Producing a scientific thesis / paper

by Aleksandar I. Ivanov
Dresden University of Technology

I have come to develop a simple and effective method for how to approach the writing of a thesis, structure it, work it, produce text and results, and control the process. Since writing a thesis can be a daunting task for newbie future scientists, I thought I'll put this little list of actions together in the hope of making people's lives easier.

A caveat emptor: Scientific working is a highly non-sequential, non-linear activity and thus the rather stiff sequential method given below is to be used to get oneself started, and not as the ultima ratio for "doing" science. There will and should be many revisions and lots of "prototyping" & feedback loops till a satisfactory result can be reached.

This note is still kept short, since it seems to be a mater of fact, that huge books about how to do scientific research, are usually written by people who have done lot's of it already. To the author here it seems the pedagogically best way to introduce scientific newcomers to their rules of play by a *short* manual, not by a 3-pounder from Karl Popper or the like. Now, let's go:

V1.0

Part I – Getting Started

1. Get in touch with the topic / question you're interested in

Read some papers, browse the Web for information, etc. about your topic.
→ Tools: Information resources, WWW, libraries, databases, special interest journals, etc.

2. Define the 1st very rough draft research question.

What is a research question? Define the target and the particular aspect of it you want to study. … A good research question needs time to emerge, hence don't feel discouraged if at first you're incapable of producing a clear cut, well posed research question.
→ Tools: Text document file (MS Word, OpenOffice, TeX whatever)

3. Produce a quick rough outline

Set up a 1st draft outline of the thesis/dissertation. Very rough, cause you probably don't know much about the topic yet.
→ Tools: Text document file

4. Control your search with a Search List & Needed Readings List

Start searching the resources (Scientific databases like EBSCO or Science Direct, Google Scholar, your local university library, the Web, etc.) and control your search using a spreadsheet "Search List", just so you know, where you've searched for what:
Results don't need to be listed, just list what search terms you've tried in the various information resources, and maybe the number of hits if you like, but not (!) the full listing of hits – you take what you need on your Reading List – see below – and leave everything else, otherwise, you'll suffer a brain overload :-)
Collect the electronic files (usually PDF and DOC files, but also data and software) on your hard drive and list the titles of documents/books you could not get but would like to look at in a "Needed Readings List" - call it NEEDED_Stuff.doc or something.
→ Tools: Search List (simple quick Excel spreadsheet), Needed Readings (text document)
→ Note: Find a naming convention for all electronic files you collect, e.g. for a paper
"Title of the paper, author, year.pdf" -> "Empirics of Growth, Dirac, 1988.pdf"

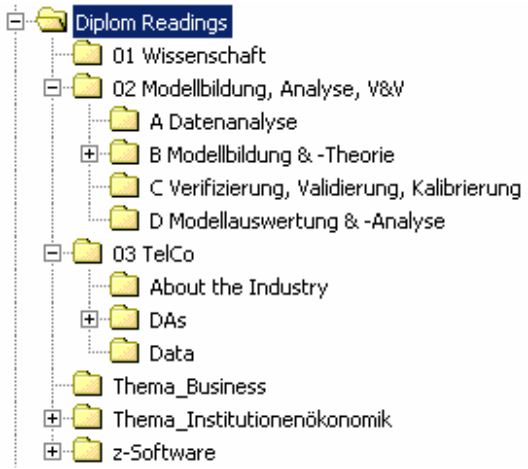
Example of a Search List:

EBSCO Business / Academic	Resultate	**JASSS**	Resultate	**SLUB library**	Resultate
mobilfunk	-	Business	-	agent-based	20
cellular agent	- (6000)	Telecommunication	-	titel:agent titel:business beliebig:simulation	100
cellular phone agent	-	cellular agent	-	titel:management beliebig:simulation	1527
customer	-	mobile phone	-	titel:unternehmensführung beliebig:simulation	50
mobile phone agent	-	mobile phone agent	-	titel:führung beliebig:simulation	2
agent-based mobile	-	agent-based +mobile	-	titel:strategie beliebig:simulation	1131
agent-based mobile comm	0			titel:strategie beliebig:{simulation, agent}	1131
agent model business	-				
agent model business management	0				
agent-based modeling business	-				
agent-based modeling management	-				

5. Build a Folder Structure on your hard drive

Now start skimming through the readings you've gathered in step 4. Take your time for this (1 - 2 weeks or as many as you need). Try to develop some sort of Folder Structure on your hard drive in which you can sort the electronic files you've gathered and named. This will be very useful, trust me. A folder structure may look like this (see below for a screenshot of a Windows Folder Structure):

