Guide to Starting a PhD

Robert Feldt
Gothenburg, Sweden
robert.feldt@gmail.com

January 12, 2012

1 Preface

This is my very personal view of what research is, how it should be carried out and how you should start your journey to become a great researcher. Others will likely have other views and they might have merit. However, this is my guide so I will summarize my views. Anyhow, you must have this in mind when reading and maybe take some things with a grain of salt. I will continuously try to improve the text as I learn more. Hopefully, some day you might not need so much salt after all... :)

I try to keep the writing style fairly informal so this is not a good 'model' for how you should write your research papers. I will refer to other and more formal papers and models as appropriate.

Note that there is quite a lot of material in here that has been collected during many years of experience. It might seem very complicated or as too much to handle. It is not. My intention is not that you need to learn all this from day one. But by telling you important things from the start you get a sense for it. We can later come back to specific areas and detail and develop them more.

This text is of course heavily influenced by my own background and the academic areas that I focus on. I will try to keep the text quite general but some sections will be specific to Software Engineering.

A guide like this can never be complete since there is almost an infinite number of things you need to learn before becoming a true 'master'. So there will definitely be important things that are missing or lacking in detail. Gradually, things will improve.

If you mail me about this document please state the date given above and/or the version number below so that I know which version you refer to. Thanks and remember that research should first and foremost be fun. The rest is details... :)

2 Introduction

Research is one of the most rewarding activities a human can take part in. Every day you have the chance to improve knowledge and find solutions to important problems and thus help make the world a better place. It is also

rewarding on a personal level since you get to learn and improve to become the best you can be. It is a great opportunity for you to develop both professionally and as a human.

A PhD project has the goal of developing you into an independent researcher. This guide tries to explain the most important information about a PhD project so that you can get the best start possible. However, in the end how good a researcher you become and thus, in a sense, how much better you can help make the world is *up to you*. Nothing you read or is told can really change the fact that you will need to mobilize your best knowledge and skills and to work very hard for a long time to really succeed.

3 What is good research?

What is good research depends on your subject area but it is always about *changing something for the better*. It can be to change our current understanding about something or to change the way in which companies or people carry out a task. When the focus is on changing understanding and improving knowledge the research is often called academic. When the focus is on improving practices the research is often called applied. Applied research done at a company is often called industrial. In practice there are no clear limits between these types of research and you do a little bit of each. Or you change between them during your project and between different studies.[1]

Good research should also be *novel*. Thus you are expected to describe what is the current state-of-the-art (SoA) or state-of-the-practice (SoP) and describe how your work is different from them.

Good research should be *relevant*. Novelty in itself is not really enough. There are simply too many problems out there and you can find something to change or improve almost everywhere you turn. Even though in the long-term almost any improved knowledge is important and will lead to something good you should try to be relevant. You can achieve this by considering what problems we face today or are likely to face within some time span. It is hard to only 'sit in your room' and guess what is relevant. You need to talk to other researchers and practitioners and learn what their problems are and what they consider important.

If you focus too much on a specific situation, company or case your results might be relevant but lack *generality*. An attribute of good research is that you are aiming for and discussing the generality of your findings. On the other hand, the general knowledge you build might need several cases, and thus reports, before the generality can be seen and discussed in a useful way. So it can be ok to have a few specific cases to build your case for generality on. However, it is not ok to skip the question about generality and just put together a string of specific cases. Another way to express this is that you should aim to uncover *principles* in your subject area:

[1] Finding the right balance between research and development/engineering can be a challenge when you are an industrial PhD student; you need time to build up (academic) knowledge and choose targets, but the company want to see results. In the rest of this text we focus primarily on non-industrial/academic PhD students; however most of what we say is relevant for both types.

Definition 3.1. A *principle* is a fundamental, primary, or general law or truth from which others are derived.

Good research is often *systematic* and *structured*. The former means that you have a clear idea of what to do and that this will clearly 'cover' the most likely relevant aspects. The latter means that there is a good 'logic' and 'flow' in what you are trying to do and how you describe it. A systematic analysis of threats to the validity of your results goes through all relevant and thinkable threats, one at a time, and analyze if they affected your study and what you did to mitigate each one. A structured paper has a clear connection from one section to the next, what you did and how you present it has a logical and ordered sequence. A systematic study can be presented in an unstructured way. A structured paper can report on an unsystematic study. Good research is both systematic and structured.

A key element of good research is that you *claim something* and that your claims are *validated*. Unvalidated claims can give inspiration to others and point to interesting areas that should be investigated. However, people will typically not be interested in unvalidated claims unless you are a world-class authority in your scientific field. When you are a distinguished professor you might sometimes be asked to 'give the big picture' or 'forecast' the future of your field, but even then very rarely. So you should always strive to have as strong evaluation and validation of your results and claims as you possibly can. Often the strength of a research paper is judged on how much and how strong validation you have done. To publish in the top venues you need to have both theoretically sound and strong and empirically supported results.

3.1 Can't I just do as the rest are/have been doing?

Software Engineering is a maturing field but, as in many scientific fields, there is still a lot of publications that are not very good. Maybe they lack validation, focus on problems that are not relevant, present only minor variations on already existing results or validate their claims only on toy problems? The aspects we discussed above sets the bar high. But why should you strive to have a higher standard? Why not do as many others and cut corners?

