Scientific Writing

CMPS 7010 Research Seminar

References

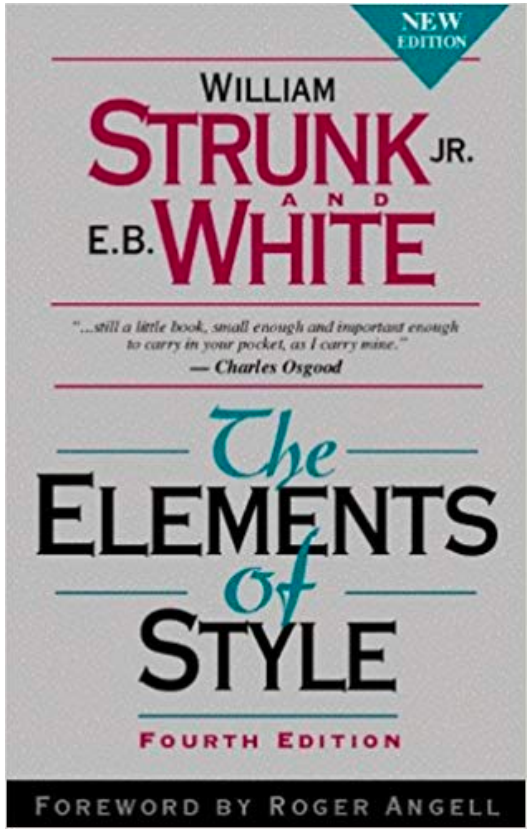
- Mathematical Writing by Donald E. Knuth, Tracy Larrabee, and Paul M. Roberts [video]
- Ten Simple Rules for Mathematical Writing by Dimitri Bertsekas
- The Science of Scientific Writing by George D. Gopen and Judith A. Swan
- How to write a technical paper by Michael Ernst
- The Elements of Style by Strunk and White
- The Not So Short Introduction to LATEX 2ε by Tobias Oetiker

Outline

- English Writing
- Mathematical Writing
- How to write a research paper in CS
- Tools
  - LaTeX and related tools
  - Figures
  - Word choice and grammar checking
  - Paper X-Ray

English Writing

- Everyone should read



Mathematical Writing

1. Symbols in different formulas must be separated by words.

Bad: Consider S_q, $q < p$.

Good: Consider S_q, where $q < p$.

2. Don't start a sentence with a symbol.

Bad: $x^n - a$ has n distinct zeroes.

Good: The polynomial $x^n - a$ has n distinct zeroes.

Mathematical Writing

3. Don’t use the symbols $\therefore, \Rightarrow, \forall, \exists, \ni$; replace them by words

4. The statement just preceding a theorem, algorithm, etc., should be a complete sentence or should end with a colon

Bad: We now have the following

Theorem. $H(x)$ is continuous.

Good: We now have the following result.

Theorem. The function $H(x)$ defined in (5) is continuous.

Better: Tie the theorem up with the previous discussion

Mathematical Writing

5. The statement of a theorem should usually be self-contained

6. The word “we” is often useful to avoid passive voice

- “we” means “you and me together”
- “I” should be avoided in most technical writing, unless the author’s persona is relevant

8. Don’t omit “that” when it helps the reader to parse the sentence.

Bad: Assume A is a group.

Good: Assume that A is a group.

- But never say “We have that $x = y$,” say “We have $x = y$.”

Mathematical Writing

9. Vary the sentence structure and the choice of words, to avoid monotony. But use parallelism when parallel concepts are being discussed.

Bad: Formerly, science was taught by the textbook method, while now the laboratory method is employed.

Good: Formerly, science was taught by the textbook method; now it is taught by the laboratory method.

11. All variables must be defined, at least informally, when they are first introduced

Mathematical Writing

12. Motivate the reader for what follows
    - When describing your own work, be humble and don’t use superlatives of praise
14. Don’t use the same notation for two different things;
    Use consistent notation for the same thing
19. Capitalize names like Theorem 1, Lemma 2, Algorithm 3, Method 4.

Mathematical Writing

22. “which” vs. “that”: use “which” only when it is preceded by a comma or by a preposition, or when it is used interrogatively

Bad: Don’t use commas which aren’t necessary.

Better: Don’t use commas that aren’t necessary

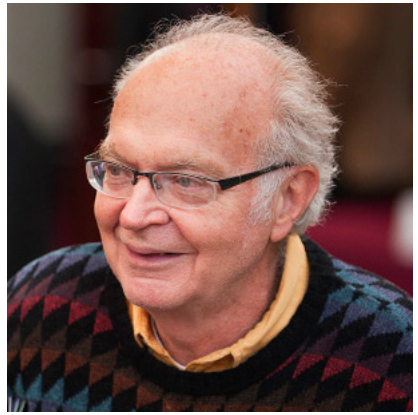
24. Don’t say “An x is y.”

