# PREPARING DOCUMENTS FOR PRINT, HTML, AND PDF

Jeff Clark
Elon University
Elon, NC 27244
email: clarkj@elon.edu
web: http://frodo.elon.edu

## Introduction

My students are used to getting information in many different media, whether it be from me in class, working at a computer, or discussing the class in groups. In the past few years I have gotten accustomed to providing my classes with course materials in several different media, and have worked to do so efficiently.

I have been using LaTeX for quite a few years now in preparation of all of my documents. In terms of sheer aesthetic appeal it is still unbeaten as a typesetting system. It makes possible the display of mathematical formulae with more ease than I ever thought possible using other commercial and academic systems. The fact that free versions of the software are available for faculty and student use is icing on the cake.

I have found that it is possible to use my LaTeX source files to generate both HTML (HyperText Markup Language) files for display on a web browser as well as PDF (Portable Document Format) files for printing from my class web pages. The techniques involved are well-known in the LaTeX community but not as well known by more casual users. This presentation will describe the reasons for using such a system as well as relevant implementation details.

## The Need for a Single Source

When working with a source document, it is very easy to make errors when maintaining multiple source files for the same content. If you type your content perfectly the first time, then you may be safe with cutting and pasting the source into different formats. For the rest of us mortals, revision is a fact of life. It is near impossible to manually keep separate files in synchronization with multiple editing sessions.

## Need for Multiple Output Media

There are two main ways to share mathematical course materials with students: on paper or in a web browser.

## PDF

Portable Data Format (PDF) is a graphic format derived from PostScript. PDF is an improvement over PostScript in providing smaller files, simpler printer commands, and a greater range of options. Both PDF and PostScript were developed by Adobe Systems [4].

PDF is not as accessible as HTML, but Adobe has made its Acrobat reader freely downloadable. PDF prints well, much better than the quality that can be found from web browsers.

## LaTeX2HTML

LaTeX, by virtue of its logical formatting structure, is wonderfully constructed to be translated into HTML. Most mathematical formulae are not representable in a web browser. LaTeX2HTML translates such formulae into small graphical images that are embedded in the web browser.

LaTeX2HTML was created by Nikos Drakos and a large number of contributors, and is now maintained by Ross Moore. It is a series of Perl scripts that use the structure of a LaTeX document to break it up into web pages in HTML. When it encounters mathematical formulas that can not be represented easily in HTML, it creates small image files from them which are then embedded within the HTML code.

LaTeX2HTML is freely available at `http://www.latex2html.org`.

### Use of Stylesheets

Stylesheets allow for the consistent appearance of content across many web pages [3]. A sample style sheet looks like the following:

```
body {
        background-color: rgb(80%,96%,93%);
        color:  black;
        font-family: georgia,  serif;
        margin-left: 15%;
        margin-right: 10%;
        }
h1, h2, h3, h4, h5, h6 {
                        font-family:  verdana,  sans-serif;
                        }
```

Print has many advantages: it is portable, more easily read, and much more compact. Most of my students read my course materials once or twice however, and would prefer not to have to carry them around. They also are prone to losing my masterpieces and having to ask for replacement copies.

My students by and large would like to choose which documents they need to have in print and which ones they will only read once. They often make this decision long after I have left campus. Having course materials accessible from my course home page gives them a great deal of freedom of choice.

Given that documents are going to appear on the course web page, it is important to provide a version that can be printed. Most printing from web pages is not compact, with formatting chosen by the browser and not by me the author. Providing specifically formatted printable versions of all web documents provides the best of both worlds.

## LaTeX

LaTeX is my preferred choice for source format for the reasons described above.

TeX is an incredibly powerful, book-quality technical text formatting system invented by a computer scientist by the name of Donald Knuth. It is capable of typesetting complex mathematical expressions as well as performing good line and page breaks, maintaining references and other structures [2].

TeX, possibly with the inclusion of a macro package, takes a plain text file of commands and uses it to produce a device-independent (DVI) file. A printer or screen previewer then processes the DVI file according to the hardware used.

TeX is very much concerned with all of the details of typesetting. In 1985, Leslie Lamport constructed a set of macros known as LaTeX. This set of macros focuses on logical formatting: sections, subsections, environments, and so forth are the basic building blocks of a LaTeX file. Style components such as margins, etc., are handled by default or user-constructed style packages.

## Output Formats

### HTML

HTML is the language of the World Wide Web, and requires no special browser features or plug-ins. For the foreseeable future, e.g., the next couple of weeks, it will continue to be the most widely accessible format for web documents.

This style sheet instructs web browsers to provide a given background color (given as percentages of red, green, and blue), to use black type (still the easiest to read), if possible to use the Georgia font but if that is not available any serif font, and finally sets left and right margins as percentages of the web browser's width.

I use stylesheets to ensure decent margins, and to color-code my classes to ensure that I have not placed class web pages in the wrong location. The appropriate use of margins can have a huge impact on the readability of your web pages.

### HTML.sty

When constructing documents for both print and HTML output, it is often desired to include sections that are handled differently. For example, where in a web page you would like to include a link under a given name, in a print version you would like to include the URL of the link itself. More generally some sections are more appropriate for one format than for both.

LaTeX2HTML comes with a package HTML.sty with a number of commands for including links, images, LaTeX-specific, and HTML-specific sections.

### pdfLaTeX

pdfLaTeX is part of a package known as pdfTeX, created by Hàn Thé Thành, that produces PDF output from TeX files and related files such as LaTeX. It is included with most LaTeX software packages by default.

pdfLaTeX can handle nearly everything that LaTeX can, and it does handle graphics of many different formats, such as JPG and GIF.

### Implementation Issues

Keeping track of all of the needed commands and options is a pain. I have been using Makefiles for quite a few years to help out. Make is a programming command that has been around in Unix environments for many years as a way of keeping track of compilation commands and dependencies. It has also been ported to the Windows environment.

My default LaTeX2HTML options request that the program produce HTML version 4.0 code; that images are generated in the GIF format; that there be no LaTeX2HTML descriptions on each page; that there be no navigation buttons; and that my course documents (which are usually short) be produced in a single page.

### Installation

The Comprehensive Tex Archive Network (CTAN) contains links for installing TeX and related programs such as LaTeX and pdfLaTeX. Its URL is

```
http://www.ctan.org
```

The `MikTeX` system works well on `Windows`, and the `tetex` system works well on `Unix` systems.

The LaTeX2HTML program can be downloaded from

```
http://www.latex2html.org
```

Installation on `Unix`-related systems is not too bad, but the installation on `Windows` systems does require a good deal of patience.

The `make` program can be downloaded from `http://www.gnu.org` and is included in many `Linux` distributions.

### Conclusion

There are many good alternatives for web document production. The combination of LaTeX2HTML and pdfLaTeX offers a good deal of flexibility and deserves to be better known.

# References

[1] Michel Goosens et al, *The LaTeX Web Companion*, Addison-Wesley, 1999.

[2] Helmut Kopka and Patrick W. Daly, *A Guide to LaTeX*, Addison-Wesley, 1999.

[3] H. W. Lie and B. Bos, *Cascading Style Sheets*, Addison-Wesley, 1997.

[4] Thomas Merz, *Web Publishing with Acrobat/PDF*, Springer-Verlag, 1997.

[5] Norman Walsh and Leonard Muellner, *DocBook: The Definitive Guide*, O'Reilly, 1999.