TDDD89 - Scientific Method

Lecture 1
Introduction

Christoph Kessler

2023

LINKÖPINGS UNIVERSITET

Part I: General Information

- Course format
- Activities
- Examination

Part II: Towards a great thesis

- What is a *good* thesis project?
- Overall thesis structure
- Thesis project types

Continued in first part of Lecture 2 (tomorrow afternoon)

Part I: General Information

Course format and staff

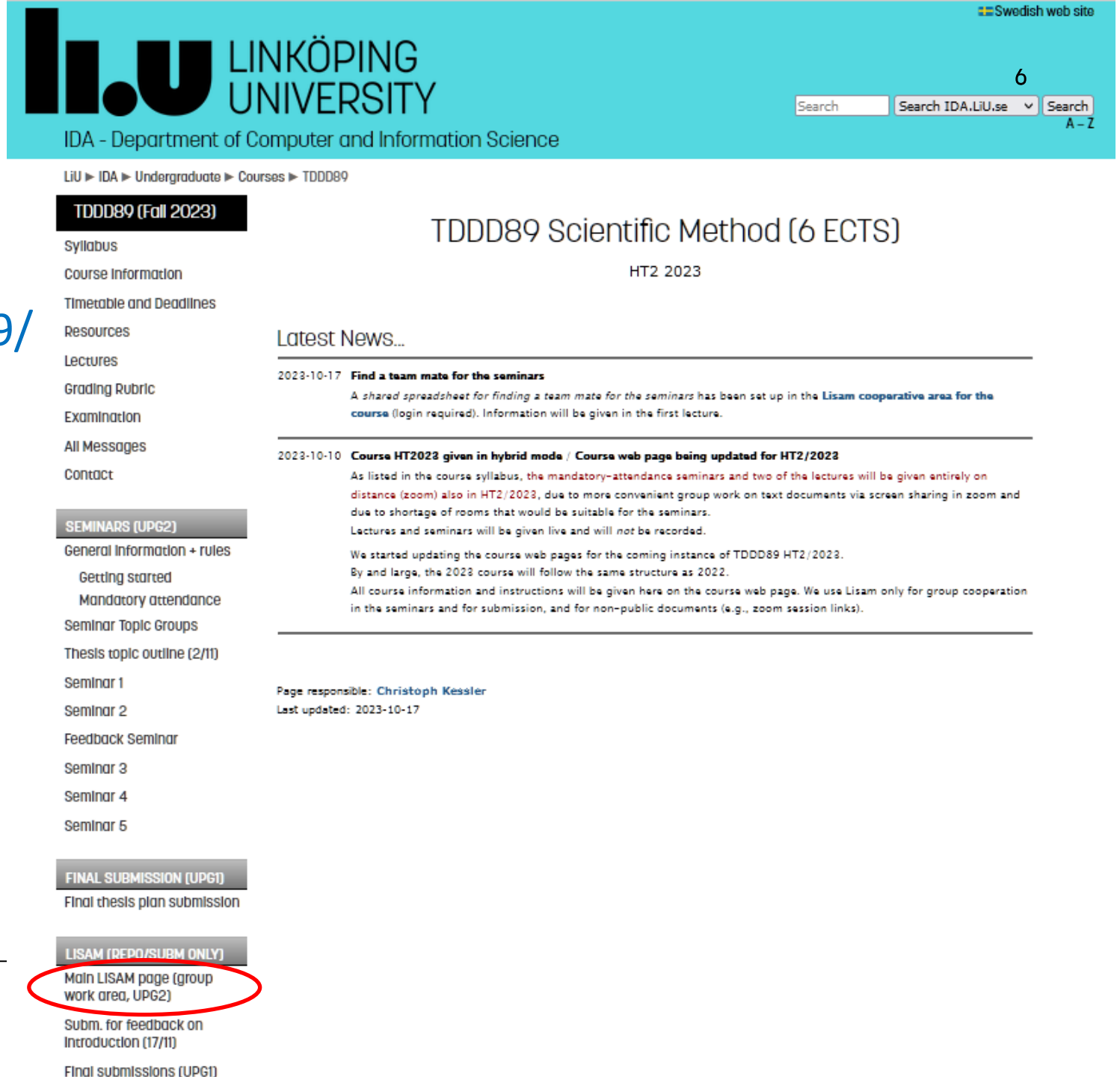
- 6 x 2h lectures
- 6 x 2h seminars
  - theme-specific groups
  - Group leaders:
    - Christoph Kessler (Group A)
    - Ali Hassan Sodhro (Group B)
    - Navya Sivaraman (Group C)
    - August Ernstsson (Group D)
    - John Tinnerholm (Group E)
    - Szilvia Varro-Gyapay (Group F)
    - Jose Antonio Hernandez-Lopez (Group G)
- include: 2 feedback sessions
  - Feedback seminar on research questions, by group leaders
  - Feedback lecture on Academic English and Stylistic Issues, with Shelley Torgnyson and C.K.

Lectures

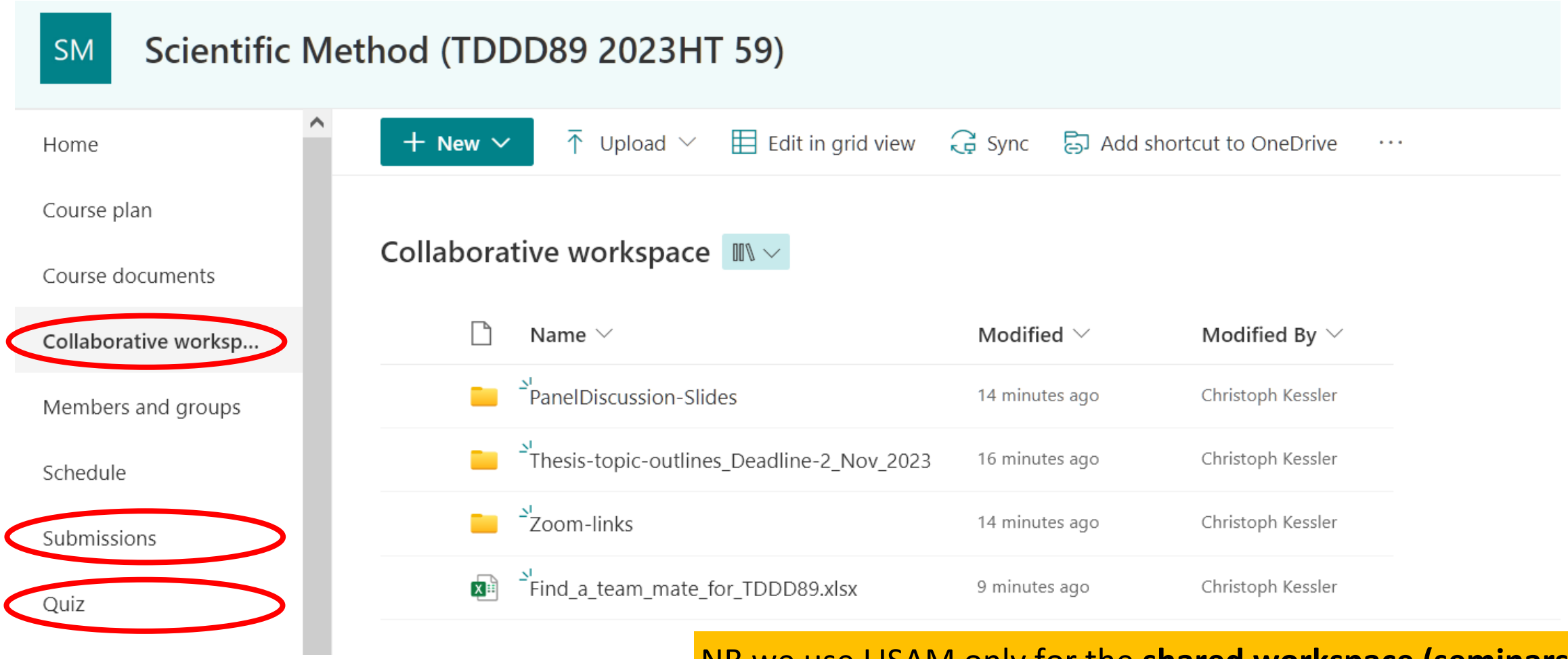
1. Introduction (Christoph Kessler)
2. Introduction (cont.): (Christoph Kessler and guest speakers)
   Common thesis types.
   Panel: Outlook to working and career paths in academic and industry R&D
   (and what the thesis topic selection may have to do with it…)
3. Literature search and evaluation (Christoph Kessler)
4. Introduction to academic writing in English (Shelley Torgnyson)
5. Scientific methods (Christoph Kessler)
6. Feedback on academic English (Shelley Torgnyson);
   Outlook to the master thesis process (Christoph Kessler)

Course web page

https://www.ida.liu.se/~TDDD89/



LISAM - Shared workspace and document submission



NB we use LISAM only for the **shared workspace (seminars)**, for sharing non-public information, and for **submissions**. All instructions and other information are provided on the course web page.

