

fifinddo

Filtering T_EX(t) Files by T_EX

Uwe Lück*

April 4, 2010

FIDO, FIND!
or:
FIND FIDO!
oder:
FIFI, SUCH!

Abstract

`fifinddo` starts implementing parsing of plain text or T_EX files using T_EX, generalizing the philosophy behind `docstrip`, based on how T_EX reads macro arguments. Rather than typesetting the edited input stream immediately, results are written to another file, in the first instance as input for T_EX. Rather than presenting a “complete study” of a computer-scientific idea, it aims at practical applications. The main one at present is `makedoc` which removes certain comment marks from package files and inserts listing commands. Parsing macros are not defined anew at every input chunk, but once before a file is processed. This also allows for *expandable* sequences of replacements, e.g., with `txt`→T_EX functionality. The method of testing for substrings is carefully discussed, revealing an earlier mistake (then) shared with `substr.sty` and L^AT_EX’s internal `\in@`.

Contents

1	Introduction: The Gnome of the Aim	2
1.1	Parsing by T _E X—are you mad?	2
1.2	Useful for	3
1.2.1	Comparisons	5
1.3	For insiders	5

*<http://contact-ednotes.sty.de.vu>

1	INTRODUCTION: THE GNOME OF THE AIM	2
2	Preliminaries	6
2.1	Head of file (Legalese)	6
2.2	Format and package version	7
2.3	Category codes	7
3	File handling	8
4	Basic handling of substring conditionals	9
4.1	“Substring Theory”	9
4.2	Plan for proceeding	11
4.3	Meta-Setup	11
4.4	Setup for conditionals	12
4.5	Setup for sandboxes	12
4.6	Getting rid of the tildes	14
4.7	Calling conditionals	14
4.8	Copy jobs	15
5	Programming tools	15
5.1	Tails of conditionals	15
5.2	Line counter	16
5.3	The “identity job” LEAVE	17
6	Setup for expandable chains of replacements	17
7	Leave package mode	19
8	Pondered	19
9	VERSION HISTORY	20

1 Introduction: The Gnome of the Aim

1.1 Parsing by T_EX—are you mad?

The package name `fifinddo` is a `\listfiles`-compatible abbreviation of ‘file-find-do’¹ (or think of ‘if found do’). `fifinddo` implements (or aims at) general parsing (extracting, replacing [converting], expanding, ...) using T_EX where `texhax` posters strongly urge to use `sed`, `awk`, or Perl. `fifinddo`’s opposed rationales are:

- It works instantly on any T_EX installation. (*Restrictions:* Some T_EX versions `\write` certain hex codes for certain characters, cf. T_EXbook p. 45,

¹‘file’ possibly for “searching T_EX(t) files” (I don’t remember my thoughts!), while there were requests for doing replacements on L^AT_EX *environments* on `texhax`. However, the package might be enhanced in this direction ... so the name may be wrong ... but now I like it so much ... Or the reason was that results are written to a *separate file*, not typeset immediately.—Let me also mention that ‘*Fiji*’ (as the package name starts) is a kind of German equivalent to the “English” ‘*Fido*’, or may have been.

I have seen this with PCT \TeX . However, some applications of `fifinddo` are nothing but technical steps where you will read the result files rarely anyway.

- You can apply and customize it like any \TeX macros, knowing just \TeX (or even only the documentation of some user-friendly extension of `fifinddo`), without the need of learning any additional script language.
- The syntax of usual utilities (e.g., “wildcards”) is sometimes difficult with \TeX files with all their backslashes, square brackets, stars, question marks ...

At least the first item is just the philosophy of the `docstrip` program, standard for installing \TeX packages; and while I am typing this, I find at least 14 other similar packages in Jürgen Fenn’s *Topic Index* of the *\TeX Catalogue*:

<http://mirror.ctan.org/help/Catalogue/bytopic.html#parsingfiles>²

(Some of them may have been *reactance* to `texhax` and other postings urging not to try something like this; some seem just to be celebrations of the power of \TeX —yes, celebrate!)

Actually, \TeX ’s mechanism of collecting macro arguments is hard-wired parsing at quite a high level. L \TeX hides this from “simple-minded” users by a convention *not* to use that full power of \TeX for *end-user macros*. Internally, L \TeX *does* use it in reading lists of options and file dates as well as to implement certain FOR- and WHILE-like loop programming structures. L \TeX ’s `\in@/\ifin@` construction is an implementation of a “ $\langle string1 \rangle$ occurs in $\langle string2 \rangle$ ” test. More packages seem to use this idea for extracting file informations, like `texshade`.³

However, such packages don’t make much ado about parsing, there seems to be no general setup mechanism as are presented by `fifinddo`. Indeed, tailoring parsing macros to specific applications may often be more efficient than a general approach.

1.2 Useful for ...

My main application of `fifinddo` at present is typesetting documentations of packages using `makedoc` which removes certain percent marks and inserts listing commands, so you edit a package file with as little documentation markup as possible. This may be extended to other kinds of documents as an alternative to `easylatex` or `wiki` (the approach of which is dangerous and incompatible with certain other things).

I have used a similar own package `txtproc` successfully, where more features were implemented for practical purposes than are here so far, yet I don’t like its implementation, want to improve it here. This package also *created batch files*, e.g., to remove temporary files. This could be used for package handling:

²Click here!

