Organizing Your Research and Developing Your Research Skills

Kelly A. Shaw
kshaw@richmond.edu

Working on research can be a lot of fun, even exhilarating when you first start on a project. There are so many new things to learn in the area you are working on and so many new techniques and experiments to try. It can also be overwhelming as you realize how many research possibilities exist and how much information there is for you to learn or how many approaches there are for you to master. Trying to do everything all at once during the flurry of new project excitement can burn you out or depress you as you realize that you will never know everything.

After this initial flush of activity, it is important to step back and recognize that research is more like a marathon than a sprint (although there can be sprint-like periods). The process of doing research is evolutionary, where one new piece of information helps lead to another over time. Progress on your research will unfold gradually as long as you persist in working on it. Similarly, your research skills and knowledge will develop over time as you are gradually exposed to more research results and research approaches and as you create and try your ideas.

This article is geared towards helping you condition yourself to do research over the long haul - more than just a few months. It addresses how to be organized in your research, how to become a more efficient researcher, and how to develop your research skills. (If you are just getting started on research, consider reading "Getting Started on Undergraduate Research" for suggestions on how to begin.)

Organizing Your Research

As you delve into research, you will discover that the volume of data and existing related work is too much to keep track of without some organization. Additionally, the complexity and usefulness of your experimental infrastructure may demand more than one individual being involved in its maintenance and modification. Acquiring some simple habits can help you manage the large amount of research materials you acquire and help you share resources.

Preventing Wheel Reinvention

Research progress proceeds over long periods of time and may include the development of experimental infrastructure as well as the creation and collection of results. A single individual does not work constantly on all of the different components of their research, making it likely they will forget the particulars of what certain components do, how results were acquired, or even where items are located. As individuals share experimental infrastructure concurrently and over time, it becomes difficult for all of these users to keep track of all of this information. Consequently, an important aspect of doing research is recognizing that other researchers will use this infrastructure. Doing your best to make that utilization easy will prevent individuals from wasting time regenerating results or functionality or from needlessly searching for materials or information.

One way to keep track of where research materials are located and to enable resource sharing is to create a central repository with a well-organized hierarchical structure. The organization will help everyone know where to find specific items. If the materials are stored on a computer, using a revision control system like *cvs* or *svn* to create this repository can be helpful. Revision control systems allow users to work on personal copies of the repository when appropriate and then propagate their changes to other users by copying their work into the central repository. These systems keep track of older versions of items in the repository, which reduces the likelihood of losing work accidentally. They also allow changes to the central repository to be reverted if those changes cause problems.

Another habit to acquire is documenting everything you produce. Experimental infrastructure and results can take a long time to construct and can be extremely complicated. What is obvious to you during creation may not be obvious to you months later or to another researcher. Consequently, whenever you create some form of tool or result, you should document it. If you write code, you should thoroughly comment that code. If you create a proof or algorithm, you should accompany it with a detailed explanation of the different steps and the reasoning for each step. Be explicit and err on the side of more description and explanation! You should also create a README explaining where to find the different components of your work as well as what those components are and how they are expected to be used.

Insuring Correctness

Whenever you create some sort of solution (e.g. algorithm, proof, software), you need to test your solution for correctness. Each test should be as simple as possible and still test the desired condition. Simple tests allow you to check the actual correctness of your results, not just whether the results are in the right ballpark. You should test your solution in both the common or expected cases as well as the anomalous or unexpected cases. This should include testing at boundary conditions.

When working on solutions that will be refined over long periods of time, it becomes extremely important to test your solution after every modification. Changes may work well for new scenarios but cause failures on previously tested scenarios. Consequently, it is essential solutions be tested on all scenarios after each change. Creating a test suite, in which you continually add new tests (but rarely remove older tests), can insure you quickly discover when modifications violate earlier correctness.

Validating the Relative Benefits of Your Approach

To convince researchers about the benefits of your approach, you need to compare your solution to existing solutions and show how and when your solution is preferable. This requires generation of results that can be directly compared to results presented in related work. When you reach the point of collecting results, you want to determine how to generate results for this comparison and build that capability into your research materials.