Resources

TDDD89 Scientific Method

Resources

- **Theses at IDA** (current page)
- **Degree projects (Master theses 30hp) at IDA (2023)** (new page on Liunet, under construction, requires LiU login)
- **Portal for thesis projects at IDA** (contains links to divisions/labs/research groups where thesis proposals are posted)
- **Theses at ISY** (also contains links to divisions where thesis proposals are posted)
- **New exjobb-project database at LiU** (still under construction, check the IDA and ISY exjobb web pages above for more options)
- **Planning report** structure for a Master's thesis (generic structure TekFak, minimum requirements. Your thesis examinator might have further requirements.)
- **Checklist for degree project at the second cycle (Master's) level**
- J. Åberg: **Introduction to scientific methods** (in Swedish)

- **ACM Code of Ethics and Professional Conduct**,
- **IEEE Code of Ethics**

Writing a thesis

- J. Åberg (2015). *Instructions for final thesis reports.* (**English**, **Swedish**).
- C. Kessler: **Stylistic advice to my students for writing a thesis**
- N. Ramsey: **Learning Technical Writing Using The Engineering Method**, Tufts University, 2016
- LiU Academic English Support (AES) (no web page available any more - see guest lecture by Shelley Torgnyson for references)
- IEEE Editorial Style Manual, official manual by the Institute of Electrical and Electronics Engineers, used by engineers in Computer Science and Electrical Engineering. (**pdf**).
- The **Academic Phrasebank** from Manchester University, UK, may be used to find synonyms and useful phrases in academic writing.
- **Advice on academic writing in English** from **Academic English Support** @ LiU
- In Swedish only: **Språkverkstäder vid Campus Valla och Campus US** har en hel del tips om både muntlig och skriftlig framställning på svenska och engelska. Språkverkstäderna är öppna för alla studenter vid Linköpings universitet som vill ha hjälp med muntlig eller skriftlig framställning på svenska och engelska.

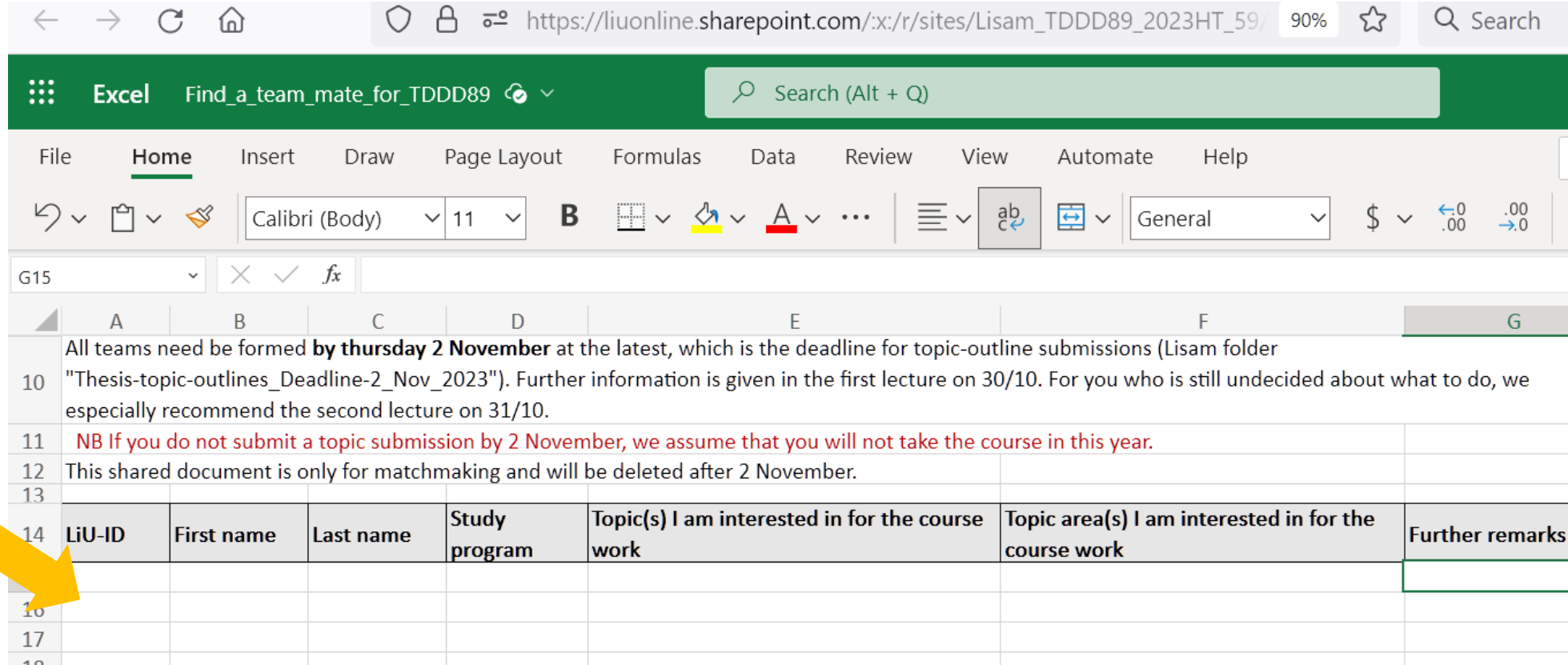
Reading papers

- S. Keshav (2007). **How to read a paper.** ACM SIGCOMM Computer Communication Review, 37(3), 83-84.
- T. Greenhalgh (1997). **How to read a paper. Statistics for the non-statistician. I: Different types of data need different statistical tests.** British Medical Journal, 315(7104), 364. (pdf)
- T. Greenhalgh (1997). **How to read a paper. Statistics for the non-statistician. II:" Significant" relations and their**

Working in pairs – Find a partner

Seminar work such as extended thesis plan writing is done in pairs.
Exception for those who already have their “real” final project and know they will do it alone.

