How to write a research paper?

Renée J. Miller

Acknowledgments

- Acknowledgements
  - These slides are an extended (updated) version of Alberto Mendelzon's well-known writing slides.
  - Other useful tips at:
    - dblab.cs.toronto.edu/~miller/Research/writing.html
  - Slides on revisions and responding to reviewers are from Patricia Arocenas
    - H. C. Williams, "How to reply to referees' comments when submitting manuscripts for publication", Journal of the American Academy of Dermatology, Volume 51, Issue 1, July 2004, Pages 79-83

How to write a research paper?

The goal of education is the advancement of knowledge and the dissemination of truth.
John F. Kennedy

1. Do good research
2. Write it up well
3. Submit it to the right place
4. Revise and/or Resubmit (?)

Part I: Doing good research

- David A. Patterson, "How to have a bad career in research/academia," www.cs.berkeley.edu/~pattrsn/talks/nontech.html

Part II: Writing it up well

If you can't say it clearly, you don't understand it yourself. *John Searle*

- Basics: grammar, spelling, mathematical accuracy
- Objectives
- Constraints
- Organization
- Style

Paper objectives

- Communicate/inform
  - What you did
  - How you did it
  - What you learned from it
- Convince/persuade
  - Why it's true/plausible/feasible
  - Why it's important
  - How it improves the state of the art

Constraints

- Audience
- Length
- Politics

Audience

- What do they know?
  - Experts in the field
  - Experts in related fields
  - General DB/OS/AI... audience
  - General CS audience
- What do they care about?
  - Theoreticians
  - System builders
  - Researchers vs. practitioners
  - Reviewers vs. readers

Length

I have only made this long because I have not had the time to make it shorter.

Blaise Pascal

- Usually constrained by the call for papers/publisher
- Be ruthless in cutting non-critical material
- Do not cut examples in favor of technical details
- Do not play margin/font size games

Politics

- Who are the authors?
  - In theory: everyone who made an intellectual contribution
    - contributing to the writing or the implementation is not enough
  - In practice: your boss? Your supervisor?
  - Order of authors
    - try to avoid controversy, discuss upfront
      - alphabetical
      - primary author first
      - reverse "academic" age
- What cannot be said?
  - Contractual limitations, e.g., your license to use software indicates you cannot publish benchmark results
  - Premature disclosure

Organization

- Typical structure
  - Title and authors
  - Abstract
  - Introduction and road map
  - Related work
  - Research description
  - Conclusions
  - Acknowledgements
  - Bibliography
  - Appendices

Title

- Title: tradeoff between specificity and length
  - "Efficient computation of approximately optimal data summaries for temporal data warehouses using Haar wavelets"
  - "Summaries in data warehousing"
  - "Wavelet summaries for temporal data warehouses"

Abstract

- Should answer the question: do I want to read this paper?
- Summarize problem and results
- Single paragraph
- No citations
- Avoid "In this paper…"

Introduction

- Often the hardest part to write
- Motivation
  - why is the problem important/interesting?
  - examples
  - applications
- Background
  - see if related work can be woven in
- Approach and results
- Roadmap

The first sentence

- Bad examples (real examples from a single SIGMOD/ PODS conference year)
  - Marketing-speak
    - Effective decision-making is vital in a global competitive environment where business intelligence systems are becoming an essential part of virtually every organization
  - Platitudes
    - The Extensible Markup Language (XML) is rapidly emerging as the new standard for data representation and exchange on the Internet.
    - The Extensible Markup Language (XML) is becoming the dominant standard for exchanging data over the WWW.
    - The Extended Markup Language (XML) is emerging as the standard for data exchange on the Web.
    - XML is becoming the new standard for the exchange and publishing of Data over the Internet.
    - XML has become an important medium for data representation...

The first sentence

I always write a good first line, but I have trouble in writing the others.

