

Discussion

On Ph.D. thesis proposals in computing science

H. C. Lauer

*Computing Laboratory, The University, Claremont Tower, Claremont Road,
Newcastle upon Tyne NE1 7RU*

A Ph.D. candidate in Computing Science at Newcastle typically comes to us with some knowledge of programming and a clear indication of high ability. But his specific background in computing may range from a broad appreciation of some of the fundamental problems of the science to a total ignorance of others; and he perhaps may have some specialised experience in some area of interest. We educate him to a level of expertise worthy of the title 'Doctor' by providing an environment in which he can learn, teach and do research and by demanding of him a thesis representing an original contribution to the science. The actual character of this educational program is, necessarily, tailored to the individual or to small groups of individuals with closely related interests. In one model for such a program the student spends the first part of his candidacy—the whole candidacy normally taking about three years, as in most British universities—working on small projects, attending lectures and doing reading to broaden his knowledge and to fill gaps in his background, and exploring the science for topics which interest him. During this time, he develops close working relationships with one or more members of staff who, in turn, agree to become his supervisors. With their help, and the help of visitors, his own colleagues and others, the candidate eventually narrows his sights to a particular area of the science as a potential source of research problems. He hones his skills to the point at which he can do original work in that area and finally defines a problem which he believes he can solve and which is suitable for presentation as a thesis.

It is at this point in his career that he ought to be able to present a *thesis proposal*. This article is concerned with the character and content of such proposals, and it concentrates on this important period of a research student's life. Obviously, the necessity or desirability of this kind of thesis proposal in different Ph.D. programs and/or different sciences is a matter for debate; but that discussion is beyond the scope of this paper. Instead, we concentrate on what we expect of the proposal and on six vital points it should address.

What is a thesis proposal?

A thesis proposal should represent a considerable effort, perhaps several months of very intensive, full-time work. It should lay the ground work for the thesis research by providing convincing arguments that the problem is worth solving and can be solved. It allows the candidate to 'stake out a claim' in a potentially crowded area. It provides a good yardstick against which the candidate can measure his own progress or lack of it, and it helps him to focus his energy when he feels he is waffling. It provides extremely useful evidence of achievement if he needs to seek additional financial support when his grant expires. Finally, it helps him to combat the common occupational affliction of Ph.D. students, namely depression.

The timing of a thesis proposal is important. For a three-year research program, it should be presented during the second year. If it is done much earlier, it is likely that the problem will not have been well-enough defined or that the candidate will not have done enough background work and/or made enough progress in the area to convince himself and others that he can solve it. If the proposal comes much later, then either there is too little time to do the work before the money runs out or it is a spurious proposal produced after the fact, when the thesis is nearly done.

The form of a thesis proposal is a matter of individual taste of the candidate, his supervisors, and the university. It may be written down in one document, presented orally in seminar, evolved by mutual agreement, or done in some other fashion. It may include research memoranda and/or published articles by the candidate (or coauthored by him). Some parts of it may be eventually included directly in the thesis. The different sections of the proposal may be done in any order, depending upon how the thesis topic was developed. But it is important that it be 'public' at least within the department, so that everyone can know what the candidate is investigating and why.

A thesis proposal in computing science should address at least the following six points:

1. A statement of the problem and why it should be solved
2. Reference to and comments upon relevant work by others on the same or similar problems
3. The candidate's ideas and insights for solving the problem and any preliminary results he may have obtained
4. A statement or characterisation of what kind of solution is being sought
5. A plan of action for the remainder of the research; and
6. A rough outline of the thesis itself.

If the candidate is unable to include and defend these six points in his thesis proposal—or indeed, if he cannot defend them at the corresponding stage in his career even if he does not prepare this kind of thesis proposal—then he is not ready to commit himself to the one or two years of blood, sweat and tears to turn it into an acceptable thesis.

Naturally, *neither his supervisor, nor the university, nor his examiners are going to hold him to the details presented in the proposal*. The nature of research in this science is that it provides the biggest surprises to those who are most strongly convinced of some fact or idea. When a lot of people are working in a given area at a lot of universities, anyone can be easily 'scooped' or may feel it necessary to revise his plan or problem in mid-stream. He may find that his original ideas do not work and he must modify his expected solution. This is perfectly acceptable, and the plan of research will have to be adapted to fit. Nevertheless, a candidate who is unable to answer the six points is not ready to embark on the work, let alone follow it, control it, adapt it and force it to some kind of conclusion.

Let us consider these points in turn.

