How to Write a Research Paper in Computer Science

Compiled by the ICCL Team
International Center for Computational Logic
Technische Universität Dresden
Germany

Sources, Literature and Links

- **Alan Bundy: How to Write an Informatics Paper**
  `http://homepages.inf.ed.ac.uk/bundy/how-tos/writingGuide.html`
- **Donald E. Knuth, Tracy Larrabee, and Paul M. Roberts: Mathematical Writing Research Report, Stanford University**
  `http://www-cs-faculty.stanford.edu/~knuth/papers/cs1193.pdf`
  Video tapes: `http://scpd.stanford.edu/knuth/index.jsp`
- **Instructions for Authors, LaTeX 2ε class for Lecture Notes in Computer Science**
  `http://www.springer.com/lncs`
- **Alan Bundy: Notes to CADE-12 Referees (1997)**
  `http://homepages.inf.ed.ac.uk/bundy/how-tos/referee-notes.pdf`

Part I: The Structure of a Research Paper

The Importance of Hypotheses

- It is not enough to describe some new technique or system, some claim about it must be stated and evaluated
- In experimental research, hypotheses typically take one of these two forms:
  - Technique/system X automates task Y for the first time
  - Technique/system X automates task Y better, along some dimension, than each of its rivals, where the dimensions are typically:
    - Behavior: X has a higher success rate or produces better quality outputs than Y
    - Coverage: X is applicable to a wider range of examples then Y
    - Efficiency: X is faster or uses less space then Y
    - Dependability: X is more reliable, safe or secure than its rivals
    - Maintainability: X is easier to adapt and extend than its rivals
    - Usability: Users find X easier to use than its rivals
- In theoretical papers, the hypotheses are the statements of theorems and the supporting evidence is their proofs

Default Structure of an Experimental Research Paper

- Title
- Abstract
- Introduction
- [Literature Survey]
- [Background]
- [Notation and Preliminaries]
- [Theory]
- [Specification]
- [Implementation]
- Evaluation
- [Related Work]
- [Further Work]
- Conclusion
- [Appendices]

Title and Abstract

- **Title**
  - **summarizes the hypothesis of the paper**

  The reader should be able to work out what the paper is about from the title
- **Abstract**
  - **states the hypothesis**
  - **summarises the evidence that supports or refute it**
  - **mentions the key contributions of the paper**

  Most readers will not read beyond the abstract

Introduction

- motivates the contribution of the paper
- places it in context
- Questions to be addressed in a succinct and informal manner:
  - What is the problem?
  - Why is the problem important?
  - What has so far been done on the problem?
  - What is the contribution of the paper on the problem?
  - Is the contribution original? Explain why
  - Is the contribution non-trivial? Explain why
- ends with a short summary of the rest of the paper
  The rest of the paper is structured as follows: In Section 2 we ...

Optional Introductory Sections

- **Literature Survey**
  - **is a broad and shallow account of the field, which helps to place the contribution of the paper in context**
    - **What are the rival approaches?**
    - **What are the drawbacks of each?**
    - **How has the battle between different approaches progressed?**
    - **What are the major outstanding problems? (This is where you come in)**
- **Background**
  - **describes previous work in more technical detail,**
  - **as far as needed for a proper understanding of the contribution of the paper**
- **Notation and Preliminaries**

Optional Sections of the Main Part

- **Theory**
  - describes the underlying theory of techniques or system
  - where appropriate, uses a mathematical style of definitions, lemmas, propositions, theorems, corollaries, remarks
  - illustrates the main definitions and theorems with simple but meaningful examples
- **Specification**
  - formally specifies techniques that underlie the implementation
  - states the requirements of the implementation
- **Implementation**
  - describes only the final state of the implementation
  - identifies the major design decisions and gives their reasons
  - describes the overall structure of the system and key algorithms in abstract form
  - illustrates the main algorithms with simple but meaningful examples