Some reasons are:

- If you really want to create new knowledge that is TRUE there are no shortcuts; if you take them you are only pretending to be a scientist/researcher, you cannot be/become one
- As the field matures the standards will rise and you better prepare as early as possible or your papers will be rejected
- Most importantly though: Why do sloppy work? Why not strive to be the best and do the best you can? Why diminish yourself and your talent/capability?

4 What you must do all the time

You must always be *writing a research paper*. Without writing papers your research is not making progress. Yes, there are many different activities involved

in creating good research, you need to read the existing literature, think hard, design the study etc. However, without writing things down they do not exist. They are good intentions and ideas but they are not tangible. Before things are tangible you cannot discuss them, weigh them against each other, identify problems or get any feedback. You must be writing all the time. Without it you will fail.

It is not ok to do the research and then write things up afterwards. It has been tested. I have tested it. It does not work. You will forget important details, you will loose good ideas and 'leads' and you cannot get feedback. Peer-review (see the Glossary in section 18) is the key quality evaluation of research, i.e. getting feedback from other experienced researchers. To ensure your submitted papers have a high quality you need to do your own, internal peer review in several steps. First you, then your supervisor(s), then maybe the examiner and other colleagues, then maybe the actual reviewers. To get this circle of quality enhancement started you must write things down. If you insist that 'you work differently' and that 'you can write things up in the end' I will not listen. You are allowed to do this when you have proven yourself, when you mature as an independent researcher. Until that time you write all the time.

5 What is the goal

The overall goal is to turn you into an independent researcher of top, international quality. *Researcher* means that you know and have developed high-quality skills in all the major activities involved in being a good researcher. *Independent* means that you can create and drive good research on your own. During your project you should surpass your supervisor(s) as an expert in your specific, subject area. No one will be happier than your supervisor, if you also surpass him/her in other aspects; since everything you do in your academic career will also shine some light on him/her... :) *Top, international quality* means that we strive to compete with the best. This has typically meant that you should be as good as the PhDs graduating from the top American universities. With a globalized world good quality can come from anywhere. In practice it is judged by how well your research measures up to the papers published in the best publication venues, relevant for your research area.

6 What you eventually need to learn

In order to become a top, independent researcher there are a number of areas where you need good knowledge, skills and attitudes. Below is just a brief (and incomplete) overview of the many aspects/concepts you will have to become proficient with (and don't worry, you will only start with a few, and not many researchers ever can master them all):

- *Searching (Databases, Search strategies)*
- *Reading (Critically, Focused, Skimming)*
- *Topic/Area (Identification, State-of-the-Art, State-of-the-Practice)*
- *Reviewing (Relevance, Novelty, Height)*

- *Writing (Language, Flow, References, Tools, Summarizing, Telling story, Reader-focused)*
- *Planning & Designing research (Brainstorming, Design of Experiments, Scoping)*
- *Thinking (Abstracting, Generalizing, Patterns/Connecting, Synthesizing, Critical, Formalizing)*
- *Methodology (Quantitative, Qualitative, Validity threats)*
- *Analysis (Statistics, Arguing)*
- *Organizing (Structuring, Eliminating, Time management)*
- *Communicating (Naming, Simplifying, Clarifying, Visualizing, Presenting, Selling)*
- *Submitting (Attribution, Rejection, Comments, Rejoinder)*
- *Collaborating (Networking, Contacts, Compromising, Adapting)*
- *Industry (NDA, Contract, Lawyers, Inertia, Legacy)*
- *Ethics (Confidentiality, Anonymity)*
- *Supervising (Focusing, Mentoring, Hinting, Cheering, Criticizing)*
- *Funding (Applications, Agencies)*
- *Venue Organization (Call for Papers, PR, Program Committee, Speaking, Questioning, Discussing)*
- *Personal (Confidence, Vision, Initiative, Passion, Balance)*
- *Politics (Academic organization, Stakeholders, Adm/support, Meetings, Decisions)*

For some of these concepts I have written specific guides and advice. Please check my home page for more information (or just ask me! :)).

7 What I Expect of You

- That you continuously write on scientific, research papers
- That we have continuous, at least weekly, contact
- That you are on time to planned meetings and activities
- That you inform me beforehand if you need to reschedule meetings
- That you give me 3-5 days to give feedback to your texts
- That you come to me when you have problems or are in doubt

- That you will gradually improve and make progress. It is ok if you are weaker in some areas but you will need to show that you take it seriously and try to improve. Over time there should be a clear progress. For example, I will not find it amusing to give the same comments on language, logic or study design again and again.
- That you drive your project forward, not wait for me to tell you what to do

The last point is very important; it is your project not mine. You must have and take the initiative. My role is primarily to give advice, support and help you out. Not dictate what you need to do.