Problem statement and background

The first obvious thing which a thesis proposal should contain is a statement of the problem to be considered, in both specific and general terms. The specific statement must deal with the specific issues in which the candidate is interested—for example, the optimisation of tables of LALR parsers. The general statement should relate the problem to the larger context of the science and show why it is worth solving. The problem statement in the thesis proposal should be directed to an audience of intelligent scientists who have no specific interest in the problem but who are interested in knowing what the candidate is doing. It should *not* be directed to the candidate's supervisors and/or to people with similar research interests.

To prepare the proposal for their benefit is to make a common

mistake. Such a proposal is filled with jargon which is private to that local group. It fails to state important constraints and frequently does not provide enough background. Sometimes the candidate assumes that his supervisors know as much about the specific area of the thesis as he does—something which is often false. Proposals which suffer these faults lack credibility, and it is difficult for the department and the examiners to evaluate the research on its merits. The candidate is then exposed to the real danger that he and supervisors may have been working happily in their own microcosm, only to find that at the end of three years he has no results which justify a Ph.D. degree.

To present the problem to the wider audience, and to justify proceeding with the work, it is necessary for the candidate to present the background to the problem and to survey related work by others. This is the second component of a thesis proposal; and in some cases, it may be included directly in the thesis. It may take any of several forms—for example, an annotated bibliography or a comprehensive summary, explanation, and analysis of existing results. It may be necessary or desirable for the candidate to include his own critical comments. For example, if the thesis is to present a new technique for solving a class of numerical problems, then this section of the proposal should review existing techniques and analyse their inadequacies.

This summary/survey/overview is not without its traps. If most of the references cited and most of the work mentioned are from in the candidate's own department (or in one other department with whom he is very 'chummy'), then there are serious grounds for questioning his breadth of knowledge and background for pursuing his problem. The danger is that people who limit their horizons to their own local environments produce very inbred research, narrow attitudes, and unacceptable theses. They tend to reinvent ideas already known elsewhere; they fail to apply techniques which could simplify their problems considerably; they often attach too much importance to minor results and do not recognise major ones worth reporting; and they write incomprehensible theses and papers which make no effective contribution to knowledge. In inbred environments, the work of other organisations is often dismissed as irrelevant or unimportant—characteristic of a disease called NIH (Not Invented Here). It is extremely important for the thesis proposal to indicate that the candidate knows about and accepts such work.

The candidate's ideas

It is hard enough to schedule 'invention' when one has some good ideas for solving a problem. It is almost impossible when he does not. Thus the Ph.D. student, who is working to a tight and very emotionally constraining time-table, needs to have some insight, some ideas, some preliminary results before he commits himself to discover more. These should be described in the third section of the thesis proposal. If he has none of significance, then his proposal is premature. For he would have no indication that the problem can capture his attention for as long as it takes to solve it and write the thesis. He would have no assurance that he is heading in the right direction, that he is capable of finding a solution.

By implication, then, the candidate must have done some successful work in the area, perhaps in collaboration with others, before the thesis proposal. This may be something like the discovery of an interesting algorithm, representation, or relation while working on one of his pre-thesis projects. He recognises this as the tip of the iceberg, the introduction to a new problem area which eventually becomes his thesis research. For example, a student simulating a well-known paging algorithm stumbles across a phenomenon quite different from that which was expected or generally accepted. This result and his subsequent explanation for it form the basis of his thesis

proposal and thesis research in memory management. They form the seed of the methods which he develops to specify and solve his problem. Without such results, a plan to investigate the area would have seemed like hot air, and his efforts would have lacked direction. But with them, the success of his research is assured and the timely completion of his thesis is much more likely.

A common situation occurs when a student proposes what seems to be a good problem to investigate, involving new broad, general models or theories. But when he is pressed, he has only some ideas about a small, special case or example. He might not even have explored these ideas fully because he regards that example as uninteresting in the context of the overall problem and those ideas as having no apparent generalisation. Some students will be able to discover the necessary general ideas, develop them and defend them. But such theses are few and far between, and their authors are typically awarded Nobel prizes and other very high distinctions. Ordinary mortals with good first class Honours degrees have no such luck and often get stuck, unable to find any other examples, applications or ideas which are substantially different from the ones they know already.

At this point, it is time to go back and look at the problem statement again. As often as not, that 'uninteresting' example may be the foundation for an interesting and valuable thesis problem in its own right. If so, it is probably a better investment of the candidate's energy to solve it, finish his thesis, and then devote his life's work to the general problem in a more relaxed fashion.