No partner for the course yet?
Use (today!) this matchmaking spreadsheet linked from the course web page:



Selecting some thesis topic to work with in this course

https://www.ida.liu.se/edu/ugrad/thesis/katalog.shtml

LINKÖPING UNIVERSITY

IDA - Department of Computer and Information Science

LiU ▶ IDA ▶ Education ▶ Undergraduate ▶ Thesis ▶ Lediga examensarbeten / Thesis opp

Final thesis at IDA

Thesis opportunities

Previous theses

WEXUPP

New application at the technical faculty (for students)

WExUpp for teachers

THESIS PRESENTATION

Upcoming presentations

Opponent available

Opponent wanted

REPORT

Templates

Instructions for final report

Report publication

PRIZES

Available prizes and awards

OTHER

Degree-project staff

Fir

Lediga exa

De flesta exjobbsförslag publiceras nu via L

Interna arbeten (Internal thesis

ADIT:

- examensarbeten/final theses hos/at Security and Networks Group
- examensarbeten/final theses hos/at Database and Web Information Systems Group
- examensarbeten/final theses hos/at Air and Ground Information Security Group
- examensarbeten/final theses hos/at Sports Analytics Group
- examensarbeten/final theses hos/at Communications for Networked Intelligent Systems Group

AIICS:

- Knowledge Processing Lab
- Natural Language Processing Group
- Representation, Learning and Planning Lab
- Theoretical Computer Science Lab

HCS:

- examensarbeten/final theses hos/at COI
- examensarbeten/final theses hos/at MDA
- examensarbeten/final theses hos/at NLPL
- examensarbeten/final theses hos/at IxS

SaS:

- examensarbeten/final theses hos/at ESLAB
- examensarbeten/final theses hos/at PELAB

Also: New LiU-central database for thesis projects, started recently: https://exjobb.liu.se

- Master Thesis information
- Student Labs

PhD studies

Division

Automatic Control

Communication Systems

When these web pages are removed, you can still find opportunities for internal and some external thesis projects on the web pages of the **different researchers and research groups at IDA and ISY**, e.g. https://www.ida.liu.se/~pelab

.liu.se/en/edu/xjobb/

NKÖPING
NIVERSITY

Engineering

raduate level » Master Thesis information

Thesis at ISY

Instructions

- Instructions for the student
- Checklist for the student
- Instructions for the opponent
- Attending presentations

When you have an idea and an outline for a thesis topic or have found an interesting prop then contact a suitable examiner within the area or one of the Directors of Studies or th thesis coordinator to discuss the topic(s).

Get inspired by the list of best theses nominated by our department:

- Awards and nominations given to previous theses at ISY

Proposals for master theses projects can be found in the theses database.

You can find more information on master theses at our divisions:

- Automatic Control
- Communication Systems
- Computer Engineering
- Computer Vision Laboratory
- Integrated Circuits and Systems
- Information Coding
- Vehicular Systems

Sample theses – define the group topic areas

IDA - Department of Computer and Information Science

LiU ▶ IDA ▶ Undergraduate ▶ Courses ▶ TDDD89 ▶ Seminars ▶ Seminar Topic Groups and Sample Master Theses

TDDD89 Scientific Method

Seminar Topic Groups and Sample Master Theses

Seminar Topic/Method Groups and Sample Master Theses

The following selected Masters' theses pertain to the different groups' topics and/or research method types. Each thesis has a number of keywords describing it by topic area and method type, and seminar groups are formed based on the similarity of the thesis topic outlines submitted by students in the first week.

- **Topic 1:** Software development processes, software quality - Case study:
  "Vertically Scaling Agile: A Multiple-Case Study" by Nicklas Östman and Rasmus Lindström, Linköping University 2017.
- **Topic 2:** Programming framework, parallel systems, compilers, metaprogramming, performance - Design, Prototype implementation:
  "SkePU 2: Language Embedding and Compiler Support for Flexible and Type-Safe Skeleton Programming", by August Ernstsson, Linköping University 2016.
- **Topic 3:** Business software, services, software contracts - Usability study, iterative development:
  "Usability of a Business Software Solution for Financial Follow-up Information of Service Contracts" by Therese Borg, Link University 2018.
- **Topic 4:** Machine learning, data mining, image processing - Experimentation:
  "3D reconstruction from satellite imagery using deep learning" by Tim Yngesjö, Linköping University 2021.
- **Topic 5:** Security - Evaluation:
  "Certificate Transparency in Theory and Practice" by Josef Gustafsson, Linköping University 2016.
- **Topic 6:** Algorithms, scheduling, embedded realtime systems, metaheuristics - Improvement, optimization, analysis:
  "Performance Optimizing Priority Assignment in Embedded Soft Real-time Applications" by Fredrik Bergstrand and Tobias Edqvist, Linköping University 2018.

Topic groups and seminar groups will be assigned based on the submitted thesis topic outlines.

Absolutely no idea for a topic? Take one of these – as a last resort, so you have something to work with during the course.

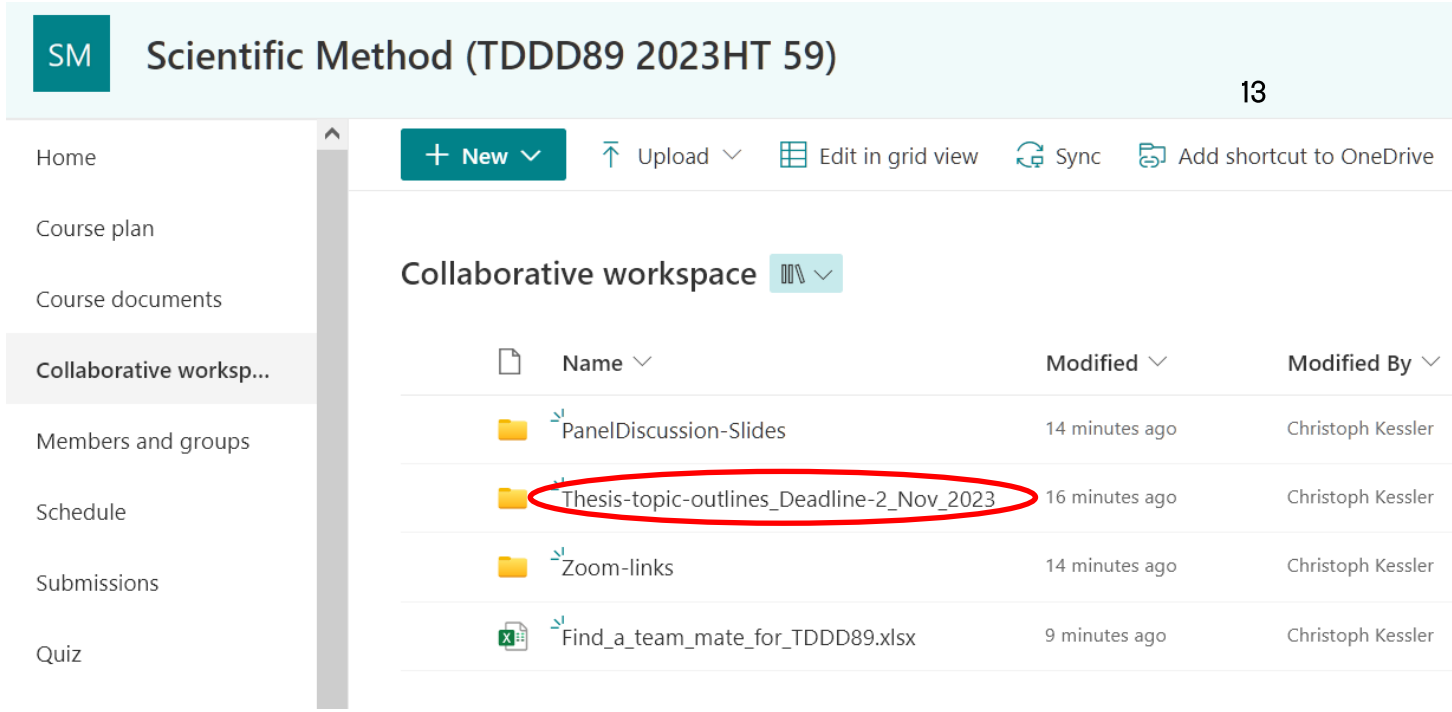
Choosing a closest thesis topic area / method type

from the list of sample master's theses

1. **Software development processes, software quality / Case study**
2. **Programming framework, parallel systems, performance / Design, prototype implementation**
3. **Business software, services, software contracts / Usability study, iterative development**
4. **Machine learning, data mining, image processing / Experimentation**
5. **Security / Evaluation**
6. **Algorithms, scheduling, embedded / realtime systems, metaheuristics / Improvement, optimization, analysis**

