
Subject:	Why \LaTeX is used for review report typesetting within the TDB project
Distribution:	Public
Issued by:	Erik Östhols on 5th July 1999
Telephone:	+33-1-45 24 10 83
Signature:	
No. of appendices:	
TDB Document number:	TDB-1999.8

1 Introduction

The Thermochemical Data Base (TDB) project review reports are currently typeset using the \LaTeX system [94LAM]. Questions are frequently asked about why we use \LaTeX . This document aims to explain why \LaTeX is used for typesetting TDB review documents.

2 About \LaTeX

What is \LaTeX ? The short answer is that \LaTeX is a system for typesetting (scientific) documents. It is really what is commonly referred to as a “Mark-up Language” (ML), a concept which is again gaining popularity with the spread of the Internet and languages such as HTML (Hyper-Text Mark-up Language) for mark-up of World-Wide Web pages, XML (eXtended Mark-up Language) and SGML (Standard Generalised Mark-up Language). The simple idea behind a ML is, that the input is a standard (ASCII) text file containing the mark-up “tags”, telling which part of the text is what (title, headings, tables, cross-references, links to other documents etc.). The interpretation of these “tags” is then a task for some computer program, which will turn the input into (hopefully) readable text e. g. on a computer screen or in a document to be printed. This is in many ways similar to a computer program, which starts its life as a source file, which is then treated by a compiler or an interpreter to produce a running program. For MLs, the source file corresponds to the input file, and the running program corresponds to the document or screen output produced by e. g. a WWW browser program. More information about the \TeX system that forms the base for \LaTeX can be found in [86KNU] and [94WAL], and \LaTeX is thoroughly described in several publications

[94LAM, 94WAL, 94GOO/MIT, 97GOO/RAH]. There are many web pages dedicated to \TeX and \LaTeX and its uses, <http://www.tug.org> provides a good starting point.

3 How does \LaTeX work in practice?

When using \LaTeX , one creates an ASCII input file (this can be done with any available text editor, though some, such as GNU Emacs, work better with \LaTeX than others), containing the necessary mark-up tags, runs it through the \LaTeX program, which checks syntax, formats the document and outputs a so-called DeVice-Independent (DVI) file. This DVI file can then be converted to PCL, PostScript (PS) or whatever the local printer understands (though for some special effects it may be necessary to go via PostScript first) using available software tools. Generally, \LaTeX works very well with PostScript as its final output, since PostScript is powerful enough to do easily do the all the things required by \LaTeX . However, since one can use the Ghostscript program (<http://www.wisc.cs.edu/~ghost>) to convert the PS output to whatever is required by the local printer, portability is not really a problem. With the `hyperref` package for \LaTeX (more about packages below), it is also very easy to produce a nicely hyperlinked, very portable PDF (Portable Document Format) file from \LaTeX input. PDF files are readable with the freely available Adobe Acrobat Reader program (<http://www.adobe.com>).

4 Which are the alternatives to using \LaTeX ?

There are few serious contenders to \LaTeX , but the obvious ones would be Microsoft Word, and, to a lesser extent, WordPerfect. Whereas these two programs have been around for fairly long, they have only in recent years become powerful enough to be serious contenders to \LaTeX for typesetting documents. HTML is not an alternative, since its present version (4.0) still very much reflects the origin of HTML as a contents-concentrated ML, with very limited possibilities for layout control and advanced scientific typesetting (it cannot even do automatic hyphenation). Possibly, this will change with time, but for now, HTML is not an alternative for typesetting TDB review reports.

5 So, why *are* you using \LaTeX instead of MS Word?

To understand why the TDB project uses \LaTeX rather than MS Word (or WordPerfect), it helps to list the advantages and disadvantages of \LaTeX .

Advantages with using \LaTeX :

- It does very high quality typesetting, decidedly better than current versions of Microsoft Word and WordPerfect.
- It is very powerful and flexible. There are really no significant restrictions on the types of formatting, the mathematical and/or chemical formulae etc. that can be typeset.
- Powerful so-called “packages” can be written, that take care of all the essential formatting, leaving the user to concentrate on the contents of the text.
- Since the input is a standard ASCII file, it is very easy to write \LaTeX interfaces for data base management systems. This is not necessarily true for e. g. MS Word, which has as its only true text file interface the so-called Rich Text Format (RTF). Unfortunately, RTF is not well standardised, and tends to change with each new version of MS Word coming out. This makes it necessary to update any RTF-dependent tools quite often.
- \LaTeX has a proven ability to deal with documents of 800+ pages, full of formulae, complicated tables, figures, cross-references and rotated material (landscape). This is not true for Microsoft Word at the time of writing. While recent (e. g. Word 97) versions of MS Word can do quite sophisticated typesetting, they still tend to have difficulties with very large manuscripts.
- Although commercial versions exist, basically \LaTeX is freely available. One downloads the source code or a pre-compiled version from the net, installs it and runs.
- \LaTeX can run on a multitude of computer platforms, notably Microsoft Windows 95/98/NT, virtually any Unix system available, VMS and MacIntosh. There are even versions available for Atari, Amiga and TOPS-20, for those who remember it.
- \LaTeX is a mature system. It is very stable and reliable, and has few bugs and problems.
- \LaTeX has a very powerful macro mechanism, which can be used to ensure consistent formatting of particular text, and to abbreviate often typed text and formulae. It should be noted, that although MS Word does have macros for inserting frequently typed text, once they are in the text, one has to search and replace them all if a change has to be made. This is not the case

with \LaTeX : one simply changes the definition of the macro, and then all occurrences of the macro text are changed automatically without the user having to do anything at all except changing the macro definition.

