Research Seminar on Software Engineering (3 ECTS)

Dietmar Pfahl
dietmar.pfahl@ut.ee
Seminar Goal

• To help you deliver a high-quality master thesis on time
  • Target: Defence in June 2021

• Focus on **Goals** and **Methodology** (to achieve the goals)
• In the Spring seminar you already learned how to do a (superficial) literature review
Thesis Template

• Link: https://www.cs.ut.ee/en/studying/guidelines-regulations

• The typical structure of a thesis consists of the following components:
  • Title page
  • Information sheet
  • Table of Contents
  • Introduction
  • Terms and notions (optional)
  • Chapters
  • Summary
  • References
  • Appendices (if relevant)
  • License
Thesis Template

- The typical structure of a thesis consists of the following components:
  - Title page
  - Information sheet
  - Table of Contents
  - Introduction
  - Terms and notions (optional)
  - Chapters
  - Summary
  - References
  - Appendices (if relevant)
  - License

<table>
<thead>
<tr>
<th>Component</th>
<th>What to Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>Why do it? (Context / Motivation)</td>
</tr>
<tr>
<td></td>
<td>What to achieve? (Goals)</td>
</tr>
<tr>
<td>Related Literature</td>
<td>What exists already?</td>
</tr>
<tr>
<td>Methodology</td>
<td>How to achieve goals?</td>
</tr>
<tr>
<td>Results</td>
<td>What was achieved?</td>
</tr>
<tr>
<td>Discussion</td>
<td>How well were goals achieved?</td>
</tr>
<tr>
<td></td>
<td>What are the limitations?</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Summary of achievements</td>
</tr>
<tr>
<td></td>
<td>What next?</td>
</tr>
</tbody>
</table>
Thesis Template

• Link: https://www.cs.ut.ee/en/studying/guidelines-regulations

• The typical structure of a thesis consists of the following components:
  • Title page
  • Information sheet
  • Table of Contents
  • Introduction
  • Terms and notions (optional)
  • Chapters
  • Summary
  • References
  • Appendices (if relevant)
  • License

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Why do it? (Context / Motivation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related Literature</td>
<td>What exists already?</td>
</tr>
<tr>
<td>Methodology</td>
<td>How to achieve goals?</td>
</tr>
<tr>
<td>Results</td>
<td>What was achieved?</td>
</tr>
<tr>
<td>Discussion</td>
<td>How well were goals achieved?</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Summary of achievements</td>
</tr>
<tr>
<td></td>
<td>What are the limitations?</td>
</tr>
<tr>
<td></td>
<td>What next?</td>
</tr>
</tbody>
</table>
Types of Theses


• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author
Types of Theses

• See:  

• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author
Types of Theses – Theoretical


• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author

More relevant for theoretical informatics
In SE/IS: Modeling, Algorithms, ??
Types of Theses – Theoretical

• See:

• Theoretical Study
  • Independent Study
  • Review-type Study
  • (Systematic) Literature Review
  • Mapping Study (=light-weight SLR)
  • Survey Study (questionnaires/interviews)

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author
Types of Theses – Applied


• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author

Design Science / Engineering / Case Study / Action Research
Types of Theses – Applied


• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author

Evaluation / Quality Assessment (Test & Static Analysis) (and Process Assessment/Evaluation)
Types of Theses – Applied

• See: 

• Theoretical Study
  • Independent Study
  • Review-type Study

• Applied Research
  • Thesis based on a software solution created by the author
  • Hardware or software testing
  • Thesis based on study aid or learning material compiled by the author

  Didactics / Evaluation based on Student Feedback
Goal Setting & Baselining

• It is important to have one (or more) goal(s) set out at the beginning of the thesis
  • Question to ask yourself: What will you have achieved when you submit?

• There are many types of goals (see next slide)
  • Type of goal depends on type of thesis and problem statement
  • Goals might be formulated as Research Questions (RQs)

• In order to be able to decide whether you have achieved the goal(s), you need to know what the starting point is - Baseline
Types of RQs

- Exploratory Question
  - Existence Question
  - Description and Classification Question
  - Descriptive Comparative Question
  - Frequency and Distribution Question

- Base-Rate Question
  - Descriptive-Process Question
  - Relationship Question
  - Simple Causality Question
  - Causality-Comparative Question
  - Causality-Comparative Interaction Question

- Knowledge Question

- Research Question

- Design Question
Exploratory Questions

• **Existence questions** - Does X exist?
  – Example: *Do code smells that trigger refactoring exist?*