https://www.ida.liu.se/~TDDD89/seminars/topicgroups.shtml

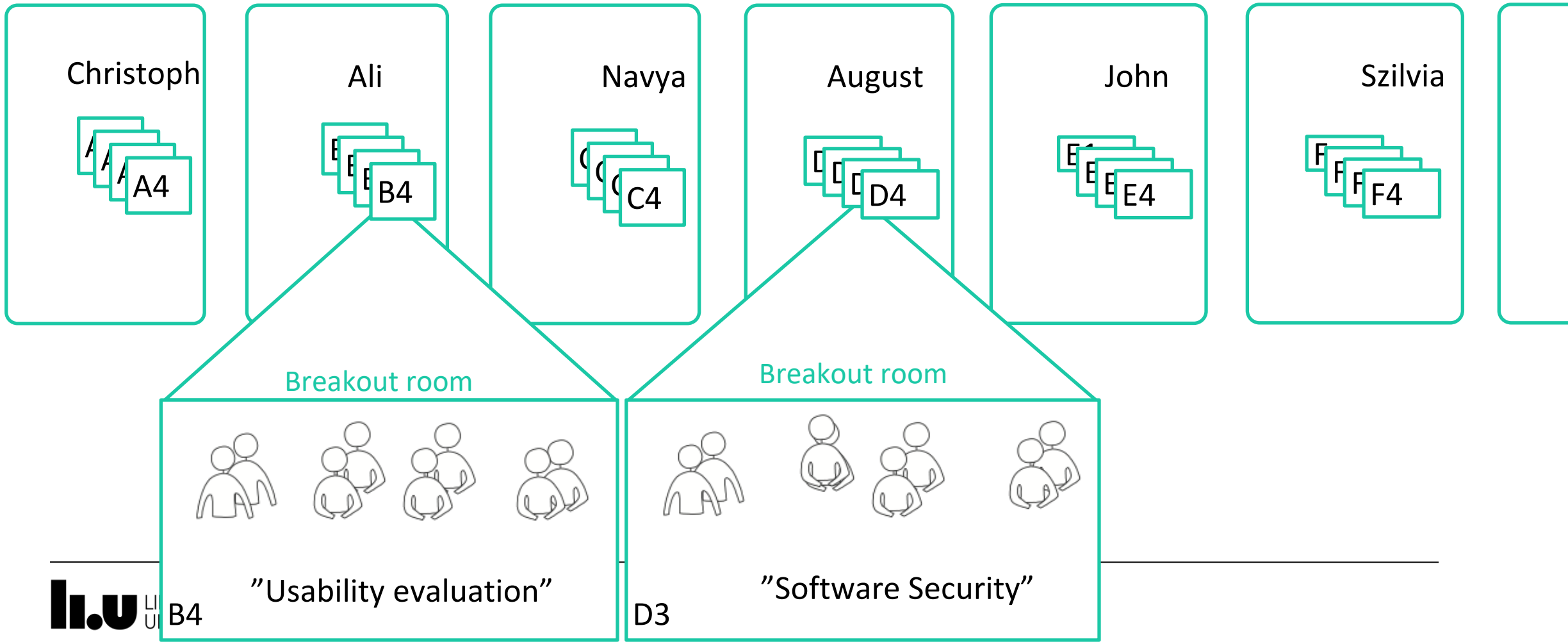
Getting started: An early deadline



- **By Thursday (!)**:
  - Find a partner to work with
  - Find a topic to work with
  - Find the closest topic area (1-6)
  - Find the closest method type (1-6)
  - Submit a ***thesis topic outline*** (**max. 1 page**)
    - File (.txt, .pdf, .docx), e.g. “liuid001_liuid002_topic.txt”
    - Put it in LISAM collaborative workspace folder “Thesisplan_topic_outlines_by_3_Nov”
- By next monday, you will be divided into seminar groups based on your selected topics / research method types.
  - We will sign you up in a webreg group (A-G, see schedule) for UPG2.

Up to 28 Seminar topic groups: {A,…,G} x {1,…,4}

(mapping to be announced early on Monday 6/11)



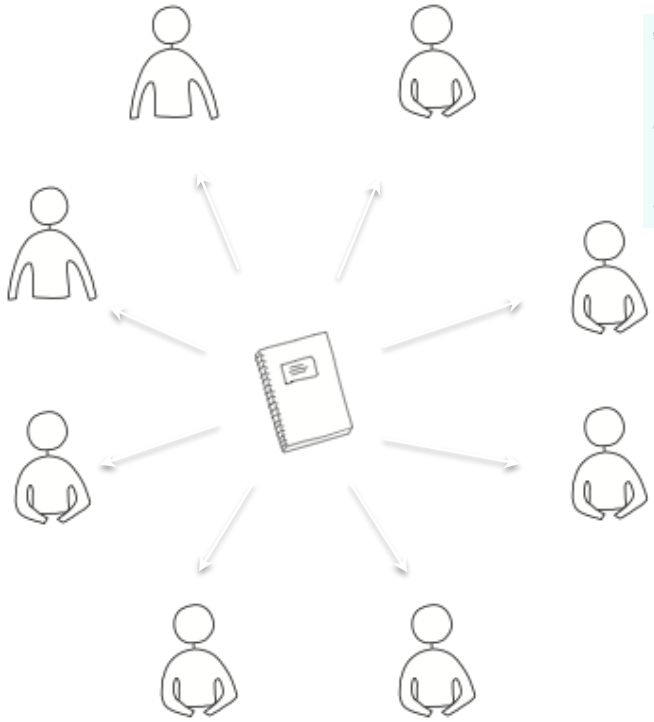
Seminars

Seminar	Read in sample thesis	Write extended thesis plan
1	Introduction, Background+Related work / Theory	
2		Introduction incl. research questions → ca. 2 pages + references
FB	*Feedback session seminar on research questions*	
3		Introduction incl. res. questions, Background, Related work → ca. 5 pages + references
4	Method, Results, Discussion, Conclusion	
5		Introduction incl. res. questions, Background, Related work, project plan → ca. 8 pages + references

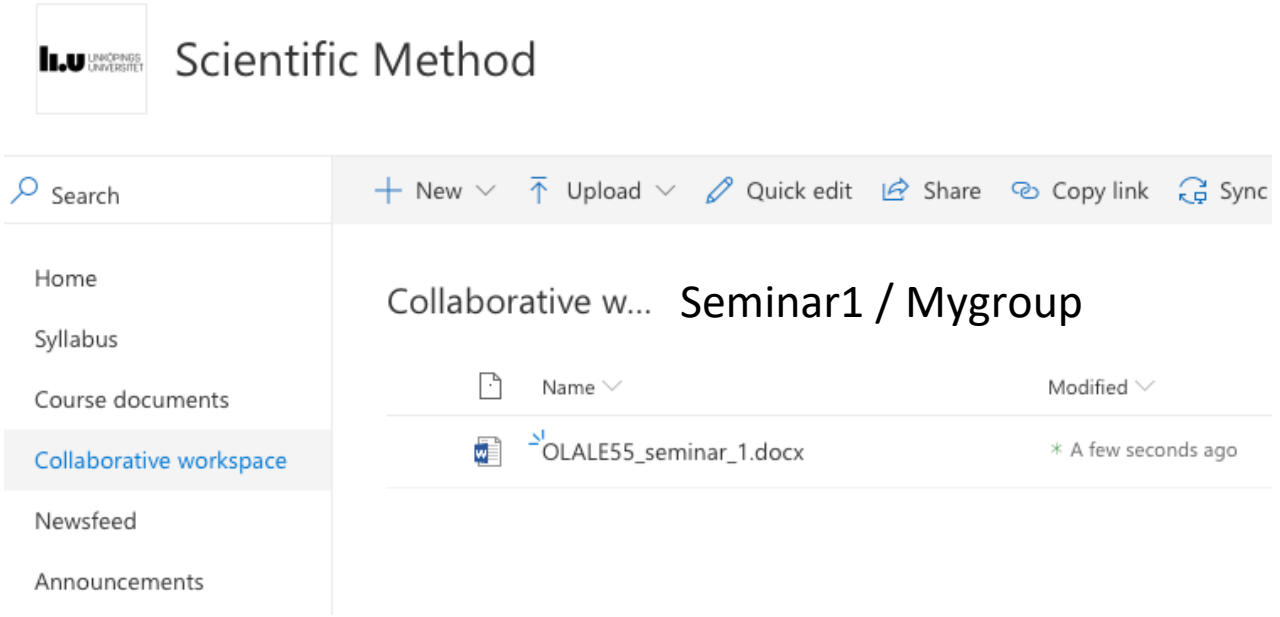
Submission of introductions 17/11 for feedback
(1) from Shelley and Brittany on academic English
(2) from your group supervisor on the research questions

