Tips for Writing Large LaTeX Documents

David Kraemer

Today

1 Introduction

I should mention that I personally think that these tips are important for writing LaTeX documents of *any* size. There is a lot to be said for developing good LaTeX habits that scale well. That said, you (read: I) can get away with a lot of dirty tricks when the text you are composing is only a handful of pages. To borrow from software engineering, it's quite easy to manage "technical debt" on small projects. Hacks and other tricks don't show their dark side to you until you've reached a scale where juggling the accumulation of bad choices begins to overpower your ability to write.

2 The Tips

2.1 Get to know a good LaTeX text editor

2.2 Try to think systematically about your document

2.3 Raise style violations to compilation errors

2.4 Write wide, not deep

Every paper, report, or book should have its own directory. I come from a programming tradition where the principle file in the project is called `main.*`, and I've kept that practice with LaTeX drafts.

There is virtually[1] no reason to write a large document in a single `.tex` file. In fact, there are several reasons not to.

If you're not using a robust text editor (see Section 2.1), consider two possible worlds. In World 1, you're looking for Chapter 6 Section 4, and you spend two or three minutes scrolling up and down a 10,000-line file until you catch it. In World 2, you have a directory structure where each chapter is its own folder,

[1] A potential downside to wide projects is that systematic changes (i.e., through regular expressions) have to be applied individually to each included file. I agree that this is somewhat tedious. However, a good editor will make this process painless, and the downsides to the deep approach are overwhelming. See Section 2.5, below.

and each section its own file in the folder. It takes all of ten seconds to locate the desired section.

Using `\include{}` properly will actually speed up the document compilation process. This is because LaTeX is a little smart, and it knows to recompile the included files *only* when these files have changed. No change, no recompilation. You'll notice that these files will have their own individual `.aux` files in their respective directories. That's the result of LaTeX precompiling these files for direct inclusion into the body of the main text.

My undergraduate computer science advisor liked to remind us that space is cheap, but confusion is expensive. I am personally the type of mind that gets easily overwhelmed by sprawling files. I start to lose the forest for the trees. It becomes harder to reason systematically about the entire document.

2.5 Learn to love regular expressions

2.6 Prefer semantic descriptions to numbers

2.7 Separate working and polished drafts

2.8 Adopt a good label convention

Every numbered item in your large LaTeX document should have a label. Chapters, sections, theorems, equations, even items in an `enumerate` list[2]. You may end up not using every label you make. But getting into the practice of writing labels is much better than the alternative. I have seen too many LaTeX documents where "`Equation (4)`" is hard coded into the text of the document, which is one of the more pungent code smells a document can have.

If everything gets a label, then we need to have some way of organizing them. Let's consider an extended example for the equation

$$\bar{c}(\mu,\nu) \stackrel{\text{def}}{=} \inf_{\pi \in C(\mu,\nu)} \int_{\mathbb{X}\times\mathbb{Y}} c(x,y)\ \pi(dx,dy) \qquad \mu \in \mathbb{P}(\mathbb{X}), \quad \nu \in \mathbb{P}(\mathbb{Y}). \tag{1}$$

For context, (1) is the cost function in optimal transport theory. Here are my suggestions:

- The beginning of a label should simply indicate the type of the item being labeled. If I'm labeling a theorem, the label begins `thm:`; if I'm labeling an equation, the label begins `eq:`; and so on. I put this at the beginning of the label, because when I am in the process of writing and I think to myself, "oh, now I need to reference this equation," the first thing that comes to mind is that the reference is an *equation*.

- The next parts of the label should indicate where the item is located in your text. In an article, that means indicating the section; in a report, thesis, or dissertation, that means including the chapter (or part)

[2] If you don't care about the number, why is it an ordered list?

that houses the item. If (1) is from a chapter on metric spaces in a section concerning optimal transport theorems, I would put the preface `eq:metric_spaces:transport` in the label.

- The actual item should be given a semantically descriptive label. The label up to this point carefully documents where the label is located, which usually will let you describe the actual item in a few short words. For 1, it's enough to add the word `cost` or potentially `cost_function` to the label text. Then our complete label is

  ```
  \label{eq:metric_spaces:transport:cost_function}
  ```

  which uniquely identifies the equation in a fashion that can be easily indexed. It's also quite descriptive, since we know that "cost function" is referring to a specific construction for optimal transport in metric spaces.

- Be consistent. The chapter

- Don't use numbers[3] in your labels. One of the most annoying and difficult things in maintaining a large LaTeX file is updating your files when the organization of the sections in a paper (or chapter) changes.

  The labeling system I've proposed above is not immune to some difficulties itself. If you swap sections in a paper, then you don't need to make any changes. But if you move content between sections, you might need to update the relevant parts of the label path. It's a little annoying and easy to forget. On the other hand, a good text editor (see Section 2.1) can make global refactoring of an individual label a one-click affair.

2.9 BiBTeX conventions to live by

3 Conclusion

[3] Semantically significant numbers get a pass. But use sparingly! English is an asset, not a liability.