• **Description and classification questions** - What is X like? / What are its properties? / How can it be categorized? / How can we measure it? / What is its purpose? / What are its components? / How do the components relate to each other?
  – Example: *What are all the types code smells in Java code?*

• **Descriptive comparative questions** - How does X differ from Y?
  – Example: *How do Java code smells differ from Python code smells?*
Knowledge and Design Questions

• **Knowledge Questions**: focusing on the way the world is
  • Questions about the normal pattern of occurrence of a phenomenon (Base-rate Questions)
  • Questions about relationships between two different phenomena (Relationship Questions)
  • Questions about causality between two phenomena (Causality Questions)

• **Design Questions**: concerned with how to do things better
Knowledge Questions

• Base-rate:
  • **Frequency and Distribution Questions** -> How often does X occur? / What is an average amount of X?
  
  Example: *How many distinct code smell types do in Java code exist? How often do they occur?*

  • **Descriptive-Process Questions** -> How does X normally work? / What is the process by which X happens? / In what sequence do the events of X occur?

  Example: *How do software developers at company XYZ test their web-applications? (i.e., what processes and tools do they use?)*
Knowledge Questions (cont’d)

• Relationship:
  • **Relationship Questions** -> Are X and Y related? / Do occurrences of X correlate with occurrences of Y?

  Example 1: *Do project managers’ claims about the intensity of testing correlate with the actual intensity of testing?*

  Example 2: *Does the occurrence of certain types of failures correlate with specific use cases?*
Knowledge Questions (cont’d)

- **Causality:**
  - **Simple Causality Questions** -> Does X cause Y? / Does X prevent Y? / What causes Y? / What are all the factors that cause Y? / What effect does X have on Y?
    
    Example: *Does the use of UML diagrams improve the quality of the design?*
  
  - **Causality-Comparative Questions** -> Does X cause more Y than does Z? / Is X better at preventing Y than Z?
    
    Example: *Does the use of UML diagrams improve the quality of the design more than other graphical design notations?*
  
  - **Causality-Comparative Interaction Questions** -> Does X or Z cause more Y under one condition but not others?
    
    Example: *Does the use of UML diagrams improve the quality of the design more than other graphical design notations in large projects, but not otherwise?*
Design Questions

-> "What is an effective way to achieve X?" / What strategies help to achieve X?"

Examples:

What is an effective way for teams to capture requirements in order to improve communication with customers?

or

What is an effective way for developers to represent design knowledge in order to improve design quality?
Methods -> Literature

- **Systematic Literature Survey:**

- **Mapping Study:**

- **Design Science:**
Milestones

• 21-Sep-2020: MSc topic and supervisor confirmed
  • e-mail from supervisor to dietmar.pfahl@ut.ee required
• 24-Sep-2020: Last possibility to withdraw from the seminar (without going through the Dean's office)
• 28-Sep-2020: Classification of thesis topic
• 19-Oct-2020: 1st draft of report
• 23-Nov-2020: 2nd draft of report + presentation slides (for presenters on 25-Nov-2020)
• 30-Nov-2020: 2nd draft of report + presentation slides (for presenters on 02-Dec-2020)
• 07-Dec-2020: 2nd draft of report + presentation slides (for presenters on 09-Dec-2020)
• 14-Dec-2020: 2nd draft of report + presentation slides (for presenters on 16-Dec-2020)
• 23-Dec-2020: Final report submission
Schedule – mandatory sessions in red

- **09-Sep-2020**: Introduction Session - look at requirements for graduation theses at ICS
  - In this session, I give an overview of the seminar course: Slides

- **16-Sep-2020**: Consultation Session (please reserve individual time slot during first 15 min of session) - first 15 min will be used for general Q/A
  - In this session, I answer questions related to the the first submission (email with confirmed topic from supervisor) and other questions about the seminar.

- **23-Sep-2020**: Consultation Session (please reserve individual time slot during first 15 min of session) - first 15 min will be used for general Q/A
  - In this session, I answer questions related to the the second submission (classification of topic according to the guidelines for graduation theses) and other questions about the seminar and its next submissions.

- No formal sessions from 30-Sep-2020 to 18-Nov-2020: Please schedule individual consulting sessions when needed. Preferably, I try to schedule individual consultation sessions on Wednesdays after 16:15. If that does not work for you, we can try to find a different time-slot. Please contact me for an individual consultation at least 24 hours ahead of time.

- **25-Nov-2020**: Presentation Session 1
- **02-Dec-2020**: Presentation Session 2
- **09-Dec-2020**: Presentation Session 3
- **16-Dec-2020**: Presentation Session 4 (if needed)
Assessment and Grading

• This seminar is a "pass/fail" course.

