`ltx2rtf`: Exporting LaTeX documents to Word addicts

Daniel Taupin
Laboratoire de Physique des Solides
bt. 510, Centre Universitaire
F-91405 Orsay Cedex
France

Abstract

`ltx2rtf` is a compiler that translates LaTeX2ϵ source text into the RTF format used by several text processors, including Microsoft Word and Word for Windows. It was written by Fernando Dorner and Andreas Granzer in a one-semester course in Vienna (Austria) and is currently found as `latex2rtf` in CTAN servers.

It was heavily corrected and adapted to LaTeX 2_ε in 1997 by Daniel Taupin. The distribution was intended mainly for use within the MS-DOS window of Win95 and Win3.11, but all sources can be compiled on UNIX computers having GCC compilers.

Introduction: The need for a converter to RTF

Like most of the audience of TUG and other TeXperts' meetings, I usually write most of my papers in LaTeX 2_ε. But problems arise when I need to transmit these documents to non-LaTeX users.

Various cul-de-sacs when transmitting LaTeX documents. Transmitting a LaTeX document to other LaTeX users is no problem, since all LaTeX formats at least recognise the 7-bit representation of accented letters. The problem arises only when the addressee is reluctant to use a LaTeX representation:

Sending plain text. This obvious (poor) solution fails because accented letters have at least three commonly used codings, the 850 for PCs, the Mac encoding and the ISO-latin1 coding, notwithstanding eastern European countries which use other ISO-8859 codings. Even the possible 7-bit bypass is often rejected since people who are not computer scientists seem allergic to the 7-bit representation "`r\'esum\'e`" instead of "résumé".

Sending a PostScript file. This is apparently the "good" solution used by everyone in scientific areas. But it may fail for several reasons:

1. All *hard*-scientists have access to at least one PostScript printer, but administrative offices, as well as most private persons, may not due to the high cost of PostScript printers.

2. Even if they can access a PostScript printer, people receiving such a file in an e-mail under Windows have no standard means to send a PostScript file to their PostScript printer: as a matter of fact, Windows provides several "drivers" for PostScript printers, but no driver which does nothing but plain transmission, which is possible only by using the UNIX `lp` or `lpr` commands, or the MS-DOS `copy` command.[1]

3. Other software can solve this problem, but you cannot reasonably ask your correspondent (all of your correspondents in the case of a mailing list) to install either GhostScript, GhostView or `prfile10`.

Sending image files. One could think of sending images of the document, rather than the text with its layout. In fact, this is rather satisfactory if the document is one or two pages: a scanner can be used to produce GIF files, or several packages are available to help the knowledgeable producer of a complex document in converting it from DVI, PostScript, PCL to GIF, a format whose advantage is being compressed. However:

1. not all addressees are aware that they could use their Netscape or Microsoft Explorer to view a local GIF file; and

2. the GIF-bitmap file of a pure text page consumes much more space that the text it contains: the size is no problem for a few pages, but it is for dozens.

How to think of portability. The sender of a LaTeX document – as well as the sender of a C or F77 source program – is therefore faced with a *portability* problem.

Unfortunately, the person who exposes these difficulties is likely to get answers of the form: "Why

---

[1] Which does not work with network connected printers.

ltx2rtf : exporting LaTeX documents to Word addicts

Daniel TAUPIN
laboratoire de Physique des Solides
bât. 510, centre universitaire
F-91405 Orsay Cedex

July 18, 1998

Abstract

ltx2rtf is a compiler that translates LaTeX 2_ε source text into the RTF format used by several text processors, including Microsoft Word and Word for Windows. It was written by Fernando DOHNER and Andreas GRANZER in a one-semester course in Vienna (Austria) and is currently found as latex2rtf in CTAN servers.

It was heavily corrected and adapted to LaTeX 2_ε in 1997 by Daniel Taupin. The distribution was intended mainly for use within the MS-DOS window of Win95 and Win3.11, but all sources can be compiled on UNIX computers having GCC compilers.

1 Introduction: the need for a converter to RTF

Like most of the audience of TUG and other TeXperts' meetings, I usually write most of my papers in LaTeX 2_ε. But problems arise when needing to transmit these documents to non-LaTeX users.

1.1 The various cul-de-sacs when transmitting LaTeX documents

1.1.1 Sending plain text

This obvious (poor) solution fails because accented letters have at least three commonly used codings, the 850 for PCs, the Mac encoding and the ISO-latin1 coding, notwithstanding eastern European countries which use other ISO-8859 codings. Even the possible 7-bit bypass is often rejected since people who are not computer scientists seem allergic to the 7-bit representation "r\'esum\'e" instead of "résumé".