8 What You Should Expect from Me

- That I continuously set aside time and give energy and show interest in you and your research project
- That we have continuous, at least weekly contact
- That I am on time to planned meetings and activities
- That I inform you beforehand if I need to reschedule meetings
- That I give relevant and detailed feedback and explain what I mean to you if it is hard to understand
- That I encourage you and give pointers and advice that can improve your research

Since I typically travel a lot I definitely will have to reschedule meetings, perform them over Skype or telephone etc. I have not found this to be a major problem. Once you have gotten started on your project it is more important that we talk every week and discuss things over email and telephone, than that we actually meet. Even so, we should of course still try to have meetings physically, as often as possible.

8.1 How I give feedback

The way I see it there are three main things that a supervisor can (explicitly or implicitly) give you feedback on: your results, your methods and yourself. For me the feedback and support I give you needs to be very different in these three areas.

For your end results, the papers, reports etc., I must at some point be ruthlessly honest and tell you what are its strengths but also its weaknesses. If you would not get negative feedback in a timely fashion you loose time since someone else is sure to give you this feedback at a later point and you will have to adapt and change accordingly. The bar when trying to improve knowledge is set high; this is not easy work! If we want to produce new research results of the highest quality we must be ruthless in finding, understanding and pointing out problems. This way we can improve the work at hand but also avoid similar mistakes in the future.

When it comes to the methods you use and activities you do to produce your results the feedback looks somewhat different. Early on I might take a more active part and direct you more in selecting methods, designing questionnaires, developing or extending tools etc. As time goes on the feedback is more iterative. You propose a design and we iterate until we are both happy. I might propose an alternative as I see fit. Overall the feedback is more iterative and supportive; we need to find a good balance between many different choices.

When it comes to yourself my feedback should be uplifting and positive. My sometimes harsh criticism can and should be directed at the results and possibly to you not having carried out an activity according to how we have planned it. But not to you as a person or to your capabilities. I very strongly believe that most people will respond positively if they are given possibilities, support and encouragement. So I strive to give that kind of feedback and to create that kind of atmosphere and 'environment' for you.

Of course I will not always succeed in giving the right feedback, keeping the types of feedback clearly separate and even if I think I do you might not interpret that in that way. A good 'middle way' is that we both try to keep these types clearly in mind, remember that there is a difference between them, and remind each other when we think feedback of the wrong type is given to/for the wrong thing.

9 What My Strengths and Weaknesses Are

I think it is good for you to know more about me. This way you can maybe better predict what I am doing and not and my reasons (if/when there are any! ;)).

I am an idea person. I love knowledge from multiple fields and I can often see connections between them. I associate easily between areas. This makes me what is often called 'creative'. I'm very good at seeing possibilities and start things up.

On the other hand, I can be quite 'lazy'. Since I have a high capacity and know that I can get things together quickly even if I'm late I tend to wait too long to get going. This can lower my quality somewhat compared to the levels I could achieve if I worked longer and in more iterations on things. I am actively working on changing this and I am making progress.

I am not very good at resolving conflicts. I generally try to stay away from 'power games' and 'political struggles'. This does not mean that I will not point out when I think you are going in the wrong direction or are not working enough. However, in general I'm more of a believer in cooperation and understanding rather than in criticism or avoidance.

When I started out in research I had a scholarship and a lot of drive. My 'seniors' left me mostly on my own and I had to 're-discover' things. I didn't understand the importance of the publication system and the focus on journals. I lost a few years of career progress because of this. However, I am picking up my pace fairly quickly now when I know better. I will make sure you do not repeat my mistakes. ;)

10 Where You Should Start

Ok, this has been a lot of information in a fairly short time span. What should you do now?

Assuming that you are starting your PhD project the most important thing is that you start writing. Together with your supervisors you should decide on a manageable first research task. The goal should be that you within 4 months have at least a 10 page, scientific paper on a topic relevant for your research project. It is very important that you get this quick start. More important than that the result is novel and relevant.

Planning your full project and deciding on courses are secondary to the task of getting started with the main activity; learning and writing about your area in a way that advances it. Also by writing early and getting feedback from me we can better understand if you have any particular weaknesses in your writing that we need to work on. This will seriously help you in your future project and career. Delaying it is a risk to your success.

10.1 Start project

To actually get going you should start right away on a start project. The type of start project can vary but it should involve writing a 10-15 page paper in a scientific article template within 4 months of starting. Our first task when starting you up is to decide on a start project. These are the main types:

- Turning your master thesis into a paper (MThesisPaper)
- Extending your master thesis project with more experiments and/or data collection so that it is publishable (MThesisExtend)
- Writing a literature or systematic review of your research area (LitRev)
- Taking a 'low-hanging fruit' (i.e. simpler and sure to give publishable result) of research from your project (MiniStudy)
- For industry research: Deepen your understanding of the company by doing an assessment or analysis of their current state (Assessment)

The first two are mainly relevant if your master thesis project was relevant for your PhD thesis project. If it was not these start projects can help train your skills but are less attractive since they would not help you towards your PhD. Also a LitRev has the advantage that you will need to learn the area in depth anyway and a good review of the previous work will have to be included in your first papers and your licentiate thesis anyway. The MiniStudy option is good since you will quickly get started on actually doing your own research. The downside is that it will typically not be possible to finish any part of your research project in just 3-4 months. If you think you can do it and you can argue for why then it is a good start; otherwise go for the LitRev.

