschaftsinformatik und MMDS

Research Group Data and Web Science
Data and Web Science Research Group

- 7 Professors
- 4 Post-docs
- 24 PhD students
Data and Web Science Research Group – Research Areas

- **Artificial Intelligence** (Prof. Heiner Stuckenschmidt)
  - knowledge representation formalisms and reasoning techniques for information extraction and integration

- **Data Analysis** (Prof. Rainer Gemulla)
  - methods for analyzing and mining large datasets as well as their practical realizations and applications

- **Natural Language Processing** (Prof. Simone Ponzetto)
  - knowledge acquisition, knowledge-rich language understanding, Computational Social Science and Digital Humanities

- **Statistical Nat. Lang. Processing** (Prof. Goran Glavaš)
  - modeling meaning of language, understanding text, and structuring knowledge from text
Data and Web Science Research Group – Research Areas

- **Image Processing** (Prof. Dr.-Ing. Margret Keuper)
  - Image Segmentation, Motion Segmentation, Efficient Video Segmentation, Semantic Segmentation, Multiple Object Tracking

- **Web-based Systems** (Prof. Chris Bizer)
  - large-scale data integration, evolution of the World Wide Web from a medium for the publication of documents into a global dataspace

- **Data Science** (Prof. Dr. Heiko Paulheim)
  - using web data as background knowledge in data mining, and data mining methods to create and improve large-scale knowledge bases
DWS Overall Research Goals:

1. methods for understanding large and heterogeneous data
2. application of these methods in different contexts
The DWS Group offers the following courses for master students:

- Relational Learning
- Data Mining and Matrices
- Semantic Web Technologies
- Computer Vision
- Web Data Integration
- Web Mining
- Image Processing
- Data Mining II
- Text Analytics
- Information Retrieval
- Decision Support
- Data Mining I
- Large-Scale Data Management
- Database Technology (MMDS)

- Offered this FSS semester.
- Offered in HWS.
IE 500: Data Mining 1

- Lecture contents – the basics of “torturing data”:
  1. Clustering: *How to automatically organize your MP3 collection?*
  2. Classification: *Will your bank grant you a loan?*
  3. Regression: How to determine the price of a house?
  4. Association Analysis: *Which products to place together in a supermarket to maximize customer purchases?*
  5. Text Mining: *Do students on Twitter like or dislike this lecture?*

- Exercises
  - Experiment with Rapidminer or Phython

- Student project:
  - Mine some data of your choice

- Teaching staff:
  - Prof. Dr. Christian Bizer (Lectures)
  - Oliver Lehmborg, Kiril Gashtevski, Daniel Ruffinelli (Exercises)
Advanced Data Mining methods
- Dimensionality Reduction
- Anomaly Detection
- Time Series Analysis
- Parameter Tuning
- Ensemble Learning
- Online Learning

Organization:
- Lectures and Exercises
- Participation in Data Mining Cup
- Opportunity to become a certified RapidMiner Data Analyst

Teaching staff:
- Prof. Dr. Heiko Paulheim (Lectures), Oliver Lehmberg (Exercises)
IE 671: Web Mining

- Approaches to mine knowledge from the Web
  - Web Usage Mining
  - Web Structure Mining
  - Web Content Mining

- Course Structure:
  - Lectures and exercises
  - Projects (during the second half)

- Teaching staff:
  - Prof. Dr. Simone Ponzetto
  - Prof. Dr. Goran Glavas
  - Dmitry Ustalov
IE 673: Data Mining and Matrices

- Matrices and tensors are powerful representations of data
  - Data points, sets, graphs, relational data, knowledge bases, ...

- Course goal: Learn how to analyze such data
  - Course covers theory and applications of dimensionality reduction, embeddings, denoising, discovery of latent structure, visualization, prediction, clustering, pattern mining, topic modelling, ...
  - Focus is on unsupervised and semi-supervised learning & matrix decompositions

Matrix representations:

<table>
<thead>
<tr>
<th>Anna</th>
<th>Bread</th>
<th>Butter</th>
<th>Beer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Charlie</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Customer transactions

<table>
<thead>
<tr>
<th>Book 1</th>
<th>Data</th>
<th>Matrix</th>
<th>Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book 2</td>
<td>5</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Book 3</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

Document-term matrix

<table>
<thead>
<tr>
<th>Avatar</th>
<th>The Matrix</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Bob</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Charlie</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Incomplete rating matrix

