

Writing Style for Computer Science

Enrico Franconi

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," (see slides in the course material)

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

On writing



Easy reading is damn hard writing.

Hawthorne

Word-smithing is a much greater percentage of what I am supposed to be doing in life than I would ever have thought.

Knuth

- •
- Most scientists can produce competent papers simply by following elementary steps:
 - create a logical organisation,
 - use concise sentences,
 - revise against checklists of possible problems,
 - seek feedback.

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
- A paper is a contribution if it has two properties: originality and validity.

Author's aims and scope

- To begin a paper, the first task is to identify your aims.
- Write down everything that motivated you to start the research.
 - What did you want to achieve?
 - What problems did you expect to address?
 - What makes the problems interesting?
- Next, define the scope of the work that you plan to write up.
 - To do so, it is necessary to make choices about what to include, and thus it is necessary to identify what might be included.



- Issues such as whether results have been critically analysed are of importance in papers,
- but there is a different emphasis for theses:
 - it is you, not the research, that is the primary object of scrutiny.

Constraints

- Audience
- Length
- Politics

- If someone dislikes anything you have written, remember that it is readers you need to please, not yourself.
- What do they know?
 - Experts in the field
 - Experts in related fields
 - General DB/OS/AI/SE/... audience
 - General CS audience
- What do they care about?
 - Theoreticians
 - System builders
 - Researchers vs. practitioners
 - Reviewers vs. readers

Write with the reader in mind

- Identify an audience, and write with an awareness of that audience.
- As you write a paper remember that, unlike you, the reader has not been thinking intensely about the material for an extended period of time.
- Also be careful of English language that is commonly used in mathematics or computer science, but is usually unfamiliar to the layman. For example, terms such as "if and only if", "contrapositive", or "nontrivial" are used so often in mathematics that one often forgets that a nonmathematician (such as a grant reviewer) may not know what they mean.
- Use an example whenever it adds clarification.

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 favour 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

Golden rule

- Tell the reader what you are going to say;
- then say it, and
- then tell the reader that you have said it.

Phases of writing

- The writing of a paper begins with a rough draft, perhaps based on notes of experiments or sketches of a couple of theorems.
- The next phase usually consists of filling out the draft to form a contiguous whole: explaining concepts, adding background material, arranging the structure to give a logical flow of ideas.
- Finally, the paper is polished by correcting mistakes, improving written expression, and taking care of layout.

Evolution of a paper

- Early drafts tend to be repetitive and long-winded.
- Another problem is that some material becomes irrelevant as the paper evolves.
- The ordering too may need to be reconsidered once the paper is complete.
 - When material is moved from one place to another, check that the text in each location is intelligible and appropriate in the new context. Beware, for example, of moving definitions of terms or of breaking the flow of an argument.
- Don't be afraid to shorten your papers: cutting will improve the quality. Edit for brevity and balance. Omit or condense any material whose content or relevance to the paper's main themes does not justify its length.

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?
- Summarise problem and results
- Single paragraph
- No citations
- Avoid "In this paper..."
- □ The more specific an abstract is, the more interesting it is likely to be. Instead of writing "space requirements can be significantly reduced", write "space requirements can be reduced by 60%". Instead of writing "we have a new inversion algorithm", write "we have a new inversion algorithm, based on move-to-front lists".

Introduction

KRD

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

Write a good introduction

- Most people who read a mathematics paper will only read the introduction and skim the theorems.
- Furthermore, when a reviewer reads a proposal for a grant or fellowship, it is the introduction which will have the most influence on the reviewer's opinion of your work.
- Consequently, you should put a great deal of time and effort into writing an effective introduction.
- Remember that in an introduction you are often trying to "sell" your work and convince others of its importance.

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 organisation

Banalities

- 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
- The opening paragraphs can set the reader's attitude to the whole paper, so begin well.

