

B.Sc. / M.Sc. Workflow, Hints, and Information

Benjamin Schiller
schiller@cs.tu-darmstadt.de

July 27, 2012

Abstract

This documents gives a rough overview over the workflow when doing a Bachelor or Master thesis with us. Also, it includes hints on how to find and read papers, how to structure your thesis, and how to properly write down your accomplishments. Furthermore, it contains some information about the tools we use, the accounts you need to create, what information we need from you, and how to proceed when handing in your thesis.

1 Rough timeline of a thesis

A thesis (very) roughly consists of four main parts: reading, solving a problem, implementation, and evaluation. While they build upon each other, there is not always a clear cut between them. Writing your thesis is an ongoing process which you should pursue from the first day, starting by annotating the first paper you read. Giving exact durations to these phases is quite impossible as it highly depends on each thesis and topic. This is why thinking about a rough timetable is the last question of your exposée as described in Section 2.

1st phase - Reading In the first part of your thesis, you get an understanding of its topic and contents by reading and researching the related work, i.e., what other have already done in your field. This gives you a better understanding of the problem we are attempting to tackle and hopefully solve in this thesis. After you have finished this part, you should have a good understanding of the area we are working on. It should then be easy to answer all the questions required for the exposée (cf. Section 2). After finalizing the exposée, we register your thesis at the Prüfungssekretariat and the official time of 3 months (B.Sc.) or 6 months (M.Sc.) starts.

2nd phase - Solving the problem After understanding the problem we are about to tackle, we start thinking about a solution. This means figuring out how to solve the problem we are investigating or at least parts of it (in order to make the world a better place).

3rd phase - Implementation In most cases, we need to implement the solution or a way to evaluate our solution / test our hypothesis, etc. After figuring out a solution to our problem / a way how to solve it or at least parts of it, you will start to implement means to show that our solution is valid.

4th phase - Evaluation The last phase of your thesis will be the evaluation of our solution. For example, evaluating multiple simulations runs of a system under investigation and drawing conclusions about the investigated system, its parameters, out hypothesis, etc.

Handing in your thesis When handing in the final version of your thesis, we need four printed copies. In two of them, a CD/DVD containing source code, thesis sources, result data sets, etc., must be included. If you want to save some money, you can print the two other copies in black/white.

The deadline for handing in the thesis is determined when registering the thesis.

Defense After handing in your thesis, you will give a presentation (defense) where you describe the contents of your thesis, what you did, what the results are, etc. This should be done roughly four to five weeks after handing in the thesis.

2 The exposée

After the first phase of your thesis, i.e., reading and understanding about your topic and the related work, you are expected to write a short exposée. It is supposed to answer the following seven questions. Answering these questions ensures that you understand the problem we are tackling and that we are on the same page on what is supposed to be done as part of your thesis.

Problem statement The problem (what /exactly/ is the problem tackled in this thesis)

Relevance / motivation Some support for it's relevance (why is it really a problem and why will solving it change the world and make it a better place?)

Background A rough overview over background (explain the basics) and related work (what do others solve and what is missing) (it can help to already roughly sketch the requirements for the system, as well as metrics that can be used to show if a solution is good or bad =, especially the metrics that will show that your solution is better than the state of the art)

Our approach A first sketch of an idea how the problem could be solved in the thesis – here the proposal will indicate which ideas the student wants to follow and how they will be implemented

Solving the problem A very short rationale why this solution will solve the problem (and how)

Evaluation An indication how the solution will be evaluated (at best: together with the expected outcome) – here it will indicate how the student will show the superiority (or characteristics) of the solution, will most probably contain two of (mathematical analysis & proofs—simulation—prototypical implementation and measurement study). Keep in mind that you will have to compare your solution to a "baseline", iow. the best existing solution, and that you will somehow need to show this in your thesis (evaluating /both/!) – unless you are the first one solve a new problem and there is no prior art (in which case it is still a sign of quality to come up with a baseline like, e.g., an upper bound (the optimal solution), or a trivial solution).

Timeplan An initial time plan: what are the major steps (and milestones) and roughly during which time will they be dealt with. Should contain "literature review", "initial design phase", "implementation", "implementing the evaluation", "evaluation", "writing", and "final revision and correction". Allow at least 2 weeks for the revision and corrections and at least 4 weeks for the "writing" (but make no mistake: state writing early, you will need at least the four weeks to make a story out of you text snippets, to rearrange content avoiding redundancy and streamlining the logic, etc.