Bad: An important method for internal sorting is quicksort.

Good: Quicksort is an important method for internal sorting, because . . .

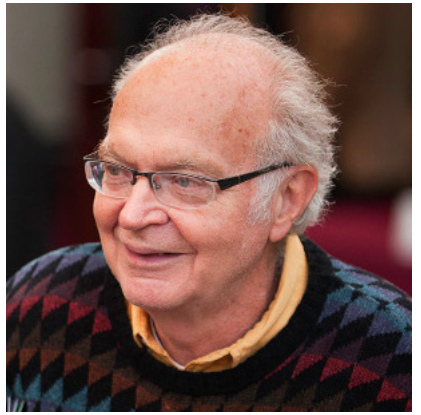
Excerpts from “Mathematical Writing”

- Computer Scientists must be careful to distinguish between mathematical notation and programming language notation.
- While it may be appropriate to use $p[r]$ in a program, in a formal paper it is probably better to use p with a subscript of r.
- As another example, it is not appropriate to use a star ($*$) to denote multiplication in a paper about mathematics. Just say xy.



Excerpts from “Mathematical Writing”

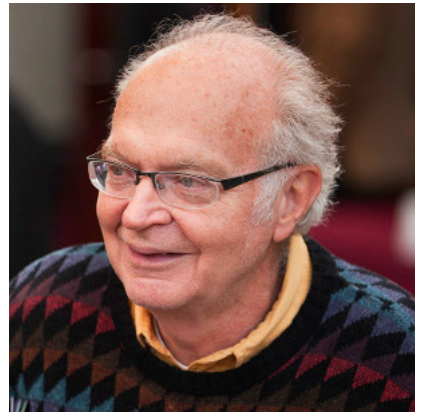
- When you use ellipses, such as $(P_1, \ldots, P_n)$, remember to put commas before and after the three dots.
- When placing ellipses between commas the three dots belong on the same level as the commas, but when the ellipsis is bracketed by symbols such as ‘+’ or ‘<’ the dots should be at mid-level.



- Numbering all displayed formulas is usually a bad idea; number the important ones only.

Excerpts from “Mathematical Writing”

- Past or present tense
- Knuth (hereafter known as Don) recommends that one of two approaches be used with respect to tenses of verbs: Either use present tense throughout the entire paper, or write sequentially. Sequential writing means that you say things like, “We saw this before. We will see this later.” The sequential approach is more appropriate for lengthy papers. You can use it even more effectively by using words of duration: “We observed this long ago. We saw the other thing recently. We will prove something else soon.”



Outline

- English Writing
- Mathematical Writing
- How to write a research paper in CS
- Tools
  - LaTeX and related tools
  - Figures
  - Word choice and grammar checking
  - Paper X-Ray

Why Should You Write Papers?

- To communicate the new knowledge you created
  - Keep the readers in mind
- To clarify your thinking
  - Help you understand your ideas and contributions more deeply
  - Identify future directions
  - Improve your research
- To get a degree/job

How to Write a Research Paper

- Learn from good papers in your area
  - but don’t plagiarize
- Keep readers/reviewers in mind
  - understand and meet the expectations of your community
- Present your argument logically and smoothly
  - State the main results before going to details
  - Provide “roadmap” of your argument
  - Use examples: e.g., a running example throughout the paper

How to Write a Research Paper

- Proofread
  - TYPOS ARE DEATH: use grammar checking tools
- Avoid words you don’t know
- Revise and Rewrite
- Solicit feedback
  - Learn from reviews (but reviewers can be wrong)
- Collaborate with more experienced researchers

Structure of a Research Paper in CS

Milestone 2

- Abstract
- Introduction
- Background/Related work
- System model & Problem formulation
- Solutions
- Evaluation
- Related work
- Conclusion
- Acknowledgement
- References
- Appendix

Introduction

- Background (1 paragraph)
- Motivation
  - What's the problem and why is it important
  - What are the challenges (why is it hard?): a simple example may help
  - Weakness of prior work
- Summarize your methods and results
- Give a bulleted list of main contributions (1 paragraph)
- Provide a roadmap for the rest of the paper (1 paragraph)

[Example 1]

➢When to write the introduction section?