Molière

- A better example
  - We study absolute and relative keys for XML, and investigate their associated decision problems.
- Be specific about your contributions

Research Description

- Core of the paper
- Ways to organize it
  - Logical chain: problem statement, previous solutions, new solution, analysis
  - From general to specific: general outline first, then fill in details
  - From simple to complex: solve easy special case first, then harder cases
  - By architecture: describe each system component in turn

Related Work

- Survey of the relevant literature
  - Don't just repeat X's contribution statement using X's terminology
  - Give motivation for X and how it differs in motivation, solution, or other characteristics from your work
- Can be all in one place or woven through paper
  - Consider creating a narrative around historical evolution of field
- Goal: substantiates novelty of the work and provides context for research

Citations

And for the citation of so many authors, 'tis the easiest thing in nature. Find out one of these books with an alphabetical index, and without any farther ceremony, remove it verbatim into your own.

Miguel de Cervantes

- What are citations for? Be sure context makes it clear...
  - To justify something you claim
  - To show you are aware of earlier work
  - To give credit where it is due
  - To let interested readers dig deeper
  - To flatter your reviewers... avoid
- Show good scholarship in using the right citations, not long (kitchen sink) lists of citations

How do you find them?

- Know the best venues in your field
  - Stay on top of reading these
- Talk to experts
  - Go to conferences (see our "Networking" session)
- Search online resources
  - DBLP
  - ACM Digital Library
  - Google Scholar (Alerts)

Plagiarism

If you steal from one author it's plagiarism; if you steal from many it's research.

Wilson Mizner

- Definition: Use without attribution
- Be careful not to lift words verbatim or close to it from other papers (even your own)
  - Use quotations and citations for verbatim passages
  - Use citations for reworded descriptions
- Discuss standards for attribution with your advisor and research group

Imitation....

I could tell you which writer's rhythms I am imitating. It's not exactly plagiarism, it's falling in love with good language and trying to imitate it..

Charles Kuralt

- Imitating the style of a well-written paper is a great way to learn how to write...
- Study how they created their argument and see if the same structure will work for you

Genius borrows nobly.
Ralph Waldo Emerson

Style

- Generic advice on style:
  - William Strunk Jr. and E.B. White, The Elements of Style (3d ed.). Macmillan, 1979. www.bartleby.com/141
  - Omit needless words
  - Prefer the standard to the offbeat
  - Vigorous writing is concise
- I also recommend
  - "Writing English as a Second Language", W. Zinsser
  - http://theamericanscholar.org/writing-english-as-a-second-language/

Citation style

- Citations are parenthetical remarks; text should be readable (and grammatically correct) without them.
- Wrong:
  - Thirty-second normal form is defined in [AO72].
  - [A072] contains a definition of...
- Right:
  - Alpha and Omega defined thirty-second normal form [A072].
  - Many researchers have studied these normal forms [A072,ABC00,XYZ+80].
- Use less cryptic citations if possible
  - [AlphaOmega 72] better than [A072] better than [14]

Mathematical writing

- Good resource: Knuth, Larrabee, and Roberts book on Mathematical Writing
  - Don't punctuate math symbols (real examples)
    - "There are 23^5 other left-deep query plans."
      - Reads like 5 is an exponent but it is meant as a footnote...
    - "... relation r. P is the next..."
      - Is this one sentence mentioning r.P or two sentences?
  - Do not start sentences with symbols even capital symbols
    - **Wrong:** f is a total function.
    - **Right**: Function f is total.
  - Avoid using notation with multiple, or (horrors!) nested, sub- or super-scripts.
  - Do not use notation for the sake of notation. Sometimes it is clearer to use prose.

Debugging your paper

- Read it
- Use a spell-checker
- Have other people read it – how?

Multi-author protocols

- Have a coordinator:
  - Ensures consistency of sections
  - Gets formatting right
  - Submits as the contact author
- Use a locking protocol/ version control software
- Use macros for visible comments
- Document your changes with comments
- Avoid non-terminating change sequences (colour → color → colour → color …)

Other writing tips

- Bullet lists are over used by many CS writers
  - Can be effective for drawing attention to a set of important statements
  - Are not an excuse for writing abbreviated or sloppy prose
  - Should be punctuated consistently
  - Should use consistent sentence or phrase structure in each item
- Enumerated nouns should be capitalized consistently (or not at all). Do not switch back and forth on a whim.
  - See Figure 1 in Appendix A.
  - We will use Function f1 in Equation 32a.
  - In our experiments, Iguana 17 performed very well.
  - Note that the words section and figure are not capitalized in English unless they are enumerated (see Section 4).