An Assessment can be a good way to get started with industry research. You quickly get to know your company and get some 'connections' to people working there. The downside is that it may not be publishable. However, you and your supervisor can often find a way to make it relevant to your future studies and to the research community.

If you go for the LitRev here are a few tips. You should:

- focus on a part of your subject area that is limited so that you can go deep in 40-150 papers
  - try different search terms etc in the major databases until you have 400-1500 hits for the title+abstract screening
  - go through your 400-1500 papers but just check the title and abstract; filter out the ones that are not relevant (you can also note which are the main non-related areas that come out in searches, if they are close to your area you might need to define more clearly what you are studying and what it is not)
  - the remaining papers are your primary 'catch', the papers that you will read in more detail, if they are much more than a 150 it is better you focus your scope further for this time and exclude more papers
  - while you read you should note which are the 'key' papers that others refer to (a paper is 'key' either if it was among the earliest, advanced understanding a lot or if it has been very influential, i.e. highly cited)
- create a taxonomy of the papers you find
  - note specific attributes/features/dimensions/variation points for the papers and approaches they describe as you read them
  - choose the 2-7 most prominent/important attributes (argue for your choice in your paper)
  - specify the levels/values for each chosen attribute (and describe them in your paper)
  - structure your description of the area and the papers you found according to your taxonomy
  - see [1] for an example
  - reasons for a taxonomy is that it gives you and others a deeper understanding of the field as well as make it clearer where much and less results have been produced, i.e. where there are gaps in our understanding (it can also create many citations since people can use your taxonomy to position themselves compared to prior art)
  - More on the reasoning behind taxonomies and classification can be found in [2] (especially in section 5.2)
- create mapping diagrams to give a visual overview of the chosen subject area
  - as introduced in [3] and exemplified in [4]
- have a detailed discussion of the different sub-areas you have identified, both the ones with much previous work and the ones where work is lacking (and discuss why you think this is the case)

11 When will you be finished?

There are no definite, quantitative rules for how much work should go into a Licentiate or PhD thesis. Basically your examiner and supervisor will consider you ready when you have shown a certain maturity as an independent researcher. However, to give you an indication a Licentiate thesis is typically made up of the contents of between 3-5 papers, most of which have been published. For a PhD thesis this would be 5-10 papers, some of which are typically papers that was also in the Licentiate. The actual number will depend on the quality of the papers and where they have been published as well as several other factors (examples are: degree of independence, degree of novelty/originality, uniqueness of your area, depth of analysis, difficulty of task (whether its empirical of theoretical) etc.).

12 The Best Software Engineering Journals

The best Software Engineering journals are typically considered to be:

- IEEE Transactions on Software Engineering (TSE) - the top one, often high requirements on validation, empirical support and statistical analysis
- ACM Transaction on SE and Methodology (TOSEM) - a little bit more technical and Computer Science-oriented than TSE, otherwise similar
- Requirements Engineering Journal (REJ) - requirements engineering only, have had a problem with long review times
- Empirical Software Engineering Journal (ESEJ) - strong and growing as the field matures, any SE with lots of empirical support
- Information Systems and Technology (IST) - general, covers both processes and softer aspects as well as more technical issues, tries to shorten review times to 3-4 months (and often succeeds)
- Journal of Systems and Software (JSS) - general, has had longer review times than IST
- Software Practice and Experience (SPE) - more technical than JSS and IST
- IEEE Software (IESW) - more 'popular' science focus but highly regarded and cited, reaches more 'practitioners'

For your specific sub-area other journals might be relevant. You need to discuss the details with your supervisor and your examiner. When you choose publication target you can look at different rankings. For example, the 'ISI Web of Science' can create ranking lists based on different criteria. The most used criteria is 'impact factor' which is the average number of citations for a paper published in the journal in a certain time period (often the last two years). It can be debated if this is a good criteria, but it is a fact that it is used.

Of growing importance in Software Engineering is the yearly ranking done by the JSS journal. They go through all papers published in six journals

relevant for 'System and Software Engineering' in five consecutive years. For 2009 the years studied were 2002-2006 so there is a three-year delay. For each paper published in the journals TSE, TOSEM, IESW, JSS, IST, ESEJ and SPE, they award:

- 1.0 points to the author of single-author papers
- 0.7 points to each author of papers with two authors
- 0.5 points to each author of papers with three authors
- 0.4 points to each author of papers with more than three authors

The top ranked researcher typically has around 10-15 points for a five year period, corresponding to 2-3 single-authored papers per year. The points are also summed up to give a ranking of research groups. The only Swedish research group in Software Engineering that have appeared on the ranking is Blekinge Inst of Tech, who was ranked 5^{th} in the 2010 ranking (and 11^{th} in the 2009 ranking). It is possible that more Swedish groups will rank high in future rankings since the development of SE has been strong in Sweden from 2004 onwards.

13 How to Choose Courses

There are four main questions you should ask yourself regarding which courses to choose (in decreasing order of importance):

- What knowledge must I have in order to do my research and finish my planned studies?
- What knowledge is relevant for my project (if not now then in the long-term)?
- What do I find interesting?
- What can be important for my future career?
- Which courses are available?