The Evaluation

- gathers evidence to support or refute the hypothesis
  - Technique/system X automates task Y for the first time
  - Technique/system X automates task Y better, along some dimension, than each of its rivals, where the dimensions are typically:
    - Behavior: X has a higher success rate or produces better quality outputs than Y
    - Coverage: X is applicable to a wider range of examples then Y
    - Efficiency: X is faster or uses less space then Y
    - Dependability: X is more reliable, safe or secure than its rivals
    - Maintainability: X is easier to adapt and extend than its rivals
    - Usability: Users find X easier to use than its rivals
- Experimental results are often best presented graphically

Optional Sections of the Concluding Part

- **Related Work**
  - **compares narrow but deeply between system X and its main rivals at their critical points of difference**
  - **belongs “logically” to the evaluation since it establishes the originality of the contribution**
  - **is different in purpose, position, breadth and depth from the literature survey**
- **Further Work**
  - **describes future research and new directions suggested by the contribution**
  - **in particular, research that would improve the evidence for/against the hypothesis**

Conclusion

- summarizes the research and discusses its significance
  - The hypothesis and the evidence for and against it are briefly restated
  - The original motivation is recapitulated
  - The state of the field in the light of this new contribution is reassessed

Appendices

- provide information whose inclusion could assist the reader in understanding or assessing the research, but
  - would detract from the flow of the main body of the paper
  - does not fit into a prescribed page number
- contain for example
  - a glossary of technical terms
  - technical background that only some readers may require
  - examples of program code
  - a trace of the program on one or more examples
  - more details of the examples evaluated and the experimental results
  - the full versions of proofs
  - an index

Part II: Important Points in Writing a Research Paper

Capitalization and Non-Capitalization

- Depends on the style, here we consider LNCS
- The following should always be capitalized:
  - All words in headings except, unless at the beginning:
    - conjunctions, prepositions
      on, of, by, and, or, but, from, with, without, under, ...
    - definite and indefinite articles
      the, a, an
  - Expressions with numbers such as Table 5, Sect. 2, Theorem 2
- The following should not be capitalized:
  - The words figure(s), table(s), equation(s), theorem(s) in the text without accompanying number
  - Figure legends and table captions except for names and abbreviations

Abbreviations

- **If abbreviations of names or concepts are used throughout the text, they should be defined at first occurrence**
  The Davis-Putnam-Logeman-Loveland Procedure (DPLL) is a ...
- **Chap(s).**, **Sect(s).**, **Fig(s).** **should always be abbreviated unless at beginning of sentence (in LNCS style)**

Miscellaneous

- **Ensure that expressions like “formula *F*” or “Sect. 1” are not disrupted by a line break: “formula~ F”, “Sect.~\ref{sec-introduction}”.**
- **Only make assertions that are true and can be verified**

  Bad: This has never been considered before.
  Good: The authors are not aware of any other work where this has been considered.

  Bad: All other works apply the slow method.
  Good: In [2,4,8] the slow method is applied.
- **Before you give a draft to your colleagues for proof reading:**
  - **Remove typos and incoherent notation because these distract the reader’s attention**
  - **Read it by yourself**

Some Important Points in Mathematical Writing (1)

- **Symbols in different formulas must be separated by words.**
  Bad: Consider S_q, $q < p$.
  Good: Consider S_q, where $q < p$.
- **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.
- **Don't use the symbols $\Rightarrow$, $\forall$, $\exists$, $\wedge$; replace them by the corresponding words (except in logic formulas, of course).**
- **The statement of a theorem should usually be self-contained, not depending on the assumptions in the preceding text.**
- **The word "we" is often useful to avoid passive voice. It should be used in contexts where it means "you and me together". "I" should be avoided, unless the author's persona is relevant.**
- **Read what you have written, and change the wording if it does not flow smoothly.**

Some Important Points in Mathematical Writing (2)