Research Description



- Core of the paper
- Ways to organise 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

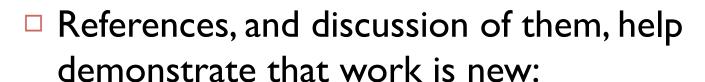
- Survey of the relevant literature
 - Don't just repeat X's contribution statement using X's terminology
 - Give motivation for X and how it differs in motivation, solution, or other characteristics from your work
- Can be all in one place or dispersed through paper
 - Consider creating a narrative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

Conclusions

- The closing section, or summary, is used to draw together the topics discussed in the paper.
- It should include a concise statement of the paper's important results and an explanation of their significance.
- This is an appropriate place to state (or restate) any limitations of the work: shortcomings in the experiments, problems that the theory does not address, and so on.
- Write "Conclusions", not "Conclusion". If you have no conclusions to draw, write "Summary".

Citations

- 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.
- References should be relevant, it should be up-to-date, it should be reasonably accessible, and it should be necessary.



Originality and references

- claims of originality are much more convincing in the context of references to existing work that (from the reader's perspective) appears to be similar.
- They demonstrate your knowledge of the research area, which helps the reader to judge whether your statements are reliable.
- And they are pointers to background reading.

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)



(1), (KRD

The quality of a paper can be reflected in its bibliography.

Good bibliography

- For example, how many references are there? This is a crude rule-of-thumb, but often effective.
- □ For some research problems there are only a few relevant papers, but such cases are the exception.
- Giving only a few references may be evidence of bad scholarship.
- Also, some authors cite a reasonable number of papers without actually citing related literature, thus disguising a core bibliography that is far too short.
- If only a couple of the references are recent, how sure can you be that the paper is valid? The author doesn't appear to be familiar with other research.
- Similarly, be suspicious of papers with no references to the major journals or conferences in the area.
- Also, some references age more quickly than others.

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

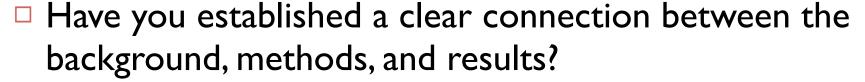
- Imitating the style of a well-written paper is a great way to learn how to write...
- Study how they created their argument and see if the same structure will work for you

Genius borrows nobly. Ralph Waldo Emerson

A writing-up checklist

- Have you identified your aims and scope?
- Are you maintaining a log and notebook?
- Does the paper follow a narrative?
- In what forum, or kind of forum, do you plan to publish'?
- What other papers should your write-up resemble?
- Are you writing to a well-defined structure and organisation?
- Have you chosen a form for the argument and results?
- Have you established a clear connection between the background, methods, and results?

A writing-up checklist



- How are results being selected for presentation?
- How do the results relate to your original aims?
- Have you used any
- unusual patterns of organisation?
- Have the results been critically analysed?
- Are the requirements for a thesis met?
- Do you and your co-authors have an agreed methodology for sharing the work of completing the write-up?

Style

- □ Generic advice on style:
 - Omit needless words
 - Prefer the standard to the offbeat
 - Vigorous writing is concise

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]

Respectable Graphs and Equations

- Use the right kind of chart.
- Avoid false precision

One of the most common statistical abuses is the presentation of averages to far more digits of accuracy than is justified by the data, with the subsequent temptation to draw conclusions about differences that are just in the noise (David Johnson, A Theoretician's Guide to the Experimental Analysis of Algorithms, www.research.att.com/~dsj/papers/experguide.ps)

- A diagram does not have to be too faithful to every detail of the concept being illustrated; fine details can always be clarified in the supporting text and even the best diagram requires some explanation
- Edward R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, 1992.

Tables

- Small tables can be part of the running text, displayed in the same way as mathematics.
- Larger tables should be labelled and positioned at the top or bottom of a page.
- Each figure and table should be numbered to allow easy reference and have a descriptive caption so that the figure is, as far as possible, independent of the text.

