Writing the Course Project for CS 581

Tandy Warnow

December 2021

Abstract

This document provides some advice to students in CS 581 on how to write their course project documents. More information on the course project requirements is available online at `https://tandy.cs.illinois.edu/CS581-Fa2021-short.html` or from the instructor.

1 Introduction

This is advice for students in my CS 581 course, who are writing their project reports. This is not advice about how to do research, but only how to write up the report. I remind all students that paraphrasing or copying text is never allowed, unless fully documented. Please review the information in my course webpage at `http://tandy.cs.illinois.edu/academic_integrity.html` about this.

Here is what the course page says about the course project:

> Course project The course requires a final project of each student, and is due on the last day the class meets. You are strongly encouraged to do a research project, but you can also do a survey paper on some topic relevant to the course material. In both cases, your project should be a paper of at least 3000 words (not including the bibliography) in a format and style appropriate for submission to a journal such as Bioinformatics (however, please provide this in a single column format, not double column). Research projects can involve two students, but survey papers must be done by yourself. (Note: When projects are done by two students, the division of the work should be communicated in the write-up, and each student should submit their own course project write-up. Please see me to discuss requirements regarding division of work and write-ups for your specific project, if this applies to you.) Grades on the final project depend upon the kind of project you do. For a research paper, your grade will be 30% writing and 70% content. If you do a survey paper, the grade will be 40% writing and 60% content. In both cases, you should include a thoughtful discussion of the relevant literature and have an appropriate bibliography. Note also

the requirements for reproducibility (for research papers) and the expectations about writing quality.

Since the writing quality is at least 30%, this document is about basic aspects of writing. So you should read your project report carefully before you submit it. And, one main piece of advice: always read the paper as it appears in the PDF rather than just looking at your latex.

Now for some specific recommendations.

2 Specific notes for projects involving two students

If you are writing the report with someone else, you should put your own name and not both names as the author on your report. In your acknowledgments section, you should mention the other person, but the text in your write-up should be done entirely by yourself. If you have a supplementary materials document, that can be written up together and shared, but the main documents should be written up entirely independently (and should not use text from the project proposal, progress reports, etc). Figures and tables can be shared, but the captions should not be shared. See me if you have questions about this.

3 Writing advice

3.1 Principles

Your report should be easy to read, easy to understand, and if you did a study then it should be fully reproducible. Don't assume the reader knows much at all about your subject! You should provide enough information for the reader to be motivated to read the paper, and enough for the reader to understand it. Don't use terminology without definition, for example. Justify all your statements with appropriate references. Just remember: prepare your paper for an audience that is not just your fellow students in the class and me.

Reproducibility is essential, so this means full information on how you generated the data and performed the analyses, and locations for datasets that you used.

3.2 Structure of the paper

As noted in the instructions for the course project, these reports should look like a submission to a journal or a conference, so they should have all the usual sections. This is not a rigid format, however, so you can modify it depending on the type of paper you are writing.

Some points before going further: all the papers will have an Abstract, an Introduction, and nearly all will have a section labelled either "Materials and Methods" or "Study Design". Nearly all will have Results, Discussion, and

Conclusions. Sometimes Results and Discussion are combined into one section. It's important to know what the purpose is of each section.

- Title and author
- Abstract: this is short, usually limited to 200 or so words, and should explain briefly what your paper is about. No citations in the abstract
- Introduction: let the reader know what the problem is that you are addressing. Do not assume they already know that much, so start more or less from the beginning.
- Study Design or Materials and Methods: This should explain enough for the reader to know what you did and interpret the results. (Sometimes full details are provided in supplementary materials, but as much as possible this section should be self-contained.)
- Results: Assuming you ran experiments, this is where you describe the results of each experiment. Typically this has figures and/or tables. Each figure or table should be discussed in detail, in a separate paragraph.
- Discussion: This is where you draw high level insights from the results reported on the experiments you performed. This is also where you make a comparison to the literature: is what you found consistent with prior studies, or different? If different, how do you explain the differences? What high-level understanding do you have get after synthesizing across your study and prior studies? What are the limitations of what you did, and so what is not yet understood?
- Conclusions: This often begins with a summary of what your did and observed, but should go beyond a summary. What are the implications for the rest of the world, given what you learned? What are some future steps to make?
- Acknowledgments (optional): This is where you can acknowledge help you received from anyone.
- Appendix (optional): When provided, this typically provides URLS for data, software, and sometimes lists version numbers and commands for software.
- Bibliography: All papers or webpages that you cite are listed here
- Supplementary materials (optional): These are not provided in the main paper, but require clicking on links to get to. Therefore, do not assume the reader will look at the supplementary materials. These generally provide additional results (figures, tables, etc.) that are not essential for the main paper. These also sometimes provides version numbers and commands for methods that you use, etc. But think of the supplement as something that is not needed in order to understand or evaluate the paper, but might

be needed for reproducibility. In general, it's better to put everything you can that is needed for reproducibility into the main paper (and the appendix is part of the main paper), but it is okay to put the exact version numbers and commands in the supplement instead. Just realize that in a journal or conference paper, the supplement is not part of the publication - it's somewhere else that the reader needs to click on to find. So it's not that easily read. Therefore, don't assume the reader will look at the supplement.

Paper presenting a new method If you are designing a new method, then you probably have either a theoretical result or an empirical study. You may have both! If you have both, then one possible structure for your paper is as follows. (Note that if you don't have any theorems, then you can omit the theorems section, and if you don't have any performance study then you can omit the Study Design and Results sections.)

