A Note on Writing and Distributing a Document with Mathematical Expressions −by using a Personal Computer−

Yuuichi KAWAGUCHI

This is a note explaining about document preparation systems for writing a document with mathematical expressions and document formats for distributing such a document using a personal computer.

Many university staffs and students often have to prepare documents that include mathematical expressions. Simple fractions such as 'one over n' can not be written using some application software programs. Various document preparation systems and computer languages are introduced and evaluated.

Distribution of a document is often required. One simple way is to print out the document and distribute it. However, it is difficult to reuse a document that has been printed out on paper. A document can also be distributed using a floppy disk or a computer network. In this note, various document formats are introduced and are evaluated.

It is concluded that LᴬTᴇX is the best document preparation system. It has a plenty of functions to enable various mathematical expressions to be written.

PDF is considered to be the best for document format. PDF is designed to be distributed via the Internet. There is an official and free viewer. There is no confusion among users.

1 Introduction

This is a note explaining about document preparation systems for writing a document with mathematical expressions and about document formats for distributing such a document using a personal computer.

Many university staffs and students often have to prepare documents that include mathematical expressions. Simple fractions such as $\frac{1}{n}$ can not be written using some application software programs. Various document preparation systems and computer languages are introduced and evaluated.

One simple way of distributing a document is to print it out and then distribute it. However, it is difficult to reuse a document that has been printed out on paper. A document can also be distributed using a floppy disk or a computer network. In this note, various document formats are introduced and evaluated.

2 Mathematical Expression

2.1 LaTeX

The LaTeX [6] system was developed mainly for expressing mathematical expressions, and it is now widely used by scientists and engineers for writing papers or books. The system is *de facto* standard, but is much different from ordinary word processing systems used on personal computers such as Microsoft Word (MS-Word).

There are two steps in writing a paper with LaTeX. The first step is writing a document. There is no need to use a special application software for writing a document using LaTeX. The document consists of sentences and commands to LaTeX (LaTeX commands). A LaTeX command works on a letter, a word, a phrase and any part of the document. For example, a piece of a document '\emph{happy}' consists of a command '\emph{ }' and a word *'happy.'* The command makes the word emphasized. When the document is printed out, the word is usually italicized like *'happy.'*

The second step is compiling the document. A program called latex reads a document consisting of sentences and commands. The program recognizes LaTeX commands and creates a file in a format called DVI (DeVice Independent). A file in the DVI format (DVI file) consists of commands (DVI commands). The commands construct pages of the document. Some programs, such as pdflatex [3] and latex2html [3], compile the original LaTeX document into a file other than the DVI file. These programs are not evaluated in this note, since they seem to be still in the developmental stage.

There are many programs that read DVI files and recognize DVI commands. For example, a program called dvips reads a DVI file and creates another sequence of Postscript [4] commands for printing devices. Other programs called xdvi and dviout.exe display the document on a computer screen. There are some programs for translating a DVI file into a file in another format. For example, dvips reads DVI commands and translates them into Postscript commands.

2.2 LaTeX Command
for Mathematical Expression

There are many LaTeX commands for writing a document. Let me focus here on LaTeX commands for writing mathematical expressions.

There are two modes in a LaTeX document. One is a mode for ordinary texts and the other is a mode for mathematical expressions (math mode). The text between \begin{*mode*} and \end{*mode*} is in math mode and is treated as a mathematical expression. There are three alternatives for *mode*: math, displaymath and equation. In

math mode, mathematical expressions are displayed together with text, like $e = mc^2$. In displaymath mode, they are displayed outside of other texts, like

$$e = mc^2.$$

Mathematical expressions in these two modes are unnumbered. In equation mode, mathematical expressions are treated the same as they are in displaymath mode, but they are numbered, like

$$e = mc^2. \tag{1}$$

The number label generated can be referred to from anyware in the document. For example, suppose that there is a piece of a document

\begin{equation}
π = 3.141592 \label{eq:pi}
\end{equation}.

The command \label{eq:pi} stores the generated number of the mathematical equation into the key 'eq:pi.' A command \ref{ } refers to a number. If a command \ref{eq:pi} is placed somewhere in the same document and if the number referred to by the key is two, then the command results in '2.', for example, the LᴬTᴇX command

Eq.~(\ref{eq:pi})
results in 'Eq. (2).'

2.3 LᴬTᴇX Command by Example

Let me show some examples of LᴬTᴇX commands for writing mathematical expressions. The sequence of LᴬTᴇX commands on the left-hand side of the symbol '⇒' results in the right-hand side when displaying or printing.