[Example 2]

Solutions

- Explain “what”
  - give pseudocode and explain it line by line; give complexity analysis
  - give proofs (or proof sketches)
  - give examples/figures
- Explain “why”
  - What makes your algorithms/proofs work?
  - What are the key insights of your approach?
  - Are there any limitation of your approach?

Related Work

- Be comprehensive
  - Help the readers understand the research effort in this area
- Compare and contrast
- Organize related works
  - Works on the same/related problems
  - Works that use the same/similar techniques
  - …
- [Example]
- When to write the related work section?

Figures

- “a picture is worth a thousand words”
- A figure should stand on its own [example 1][example 2]
  - Use captions/legends/labels to explain the context/symbols/content/
- Use large enough font size
- Put figures (tables, algorithms, etc.) at the top of the page

Paper Writing Ethics

- Start early
- Have a writing plan
- Work with your advisor
  - Be responsive
  - Never send an unpolished draft to your advisor
  - The first complete draft should be ready at least 2 weeks before the deadline
  - Learn from your advisor's feedback & revision
- Adhere to the highest standards of integrity

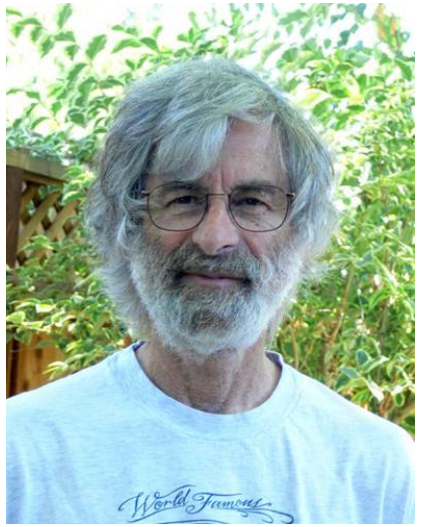
Outline

- English Writing
- Mathematical Writing
- How to write a research paper in CS
- Tools
  - LaTeX and related tools
  - Figures
  - Word choice and grammar checking
  - Paper X-Ray

LaTeX – A document preparation system

- TeX is created by Donald Knuth in 1978
- LaTeX is created by Leslie Lamport in 1983

By Flickr user Jacob Appelbaum, uploaded to en.wikipedia by users BeSherman, Duozmo - Flickr.com (via en.wikipedia), CC BY-SA 2.5, https://commons.wikimedia.org/w/index.php?curid=1303242



Copyrighted free use, https://commons.wikimedia.org/w/index.php?curid=487587

LaTeX – A document preparation system

- Pronunciation
  - Leslie Lamport: “One of the hardest things about LaTeX is deciding how to pronounce it. This is also one of the few things I'm not going to tell you about LaTeX, since pronunciation is best determined by usage, not fiat. TeX is usually pronounced *teck*, making *lah*-teck, and *lay*-teck the logical choices; but language is not always logical, so *lay*-tecks is also possible.”
  - https://www.youtube.com/watch?v=QGWG5Ajbq8o
  - https://www.youtube.com/watch?v=8HuwiBPLV3A

LaTeX – A document preparation system

- LaTeX is NOT a WYSIWYG word processor
  - WYSIWYG ("what you see is what you get"): Microsoft Word, Apple Pages
- LaTeX is
  - a high-level, descriptive markup language built on top of TeX
  - Tex – typesetter; LaTeX – typographic designer
  - separating presentation/style from the content
  - extensible: 5753 packages in CTAN (The Comprehensive TEX Archive Network)
  - the de facto standard for scientific writing

LaTeX – A document preparation system

- TeX distributions: MiKTeX, TeX Live, MacTeX, etc.
  - essential tools: BibTeX, pdfTeX, simple text editor, etc.
  - package management

- LaTeX editors: WinEdt (Windows only), TeXstudio, TeXShop, Overleaf (online), etc.
  - spell checking, syntax highlighting, access to common math symbols/text formatting, …
  - some are partial-WYSIWYG

Figures

- Build “publication-quality” figures
- data plots
  - plot packages: Python, MATLAB, R, etc.
  - gnuplot
- images
  - WYSIWYG: PowerPoint, Microsoft Visio, …
  - MetaPost: generating figures from an algebraic/geometric description
  - bitmap vs. vector images

Paper Templates

- IEEEtran LaTeX class
  - Ex: \documentclass[10pt, conference, letterpaper]{IEEEtran}
- ACM LaTeX template
- Check conference/journal websites for templates/author kits

Grammar Checking

- English-Corpora.org
- Grammarly

Paper X-Ray

- https://www.acemap.info/