

A Corpus of Research-Based and Group-Based Qualification Works: From Proposal to Submission, to a Peer-Reviewed Paper

Prof. Dr. habil. Ansgar Scherp

Tag der Lehre, 19. October 2023

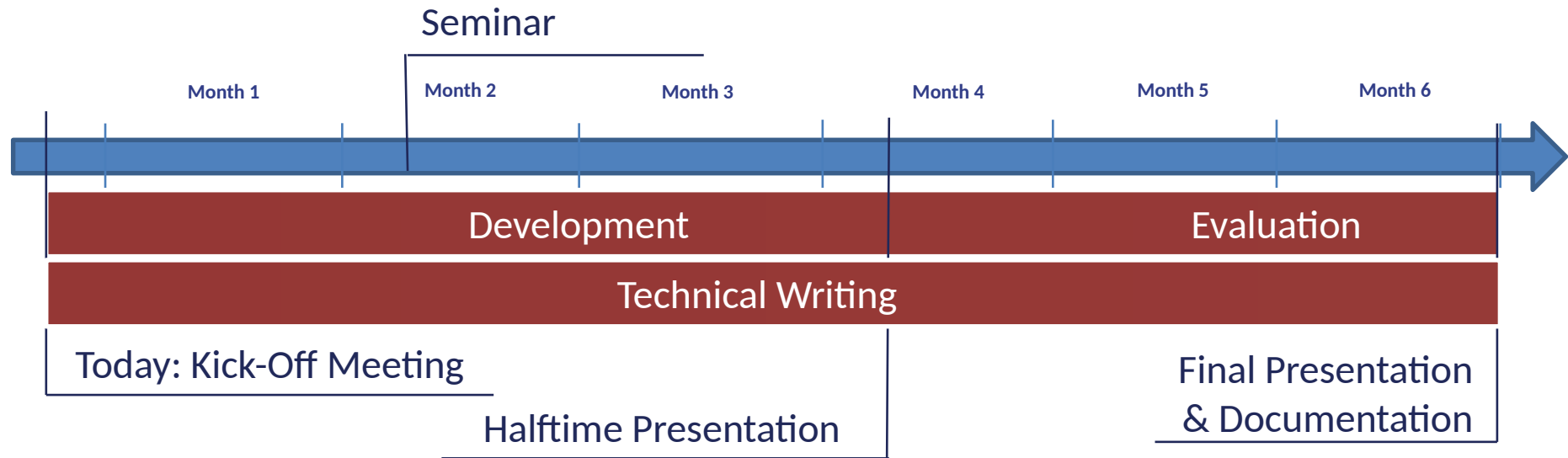
Ulm University, Ulm, Germany

Research- & Group-based Projects

- Students' assignments are a data analysis tasks (text, graph) or advanced application development
- Conducted in a small group of students during the term
- Groups independently plan, design, implement, and evaluate their solution parallel to the weekly lectures
- Results are written up in a scientific report !
- Variant as BSc or Msc thesis: **solo-work, but same idea**
- **Pro and Con, Rationale** → see TdL 2021 Presentation

Project Timeline (6 Months)

Also a 12 months option available!



- Months 1-4 are for development, months 5+6 for evaluation
 - Yes, this means: “feature freeze” after month 4 !
- Scientific paper and technical paper writing from early on !
- Seminar phase (optional)

How Did We Start?

- Introduced research- & project-based group assignments in 2008, *replacing traditional exercise sheets*
- *Initially*, with final oral exam, later: coursework only

Examples of IMS 2008/2009

- "Park Problem Lösung" by R. Bing, U. Buchholz, S. Dittmann, and R. Schrage
→ Interactive game to identify parking violators
- "Toaster-Director" by F. Altherr, C. Effert, D. Eissing, and S. Thierfelder
→ Highly-interactive tutor for a toaster

Note: At that time, application oriented, thought this motivates students. But with paper in mind, e.g., Related Work, Innovation, etc.