Experiments' results

- When describing the outcomes of an experiment, don't just compile dry lists of figures or a sequence of graphs.
 - Analyse the results and explain their significance,
 - select typical results and explain why they are typical,
 - theorise about anomalies,
 - show why the results confirm or disprove the hypothesis, and
 - make the results interesting.
- □ That is, motivate the work.

Theorems and proofs

- Essential in theory papers, optional in practical papers
- What is a theorem?
 - Lemma: useful fact that will be used later. Too easy/ specific to be a theorem
 - Fact, Observation: baby lemmas
 - Theorem: an important and general fact that requires proof

KRDE

Mathematical writing

- Good resource: Knuth, Larrabee, and Roberts book on <u>Mathematical Writing</u>
 - Don't punctuate math symbols (real examples)
 - "There are 23⁵ 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.

Abbreviations

- Do not use common blackboard abbreviations.
- For example, write "if and only if" rather than "iff", and "without loss of generality" rather than "WLOG".
- □ This also applies to symbols such as ∀ and ∃.
 Unless one is writing a paper in mathematical logic, one should write out "for all" and "there exists".

Latin abbreviations



Abbreviation	Latin term	English translation
i.e.	id est	that is
e.g.	exempli gratia	for example
cf.	confer	compare
n.b.	nota bene	note well (or just note)
q.v.	quod vide	which see
q.v. viz.	videlicet	namely
et al.	et alii	and others

Contractions



- Do not use contractions in formal writing.
- □ Thus words such as "don't", "can't", "l'm", and "we've" should be written out.

Anaphoric references

- Contextual information can be forgotten between paragraphs, and references between paragraphs can be difficult to follow.
- For example, if a paragraph discusses a fast sorting algorithm, the next paragraph should not begin "This algorithm" but rather "The fast sorting algorithm".
- Check carefully for ambiguity. It is often hard to detect in your own text because you know what is intended.

Verb tense

- In science writing, most text is in past or present tense. Present tense is used for eternal truths. Thus we write "the algorithm has complexity 0(n)", not "the algorithm had complexity 0(n)". Present tense is also used for statements about the text itself. It is better to write "related issues are discussed below" than to write "related issues will be discussed below".
- Past tense is used for describing work and outcomes. Thus we write "the ideas were tested by experiment", not "the ideas are tested by experiment". It follows that occasionally it is correct to use past and present tense together.

Use of words

- Qualifiers such as "very" and "quite" should be avoided altogether, because they are in effect meaningless.
- Other words of this kind are "totally", "completely", "truly", "highly", "usually", "accordingly", "certainly", "necessarily", and "somewhat".
- Use "which" only when it cannot be replaced by "that".
- Use "may" to indicate personal choice, and "can" to indicate capability.
- Avoid exclamation marks!

Dash



- Note that there are three different "dash" symbols:
 - the hyphen "-" used for joining words,
 - the minus sign or en-dash "-" used in arithmetic and for ranges such as "pages 101-127", and
 - the em-dash "—" used for punctuation.

Footnotes



If you think some text should be relegated to a footnote, perhaps it can be deleted.

DON'T use capitals for emphasis



Writing flow

- Make sure your writing flows.
- Avoid writing a succession of loose sentences. Particularly when writing proofs, it is easy to become so engrossed in the mathematics that one forgets to pay attention to English style.
- The result is often a proof that reads "...and then...
 and then...and then...".
- □ Try to use a variety of words in proofs, such as "therefore", "consequently", "it follows that", "we see", "hence", or "thus".

Primadonna

- Don't make excessive claims about your own work.
- Phrases such as "our method is an ideal solution to these problems" or "our work is remarkable" are not acceptable

Debugging your paper

- Read it
- Use a spell-checker
 - Readers tend to judge statements to be wrong if they contain numerous spelling errors
- □ Have other people read it how?

Check for <u>COHERENCE</u>

Order

- Organize your paper in an order that makes the exposition clear.
- □ This will not usually be the order of discovery.
- Often when proving theorems one will first obtain a collection of results, and then later prove a theorem or create a theory which encapsulates these results as special cases.
- When writing up these results, one may want to first prove a general theorem and then obtain the special cases as corollaries.
- On the other hand, it may be appropriate to begin with a few specific examples which identify the important concepts and motivate the more general work to follow.
- (In either case, however, one would not want to prove a specific theorem first, and then a more general theorem later, since this would result in unnecessary repetition.)