Often you have to make a trade-off between the 'ideal' answers and what is practical.

In general there are three types of courses you can take: PhD courses, Master courses and Reading courses.

PhD courses are specifically created for PhD students. Compared to Master courses they typically involve many fewer students and are more demanding. Often the teacher will assume you have read all the material and when you meet the focus is on discussions and actual work. You also typically do a project or write a report within the course. It is a good thing if the project can be set up so that it can be of value to your PhD project.

Master courses are 'normal' courses given at a university. You often must have higher grades and do additional work to get them accepted as courses towards your PhD.

Reading courses are individual courses set up for just you or possibly a few more PhD students. You often have a lot of flexibility in adapting these courses to your specific needs or your project. Often your supervisor(s) or examiner gives you tasks and you then report to them or discuss directly with them when you have finished different parts. You often need to 'drive' a reading course forward yourself, more than for the other types of courses.

It is generally not expected that you have too many Reading courses but the decision rests with your examiner. He has all the power to decide over which courses are ok to include in your PhD.

14 Psychological Aspects

14.1 It requires hard work

It is not possible to become a PhD without having stamina and really putting in a lot of hard work. The idea of a 'lonely genius' is quite false. Most successful researchers both had talent and worked very hard to develop it and actually get results. You have to be prepared at times to make sacrifices and really put in that little extra to submit on time, do the extra round of validation, add another section to better explain the approach etc.

14.2 It can be a sole quest

Many PhD projects can be 'social' in the sense that you work with other colleagues and students. However, doing a PhD is ultimately a sole quest in that you have the responsibility for your own project and results and your thesis should be yours and yours only. At times this can make you isolated in the sense that you have to work 'on your own' for some extent of time. To cope with it you should of course be sure to take time off, socialize with others during lunch time etc. However, there is no avoiding that at times you just have to accept the fact that doing a PhD can be quite a solitary activity.

14.3 You will face rejection

When you submit a paper to a conference or journal you will sometimes be rejected. This can be very tough on PhD students, especially new ones.

Valid criticism should be taken seriously. This is the cornerstone of science; through peer-review we increase quality and ensure a certain minimal level is reached. You should be thankful to reviewers that give serious and extensive feedback.

It is very rare that any criticism is unfounded; more commonly the paper is not clear enough so the reviewer has misunderstood what you intended to say. Thus, you should consider all feedback valid and consider how you can change the paper to avoid other readers making the same mistake.

You should know that there is an element of chance/lottery in any publication process. For the top journals and conferences it is less likely since they do very thorough reviewing with multiple reviewers and discussions. However, in many cases, you might simply be unlucky. The specific reviewer you got had a bad day and over-reacted to your paper. Or she did not consider your topic

important even if it fit the scope of the venue. Or the competition was very fierce; your paper was among the best of its category but there were as good papers who happened to get slightly better reviews.

There can be many reasons. What you should focus on is to go through the feedback you get and objectively put yourself in the 'mind' of the reviewer. Is there something to the comment? What lead the reviewer to see it this way? What could I have done differently to address the problem? By considering the feedback honestly and as objectively as possible you will also know how valid it is and what you need to do to address it. And then you press on by improving and sending to another suitable venue. Don't take it to personally; no system is perfect. We can try but there is always an element of chance. So focus on improving as many aspects you can; the higher the quality of your results the less likely you are to get rejected. And good work will always find its way to publication.

14.4 You will get stuck

Sometimes you will feel 'stuck', i.e. without knowing what is the next step or how to solve a particular problem. It is normal, don't take it personally.

There are multiple coping strategies for this problem. One is to take a break. Do something else for some time and see if the problem just disappears or if a solution just 'pops up' in your mind. This can be helpful since your subconciuos mind continues working on the problem 'behind the scenes'. Or gives you a new angle/perspective on the problem.

Another strategy is to explain your problem to someone else. When you 'spell it out' this can often help clarify things and possibly help solve at least some parts.

A third strategy is to do some physical exercise. This not only boosts your brain power in the long term but can also act as a kind of 'power break' and speed up your 'processing' of the problem.

Of course there are many more strategies and you will find your own. Experiment. But remember that you are not alone in this; all researchers go through this and quite often. See it as a natural part of the process.

14.5 You might fail

In any creative and innovative undertaking there is a risk that you might fail, that the thing will not work as you planned etc. This quote says it in a good way:

> 'As soon as you say 'failure is not an option,' you've just said 'innovation is not an option.' - Seth Godin, The Flip Side

However, there is a difference between temporary, intermittent failure and overall, final failure. Even if you fail in one step, if you continuously get up again and try in a different way you are very likely to succeed. What you cannot do is to be paralyzed by the fact that things might not work out the way you thought. Adapt to what you actually see, decide on a new course of action and move on.

14.6 Dare to stay in uncertainty

I have learnt that there is a very important thing one must do to be able to find good and creative solutions: Stay in uncertainty!