1. Abstract (one paragraph)
2. Introduction
3. Your new method. Your description of your method needs to be clearly stated in your main paper. This can be made easy by using figures, or bulletized lists for the steps. Be clear about the input/output for the method. If you have theoretical results, then put them here (e.g., statistical consistency, running time, etc.) as theorems with proofs.
4. Study Design
    - Overview of study: What you did at a high level (list of experiments)
    - What methods you compare to
    - What datasets you use
    - What criteria you use
5. Results
6. Discussion
7. Conclusion
8. Acknowledgments (this is an optional section, but it is an opportunity for you to list your collaborators if any, and who helped you with the project if relevant)
9. Appendix (optional)
10. Bibliography
11. Supplement (optional)

Paper comparing existing methods If you are comparing existing methods, then you probably have an empirical study. It is also possible you'll have a comparison with respect to theoretical guarantees. If you have both, then one possible structure for your paper is as follows. (Note that if you don't have any theorems, then you can omit the theorems section, and if you don't have any performance study then you can omit the Study Design and Results sections).

1. Abstract (one paragraph)
2. Introduction
3. Materials and Methods (describe the methods you are comparing and what you are going to do to compare them)
4. Results
5. Discussion
6. Conclusion
7. Appendix (optional)
8. Acknowledgments (optional)
9. Bibliography
10. Supplementary materials (optional)

Paper doing a survey If you are surveying the literature on a problem, then depending on what you are surveying how you structure the paper could change. If you are surveying methods for a specific task, then to a large extent your paper structure will look like that of the previous section. If you are surveying methods that use some specific technique, probably have an empirical study. It is also possible you'll have a comparison with respect to theoretical guarantees. If you have both, then one possible structure for your paper is as follows. (Note that if you don't have any theorems, then you can omit the theorems section, and if you don't have any performance study then you can omit the Study Design and Results sections).

1. Abstract (one paragraph)
2. Introduction
3. Materials and Methods (describe the methods you are comparing and what you are going to do to compare them)
4. Results
5. Discussion
6. Conclusion

7. Appendix (optional)
8. Acknowledgments (optional)
9. Bibliography
10. Supplementary materials (optional)

4 Writing points

This section is about writing advice in general, from very low-level (spelling, etc.) to somewhat higher level (paragraph structure). A major piece of advice is to print the document and look at it how it appears. Don't just look at your latex.

1. Paragraph structure: Each paragraph should have one point, not several. You should divide a paragraph into two or more paragraphs if there is more than one point. (You can read up on paragraph structure and topical sentences, etc., online.)

2. Justify all your statements. If you make a statement that you think your own study provides justification for, make sure to point to the figure or table that provides that evidence. If you make a statement that is justified by some prior study, make sure to cite that prior study. Importantly, note that essentially all points need to be justified, unless well known. For example, a comment that "phylogeny estimation is challenging" *does* need references to support the statement, but "New York City is a major city" does not need a reference.

3. Low-level writing: Be careful about spelling, punctuation, etc. Look for capitalization that isn't needed, or that is needed. Look for spaces accidentally introduced within words (or phrases). Look for words that have been accidentally combined due to lack of space. Check for directions of quotation marks (specifically, look at the leftmost quotation marks and make sure they appear as you want them).

4. Number your figures and tables so you can refer to them. This is done easily in latex. See, for example, `https://www.overleaf.com/learn/latex/Referencing_Figures`

5. Each figure or table should have its own paragraph for the discussion. But you need to reference the figure or table to do this well.

6. Make sure all terms you use are defined. Some terms do not need to be defined, but only if you are very sure that the reader should know this. For example, "phylogeny" doesn't need to be defined. Nor does "DNA" or "nucleotide". But many other terms will need to be defined, because not all readers will know them. Examples of terms that not all readers are

expected to know are "polytomy", "Robinson-Foulds distance", "SPFN", "modeler score", etc.

7. Citation practice. Lots of things to say here.

   - The first time you mention a method or paper, you need to cite it. The citation typically goes immediately after the text where it is mentioned but is sometimes later in the sentence; in any event, it is before the period at the end of the sentence.
   - You only need "et al." for three or more authors. If there are two authors, it's "Johns and Wiens" rather than "Johns et al."
   - Any paper you cite should be one that you have at least skimmed so you know what's in it. For example, if you are citing a paper with the aim of supporting a statement that phylogenies are useful for inferring protein structure and function, then make sure that the paper actually is about that.
   - Try to cite the original papers. If you want to cite a paper for a given dataset, cite the first paper that provided the dataset. Similarly, if you want to cite a paper that shows that maximum likelihood phylogeny estimation is NP-hard, find the paper that proves the NP-hardness result. Think of citation as giving credit for an idea – the group that came up with the idea should get the credit, not the people who cite them (or the people who cite the people who cite the people who site them).

8. Bibliography list.

   - If you are citing any preprint, make sure it isn't already available in a peer-reviewed form. Preferentially cite the journal version rather than the conference version. Make sure to include a DOI or URL for a paper in preprint form. Include all the authors unless there are more than 7.
   - Check the capitalization (for example, you would not want to write "Raxml" since it should be "RAxML"); learn how to enforce capitalis in bibtex by using "{RAxML}" instead of writing "RAxML."
   - Be careful to look at the PDF and not just at your bibtex. Sometimes there is information in your bibtex that doesn't show up in the PDF – check that your printed document has everything you want it to have.