Reading Literature for Research

This is a personal view on reading research papers with the purpose of getting started with some research area. It seems to work for some of my students - feel free to use it to get started in your own reading, but know that you may have to come up with your own techniques in addition to these.

- 0. **Choosing "Seed Paper".** Very often, we have a rough idea of the area in which we want to (or are being asked to) do research in, and very often, we have at least one research paper in mind that epitomizes this research area. Not necessarily the first paper in this area, not necessarily the latest or the best, just a good one. We shall call this your "seed paper". Sometimes the seed paper will be assigned to you by an advisor or a superior.

- 1. **Focusing Literature Survey.** For this paper, do the following:

    - a. Get familiar with some Bibliographical indexing tool - I wholeheartedly suggest BibTeX, which accompanies LaTeX. Start the bibliography of your research project by including the full citation for your "seed paper."

    - b. Understand well the basic design problem that the paper addresses. Briefly describe the problem (in half a page or less).

    - c. State the research problem in one sentence. To do this, you may have to generalize/broaden the problem somewhat, so that you are describing a narrow problem area, rather than a specific problem.

    - d. Find the most prestigious journal in your research area, for my area this is probably IEEE/ACM ToN. Make a list of all the papers published in that journal over the past one year.

      For each paper, answer the question "Does this paper address the same research question (part c) as the seed paper?" For most papers on your list, you will only need to read the title, or a few sentences in the abstract or at most the entire abstract, before ruling them out. For papers which seem as if they might be related to the seed paper, you should delve deeper to decide. You will have to read beyond the abstract for these - you should definitely read the introduction section, and the context or background section is there is one. You may have to read the problem definition or problem formulation section, but try to avoid it. Under no circumstances should you read beyond the problem formulation section. At this time, you must have a definite "yes" or "no" answer for each of the papers on your list. For each paper you answer "no" to, enter them in a separate "reject" bibliography, and keep a one sentence description of your reason for so answering. For each "yes" paper (if any), enter them in your research bibliography.

      Note: As a guideline to the effort you should be expending at this stage, I suggest you should not spend more than 2 to 5 minutes on considering any individual paper before deciding your answer.

    - e. If you answered "no" to every paper, consider revisiting your answer to part (c) to see if you were too specific with the research question; however, if you have good reasons for each "no" answer then you are done. If you answered "yes" to 5 or more papers on your shortlist, you have definitely made the research question too general - go back to part (c), make the question more specific, and repeat.

- 2. **Expanding Literature Survey.** Not everything gets published in one top journal in the field. Realistically one should be aware of papers published in several journals and conference proceedings. In order to do this, we cannot follow the model used above (starting from the set of all papers and narrowing down), because with several good conferences and journals over several years, we are talking of several thousand papers in the original list. In this case, we have to expand our list of papers rather selectively.

    - a. Identify two or three keywords or keyphrases that describe the research question you are homing in on. The seed paper and the other papers on your current list may actually provide keywords, which might help. Conduct a search on the INSPEC database (available through the D H Hill Library's website) with combinations of these keywords and keyphrases. Search over approximately the last 5 years of the database.

    - b. Repeat your keyword searches independently on the IEEE Xplore database and the ACM Digital Library database, and report if you turn up any papers that your search on INSPEC did not turn up.

    - c. Repeat part (d) of Step 1 with the papers that you turned up. At this time, you should have between 5 and 20 papers that you have classified "yes". We shall call this your "core list".

    - d. Read the bibliographies of each paper on your core list, and make a list of papers that also address the core research topic, and appear in the bibliographies of more than one of the core list papers. Add these papers to your core list.

    - e. By this time, from the list of authors of the papers on your core list, you have a good idea of who the most active or influential researchers in your research area are. Visit their websites - they will usually have a list of their own publications. If you find any paper that addresses the core research topic and that you have not found so far, add them to your core list.

    - f. So far we have searched backwards. Now we have to try forward searching, which is more difficult. Fortunately, it is much less difficult now than it was even 10 or 15 years ago. There are two fairly good public citation search engines - Citeseer and Google Scholar. Identify one or two papers on your core list that have the maximum citations within the core list. Conduct a search on Citeseer and Google Scholar to find other papers that refer to one or more of these papers, and that you have not already encountered before. For each such new paper, repeat part (d) of Step 1. If you answer "yes" to any of these papers, add them to the core list.

    - g. Enter all the papers on your core list into your bibliography.

      Note: Retain all the intermediate lists you have made, and save all the keyword searches you performed.

- 3. **Reading and Remembering.** Now you have to read all the papers in your core list. There is no shortcut and no way around; fortunately it gets much easier after the first couple of papers in any research area. I have only a couple of suggestions on reading.

    - a. Research papers and presentations, textbooks, almost all archived material is presented sequentially, but this is not how the human mind works in understanding, learning, or creating. When reading, you have to jump back and forth a little bit. But if you keep

referring back, you may never make much progress. One technique that helps is to underline or highlight the key things, such as definitions, notations, formulae, as you come to them, because these are things you are likely to want to refer back to from later in the paper. Another is divide-and-conquer; which in this context means if you do not understand some part of it on repeated reading, then leave it behind for the moment. Just make sure you understand what the consequence of that part is for the rest of the paper, then take the author's word for it, and move on. Re-visit it later; it might be much easier to understand in the light of the understanding of later parts of the paper, or a different paper.