³<http://ctan.org/pkg/texshade>

typset the documentation at the desired place in the tree, write the packages to another, write a batch file to remove files that are not needed any more after installation (cf. `make`).

I used `txtproc` also for *large-scale substitutions* (it had been decided to change the orthography in a part of a book). Other large-scale substitutions may be:

- inserting `\index` commands;
- inserting (soft) hyphenation commands near accents;
- manual umlaut-conversion.⁴
- typographical (or even orthographical) corrections (same mistake many times on each of hundreds of pages). You may turn `...` into `\dots` and `etc.` into `etc.\ etc.`⁵ This could replace packages like `easylatex`,⁶ `txt2latex`,⁷ `txt2tex`⁸ in a customizable way, using, e.g., the “correct” hook from `makedoc.sty` as exemplified in `mdoccorr.cfg` (see examples section of `makedoc.pdf`). You should find `fdtxtex.tpl`, a `fifinddo` script to try or apply `\MakeDocCorrectHook` from `mdoccorr.cfg`, as well as `fdtxtex.tex` that runs a dialogue for the same purpose if you can manage to run it (WinShell?). You can then try to create your own `\MakeDocCorrectHook`. Section 6 provides setup for macros of this kind.
- as to `easylatex` again, *lists* could be detected and transformed into \LaTeX list commands. This could re-implement the lists functionality of `wiki.sty` that is somewhat dangerous.
- introduce your own *shorthands* to be expanded not as \TeX macros, but by text substitution;

In certain cases, insertions deteriorate readability, hyphenation corrections even make text search difficult. It is therefore suggested to

1. keep editing the file without the insertions,
2. run the script (commands based on `fifinddo`) for insertions in the preamble of the main file (“`\jobname.tex`”, maybe `\input` the script file) and
3. `\input` the result file within the document environment.

In general, differences to “manual” replacing by the substitution function of your *text editor* is that

- you first keep the original version,

⁴If you know the “names” of the encodings, Heiko Oberdiek’s `stringenc` may be preferable.

⁵But what when a new sentence is starting indeed? Well, `cf.` is an easier example.—`etc.` even showed a problem in `niceverb`. `mdoccorr.cfg` replaces `etc.` only, so you can keep the extra space by a code line break.

⁶<http://ctan.org/pkg/easylatex>

⁷<http://ctan.org/pkg/txt2latex>

⁸<http://ctan.org/pkg/txt2tex>

- you can check the resulting file before you replace the original file by it,
- you can store the replacement script in order to check for mistakes at a later stage of your work,
- you can do *all* the replacements in *one run* (by *one* script to check for mistakes),
- you can store replacement scripts for future applications, so you needn't type the patterns and replacement strings anew.

1.2.1 Comparisons

It should be noted (perhaps here) that the present approach to parsing is a quite *simple* one and in this respect much different to the string handling mechanisms of `stringstrings`,⁹ `ted`,¹⁰ `xstrings`¹¹ (as I understand them, perhaps also `coolstr`¹²) which are *much more powerful* than what is offered here—but perhaps slow and for practical applications possibly replaceable by the present approach. *Expandable replacement* seems not to exist outside `fifinddo` (2009/04/13).

Much is missing, I know.¹³ I am just implementing what I actually need and what could show that this approach is worth being pursued.

1.3 For insiders

Warning: You may (at least at the present state of the work) have little success with this package, if you don't know about T_EX's category codes and how T_EX macros are defined. The package rather provides tools for package writers. You may, however, be able to run other packages which just load `fifinddo` as required background.

That `fifinddo` acts on “T_EX(t)” files or so means that (at present) I think of applications on “plain text” files which will usually be T_EX input files. “At present” they are read without “special characters,” so essentially category codes of input characters are either 11 (“letter”) or 12 (“other”). This way some things are easier than with usual T_EX applications:

1. You can “look into” curly braces and “behind” comment characters.
2. There are exact or safe tests especially of *empty macro arguments* that are “expandable,” i.e., they are “robust,” don't need assignments, can be executed in `\writeing` and in `\edef` definitions. “Usually,” the safe way to test emptiness is storing a macro argument as a macro, say `\tempo`, in order to test `\ifx\tempo\empty` where `\empty` has been defined by `\def\empty{}` in the format. But this requires some `\def\tempo{#\langle n \rangle}`

⁹<http://ctan.org/pkg/stringstrings>

¹⁰<http://ctan.org/pkg/ted>

¹¹<http://ctan.org/pkg/xstrings>

¹²<http://ctan.org/pkg/coolstr>

¹³There is more in my badly implemented `txtproc.sty`.

which breaks in “mere expanding” (T_EX *evaluates* `\tempo` instead of defining it). An *expandable* test on emptiness is, e.g. `\ifx$#<n>$`, where we hope that it becomes `\iftrue` just if macro argument `#<n>` is empty indeed. However, “usually” it may *also* become `\iftrue` when `#<n>` starts with `$`—if the latter has category code 3 (“math shift”). But `fifinddo` does not assign category code 3 to any character from the input file! Therefore `\ifx$#<n>$` is `\iftrue` *exactly* if `#<n>` is empty.

3. You can avoid interference with packages that are needed for typesetting. You can do the “preprocessing” in one run with typesetting, but you should do the preprocessing before you load packages needed for typesetting. One may even try to keep the macros and settings for preprocessing local to a group.

