Hints for writing a PhD thesis
– A pattern oriented approach –

Daniel Ratiu, Stefan Wagner, Martin Leucker
Scope of this Presentation

➢ Syntax vs. Semantics
  ➢ How to write the thesis (syntax)
  ➢ NOT how to produce the content that will be written down (semantics)

Hints to avoid turning good research results into a bad thesis.

➢ Focus on theses in software engineering
  ➢ There are different kinds of theses in different domains
    ➢ With different structure, argumentation line, requirements, etc

➢ Fact: there is no unique recipe
  ➢ Each thesis is unique
  ➢ There are however some patterns that many good theses follow

As side effect, you may infer hints to do research in a target-oriented manner.
Scope of this Presentation

Entry point
Ideas, results, experiments
Papers, technical reports, written fragments

Desired result

Fact: on the topic of your thesis YOU are the world expert

In the general field of the thesis are however other experts. There are some thesis writing patterns that you can use to convince the others.
What is a Thesis?

➢ An exposition of **original** piece of research
➢ A research **report**
➢ A **contribution** that you make to the research community
➢ What you produce after **3+ years of work**
➢ The product of a **one man show**
➢ (Maybe) the largest piece of work that you will ever do

“A thesis for the PhD must form a distinctive contribution to the knowledge of the subject and afford evidence of originality shown by the discovery of new facts and/or by the exercise of independent critical power.”

*(University of London regulations)*

➢ A **BIG** document
  ➢ Usually of ca. 150 pages, upper limit is > 300 pages
Who will read your thesis?

➢ You :-)
➢ Your professor
➢ Your examiner
➢ Your colleagues at the chair
➢ Your friends / acquaintances
➢ Other researchers from your field
  ➢ The experts look target-oriented to find what is interesting for them
    ➢ The expert will rather read your papers
  ➢ The fresh PhD students will read the entire thesis
    ➢ Write the thesis in a less dense style than you write papers
Why is the Thesis Writing Difficult?

- The content of the document should reflect many years of work
  - ... in a convincing manner, systematic,
- The document should be written in a scientific manner
  - with clear argumentation, rigorous,
- The document to be produced is big
  - Writing > 150 pages that fit together is really demanding
    - You actually write 450 and only 150 “will stay”
- The quality of the document
  - There is no upper bound for the quality of your thesis (is a strive to excellence)
  - The lower bound is NOT low at all
- The time frame in which you write will span over many months
  - The terminology from the beginning changes
  - You “forget”, or are not anymore “happy” with what you wrote 4 months ago
  - Writing is often a back and forth process
Marathon vs. Sprint

- Thesis writing is more like a marathon, paper writing is more like a sprint

Consequence 1 – **One full day myth**
- If you invest one full day into a paper, you can significantly improve it
- If you invest one full day in the thesis, the improvements will be minor

Consequence 2 – **The sprint does not help much in a marathon**
- There are no overnight surprises, you will KNOW when you are in the middle, towards the end or when you are finished
How to get closer?

➢ **Context:** a PhD thesis is a “big thing”

➢ **Problem:** how can I get closer?

➢ **Patterns:**

  ➢ Publish (several) papers about your research
    ➢ It offers a starting point in scientific writing
    ➢ It is good to have the main contributions backed-up by papers
    ➢ Workshop paper(s) → conference paper(s) → journal paper
      ➢ The journal paper requires a longer time frame to get accepted

  ➢ Start rather soon with the thesis writing
    ➢ Writing is anyway an iterative process that takes > 9 months
    ➢ Maybe based on a (journal) paper you already published

  ➢ Start your research from a project done
    ➢ Synergy effects between project and research work

**Fact:** as you write the thesis, new ideas might come
Your thesis can contain > 30% new material as compared to the starting point
Try to describe your thesis at different levels of detail
   ➢ … during a trip in an elevator (1 minute)
   ➢ … at a coffee (5 minutes)
   ➢ … in a long presentation (60 minutes)

Try to describe your work to different persons – while drinking coffee with
   ➢ … a common person from the street
   ➢ … a computer scientist
   ➢ … an expert in your field of research

Talk with other experts about what you do (e.g. at conferences, this is what “social events” are meant for)
   ➢ This helps you refine your ideas and argumentation line
The Process of Getting Prepared

➢ Context: the time before you really start writing

➢ Problem: how do I start?