- e = mc^2 $\Rightarrow e = mc^2$
- a_1, \ldots, a_n $\Rightarrow a_1, \ldots, a_n$

- \bar{x} =
  \frac{1}{n} \times
  \sum_{i=1}^n x_i^2

$$\Rightarrow \bar{x} = \frac{1}{n} \times \sum_{i=1}^n x_i^2$$

- \int_{-\infty}^{+\infty}
  \frac{1}{\sqrt{2\pi\sigma^2}}
  e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 1

$$\Rightarrow \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 1$$

2.4 Discussion
2.4.1 HTML and MathML

A system like LᴬTᴇX, in which a document consists of sentences and commands, is called a mark-up system or, in general, a mark-up language. Some mark-up languages enable mathematical expressions to be written.

The hyper-text mark-up language (HTML) [9] is mainly used for constructing a web page. HTML commands, which are called 'tags,' are classified in two categories. One category is commands for specifying a logical structure of the document. For example, the <H1> tag specifies that a sentence is a headline. Display effects, which mean how the headline sentence is displayed on a computer screen, are left to web browsers. A browser can display a headline in, for example, bold-faced letters or underlined letters.

Commands for specifying a logical structure are placed into HTML documents. Commands for controlling display effects are separated from the HTML document and placed into another document, called 'style sheet' [7, 8]. The style sheet is referred to by HTML documents.

There are many HTML commands, but there are few commands for writing mathematical expressions.

The mathematical mark-up language

(MathML) [1] is a system for writing mathematical expressions in a web page. The system was designed on the basis of LATEX. Mathematical expressions written in MathML are interpreted as mathematical expressions by a web browser supporting MathML. However, MathML is still not widely supported.

2.4.2 Microsoft Word

Microsoft Word (MS-Word) is a very popular an application software for word processing. It is so-called WYSIWYG (What You See Is What You Get). This is the main point that differs from LATEX. WYSIWYG applications use a graphical user interface (GUI), which is a set of menus, buttons and so on.

The GUI of MS-Word has frequently been changed by its developer, Microsoft Corporation. When a new function is added to MS-Word, a new menu or button is added to the GUI in order to be able to use the new function. Such a change occurs almost every year. Thus, users are always unfamiliar with MS-Word. One of the main problems with MS-Word is that there are too many functions. Most MS-Word users probably do not know how to use all of the functions. In contrast, it is difficult to change the LATEX system. The newest LATEX system, developed in 1994, is LATEX2ε. Bugs in the system are continuously being removed. Donald E. Knuth, the developer of the TEX system, which is a base system of LATEX, has announced that he will do no further development of TEX[*1]. Thus, most users of LATEX are very familiar with its functions.

MS-Word itself has few commands for writing mathematical expressions. It includes a equation editor. The equation editor has functions for writing mathematical expressions. It produces a graphical image of a mathematical expression. The image can be pasted into a document written by using

MS-Word. The old version of the equation editor is included in MS-Word. Users can purchase a newer version, called MathType, from a web site[*2].

2.4.3 MathType

MathType is a GUI for writing mathematical expressions. A user constructs a mathematical expression by using menus, buttons and so on. MathType creates a graphical image of the mathematical expression with GIF or EPS format or creates a sequence of LATEX commands for writing the mathematical expression. Many application software programs for word processing enable graphical images to be included in a document. There is a LATEX command for inserting and manipulating graphical images in a LATEX document. MathType makes it possible to write mathematical expressions intuitively without logical thinking.

Some people like intuitive functions and other people like logical functions. Mathematical expressions written using MathType are not different from those with LATEX. MathType is just a GUI.

2.5 Conclusions

It is difficult to write mathematical expressions using MS-Word, even with MathType, because functions of such a WYSIWYG system are implemented by a GUI that is intuitive but not logical. The GUI of MS-Word frequently changes. The intuition of a user is not effective when the GUI changes.

It is easy to write mathematical expressions by LATEX. It does not change easily. Users can become familiar with LATEX after a period of time. The knowledge about LATEX commands is effective for a long time.

MathML is a new system for writing

[*1] http://www.tug.org/faq/

[*2] http://www.mathtype.com/

mathematical expressions, but this system has not yet been evaluated.

3　File Format for Distribution

3.1 Adobe PDF

The portable document format (PDF) [5] was developed by Adobe Systems Incorporated. Postscript (PS) was developed by Adobe before the development of PDF. PS was originally designed to control printing devices. PDF was developed based on PS.