During your initial reading of papers, you probably read about existing approaches. The methodology details in those papers provide you with the information (e.g. configuration of test environment or inputs) you need to recreate the test scenarios for your comparison. You want to test your solution using the same test setups used by these researchers. If you are lucky, there will be consensus about the test scenarios used in the related work; if not, choose scenarios used by solutions considered to be the "best" solutions in the existing research.

The environment that your solution is targeting may be sufficiently different from the testing environments used in related work. If this is the case, you may need to implement the solutions proposed in the related work and collect results for those solutions in this different environment. It is important to implement these existing solutions as faithfully as possible. If aspects of the approaches are not well specified, determine what the optimistic and pessimistic approaches would be. You can then either acquire results for both cases or gather results that would be considered optimistic. By comparing your work to an optimistic version of existing research, you insure that your solution will look that much more impressive.

Keeping Track of Related Work

When you first started your research project, you probably read a number of papers, learning about the different issues and solutions proposed by other researchers. As you continue doing research, you will

continue to read papers, learning more about this subject and related topics. If you plan to present your research results (e.g. poster, paper, or presentation), you will need to be able to explain how your research fits into the context of related work, citing specific papers to support your argument. A little effort can help you organize your thoughts about how different research approaches (including your approach) fit into the larger problem space.

After reading a research paper, it is useful to reflect on the paper's contributions, weaknesses, and future work and how the paper relates to your research. Because you will want to remember and potentially revisit your insight into this work, you should create a written reflection or critique that you can refer to later. You should keep these critiques in a single location, such as a research notebook, to prevent them being misplaced.

It is also a good idea to keep an ongoing bibliography of papers that you have read in order to make it easy for you to cite related work in your final presentation. You want to keep information that would be used for a citation (i.e. authors, title, publication, publication date) as well as a summary of the paper. If you have written a critique of the paper, indicate the location (e.g. page number or date in a research notebook) of the critique in your bibliography notes so you can easily refer back to your full critique. Computer scientists frequently store these citations in *BibTeX* in order to easily reference them from papers written using *LaTeX*.

Improving your Research Efficiency

Because research is open-ended, it can be easy to get distracted from your primary research goal by interesting tangents. Additionally, indecision and the absence of well-defined goals and deadlines can impede students from making measurable progress. Adopting habits that add structure to your activities and taking advantage of feedback opportunities can help you overcome these challenges and improve your research efficiency.

Learning from Others

As you do research, you are going to have questions, run into problems, or not know how to proceed. When should you ask for help? You do not want to spend hours searching for an answer that someone can provide in a short email, but you also do not want to ask for help if you can quickly find the answer yourself. Similarly, you do not want to tie yourself up in mental knots and not make progress because you are unsure which approach to take. The answer depends on your distinguishing when asking for help will be better in terms of your research efficiency and skill development than finding answers on your own.

Questions usually relate to either specific details or to high-level design. Many detail-oriented questions should be easy to answer independently. If you are using a tool, searching the online documentation or looking through code comments may provide answers to many questions. You can also search the online archive of the tool's user mailing list or enter the text of the tool's error message into an internet search engine. Similarly, questions about how to do something in a programming language or the standard way of using a mathematical or statistical technique can be explored via textbooks or online resources. When these approaches fail, consider asking research group members and then your research adviser. However, if a colleague can immediately and effortlessly provide an answer versus your taking hours to find the answer, you should ask the colleague.

When you are unsure which path to pursue, you are dealing with high-level design questions. You want to discuss these concerns with other researchers, including research group members and your research adviser. For example, they may be questions about missing pieces of an algorithm or protocol or

questions about how two system components function together in real systems. Or, they may be questions about how to design experimental tests. You should take the time to think about the problem and its possible solutions, including advantages and disadvantages of each approach. This prepares you for explaining your ideas to researchers who have limited knowledge about your specific approach, but nonetheless can provide insightful suggestions. When a researcher responds with a solution, ask them to explain their reasoning so that you can understand their thought process. This will help develop your skills at creating your own solutions.

Finally, you should not be scared of discussing your research with other researchers. It is an important aspect of the research process. Other people can help you see problems and brainstorm solutions. Do not wait until you have a perfect solution or perfect results! As soon as you have analyzed your results, you should share them with other researchers to see if they agree with your observations and conclusions and whether they have any other feedback. They may point out a flaw in your algorithm or an exciting new implication for this work that you had not realized.