Part III: Submit to the right place

- You should have read dozens of papers from a venue before submitting there
  - Understand the audience
  - Understand the venue's conventions/expectations
  - Understand their process of selection

The process of selection

- Anonymous peer review
- Conference
  - Program committee
  - Blind
    - Only reviewers are anonymous
    - Reviews know authors but authors don't know reviewers
  - Double blind
    - Both reviewers and authors anonymous
  - Fixed upper bound on acceptances
- Journal
  - Editor and referees

Part: IV Revise and/or Resubmit

- What if your paper is reject?
  - Rant to your office mates
  - Rant to your family, friends, neighbors...
  - Tell everyone you are going to drop out and become a barista...
- Put review in draw for at least a week
- Reread reviews and incorporate them
  - They're right, I'll fix it
  - They didn't get it, how can I rewrite so they will?
- Try again, repeat until …
- Remember: some famous papers were rejected (e.g., DataCube, B+-tree)

If Revisions are allowed...

- Three Golden Rules
  - Respond thoroughly
  - Respond politely
  - Answer with evidence, my dear Watson!
- Well maybe 4:
  - Do not underestimate the task of revising the paper, and crafting the response letter

Thanks to Patricia Arocena for these slides

Rule 1: Respond Completely

- All reviewer's comments should be addressed, and responded to in sequence
  - Think on how to make ***their job easier***
    - "We first address the three important issues mentioned by the meta-reviewer, and afterwards present additional clarifications."
- Itemize the reviewers' comments (e.g., Reviewer 1, Comment 1.1, 1.2, and so on).
  - Use headings, bold, and italics to highlight them
- Include context
  - paraphrase their comments, then include response
  - be open to comment

Rule 2: Respond Politely

- Avoid confrontation
  - Even if reviewer is wrong
  - Avoid opening phrases such as "We totally disagree …"
  - Give and take: first find some common ground, then start with phrases such as "The referee is right to assert that … However, we would like to point out …"
- Provide, if possible, an escape clause (think about why he or she may have made a conflicting assertion):
  - "The misunderstanding may stem from a sentence in ... We have reworded to make it clear."
- Resist the temptation of using sarcasm in your replies
  - "If the referee had bothered reading the paper …"
  - Try this instead: "We agree that this is an important point and we have addressed it in page 7, paragraph 5."

Rule 3: Answer with Evidence

- Say why you disagree, provide a coherent argument, and back it up with facts!
- State which sections have been revised/added and why
  - "We added Section 5.3. Here we show examples of X … Moreover, we prove that …"
  - "We added a series of experiments using real-life data that show..."