➢ Patterns:
  ➢ Read several theses that fit your background, that you like, that **FIT** you
    ➢ Look at their structure, argumentation line, etc
    ➢ Try to see the big picture behind the details, discover patterns that fit you
    ➢ Both from your older colleagues and from outside
  ➢ Have the core of the thesis be based on an already published material
    ➢ If you don't have anything published, try to do this
  ➢ Write an extended abstract about your ideas
  ➢ Start inside-out (from the core concepts outwards)
Get the right tools

➢ **Context:** you write your dissertation on a computer

➢ **Problem:** how to write productively
  ➢ … and not to loose the dissertation :-) 

➢ **Patterns:**
  ➢ At our chair: get a Subversion repository
    ➢ Have a subversion explicitly for your dissertation or your personal subversion with an explicit sub-directory
    ➢ Commit and comment everything as if you'd work in a team
  ➢ Have a functioning backup system
    ➢ E.g., on a Mac: use Time Machine and test your backups
    ➢ Run it hourly to recover stuff you change or delete accidentally
  ➢ Use LaTeX
    ➢ Handles large documents great (figures, references)
    ➢ Typesets equations beautifully
    ➢ E.g., on a Mac: use TexShop and Bibdesk (SyncTex!)
I don't find the time for writing

➢ **Context:** project work, teaching, girl/boy friend, … all take time

➢ **Problem:** I have no time for writing

➢ **Patterns:**

   ➢ Define time as your dedicated writing time
      ➢ Mark it as such in your calendar, e.g., every day from 9 - 11 or 7.30 – 10 :-)  
      ➢ Teach your colleagues to respect your writing time
   ➢ Know your priorities – time allocation is a matter of defining your priorities
      ➢ After all, we all have 24 hours a day, 7 days a week

   ➢ **Fact:** you are the most interested person about your thesis

   ➢ Research is not really a 8 hours/day job
      ➢ It is rather a state of mind – good ideas might come Sundays at 7.00 AM
      ➢ Have any time something with you to make notes

➢ **Antipatterns:**

   ➢ Work mornings for the projects and (hopefully) in the evening for the thesis
      ➢ Many times you will be too tired and have no energy for your thesis
I'm stuck and get nothing done

- **Context:** You have problems getting words on the screen / on paper
- **Problem:** Writer's block
- **Patterns:**
  - Set *explicit goals*
    - Start a spreadsheet with your writing goals
    - Define the goals in words per day as well as content-wise
    - Increase your word goals
  - **Write anyway**
    - Don't think you need to get it right the first time
    - Write a first draft, improve it the next day
  - Alternate working modes: 1) *writing mode*, 2) *censuring mode*
  - Alternate activities, read some related work, do some experiments, program smth.
  - **Change the environment** – e.g. work in Staatsbibliothek, at Starbucks, …
  - Take a break (for a day), take a sleep, do some sport, etc.
I'm afraid that what I currently do does not fit

- **Context:** You are writing
- **Problem:** You have doubts that what you've just written is not ok

- **Patterns:**
  - Find a *spare partner*
    - A colleague that is also writing, somebody that finished, somebody that is especially critical
    - Give her/him the stuff you wrote – ask for a critical opinion
  - Play “devil's advocate” with yourself
    - Simply in a dialectical manner
  - Leave the part you work on apart for the moment, focus on other things, *come back later* with a fresh view
  - Search for other works that do similar things
    - how do they motivate?
    - how is their argumentation line?
I don't know how to start

- **Context:** you want to start
- **Problem:** with what do I start?
  - What do I cover in what sequence?

