$L_2\mathcal{A}$ — a LaTeX detergent

Gisle Hannemyr
Norwegian Computing Center

October 19, 2008

Introduction

$L_2\mathcal{A}$ is a filter to "detexify" texts. That is, it attempts to remove LaTeX markup commands, leaving only the body of text. It is intended to be used when journal editors request plain ASCII text for typesetting, or when you want to post a plain version of a LaTeX document on an electronic conference system.

The author's address is:

    Gisle Hannemyr
    Norwegian Computing Center
    P.O. 114 Blindern
    N-0314 Oslo 3
    Norway

EAN:	C=no;PRMD=uninett;O=nr;S=Hannemyr;G=Gisle (X.400 SA)
	gisle.hannemyr@nr.no (RFC-822)
Inet:	gisle@ifi.uio.no
UUCP:	...!mcsun!ifi!gisle

Description

$L_2\mathcal{A}$ is a filter. It reads from standard input and writes to standard output. Typical usage would be:

```
l2a < foobar.tex > foobar.txt
```

It accepts three switches:

`-a` displays copyright etc.

`-h` displays brief help

`-n` uses Norwegian for texts etc.

Current state

$L_2\mathcal{A}$ handles a subset LaTeX. Most if the missing things I plan to add when I need them, but there are some features of LaTeX (e.g. the \kill function) that I can not see how can be handled with lex. Some manual polishing of the output will always be required.

Some Norwegian bias is present in the source code. In particular some Scandinavian characters are translated to their counterparts in the Norwegian/Danish version of the ISO 7-bit character set.

$L_2\mathcal{A}$ works for my style of LaTeX usage, but proably barf when fed other people's input. Consider the current state of $L_2\mathcal{A}$ as a starting point: If you want to use it, then it is up to you to hack it into shape for your style. Btw. if you teach $L_2\mathcal{A}$ new tricks, I would like to get back a copy of your enhancements...

Diagnostics

Unrecognized markup commands generate an error message on the screen. They are also retained in the text, enclosed in a brackets looking like this: @( )@. This style of bracketing was chosen so that it should be simple to

use a text editor to search the output file for these commands and edit the context they appear in.

Footnotes

$L_2\mathcal{A}$ does not recognize such advanced concepts as a "page". To avoid having do deal with pages, it will transform *footnotes* to *endnotes* (i.e. the footnotes are moved to the end of file, and renumbered). $L_2\mathcal{A}$ will take care of the renumbering. It will insert numbers in angle brackets (e.g. $\langle 3 \rangle$) to number the footnotes in the text.

Tables, figures and captions

Tables and figures are stripped from the text. They are however clearly outlined with lines like this[1] in the text:

```
<<=============== NB! Typeset as table. NB! ================>>
<<=========== NB! Please insert figure here. NB! ===========>>
```

This should make the missing bits stand out to the dullest of editors.

Captions for figures and tables are marked like this, with the actual text of the caption on the next line.

```
  --------------- Caption for figure or table: -------------
```

A line like this is used to mark when end of the point where the table or figure should be inserted.

```
<<=============================================================>>
```

Please equip the editor with scissors and glue and refer him/her to the paper version typeset with LaTeX to find the actual figures and tables (or submit them on separate sheets of paper).

---

[1]The text inserted is actually dependent on which language you have selected.

To do

$L_2\mathcal{A}$ is far from complete. It is very weak as far as mathematical mode are concerned (I don't write much mathematics).

However, I hope the current version still are of some use. The lex source is easy to maintain. I suggest that users add the stuff they need when they need it.

The most urgent thins on the "to do" list is functions to handle included files, cross-references, citations and bibliographies.

$L_2\mathcal{A}$ should also have an option to use an 8-bit character set (ISO 8859/1) for accented characters.

Finally, I nurse a secret dream of having $L_2\mathcal{A}$ generate *WordPerfect* or *Word* files preserving italics etc.

Enjoy!