How to Write a Research Paper in Computer Science

Compiled by Christoph Wernhard International Center for Computational Logic Technische Universität Dresden Germany

Version: 6 October 2014 16:00

- Sources, Literature and Links
- Preliminary Note: The Idea of Formalized Scientific Knowledge
- I: The Structure of a Research Paper
- II: Miscellanous Important Points
- III: Important Points in Mathematical Writing
- IV: Citations and Bibliography
- V: Criteria for Assessing the Quality of the Presentation
- Summary

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

Preliminary Note: The Idea of Formalized Scientific Knowledge

```
Definition 1 \forall x \forall y \text{ subset}(x,y) \leftrightarrow_{\text{def}} \forall z \ z \in x \rightarrow z \in y.

Definition 2 \forall x \forall y \forall z \ x \in \text{union}(y,z) \leftrightarrow_{\text{def}} x \in y \lor x \in z.

Theorem 1 \forall x \forall y \text{ subset}(x, \text{union}(x,y)).

Proof

1. \forall x \forall y \text{ subset}(x, \text{union}(x,y)) \text{ theorem statement}

2. iff \forall z \ z \in x \rightarrow z \in \text{union}(x,y) \text{ expanding Def. 1}

3. iff \forall z \ z \in x \rightarrow z \in x \lor z \in y \text{ expanding Def. 2}

4. iff true. logical equivalence

Theorem 2 \forall x \text{ subset}(x, \text{union}(x,x)).

Proof
Follows immediately from Theorem 1.
```

Theoretical Research

- ► Definitions introduce concepts
- ► Theorems make claims about concepts
- Proofs verify theorems

Practical Research

- ► Techniques or systems are defined
- ► Hypotheses make claims
- ▶ Empirical evaluation

A network emerges, created by many people over a long period of time

Part I: The Structure of a Research Paper

- The Importance of Hypotheses
- Default Structure of an Experimental Research Paper
 - ▶ Title and Abstract
 - Introduction
 - Optional Introductory Sections
 - Optional Sections of the Main Part
 - Theorem Terminology
 - ▶ The Evaluation
 - Optional Sections of the Concluding Part

4

- Conclusion
- Appendices
- Excercise

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

Theorem Terminology

Definition

A statement (definiens) that specifies the meaning of a term (definiendum)

▶ Theorem

A statement that follows from definitions and other theorems

- ▶ Comes in different kinds, distinguished by role; Distinction is often blurred
 - Theorem (in the narrow sense)
 Important theorem, a result of the paper
 - Lemma Important stepping stone used to prove important theorems
 - Proposition Theorem of no particular importance (e.g., stating a property of a concept)
 - Corrollary Theorem that follows with little of no proof from a definition or theorem
- Proof
- ► Remark

Can be used to introduce implicit preconditions, introduce ways of speaking, clarify properties

▶ Example

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

Conclusion: 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]

Exercise

- Inspect some papers from your field with respect to the presented issues:
 - What is the hypothesis?
 - How is it expressed?
 - What are the key contributions of the paper?
 - How and where are they expressed?
 - Do they seem original and non-trivial?
 - What is the context of the paper?
 - How is the paper placed into it?
 - ▶ How are the claims made in the paper evaluated?

Part II: Miscellaneous Important Points

- Capitalization and Non-Capitalization
- ► Abbreviations, References to Numbered Items
- ▶ Your Assertions must be True and Verifiable
- General Hints

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, References to Numbered Items

- 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)
- ► Ensure that expressions like "formula F" or "Sect. 1" are not disrupted by a line break: "formula \$F\$", "Sect. \ref{sec-introduction}".

Your Assertions must be True and Verifiable

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.

General Hints

- 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



Part III: Important Points in Mathematical Writing

Some Important Points in Mathematical Writing (1)

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.

3. Don't use the symbols \Rightarrow , \forall , \exists , \land ; replace them by the corresponding words (except in logic formulas, of course).

Possible exceptions: Symbols \Leftarrow and \Rightarrow for "Left-to-right" and "Right-to-left" in bullet position to indicate subproofs. Symbols \forall , \exists , \land if consistently and unambiguously used, e.g., in tabular proof presentations.

- 4. The statement of a theorem should usually be self-contained, not depending on the assumptions in the preceding text.
- 5. 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. Even in that case "the author" or "the first author" might be better.

Some Important Points in Mathematical Writing (2)

- Read what you have written, and change the wording if it does not flow smoothly.
- 7. 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.

- 8. 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."
- 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.

10. 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.