The shape of the solution

The most important part of the thesis proposal is a statement of what kind of solution to the problem is expected—i.e. a characterisation of the stopping condition of the project. This, more than anything else, will help the candidate estimate the value of his efforts to separate the chaff from the wheat, to allocate his time. Without such a characterisation, the candidate has no good way of knowing when to stop and submit. He cannot measure how far towards his goal of a Ph.D. degree he has progressed. He might even discover a satisfactory solution to his problem and not perceive that he has. With a characterisation, he will know where he stands during his research, and he will be able to argue convincingly at the appropriate time that he has done what he set out to do.

Occasionally, a research student will say 'I know precisely what problem I want to solve. I have no idea of what the solution will be, but I will certainly recognise it when I've got it. After all, this is research. So how can I possibly give a characterisation of the solution beforehand?' That is, he thinks he is an exception, but if he cannot characterise his expected solution, how can he recognise it? More likely, he has not specified his problem sufficiently precisely, or he has not yet done enough preliminary work and obtained some preliminary results in the area of the problem. In either case, he must do more legwork before presenting his thesis proposal.

Sometimes it is easy to characterise the solution, particularly in the light of preliminary results. For example, a candidate developing a new analytic model to describe message traffic among communicating machines would expect to prove some theorems about the model, validate it empirically against some existing systems, construct some algorithms based on it for calculating the performance of similar systems with different parameters, and argue by example that they are useful in the design and understanding of future systems. At other times, it is much harder to be so specific about a stopping condition. It may also be necessary to change it as the research progresses. However, a moving target is better than no target at all (providing that it is not moving so fast that the candidate cannot catch it).

Plan of action and outline of the thesis

The last two points which a thesis proposal should address are almost, but not quite, afterthoughts. After the candidate knows what he wants to do, has some background to allow him to do it, has done a little bit, and has some idea where it will take him, he had better draw up a plan of action. This section of the thesis proposal is like a road map and timetable of how he will travel during the remainder of his research. If it is carefully and realistically prepared, it will expose to him any hazard of trying to do more than he reasonably can before he runs out of steam. Obviously this plan, like everything else in the proposal, is subject to change as new results are obtained and new ideas gained. But some plan is better than no plan.

Finally, it is always useful when doing research to keep in mind how it is to be reported, what issues will be emphasised, and what will be de-emphasised. Thus, the thesis proposal should contain a rough outline of the thesis itself, preferably in terms of the expected solution to the problem. This will have at least a small impact on the shape of the research, and it will provide a set of good guidelines when the candidate decides that it is time to 'write it all up'.

The thesis itself

It is almost impossible to define what a Ph.D. thesis in Computing Science ought to be. Neither can we characterise the differences between an acceptable one and an unacceptable one. No one can present the candidate with a prescription for success when he embarks on his studies. We cannot predict who among the entering research students will succeed, who will lose interest and drift away, who will work hard for three years at what they perceive to be genuine research, only to leave in great bitterness after discovering that they have nothing to present in theses. There are no formulas which tell us how to conduct research in this science, what steps to take, what things to avoid. The same road can lead to progress and results for one

person and to disaster for another.

It follows that the thesis proposal as we have described it is not a guarantee of anything and may not always be appropriate. But it helps, particularly when the problem, the investigation and the expected results are ill-defined. By considering his research in terms of the guidelines we have presented, the candidate, and his supervisors, will go a long way toward developing the sensitivity and awareness necessary to make the research lead to a successful thesis. It is an effort not to be undertaken lightly.

Note and acknowledgement

In this note, I have attempted to set down some personal ideas about Ph.D. thesis proposals, what I think they ought to be, and what I feel they ought to contain. These ideas have evolved from my own experience in doing a thesis, from observation of colleagues during my post-graduate days, from supervising Ph.D. students here at Newcastle, from analysing why some apparently brilliant students never finish, and from dozens of conversations with my students, colleagues, teachers and friends. I have come to expect and demand that my own research students use the guidelines which I have outlined here when they define their thesis topics and prepare their proposals. When other students and colleagues seek my comments and advice about thesis topics and projects, I ask the same questions and apply the same criteria. I offer these thoughts to you for what they are worth—whether you be student or teacher—in the hope that you, your supervisors, and/or students will derive at least some small benefit from them.

I must acknowledge my deepest debt to Professors Brian Randell, William Lynch and Bernard Galler, who have taught me enough to be able to recognise a good thesis topic when I see one and to be able to head off at least a few bad ones before the student gets too committed.