1.1.2 Sending a PostScript file

This is the apparently "good" solution used by everyone in scientific areas. But it may fail for several reasons:

1. All *hard*-scientists have access to at least one PostScript printer, but administrative offices, as well as most private persons, do not due to the high cost of PostScript printers.

2. Even if they can access a PostScript printer, people receiving such a file in an e-mail under Windows have no standard means to send a PostScript file to their PostScript printer: as a matter of fact, Windows provides several "drivers" for PostScript printers, but no driver doing nothing except plain transmission which is possible only by using the UNIX lp or lpr commands, or the MS-DOS copy command[1].

3. Other software can solve this problem, but you cannot reasonably ask your correspondent (all of your correspondents in the case of a mailing list) to install either GhostScript, GhostView or prfile10.

[1] Which does not work with network connected printers.

1

2 The idea of ltx2rtf: using Word as a DVI driver

When sending a document to a variety of addressees, one should think of which software is most widespread among them; the answer is "thanks to Micro$oft's powerful advertisements, they all possess[2] a version of Word[perfect] which can read RTF files.

Anyway, even with its deficiencies, latex2rtf produces a RTF file which is quite satisfactory in the sense that it can be processed using Word, and nicely printed after several manual corrections, without the need to retype the whole of the text and add the font changes.

3 ltx2rtf

3.1 Features

In the same way as latex2html, ltx2rtf compiles the LaTeX source and directly produces RTF output, instead of HTML.

- Part, chapter and section numbering all use Word (and RTF) built-in macros to provide section numbers which can be updated when inserting new sections (i.e. "title" levels) as provided by Word. Optionally, these numberings can be computed by ltx2rtf itself, in such a way they are frozen for further Word updates.

- Conversely, enumerate environments produce frozen numbers, mainly because Word's built-in features inhibit unnumbered paragraphs within that environment.

- Western European accented letters – including capitals – are correctly treated, including the famous ISO-latin1 excluded "œ", the german cs-zet "ß", the latine "æ", the scandinavian "ø" and the spanish "ñ" (as well as Ñ).

But the deficiencies of our Word 6.0 make environment multicol unreliable.

3.2 Implementation

3.2.1 Maths handling

3.2.2 Quality of the result

From the LaTeX-er's view point, output (text and moreover maths) is much better than the results obtained by average 'Wordists', especially with respect to lists. Fox example the famous formula using the discriminant $\sqrt{b^2-4ac}$ of the second degree equations:

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

3.3 Availability

Software can be obtained from:

ftp://ftp.lps.u-psud.fr/pub/ltx2rtf/ltx2rtf.zip

4 Conclusion

In the same way that DVIPS is not intended to help typesetters moving from LaTeX to PostScript, ltx2rtf is not intended to help them moving from LaTeX to Word, but to help them in sending or posting nicely typeset papers, thus multiplying by tens the number of persons able to display and print it on their own Microsoft-addicted devices.

[2] Whether they actually bought the license is the addressee's problem, not mine.

2

Figure 1: Example LaTeX document.

don't you get rid of your Windows system and move to UNIX?", "Why don't you discard your old Epson printer and have a PostScript printer?", "Why don't you move from Microsoft's text editors and use LaTeX?", "Why don't you install GhostScript, GhostView, prfile10, or a Linux partition to your PC?", etc.

All these common sense answers are right, but they just forget one thing: the problem is not with **my** personal installation when sending/mailing a document, the problem is with the installation of the addressees, whose skill I perhaps do not know at all, and who are probably unable to install software other than what they got when buying their personal computer or when registering on some multi-user workstation.

The idea of ltx2rtf: Using Word as a DVI driver

When sending a document to a variety of addressees, one should think of which software is most widespread among them; the answer is "Thanks to Micro$oft's powerful advertisements, they all possess a version of Word[perfect] which can read RTF files.[2]

In fact, whatever many people claim about Microsoft's way of managing its software, RTF specifications are published by this company, and are available at: ftp://ftp.microsoft.com/Softlib/MSLFILES/GC0165.EXE, a self-extracting zipped file yielding a *.DOC file.[3] Thus, using this specification file (130 pages) and testing the actual behaviour of Word,[4] one can obtain a means of producing RTF from a LaTeX source.

This was attempted in 1994 by students at an institution which appears to be a Technical University in Vienna (Austria) and widely posted on CTAN under the name latex2rtf. Their translator is provided as several C source files which can be easily compiled with a satisfactory "makefile".

[2] Whether they actually bought the license is the addressee's problem, not mine.

[3] Unzipping it seems however to fail since that last posting. No comment...

[4] An old Word 6.0 did not exactly respect the specifications...