- 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.
- Vary the sentence structure and the choice of words, to avoid monotony. But use parallelism when parallel concepts are being discussed.
- Don't use the style of homework papers, in which a sequence of formulas is merely listed. Tie the concepts together with a running commentary.
- Try to state things twice, in complementary ways, especially when giving a definition. This reinforces the reader's understanding.
- Motivate the reader for what follows.
- Many readers will skim over formulas on their first reading of your exposition. Therefore, your sentences should flow smoothly when all but the simplest formulas are replaced by "blah" or some other grunting noise.
- Resist the temptation to use long strings of nouns as adjectives: consider the packet switched data communication network protocol problem. In general, don't use jargon unnecessarily. Even specialists in a field get more pleasure from papers that use a nonspecialist's vocabulary.

Some Important Points in Mathematical Writing (3)

- **Don't use the same notation for two different things.**

  Use consistent notation for the same thing when it appears in several places. Don't say "A_j for $1 \leq j \leq n$" in one place and "A_k for $1 \leq k \leq n$" in another place.

  Typographic conventions (like lowercase letters for elements of sets and uppercase for sets) are also useful.

- **Set element notation can be used to avoid subscripted subscripts. It is often troublesome to start out with a definition like "Let $X = \{x_1, \ldots, x_n\}$" if you're going to need subsets of X, since the subset will have to defined as $\{X_{i_1}, \ldots, X_{i_m}\}$, say. Also you'll need to be speaking of elements x_i and x_j all the time. Don't name the elements of X unless necessary. Then you can refer to elements x and y of X in your subsequent discussion, without needing subscripts; or you can refer to x_1 and x_2 as specified elements of X.**

- **Display important formulas on a line by themselves. If you need to refer to some of these formulas from remote parts of the text, give reference numbers to all of the most important ones, even if they aren't referenced.**

Some Important Points in Mathematical Writing (4)

- **Sentences should be readable from left to right without ambiguity. Bad examples:** "Smith remarked in a paper about the scarcity of data." "In the theory of rings, groups and other algebraic structures are treated."

- **Small numbers should be spelled out when used as adjectives, but not when used as names.**

  Bad: The method requires 2 passes.
  Good: Method 2 is illustrated in Fig. 1; it requires 17 passes.
  The count was increased by 2.
  The leftmost 2 in the sequence was changed to a l.

- **The opening paragraph should be your best paragraph and its first sentence should be your best sentence. If the beginning flows smoothly, the reader will be hooked and won't notice occasional lapses in the later parts. Probably the worst way to start is with a sentence of the form** "An x is y." **For example,**

  Bad: An important method for internal sorting is quicksort.
  Good: Quicksort is an important method for internal sorting, because
  Bad: A commonly used data structure is the priority queue.
  Good: Priority queues are significant components of the data structures needed for many different applications.

Part III: Criteria for Assessing the Quality of the Presentation

- **Organization**
- **Readability**
- **English**

Assessing the Organization of the Presentation

- Is all the material present?
- Is any material unnecessary?
- Is the material in the right order?
- Is the paper self-contained?
- Are the title, abstract and conclusion an accurate precis of the work done?
- Is terminology defined before it is used?
- Is there an adequate use of diagrams, examples and formalism?
- Is related work discussed?
- Is the bibliography adequate?

Assessing the Readability of the Presentation

- Is the meaning always clear?
- Are the arguments crisp or woolly?
- Are there any cryptic or obtuse sentences?
- Does the paper ramble?
- Did the paper hold your attention?

Assessing the English of the Presentation

- Are there any spelling errors?
- Are there any typographical errors?
- Are there any grammatical errors?

Summary

- The importance of the hypothesis
- Default structure:
  Title / Abstract / Introduction / [Literature Survey] / [Background] / [Notation and Preliminaries] / [Theory] / [Specification] / [Implementation] / Evaluation / [Related Work] / [Further Work] / Conclusion / [Appendices]
- Important details:
  Capitalization / Abbreviations / Mathematical Writing
- Quality criteria of a written presentation:
  Organization / Readability / English