PPL

Park - Problem - Lösung



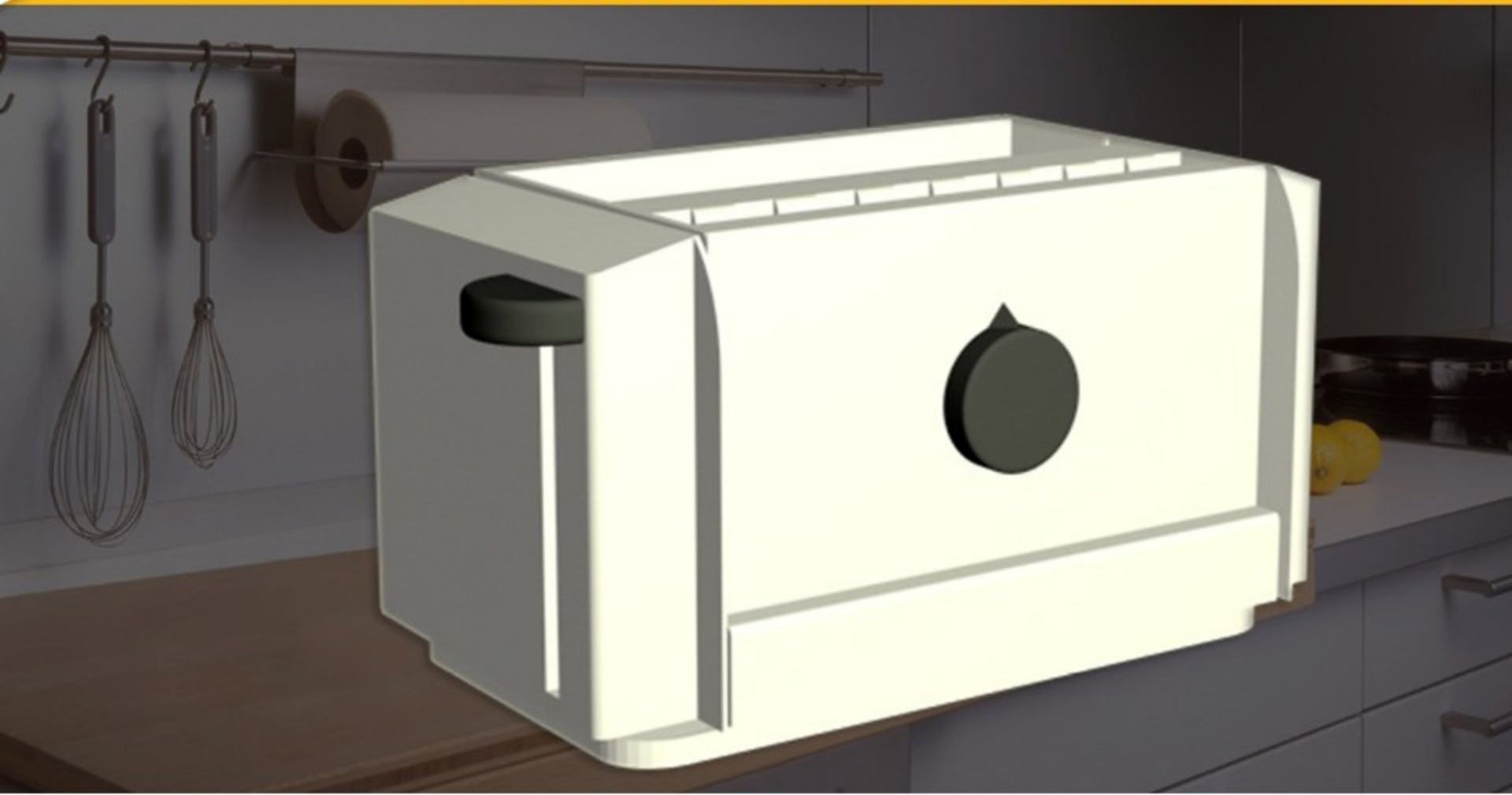
identifiziere die Falschparker



Toaster Director



Florian Altherr, Carsten Effert, Daniel Eissing, Susanne Thierfelder



Where are We Today?