Disadvantages with using \LaTeX :

- Since \LaTeX is not a true WYSIWYG tool, it is necessary to run the \LaTeX program to see the results of changes made.
- Package writing is fairly difficult, and not a task for novices.
- Generally, \LaTeX takes longer to learn than WYSIWYG systems such as MS Word.
- \LaTeX is not as widespread as e. g. MS Word, meaning that \LaTeX files often cannot be distributed directly.
- Conversion between \LaTeX and other systems, such as MS Word, is difficult, often impossible.

The deciding factors for using \LaTeX rather than MS Word are the following:

- Its known stability and ability to handle large, complex manuscripts such as the TDB review reports. I have yet to see anyone typeset 800+ pages of manuscript with tables, figures and cross-references with MS Word.
- \LaTeX uses standard text files as input, which makes it easy to interface with data base management programs. This is in contrast with MS Word (see above).
- There is a lot of material already in \LaTeX format in the project, including the reference publication database. Converting all of this would be very time-consuming.
- \LaTeX 's typesetting quality is superiour.

Having said all this, it should also be pointed out that there are many tasks for which \LaTeX is not well suited. Basically, \LaTeX is the typesetting equivalent of a professional, high-price mirror-reflex camera with lots of optional lenses and extra gadgets. If you merely want something to take pictures of your summer vacation, you don't need an expensive, heavy, complicated professional camera such as a Hasselblad. You bring your compact camera with built-in flash and happily snap some shots when you feel like it. Similarly, if you are only writing a letter, or want to produce a not too long manuscript without too many complicated formulae, you don't want to expend the extra effort it takes to use \LaTeX . \LaTeX is meant for very

high quality typesetting of extensive and/or complicated material, not for writing a one page PM with text only. Likewise, \LaTeX is not very good for producing material that requires a lot of visual formatting, such as slides (even though it can of course be used for that too).

6 Some common misconceptions about \LaTeX

Conception: \LaTeX is very old and cumbersome, and nobody uses it anymore.

Truth: Although \TeX and \LaTeX have been around for a number of years now, new packages and additions keep appearing, and the TUG (\TeX User Groups) in especially the U. S., Germany, France and the U. K. are very active. It is standard for typesetting several scientific journals, and used in many universities all over the world. New books about \LaTeX keep appearing [94GOO/MIT, 97GOO/RAH]. With today's fast computers and nice viewing tools, working with \LaTeX is not really particularly cumbersome.

Conception: It only runs on VAX/Unix systems.

Truth: No, it runs on virtually any computer system. See above.

7 How do I convert documents to/from \LaTeX ?

This is tricky. One problem is that \LaTeX has more special symbols and can do more complicated typesetting than almost any other available program. Often, it is simply not possible to make a good conversion from \LaTeX to other document formats because these other document formats just cannot do the things \LaTeX can. The most common question is how to convert between \LaTeX and MS Word. Since the MS Word document format is binary and not really open, and also changes with each new version of MS Word, trying to convert directly to MS Word format is usually not an option. Converting to Rich Text Format (RTF) is also difficult, because, as noted above, RTF keeps changing. Microsoft have not declared that they will keep backward compatibility with older versions of RTF. However, there is a recent program called `rtf2latex2e` that claims to be able to understand the latest versions of RTF. Still, the best long term bet for converting between \LaTeX and MS Word is probably to go via HTML, since this is an open, standardised format. Newer versions of MS Word can both read and write HTML directly. Converting from \LaTeX to HTML is probably best done with the \LaTeX package and program `hyperlatex`, which imposes restrictions on what one writes in \LaTeX so as to make a conversion to HTML, with its severe restrictions on layout, possible. For converting from HTML to \LaTeX , there are some tools available. These are best found through <http://www.tug.org>, which

has pointers to lists of conversion tools. The book “The L^AT_EX Web Companion” [99GOO/RAH] gives a good overview of techniques for converting L^AT_EX to HTML and for creating PDF files from L^AT_EX source.

References

- [86KNU] Knuth, D. E., The T_EXbook, Reading, Mass.: Addison-Wesley Publ. Co., 1986..... 86KNU, 9015
- [94GOO/MIT] Goossens, M., Mittelbach, F., Samarin, A., The L^AT_EX companion, Reading, Massachusetts: Addison-Wesley Publ. Co., 1994, 528 pages. 94GOO/MIT, 9012
- [94LAM] Lamport, L., L^AT_EX: A document preparation system, Reading, Massachusetts: Addison-Wesley Publ. Co., 2nd ed., 1994, 272 pages..... 94LAM, 9011
- [94WAL] Walsh, N., Making T_EX work, Sebastopol, California: O’Reilly & Associates, Inc., 1994, 483 pages. 94WAL, 9014
- [97GOO/RAH] Goossens, M., Rahtz, S., Mittelbach, F., The L^AT_EX Graphics Companion, Reading, Massachusetts: Addison Wesley Longman, Inc., 1997, 554 pages..... 97GOO/RAH, 9013
- [99GOO/RAH] Goosens, M., Rahtz, S., The L^AT_EX Web Companion, Reading, Massachusetts: Addison Wesley Longman Inc., 1999, 522 pages. 99GOO/RAH, 9037