LaTeX News

Issue 33, May 2021

Contents

Introduction

to be written

Extending the hook concept to paragraphs

Largely triggered by the need for better control of paragraph text processing, in particular when producing tagged PDF output (see [1]), we have extended the paragraph processing of LaTeX so that the kernel gains control both at the start and the end of each paragraph. This is done in a manner that is (or should be) transparent to packages and user documents.

Beside the internal control points for exclusive use of the LaTeX kernel we also implemented four public hooks that can be used by packages or user via the hook management declarations to achieve special effects or implement manipulations that in the past were only possible through redefinitions of `\everypar` or `\par` with the usual issue that such changes would conflict with changes in other packages.

The documentation of the hooks together with a few examples is provided in `ltpara-doc.pdf` and for those who want to study the (quite interesting) code is found in `ltpara-code.pdf`. At some point in the future it will get fully merged into the LaTeX kernel and will then appear in `source2e.pdf`.

Other changes to the LaTeX kernel

Adjusting `itemize` labels with `\labelitemfont`

The command `\labelitemfont` was in fact already introduced with the LaTeX release 2020-02-02, but back then we forgot to describe it, so we do this now. Its purpose is to resolve some bad formatting issues with the `itemize` environment and at the same time make it easier to adjust its layout if necessary. What could happen in the past was the the `itemize` labels, e.g., the •, would sometimes react to surrounding font changes and could suddenly change shape, for example to •.

Now `\labelitemfont` is applied to each label defaulting to `\normalfont` which will prevent this behavior. By choosing a different settings other effects can be achieved, for example

```
\renewcommand\labelitemfont
   {\normalfont\fontfamily{lmss}\selectfont}
\renewcommand\labelitemfont
   {\rmfamily\normalshape}
```

The first will take the symbols from Latin Modern Sans so that you get ▪, −, ∗ and ·, while the second variant freezes the font family and shape, but leave the series variable, so that an `itemize` in a bold context would show bolder symbols. Making it empty would give you the buggy old behavior back. *(github issue 497)*

A note on file names with spaces, dots or UTF-8 characters
In one of the the recent LaTeX releases we improved the interface for specifying file names so that they can now safely contain spaces (as is common on Windows but also elsewhere), UTF-8 characters outside the ASCII range as well as names with several dots in it. In the past this was only possible by applying a special syntax (in cases of spaces), not at all for most UTF-8 characters and not in all circumstances for files with several dots.

However, TeX has a built-in rule saying that you can leave out the extension if it is `.tex`. Because of that `\input{file}` or `\input{file.tex}` both load `file.tex` if it exists. While this is convenient most of the time it is a little awkward in some scenarios (for example, when both `file` and `file.tex` exist) and also when you manually try to implement that rule.

LaTeX therefore had one special syntax for `\include` and `\includeonly`: they always expected that their arguments contains a file name[1] without its extension, which had to be `.tex`. Thus when you mistakenly wrote `\include{mychap.tex}` (for example, when you changed from `\input` to `\include` somewhere), LaTeX went ahead and looked for the file `mychap.tex.tex` for inclusion and tried to write support information to the file `mychap.tex.aux`. The reason was that `\include` had to construct both physical file names from the argument and it didn't bother to do something special about the extension `.tex`.

As a side effect of the new implementation this has now changed and the argument of `\include` now gets the extension `.tex` removed if it was used. Thus `\include{mychap.tex}` now loads `mychap.tex` and no longer looks for `mychap.tex.tex`. *(github issue 486)*

`\end{document}` should always start in v-mode
Until now `\end{document}` executed the code from the `\AtEndDocument` hook as its first action. This meant that it was executed in horizontal mode if the user left no empty line after the last paragraph. As a result one could get a spurious space added, for example, when that code contained a `\write` statement. This was fixed and now `\enddocument` first issues a `\par` to ensure that it always starts out in vertical mode. *(github issue 385)*

Allow extra space between name and address in letter class
The `\opening` command in the the letter class expects the name and address to be separated by `\\` but it didn't allow to use an optional argument at this point to add some extra space after the name. The coding has now been slightly altered to allow for this. *(github issue 427)*

Add a Lua callback to ltshipout to provide a uniform location for applying custom attributes
Just before shipping out a page, a new LuaTeX callback `pre_shipout_filter` is now called to allow final adjustments to the box to be shipped out. This is particularly for LuaTeX packages which flag certain elements of the page (e.g. using attributes or properties) in order to apply certain effects to these elements at shipout. An example for this is the luacolor package which could insert the color commands using this callback.

Improved copy&paste support for pdfTeX documents
When compiling with pdfTeX, additional information is added to the PDF file to improve copying from and searching in text. This especially allows ligatures to copy correctly from pdfTeX generated PDF files in most cases.

Since this has been integrated into the kernel, most documents should no longer need to load the cmap package or input `glyphtounicode`. *(github issue 465)*

Provide a hook in `\selectfont`
After `\selectfont` has altered the font we run a hook so that packages can make final adjustments. This functionality was originally provided by the everysel package, the new implementation is slightly different and uses the standard hook management. *(github issue 444)*

Delay change of font series and shape to `\selectfont` call
With the NFSS extensions introduced in 2020 the font series and shape settings be be influenced by changes to the font family. The setting is therefore delayed until `\selectfont` is executed to avoid unnecessary or incorrect substitutions that may otherwise happen due to the order of declarations. *(github issue 444)*

Allow `\nocite` in preamble
A natural place for `\nocite{*}` would be the preamble of the document, but for historical reasons LaTeX issued an error message if it was placed there. From the new release on it is now allowed in the preamble. *(github issue 424)*

Shipping out a page while bypassing hooks
In the 2020 October release several hooks were added to the page shipout process, e.g., to add some background or foreground material to some or all pages. We now also added a `\RawShipout` command that bypasses most of these hooks during the shipout. Some essential internal bookkeeping still takes place such as updating the `totalpages` counter or adding `shipout/firstpage`

[1] In case of `\includeonly` a comma separated list of such names.

or `shipout/lastpage` material if the page happens to be the first or last.