Creating goals and deadlines

Some students struggle to make significant research progress because research frequently does not have hard deadlines. To avoid this pitfall, you need to create your own research deadlines and treat them as equally as important and firm as course deadlines. One way of creating deadlines is to divide tasks into a hierarchy of goals, each with its own deadline. For example, you may want to obtain some result by the end of the semester. That goal can be decomposed into subcomponents of work that must be completed to reach the larger goal. For each subcomponent, you want to determine how long it should take to complete it. You then want to break these subcomponents into finer grained tasks and set deadlines for these smaller tasks. Your objective is to create sets of tasks where each can be completed in approximately a week. You can then decompose those tasks into smaller tasks that can be completed in a day or two.

The hardest aspect of this process is judging how long each task will take. In order to accomplish a given task, you may need to learn about something else (e.g. an existing tool or technique) or you may need to correct a mistake that you only discover when tackling this new work. Additionally, you may overestimate how much time you have to work on a problem in a given week or how quickly you work. With experience, you will improve your ability to judge the amount of time required of each task. Until then, be conservative in your estimates! You do not want to have the feeling of never meeting expectations.

Low productivity may also be the cause of missed deadlines. One way to evaluate your productiveness is to keep a log of how you spend time dedicated to working on research. Keeping track of time in 15 minute increments will enable you to see if you are getting distracted from your research. It can also help you identify whether or not you are spending time learning non-pertinent information. Just because you are working does not mean the work you are doing is helping you make research progress; you need to stay focused on tasks that move your research forward.

Tracking Your Progress

One pitfall related to setting many small deadlines is that it can be easy to lose track of the big picture. Failing to remember what you are trying to achieve at a high level can easily erode motivation and productivity. Thus, in addition to making research progress, you need to keep track of your research progress towards those high-level goals. Doing so will keep you motivated and focused on completing tasks that actually need to be accomplished. Additionally, if you have kept track of your activities, the locations of different pieces of work or results, and the observations and conclusions from your work in an orderly manner, it will be relatively easy to construct a final presentation about your research.

Keeping a research notebook
A traditional way of tracking research progress is use of a research notebook. A research notebook may be a physical notebook or it may be a computer file or a wiki or a blog. You should choose whatever medium works for you.

Research notebooks contain items related to your research with each entry being dated. The content of entries will vary widely depending on what you are working on. An entry may contain information you have learned in order to move forward on a problem. It may include your write-up of related research or the particulars of some tool or technique. Depending on the research area, it may include proof details or information about experiments such as the experimental setup and results. Those results may be accompanied with analysis and any related intuition, insight, or questions.   Another type of entry may specify precise steps that need to be performed next and any information that you need to proceed.

The research notebook can also be a great place to flesh out ideas or concerns. For example, you might write an argument for your approach to your research or about the assumptions you are making. This can help you clarify why you are pursuing a specific research question in a certain way and why it is better than alternative approaches. You can also brainstorm complementary or alternate approaches for future work. Writing these thoughts down may help you think of issues you have not considered previously. It can be useful to occasionally review entries in your notebook to remind yourself of the insights you have had.

While a research notebook is good for tracking daily progress, you might find it difficult to refer back to specific information quickly. For example, you may have accumulated notes about how to use a tool over several months' worth of entries. At some point, you may want to consider consolidating those notes into a single document. Or, you may want to create an index to write-ups of related work written in your research notebook. Over time, you will determine how best to organize and use your research notebook, but it is probably best to initially write everything related to your research in your notebook.

Creating weekly reports and research meetings
At the end of every week, you should review the progress you made and how it contributes or moves your research forward with respect to your high-level goals. You should also write down the next steps for your research. It may be useful to do this review in the form of specific questions that you answer. William J. Dally at Stanford University asks his research students to send him weekly email answering the following questions:
1. What were your goals for last week?
2. What did you accomplish?
3. What problems did you encounter, how did you deal with them, and what did you learn?
4. What are your goals for next week?
5. What are your longer-term (month, quarter, year, until graduation, 10 years, life) goals?

The answers need not be long. The point of keeping a written log of your answers is to keep you motivated and to make sure your short-term goals are aligned with your long-term research (and life) goals.

