

Writing About Research
Or, the Art of WAR

Ted Pedersen
September 2003

There are a number of good books available that talk about writing for research. I recommend that you find one that you like, and that you read it carefully and follow it's advice (let me know whose advice you are following before you get too deeply into this). One book I like fairly well is called "The Craft of Research", by Wayne C. Booth, et. al. If you aren't able to find a guidebook on your own, I recommend this one to you.

The important thing about your writing is that it be clear. The particular style or organization is up to you, but it must end up making sense. We presume that as graduate students you are able to write in a grammatical and organized way. If this is not the case then you probably should not have been admitted to the program, so you can either work to correct these problems or resign in disgrace.

Here is my most basic and important rule. When you are writing, do not refer to any other books, articles, or papers. Do not, ***under any circumstances***, read other people's papers while you write. You will almost certainly plagiarize. If you can't write about a topic without looking at published sources or descriptions, then you don't understand it well enough to have any business writing about it. You will of course need notes of your own experimental results and algorithms. These should be ideas and materials that are original to you. Any notes derived from other sources should be done with the utmost care to avoid plagiarizing. The "Craft of Research" has some ideas about how to do this that I suggest you follow. I have some extended examples of plagiarism found here: <http://www.d.umn.edu/~tpederse/Pubs/plag.htm> Please make sure you read over this very carefully, and let me know if there is anything at all that is unclear.

My best advice as to writing is to *write within yourself*. Do not try and sound like me, do not try and sound like anyone but yourself. Write what you know and what you understand and tell me what you have learned and what you think. Do so clearly and concisely. Be formal in your writing, but not excessively so. Introduce notation only when you must, and make sure that it is easy to understand and consistent. Define terms and ideas as they are introduced. Theses should be relatively short, so don't repeat yourself. Say things once, say them well.

How do you know if your writing is any good? Does it pass the reading aloud test. Can you read your paper to someone and have them follow it relatively easily without having a paper copy in front of them? If so, your paper is probably pretty well organized. You have probably started with more general ideas, and then gotten more specific. You have probably defined terms as you introduced them. You probably have not written things in such a way that you force the reader to turn back and forth in your paper, or have them make notes to themselves as they read in order to understand your presentation. You must strive to build a clean and coherent representation of your ideas in the mind of your

reader. Bad organization, poor grammar, and spelling errors chip away very quickly at that model, and ultimately make it impossible for the reader to construct an elegant structure in their mind regarding your ideas. Instead, they may be left with a tiny broken down shack with no running water, and this is not a pleasant thing to have in your mind.

Your reference list should be honest. In other words, cite only those books, papers, and articles that you have actually read (or at least skimmed) and that have actually provided insight into your work, and given you ideas. Don't pad your reference lists. Any reference mentioned in a paper or your thesis will be assumed to be known to you, so you may expect questions on why you cited it and what it contains during your defense. I may also ask you to produce copies of any reference you cite, so please make sure that you keep track of your references. In general, books, journals, and conference papers are good sources of information. Workshop papers, book reviews, technical reports, and material on web pages are not usually good. There are exceptions to this of course, but as a rule of thumb this is good.

For our discipline, the journal "Computational Linguistics" is the premiere forum for published research. Other reliable journals include "Machine Translation" and the "Journal of Natural Language Engineering". Conferences that usually contain good information include the annual meetings of the Association for Computational Linguistics, (ACL, NAACL, and EACL), the biennial International Conference on Computational Linguistics (COLING), and the annual conference on Empirical Methods in Natural Language Processing (EMNLP). All papers appearing in ACL related events are available at the ACL/LDC Repository (<http://acl.ldc.upenn.edu>) You can find out all about ACL related conferences at the ACL web site (<http://www.aclweb.org>).

Conferences on Artificial Intelligence (AAAI and IJCAI), machine learning (ICML), or data mining (KDD) also contain high quality publications on NLP or closely related issues. The Journal of Artificial Intelligence Research sometimes has NLP related material that is quite useful (<http://www.jair.org/>). We are also starting to see NLP papers creeping into computational biology and bioinformatics. These are less likely to serve as references in your work, but nonetheless it is interesting to see how these ideas are being applied in a rather new area.

Writing Introductions, esp. for theses and thesis proposals

Again, I think the book *The Craft of Research* may be helpful, in that it goes into a bit of detail about writing introductions. One of their suggestions is that you might want to start with an outline. (They make the point though that one should not be a slave to your outline, or spend a great deal of time coming up with a very formal outline.) This is entirely appropriate, and might help to structure things. As I've said before, your thesis will be a short enough document where there is no need to repeat much of anything, so careful organization is important.

