

How to Read a Computer Science Research Paper

by Amanda Stent

Where are CS research papers found?

CS research papers may be published as: technical reports, conference papers, journal papers or book chapters. Frequently, an author will write a technical report that expands on the information in a conference paper. Results from several conference papers may be combined and expanded into a journal paper. A conference paper may also be expanded later into a book chapter. Finally, PhD dissertations are frequently revised into either journal papers or book chapters.

The timeline, then, is usually: conference paper, technical report, journal paper, book chapter. To get the most recent, “hot off the presses”, information, look for conference papers. To get more considered papers with more explanation, look for journal papers.

Conference papers and most journal papers are “peer-reviewed”, that is, have been examined by other computer scientists prior to publication. Technical reports are typically not peer-reviewed, but are still excellent sources of detailed information about algorithms and computer systems.

More and more authors post their papers on their personal web sites. If an article is posted on the web, it will probably end up in CiteSeer (www.cite-seer.com). Other places to look for papers include Science Index (available through the library) and preprint servers available from journal publishers such as Elsevier and Kluwer, which may even post copies of papers not yet published!

What are the different types of CS research paper?

There are three basic types of CS research paper: theoretical, engineering and empirical.

A theoretical paper describes a theory or algorithm or provides a mathematical proof for some hypothesis.

An engineering paper describes an implementation of an algorithm, or part or all of a computer system or application. Engineering papers are now frequently required to include descriptions of system evaluation.

An empirical paper describes an experiment designed to test some hypothesis.

All three types of paper are found in computational linguistics.

How can I tell whether a research paper is good before I read it?

The short answer to that is, it is always best to read it. However, every year thousands of papers are published by the ACM, IEEE, AAAI, ACL and other professional societies. You can't read them all!

Usually, you want to read a research paper either because you have a specific problem you need to solve, or to keep up with your field. In either case, as you read research papers you will begin to get an idea of which venues and which researchers publish good research in your area. Citeseer can help with this, as it tracks citations (so that you can find out what papers other researchers consider to be excellent or at least popular).

In addition to learning what are good publication venues in your field and checking citations, you can sometimes tell if a paper is likely to be good by looking for structural clues in the paper itself. Here are some indicators of a good research paper:

- 1) The problem the paper addresses is clearly stated, both in the abstract and early on in the paper itself. The technical importance and broader impacts of the paper are described.
- 2) The paper includes a clear description of the experiment, system or theory the problem addresses. This is usually the second section of the paper.
- 3) The paper describes and analyzes the results of the work described (either experimental or evaluation results).
- 4) The authors have some sound, non-trivial ideas for future work. This usually appears at the end of the paper.
- 5) Related work is described and cited correctly. You can get an idea of this by looking at the list of references at the end of the paper.

If you know that a researcher has been working in an area for awhile, that is usually an indicator that the research is sound; however, do not underestimate the contributions of people new to a field or the impact of politics on research.

How should I read a research paper?

First, skim the paper to check that it has the elements described above (clear problem statement, clear description of method or system, clear presentation of results, appropriate citation of related work). At the end of this quick skim you should be able to write down the type of the paper (theoretical, engineering or empirical), the area of computer science the paper is about, and the problem the paper addresses.

Now read the introduction carefully, looking for the problem statement, theoretical importance and broader impact. Sometimes related work is also addressed in the introduction. Consider: are the author's claims reasonable and realistic? Do the authors present evidence that they know why they are doing this piece of research? Do they have an idea of the larger picture?

Now read the methods section carefully (this section might be called “experiment” or “system description”, or may involve statement of theorems and their proofs). Ask yourself:

1. Can you think of counter-examples for examples given?
2. Is the approach clearly described? Can you outline the steps or summarize the approach?
3. Does the work address the problem stated earlier in the paper?
4. Does the approach seem to require unreasonable amounts of human guidance?
5. Does the approach seem objective?

Computer science papers are often written in English by non-native speakers of English. Syntactic errors or awkwardness of phrasing do not indicate that the research is bad; you should try distinguish between the writing style and the research itself. If you get confused, you can reread or write down your questions to check in other work by the same authors.

If, at the end of this section, you think the approach is sound (even if you disagree with it!), read the results, evaluation or analysis section. (If there is no such section and the paper dates from post-1995, there is a problem!) Ask yourself:

1. Do these results address the problem stated at the start of the paper?
2. Do the authors use commonly-accepted terminology and methods?
3. What is the conclusion of the research?

Finally, look at the conclusions and future work (last section of the paper). These can be good sources of topics for your own research.

A brief note about older research papers: computer science is a rapidly evolving field and is still very young compared to many of the fields closest to it (mathematics, engineering). Research papers that date from before the 1990s may appear very different from newer papers. They may be more speculative, include less evaluation or no evaluation, and typically involve the use of much less data. There are very practical reasons for this (slower computers, less data available). These are not reasons to discount these papers!

How can I remember the papers I have read?

I strongly recommend that you make an electronic file for your own bibliography. A BibTeX file is a good idea. After you read a paper, if you think it is worth remembering, write an entry for that paper in your bibliography file. You should note: authors' names, paper title, how paper was published (conference proceedings, journal etc.), date of publication and page numbers (if possible). Add a 2-3 sentence description of the paper in which you summarize for yourself the problem addressed by the paper, the solution proposed or results learned, and the future work proposed.

Some people publish their bibliographies, see: <http://liinwww.ira.uka.de/bibliography/>.