This write-up can also help provide you with a starting point for discussion with your research adviser about your progress. If your discussion with your adviser alters your next steps, you should create a revised version of your write-up immediately after your meeting and potentially send that updated version to your adviser to make sure both of you are clear on your plans.

Incrementally creating a research summary

Another way to stay motivated is to create a summary of your research that you gradually add to and refine as your research progresses. For example, create a write-up either in the form of a paper or presentation slides. Common components include problem, related work, insight, proposed solution, experimental framework, experiments and what constitutes success, results, future work, and conclusions.

Before you actually start doing experiments, you should write summaries for the first four or five sections. As you collect results, you can add them into this framework. Also, as your understanding of your research matures, you should update previously written sections with your enhanced understanding. Revisiting these sections will have the added benefit of reminding you of your high-level goals.

Over the course of your research experience, you should share these write-ups with other researchers and your adviser so they can give you feedback that improves your understanding and your presentation skills. At the end of your research experience, you will then have a completed presentation.

Further Developing Research Skills

When you begin doing research, you may do things because your research adviser suggested you do them, and you may not understand the rationale behind the suggestions. It may seem like your research adviser magically knows what to recommend next.

Doing research well is not magic. Research is a process that can be learned like any other process, such as piano playing. It takes years of lessons and practice to become a skilled pianist, and it involves becoming good at a variety of sub-component tasks. All of this is true for research too. You get better at doing research with practice, so you should not stop doing research if your first attempt does not turn out perfectly. Consequently, you should dedicate a small amount of time each week to enhancing your knowledge of doing research in your area or developing other useful skills.

Taking time to learn about the process can make your research process more understandable. One way to do this is to read books that explain various aspects of the research process. For example, David J. Lilja's <u>Measuring Computer Performance: A Practitioner's Guide</u> describes approaches for doing computer systems research. A complementary way of learning about the process is to be cognizant of the approaches and techniques used to accomplish different tasks in the papers you read about related work.

For example, you may notice that the methodologies used across different papers vary, allowing researchers to explore different questions and solutions. The type of infrastructure (i.e. tools) used to collect data or perform experiments may enable different types of questions to be asked. Different workloads or inputs may be chosen to reflect the needs of different application domains. Or, different mathematical techniques may enable certain approaches to mathematical work.

Similarly, it can be instructive to note the set of experiments used and try to understand the reason for those choices. For example, what inputs or parameters were changed for each experiment and why? If results were sampled, what is the rationale for doing so and how was the sampling conducted? If inputs were randomly generated, what distributions were used to generate the inputs? In addition to the primary metric used to evaluate a research idea, what supporting results were collected and did they support the authors' argument?

It can also be useful to learn about techniques used for evaluating and presenting results. For example, what statistics were used to evaluate the results? Was a more complicated analysis technique such as clustering used? What is clustering and when is it appropriate to use? How were data presented in graphs? What was the benefit of showing data in three dimensions versus two?

Knowing about the many tools and approaches available can make it easier to understand future papers you read. More importantly, this knowledge can help you decide which, if any, of the approaches you have previously seen are appropriate to use in your research. You may realize you lack some background knowledge or skill that would improve your research effectiveness. For example, you may need to improve your software development skills or learn statistics to improve your ability to analyze results. If gaining proficiency in an area will make it easier or faster for your research to progress, you should dedicate a small portion of time every week to learning that topic. By observing how other researchers perform research, you can discover how to develop your skills to become a stronger and more productive researcher.

Final thoughts

Research can be enjoyable for both the student and the research adviser, especially when progress is being made. Once you have moved beyond the initial phase of starting a new project, becoming organized in the allocation of your time will make your effort more productive. This paper has presented some suggestions on how to add order to a process that can initially seem very fluid and poorly defined. While you need not incorporate all of the ideas at once, consider incorporating one or two that would help with areas you consider weaker in your research experience.

References

Shaw, K. A., "Getting Started in Undergraduate Research," in *IEEE Potentials*, vol. 32, issue 3, pp. 9-17, June 2013.

Lilja, D., Measuring Computer Performance: A Practitioner's Guide, Cambridge University Press, 2005.