- **Patterns**
  - Create the directory “thesis” :-)  
  - **Start with an outline**
    - Put things into place on a high level
    - Develop a central thread, a central theme
      - e.g. based on what you already published
  - Try **writing an early abstract** that gives an overview of your finished dissertation to get your thoughts in order
  - Make a **first version of the “Introduction”** to organize your ideas
    - Don't be afraid to do it wrong, it will be refined iteratively anyway as you advance with the content
The Content Items of (every) Thesis

➢ Motivation
  ➢ Why do you think it is important what you did?
➢ Problem statement
  ➢ A clear statement about the problem that you solve
➢ Contributions
  ➢ An explicit list with your contributions
➢ Theoretical background
  ➢ What should a freshman know to understand what you do?
➢ YOUR RESEARCH
➢ Validation (experiments, case-studies, proofs)
  ➢ Show that your approach really solves the problem, that you fulfill the claims
➢ Discussions, variation points, critical view, limitations
➢ Related research (you are NOT alone)
➢ Conclusions and future work
What is a good motivation?

➢ **Context:** you are writing the first parts

➢ **Problem:** how to present the need for my concept?

➢ **Patterns**

  ➢ You can allocate an entire chapter for the motivation
  
  ➢ Present both state of research and state of the practice (if possible)

    ➢ “State of the practice”
      
      ➢ published empirical studies done by you or by others
      
      ➢ documents of well-known organizations (e.g. Standish Group)

    ➢ “State of research”
      
      ➢ use the ICSE road-maps, papers about challenges and visions published in highly ranked conferences and journals

  ➢ Try to find other researchers that also think that your work is important

    ➢ e.g. in “future work” of existent papers, “survey and future directions” papers, “great challenges” papers, etc

    ➢ If possible, give quotes

After the motivation you should derive the issues that you solve.
How to pinpoint what you did?

➢ **Context:** You are writing the introduction

➢ **Problem:** You did many things but you have to say them explicitly

➢ **Patterns:**

   ➢ Make the problem explicit – “Problem statement”
     ➢ Try to formulate the problem as clear as possible, in one or few sentences
   ➢ Make your thesis explicit
     ➢ Thesis: an affirmation stating how your work solves the problem
     ➢ … btw. “dissertation” means the whole document
   ➢ Define the scope, limitations, do not make uncovered claims
   ➢ Make the assumptions explicit from the beginning
   ➢ Contributions list
     ➢ What are the new things that you did?

**Fact:** contributions are the interface of your thesis
The List of Contributions is the Interface of a Thesis

- **Contributions** represent the “interface” of the thesis
  - Are the contributions enough? (“requirements”)
  - Are they clearly proven / sustained by the dissertation? (“implementation”)

- The contributions represent the original part of your work
  - A thesis has more contributions (however, some are more important than others)
  - Every part of the thesis should have something related to a contribution
    - Its motivation, related work, validation, etc.

- Possible kinds of contributions
  - Formalization, proof,
  - New model
  - New method, new methodology,
  - Improved method, methodology, formalization, proof
  - Empirical evidence
  - …
How to Make the “Related Work”?

➢ **Context:** finding “friends” and “enemies” :-)

➢ **Problem:** how can I write the related work

➢ **Patterns:**
  ➢ There are two types of related work
    ➢ “Friends” – motivate your approach
    ➢ “Enemies” – are in competition with you
  ➢ For each related work discuss its limitations and how you approach them
  ➢ Make a bigger picture where you organize the related approaches
    ➢ Categorize the work in logically coupled groups

➢ **Antipatterns:**
  ➢ “There is no related work syndrome”
    ➢ It is, always since everybody has neighbours !!!
  ➢ Write about irrelevant work for your approach
  ➢ Leave related work undiscussed
How do I convince others that my approach works?