Once there may be an option to read input with some usual T_EX category codes as well, it may be useful to (some of)

- avoid matching substrings of control words,
- skip blank spaces as T_EX does it usually,
- catch *balanced* input pieces,
- ignore comments,
- ignore certain characters.

The essential approach of `fifinddo` to looking for single strings is described in some detail in section 4.

The implementation of `fifinddo` is as follows. User commands are specially highlighted (boxed/coloured), together with their syntax description.

2 Preliminaries

2.1 Head of file (Legalese)

```

1  %% Macro package 'fifinddo.sty' for LaTeX2e,      %% FIDO, FIND!
2  %% copyright (C) 2009 2010 Uwe L"uck,
3  %%   http://www.contact-ednotes.sty.de.vu
4  %% -- author-maintained in the sense of LPPL below --
5  %% for processing tex(t) files
6  %% (checking, filtering, converting, substituting, expanding, ...)
7
8  \def\fileversion{0.4a} \def\filedate{2010/04/04}
9
10 %% This file can be redistributed and/or modified under
11 %% the terms of the LaTeX Project Public License; either
12 %% version 1.3a of the License, or any later version.
13 %% The latest version of this license is in

```

```

14 %%
15 %%      http://www.latex-project.org/lppl.txt
16 %%
17 %% We did our best to help you, but there is NO WARRANTY.
18 %% Please report bugs, problems, and suggestions via
19 %%
20 %%      http://www.contact-ednotes.sty.de.vu
21 %%
22 %% For the full documentation, look for 'fifinddo.pdf'.
23 %% Its source starts in 'fifinddo.tex'.

```

2.2 Format and package version

```

24 \NeedsTeXFormat{LaTeX2e}[1994/12/01]
25 % 1994/12/01: \newcommand* etc.
26 \ProvidesPackage{fifinddo}[\filedate\space v\fileversion\space
27      filtering TeX(t) files by TeX (UL)]

```

2.3 Category codes

We use the “underscore” as “compound identifier.”

```

28 \catcode'\_ =11 %% underscore used in control words

```

`\MakeOther` is a synonym for `\@makeother`, needed for matching special characters from the input file. It is exemplified by `\fdPatternCodes` which is the default of `\PatternCodes`. The latter is used in setup macros for reading patterns. We offer `\SetPatternCodes{commands}` (redefining `\PatternCodes`) and `\ResetPatternCodes` (for returning to `\fdPatternCodes`) so setup scripts such as `mdocorr.cfg` have shorter lines.

```

29 \@ifundefined{MakeOther}{\let\MakeOther\@makeother}{}
30 \newcommand*{\fdPatternCodes}{\MakeOther\&\MakeOther\$}
31 \newcommand*{\SetPatternCodes}{\def\PatternCodes}
32 \newcommand*{\ResetPatternCodes}{\let\PatternCodes\fdPatternCodes}
33 \newcommand*{\PatternCodes}{} \ResetPatternCodes
34 %% TODO adding/removing; '*' may be wrong 2010/03/29

```

It would be bad to have `\MakeOther%` and `\MakeOther\` here in that this may have unexpected, weird effects with arguments of setup macros. (With `\MakeOther\` you must not indent within a setup command, and if you add `\MakeOther%` the setup command must stay in one line.) Therefore neither `\dospecials` nor `\@sanitize` are used. Curly braces remain untouched as default delimiters in setup macros. For matching them, you must use `\MakeOther\{` and `\MakeOther}` in your `\PatternCodes`, or `\Delimiters` to introduce new ones at the same time, e.g., `\Delimiters\[\]`:

```

35 \newcommand*{\Delimiters}[2]{%
36   \MakeOther\{\MakeOther}\catcode'#1=1\catcode'#2=2\relax}

```

For replacing strings or for defining other strings of “other” characters by `\edef`, you can use some L^AT_EX constructs—here are copies `\PercentChar` and `\BackslashChar` of them (do you need more?):

```

37 \newcommand*{\PercentChar}{} \let\PercentChar\@percentchar
38 \newcommand*{\BackslashChar}{} \let\BackslashChar\@backslashchar

```

3 File handling

```

39 \newwrite\result_file %% or write to \@mainaux!?

```

`\ResultFile{⟨output⟩}` opens (and empties) a file ⟨output⟩ to be written into.

```

40 \newcommand*{\ResultFile}[1]{%
41   \def\result_file_name{#1}%
42   \immediate\openout\result_file=#1}

```

`\WriteResult{⟨balanced⟩}` writes a ⟨balanced⟩ line into ⟨output⟩ (or more lines with `^^J`).

```

43 \newcommand*{\WriteResult}[1]{%
44   \immediate\write\result_file{#1}}

```

`\WriteProvides` writes a `\ProvidesFile` command to the opened ⟨output⟩ file. This should be used when ⟨output⟩ is made as L^AT_EX 2_ε input.

```

45 \newcommand*{\WriteProvides}{%
46   \WriteResult{%
47     \string\ProvidesFile{\result_file_name}%
48     [\the\year/\two@digits\month/\two@digits\day\space
49     automatically generated with fifinddo.sty]]}%