<table>
<thead>
<tr>
<th>Saarbrücken</th>
<th>Jan</th>
<th>Jun</th>
<th>Sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helsinki</td>
<td>6.5</td>
<td>10.9</td>
<td>8.7</td>
</tr>
<tr>
<td>Cape Town</td>
<td>15.7</td>
<td>7.8</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Cities and monthly temperatures
IE 673: Data Mining and Matrices

- Instructor: Rainer Gemulla
- Tutor: Yanjie Wang
- 2 SWS lecture, 2 SWS tutorium, 6 ECTS
- IE 500 Data Mining I recommended
- Gain hands-on experience
  - Smaller exercises to deepen lecture material
  - Homework assignments to analyze real data
  - Learn R
- Passing requirements
  - Regular assignments
  - Final exam or oral examination
Understanding end-to-end search systems
IE 663: Web Search and Information Retrieval

- **Lecture contents**: understanding search systems
  - Boolean and vector space retrieval models
  - Probabilistic and lang. modeling retrieval
  - Semantic and Latent Retrieval
  - Web search: Link-based algorithms

- **Teaching staff**:
  - Goran Glavaš (Lectures)
  - Robert Litschko (Exercises)

- **Team Project**:
  - Build your own search engine!
CS 646: Higher Level Computer Vision

- Lecture contents
  - Object Detection
  - Semantic Image Segmentation
  - Optical Flow
  - Video and Motion Segmentation
  - Deep Learning for Computer Vision

- Organization
  - Lectures and Exercises
  - Gain practical python and Matlab coding experience in the exercises

- Teaching Staff
  - Margret Keuper (Lectures and Exercises)
CS 460: Database Technology

- MMDS fundamental course
- Relational databases are the most prominent data storage paradigm
- Understand the concepts of RDBMS
  - Relational modeling and normal forms
  - Query processing and optimization
  - Transactions, concurrency, and recovery
- Teaching staff:
  - Prof. Dr. Heiko Paulheim (Lectures)
  - Sven Hertling (Exercises)
CS 704: Artificial Intelligence Seminar

Topic: Analyzing Financial Networks

Lecturer:

Dr. Ioana Karnstedt-Hulpus
ioana@information.uni-mannheim.de

Prof. Dr. Heiner Stuckenschmidt

Network Types

• Interbank Networks
• Investment Networks
• Personal Networks

Methods

• Macro-Structures
• Neighbourhoods
• Centrality
• …
In this seminar, you will
- Learn about advancement research topics in NLP
- Read, understand, explore, and present scientific literature

This term: **Statistical Machine Translation**

Check out website and register **asap (Feb 13)**

Instructors: Sanja Stajner, Simone Ponzetto
In this seminar, you will

- explore and experiment with a popular machine learning platform of your choice,
- solve a small, self-defined machine learning problem with this platform,
- give an overview over the platform, your problem, and your solution

- Prerequisites: Data Mining I, suitable programming experience
- Check out website and register until Feb 11
CS 715: Large-Scale Data Integration Seminar

• Covers topics such as
  • Data Search
  • Holistic Schema Matching
  • Collective Instance Matching
  • Truth Discovery
  • Set Completion

• Instructors:
  • Christian Bizer, Anna Primpeli,
  • Oliver Lehmberg, Yaser Oulabi

• Prerequisites:
  • Web Data Integration recommended

• Register until Feb 16th
Announcement: New Lecture in Fall 2018

New Lecture on Relational Learning

Lecturer: Prof. Dr. Heiner Stuckenschmidt
Tutorials: Dr. Christian Meilicke & Manuel Fink

Learn how to combine the
*representational power of logics* with the
*data mining capabilities of machine learning*

Learn how to learn expressive rules as the one below
given the input on the right:

*A molecule is active IF it contains a ring R of size 5 and atoms Y and Z that are connected by a double bond, such that Y also belongs to R.*
The DWS group continuously hires good students.

- To work on:
  - Data and Web Mining projects
  - Information Extraction and Integration projects
  - Knowledge Representation and Reasoning projects
  - Implement open source tools

- 30-60 h/month contracts are possible.

- Contact PostDoc or Professor responsible for the project/area that you are interested in.
  - Include CV and overview about your marks.

- Good start for writing your master thesis within group.