- **Context:** you developed a model, theory, method, methodology, etc
- **Problem:** how do I clearly show that your approach works?
- **Patterns:**
  - Provide **validation for each listed contribution**
  - Link explicitly each part of your validation to the contribution that you validate
    - Provide an explicit list of **research questions**
  - In the case of empirical validation **discuss the threats to validity**
    - Internal validity – what factors did influence my experiments?
    - External validity – in what measure can I generalize the results?
  - **Discuss explicitly the limitations**, what you learnt, how you can extend the approach
  - Use **accepted benchmarks**, metrics, etc
Contributions and their Possible Validations

➢ **Formalization, proof**
  - The proofs are validation per se
  - Errors identified in the work of others by using the formalization show the added value of the formalization
  - New theorems, new analyses that can be done due to the formalization show its added value

➢ **New methodologies**
  - Show the systematic parts added due to the method
  - Show where others failed

➢ **Improved method, approach**
  - Show in that measure it is improved w.r.t. the existing approaches (benchmarks)

➢ **New model**
  - Two staged validation:
    - Stage 1: use several use-cases to show the applicability of your main concept
    - Stage 2: validate each of these use-cases with a case-study
Writing the Evaluation

➢ **Context:** you are writing the evaluation

➢ **Problem:** how to evaluate the results?

➢ **Patterns:**

   ➢ Define a set of criteria that your concept should fulfill
     ➢ Use these criteria in your experiments

   ➢ Write a list of research questions
     ➢ Show what you investigate in the experiment
     ➢ Show how are these research questions related to the list of contributions

   ➢ Give a clear description of the chosen case-studies
     ➢ Others might be interested to redo your experiments

   ➢ List the objectives of the case-study

   ➢ Say explicitly why are the chosen case-studies relevant
     ➢ Say why did you choose these case-studies (and not others)
How can theses be read?

- **Sequentially** – almost nobody will read your dissertation like this
- **Randomly** – a few pages from here, a few from there
- **Starting from “contributions”**
  - Read where are they motivated
  - Read where are they implemented
  - How are they validated
  - How is the related work w.r.t. these contributions done
- **Starting from the validation**
  - Read backwards the thesis (“what is not validated does not exist”)
- **Starting from conclusions**
  - Read backwards the validation, theory, motivation

**Fact:** Usually, a thesis is read mixed
Nobody has the time and energy to read the thesis like a novel.
How to make the document readable?

- **Context**: a PhD thesis is a **BIG** document
- **Problem**: how can I facilitate its reading?

- **Patterns**:  
  - Provide a road-map of the document  
    - Present the big picture  
  - Make each chapter self-contained:  
    - write an abstract that shows its role in the big picture,  
    - write a summary that enumerates the most important issues  
  - Link explicitly the contributions to chapters  
  - Am I allowed to make jokes?  
    - YES, whenever they are relevant and smart (they make the thesis easier to read)  
  - Use a running (toy) example  
    - this is however NOT the validation
Generally “short is beautiful”
  ➢ Shortness is NOT an excuse for gaps!!!
➢ Length is NO replace for contributions
➢ How much can I write?
  ➢ You can write (almost) as much as you want as long as you really need that
➢ Aim at high cohesion between different parts
➢ Publish articles about parts of the thesis
  ➢ Represent a validation of the results
➢ Start with the writing rather early
  ➢ You can always change the document, however, you need to have what to change
➢ Make your bibliography look good
  ➢ Consistent citation style, pay attention to details, offer complete information
➢ Ask more people to review (parts of) your thesis
  ➢ More small feedbacks together means a big feedback
Final Antipatterns

- Some parts of the thesis (e.g. sections originating from published papers) are much much better than the other parts
  - This is per se not bad but obscures the other parts
- Document is heterogeneous
- There are no links between sections
- There is no argumentation line
- Uncovered claims
  - Are the easiest to refute (e.g. by playing “devil's advocate”)
- Pictures, work of others taken without a reference
  - This is really really problematic
- Citations that are weak from a scientific point of view
  - Wikipedia, obscure local workshops, etc.
- Argumentation referring to space limitations – “Due to space limitations ...”
  - The limits of a PhD document are very generous!!!
The End of the “Thesis Writing Marathon”