

Organization

Dr. Steffen Herbold
herbold@cs.uni-goettingen.de

Outline

- Research Group
- Course Administration
- Contents of the Course
- Literature Hints

Research Group

Software Engineering for Distributed System



Jens Grabowski



Steffen Herbold



Annette Kadziora



Heike Jachinke



Gunnar Krull



Ella Albrecht



Johannes Erbel



Fabian Gumz



Patrick Harms



Verena Herbold



Fabian Korte



Philip Makedonski



Alexander Trautsch



Fabian Trautsch

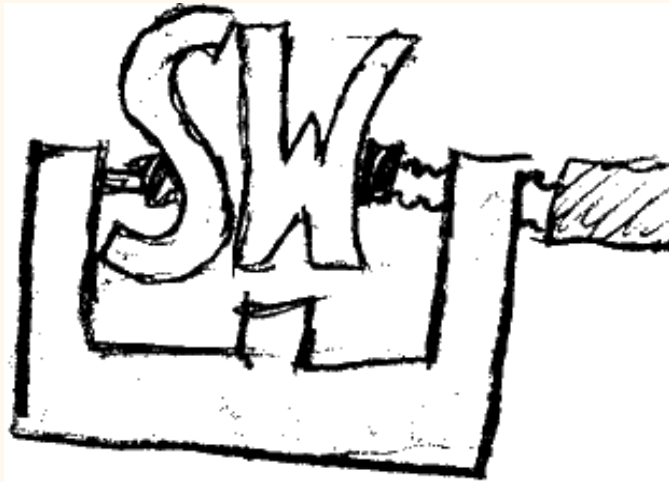
Research Group

- Student assistants:
 - Studierendenbibliothek
 - Continuously 2-3 students for research activities
- Students projects, Bachelor and Master thesis:
 - continuously >5 students

Contact us if you are interested!

- Supported by:
 - Annette Kadziora
 - Dipl.-Ing. (FH) Gunnar Krull
- Web: <http://www.swe.informatik.uni-goettingen.de>

Current Research Topics



- Data Science
 - Methods, Applications, Infrastructures
- Quality Assurance (QA)
 - Managed Software Evolution
 - Simulations of projects
 - Usability Engineering
 - AR, MR, VR
 - Test Languages
 - TTCN-3, TDL
 - QA Cloud Systems
 - Optimized deployments
 - Software Mining
 - Defect Prediction
 - Test usage
 - Developer types
 - Topic modeling

Outline

- Research Group
- **Course Administration**
- Contents of the Course
- Literature Hints

Administration

- Time and Place

- Lecture

- Tuesdays, 14:15 – 15:45 o'clock
 - Room: Provisorischer Hörsaal B

- Exercise

- Thursdays, 13:15 – 15:45 o'clock
 - Room: Ifi -1.101

- Modules and Credits

- 5 ECTS

- M.Inf.1151: Data Science und Big Data Analytics
 - 2.17 Data Warehousing and Data Mining Techniques

Exercises

- Application of concepts from the lecture
 - Lecture sheets as Jupyter Notebooks
 - Mostly programming
 - Python or R
 - Should be solved on your own
- Exercise Sessions
 - Discussion of solutions
 - Help in case of problems
 - NOT for solving the exercises!
 - Attendance optional

Data Science Project

- Requirement for the exam
 - Mandatory for everyone!
 - No project = no exam
- We will provide a dataset
- You will have to solve a task
 - Probably in groups
 - Depends on the number of participants
- Starts in December and goes until the end of the term
- Details for the registration, etc. will follow in November

Examination

- Written exam after completion of the lecture
 - Preliminary date:
- Participation in the project work precondition for the exam

Communication and Course Materials

- Announcements and course materials are distributed via Stud.IP
- Course material
 - Electronic versions (PDF) of the slides
 - Exercise sheets (Jupyter Notebooks)

Outline

- Research Group
- Course Administration
- **Contents of the Course**
- Literature Hints

Contents

- General Concepts and Definitions
 - Big data, data science, business intelligence
- Process of projects
 - Phases, roles, deliverables
- Methods for data analysis
 - Exploration
 - Associations, clustering, classification, regression, time series
 - Statistical tests
- Big data processing
 - Map/Reduce
 - Apache Spark

Outline

- Research Group
- Course Administration
- Contents of the Course
- **Literature Hints**

Literature Hints (1/2)

- Machine learning in general
 - T. Hastie, R. Tibshirani, J. Friedman: *The elements of statistical learning data mining, inference, and prediction*, Springer, 2009
 - D. J. C. MacKay: *Information theory, inference, and learning algorithms*, Cambridge Univ. Press, 2008
 - T. M. Mitchel: *Machine Learning*, McGraw Hill, 1997
- Scikit-Learn+TensorFlow
 - A. Géron: *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly, 2017
- R
 - M. Kuhn, J. Kjell: *Applied Predictive Modeling*, Springer, 2013
- Weka
 - I.H. Witten, E. Frank, M.A. Hall, C.J. Pal: *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016

Literature Hints (2/2)

- Parellel Programming
 - Peter S. Pacheco: *An Introduction to Parallel Programming*, Morgan Kaufmann, 2011
- Big data processing
 - Donald Miner, Adam Shook: *MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems*, O'Reilly, 2012
 - Tim White: *Hadoop: The Definitive Guide*, O'Reilly, 2012
- DISCLAIMER:
 - This kind of literature is horribly fast outdated