3 Finding and reading papers

3.1 Publication formats

The following list gives a quick overview over the different ways, how scientific results can be published.

- **books / book chapter** - detailed information, mostly very good quality, not necessarily up-to-date
- **journal article** - strictly reviewed, long duration until publication, potentially outdated
- **conference paper** - quality depends on the conference
- **conference poster** - very new but potentially uncertain results / rough ideas
- **workshop paper** - often work-in-progress, quality depends on associated conference
- **technical report** - detailed description of a problem and its solution, not peer-reviewed
- **webpage** - not peer-reviewed, can only be quick notes

3.2 Finding additional sources

In essence, there are three ways to find publications on a given topic: searching by keywords, backwards referencing, and forward referencing.

The easiest way is to use the search functionalities of a publication database like *Google Scholar*, *ACM*, or the proceedings of relevant conferences. Always keep in mind that not everyone uses the same terminology and keywords even though working on the same topic. Hence, always think about different keywords that might point to publications from the area you are working on. The following list includes the most important portals for finding papers:

- ACM - <http://portal.acm.org/portal.cfm>
- IEEE Xplore - <http://elib.tu-darmstadt.de/ieee/Xplore/dynhome.jsp>
- ULB - <http://www.ulb.tu-darmstadt.de>
- Libraries - <http://www.wbs.cs.tu-berlin.de/bibliotheken>
- SpringerLink - <http://www.springerlink.com>
- Science Direct - <http://www.sciencedirect.com>
- CS Bibliographies - <http://liinwww.ira.uka.de/bibliography>
- DBLP - <http://dblp.mpi-inf.mpg.de/dblp-mirror/index.php>
- Google Scholar - <http://scholar.google.com>

- CiteseerX - <http://citeseerx.ist.psu.edu>
- Zentralblatt Math - <http://www.zentralblatt-math.org/zmath/en/>
- Arxiv - <http://arxiv.org>
- Conference webpages ...¹
 - always prefer the good conferences: INFOCOM, SIGCOMM, CCS, S&P (Oakland), WWW, Usenix, PETs, STOC, FOCS, ...

When you have found a paper that is relevant for your work, you should also look at its references. Since most papers contain information about related work that has been done before, the bibliography should contain the most important works of the same field. Therefore, it is always a good idea to simply look at the bibliography of relevant papers and investigate the papers that have been referenced often.

The other way around would be to look at the papers that reference (potentially older) publications that are relevant to your work. Most portals provide a list of publications that cite a specific paper, e.g., in Google Scholar under the link *Cited by*.

Very often, you will find a paper with a title that sounds very promising and like a perfect fit to your topic. Since not all of these papers are in fact relevant for your work, before printing and potentially reading a paper completely, skim over the abstract to check if the content of the paper really contains what the title (in your context) seems to promise. In case the abstract sounds promising as well, read the summary and conclusion next before including a paper into your reading queue.

3.3 Reading papers

Reading papers:

- decide on a set of questions you want to be answered by this paper
- write down/mark the answers to those questions + other interesting stuff you find
- relate the paper to others, categorization
- you can (but don't have to) use tools for organizing literature (e.g. Mendeley <http://www.mendeley.com/>)

Writing the final version:

- decide on structure:
 - abstract, introduction and conclusion are necessary
 - possibly a background section introducing concepts necessary to understand the actual content: notation, mathematical concepts and theorems, basic terms
 - the remaining structure should classify the papers you read in some way
 - most cases: one general section introducing the different categories and one (sub)section per category
 - alternative: discuss analytical ideas in one section, practical evaluation in a second section
 - alternative: discuss ideas in one section and results in a second
- write each section down in bullet points usually starting with the 'content sections', organise bullet points into groups, so that each group contains one basic thought (which will later be one paragraph)
- write down the full text
- several iterations of corrections:
 1. structure, overall content (ideally before writing down the full text!)
 2. content, sense, understanding
 3. grammar and spelling

LaTeX:

- compile regularly, at latest after each paragraph, after each (long) mathematical equation and whenever you try something new
- behavior varies between different operating systems, LaTeX distributions and Editors, so make sure you know your system setting when googling for help
- Guide: <http://en.wikibooks.org/wiki/LaTeX>
- Forum: <http://www.latex-community.org/forum/>

¹http://en.wikipedia.org/wiki/List_of_computer_science_conferences

4 Outline of a thesis

Tips on scientific writing

- <http://homepages.inf.ed.ac.uk/bundy/how-tos/writingGuide.html>
- <http://www-net.cs.umass.edu/kurose/writing/>
- <http://www.cs.columbia.edu/~hgs/etc/writing-style.html>
- Book: *Zen - or the art of motorcycle maintenance*, *AUTHOR TODO*²
- Book: *The craft or research*, *AUTHOR TODO*³

Outline for a computer science paper

The following list gives an overview over a good outline for a computer science paper (according to Al Bundy)⁴.