then revise for UPG1 submission

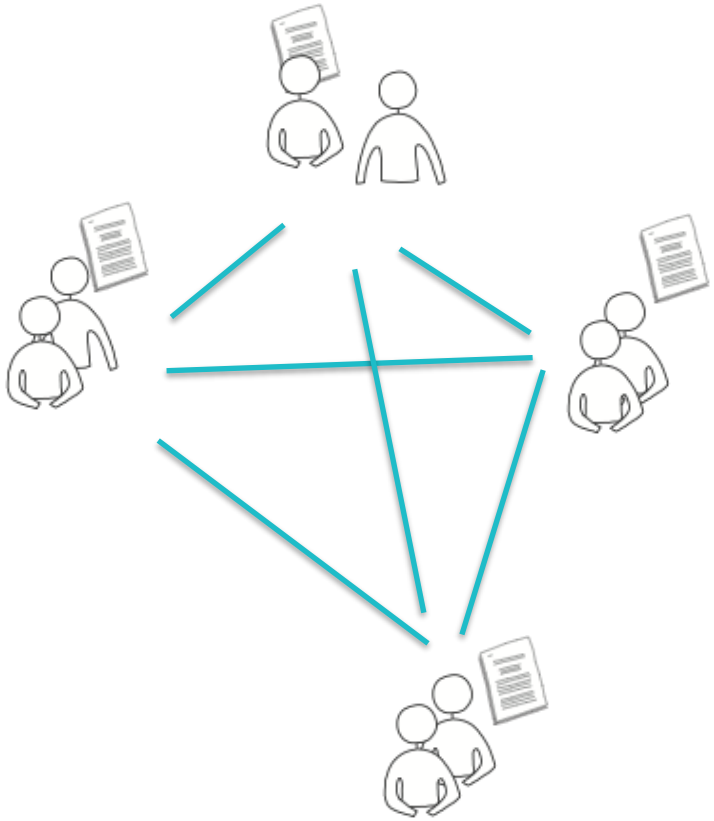
Seminars 1 and 4: Reading the sample thesis



”Are the research questions in the published thesis easy to find, clear and with a reasonable scope, as required by the *instructions for final thesis reports* ?” …

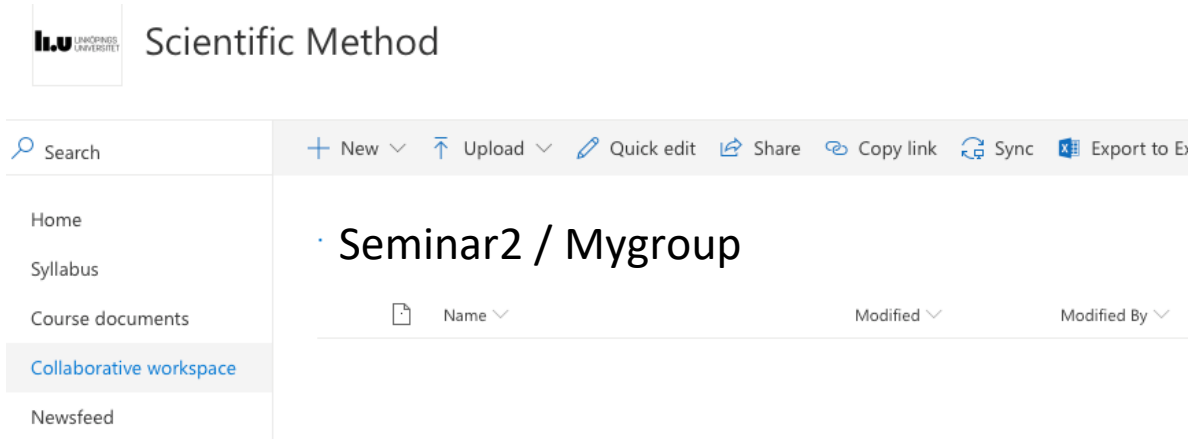


Seminars 2, 3, 5: Writing an extended thesis plan + feedback



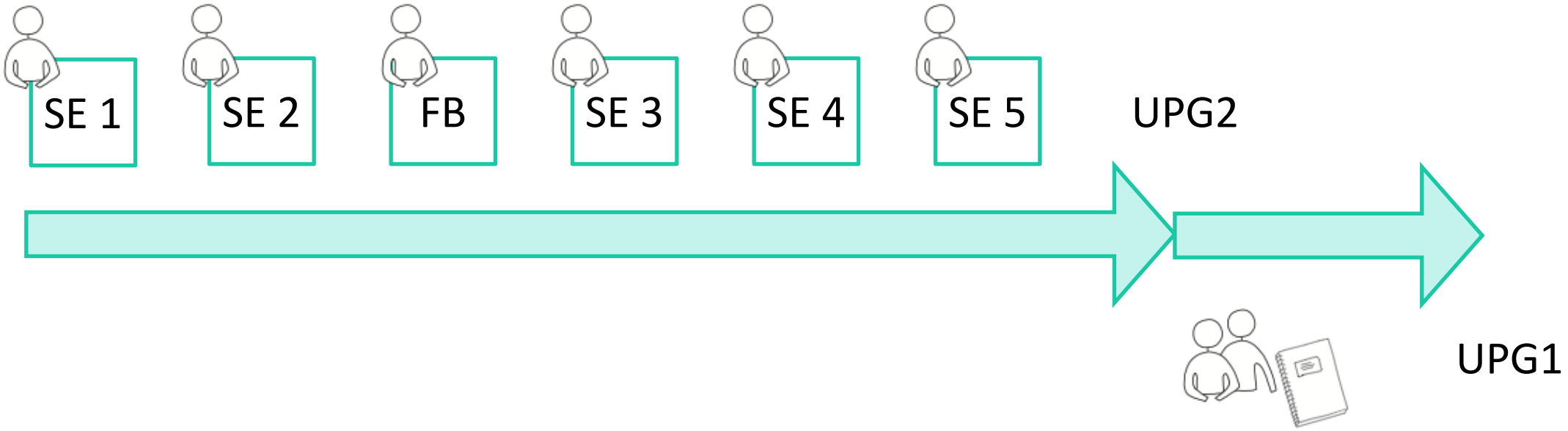
1

2

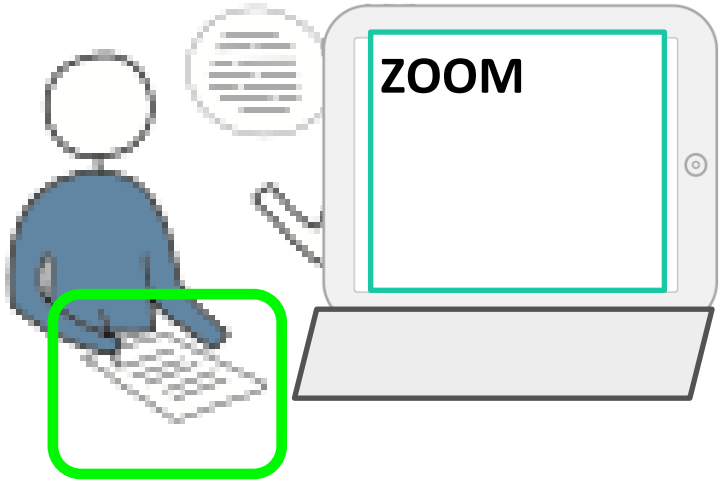


Examination

- **UPG1**: *Extended thesis plan*, to be finished at the end of the course
- **UPG2**: Preparation and participation in seminars during the course



Seminars



Avoid distractions.