1 Theoretical background,
2 Building the model / theory
  2.1 data analysis
  2.2 modelling: parts of the model
  2.3 modelling: verification & validation
3 Using a model / testing it
  3.1 Analyzing the model / theory
  3.2 Statistical analysis of model / theory outcomes / predictions
  3.3 Contrast model / outcomes outcomes with real life data / statistical testing... (see 3.3)
4 Conclusions
0 General Information, Miscellaneous



Example of an *evolving* Folder Structure in Windows

If there are files you cannot assign to your folders, put them in a General Info folder or something of that sort. Later you may realizing that there's yet another folder needed to complete the Folder Structure.
Cross-referencing: In Windows or Linux also use links to create multiple references for a file in different folder locations! Example: You may put a PFD-file called, say, "Synergetics, Haken, 1970.pdf" in some folder named *FOLDER_1* and, if you like, a link (in Windows a link is another file with extension *.lnk,* which you may name as you wish, eg. "Synergetics, Haken, 1970.pdf.*lnk*" – *lnk*-files are usually only 1 KB large) in another folder named *FOLDER_2*.

→ Tools: Your computer's hard disk

6. Build a Reading List

Now build your Reading List file according to the same structure (put each reading entry under a topic-heading of your folder structure). Mark the most important readings with some symbol so you know what your first reads are. You don't need to include all the stuff you've found, maybe just the best ones, but keep everything you've ever downloaded in the folder structure.
This Reading List serves the following functions: Gives you an overview of the stuff you've gathered, your reading workload and reading priorities / schedule, and it should be updated with the new readings you will come up with in the process.
The hard drive folder structure is simply a convenience tools for looking up files and it will become an information resource when more and more electronic files come together. So don't delete the stuff.

→ Tools: Hard drive folder structure, text document or spreadsheet for the Reading List

Excerpt from a Reading List: Note the info tags I am using in the front or back of some of the entries. You can design your own.

- Ebeling: Komplexe Strukturen UG 3000 E15 (ZB)

- Leiber: Kosmos, Kausalität und Chaos - naturphilosophische, erkenntnistheoretische und wissenschaftstheoretische Perspektiven - CC

- Berger, Thomas (Bonn) Agentenbasierte räumliche Simulationsmodelle in der Landwirtschaft, ZB 55000-168, Drehpunct

- Follesdal, D./Walloe, L./Elster, J.: Rationale Argumentation. Ein Grundkurs in Argumentations- und Wissenschaftstheorie.

- Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View

HANDAPP (Prof. Computational Physics)– Badii: Complexity - anrufen, ob ich anschauen kurz ausleihen darf?

ORDERED - Simon, Herbert: Models of man (1957!)

ORDERED - J. Kleijnen, Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: J. Banks, Editor, *Handbook of Simulation*, Wiley, New York (1998), pp. 173–223 (Chapter 6). → Vorbestellt, autom. Bereitstellung in SLUB (eMail)

One might also want to merge the Needed Readings List with the Readings List. Personally, I did not do it, because my Needed Readings List became somewhat crowded and is in constant flux, plus I prefer to have a nice looking Readings List which fits on less then say 5 pages.

7. Taking Reading Notes in a Systematic Way

By this time you must have read lots of stuff. Now, at the latest, is the time to start taking notes in a systematic way. For this to happen, first refine your research question if needed, then refine your outline using your improved understanding and knowledge of the topic.
Now, start your reading using the Reading List and take your reading notes like this:
Copy the thesis/dissertation outline into another text document file (MS Word, OpenOffice, whatever) and put each note you make directly under the heading it corresponds best to (use cross-references extensively → this enhances your notes as if they were a HTML-document or a mind map). So what you do is start filling up the outline, which previously had no other content than headings & subheadings, with text/notes.
When taking a note always add the source reference after or before the note *and* each time you've added a *new* source, add it also to the *References* section of your outline.
You get the idea: While taking notes, you automatically start producing content for the chapters of your thesis plus the References are growing automatically. Nice, huh?!

→ Tools: text document file plus work with the tools/files you've developed in the previous steps (Reading List, Outline, Search List, etc.)