Title - ideally the title should state the hypothesis of the paper

Abstract - state hypothesis and summarise the evidence that supports or refutes it

Introduction - motivate the contribution!

Literature Survey - broad and shallow account of the field, rival approaches, drawbacks of each, major outstanding problems

Background - states previous work in more detail, where this is necessary for understanding

Theory - underlying theory, definitions, theorems etc.

Specification - requirements and specs of implementation

Implementation Evaluation Related Work - narrow but deep comparison with main rivals

Further Work Conclusion - summarise research, discuss significance, restate hypothesis and the evidence for and against it, - recapitulate original motivation, reassess the state of the field in the light of this new contribution

Appendices

Plagiarism

- [TODOLINKTUDINFFB](#)
- http://www.williamstallings.com/Extras/Writing_Guide.html

Structural hints

The following lists gives a rough overview on how to structure sections, paragraphs, and sentences and how to deal with different levels of abstraction.

Sentences A sentence presents a single statement, e.g., $x = y$, *red is not blue*, $z < 1 \Rightarrow x = y$.

Paragraphs A paragraph contains all statements belonging to the same idea or concept. For example, $x > 2$ and $y = 7 \Rightarrow x > 10$ should be grouped in a single paragraph. Separate paragraph on the other hand should be used for statements like “*the weather is nice*” and “*elephants are huge*”.

Sections A section combines all statements that belong to the same idea / line of thoughts. In most cases, this implies that they are on the same level of abstraction. Subsections should logically divide the content of a section and build upon each other, e.g., “preliminaries for X”, “approaches for X”, “new ideas”.

Abstraction levels Inside a paragraph, always stay on the same level of abstraction, e.g., design, specification, implementation. In case you change the level of abstraction inside a section, do not jump between these levels. Either move from a high level of abstraction to a low level or the other way around (e.g., design \rightarrow specification \rightarrow implementation or implementation \rightarrow specification \rightarrow design).

Reader’s digest The first part of every section summarizes what the content of this section is, e.g., “In this Section, we motivate our work, describe the problem space, and solve the problem”. This allow the reader to quickly get an idea about the content of a section.

²TODOLINKTUDINFFB

³TODOLINKTUDINFFB

⁴<http://homepages.inf.ed.ac.uk/bundy/how-tos/resbible.html>

4.1 Abstract

The abstract of a thesis consists of the following four points (one paragraph each):

1. Problem statement: What is the problem we are trying to solve? (One problem is more than sufficient for a thesis!)
2. Relevance: Why is this problem really a problem?
3. Response: What is our solution to the problem?
4. Confidence: how do we show in this thesis, that our solution is good?

4.2 Introduction

- Broad Topic, potentially little broad background
 - topic, special problem we're looking at, motivation
 - probably more background for our problem (why is it actually hard?)
 - broad background, general definitions
- our goal and our claims (what are we solving in this work?)
 - Our goal, research question, motivation and relevance (Why is it a problem the reader should care about? Why is it hard?)
- Requirements for our solution
 - Requirements for a good solution
- Which metrics can we use to show the quality/quantity of our solution?
 - probably rough definition of metrics
 - Metrics to measure how good a solution is
- If space missing the related work may be presented in a paragraph here
- Summary of our solution
 - Overview of our solution and first confidence (how do we show that it's good?)
- Our contributions in this paper
- outline of the paper / reader's digest
- Related work (can be done together with literature survey)
 - STATE HOW THE RELATED WORK RELATES TO YOUR WORK!! (how is it similar, how is it different?)
 - Related work is not your enemy, but gives you "the shoulders of giants" you can stand on
 -
- –
-
- reader's digest

4.3 Background

- Literature Survey and Background
-
-
-

4.4 Theory (probably)

-
-
-
-

4.5 Specifications

-
-
-
-

4.6 Implementation

-
-
-
-

4.7 Evaluation

-
-
-
-

4.8 Summary & conclusion

-
-
-
-

5 Writing proper English

In general, it is up to you to choose the language of your thesis. Since basically everything in computer science is done in English and there are many technical terms that do not have German translations, I strongly recommend to write in English.