Figure 2 content:

ltx2rtf : exporting LATEX documents to Word addicts

Daniel TAUPIN
laboratoire de Physique des Solides
bât. 510, centre universitaire
F-91405 Orsay Cedex

Abstract
ltx2rtf is a compiler that translates LATEX2ε source text into the RTF format used by several text processors, including Microsoft Word and Word for Windows. It was written by Fernando DORNER and Andreas GRANZER in a one-semester course in Vienna (Austria) and is currently found as latex2rtf in CTAN servers.
It was heavily corrected and adapted to LATEX2ε in 1997 by Daniel Taupin. The distribution was intended mainly for use within the MS-DOS window of Win95 and Win3.11, but all sources can be compiled on UNIX computers having GCC compilers.

1 Introduction: the need for a converter to RTF
Like most of the audience of TUG and other TEXperts' meetings, I usually write most of my papers in LATEX2ε. But problems arise when needing to transmit these documents to non-LATEX users.

1.1 The various cul-de-sacs when transmitting LATEX documents

1.1.1 Sending plain text
This obvious (poor) solution fails because accented letters have at least three commonly used codings, the 850 for PCs, the Mac encoding and the ISO-latin1 coding, notwithstanding eastern European countries which use other ISO-8859 codings. Even the possible 7-bit bypass is often rejected since people who are not computer scientists seem allergic to the 7-bit representation "r\'esum\'e" instead of "résumé".

1.1.2 Sending a PostScript file
This is the apparently "good" solution used by everyone in scientific areas. But it may fail for several reasons:
1. All *hard*-scientists have access to at least one PostScript printer, but administrative offices, as well as most private persons, do not due to the high cost of PostScript printers.
2. Even if they can access a PostScript printer, people receiving such a file in an e-mail under Windows have no standard means to send a PostScript file to their PostScript printer: as a matter of fact, Windows provides several "drivers" for PostScript printers, but no driver doing nothing except plain transmission which is possible only by using the UNIX lp or lpr commands, or the MS-DOS copy command[1].
3. Other software can solve this problem, but you cannot reasonably ask your correspondent (all of your correspondents in the case of a mailing list) to install either GhostScript, GhostView or prfile10.

[1] Which does not work with network connected printers.

2 The idea of ltx2rtf: using Word as a DVI driver
When sending a document to a variety of addressees, one should think of which software is most wide-spread among them; the answer is "thanks to MicroSoft's powerful advertisements, they all possess[2] a version of Word[perfect] which can read RTF files.
Anyway, even with its deficiencies, latex2rtf produces a RTF file which is quite satisfactory in the sense that it can be processed using Word, and nicely printed after several manual corrections, without the need to retype the whole of the text and add the font changes.

3 ltx2rtf

3.1 Features
In the same way as latex2html, ltx2rtf compiles the LATEX source and directly produces RTF output, instead of HTML.
– Part, chapter and section numbering all use Word (and RTF) built-in macros to provide section numbers which can be updated when inserting new sections (i.e. "title" levels) as provided by Word. Optionally, these numberings can be computed by ltx2rtf itself, in such a way they are frozen for further Word updates.
– Conversely, enumerate environments produce frozen numbers, mainly because Word's built-in features inhibit unnumbered paragraphs within that environment.
– Western European accented letters – including capitals – are correctly treated, including the famous ISO-latin1 excluded "œ", the german es-zet "ß", the latine "æ", the scandinavian "ø" and the spanish "ñ" (as well as Ñ).
But the deficiencies of our Word 6.0 make environment multicol unreliable.

3.2 Implementation

3.2.1 Maths handling

3.2.2 Quality of the result
From the LATEX-er's view point, output (text and moreover maths) is much better than the results obtained by average 'Wordists', especially with respect to lists. Fox example the famous formula using the discriminant $\sqrt{b^2 - 4ac}$ of the second degree equations:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

3.3 Availability
Software can be obtained from:
ftp://ftp.lps.u-psud.fr/pub/ltx2rtf/ltx2rtf.zip

4 Conclusion
In the same way that DVIPS is not intended to help typesetters moving from LaTeX to PostScript, ltx2rtf is not intended to help them moving from LATEX to Word, but to help them in sending or posting nicely typeset papers, thus multiplying by tens the number of persons able to display and print it on their own Microsoft-addicted devices.

[2] Whether they actually bought the license is the addressee's problem, not mine.

Figure 2: Example output from `latex2rtf`.

The C coding is clean and well structured but, unfortunately, the students did not have a knowledge of LATEX of the same quality as their C programming skill; thus many things had to be revised concerning font management, sectioning, `itemize`, `enumerate`, `description` and `tabular` environments, notwithstanding LATEX2ε more recent specifications.