Other advice

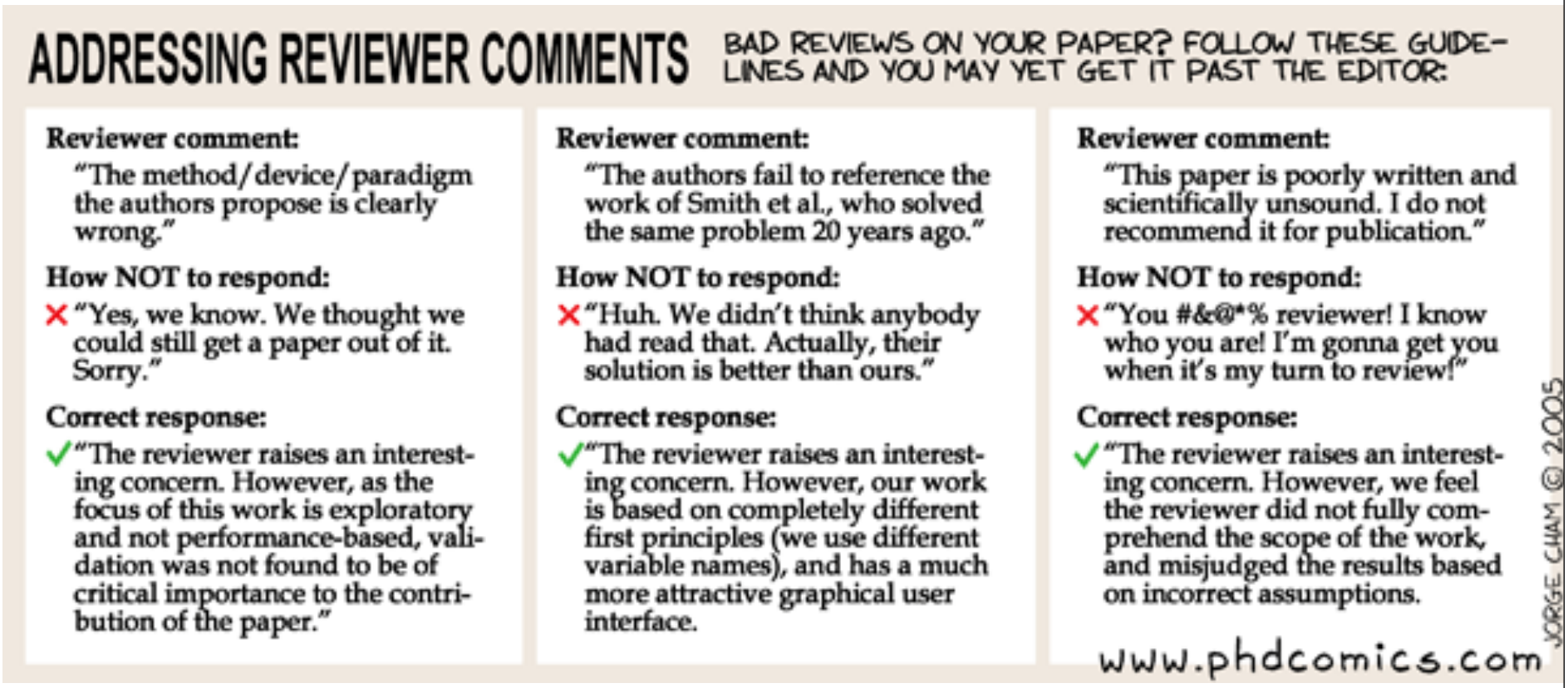
Consistency is the last refuge of the unimaginative.
Oscar Wilde

- Be consistent in structure and style
  - Unless you are using inconsistency deliberately and with intention (e.g., to draw attention to something)
  - Too often inconsistency is just laziness and is distracting to your readers
- Take joy and pride in a well-crafted, clean, clear argument
- Time time for your writing and use the writing process to improve your research!

Resources

- Books
  - Justin Zobel, Writing for Computer Science: The Art of Effective Communication. Springer, 1997.
    - 2nd edition (April 27, 2004)
  - Strunk and White, The Elements of Style
- I also recommend
  - "Writing English as a Second Language", W. Zinsser
  - http://theamericanscholar.org/writing-english-as-a-second-language/
  - www.cs.toronto.edu/~miller/Research/writing.html

A final thought....



ADDRESSING REVIEWER COMMENTS

BAD REVIEWS ON YOUR PAPER? FOLLOW THESE GUIDELINES AND YOU MAY YET GET IT PAST THE EDITOR:

Reviewer comment:
"The method/device/paradigm the authors propose is clearly wrong."

How NOT to respond:
✗ "Yes, we know. We thought we could still get a paper out of it. Sorry."

Correct response:
✓ "The reviewer raises an interesting concern. However, as the focus of this work is exploratory and not performance-based, validation was not found to be of critical importance to the contribution of the paper."

Reviewer comment:
"The authors fail to reference the work of Smith et al., who solved the same problem 20 years ago."

How NOT to respond:
✗ "Huh. We didn't think anybody had read that. Actually, their solution is better than ours."

Correct response:
✓ "The reviewer raises an interesting concern. However, our work is based on completely different first principles (we use different variable names), and has a much more attractive graphical user interface.

Reviewer comment:
"This paper is poorly written and scientifically unsound. I do not recommend it for publication."

How NOT to respond:
✗ "You #&@*% reviewer! I know who you are! I'm gonna get you when it's my turn to review!"

Correct response:
✓ "The reviewer raises an interesting concern. However, we feel the reviewer did not fully comprehend the scope of the work, and misjudged the results based on incorrect assumptions.

www.phdcomics.com

JORGE CHAM © 2005

UofT:DB Group