Crafting Scientific Papers

Mark Johnson

Macquarie University
Sydney, Australia

October 21, 2014

MACQUARIE UNIVERSITY

Outline

MACQUARIE UNIVERSITY

Talk summary

- Researchers' job: change the way people do (or think about) something
- Your work is not done until you've communicated your results
- Written publications are the primary way we communicate our results
- Figure out your document's purpose, and design it to achieve that purpose
- Scientific ideas are hard to understand: it's hard work to craft a document that's easy to read
- This talk gives suggestions about how to craft scientific papers and talks

Outline

Why are you writing this document?

- Figure out what goals you want your document to achieve, and *design the document so it achieves your goals*
  - if your paper will have no impact, *don't bother writing it*
  - in undergraduate study, the purpose is e.g., to demonstrate you have mastered a subject, so you will pass the unit
  - the purpose of a scientific paper is to *change the field in some way* (e.g., adopt a new technology, or at least think differently about a problem)

$\Rightarrow$ A "brain dump" might be ok for an undergrad paper, but almost never is appropriate for a research document

- Write your document with your intended audience in mind
  - figure out what information the readers need in order to do what you want, and give it to them
  - for applications and grant proposals, try to get the reviewing criteria

What's the 1-sentence summary?

- We summarise other people's papers in a single sentence, so it's only reasonable to expect they'll summarise our papers in a single sentence too

$\Rightarrow$ *Figure out what your paper's 1-sentence summary will be*, because otherwise other people will do it for you (probably badly)

- Put your 1-sentence summary into your conclusion and your abstract and/or your title (maybe slightly changing the wording)

Differences between papers and conference talks

- *Research papers should be self-contained*
  - a paper should contain *all the information needed to convince an expert in the field*
  - you may need appendices or supporting documentation
- In computer science, many of our papers appear in conference proceedings, and are associated with a conference presentation
- Most conference talks aren't long enough to present all the material from the paper
  - and a talk usually isn't a good place to present a details of a proof or new algorithm

$\Rightarrow$ Use the conference presentation as *an advertisement for the paper* in the conference proceedings

  - most people only read a small fraction of the papers in the proceedings

Outline

The components of a paper

- The conventional paper structure in many fields is:
  - Title, authors, affiliations
  - Abstract
  - Introduction (including summary of prior work; this can be its own section)
  - one or more sections explaining the paper's contributions (this is the meat of the paper)
  - Experimental Results
  - Conclusion
  - Acknowledgements (e.g., funding agency)
  - Bibliography
- If you're writing a paper in an area for a first time, copy the structure of good papers in your field
- Use a standard document structure unless there's a good reason for using a different structure

Sections and subsections

- You should structure your papers and talks into sections that reflect the logical structure
  - section headings help make clear the goal(s) of this part of the paper/talk
- In a paper (but usually not in a talk) it's often good to divide sections into subsections
  - usually it doesn't make sense to use subsubsections
  - you can use numbered or bullet points or "descriptions" to provide structured lists
- A book (e.g., a thesis) should have chapters as well as sections (and subsections)

Why discuss prior work?

- A good discussion of prior work should:
  - convince the readers that you understand the field well enough not to have missed something important

  $\Rightarrow$ you should discuss work that readers are likely to think is relevant
  - help the readers understand your contribution by relating it to something they already understand
- Provide *full bibliographic references*, so readers can find it
- Be very clear about *what is prior work and what is your innovation*
  - explicitly compare your work to close prior work
  - sometimes it may be useful to delay discussion of prior work until after you've presented your contribution

Outline

What goes into a paper?

- A paper is usually designed to *convince the reader of something*
  - E.g., a scientific claim is true
  - E.g., a new algorithm is a better way of solving some problem than an old one
- It is *not* a description of how you did your research
  - real research is messy, full of ideas that didn't work, etc.

  $\Rightarrow$ in general, *the order in which you did the research won't be the same as the order in which you present the material*
  - nobody cares about your research experience (unless you're getting a Nobel prize)
- Only include content that is *relevant to the paper's goals*

  $\Rightarrow$ the paper shouldn't be a "brain dump"
  - when writing something in a paper, always ask: *does the reader need to know this?*
- Figure out the paper's goals, then work out *what you have to tell the reader to convince them your claims are correct*

Why plan your research paper?