I have seen that many students and colleagues are in love with certainty and get really annoyed when things are unclear or unknown, i.e. uncertain. To be able to really 'turn all stones' and consider not only one or a few solutions you have to be able to 'sit' with the uncertainty. To really stay in that state where you have not yet decided which path to take, which alternative is the best or which design should be selected. If you rush out of the uncertain state you will miss good solutions. Instead you should train yourself to be able to stay there for days or even weeks at a time. So that you can be sure you have really done your best to penetrate the issue/problem at hand before you move on.

14.7 Reading and/or Doing

Reading the work of others and understanding the state-of-the-art is a key activity in research. However, it is hard to really learn something without doing. Don't get stuck forever in searching and reading. Sometimes you need to act to start a cycle of learning/judging/adapting/improving.

14.8 Why good research is not enough

Research should ideally only be about the results themselves, i.e. judged on the strength of the theoretical or empirical validation and support. However, you will learn that it is also about name and recognition and even about doing stuff that others can recognize as and think is important (vs. doing stuff that really is important!). This is natural since the value of something is unknown if it just sits in your drawer. Thus publication really is a very important part of doing good research and to be successful you will also have to market yourself and your research:

$Success = Making + Marketing$

where, in our case, the marketing is to publish and go to conferences and to talk about your results with others etc. For industry research it is crucial that you can market your -self/project/study to companies to even get access to their systems/data/people.

For more info see:

http://lateralaction.com/articles/who-you-are/

You can be an idealist and fight this system in silence but if you do you will ultimately be less successful as a researcher. Don't fight the system; play along while still being yourself or striving to be what you want to be.

15 Practicalities

It is not crucial which tools you use in your research but it is important. If you do not already have a tried and tested work-flow for producing research papers and having control over your files you can save time by following my recommendations.

Your papers should look sharp. This means that references should always be full and correct, formatting should be impeccable and figures should be clear and high-res. You do not want reviewers to get annoyed on any of this 'easy' stuff; they should focus on your content.

Microsoft Word is very widely used and for many uses it works well. For producing research papers it does not. You have to decide on an external tool to manage references and often Ms Word will not format things perfectly anyway. I use Word for funding applications and internal memos etc but not for any real research. I did use it up until my licentiate degree. I had so many problem with it at that point that I would not do it again. For writing research papers and reports I use Latex. For managing references I use its accompanying Bibtex.

In much of our work statistical analysis of data we gather is critical. The open-source *R* software for statistical computing is growing in importance and power each day. I suggest you learn it early and then stick to it during your project. When it comes to statistical methods, parametric methods are often not enough. You should check out non-parametric statistics and also robust statistical methods since we often have less data than other academic fields and data that is not normally distributed.

All your files and documents should be kept in a version control system[2]. This way you will never loose any files or their histories. The version control will also give you fine-grained backup. I have all my files, excluding video footage, music and photos, in a single git repository. I have tried most of the available version control systems and git is best. You should use git to; then we can share a repository for all files related to our research papers.

For producing figures I use OmniGraffle. It is a Mac program so you cannot use that if you are on a PC. However, you probably should be on a Mac also. I have tried almost all computers and OS's and found that Mac OS X is great since it gives you the power of a Unix (shell etc) with a very nice looking GUI. You will also have less problems with viruses. It is the tool of knowledgeable computer professionals, imho.

However, you are of course allowed to use any tools you are used to and feel are up to the task. I will not accept sub-par results in these areas though, since it will affect your chances of success, so pick wisely.

Naming of files is important. You have to think about the recipient. Don't call your paper 'paper.pdf' or something similarly non-descriptive. Don't use a name that is too common so that the file is lost among many others. Structure on your name first, then describe what it is shortly and possibly add a date marker. Underscores instead of spaces is maybe pushing it too far, but I am old school and want to respect people that are still using Unix or other systems that traditionally never played well with spaces in file names.

16 Writing a Research Paper

16.1 Who should be an author?

One of the main ways to assign credit in academia is through the list of authors on papers. It is thus very important to have the right authors credited on your papers. If you 'forget' someone this could lead to a conflict; at the very least

[2]And I mean *all* files, with large video files being the only exception

you will have to discuss with them and 'clear the air' afterwards, at the worst you have created a life-long, academic enemy.

The simple rule is that the people that have helped make a paper possible should be credited. And they should be credited in the order from larger contribution to smaller. Below we outline this in more detail; this is based on the rules and process for authorship that was developed by the seniors of the SERL group at BTH in October 2010[3].

A person should be one of the authors of a paper if they have made a substantial contribution to

- the *research* **OR**
- the *drafting* of the article, **AND** have
- *reviewed* the paper (for important intellectual content) **AND** have
- given a final *approval* of the version to be published.

Note that the structure is (Research **OR** Drafting) **AND** Review **AND** Approval. If a person has contributed to the research or the drafting of the paper they should be given the chance to be an author. They take that chance by reviewing and approving the paper before submission.

The contribution to the research depends on the type of paper. For an empirical study it is enough with a substantial contribution to the

- 1) conception/idea **OR**
- 2) design of study **OR**
- 3) acquisition of data **OR**
- 4) analysis **OR**
- 5) interpretation of data/results.