```

`\ProcessFileWith{⟨input⟩}{⟨loop-body⟩}` opens a file ⟨input⟩ and runs a loop on its lines the main body of which is ⟨loop-body⟩. When the ⟨loop⟩ starts, a new line of ⟨input⟩ is stored as macro `\fdInputLine`.

```

50 \newcommand*{\ProcessFileWith}[2]{%
51   \openin\@inputcheck=#1%
52   % \ifeof\@inputcheck %% bad 'exists?' test
53   % \PackageError{fifinddo}{File '#1' not here}%
54   % {Mistyped?}%
55   % \else
56   \global\c@fdInputLine=\z@ %% line counter reset
57   \begingroup
58   \MakeOther\{\MakeOther\}\@sanitize
59   %% from docstrip.tex:
60   % \MakeOther\^^A\MakeOther\^^K%% irrelevant, not LaTeX
61   \endlinechar\m@ne
62   %% <- cf. TeXbook "extended keyboards" up-/downarrow
63   %% -> "math specials", cf. "space specials"
64   \MakeOther\^^I% ASCII horizontal tab -- guessed!? ^^L!?

```

With v3.1, we support non-ASCII:


```

65     \count@=128
66     \loop
67     \ifnum\count@<\@ccclvi
68     \catcode\count@=12
69     \advance\count@\@ne
70     \repeat
71     \loop \ifeof\@inputcheck \else
72     \read\@inputcheck to \fdInputLine
73     \ignorespaces #2%
74     \repeat
75     \endgroup
76 % \fi
77 \closein\@inputcheck}

```

`\CloseResultFile` closes $\langle output \rangle$.

```
78 \newcommand*{\CloseResultFile}{\immediate\closeout\result_file}
```

TODO: write EOF for debugging!?!—Peter Wilson’s newfile provides more powerful file handling.

TODO move theory to ffinddo.tex 2009/04/12

4 Basic handling of substring conditionals

4.1 “Substring Theory”

*I wished I could study string theory,
but I only could study substring theory.*

A \TeX macro, say, `\find` whose parameter text (cf. \TeX book p. 203) starts with `#1<pattern>#2&` stops \TeX with an error if it does not find $\langle pattern \rangle$ and then `&`. Otherwise we have a situation `\find<split1><pattern><split2>&`, and `\find` reads $\langle split1 \rangle$ as `#1` and $\langle split2 \rangle$ as `#2`. An important point to note is that $\langle split1 \rangle$ will not contain $\langle pattern \rangle$, but possibly $\langle pattern \rangle$ has more occurrences in $\langle split2 \rangle$. In this sense, `\find` uses the *first* occurrence of $\langle pattern \rangle$ it finds in order to delimit `#1`. Finding the *last* occurrence of $\langle pattern \rangle$ therefore needs a special idea.

In order to use `\find` for a test whether $\langle pattern \rangle$ is in $\langle target \rangle$, we build a “sandbox” `\find<sand>&`, where $\langle sand \rangle$ contains $\langle target \rangle$ and additionally $\langle pattern \rangle$ —as a “dummy;” so `&` delimits the search and `\find` finds $\langle pattern \rangle$ either in $\langle target \rangle$ or somewhere else before `&`.

Consider the simple sandbox `\find<target><pattern>&`. We can test `#1` and `#2` on being empty by `\ifx$#1$` and `\ifx$#2$`. If `#2` is empty, $\langle pattern \rangle$ is *not* in $\langle target \rangle$. If `#1` is empty at the same time, $\langle target \rangle$ is empty. If `#1` is empty and `#2` is not, $\langle pattern \rangle$ *starts* $\langle target \rangle$! This can be used to implement Wikipedia-like lists and to distinguish package code from comments in `makedoc`.

If #2 is *not* empty, $\langle pattern \rangle$ occurs in $\langle target \rangle$ —or this once was *thought*, some time in developping the present package, as well as in the version of `substr.sty` marked 2005-11-29,¹⁴ try (if that version still is installed)¹⁵

```
\IfSubStringInString{\langle str1 \rangle \langle str2 \rangle \langle str1 \rangle}{\langle str1 \rangle \langle str2 \rangle}{YES}{NO}
```

which works *verbatim* as well as considering $\langle str1 \rangle$ and $\langle str2 \rangle$ *placeholders*, e.g., for

```
\IfSubStringInString{day_after_day}{day_after_}{YES}{NO}16
\IfSubStringInString{AMSTERDAM}{AMSTERD}{YES}{NO}
\IfSubStringInString{TORONTO}{TORON}{YES}{NO}
\IfSubStringInString{bonbon}{bon}{YES}{NO}17
\IfSubStringInString{bonobo}{bono}{YES}{NO} (an ape)
```

or `\IfSubStringInString{ionization}{ionizat}{YES}{NO}`.¹⁸ Same with L^AT_EX's internal `\in@`.¹⁹

```
\makeatletter\in@{bonbon}{bon}\ifin@_YES\else_NO\fi\makeatother
```

In general, the previous approach *fails if and exactly if* $\langle pattern \rangle$ has a *period* p —less than its length—in the sense of that the p th token to the right or left of each token in $\langle pattern \rangle$ is the *same* token. `AMSTERDAM` has a period 8, `day_after_day` 10, `bonbon` 3, `bonobo` 4. There is a counterexample $\langle target \rangle$ of length p iff $\langle pattern \rangle$ has period p , namely the first substring of $\langle pattern \rangle$ having length p . If the length of $\langle pattern \rangle$ exceeds a multiple mp of its period, the first mp tokens of $\langle pattern \rangle$ form a counterexample $\langle target \rangle$.