- High level goal: *make sure you've already told the reader what they need to know in order to understand what you're about to tell them now*
  - identify your target audience: usually researchers in the field who don't know the specific topic of the paper
  - it's fine assume standard background knowledge
  - include citations to textbooks or survey papers where reader can find any necessary background information
- Use a non-standard document structure if it enables you to present material in a more coherent order
  - explain to reader why you're doing this

Drafting the paper

- There are many ways to do this; I usually write papers backwards
- Draft a bullet point outline, with roughly one bullet point per paragraph in the final paper
  - draft the conclusion first, one bullet point for each high-level point that the paper makes
  - pick a title and write the abstract (about a paragraph)
  - draft the results section; work out what tables or figures are needed to provide evidence for the conclusion (this may identify additional results you'll need for your paper!)
  - draft bullet points for the rest of the results section, indicating e.g., the experimental methods you'll need to explain
  - draft bullet points for the content sections of the paper (this may force you to change the results and/or conclusion section)
  - draft the introduction and prior work section(s)
- Now work forwards, replacing the bullet points with text

Explain everything three times

- Scientific papers are hard to read – do what you can to make yours easier to read!
- In general, *explain everything three times*:
  - tell the reader in general terms what you're about to explain
  - then explain the material in detail
  - then summarise for the reader what you've just told them
- Don't worry about making it too easy!
- Papers and monographs already have this structure (Introduction, Conclusion)
- It's good to do this within the section and subsection level too; i.e., have an introductory paragraph at the beginning of each section explaining what the section contains, and how it relates to the point of the paper

Running worked examples

- Detailed, worked examples can help readers understand difficult concepts
- Introduce examples immediately after a complex definition or part of the paper
- It's a good idea to have two or three examples that you refer to and build on throughout the paper
- The examples can be artificial, chosen to provide simple demonstrations of key points of the paper
- The examples can also help tie the subsections and sections of the paper together
- *Identifying the examples you will use is a key part of designing a paper*

Outline

A scientific paper isn't a murder-mystery

- Don't "hold back the best for last": most readers won't get to the end of the paper anyway

$\Rightarrow$ Trumpet important points in the abstract and the introduction, maybe even in the title

  - e.g., a good title might be "Part-of-speech tagging improves named entity extraction"
  - even then, people will misunderstand the point of the paper

Make your paper as easy to read as possible

- Scientific papers are hard to read because the ideas are complicated.

$\Rightarrow$ Make your text as easy to read as possible. Use simple words and short sentences if you can.

- You're not writing a novel: avoid rhetorical flourishes
  - ignore your high-school writing advice
  - use the same word to describe the same concept throughout the paper
  - give examples of complicated definitions or concepts
- Use *consistent formatting throughout* (easy with LaTeX)
- Write in *standard English*
  - many readers are not native speakers
  - make your document *as simple as possible* (but not simpler!)

Citations and cross-references

- *It is crucial to cite the right publications*
  - if you miss obviously relevant papers, your readers will think you don't understand the topic
  - you can cite background material your readers should know, rather than explaining it in your paper
- Use *cross-references within your document* where appropriate
  - cross-references are not appropriate in a talk (repeat the material instead)
  - in a complex document, use cross-references to help reader find key *definitions*, *running examples* and *results*
  - in LaTeX, use `\ref` and `\pageref` (except in conference proceedings)
  - to create hyperlinked cross-references, use the `hyperref` package, with `\autoref` and `\autopageref` commands

Proof-reading your paper

- First drafts are almost never very good

$\Rightarrow$ Plan to re-read and polish your paper (several times if it is important)

- After you've re-read a paper several times you lose the ability to see problems in it
  - give it to a friend to proof-read (in general, another person can only give you advice on a paper once)
- After a week or so, you start to see problems in your paper again

  $\Rightarrow$ if a document is important, give yourself several weeks to write it
- *Use a spelling corrector*. E.g., ispell or aspell for Emacs.

Layout and fonts

- Most conferences will supply a LaTeX sty file that specifies layout, fonts, etc
  - there's usually a rigorously-enforced *page limit*
- If there's no specified formatting, use a conventional layout and font
  - *don't squeeze much onto a page* (both papers and talks)
  - fancy fonts get tiresome quickly
- For slides, "less is more": *use at least 18 point font*

Outline

MACQUARIE UNIVERSITY

Summary and conclusions

- Your publications are the primary way your research will become known

$\Rightarrow$ *Quality documents are critical for your scientific reputation!*

- Identify the document's goals, and *engineer the document to achieve those goals*
- Use *a standard document structure* unless you have a good reason not to
- Make sure the reader doesn't have to be clairvoyant
- Explain everything three times
- Use examples to clarify difficult concepts

$\Rightarrow$ Structure documents back-to-front, then write front-to-back

- Usually, make conference talks *advertisements for the paper*