A very good and short book with basic and very useful rules and hints is *The Elements of Style* by William Strunk Jr. and E.B. White⁵ (cf. Figure 1). All basic rules are available in various online resources⁶. But for a price of less than 10 Euro, I consider it a good investment⁷.

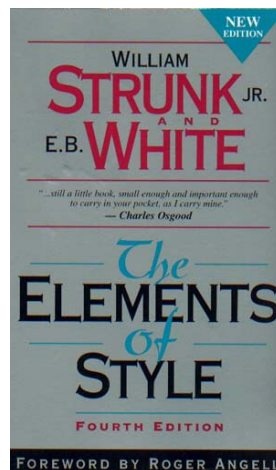


Figure 1: The Elements of Style

5.1 Some rules

The following list contains some important rules (some of them included in Strunk's basic rules) which are disregarded quite often.

- *my father's dog, the inhabitants' opinion, Hans's car*
- *it's* is an abbreviation for *it is* (should not be used!) and differs from *its*, the possessive pronoun
- *a, b, and c / a, b, or c* in clear contrast to the German *a, b und c / a, b oder c*
- *i.e.* (id est) does not introduce an example but a clarification

⁵http://en.wikipedia.org/wiki/The_Elements_of_Style

⁶<http://www.cs.vu.nl/~jms/doc/elos.pdf>

⁷http://www.amazon.de/The-Elements-Style-E-White/dp/020530902X/ref=sr_1_1?ie=UTF8&qid=1341214949&sr=8-1

- *e.g.* (exempli gratia) introduces an incomplete list of examples (can be only one), never to be ended by etc.
- *cf.* (confer) is used to compare a statement to another part, e.g., a Figure presenting evidence for your statement (do not use cp.)
- *i.e.* and *e.g.* are always enclosed in commata: *my father, i.e., the guy over there* (some style guides only mention the first comma)
- *Figure 2.3, Listing 1.2, Table 8.2.3, Section 4.3, Chapter 1*, etc. must be capitalized (basically every term that indicates / is followed by a reference to another part of the document)
- there must be a reference to every table, figure, etc. that is included in your thesis, unreferenced elements need to be removed (or referenced)
- figures, tables, listing, etc. are not the carrier of information, they are only used for clarification and details that are not required for the overall understanding of what you are talking about. the crux of what you are talking about should always be contained in the written text so that figures could be removed without losing the basic meaning of what you want to express
- *we* are explaining, performing, evaluating, etc. (do not use *I* or passive statement, like, e.g., *in Section 2.3, XYZ is explained*)
- stick to present tense, e.g., *we describe* instead of *we will describe*, *we evaluate XYZ* instead of *we evaluated XYZ*
- do not start sentences with *To ...*, use *In order to ...* or a different statement instead
- when writing about a person / user in general, do not use *he*, use *she* instead

5.2 Some resources

Some interesting (and potentially funny) resources for writing proper English:

- <http://theoatmeal.com/comics/ie>
- <http://theoatmeal.com/comics/misspelling>
- <http://theoatmeal.com/comics/semicolon>
- <http://theoatmeal.com/comics/apostrophe>

6 Accounts

In order to comply with the infrastructure of our group / that our group uses, you need to create some accounts to allow for a seamless workflow.

6.1 Github

Version control is very important when developing any kind of software. Because of the many drawbacks of SVN, I decided to use git exclusively. All projects are currently hosted on Github ⁸. In order to use this infrastructure, you need to create an account. The most important feature of Github is the use of issues allowing for a very slick workflow for reporting, discussing, and fixing errors as well as feature requests (cf. Figure 2).

If you are developing for GTNA ⁹, please follow the instruction provided in the repo to create an own repository as a fork of the main project ¹⁰.

In case you are starting a new project, I will create a repo for you in the Github organization of our group ¹¹. Please create the following branches for storing the respective data:

- *master* - code (of course other branches as well)
- *doc* - documentation files for the code
- *docs* - documents like lists, pdfs, etc.
- *thesis* - you thesis (latex code, images, other sources, no binaries except for thesis.pdf)

Please refer to the following guide on how to create an empty branch that does not branch from the master ¹².