Therefore, a sandbox must have something between $\langle target \rangle$ and $\langle pattern \rangle$.²⁰ We choose `\find<target>~<pattern>$&` as standard. The `$` will be used as an argument delimiter to get rid of the dummy $\langle pattern \rangle$ in $\langle split2 \rangle$, as well as to decide whether the match was in $\langle target \rangle$ or in the dummy part of the sandbox. The `$` can be replaced by another tilde `~` in order to test whether $\langle target \rangle$ *ends* on a $\langle pattern \rangle$, defining a macro like `\findatend` whose parameter text starts with `#1<pattern>~#2&`.

¹⁴<http://ctan.org/pkg/substr>. `substr` does not change category codes as `fifinddo` does and uses `\@nil` as delimiter instead of our `&`.

¹⁵The “feature” has been fixed with v1.2 as of 2009/10/20 of `substr.sty`.

¹⁶Likewise `t^ete-\‘a-t^ete ...`

¹⁷Polynesian: `aku aku, rongorongo, wiki wiki ...`

¹⁸Read `substr.sty` or try “normal” things to convince yourself that the syntax indeed is `\IfSubStringInString{\langle pattern \rangle}{\langle target \rangle}{\langle yes \rangle}{\langle no \rangle}`.

¹⁹`\in@` has been fixed after my warning on Heiko Oberdiek’s proposal—at least in the repository.—On 2009/04/21 I learn from Manuel Pégourié-Gonnard that the first versions of his `ted` had a similar bug, fixed on v1.05 essentially like here; Steven Segletes confirms that his `stringstrings` doesn’t suffer the problem (returning positions of substrings and numbers of occurrences).

²⁰Must? Actually, I preferred this solution to other ideas like measuring the length of $\langle split2 \rangle$.

4.2 Plan for proceeding

When we check a file for several patterns, we seem to need *two* macros for each pattern: one that has the pattern in its parameter text and one that stores the pattern for building the sandbox.²¹ We use a separate “*name space*” for each of both kinds. The parsing macro and the macro building the sandbox will have a common “*identifier*” by which the user or programmer calls them. Actually, she will usually (first) call the sandbox box builder. The sandbox builder calls the parsing macro. When *all* occurrences of a pattern in the target are looked for, the parser may call itself.

Actually, the parsing macro will execute certain actions depending on what it finds in the sandbox, so we call it a “*substring conditional*”. It may read additional arguments after the sandbox that store information gathered before. This is especially useful for designing “*expandable*” chains (sequences) of conditionals where macros cannot store information in macros. The macro setting up the sandbox will initialize such extra arguments at the same time.

It may be more efficient *not* to use the following setup macros but to type the macros yourself, just using the following as templates. The setup macros are especially useful with patterns that contain “special characters,” as when you are looking for lines that might be package comments.

4.3 Meta-Setup

A setup command $\langle setup-cmd \rangle$ will have the following syntax:

$$\langle setup-cmd \rangle \{ \langle job-id \rangle \} [\langle changes \rangle] \{ \langle pattern \rangle \} \langle more-args \rangle$$

$\langle changes \rangle$ will, in the first instance, be category code changes for reading $\langle pattern \rangle$ overriding the settings in `\PatternCodes`. They are executed after the latter in a local group. It may be safer to redefine `\PatternCodes` instead of using the optional $\langle changes \rangle$ argument.

A macro

$$\backslash\text{StartFDsetup}\{ \langle do-setup \rangle \} \{ \langle job-id \rangle \} [\langle changes \rangle]$$

shared by setup commands may read $\langle job-id \rangle$ and $\langle changes \rangle$ for $\langle setup-cmd \rangle$. $\langle do-setup \rangle$ will be the macro that reads $\langle pattern \rangle$ (and more) and processes it. It must contain `\endgroup` to match `\begingroup` from `\FD_prepare_pattern`. $\langle job-id \rangle$ is stored in a macro $\backslash\text{fdParserId}$. The default for $\langle changes \rangle$ is *nothing*.

```

79 \newcommand*\StartFDsetup}[1]{%
80   \let\FD_do_setup#1%
81   \afterassignment\FD_prepare_pattern
82   \def\fdParserId}
83 \newcommand*\FD_prepare_pattern}[1][ ]{%
84   \begingroup \PatternCodes #1\FD_do_setup}

```

²¹If it were for the pattern only, the parsing macro might suffice and the macro calling it might extract the pattern from a “dummy expansion.” Somewhat too much for me now; on the other hand the calling macro also hands some “current” informations to the parsing macro—oh, even this could be handled by a general “calling” macro ...

So $\langle setup\text{-}cmd \rangle$ should be set up about as follows:

```
\newcommand*{\langle setup\text{-}cmd \rangle}{\StartFDsetup\langle do\text{-}setup \rangle}
\newcommand*{\langle do\text{-}setup \rangle}[\langle args \rangle]{\langle action \rangle}
```

$\langle do\text{-}setup \rangle$'s first argument will be the $\langle pattern \rangle$ argument of $\langle setup\text{-}cmd \rangle$.