Have your solution ready in Lisam.
Use screensharing.

Take notes!

Webcam on during seminar group work

Recording is not permitted

Final submissions

End of HT2

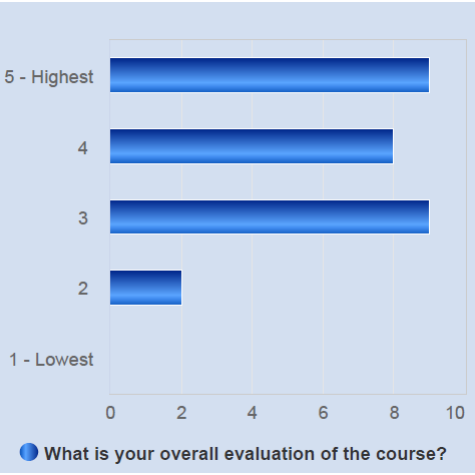
January 11, 2024

Workload distribution

160h total:
Plan your time,
look ahead and
read the course description document thoroughly

Seminar	Work %
1	20
2	15
3	20
4	20
5	15
Final submission	10

Recent changes based on course evaluation feedback



2022: 3.86

- Widely appreciated and kept: the **hybrid** format (see syllabus)
  - *Seminars* in zoom breakout rooms
    - Was perceived by many as BETTER than physical seminars (screensharing is convenient, less background noise, no need to move tables around)
      - provided that one can *see* each other during discussions and microphones are working properly.
  - Some *lectures* remain virtual, too
    - helps to reduce schedule conflicts as LiU does not have enough large lecture halls
- The *panel discussion* was appreciated and kept in its early position (Lecture 2) to possibly help with the course topic selection before the deadline 2/11
- Compensation assignment for missed seminars and late preparation hand-in has been updated/clarified
- Reading *Method* chapter of the sample thesis moved since 2021 from Seminar 1 to Seminar 4
  - to reduce time pressure before Seminar 1 deadline
- Seminars moved from Tuesdays to Thursdays / submission deadlines moved from Sundays to Tuesdays
  - Less stressful, kept also for this year
- We admit singleton groups if you already know your *real* thesis topic and that you will do it alone.

Part 2: Getting Started

Towards a Great Thesis

What is a great thesis?

Thesis = project results + written presentation

Example:

- *A working, interesting application with proven and general value*
  - A well-described *application*
  - of general *interest*
  - and with a clear description of ”*proven*” and ”*value*”

What is a great thesis?

Thesis = project results + written presentation

- Includes an **evaluation** with
  - **general** and
  - **interesting** results
  - that others can **use**
  - that others will **believe**

What is a great thesis?

Thesis = project results + written presentation

- An authoritative report
  with a good **focus** (questions!)
  and **results** that answer the questions
  through a transparent, thorough **description of the process**

Getting started: Elaborate the requirements → find the research problem / research questions

Precise requirement – finding *relevance* more important

Why?

”Evaluate algorithms to be used for image clustering”

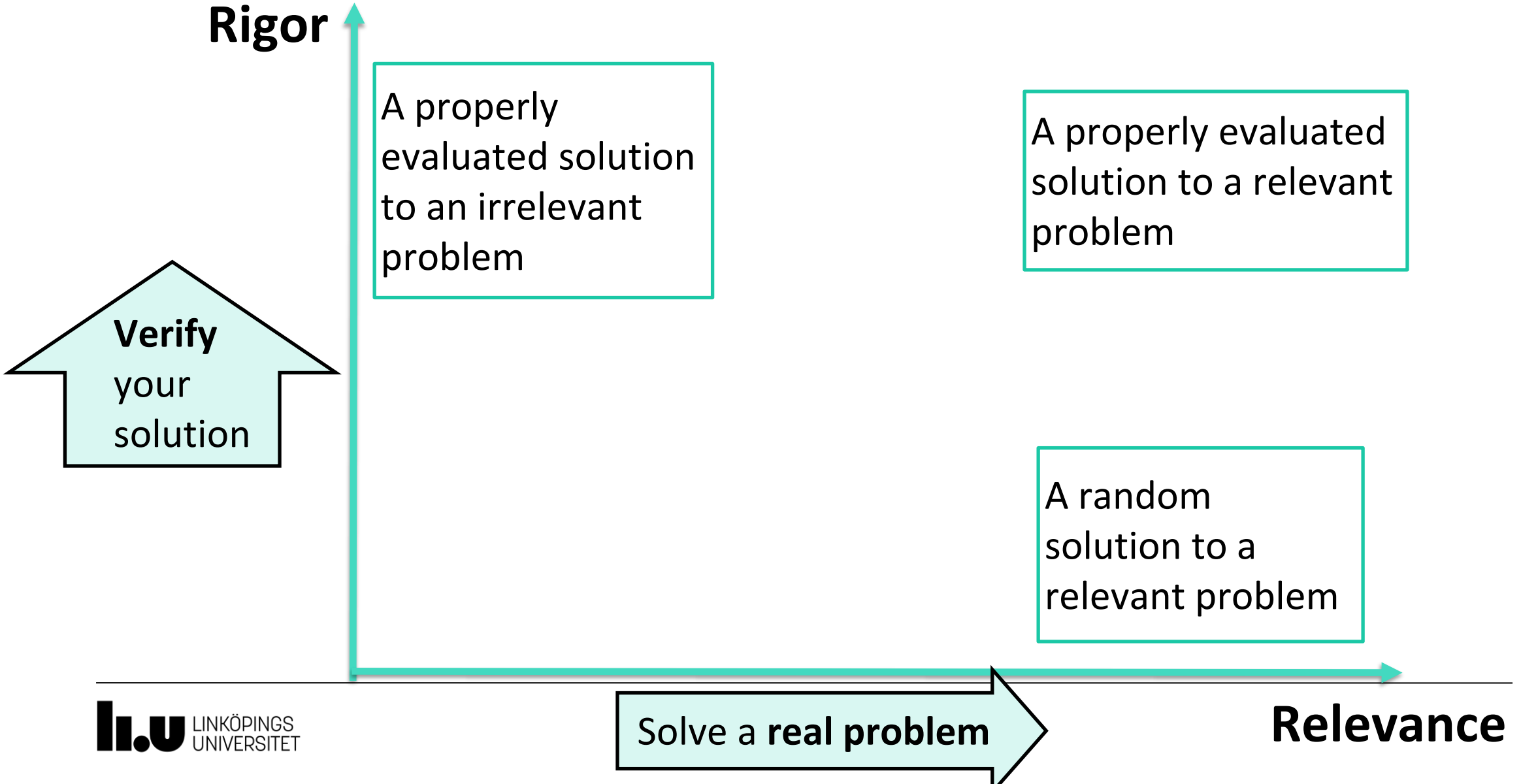
”Determine whether clustering algorithms can be used to detect activities in sets of images”

Vague requirements – finding *focus* and *rigor* more important than *relevance*

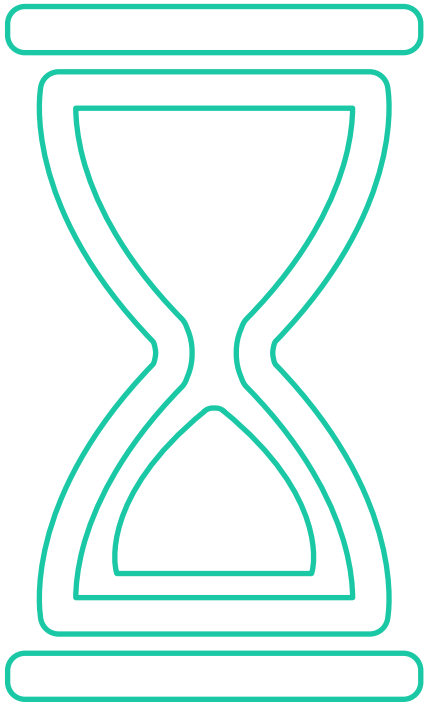
How?