Your introduction should be exactly that - an introduction to your thesis. Write it so that someone who knows very little about our area can understand what you are doing. It's

important to remember that you are not writing this for me - the introduction is for that committee member who may have little background in this area, or for your fellow students (not in the nlp group) who might be interested.

One of my thoughts when writing an introduction is that I would like my mother or father to be able to understand it. This is actually a rather nice goal because your mother and/or father would probably enjoy reading about your thesis topic, and while they won't want to read about your algorithms and so forth, a general overview that explains what problem you have tried to solve, why it really matters that anyone solve this problem, and how you went about doing this might be rather satisfying for them.

In a more pragmatic sense, consider the introduction as an executive summary that you could give to a potential employer so as to explain to them what it is you did your thesis on. Imagine taking the introduction, printing it up separately and distributing it attached to your resume. I am not suggesting that you do this of course, but this is meant to give you an idea of what the introduction should achieve. The important point is that it should be relatively intuitive and it should be self-contained.

So, the introduction to your thesis (or your thesis proposal) must have the following goals:

1) Describe the problem you are solving.

This must be very intuitive and written in an engaging way that will make the reader interested in what you are doing. What's the problem? Be specific, use examples, make the examples interesting and compelling. (I am using the term "solve" quite loosely here of course.)

2) Explain why you want to solve this problem.

This is where you motivate that the problem is worth solving, and you can describe the potential impact of your work. Imagine that you completely solve your problem. How is the world a different and better place? What can we do now that we couldn't before your thesis?

3) Explain the approach you take to solving this problem.

This is not the place for algorithms, instead simple examples are best. You should explain the general ideas that underly your approach and make you believe that it is sound. Your mission here is to convince the reader that your approach is sensible and reasonable.

4) Explain how you know that you solved it (evaluate).

A thesis claims to make some contribution. Here you tell us how you know that you did what you claim. This section will be hard to write as of now, but you can at least

summarize how you are planning to evaluate your solution, and how you will know that you have made progress.

5) A formal statement of your thesis (ie the thesis statement).

What hypothesis underlies your research? What is the question that drives your research? What question are you seeking to answer? This must be specific and it is the one part of the introduction that should be technical to a degree. The question should be on that is interesting regardless of the outcome. For instance, "Can I implement the algorithm of Hastings?" is a horrible thesis statement/hypothesis because the answer is utterly dull in both the affirmative and the negative. I believe that your Craft of Research book should be of some help in thinking about thesis statements. You should come up with a thesis statement that reflects your understanding of your research now. I will find that interesting, and we can refine it as we go along.

The above is not intended to serve as an outline. You must organize your introduction in a style that suits you. These are just items that you must make sure you address.

The introduction should be unique text. In other words, don't cut and paste text from the interior of the thesis in the introduction, and vice versa. You are writing the introduction for the novice who is trying to decide if they really want to read about your research. Make it engaging, exciting, and indeed entertaining. Then, when they get to the interior sections of the thesis you can hit them with the details. They will want them at that point. They do not want them in the introduction.

The introduction is probably the hardest part of the thesis to write, it is also, along with your the conclusions that you draw from your resarch, the most important.

I urge you to refer to books like the Craft of Research to get ideas on how to do this. You are also welcome to look at other theses and dissertations to see how they organize. Keep track of the outside sources you are looking at in terms of getting ideas for writing - I would like to know what they are, especially if you find them helpful. There are quite a few books about writing and technical writing that you can draw upon as well. I understand you probably have not written like this before, so you should seek out as much help as you can from these kinds of external sources.

How to Do Research

This is a tricky question, and we'll work on this throughout your time here. However, I firmly believe there is a connection between how you think about your writing and how you think about your research. (That seems obvious now that I say it). What I mean is that when you think about your research, you should think about how you are going to write about it to make it compelling, interesting, and important. If you can't think of any way to do that, it might be that the question you are researching is not terribly interesting, or you don't understand it very well yet.

Another book that I like very much and find somewhat inspiring (really) is called “Advice For a Young Investigator” by Santiago Ramon y Cajal. The author is often called the founder of neuroscience (1852-1934) and he talks about the challenges of doing research, and gives some ideas for how to think about it (and how to do it). The UMD library has this available for electronic checkout.

Technical Issues

Your thesis should be written using LaTeX, a Unix/Linux word processing system. You should start to get used to LaTeX now by using it for your writing. You should find a latex source file from a previous thesis and use that as a guideline for your own work. We have such examples/templates available in /home/cs/tpederse/mypublic. A good Latex book is: LaTeX: A Documentation Preparation System User's Guide and Reference Manual by [Leslie Lamport](#). You can find quite a bit of information on TeX and LaTeX at <http://www.tug.org/> and <http://www.latex-project.org/>