Testing sidebar

When you create some solution or infrastructure on which to perform experiments, you need to insure that it is as correct as you can make it. Depending on the type of research, insuring correctness in your setting may be a very different process than what is required by researchers in a different setting.

For example, if you are performing experiments in a physical lab setting, you will want to make sure that your equipment is producing the expected results for known test cases, such as for a calibration standard or a known chemical compound. That way, you will know that all of the individual pieces of equipment are functioning correctly and at the level of accuracy needed for you to collect reliable results. Your tests could also include calibrating equipment under a variety of conditions, perhaps checking the full range of large and small signals over which it will be used, and perhaps accounting for environmental variables like accidental bumps or changes in ambient temperature.

However, if you are using software to model the different components of a particular system, the required testing process will be quite different. For example, if you have created software to simulate a specific chip architecture, you likely will have created software to model the processors, caches, interconnection network, and memory. Each of those individual components needs to be tested for correctness. Additionally, the functioning of those individual components together to model the entire system must be tested. For example, basic testing of the cache may involve creating a small sequence of read and write data accesses which is sent to the cache. Because the sequence of requests is small, you can determine what exactly should happen in the cache on every data access and verify that the cache is acting correctly on every data access. You would want to try different streams of requests to insure the cache works correctly for common data access patterns as well as for pathological data access patterns. Knowing that these simple cases work will build your confidence in your cache simulator. By connecting this cache to one other cache via an interconnection network, you can create a small scale system that enables you to hand verify the correctness of your cache coherence protocol. You can then initiate larger tests with more caches where you cannot track each individual data access, but you can collect aggregate information and confirm that information agrees with the understood behavior of cache coherent systems. The key component being that you test for exact correctness on a small scale and then look for adherence to well-established expectations for larger, more complex tests.

Testing the correctness of an algorithm may require an entirely different approach to the two mentioned above. The approach taken to testing may include proving the complexity of the algorithm. It may also include evaluating the correctness of your approach for different sets of assumptions. It may be the case that your algorithm was originally created to solve a very specific problem that occurs only under very specific conditions. While you may not intend to use your algorithm under any other circumstances, it is still important to evaluate whether the algorithm continues to work well under broader conditions or completely separate conditions. Specifically, you will likely need to prove what the boundaries are for which your algorithm is guaranteed to be both sound and complete. This understanding will enable you or other researchers to apply your approach as appropriate to future problems or stop using it when it is inappropriate for a given situation.

<u>Methodology sidebar</u>

Methodology is a crucial piece of any research and needs to be well thought out and well documented in order to convince readers of the validity of your results. However, appropriate methodology depends completely on the specific type of research being performed and is, therefore, hard to generalize.

For example, in a physical lab setting, there may be several ways to collect data about some phenomenon. Of the existing approaches, one may be more applicable if you are researching a particular metric or result than the other approaches. Or, one technique may be more expensive (in terms of time or money) than is necessary to collect the data needed to support your hypothesis. Or, one approach may not collect results at the level of accuracy needed by your research. All of these considerations, in addition to availability of lab equipment, may influence what questions you can ask in your research and which techniques you use to conduct your research.

In other research areas, software tools may be used to conduct research instead of physical equipment. For example, in computer systems work, researchers frequently create software to model the system under study (i.e. simulators) and then use data sets or software applications as inputs to that system (i.e workloads). What software you choose to model the system has a huge impact on what types of research you can pursue (e.g. chip multiprocessors versus graphics processors) and how accurately you can model the components of that system (e.g cycle accurate network simulation). Additionally, the inputs to the system define what types of problems you are tackling as some workloads are geared for graphics domains while others are geared for database and other commercial workloads.

For more theoretical work, the methodology may include which type of mathematical approaches are considered valid for modeling different environments or problems. For example, if you choose to use a queuing theoretic approach to analyzing a problem, you need to choose the appropriate model (e.g. open, closed, or mixed) for the system you are researching. Similar to computer systems work, you may also have to choose inputs for your model (e.g. arrival and service distributions) that best approximate the activity in the system you are trying to analyze. Additionally, the models initially chosen may be simple but less realistic. However, as you delve deeper into your research and your questions become more precise, it may be necessary to start using more complex and realistic models.