Please note that you can register as a student on Github ¹³. This will give you a free micro plan (instead of paying the 7\$ per month) which allows you to create five private repositories ¹⁴.

⁸<https://github.com/>

⁹<http://www.p2p.tu-darmstadt.de/research/gtna/>

¹⁰<https://github.com/BenjaminSchiller/GTNA/blob/master/CONTRIBUTING.md>

¹¹<https://github.com/organizations/P2PNetworksTUD>

¹²<http://www.bitflop.com/document/116>

¹³<https://github.com/edu>

¹⁴<https://github.com/plans>

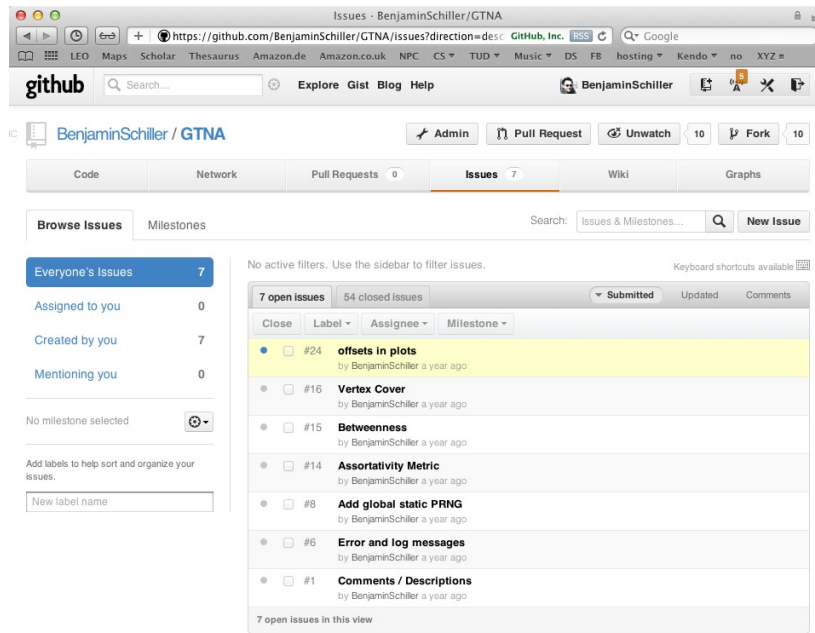


Figure 2: Github issues for a repo

6.2 Mendeley

*Mendeley*¹⁵ is basically a social network for researchers. But more importantly, it is a great way to organize articles, papers, etc. that you find and read during the first part of your thesis. Besides a web interface, *Mendeley* offers desktop applications in Windows, Mac, and Linux that allow you to organize papers, annotate them, and add notes to summarize their content and write down important facts about a paper you read (cf. Figure 3).