- b. Every time you read a paper, annotate its bibliography entry to build up an annotated bibliography. An annotated bibliography is a regular bibliography (list of cited references), with a descriptive short note (one or two paragraphs, not more than one page) written by you documenting your understanding of this paper. It is hard to believe after you just spent hours reading and understanding every detail in some paper, but you will forget all about that paper in a matter of months or even weeks. An annotated bibliography saves you from having to undertake the entire effort again - usually reading the annotation is enough, or at most a few minutes worth of the actual paper in addition. More details are available at various sources, for example, D. H. Hill Library has a webpage on annotated bibliographies. Also see the **Testing Your Comprehension** section below.

- c. Re-run all the literature searches that you made to come up with your core list of papers. The intent is to become aware of research work on your topic that may be have been published since you started work on it. A secondary goal is to tune your literature survey to your research topic, if that has taken a turn.

Testing Your Comprehension

(Note: For those interested: The questions below were obtained by consideration of the "Anatomy of a Research Paper" presented above, and Bloom's Taxonomy of the cognitive domain.)

When we take the effort to read a paper, we want to gain knowledge from the exercise. To test whether you have gotten your effort's worth from reading a paper, you can ask yourself the following questions - you should be able to answer these at satisfying level of details. If your answer is very general, and does not satisfy you, then you have some more reading to do. Perhaps you need to read other background material first, or just invest more effort in reading the paper itself. At times, going on to read other related papers can help, because you may understand something when it is explained by some other author in a slightly different way, or something else, perhaps a conversation with a colleague, might explain it for you. But be sure to keep track of what you have "skipped over" like this in any paper, so that you always know what you have and have not mastered in a given paper. (The annotated bibliography mentioned above is a good place to do this.) Be sure to come back to such unresolved questions, if the next few papers do not help you.

Answering these questions in sequence is a good way to develop a short precis of the paper:

- If you condense your answers to these questions in a few sentences altogether, it may be appropriate as an entry for that paper in the "Related Work" section of a research paper that you yourself are writing;
- A few sentences in response to each question might be appropriate for your annotated bibliography;
- The same as above, but stressing the critical evaluation aspect, may be appropriate as a formal review of the paper for a conference or journal;

- A couple of paragraphs altogether may be appropriate as an entry for that paper in a survey or research report you are developing.

As you develop answers to these questions while you are reading the paper, make sure to record them - writing in the margin with a pencil is okay, but nothing beats typing it into an annotated bibliography.

1. What is the problem the authors address in the paper?

   - What is the domain of the problem? What is the aspect of the problem that the authors want to focus on? Why; is there a reason to believe this aspect is more or less important than others? What other papers do we know of that deal with the same problem, and how does the particular flavor or aspect of the problem in this paper compare with those?

2. How do the authors formulate/model the problem?

   - Does it fall into any common class - graph model, ILP, etc.? Any sub-class, e.g. layered graph formulation? Is it clear why the problem is modeled this way? Is this formulation the only one or obviously the most appropriate one possible, or could it have been modeled equally easily as something else? Does the rest of the paper utilize or depend heavily upon the formulation?

3. What is the solution proposed or result offered?

   - What is the nature of the solution - an algorithm, a formula that is derived, a mathematical proof of some assertion, something else? How can the solution approach be demonstrated on small-scale or "toy" problem instances? How easy is it to apply the solution for large or practical instances - how scalable, how quick? How do these metrics and characteristics compare with those of other solutions or solution approaches to the same or similar approaches in other papers? Does the solution use composition of various approaches? Can the solution approach be used in conjunction with or combined with other existing or conceivable solution approaches?

4. What evidence have they provided in favor of their solution approach?

   - Has the performance of the approach been measured against absolute or relative known or provable results about the solution? Are there any guarantees of the solution obtained using the proposed approach - is it exact/optimal, or provably good, or provably probably good, or any other type of performance guarantee? Do the authors offer results of the performance of the approach, from either simulation or experiment? Was the approach implemented or realized in some realistic domain?

5. How does the solution compare with other possible approaches?

   - Do the results show whether the proposed approach performance compares better or worse with existing ones? What other approaches should the performance be compared with? How can one obtain such a comparison? Do there appear to be any issues with the performance that have not been touched upon by the results presented?

6. How dependable do you find the evidence advanced?

   - Are the conditions under which the results were obtained described sufficiently thoroughly? Are they repeatable? Are they realistic? What measures did the authors take to attain

objectivity? How do they compare with the conditions and measures offered by other papers dealing with the same or similar problems?

Additional Resources

- A [paper in ACM CCR, July 2007, by Keshav](#) provides a three-pass paper reading method. The references cited in this paper are also useful. You need access to the ACM Digital Library to use the link above (automatically available from NCSU); the citation of the print copy is: "How to read a paper", by S. Keshav, in *ACM SIGCOMM Computer Communication Review*, Volume 37 , Issue 3 (July 2007), Pages: 83 - 84, April 2007, ISSN:0146-4833.