4.4 Setup for conditionals

`substr_cond` is the “name space” for substring conditionals. A colon separates it from “*job identifiers*” in the actual macro names.

```
85 \def\substr_cond{substr_cond:}
```

`\MakeSubstringConditional{\langle id \rangle}[\langle changes \rangle]{\langle pattern \rangle}` starts the definition of a conditional with identifier $\langle id \rangle$ and pattern $\langle pattern \rangle$. $\langle changes \rangle$ optionally add commands to be executed after `\PatternCodes` in a local group.

```
86 \newcommand*{\MakeSubstringConditional}
87     {\StartFDsetup\mk_substr_cond}
```

`\begingroup \mk_substr_cond{\langle pattern \rangle}` can be directly called by other programmer setup commands when `\fdParserId` and $\langle pattern \rangle$ have been read.

```
88 \def\mk_substr_cond #1{%% #1 pattern string
89     \endgroup \@namedef{\substr_cond \fdParserId}##1##2&}
```

This really is not L^AT_EX. We are starting defining a macro `\substr_cond:\langle id \rangle` in primitive T_EX with `\def` in the form

```
\def\substr_cond:\langle id \rangle#1\langle pattern \rangle#2&
```

where `\csname` etc. render ‘: $\langle id \rangle$ ’ part of the macro name. The user or programmer macro produces the part of the definition until the delimiter `&` to match the sandbox. You have to add (maybe) `#3` etc. and the $\{\langle definition\text{-}text \rangle\}$ just as with primitive T_EX.

4.5 Setup for sandboxes

There was a *question*: will we rather see *string macros* or *strings from macro arguments*? The input file content always comes as `\fdInputLine` first, so we at least *must account* for the possibility of string macros as input.

One easy way to apply several checks and substitutions to `\fdInputLine` before the result is written to $\langle output \rangle$ is `\let\OutputString\fdInputLine` and then let `\OutputString` be to what each job refers as *its* input and output, finally `\WriteResult{\OutputString}`. (`\fdInputLine` might better not be touched, it could be used for a final test whether any change applied for some message on screen, even with an entirely expandable chain of actions.) This way each job, indeed each recursive substitution of a single string must start with expanding `\OutputString`.

On the other hand, there is the idea of “expandable” chains of substitutions. We may, e.g., define a macro, say, `\manysubstitutions{<macro-name>}`, such that `\WriteResult{\manysubstitutions{\fdInputLine}}` writes to `<output>` the result of applying many expandable substitutions to `\fdInputLine`. Such a macro `\manysubstitutions` may read `\fdInputLine`, but it must not redefine any macros. Instead, the substitution macros it calls must read results of previous substitutions as *arguments*.

Another aspect: the order of substitutions should be easy to change. Therefore expanding of string macros should rather be controlled by the way a job is called, not right here at the *definition* of the job. For this reason, a variant of the sandbox builder expanding some macro was given up.

`setup_substr_cond` is the name space for macros that build sandboxes and initialize arguments for conditional macros.

```
90 \def\setup_substr_cond{setup_substr_cond:}
```

```
\MakeSetupSubstringCondition{<id>}[<changes>]{<pattern>}{<more-args>}
```

—same `<id>`, `<changes>`, `<pattern>` as for `\MakeSubstringConditional` (this is bad, there may be `\MakeSubstringConditional*{<more-args>}`)—creates the corresponding sandbox, by default without tilde wrap. `<more-args>` may contain `{#1}` to store the string that was tested, also `{<id>}` for calling repetitions and `{<pattern>}` for screen or log informations.

```
91 \newcommand*\MakeSetupSubstringCondition
92       {\StartFDsetup\mk_setup_substr_cond}
```

`\mk_setup_substr_cond{<pattern>}{<more-args>}` can be directly called by other programmer setup commands after `\fdParserId` and `<pattern>` have been read:

```
93 \def\mk_setup_substr_cond #1#2{%% #1 pattern string,
94   %% #2 additional arguments, e.g., '{#1}' to keep tested string
95   \endgroup
96   \expandafter \edef
97     \csname \setup_substr_cond \fdParserId \endcsname ##1{%%
98     \make_not_expanding_cs{\substr_cond \fdParserId}%%
```

By `\edef`, the name of the substring conditional is stored here as a single token. The rest of the sandbox follows.

```
99       ##1\noexpand~#1\dollar_tilde&#2}%
100   \let\dollar_tilde\sandbox_dollar}
```

If a tilde `~` has been used instead of `$`, the default is restored.

```
101 \def\sandbox_dollar{$}
102 \let\dollar_tilde\sandbox_dollar
```

The following general tool `\make_not_expanding_cs` has been used (many definitions in `latex.ltx` could have used it):

```
103 \def\make_not_expanding_cs#1{%
104   \expandafter \noexpand \csname #1\endcsname}
```

4.6 Getting rid of the tildes

`\let~\TildeGobbles` can be used to suppress dummy patterns (contained in `\split2`) in `\writeing` or with `\edef`. ... will probably become obsolete ... however, it is helpful in that you needn't care whether there is a dummy wrap left at all. (2009/04/13)

```
105 \newcommand{\TildeGobbles}{} \def\TildeGobbles#1${}
```