- Introduced research- & project-based group assignments in 2008, *replacing traditional exercise sheets*
- *Initially*, with final oral exam, later: coursework only
- 15 years later
 - Many successful works bc. of scientific rigor
 - Many works also got published (ratio increases!)
 - What was “unthinkable” in beginning is now standard
- Still: there is a lot of room for improvements

Corpus of Qualification Works

- Natural question: Can you give us some examples
- So, we collected (selected) over the last 10+ years
- Covers 15+ examples of different types of modules
 - Coursework during lecture
 - Project groups (run for six or twelve months)
 - Bachelor's and Master's theses
 - Seminar theses
- Overall: 20+ publications, incl. at A-ranked conferences, best paper nominations, best-paper awards

RESEARCH-ARTICLE



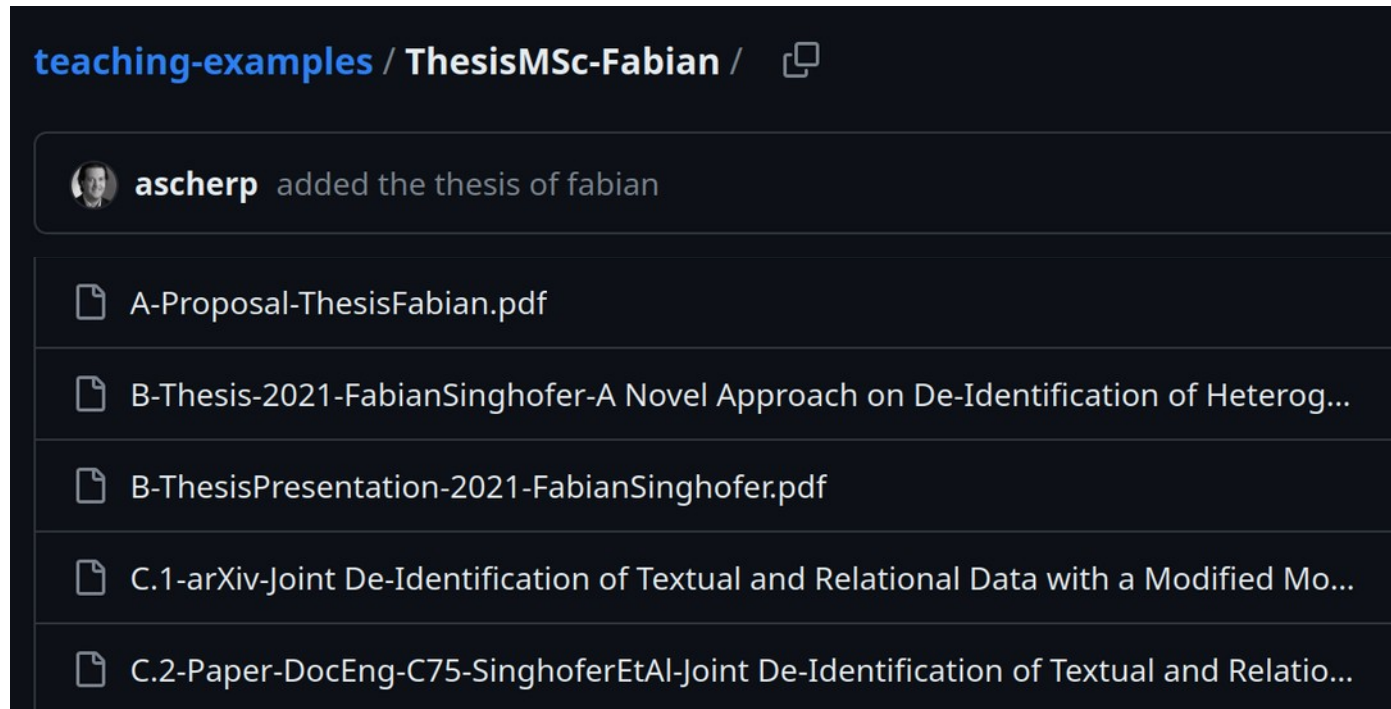
A novel approach on the joint de-identification of textual and relational data with a modified mondrian algorithm

Authors:  [F. Singhofer](#),  [A. Garifullina](#),  [M. Kern](#),  [A. Scherp](#) [Authors Info & Claims](#)

Corpus of Qualification Works (2)

- Allow insights at different stages of a work:
 - (A) writing the proposal,
 - (B) submission to the University,
 - (C) submission at arXiv,up to the camera-ready versions

- Example



The screenshot shows a GitHub repository interface. At the top, the repository name is 'teaching-examples / ThesisMSc-Fabian'. Below this, a commit by user 'ascherp' is shown with the message 'added the thesis of fabian'. The commit includes a list of files:

- A-Proposal-ThesisFabian.pdf
- B-Thesis-2021-FabianSinghofer-A Novel Approach on De-Identification of Heterog...
- B-ThesisPresentation-2021-FabianSinghofer.pdf
- C.1-arXiv-Joint De-Identification of Textual and Relational Data with a Modified Mo...
- C.2-Paper-DocEng-C75-SinghoferEtAl-Joint De-Identification of Textual and Relatio...