Some Important Points in Mathematical Writing (3)

- 11. Motivate the reader for what follows.
- 12. Keep the reader in mind:
 - What does the reader know so far?
 - What does the reader expect next and why?
 - ▶ How could the reader misunderstand me here?
- 13. Try to state things twice, in complementary ways, especially when giving a definition. This reinforces the reader's understanding.
- 14. All variables must be defined, at least informally, when they are first introduced.
- 15. 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.
- 16. 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.

```
\[\] \begin{equation} \end{equation}
```

Some Important Points in Mathematical Writing (4)

17. 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 $1 \le j \le n$ " in one place and " A_k for $1 \le k \le n$ " in another place.

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

- 18. 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.
- 19. 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 (5)

 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.

21. 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 IV: Citations and Bibliography

- Purposes of a Citation
- A Citation Gives Credit
- Some Ways to Embed Citations in the Text
- ▶ The Components of BibTeX
- Some Important LaTeX Bibliography Styles
- Verbosity vs. Brevity
- About Some Types of References Supported by BibTeX
- Referencing Online Resources
- Possible Conventions
- Hints

Purposes of a Citation

- Giving Credit
 - Upholding intellectual honesty
 - Attributing prior or unoriginal work and ideas to the correct sources
 - Acknowledging the relevance of the works of others to the topic at the spot where the citation appears
- ▶ Helping the reader to identify and relocate the source work
 - Allowing the reader to determine independently whether the referenced material supports your argument in the claimed way
- Providing evidence that your position is well-researched
 - Helping the reader to gauge the strength and validity of the material you have used

A Citation Gives Credit

- Proper citations should be given for all material that is not by yourself, except common knowledge as can be found in standard textbooks
- Cite the original publication about the referenced work
 Do not cite where you have first read about it (e.g., Wikipedia, a textbook)
 Citing a work does not imply that you have actually read it
- Sometimes there are different versions of a work. The "most valuable" version is typically the main reference to be cited:
 - journal article > conference proceedings > workshop proceedings > technical report
 - It can be useful for historic accuracy to cite also earlier versions
- ► Citation indexes are used as a measure for scientific excellence, e.g.:
 http://academic.research.microsoft.com/Author/1085332/franz-baader

Some Ways to Embed Citations in the Text

- So far, improvements have been considered, such as symmetry breaking [2,6], and...
- As described in [2], ...
- However, the approach of [2] is superior in the case where...
- As Drescher et al. [2] describe,...

The Components of BibTeX

- ► A database (BibTeX file) with entries for publications
- ► The \cite command to make citations
- ► The \bibliographystyle command to specify how references and the bibliography should be rendered
- The \bibliography command: it is automatically replaced by the bibliography, which contains just the cited items in nicely rendered and properly sorted form

Some Important LaTeX Bibliography Styles

- In the mentioned styles, the bibliography is sorted according to author name, year, title
- ▶ plain: References like [1], [2], [3]
- ▶ abbrv: Similar to plain, but all first names of authors are abbreviated
- ▶ alpha: References like [Dre11], [DE11], [DEF+11b]
 - Helps the reader to get an idea about authors and date of the referenced works
- For examples of these and further styles see: http:
 - //amath.colorado.edu/documentation/LaTeX/reference/faq/bibstyles.pdf

Verbosity vs. Brevity

- Depends on where it is published
- Consistency among your references is important
- ▶ Found BibTeX entries (e.g., DBLP, CiteSeer) often have to be adapted
- Verbose version
 - [DEF+11f] Christian Drescher, Thomas Eiter, Michael Fink, Thomas Krennwallner, and Toby Walsh. Symmetry breaking for distributed multicontext systems. In James P. Delgrande and Wolfgang Faber, editors, Logic Programming and Nonmonotonic Reasoning 11th International Conference (LPNMR 2011), volume 6645 of Lecture Notes in Computer Science, pages 26–39. Springer, 2011.
- First names abbreviated; Editors of proceedings volume omitted
 - [DEF+11a] C. Drescher, T. Eiter, M. Fink, T. Krennwallner, and T. Walsh. Symmetry breaking for distributed multi-context systems. In Logic Programming and Nonmonotonic Reasoning 11th International Conference (LPNMR 2011), volume 6645 of Lecture Notes in Computer Science, pages 26–39. Springer, 2011.

