

Tips for Writing Technical Papers

[Jennifer Widom](#), January 2006

Here are the notes from a presentation I gave at the [Stanford InfoLab](#) Friday lunch, 1/27/06, with a few (not many) revisions when I reprised the talk on 12/4/09, and no revisions for the 10/19/12 revival. The presentation covered:

- [Paper Title](#)
 - [The Abstract](#)
 - [The Introduction](#)
 - [Related Work](#)
 - [The Body](#)
 - [Performance Experiments](#)
 - [The Conclusions](#)
 - [Future Work](#)
 - [The Acknowledgements](#)
 - [Citations](#)
 - [Appendices](#)
 - [Grammar and Small-Scale Presentation Issues](#)
 - [Mechanics](#)
 - [Versions and Distribution](#)
-

Running Example

As a running (fictitious!) example, suppose you've designed and run experiments with a new algorithm for external multipass merge-sort. Your algorithm reduces the complexity from $O(n \log n)$ to $O(n)$, under the premise that it's acceptable to have some bounded "unsortedness" in the result. You plan to write up the results for submission to a major conference.

Note: This example was used throughout the live presentation but I haven't followed through much in these notes. Thus, the notes include several exercises for the reader.

Paper Title

Titles can be long and descriptive:

- *Linear-Time External Multipass Sorting with Approximation Guarantees*

or short and sweet:

- *Approximate External Sort*

Here's a middle-of-the-road length, plus a cute name that sticks in people's minds:

- *Floosh: A Linear-Time Algorithm for Approximate External Sort*

The Abstract

State the problem, your approach and solution, and the main contributions of the paper. Include little if any background and motivation. Be factual but comprehensive. The material in the abstract should not be repeated later word for word in the paper.

(**Exercise:** Write an abstract for the multiway sort example.)

The Introduction

The Introduction is crucially important. By the time a referee has finished the Introduction, he's probably made an initial decision about whether to accept or reject the paper -- he'll read the rest of the paper looking for evidence to support his decision. A casual reader will continue on if the Introduction captivated him, and will set the paper aside otherwise. Again, *the Introduction is crucially important*.

Here is the [Stanford InfoLab](http://cs.stanford.edu/people/widom/paper-writing.html)'s patented five-point structure for Introductions. Unless there's a good argument against it, the Introduction should consist of five paragraphs answering the following five questions:

1. *What is the problem?*
2. *Why is it interesting and important?*
3. *Why is it hard?* (E.g., why do naive approaches fail?)
4. *Why hasn't it been solved before?* (Or, what's wrong with previous proposed solutions? How does mine differ?)
5. *What are the key components of my approach and results?* Also include any specific limitations.

(**Exercise:** Answer these questions for the multiway sort example.)

Then have a final paragraph or subsection: "Summary of Contributions". It should list the major contributions in bullet form, mentioning in which sections they can be found. This material doubles as an outline of the rest of the paper, saving space and eliminating redundancy.

(**Exercise:** Write the bullet list for the multiway sort example.)

Related Work

The perennial question: Should related work be covered near the beginning of the paper or near the end?

- **Beginning**, if it can be short yet detailed enough, or if it's critical to take a strong defensive stance about previous work right away. In this case Related Work can be either a subsection at the end of the Introduction, or its own Section 2.
- **End**, if it can be summarized quickly early on (in the Introduction or Preliminaries), or if sufficient comparisons require the technical content of the paper. In this case Related Work should appear just before the Conclusions, possibly in a more general section "Discussion and

Related Work".

The Body

Guideline #1: A clear new important technical contribution should have been articulated by the time the reader finishes page 3 (i.e., a quarter of the way through the paper).

Guideline #2: Every section of the paper should tell a story. (Don't, however, fall into the common trap of telling the entire story of how you arrived at your results. Just tell the story of the results themselves.) The story should be linear, keeping the reader engaged at every step and looking forward to the next step. There should be no significant interruptions -- those can go in the Appendix; see below.

Aside from these guidelines, which apply to every paper, the structure of the body varies a lot depending on content. Important components are:

- **Running Example:** When possible, use a running example throughout the paper. It can be introduced either as a subsection at the end of the Introduction, or its own Section 2 or 3 (depending on Related Work).
- **Preliminaries:** This section, which follows the Introduction and possibly Related Work and/or Running Example, sets up notation and terminology that is not part of the technical contribution. One important function of this section is to delineate material that's not original but is needed for the paper. Be concise -- remember Guideline #1.
- **Content:** The meat of the paper includes algorithms, system descriptions, new language constructs, analyses, etc. Whenever possible use a "top-down" description: readers should be able to see where the material is going, and they should be able to skip ahead and still get the idea.

Performance Experiments

We could have an entire treatise on this topic alone and I am surely not the expert. Here are some random thoughts:

- Many conferences expect experiments.
- It's easy to do "hokey" or meaningless experiments, and many papers do.
- It's easy to craft experiments to show your work in its best light, and most papers do.
- What should performance experiments measure? Possibilities:
 - Pure running time
 - Sensitivity to important parameters
 - Scalability in various aspects: data size, problem complexity, ...
 - Others?
- What should performance experiments show? Possibilities:
 - Absolute performance (i.e., it's acceptable/usable)

- Relative performance to naive approaches
- Relative performance to previous approaches
- Relative performance among different proposed approaches
- Others?