For other types of studies the actual steps might be different. Use your common sense to judge which are the steps of the research and what should constitute a substantial contribution. If you think that a person have reason to believe they have made such a contribution you should invite them as authors or discuss with them why you do not consider their contribution as substantial.

Academics must review and give comments on a paper to be authors; for industrial authors you might not expect as many or as rigorous comments. All authors listed on a paper should give their final approval to the paper before the paper is published.

[3]The following people should thus be credited for these rules: Claes Wohlin, Jürgen Börstler, Robert Feldt, Cigdem Gencel, Tony Gorschek, Ludwik Kuzniarz, Kai Petersen, Darja Smite, Mikael Svahnberg and Richard Torkar

16.2 Structure of a Research Paper

This is the classical structure of a research paper:

- **Title** - much more important than you think, should be as short and 'catchy' as possible while still representing the work in a good way.
- **Abstract** - a structured abstract is often good (informally even if not allowed).
- **Introduction** - must give some context and MUST point out a problem/misunderstanding/gap as well as motivate why it is important. Research questions often at the end of this section.
- **Background / Related work** - describe related previous work and describe how this study differs from them and adds to them.
- **Design/Methodology** - of study, context for company/products etc.
- **Results** - 'raw' results in tables, diagrams etc but also text to describe how to interpret the raw results. Should not really discuss what the results mean though just ensure the reader can understand them. Sometimes the Analysis is part of the results section though and then you can have more information here.
- **Analysis** - what the 'raw' results means/says.
- **Discussion** - what it means in a larger context and in relation to goals, limits and limitations, validity threats.
- **Conclusions** - nothing new, summarizes previous, similar to an extended abstract with more details on results and what they mean. Should be possible to read without having read the rest of the detailed sections.
- **References** - important, not too much books, no web/URL references, should have recent/relevant work and the 'key/classic' papers, always refer to your own work that is related to this study.

One thing to note is that there is a difference between two groups of sections. The first group includes the sections Title, Abstract, Introduction and Conclusions and can be called the 'outward-focused' group. This means that they should be written in a way that a reader that comes from the outside without having understood all aspects of the paper can still read and understand these sections. This is needed since many people search for papers based on their titles, filter them for relevance based on the abstract and then directly read the conclusions or read the conclusions after first having read the introduction. For example, this means that you should not refer to research questions in the conclusions since the reader might not yet have read what they are.

The second group of sections can be more 'inward-focused' and assume that the reader reads them more in a linear order.

You should also try to reference relevant papers from your research group. Citation scores are increasingly important and since your group often affects you, you should acknowledge their papers and contribution when possible.

16.3 Research questions

It is essential in any research study to have concrete questions that the study should answer: the research questions. Research questions are typically created very early in the life of a study, already at the conception or design stage. They are then stated clearly at the end of the introduction section of your paper and you make sure to answer all the questions in your discussion/analysis.

Research questions (RQs) should be *interesting*, *answerable*, *brief* and in a *logical order*.

Interesting means that they are questions that academics or industrial partners are actually interested in. Of course this has to do with both their content (i.e. they are about something relevant) but also about the actual formulation (how you write them). You can never 'save' a question that is not relevant by formulating it in a more interesting way.

Of course the questions all need to be answerable. Don't aim too high and try to solve too general problems in one go. It is often easier to get a paper published if it has a strong focus on particular questions; if your RQs are too general or long this is a sign your study do not yet have the right focus.

It is hard to say when research questions are 'exactly' right and at the right level of abstraction. But you should try to keep them brief without too much text. Sometimes you will have to have one or two main questions and then have more specific sub-questions 'under' the main ones. However, you do not want to make this too complex; try to keep to a straight list of 1-5 research questions (2-3 is the most common) without having to detail them too much. You can rather discuss additional detail in the surrounding text.

The order of research questions is important; things that you want to know the answer on before other questions should be numbered and listed before. They should thus make sense, taken as a whole.

To give an example of this let me show you some research questions that one student proposed[4] at an early stage of a research study and how I then helped him refine them. This was a study where different techniques for predicting SE quality (Q below) was compared both between themselves and to human experts doing the same predictions. The study also included interviews and a questionnaire to the engineers on the usefulness of this type of prediction and how the techniques can be made more useful to the engineers. The first set of questions was:

- RQ1. What is the prediction performance of different techniques in predicting FST for each test phase multiple weeks in advance by making use of data about project progress, testing progress and fault-inflow from multiple projects?
- RQ2. What is the usefulness of predicting FST for different test phases?
- RQ3. What criteria are important in making the prediction techniques usable?

The main problem here is actually about the content (RQ1 hides the most interesting/relevant question and has actually two RQs in one), but there is

[4]This is not a criticism to that student in any way but I need to give a real-world example :)

also too much information (in RQ1) and the order is not logical (rest of the RQs are meaningsless if answer to RQ2 is 'not useful'). So my solution was to move RQ2 up front and then subdivide RQ1 into its two sub-questions. The most interesting of those (for industry) I put before the other. Finally I took away some unnecessary information from RQ1 (test phases and multiple weeks are all important and qualifying information but can be in surrounding text IMHO) and made all of the questions more industrially connected and changed language to direct verbs to end up with the new list:

- RQ1. How useful is it to industrial software engineers to predict FST for different test phases?
- RQ2. Can machine learning techniques better predict FST than human experts?
- RQ3. How does different machine learning techniques compare in FST prediction performance?
- RQ4. What criteria are important in making prediction techniques usable in industrial practice?