Verbosity vs. Brevity (2)

- Common terms abbreviated:
 Acronyms of well known series
 Abbreviated journal names (lists of abbreviations can be found in the Web)
 - [DEF+11b] C. Drescher, T. Eiter, M. Fink, T. Krennwallner, and T. Walsh. Symmetry breaking for distributed multi-context systems. In Logic Programming and Nonmonotonic Reasoning 11th Int. Conf. (LPNMR 2011), volume 6645 of LNCS, pages 26–39. Springer, 2011.
- Short name of the conference
 - [DEF+11b] C. Drescher, T. Eiter, M. Fink, T. Krennwallner, and T. Walsh. Symmetry breaking for distributed multi-context systems. In Logic Programming and Nonmonotonic Reasoning (LPNMR 2011), volume 6645 of LNCS, pages 26-39. Springer, 2011.
- Only the acronym of the conference; Series and publisher omitted
 - [DEF+11d] C. Drescher, T. Eiter, M. Fink, T. Krennwallner, and T. Walsh. Symmetry breaking for distributed multi-context systems. In LPNMR 2011, pages 26–39, 2011.

About Some Types of References Supported by BibTeX

- book author, title, edition (if not the first), [volume], address (i.e. place of publication), publisher, year
- article Article in a journal author, title, journal, volume, number, year, [month], pages
- inproceedings Article in conference proceedings author, title, year, booktitle (i.e. title of the proceedings volume), pages, address, publisher, [series], (possibly also volume editors)
- techreport Technical report author, title, institution, year, [series], [number], [url]
- mastersthesis, phdthesis author, title, school, address, year
- Unpublished work given to you by others:
 Cite with misc as personal communication author, year
- ► Unpublished papers by yourself:
 Add note like submitted or presented at ...

Referencing Online Resources

Adding a URL

BibTeX field url is supported by some styles (see DOI below)

Another possibility is note = "Available \url{http://...}"

This requires the package hyperref

► Adding a DOI (Digital Object Identifier)

http://dx.doi.org/10.1017/S1471068410000220

BibTeX field doi is supported by some styles, e.g., alphaurl http://www.tex.ac.uk/tex-archive/bibliography/bibtex/contrib/urlbst/alphaurl.bst and also the natbib package

- Citing a Web Resource
 - author (if known)
 - title
 - note = Internet: URL, DATE-UPDATED, accessed DATE-ACCESSED (DATE-UPDATED only if known)

Hints

- ▶ Bibliography entries that are not referenced can be produced with \nocite
- ► The proper dash for number ranges is "--" pages = {323--345}
- ▶ Protect proper names etc. in titles of articles against downcasing by BibTeX: {SAT}, {B}oolean

But do not protect the whole title against downcasing

- ▶ Prevent wrong hyphenation of words in other languages Tech\-ni\-sche
- ▶ Prevent pagebreaks in acronyms SAT~14, \mbox{SAT 14}
- ▶ Bibtool is useful to merge and maintain BibTeX databases http://www.ctan.org/tex-archive/biblio/bibtex/utils/bibtool/
- Check the warnings issued by BibTeX
- Use uniform conventions for all entries in your bibliography
- Proof-read also the bibliography

Possible Conventions

- Usually you have to specify some conventions that are consistently obeyed in your or your group's bibliography
- Examples of possible such conventions that apply independently of space restrictions:
 - No address for publishers well-known in computer science such as Springer, Elsevier, Morgan Kaufmann
 - 2. Acronyms for well-known series such as LNCS
 - 3. Subseries LNAI is written LNCS (LNAI) in accord with Springer's practice
 - First names should always be supplied BibTeX generates abbreviations, depending on the style
- Examples of conventions that apply under moderate space restrictions:
 - No editor for conference and workshop proceedings (but of course for other types of collections of papers)
 - 6. Names of conferences (booktitle of inproceedings) written as

Booktitle of the Proceedings Volume (ACRONYM)

Logic Programming and Nonmonotonic Reasoning (LPNMR 2011)

Possible Conventions II

- Debatable examples of possible conventions under moderate space restrictions:
 - 7. Common abbreviations such as Proc., Int., Conf.
 - Established "long" abbreviations for journal names
 J. Autom. Reasoning, Artif. Intell., Theor. Comput. Sci., J. ACM in contrast to JAR, AI, TCS, JACM

PRO: DBLP uses such "long" abbreviations, there are also lists of them, they are understood also by people outside of the core community

- 9. Dropping phrases like Proc. of the 11th Int. Conf. on
 - CON: Titles of books should be respected
- It may be useful to use BibTeX macros @string{Al="Artificial Intelligence"} iournal = Al

CON: Entries are not self-contained (e.g., on the homepage)

Part V: 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
- Citations and bibliography:
 Purposes: giving credit, allowing to relocate, show that work is well-researched
 / Types of references and their support by BibTeX / Need for conventions
- Quality criteria of a written presentation: Organization / Readability / English