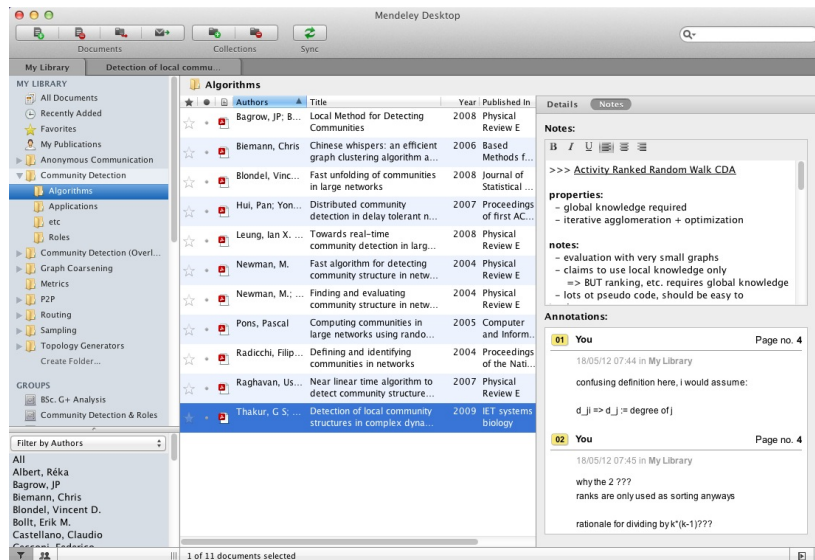


Figure 3: Mendeley Application on the Mac

During the first phase of your thesis, you will read a lot of papers. *Mendeley* can help you to organize them and also store the information you found about the papers' content. In order for me to see the paper you found and also see your annotations, please create a private (important!) group in Mendeley and invite me to it (you need to invite me with the address `benni@codingz.de` for it to work). Besides annotations, I highly encourage that you add a quick summary of the paper in the notes section for each paper you have read. This can be as short as explaining parts of an algorithm in detail or quickly describing why the paper is not really interesting for your topic. This note should always include what the problem is the authors address (e.g., build the perfect Darknet) what is done in this paper (e.g., a new algorithm for XYZ), and how their solution is evaluated (e.g., simulation, theoretical, etc.).

While there are also mobile application for Android and iOS devices, they don't allow for annotations on the respective device. Since I like reading on the iPad and want to have digital annotations in the papers without

¹⁵<http://www.mendeley.com/>

transferring them in the Mendeley desktop application later, I use the following workflow for reading papers (iOS of course, but should work with similar applications for Android):

1. Create a *README* folder in *Dropbox* ^{16 17}
2. Organize and categorize papers in subfolders
3. Import the *README* folder with *GoodReader* ¹⁸ (and sync regularly)
4. Open, read, and annotate the paper with *GoodReader's* annotation tool
5. Sync annotated paper back to *Dropbox*
6. Add paper from *README* to *Mendeley* using the desktop application

6.3 Wunderkit

Managing tasks for a thesis and organizing the notes during and after meetings is quite annoying. While sending emails is ok, finding and organizing the information in a suitable fashion is quite impossible. A nice tool that allows the collaborative administration of notes and tasks is *Wunderlist* ¹⁹ from 6 *Wunderkinder* ²⁰.

Currently, standalone applications are only available for Mac and iPhone (cf. Figure 4). But there is a nice web application that offers the same functionalities for non-Mac users.

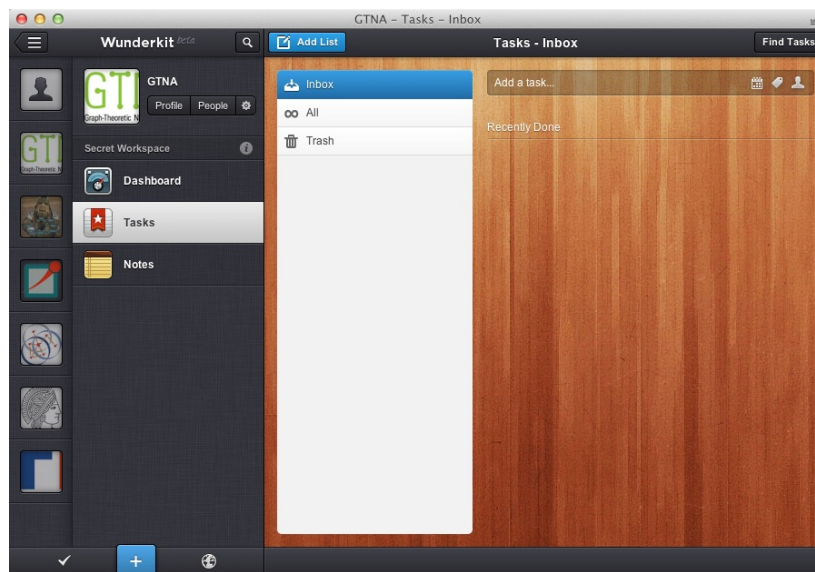


Figure 4: Wunderkit Application on the Mac

In order to manage notes and tasks for your thesis, I create a new project in *Wunderkit* and invite you to it. This way, we can edit notes together, create new tasks, and complete them.

6.4 TK VPN

In case you are going to perform computations on one or multiple of our machines, you will need to be able to connect to them. Most machines are configured to only accept SSH connections from IP addresses from the TUD's range or even more restricted to our own pool of addresses. Therefore, it is necessary that you obtain a TK account allowing you to use the TK VPN.

The settings for connecting to TK's VPN are as follows:

- Server: *vpn.tk.informatik.tu-darmstadt.de*
- Username: *TK account*
- Password: *TK password*
- Domain: *TK*
- Type: *PPTP*

On Mac I have an option to send all traffic over the VPN connection. Assuming the existence of a similar option for other operating system, I advise you to NOT check this box. Since the VPN is only required to have a TU IP address for the communication with our servers, other traffic should not be sent over the VPN which would be rather slow!