”Find activities in sets of images”

Relevance vs. Rigor



Thesis outline



Hourglass model for technical reports:
Usually, most specific in the middle
(details, technicalities)

Why should even I read this thesis?

What have you studied here?

What does this relate to?

Can I trust you?

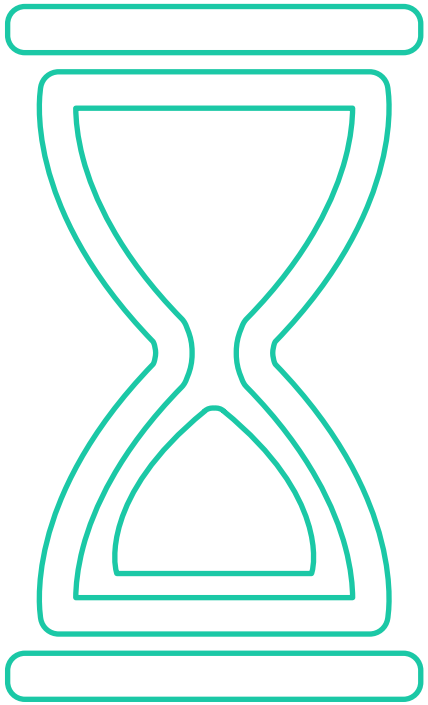
What is built?

What have you found?

How can we explain the results?

How can I use these results in my work?

Thesis outline



Hourglass model for technical reports:
Usually, most specific in the middle
(details, technicalities)

Abstract

Research Questions

Background, Related Work / Theory

Method

Implementation

Results

Discussion

Conclusion

Thesis outline

Remark: This very generic thesis structure, which we use in this course, is not necessarily the best choice for the *chapter* structure and -titles of your thesis.
For your final thesis, you should find a more specific chapter structure and titles, but make sure to duly cover all these *aspects* somewhere.

- Abstract
- **Research Questions**
- Background, Related Work / Theory
- Method
- Implementation
- Results
- Discussion
- Conclusion

What is a good research question in Computer Science?

Question type	Example question	Type of answer

Different research questions in Software Engineering, adapted from: Mary Shaw: Writing Good Software Engineering Research Papers: Minitutorial. Proc. 25th International Conference on Software Engineering, ICSE '03, pages 726-736, 2003. IEEE Computer Society.

What is a good research question in Computer Science?

Question type	Example question	Type of answer
Means of development	What is the most efficient software testing method for a small team developing a mobile application?	Procedure
General method for analysis	How can one verify conformance to real-time constraints in a multi-threaded embedded system?	Analysis method
Specific evaluation of systems	When is PhoneGap more economical to use than NativeGoo for cross-platform mobile development?	Empirical predictive model based on data
Generalization or characterisation	Given recent results in tuning deep neural networks, which meta-heuristics should be used for exploring the parameter space?	Classification, taxonomy
Feasibility study	Can one automate a car?	Specific implementation, empirical observations

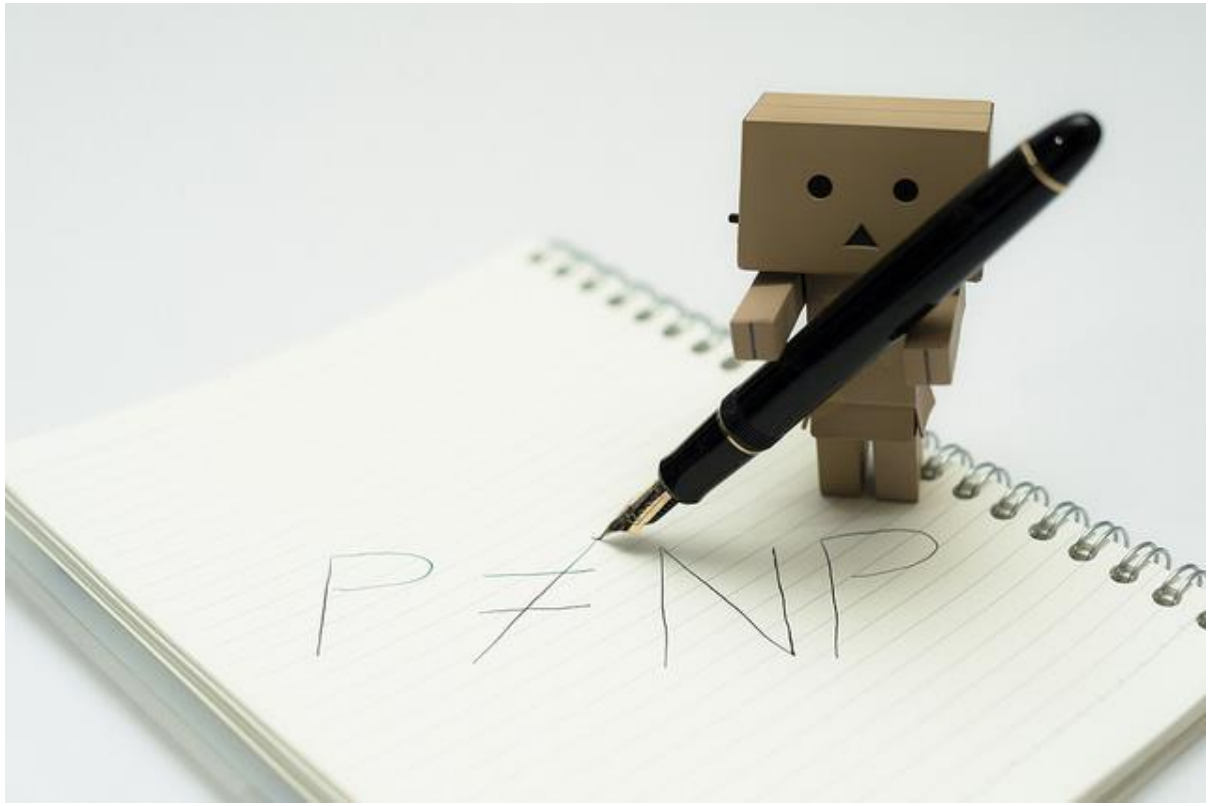
Different research questions in Software Engineering, adapted from: Mary Shaw: Writing Good Software Engineering Research Papers: Minitutorial. Proc. 25th International Conference on Software Engineering, ICSE '03, pages 726-736, 2003. IEEE Computer Society.

Towards A Great Thesis

I - The Problem

Start from a problem that people may be interested in (at least a few more than those who gave you your task)

→ Motivation and Aims
→ Formulate explicit research questions



Evaluating a research question

- ”How can one construct a web application?”
- ”Is it possible to construct a mobile application for functionality X?”
- ”How can one create a usable website?”

Is the answer any of the following?

By writing what has already been written many times before

Yes, there is no reason to believe it could not be done.

By adhering to published design guideline Y

Then come up with a better question…

Arriving at interesting research questions

- Start exploring a field:
  - ”How can fuzz testing be used to find bugs in concurrent embedded software?”
- Then, use existing technical approaches and theoretical models:
  - ”Combining with dynamic and static analysis of schedulability of embedded real-time systems”
  - ”What is the efficiency of AFL at finding timing-related errors in concurrent software?”

The journey to a better question

Question	Approach	Objection
How can we automate testing?	Applying automatic generation of test cases	We end up with 10^7 test cases, only some of which are necessary...

