



## How to write a research paper?

Winter 2012

**Renée J. Miller**

## Acknowledgments

2

- Acknowledgements
  - ▣ These slides are an extended (updated) version of Alberto Mendelzon's well-known writing slides.
  - ▣ Other useful tips at:
    - [dmlab.cs.toronto.edu/~miller/Research/writing.html](http://dmlab.cs.toronto.edu/~miller/Research/writing.html)

11/16/11



## How to write a research paper?

3

*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

11/16/11



## Part I: Doing good research

4

- David A. Patterson, "How to have a bad career in research/academia," [www.cs.berkeley.edu/~pattsrn/talks/nontech.html](http://www.cs.berkeley.edu/~pattsrn/talks/nontech.html)

11/16/11



## 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

11/16/11



## 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

11/16/11



## Constraints

- Audience
- Length
- Politics

11/16/11



## 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

11/16/11



## 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

11/16/11



## 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
      - 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

11/16/11



## Organization

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

11/16/11



## 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”

11/16/11



## Abstract

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

11/16/11



## 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

11/16/11



## The first sentence

15

- 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...

11/16/11



## The first sentence

16

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

- A better example
  - ▣ We present algorithms to label the nodes of an XML tree which is subject to insertions and deletions of nodes.
- Be specific about your contributions

11/16/11



## 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

11/16/11



## Related Work

18

- 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 you
- 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

11/16/11



## 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?
  - 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

11/16/11



## How do you find them?

20

- You tell me....
- The DBLP Computer Science Bibliography
  - ▣ [www.informatik.uni-trier.de/~ley/db](http://www.informatik.uni-trier.de/~ley/db)
- Know the best venues in your field
  - ▣ Stay on top of reading these

11/16/11



## 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]

11/16/11



## Mathematical writing

22

- 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?

11/16/11



## Multi-author protocols

- Have a coordinator:
  - Ensures consistency of sections
  - Gets formatting right
  - Submits, is contact author
- Use a locking protocol (may be check-in/check-out software)
- Use macro for visible comments
- Document your changes with comments
- Avoid non-terminating change sequences (colour → color → colour → color ...)

11/16/11



## Other writing tips

25

- 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  $f_1$  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).

11/16/11



## Part III: Submitting to the right place

11/16/11



## 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

11/16/11



## What if it's rejected?

- Read reviewers' comments
  - Rant to your office mates
  - Rant to your family, friends, neighbors...
- 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)

11/16/11



## Other advice

29

*Genius borrows nobly.*  
Ralph Waldo Emerson

- Find some one whose writing style you admire
  - ▣ Copy it (the style) shamelessly
- 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

*Consistency is the last refuge of the unimaginative.*  
Oscar Wilde



## 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*
- Online
  - [www.cs.toronto.edu/~miller/Research/writing.html](http://www.cs.toronto.edu/~miller/Research/writing.html)

11/16/11