To me this is a clear improvement.

17 Links to Learning More

- Mary Shaw on 'What makes good research in Software Engineering' and 'Writing good SE research papers'
- Find good 'model' papers. Learn from the best. Identify top scientists (may not need to be in your area) and study their papers. Study both language, structure, study design, data collection, discussion and conclusions. Different people have different strengths. Check for example Alan Kay, Richard Hamming, Claude Shannon, Fred Brooks, Gary King...
- 'How to write a good research paper', Simon Peyton Jones
- Donald Knuth's 'Mathematical Writing', where the first three sections gives a great introduction to technical writing
- Michael Ernst on 'Choosing a venue: conference of journal?'
- Alan Smith on 'The Task of the Refereee'
- 'How to give a good research talk', Simon Peyton Jones, John Launchbury, John Hughes, SIGPLAN Notices 28(11), Nov 1993
- Richard Hamming on 'You and your research'
- Dave Patterson on 'Having a bad career in academia'
- Gary King on 'Publication, Publication' which (in the latter parts) give concrete and good advice on how to write publishable papers

Just search to find them. Also check my home page where I link to relevant resources.

18 Glossary of Key Concepts

Study - A sub-project aimed at answering a specific set of research questions. Consists of at least one but often several individual papers.

Paper - A single scientific report/document. The default unit of work in research. The strength of your results is often judged on the number of papers and the strength of the venues where you have published them.

Publication - A paper that has been published. There are different types of publication venues. Until a paper has been published it is unpublished and have less of a scientific value.

Research Question - A question that a research study (or project) aims to answer. Each study often have one main research question and then between 0-5 sub-questions, whose answer would lead to an answer for the main one.

19 Acknowledgments

I want to thank Martin Ivarsson, Joakim Pernstal and Ali Shahrokni at Chalmers, Ehsan Ahmad at Air University and Richard Torkar at Blekinge Institute of Technology for comments on early versions of this text.

20 History of this document

- 2012-01-12, version 0.9.1: Added a note that one should not only investigate parametric statistical methods.
- 2012-01-12, version 0.9: Added section 14.5 describing that you might fail in parts of your project as well as section 14.6 on staying with uncertainty. Also added section 8.1 on how I (try) to give feedback. Also added a recommendation to use R for statical analysis.
- 2010-11-06, version 0.8: Added the sections on paper authorship and research questions. Also expanded somewhat on the description of the structure of a research paper.
- 2010-08-11, version 0.7.1: Finished the section on 'You will get stuck' which was incomplete
- 2010-03-02, version 0.7: Switched to memoir Latex class for more stylish and shorter PDF
- 2010-03-01, version 0.6.6: Minor fixes after Richard Torkar's comments
- 2010-02-16, version 0.6.5: Added example of what is systematic and structured research based on Ehsan's comments
- 2010-02-15, version 0.6.4: Added link to Ulf Lindqvist's thesis with more info on creating taxonomies

- 2010-02-10, version 0.6.3: Minor language fixes and added proper references to some of the papers
- 2010-01-21, version 0.6.2: Minor updates based on comments from Joakim Pernstal and Ali Shahrokni
- 2009-11-11, version 0.6.1: Minor updates based on comments from Martin Ivarsson
- 2009-11-10, version 0.6: Added overview list of some of the aspects to learn (not happy with it, I have better ways to structure this in old notes, find them!), file naming advice added in 'Practicalities'.
- 2009-06-29, version 0.5: Refined formulations and added the JSS ranking in 'The Best Software Eng...' and the text in 'What you should expect from me'.
- 2009-05-22, version 0.4: Added 'Practicalities'.
- 2009-05-11, version 0.3: Added text on start project and how to do the LitRev.
- 2009-05-07, version 0.2: Added 'Reading and/or Doing' and 'When will you be finished?'.
- 2009-05-06, version 0.1: Added an intro and goals for the guide and outlined some major sections. Wrote the 'What is Good Research', 'What is the goal', 'What you must do all the time', 'How to Choose Courses' and 'My Strengths and Weaknesses' sections. Started on others.

Bibliography

[1] S. Axelsson. Intrusion detection systems: A survey and taxonomy. Technical report, Dept. of Computer Engineering, Chalmers University, 2000.

[2] U. Lindqvist. *On the Fundamentals of Analysis and Detection of Computer Misuse*. PhD thesis, Dept. of Computer Engineering, Chalmers University of Technology (URL: http://www.csl.sri.com/people/lindqvist/ulphd.pdf), 1999.

[3] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. Systematic Mapping Studies in Software Engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering*, 2008.

[4] D. Šmite, C. Wohlin, T. Gorschek, and R. Feldt. Empirical evidence in global software engineering: a systematic review. *Empirical Software Engineering*, 15(1):91–118, February 2010.