• Assessment criteria of the seminar
  • Presence during mandatory sessions (if online session, then video must be switched on)
  • Punctual submission of information / documents
  • Quality of final report -> Grades 1-5 possible
  • Quality of presentation (incl. slides) -> Grades 1-5 possible
    • Quality characteristics: (1) Use of time; (2) Slides (text and visuals); (3) Content; (4) Presentation
  • Giving feedback to at least three other students during/after presentation sessions
    • Quality characteristics: (1) Use of time; (2) Slides (text and visuals); (3) Content; (4) Presentation
  • See scoring rubric on next slide
# Presentation Scoring Rubric

(Rate each category 0-5, with 5 being the highest possible score)

<table>
<thead>
<tr>
<th>Item being evaluated</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Points earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Point reductions if an item is missing or incomplete</strong></td>
</tr>
<tr>
<td>More than 50% time overrun or less than 50% of time used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No time overrun [1] &amp; no unused time [1] / content appropriately distributed over time [2] / no rushing [1]</td>
<td>5</td>
</tr>
<tr>
<td>Slides (text and visuals)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Point reductions if an item is missing or incomplete</strong></td>
</tr>
<tr>
<td>Letters too small to read and slides too full with text / no visuals or visuals are misleading and overly complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Large enough letters [1] / Slides not overfull with text [2] / visuals support content [2]</td>
<td>5</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Point reductions if an item is missing or incomplete</strong></td>
</tr>
<tr>
<td>Presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Point reductions if any item is missing or incomplete</strong></td>
</tr>
<tr>
<td>Presenter needs to read from slides / the flow of the presentation is lost several times / presenter cannot be heard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Presenter seems to understand the content and uses the slides as support (no reading from slides) [3] / Clearly understandable voice [2]</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>? /20</td>
</tr>
</tbody>
</table>
Assessment and Grading (cont’d)

Grading Scheme for Presentation & Final Report

• Grade 1 (very poor) should be used only in exceptional cases: if the thesis clearly violates the basic principles of an academic work (for example, in the case of plagiarism). Grade 1 in any of the criteria automatically means F (fail) as a final grade.

• Grade 2 (poor) is a negative grade which should be used if the thesis does not meet the minimal requirements established to the criterion concerned. Grade 2 in two criteria automatically means F (fail) as a final grade.

• Grade 3 (barely acceptable) is the lowest positive grade which should be used in the case when the work meets the minimal requirements of the criterion, but contains some substantial shortcomings.

• Grade 4 (good) means that the thesis is a good work without major shortcomings.

• Grade 5 (very good) requires that the work is almost flawless according to the given criterion and is outstanding in some aspect.
Assessment and Grading (cont’d)

How to **definitely** fail the course?

• Not being present in **all** mandatory sessions (without sufficient justification; sufficient justification is a certificate from a medical doctor saying that you cannot be at work).

• Missing at least 2 milestones by more than 24 hours - no further exceptions will be granted!

• Not submitting **all** requested information items and documents (topics, supervisors, reports, slides, feedback).

• Not giving a presentation

• Not giving feedback to at least three other students' presentations

• Having a grade below '3' in the final report or the presentation
Communication

• Use Slack

se-seminar2020.slack.com
Hints & Tips – Technical Writing

Importance of good writing

• It helps your readers to:
  • distinguish between more important and less important material
  • quickly find specific information
  • grasp the flow of argument, the assumptions (incl. previous/related work), and the context

• Remember: Often you don’t have control over the recipients of your document
Hints & Tips – Technical Writing

Style issues:
• No colloquial expressions or slang!
• Appropriate usage of terminology
• Consistent wording (one concept – one expression)
• Short sentences
• Normally: result-oriented (not process-oriented)
• Introduce abbreviations before first usage
• Neutral style instead of “we ...”, “our ...”, etc.
• After each headline there should be some text ...
Hints & Tips – Technical Writing

Figures and tables:

• Have a caption (incl. brief description and numbering)

• Are referenced from within the related text

Fig. 1. A sample GQM.
Hints & Tips – Technical Writing

References:

• All literature included in the list of references must be referenced (at least once)

References:


those that refer to the way software relates to its development or operational environment. Process attributes are discussed in Section 9. Remarks on the practical application of software measurement are in Section 10. Possible future developments are discussed in Section 11.

Good surveys of the state of the art and on-going research can be found in [1, 2].
How bibliography needs to be organized? References should appear as ...