The Conclusions

In general a short summarizing paragraph will do, and under no circumstances should the paragraph simply repeat material from the Abstract or Introduction. In some cases it's possible to now make the original claims more concrete, e.g., by referring to quantitative performance results.

Future Work

This material is important -- part of the value of a paper is showing how the work sets new research directions. I like bullet lists here. (Actually I like them in general.) A couple of things to keep in mind:

- If you're actively engaged in follow-up work, say so. E.g.: "We are currently extending the algorithm to... blah blah, and preliminary results are encouraging." This statement serves to mark your territory.
- Conversely, be aware that some researchers look to Future Work sections for research topics. My opinion is that there's nothing wrong with that -- consider it a compliment.

The Acknowledgements

Don't forget them or you'll have people with hurt feelings. Acknowledge anyone who contributed in any way: through discussions, feedback on drafts, implementation, etc. If in doubt about whether to include someone, include them.

Citations

Spend the effort to make all citations complete and consistent. Do *not* just copy random inconsistent BibTex (or other) entries from the web and call it a day. Check over your final bibliography carefully and make sure every entry looks right.

Appendices

Appendices should contain detailed proofs and algorithms only. Appendices can be crucial for overlength papers, but are still useful otherwise. Think of appendices as random-access substantiation of underlying gory details. As a rule of thumb:

- Appendices should not contain any material necessary for understanding the contributions of the paper.
- Appendices should contain all material that most readers would not be interested in.

Grammar and Small-Scale Presentation Issues

In general everyone writing papers is strongly encouraged to read the short and very useful [The Elements of Style](#) by Strunk and White. Here's a random list of pet peeves.

- Just like a program, all "variables" (terminology and notation) in the paper should be defined before being used, and should be defined only once. (Exception: Sometimes after a long hiatus it's useful to remind the reader of a definition.) Global definitions should be grouped into the Preliminaries section; other definitions should be given just before their first use.
- Do not use "etc." unless the remaining items are completely obvious.
 - Acceptable: *We shall number the phases 1, 3, 5, 7, etc.*
 - Unacceptable: *We measure performance factors such as volatility, scalability, etc.*

(**Exercise:** The above rule is violated at least once in this document. Find the violations.)

- Never say "for various reasons". (Example: *We decided not to consider the alternative, for various reasons.*) Tell the reader the reasons!
- Avoid nonreferential use of "this", "that", "these", "it", and so on (Ullman pet peeve). Requiring explicit identification of what "this" refers to enforces clarity of writing. Here is a typical example of nonreferential "this": *Our experiments test several different environments and the algorithm does well in some but not all of them. This is important because ...*

(**Exercise:** The above rule is violated at least once in this document. Find the violations.)

- Italics are for definitions or quotes, not for emphasis (Gries pet peeve). Your writing should be constructed such that context alone provides sufficient emphasis.

(**Exercise:** The above rule is violated at least once in this document. Find the violations.)

- People frequently use "which" versus "that" incorrectly. "That" is defining; "which" is nondefining. Examples of correct use:
 - *The algorithms that are easy to implement all run in linear time.*
 - *The algorithms, which are easy to implement, all run in linear time.*

Mechanics

- Always run a spelling checker on your final paper, no excuses.
- For drafts and technical reports use 11 point font, generous spacing, 1" margins, and single-column format. There's no need to torture your casual readers with the tiny fonts and tight spacing used in conference proceedings these days.
- In drafts and final camera-ready, fonts in figures should be approximately the same font size as used for the text in the body of the paper.
- Tables, figures, graphs, and algorithms should always be placed on the top of a page or column, not in the body of the text unless it is very small and fits into the flow of the paper.
- Every table, figure, graph, or algorithm should appear on the same page as its first reference, or on the following page (LaTeX willing...).

- Before final submission or publication of your paper, *print it* once and take a look -- you might be quite surprised how different it looks on paper from how it looked on your screen (if you even bothered to look at it after you ran Latex the last time...).

Versions and Distribution

- Many papers have a submitted (and later published) conference version, along with a "full paper" technical report on the web. It's important to manage versions carefully, both in content and proliferation. My recommendation is, whenever possible, for the full paper to consist of simply the conference version plus appendices. The full paper should be the only public one aside from conference proceedings, it should be coordinated with latest (final) conference version, and modifications to the full paper should always overwrite all publicly accessible previous versions of it.
- I believe in putting papers on the web the minute they're finished. They should be dated and can be referenced as technical reports -- it's not necessary to have an actual technical report number. Never, ever put up a paper with a conference copyright notice when it's only been submitted, and never, ever reference a paper as "submitted to conference X." You're only asking for embarrassment when the paper is finally published in conference Y a year or two later.