¹⁶<https://www.dropbox.com/>

¹⁷<http://itunes.apple.com/us/app/dropbox/id327630330?mt=8>

¹⁸<http://itunes.apple.com/us/app/goodreader-for-ipad/id363448914?mt=8>

¹⁹<http://wunderkit.com/>

²⁰<http://www.6wunderkinder.com/>

6.5 BigCluster

In order to perform computation, you will most probably get an account on one of our machines, called *BigCluster*. The *BigCluster* is a blade server (HP ProLiant DL385 G7) with the following rough specs (cf. Figure 5):

- IP: *130.83.163.152*
- Username: *your username*
- Password: *your password*
- Requirement: IP address from TK's pool (e.g., using TK VPN)
- Processors: *(x2): AMD Opteron Modell 6174 (12-Core, 2,2 GHz, 12 MB L3) each*
- RAM: *16GB + 96GB (8 x 2GB + 12 x 8GB)*
- HDD: *3x500 GB SCSI (Controller: Smart Array P410i). RAID 5*
- Optical drive: *DVD/CD-RW*
- OS: *Debian Squeeze*



Figure 5: htop on *BigCluster*

Most of the time, multiple students and members of our group are performing computations on the machine. We allocate resources on demand. After talking to the others, I will simply tell you how many CPUs and how much RAM you can use.

6.6 IM

Often, it makes a lot of sense to be able to quickly communicate using one of the many instant messaging tools out there. Therefore, please add me in one of the following IM clients:

- Jabber: *benjamin.schiller@jabber.ccc.de*
- ICQ: *104659547*
- Skype: *badtaste232*

7 Emails & filenames

Whenever you write an email regarding your thesis, please use a subject in the form as shown in equation 1.

$$\text{\$subject } (\{B.Sc.|M.Sc.|Sem.\} \text{\$thesisTitle}) \quad (1)$$

When attaching files to an email, like, e.g., a current version of your thesis, please name them in the form as shown in 2.

$$YY-MM-DD\text{\$filename} \quad (2)$$

8 Required actions / information

The following list contains everything you need to prepare before starting your thesis:

- Signup for an account on Github / Mendeley / Wunderlist
- Create a private (!) group in Mendeley and invite me to it (use *benni@codingz.de* for the invitation!)

The following list contains all information I need from you before starting your thesis:

- Name
- Email
- Student ID (Matrikelnummer)
- Course of studies (Studiengang)
- Current term (Fachsemester)
- Study and examination regulations (Prüfungsordnung)
- Github / Mendeley / Wunderkit account name (or email address used for registration)
- IM account name(s)

9 Always keep in mind...

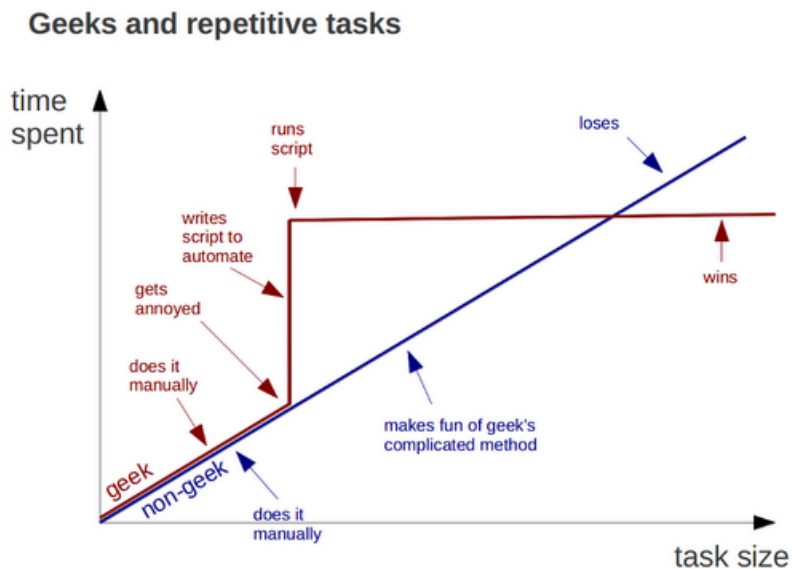


Figure 6: Sometimes it is good to be a geek