The journey to a better question

Question	Approach	Objection
How can we automate testing?	Applying automatic generation of test cases	We end up with 10^7 test cases, only some of which are necessary...
How can we select *relevant* test cases?	Applying statistical/ML clustering techniques	Black-box solution with no known accuracy, we need traceability
How can we *automatically prioritize* test cases?	Optimize based on historical records to maximize *a*verage *p*ercentage *f*aults *d*etected (APFD)	No weight given to critical functionality
What do we really want to *optimize*?	Listen, observe, collect data	Requires an open mindset

2 - the Theory

“Theory”:
Rooting your work in your subject area’s scientific body of knowledge

→ create *trust* in your work

- **Background**
- **Formalization/Modeling**
- **Related Work**

“In God We Trust – All Others Must Bring Data!”

--- W. Edwards Deming

Convince the reader!

- In many cases, you will produce some code as part of your thesis project. This is not enough for your thesis work, though.
  - Is this a *relevant* problem?
  - Is this work based on the *state-of-the-art techniques* in the field?
  - Does it *improve* or *generalize* over previous work? And how?
  - Are all claims made *proven* in the thesis, or based on *trustworthy sources*?
- You need to convince your (critical) audience that you have done a great job!

  They do not know you …

  … nor your supervisor

  … nor your company

  so better be convincing!
- By anchoring your work in the scientific body of knowledge of your field

Theory

Compare these two claims:

”There are seven dimensions of usability”

”NP-hard problems are at least as hard as the hardest NP problems”

Theory: Characterization of knowledge, accumulated through scientific studies, published in peer-reviewed journals and conferences

Criteria:

- Validation
- Does it describe the world?
- Is it proven?

Modeling / Formalization

- **Isolate** and **abstract** the core problem
- **Models = abstractions**
  - Distill and formally describe the main relevant properties of a complex real system
  - Example: Models of computation, CPU performance / energy models
  - Formalization: Key parameters, set theory, equations, constraints, graphs, abstract data types, formal logic reasoning, …
- Empirical observations are based on expectations, informed by theoretical frameworks (models):
  - When reading **power consumption** values of a modern CPU, we expect that it depends on the CPU's different **power states** (e.g., voltage/frequency level)
- Based on observations, you can **test claims** made by your theory

→ **Generalizability** of your result beyond the concrete problem instance

“There is Nothing so Practical as a Good Theory”

--- Kurt Lewin, 1890-1947, social psychologist

How to convince the reader (aka. a Method)

In industry as well as in academia,
well-presented working solutions
based on widely trusted state-of-the-art techniques
are the best means of convincing the reader.

In the scientific community, we also require
claims about solutions to be **sound**,
so others can rely on them in their work.

Type of Method used to obtain and validate results	**Example scenario**

Type of Method used to obtain and validate results	**Example scenario**
Analysis	I have conducted a formal analysis of my algorithm, and have proved that it solves the TSP in $O(N^2)$, thus proving that P=NP
Evaluation	I have compared code review results with interviews and surveys, and found that if you test software, you are more likely to find faults than if you do not
Experience	I report on experience with aligning software architecture with code based on 20 years of software development for Swedish Social Security and have found that software architectures are essentially useless
Example/Prototype	This tool can recognize location and pose, so when you enter a bathroom it will give you valuable advertisements on Facebook while sitting down.

Method

For theses that involve creating a product / prototype, you need to **assess the external or internal qualities** of what you produce (e.g. usability, correctness, or scalability), using **qualitative** or **quantitative methods**.

Other types of theses that do not involve creating products/prototypes (e.g., systematic literature reviews) have their own established criteria for assessment.

You want to ensure that what you are studying does represent reality.
This is called ***external validity***.

Reproducible Results

All details about your **assessment method** need be carefully documented in the thesis

- e.g., test data sets, benchmarks, code, system, experimental setup

Allows others to follow the same setup to obtain the same data

→ **Reproducibility** increases trust in your work.

Algorithms	Implementations	Number of runs	Hyper-parameters
nearest centroid classifier	NearestCentroid	3	metric∈{"l1";"l2";"cosine"}
naive Bayes classifier (multinomial distribution)	MultinomialNB	22	alpha∈np.linspace(0,1,11) fit_prior∈{True;False}
Linear SVM	LinearSVC	20	C∈np.logspace(-4, 4, 10) loss∈{"squared_hinge";"hinge"} class_weight="balanced"
Logistic Regression	LogisticRegression	40	dual=False C∈np.logspace(-4, 4, 10) class_weight="balanced" solver∈{"newton-cg";"sag";"lbfgs"} multi_class="multinomial" dual=True C∈np.logspace(-4, 4, 10) class_weight="balanced" solver="liblinear" multi_class="ovr"
Perceptron	Perceptron	13	penalty∈{"l2";"elasticnet"} alpha∈10.0**-np.arange(1,7) class_weight = "balanced" penalty=None class_weight="balanced"
Stochastic gradient descent	SGDClassifier	120	loss∈{"hinge";"log";"modified_huber"; "squared_hinge";"perceptron"} penalty∈{"l2";"elasticnet"} alpha∈10.0**-np.arange(1,7) class_weight="balanced" average∈{True;False}

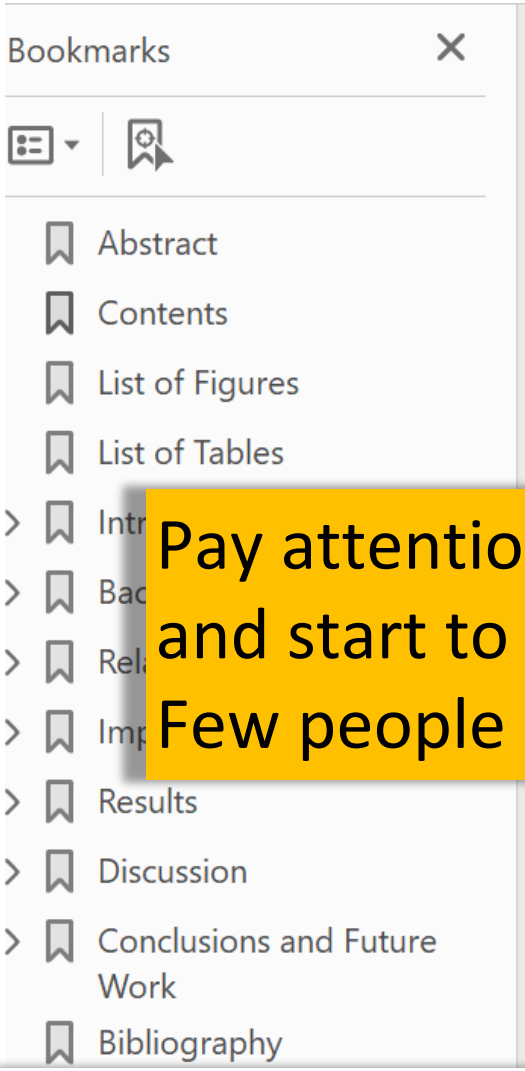
Table 3.8: The different configurations of the experiment 4

LINKÖPINGS UNIVERSITET

More about Research Methods in Lecture 4 ...

Thesis structure and format

For theses in mathematics, computer science and engineering I recommend using LaTeX.



Pay attention to the structure, formatting and typesetting of your thesis, and start to write early.
Few people have accurate estimates of how long it takes to write a thesis.

[45] B. Le Gal and C. Jego. "High-Throughput LDPC Decoder on Low-Power Embedded Processors". In: *IEEE Communications Letters* 19.11 (Nov. 2015), pp. 1861–1864. ISSN: 1089-7798. DOI: 10.1109/LCOMM.2015.2477081.

Summary so far:

A great thesis:

- An interesting problem
- A convincing theory
- A reliable method
- A working solution
- Established effects
- Great presentation

To be continued in the next lecture …

- **Engineering vs. Science?**
- **Common thesis types**

- Also:
  Panel discussion:
  Outlook to a professional career in industrial or academic R&D
  (and what the thesis topic selection may have to do with it…)
  - Important for answering preparatory questions before Seminar 1

- On zoom, tomorrow Tuesday 15:15 – zoom link see Lisam cooperative area

Acknowledgements

- Some slides are based on a previous version courtesy of Ola Leifler, IDA, Linköping University