Robust commands in filename arguments

The filename handling has been modified so that `\string` is applied while normalizing robust commands while determining the file name. Previously `\input{\sqrt{2}}` would cause LaTeX to loop indefinitely. With the new behavior it accesses `sqrt {2}.tex`. *(github issue 481)*

Additional support for Unicode characters from the Latin Extended Additional block

LaTeX is quite capable of typesetting characters such as "ṃ", but until now it lacked the Unicode mappings for some characters that are used to write Sanskrit words in Latin transliteration (as seen in books about yoga, Buddhist philosophy, etc.). These have now been added so that such characters can be entered directly instead of resorting to `\d{m}` and so forth. *(github issue 484)*

Always have color groups set up

To use color in LaTeX certain constructs, especially boxes, need an extra layer of groups to ensure that the color setting does not *escape* and continue outside the box when it shouldn't. To arrange for this the LaTeX kernel defined a number of commands, e.g., `\color@begingroup` to be used in such places. They have been initially no-ops and only the color packages redefined them to become real groups. This arrangement complicates the coding as one has to account for a group being there (or not there) depending of what is loaded in the document. So now the kernel already adds the groups. *(github issue 488)*

Execute `\par` at the end of `\marginpar` arguments

In preparation for tagged PDF it is important to properly tag all paragraphs and this requires running code at the beginning and end of each. At the end of a paragraph this is done inside the `\par` command, but the way `\marginpar` was coded, LaTeX ended the marginal note without ever explicitly calling `\par`. This has now been changed. *(github issue 489)*

Producing several footnote marks to one footnote

It is sometimes necessary to reference the same footnote several times, i.e., produce several footnote marks with the same number or symbol. This is now always possible by placing a `\label` into the `\footnote` and reference it with the command `\footref` elsewhere. This way marks refering to footnotes anywhere on the page (including those in `minipage`s) can be generated. In the past this command was only available with certain classes or when loading the footmisc package. *(github issue 482)*

Providing the raw option list of packages or documentclass to key/value handlers

LaTeX 2_ε has always normalized space in option lists so `\documentclass[ a4paper , 12pt ]{article}` processed the intended options `a4paper` and `12pt`.

Unfortunately the mechanism used was designed for the simple option names of the standard option processing. Many classes and packages now use extended *keyval* processing, however this white space normalisation makes this difficult: `[bb=1 2 3 4]` which might be expected to pass a bounding box of four numbers is normalised to `bb=1234` and `[bb={1 2 3 4}]` which might be expected to quote the spaces results in low level TeX parsing errors.

For compatibility reasons, the standard option processing has not been changed however the original un-normalised package and class option lists are now saved. They are not used in the standard processing, however extended package option systems may use these "raw" option list macros if they are defined.

The one change affecting the standard processing is that the low level error mentioned above is now avoided as values (any tokens to the right of an = sign) are removed from consideration from the "unused option list". In this release `clip=true` and `clip=false` both contribute `clip` to the list of options that have been used. *(github issue 85)*

Poor man's `\textasteriskcentered` if missing

The `\textasteriskcentered` symbol, used as part of the set of footnote symbols in LaTeX, is assumed to be implemented by every font in the `TS1` encoding (when pdfTeX is used) or in the `TU` encoding for the Unicode engines. Unfortunately, that assumption is not correct for all fonts, for example, for the `stix2` fonts don't offer the glyph, with the result that one gets missing glyphs when using `\thanks` etc.

For that reason the definition for `\textasteriskcentered` was altered to check if there is a glyph in the right position and if not a normal "*" is used, slightly enlarged and lowered. That may not be perfect in all cases, but certainly better than nothing show up. *(github issue 502)*

Provide more "dashes" in encodings `OT1`, `T1` and `TU`

When pasting in text from external sources one sometimes encounters the Unicode characters `"2011` (non-breaking hyphen), `"2012` (figure dash) and `"2015` (horizontal bar) in addition to the common `"2013` (en-dash) and `"2014` (em-dash). In the past the first three characters produced an error message when used with pdfTeX. Now they typeset an approximation (as they are unavailable in `OT1` or `T1` encoded fonts used by pdfTeX), e.g., the figure dash is approximated by an en-dash.

In Unicode engines they either work (if contained in the selected Unicode font) or typeset nothing and produce a "Missing character" warning in the log file.

However, what works in all engines now, is to access the characters via the command names `\textnonbreakinghyphen`, `\textfiguredash` and `\texthorizontalbar`, respectively. *(github issue 404)*

. . .
to be written *(github issue xxx)*

Changes to packages in the graphics category

. . .

(github issue xxx)

Changes to packages in the tools category

layout: Support Japanese as a language option
The package now recognizes `japanese` as a language option. *(github issue 353)*

longtable: General bug fix update
Minor update to longtable to fix bugs reported. Notably the possibility of incorrect page breaks if floats appear on the same page that a `longtable` starts. As this may affect page breaking in existing documents, a rollback to longtable 4.13 (`longtable-2020-01-07.sty`) is supported. *(gnats issue tools/3512)*

Changes to packages in the amsmath category

. . .

(github issue xxx)

References

[1] Frank Mittelbach and Chris Rowley: *LaTeX Tagged PDF — A blueprint for a large project.* `https://latex-project.org/publications/indexbyyear/2020/`