Acrobat Reader, which is a document viewer with a PDF format (PDF document), is distributed freely by Adobe. For the same content, a PDF document is smaller than a document with PS format (PS document). PDF has functions for encryption and addition of a digital signature of the writer. PDF documents are suitable for distributing via a computer network, especially the Internet.

It is difficult to create a PDF document. The best way is to use Acrobat Distiller by Adobe, but it is expensive for personal use. There is a program called pdflatex, which is free and distributed with LᴬTᴇX in many cases. In general, a LᴬTᴇX document is translated into a DVI format by the command latex, the DVI format is translated into a PS document by the command dvips and then the PS document is translated into a PDF document by Acrobat Distiller. The program pdflatex can directly produce a PDF document from a LᴬTᴇX document, but sometimes fails. It is still in the developmental stage.

The author uses Acrobat Distiller for creating a PDF document since it is the most reliable software currently available.

3.2 Discussion
3.2.1 HTML

An HTML document is directly displayed on a computer screen by a web browser. There is no need to compile it to another format. An HTML document can be distributed as it is.

As mentioned above, however, an HTML document has few functions for writing mathematical expressions.

3.2.2 Rich Text Format

The rich text format (RTF) [2] was developed by Microsoft Corporation. It was only a format for Microsoft's application software. The specifications of RTF have been made available, and some application software programs that were not by Microsoft use RTF.

MS-Word and wordpad.exe, which is an application software of Microsoft Windows, can open a document with RTF format (RTF document) and can save a document with RTF format. Star Office*3, which was developed by Sun Microsystems, also have those abilities for an RTF document. Many other application software programs support RTF documents. However, those application software programs deal with an RTF document in different ways. There are some differences in displaying and printing out.

If Microsoft were to distribute a free and authorized RTF viewer that can run in Microsoft Windows, Apple Mac OS, UNIX and so on, then the degree of confusion regarding RTF decreases. There is, however, no authorized viewer.

3.3 Conclusions

HTML is not good at distributing a document with mathematical expressions. RTF has the ability to write mathematical expressions. Many application software programs support RTF. There is confusion regarding RTF since there is no authorized official viewer.

PDF also has the ability to write mathematical expressions. Adobe distributes a free and authorized viewer, Acrobat Reader, at a

---

*3 http://www.openoffice.org/

web site*⁴. PDF is good at distributing a document with mathematical expressions.

4 Conclusive Remarks

In conclusion, L^AT_EX is the best system for document preparation. It is easy to use and has many functions for mathematical expressions.

PDF is the best system for a document format. This system was designed to be distributed via the Internet. There is a free and authorized viewer. There is no confusion among users.

New technology is continuously being developed, and better systems should become available in the near future.

Acknowledgment

References

[1] Ron Ausbrooks, Stephen Buswell, Stephane Dalmas, Stan Devitt, Angel Diaz, Roger Hunter, Bruce Smith, Neil Soiffer, Robert Sutor, and Stephen Watt. *Mathematical Markup Language (MathML) Version 2.0.* W3C, February 21 2001. (http://www.w3.org/TR/MathML2/).

[2] Microsoft Corporation. Rich text format (RTF) specification, version 1.6. (http://msdn. microsoft.com/library/en-us/dnrtfspec/html/ rtfspec.asp).

[3] Michel Goossens and Sebastian Rahtz. *The L^AT_EX Web Companion.* Addison-Wesley, 1999.

[4] Adobe Systems Incorporated. *PostScript language reference manual,* 3rd edition, 1999. (http://www.adobe.com/supportservice/ devrelations/).

[5] Adobe Systems Incorporated. *PDF Reference,* 2nd edition, 2000. (http://www.adobe.com/ supportservice/devrelations/).

[6] Leslie Lamport. *A Document Preparation System L^AT_EX.* Addison-Wesley, 2nd edition, 1994.

[7] Håkon Wium Lie and Bert Bos. *Cascading Style Sheets, level 1.* W3C, revised January 11 1999 edition, December 17 1996. (http://www.w3.org/ TR/REC-CSS1/).

[8] W3C. *Cascading Style Sheets, level 2 CSS2 Specification,* May 12 1998. (http://www. w3.org /TR/REC-CSS2/).

[9] W3C. *HTML 4.01 Specification,* December 24 1999. (http://www. w3.org/TR/html401/).

---

*⁴ http://www.adobe.co.jp/products/acrobat/ readstep.html

*⁵ Title: *A Development of On-Line Documents for Subjects Related to Statistics*