References should appear as …

• **Book with one author:**
  • Author, A. A. (2005). Title of work. Location/City, State: Publisher.

• **Book with two authors:**

• **Book with more than two authors:**

• **Journal article:**

• **A publication in press:**
How bibliography needs to be organized?

References should appear as …

References should appear as …

• **Journal article:**

• **A publication in press:**

• **Report from a university:**

• **Published proceedings:**
How bibliography needs to be organized?

References should appear as …

References should appear as …

- **Unpublished doctoral dissertation or master’s thesis:**

- **A presented paper:**

- **Web site:**
In-Text Citations

• In-text citations should appear with author surname followed by publication year in parentheses
  (Brown, 2002)

• Citing several references in-text:
  In most organizations, data resources are considered to be a major resource (Brown, 2002; Krall & Johnson, 2005; Smith, 2001).
  Brown (2002) states that the value of data is recognized by most organizations
  “In most organizations, data resources are considered to be a major organization asset” (Smith, 2001, pp. 35-36) and must be carefully monitored by the senior management.
  Brown (2002) states that “the value of data is realized by most organizations” (p. 45).

• If you have organized the citations with number in brackets:
  In most organizations, data resources are considered to be a major resource [15, 30, 84].
Requirements for Thesis Draft

- The final draft (in total around 12-15 pages) should contain the following
  - Title page
  - Abstract -> results part and conclusions can be empty
  - Table of Contents
  - Introduction / motivation / problem statement (section Introduction) -> 1-2 pages
  - Literature review (section Related Work) -> preliminary summary of core literature -> 2-5 pages;
  - Baseline (section Background) -> if you have a design science thesis: must contain baseline -> 2-3 pages
  - Plan of the contribution (section Method) -> must be detailed enough to understand what exactly will be done to achieve the goals -> 4-7 pages
  - Intermediate results (section Results) -> can be empty
  - Discussion of results / limitations (section Discussion) -> can be empty
  - Summary and conclusions (section Conclusion) -> can be empty
  - List of references -> at least 5 references expected (must be well-formed)
Evaluation of Thesis Drafts

• The lecturer will grade the drafts using the evaluation criteria 1-3 available at

• Additional rules
  • Draft contains a copy/paste of more than 10 consecutive words from any source → -1 point per copied fragment
  • Reproducing/re-drawing 1 or 2 pictures or tables is tolerated, if you explicitly cite the origin of the picture/table in the caption, else -1 point
Tips for Writing a Thesis

How theses get written: some cool tips ...

• http://www.cs.toronto.edu/~sme/presentations/thesiswriting.pdf
Hints & Tips – Presentation

• What to present:
  • Problem Statement (Motivation)
  • Goals (RQs)
  • Baseline
    • Related work done by others or
    • Current state-of-practice (if thesis done in a company or building on top of existing materials)
  • Method
    • Steps how you plan to achieve the goals (construction / evaluation)
    • Already achieved results (if any)
  • Plan of completion (schedule until submission)
Hints & Tips – Presentation

• Evaluation criteria (up to 1 point per criterion):
  • **Slides**: Is the amount of text appropriate? Are figures and tables used appropriately where possible?
  • **Public speaking/oratory**: Is the posture appropriate? Is the voice level and intonation engaging? Is the gesturing appropriate? Is the rhythm engaging?
  • **Structure**: Is the structure of the presentation clear? Are the transitions between logical parts of the presentation clearly articulated?
  • **Content**: Does the presentation highlight the main points of the paper? Is the level of detail appropriate? Are examples used appropriately?
  • **Timeliness**: Does the presentation use the allocated time appropriately? Is the time limit respected?
Hints & Tips – Presentation

• How to present:
  • Use the time – but don’t exceed (12 min presentation + 3 min discussion)
  • Balance content (per slide and across whole presentation)
  • Speak clear (not too fast / not too low)
  • Let the slides support your talk – don’t use slides to write down what you say
  • Use visuals/diagrams/tables where appropriate
How **not** to do a presentation ...

• Don McMillan: Life After Death by PowerPoint
  • [http://www.youtube.com/watch?v=WGiePuNFXwY](http://www.youtube.com/watch?v=WGiePuNFXwY)
  • [http://www.youtube.com/watch?v=zDvm1PVtgWo](http://www.youtube.com/watch?v=zDvm1PVtgWo)
  • [http://www.youtube.com/watch?v=lpvgfmEU2Ck](http://www.youtube.com/watch?v=lpvgfmEU2Ck)