Linearity



- At each part of a paper you should consider
 - what the reader has learnt so far,
 - whether this knowledge is sufficient to allow understanding of what follows, and
 - whether each part follows from what has already been said.

KRD

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



- □ 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 capitalised 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 capitalised in English unless they are enumerated (see Section 4).

- □ Are the titles and headings consistent with the content?
- Have all terms been defined?
- Is the style of definition consistent?
- For example, were all new terms introduced in italics, or only some?
- Has terminology been used consistently?
- Are defined objects always described in the same way?
- For example, if the expression "all regular elements E" has been used, is "regular" implicit in the expression "all elements E"?
- Are abbreviations and acronyms stated in full when first used?



- Are the full statements subsequently used unnecessarily?
- Are any abbreviations used less than, say, four times?
- □ If not, can they be removed?
- Do all headings have maximum or minimum capitalization?
- Has a term been capitalized in one place and not in another'?
- Is the style and wording of headings and captions
- consistent?
- □ Are names always used in the same way?
- Has a consistent convention been used for the formation of new names?
- Is spelling consistent?

- □ What about "-ise" versus "-ize", "dispatch" versus "despatch", or "disc" versus "disk"?
- □ Is tense used correctly?
- □ Are references discussed in a consistent way?
- Have bold and italic been used logically?
- Are any words hyphenated in some places but not others?
- □ Have units been used logically?
- If milliseconds have been used for some measurements and microseconds for others, is there a logical reason for doing so?
- Is the reason clear to the reader?
- Has "megabyte" been written as "Mb" in some places and "Mbyte" in others?
- Are all values of the same type presented with the same precision?

- Are the graphs all the same size?
- □ Are the axis units always given?
- If, say, the x-axes on different graphs measure the same units, do the axes have the same label?
- □ Are all tables in the same format?
- Does the use of double and single lines follow a logical pattern?
- Are units given for every value?
- Are labels and headings named consistently?
- If, say, columns have been used for properties A to E in one table, have rows been used elsewhere?
- That is, do all tables have the same orientation?
- □ Has the same style been used for all algorithms and programs?

- □ Is there a consistent scheme for naming of variables?
- Do all pseudocode statements have the same syntax?
- Is the use of indentation consistent?
- In the references, has each field been formatted consistently?
- Have italics and quotes been used appropriately for titles?
- Is capitalization consistent?
- Are journal and conference names abbreviated in the same way?
- Is the style of author names consistent?
- Has the same core set of fields been provided for each reference of the same type?
- Is formatting consistent?



- Are some displays centred and others indented'?
- Do some sections begin with an unindented paragraph and others not'?
- Do the parentheses match?

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

A place for every paper & every paper in its place

Conferences:

- SIGMOD, VLDB, ICDE
- PODS, ICDT
- DBWeb other workshops

_

Journals

- Computing Surveys, JACM
- PVLDB,TODS
- IEEE TKDE
- CACM, SIGMOD Record

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

68

- Rant to your family, friends, neighbours...
- □ 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

Rule I: 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 I, Comment I.I, I.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



- Book
 - Justin Zobel, Writing for Computer Science: The Art of Effective Communication. Springer, 1997.
 - 2nd edition (April 27, 2004)

A final thought....



ADDRESSING REVIEWER COMMENTS BAD REVIEWS ON YOUR PAPER? FOLLOW THESE GUIDE-

Reviewer comment:

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

How NOT to respond:

X "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:

X"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

Acknowledgments

- □ This lecture contains material from:
 - Alberto Mendelzon
 - Renée Miller
 - Richard Zanibbi
 - Justin Zobel
 - Dimitri Bertsekas
 - Mark Tomforde
 - ...