Corpus of Qualification Works (3)

- What works well = best practices
 - Focus on the topic ... try not to solve too many things
 - Take writing, especially the iterations, serious
 - Incorporate feedback on procedure and writing
 - Understanding the methods; reading the papers
 - Be good at coding and debugging!
 - Concentrated and persistent working habit
- What did not work? → negate the statements
- Writing way too little/bad, or also: way too long
- Note: Negative experimental results, but scientific rigor
= 1.0 grade, but not published :(

What Happens Post-Submission?

- Good submissions would be
 - Carefully checked by me and opt. PhD student
 - Prepared (=edited) to be published as preprint
 - Submitted to a conference
- Note: even with very good submissions, there is a substantial amount of work to be done
- If well prepared, i.e., student made several very good read-write iterations during the thesis, this is ~40hours
- Thus, a good estimate is **1 week of my time** on editing
- Is it worth the effort? Judge yourself → see papers

Summary: Corpus of Qual. Works

- **Lectures:** example reports, research proposal
- **Projects:** one resulted in two papers, another *set a research direction* for the group for the next 10 years, another was directly written into a PhD student's paper
- **Seminar:** how to well reflect and summarize literature
- **MSc theses:** track the evolution from (A) proposal, to (B) submission, to (C) arXiv + subsequent paper
- **BSc theses:** only came this year in Ulm, long though this is not possible; one thesis at A conference (after shepherding → [click here](#)) serves as *example for PhDs*
- Consent: all students gave (happily) their OK, when asked
- **MSc thesis does also prepare for a life as PhD student**

What else is in the Corpus?

- Apart from examples of student submissions
- **Writing template:** the corpus also contains a regularly updated template for writing a scientific paper in AI/ML
→ see: <https://tinyurl.com/dsbda-template> (latest)
- The official URL to this Overleaf template is: Template

DSBDA_TemplateForPaper_Annotated

[Open as Template](#) [View Source](#) [View PDF](#)

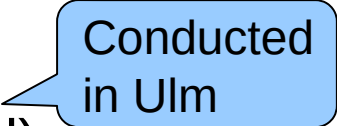

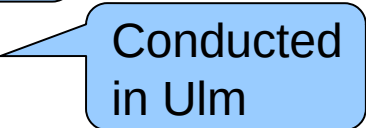
Author: Ansgar Scherp

DSBDA Template: The Name of the Title is Hope
<https://tinyurl.com/dsbda-template>

Ben Trovato* G.K.M. Tobin* trovato@corporation.com webmaster@marysville-ohio.com Institute for Clarity in Documentation Dublin, Ohio, USA	Lars Thorvöld The Thorvöld Group Hekla, Iceland larst@affiliation.org	Valerie Béranger Inria Paris-Rocquencourt Rocquencourt, France
Aparna Patel Rajiv Gandhi University Doimukh, Arunachal Pradesh, India	Huifen Chan Tsinghua University Haidian Qu, Beijing Shi, China	Charles Palmer Palmer Research Laboratories San Antonio, Texas, USA cpalmer@prl.com

- What's next? Make the template smarter :)
- **Questions?**
Ansgar Scherp, ansgar.scherp@uni-ulm.de

Success Stories

- Successfully applied in the past for publications
- Success stories are, e.g.,
 - MSc Thesis Fabian Singhofer [DocEng '21] (B ranked),
Best paper award!
 - Project STEREO [iiWAS' 21] (C ranked)
 - Project Text Summarization [iiWAS' 21] (C ranked)
 - MSc Thesis Ishwar Venugopal [IJCNN '21] (A ranked),
<https://arxiv.org/abs/2102.07838>
 - MSc Thesis Morten Jessen [DocEng '19] (B ranked),
Best student paper award!
 - MSc Thesis Florian Mai [JCDL '18] (A* ranked),
<https://arxiv.org/abs/1801.06717>
 - Project Quadflor: [KCAP '17] (A ranked)
 - MSc Thesis Gregor Große-Bölting [KCAP '15]:
Best paper nomination!