Anyway, even with its deficiencies, `latex2rtf` produces a RTF file which is quite satisfactory in the sense that it can be processed using Word, and nicely printed after several manual corrections, without the need to retype the whole of the text and add the font changes.

ltx2rtf

Features. In the same way as `latex2html`, `ltx2rtf` compiles the LATEX source and directly produces RTF output, instead of HTML.

Part, chapter and section numbering all use Word (and RTF) built-in macros to provide section numbers which can be updated when inserting new sections (i.e. "title" levels) as provided by Word. Optionally, these numberings can be computed by `ltx2rtf` itself, in such a way that they are frozen for further Word updates. Conversely, `enumerate` environments produce frozen numbers, mainly because Word's built-in features inhibit unnumbered paragraphs within that environment.

Western European accented letters – including capitals – are correctly treated, including the famous ISO-latin1 excluded "œ". In the same way, additional abbreviation features provided by Bernard Gaulle's `french.sty` and Daniel Taupin's `smallcap.sty` (which enables a `\scfamily` command instead of `\scshape` to provide bold and/or slanted small capitals).

Implementation.
Basic.

1. The input code can be either 7-bit, or ANSI (ISO-latin1) or 850. The Mac coding is not yet implemented but doing that would not be a problem.

2. The source can be compiled with any GCC compiler (no serious problems with other normal C compilers). We tested it mainly with the DJGPP port of GCC to DOS (native, Win3.11 and Win95).

3. Nothing more is needed, as long as one does not want to translate maths.

4. Maths are tentatively translated using the few RTF mathematical features such as raising parts of the text and changing fonts (size and shape).

Maths handling. Two options are provided for maths handling.

1. The `-m` option uses LaTeX-ing for *displayed equations*, namely those enclosed with `$$` (`equation` environment in the future). Then, nearly in the same way as `latex2html`:

   - `ltx2rtf` calls `latex` to produce a DVI file for each equation;

   - `ltx2rtf` calls an external procedure (`DVI2PBM.BAT` under MSDOS) which, in turn, either calls emTeX's `dvidrv dvidot` to produce PCX files and then calls NETPBM routines to convert the PCX to PBM, or calls DVIPS to produce a PostScript file and then GhostScript to produce a PBM file; [5] and

   - finally, the PBM file is read by `ltx2rtf` itself and converted to "wbitmap" as specified in the RTF specification document.[6]

2. The `-M` option not only uses LaTeX-ing for *displayed equations*, but also for single `$`-enclosed mathematical text.

Quality of the result. From the LaTeX-er's view point, output (text and moreover maths) is much better than the results obtained by average 'Wordists', especially with respect to lists.

Therefore the RTF produced is very good when one wants to e-mail a LaTeX-typeset text to unknown (or known) addressees whose probability of possessing Word is 95%, but of having at least DVI printers/viewers or easy access to PostScript printers is only 5%.

The inconveniences. From the producer's viewpoint, one sees the same installation difficulties as with `latex2html` with the exception that neither Perl nor GDBM/DBM are needed.

But more major inconveniences are seen from the addressee's viewpoint:

- Since Microsoft now rules the computation and networking world, any apparent deficiency in its products must be considered as "not a bug,

but a feature". Therefore, any layout different from what a Wordist usually gets (think of no paragraph hanging indentation in hierarchical lists) may be considered as a negative feature, in the same way as accented capitals or non-english characters, which are so difficult to type (4 clicks and 3 mouse moves to produce æ or œ in French).

- Even worse, the LaTeX-like layout takes advantage of the powerful basic commands of RTF – something near to TeX primitives plus some plain-TeX facilities – but such results are nearly impossible to obtain with Word's ready-made clicking commands.

  The reason is that the `ltx2rtf`-generated format ("format" in the Word sense) is definitely different from those which are provided as standard in Word's clicking windows. This results in the impossibility for the addressee to modify the RTF file except for pure text corrections and, perhaps, changes in sectioning (but not in section numbering style).

- Obviously, math parts are frozen as images which can only be removed, moved, enlarged or shrunk, but not edited.

Availability. The software can be obtained from: `ftp://ftp.lps.u-psud.fr/pub/ltx2rtf/ltx2rtf.zip`.

Conclusion

In the same way that DVIPS is not intended to help typesetters moving from LaTeX to PostScript, `ltx2rtf` is not intended to help them moving from LaTeX to Word, but to help them in sending or posting nicely typeset papers thus multiplying by tens the number of persons able to display and print it on their own Microsoft-addicted devices.

---

[5] Thanks to Emmanuel Bigler who provided this alternate solution.

[6] Other picture specifications are described, but they all fail with Word 6.0; therefore we kept to the only one succeeding.