8. In the Loop

So, now you can iterate some or all of the previous steps, work through the readings, keep updating & refining your research question and outline, going back to information resources for more info, and so on. It is a good idea to print out your systematic notes from time to better be able to rearrange. At some point you will start doing your own original research (design a theory, produce a model, etc.) and incorporate your results into your writing under the corresponding headings (the "My Theory / Results" chapters). You will start transforming the notes into publishable text paragraphs and then some day ... your thesis/dissertation will be there ;-)

Remarks:

- There's also software, which supports steps 5, 6 and 7, e.g. Citavi (Free) and Visual Composer (40 - 80 EUR), but I feel they are too cumbersome and less flexible then a simple text document, which you have complete control over. I have tried them, and prefer to use the more flexible and quicker Word and Spreadsheet files as tools. Also, you have complete control over your working style, while the software confines you to its build in functionality only. Mind Manager or Free Mind could be beneficial, too. Check, however, whether they have sufficient exporting capabilities, so you can use the content you create in one of them in other software as well (e.g. Word or Excel).

- You may also feel the need to open up additional lists, e.g. a list of possible analysis methods, a list of important key words, etc. Or you can add more information to the Reading List, like a File Numbering or a Tag System about the Content of some File (*TSCF*), which might come in handy when using a spreadsheet and the atuo filter funtion, etc. pp.
It's all up to you.

- Also, it is possible to introduce time tables to support & control your progress, e.g. a Reading Schedule, a Global Project Schedule, etc.

- Everyone should modify the method described above to fit his/her own needs. This is merely a proposal for "how to ...". Also, this method is described for the computer only, but you can equally well do all of this using pen & paper. The advantage of the computer is, that you can search all your records, files, etc. very quickly (in Windows: Strg+F) and you can carry them around, send them, reproduce them, etc. etc. - it's the magical world of digital ☺

Part II – Hard Work

You could now create a **Global Working List**, possibly amended by a Global Time Schedule.

A Global Working List for someone who is attempting to develop a new theory or model could look like this:

1. PREPARATION PHASE

   This is exactly the first run of Part I up to point 6: Reading List, Outline, information gathering, talk to people, scope & forming the research question, etc.)

2. READING PHASE

   Well, this should be Part I, from point 7 onwards.

   Optional: *Looking at existing models and software systems while reading makes sure that your own skills practicing, experimentation, deepening your knowledge and modelling come in all together - naturally leading to the 3rd phase here:*

3. PROTOTYPING & EXPERIMENTS WITH YOUR MODEL / THEORY / WHATEVER YOU'RE DOING

4. FINALIZING YOUR MODEL

   Verification & Validation, Calibration, etc. – all the stuff a modeller needs to do.

5. CONDUCT THE ANALYSIS ACCORDING TO YOUR RESEARCH QUESTION

6. WRITE THE RESULTS & COMPARISON CHAPTERS

7. POLISH OR START WRITING THE INTRODUCTORY CHAPTERS, CHAPTERS ON PHILOSOPHY OF SCIENCE, OTHER STUFF YOU NEED, RESPECTIVELY

8. DRAFT A FINAL CHAPTER (THE CONCLUSIONS)

9. REVISE YOUR WORK, MAYBE CHANGE THE STRUCTURE OF THE WORK / CHAPTERS, POLISH THE LANGUAGE, ETC.

10. FINE TUNINGS, SUBMIT YOUR FINAL WORK, THINK ABOUT YOUR FUTURE CAREER WHILE WAITING FOR THE ORAL EXAMINATION / GRADING

As you can see, you should allocate substantial time to revisions and improvement of structure, fine tunings, improving the writing, etc.

Layers

I would also like to add another dimension to the simple linear structure above, which is especially useful if one is not sure, what level of depth and/or breadth (complexity) to chose for the work . It is useful to slice the topic into layers of complexity. Layer one could be a

grounding and the minimum which is to be achieved in one's work, e.g. a simple model with 2 parameters. Everything described so far, can be applied while working in that layer.
Then if in the process one decides to do more, just "switch" to a more complex layer, e.g. decide to make a model with 6 parameters, and non-linear equations. When switching to a more complex level, basically all the steps of your work will take more time to complete due to the increased complexity.
If that isn't enough, just switch to layer three: A model with 6 parameters, non-linear equations, adaptive intelligent objects, etc. – sky's the limit (or actually our brains).
Layers help to create clarity in your work, and avoid that you get confused with complexity. Also, different levels of complexity allow for different Research Questions, which must be reviewed every time you switch levels! That's important, else you'll be doing more complex models but not answering more interesting questions – you have to ask them first, i. e. adjust your Research Questions.