`\RemoveDummyPattern` is used to remove the dummy pattern *immediately*, not waiting for `\writeing` or other “total” expansion:

```
106 \newcommand{\RemoveDummyPattern}{} \def\RemoveDummyPattern#1~#2${#1}
```

`\RemoveDummyPatternArg<macro>{<arg>}` executes `\RemoveDummyPattern` in the next argument:

```
107 \newcommand*{\RemoveDummyPatternArg}[2]{%
108 \expandafter #1\expandafter {\RemoveDummyPattern #2}}
```

`\RemoveTilde` is used to remove the tilde that separated the dummy pattern from `\split1`.

```
109 % %% An alternative policy is to pass
110 % %% <target> (as an argument) to the parsing macro.
111 \newcommand{\RemoveTilde}{} \def\RemoveTilde#1~{#1}
```

`\RemoveTildeArg<macro>{<arg>}` executes `\RemoveTilde` in the next argument:

```
112 \newcommand*{\RemoveTildeArg}[2]{%
113 \expandafter #1\expandafter {\RemoveTilde #2}}
```

4.7 Calling conditionals

`\ProcessStringWith{<target-string>}{<id>}` builds the sandbox to search `<target-string>` for the `<pattern>` associated with the parser-conditional that is identified by `<id>`, the sandbox then calls the parser.

```
114 \newcommand*{\ProcessStringWith}[2]{%
115 \csname \setup_substr_cond #2\endcsname{#1}}
```

`\ProcessExpandedWith{<string-macro>}{<id>}` does the same but with a `macro` (like `\fdInputLine` or `\OutputString`) in which the string to be tested is stored.

```
116 \newcommand*{\ProcessExpandedWith}[2]{%
117 \csname \setup_substr_cond #2\expandafter \endcsname
118 \expandafter{#1}}
```

I would have preferred the reversed order of arguments which seems to be more natural, but the present is more efficient. Macros with reversed order are currently stored after `\endinput` in section 8, may be they once return.

Anyway, most desired will be `\ProcessInputWith{<id>}` just applying to `\fdInputLine`:

```

119 \newcommand*\ProcessInputWith}[1]{%
120   \csname \setup_substr_cond #1\expandafter \endcsname
121   \expandafter{\fdInputLine}}

```

(Definition almost copied for efficiency.)

```

122   %% TODO: error when undefined 2009/04/07

```

4.8 Copy jobs

A job identifier $\langle id \rangle$ may also be considered a mere *hook*, a *placeholder* for a parsing job. What function actually is called may depend on conditions that change while reading the $\langle input \rangle$ file. `\CopyFDconditionFromTo{ $\langle id1 \rangle$ }{ $\langle id2 \rangle$ }` creates or redefines a *sandbox builder* with identifier $\langle id2 \rangle$ that afterwards behaves like the sandbox builder $\langle id1 \rangle$. So you can store a certain behaviour as $\langle id1 \rangle$ in advance in order once to change the behaviour of $\langle id2 \rangle$ into that of $\langle id1 \rangle$.

```

123 \newcommand*\CopyFDconditionFromTo}[2]{%
124   \expandafter \let
125   \csname \setup_substr_cond #2\expandafter \endcsname
126   \csname \setup_substr_cond #1\endcsname}

```

(Only the *sandbox* is copied here—what about changing conditionals?)

An “almost” example is typesetting documentation from a package file where the “Legalese” header might be typeset verbatim although it is marked as “comment.” (The present example changes “hand-made” macros instead.)

This feature could have been placed more below as a “programming tool.”

5 Programming tools

5.1 Tails of conditionals

When creating complex *expandable* conditionals, this may amount to have primitive `\if ... \fi` conditionals nested quite deeply, once perhaps too deep for T_EX’s memory. To avoid this, you can apply the common `\expandafter` trick which finishes the current `\if ... \fi` before an inside macro is executed (cf. T_EXbook p. 219 on “tail recursion”).

Internally tests whether certain strings are present at certain places will be carried out by tests on emptiness or on starting with `~`. E.g., “`#1 = $\langle split1 \rangle$ empty`” indicates that either the $\langle pattern \rangle$ starts a line or the line is empty altogether (this must be decided by another test).

`\IfFDempty{ $\langle arg \rangle$ }{ $\langle when-empty \rangle$ }{ $\langle when-not-empty \rangle$ }` is used to test $\langle arg \rangle$ on emptiness (without expanding it):

```

127 \newcommand*\IfFDempty}[1]{%
128   \ifx#1$\expandafter \@firstoftwo \else
129     \expandafter \@secondoftwo \fi}

```

`\IfFDinputEmpty{ $\langle when-empty \rangle$ }{ $\langle when-not-empty \rangle$ }` is a variant of the previous to execute $\langle when-empty \rangle$ if the loop processing $\langle input \rangle$ finds an empty line—otherwise $\langle when-not-empty \rangle$.

```

130 \newcommand*\IfFDinputEmpty}{%
131   \ifx\fdInputLine\@empty \expandafter \@firstoftwo \else
132   \expandafter \@secondoftwo \fi}

```

`\IfFDdollar{<arg>}{<when-empty>}{<when-not-empty>}` is another variant, testing *<split2>* for being \$, main indicator of there is a match anywhere in *<target>* (as opposed to starting or ending match):

```

133 \newcommand*\IfFDdollar}[1]{%
134   \ifx$#1\expandafter \@firstoftwo \else
135   \expandafter \@secondoftwo \fi}

```

It is exemplified and explained in section 6. (The whole policy requires that `~` remains active in any testing macros here!)

However, you might always just type the replacement text (in one line) instead of such an `\If ...` (for efficiency ...)

If expandability is not desired, you can just chain macros that rework (so re-define) `\OutputString` or so.

2009/04/11: tending towards combining ... Keeping empty input and empty arguments apart is useful in that *one* test of emptiness per input line should suffice—it may be left open whether this should be the first of all tests ...

5.2 Line counter

A L^AT_EX counter `\fdInputLine` may be useful for screen or log messages, moreover you can use it to control processing of the *<input>* file “from outside,” not dependent on what the parsing macros find. The header of the file might be typeset verbatim, but we may be too lazy to define the “header” in terms of what is in the file. We just decide that the first ... lines are the “header,” even without counting just trying whether the output is fine. It may be necessary to change that number manually when the header changes.

You also can insert lines in *<output>* which have no counterpart in *<input>*—if you know what you are doing. With `makedoc`, there is a hook `\EveryComment` that can be used to issue commands “from outside” at a place where executing the command is safe or appropriate.

```

136 \newcounter{fdInputLine}

```

You then must insert `\CountInputLines` in the second argument of `\ProcessFileWith` (or in a macro called from there) so that the counter is stepped.

```

137 \newcommand*\CountInputLines{\global\advance\c@fdInputLine\@ne}

```

At present the counter is reset by `\ProcessFileWith`, this may change.

`\IfInputLine{<relation>{<number>}{<true>}{<false>}}`, when called from the processing loop (second argument of `\ProcessFileWith`) issues *<true>* commands if `\value{fdInputLine}<relation>{<number>}` is true, otherwise *<false>*. *<relation>* is one out of `<, =, >`.

```

138 \newcommand*\IfInputLine}[1]{%
139   \ifnum\c@fdInputLine#1\relax \expandafter \@firstoftwo
140   \else \expandafter \@secondoftwo \fi}

```


5.3 The “identity job” LEAVE

The job with identifier `\LEAVE` *leaves* an (expandable) chain of jobs (as expandable replacement in section 6) and *leaves* the processed string without changing it and without the braces enclosing it:

```
141 \expandafter \let
142   \csname \setup_substr_cond LEAVE\endcsname \@firstofone
```

I.e., `\ProcessStringWith{<string>}{LEAVE}` expands to `<string>` ... (Indeed!)

6 Setup for expandable chains of replacements

By the following means, you can create macros (`\Transform` among them) such that, e.g.,

```
\edef\OutputString{\Transform{<string>}}
```

renders `\OutputString` the result of applying a chain (sequence) of stringwise replacements to `<string>`. You can even write a transformed input `<string>` to a file without defining anything anything after `\read_upto...` In this case however, you don’t get any statistical message about what happened or not. With `\edef\OutputString` you can at least issue some `changed!` or `left!` (maybe `\message{!}` vs. `\message{.}`). There is an application in `makedoc` for “typographical upgrading” from plain text to T_EX input.

`\repl_all_chain_expandable` will be the backbone of the replacements. It is called by some parsing macro `<parser>` and receives from the latter `<split1> = #1` and `<split2> = #2`. `#3` is the result of what happened so far.

```
143 \def\repl_all_chain_expandable#1#2#3#4#5#6{%
144   %% #1, #2 splits, #3 past, #4 substitute,
145   %% #5 repeat parser, #6 pass to
146   % \ifx~#2\expandafter\@firstoftwo\else\expandafter\@secondoftwo\fi
```

The previous line would be somewhat faster, but let us exemplify `\IfFDdollar` from section 5.1 instead:

```
147 \IfFDdollar{#2}%
```

If `#2` starts with `$`—with category code 3, “math shift”!, it *is* `$`, due to not reading `$` from input with its standard category code 3 and the sandbox construction (where `$` appears with its standard category code). And this is the case *exactly* when the `<pattern>` from `<parser>` didn’t match, again due to the input category codes. Now on *no* match, the sandbox builder `#6` is called with target string `#3#1` where the last tested string is attached to previous results. The ending `~` is removed, `#6` inserts a new wrap for the new dummy pattern.

```
148   {\RemoveTildeArg #6{#3#1}}%
```

Otherwise ... the *sandbox builder* `<sandbox>` (that will be shown below) that called `<parser>` initialized `#5` to be that `<parser>` itself. (`<parser>` otherwise wouldn’t know who it is.) So `<parser>` calls itself with another sandbox `#2&`. Note that `#2` contains ‘`~<pattern>$`’ due to the initial `<sandbox>` building.

149 `{#5#2#{#3#1#4}#{4}#5#6}}`

#4 is the replacement string that $\langle sandbox \rangle$ passed to $\langle parse \rangle$. The first argument after the $\&$ is previous stuff plus the recently skipped $\langle split1 \rangle$ plus #4 replacing the string $\langle pattern \rangle$ that was matched.

Finally, #5 and #6 again “recall” $\langle parser \rangle$ and the sandbox builder to which to change in case of no other match.

```
\MakeExpandableAllReplacer{<id>}[<chng>]{<find>}{<replace>}{<id-next>}
```

creates sandbox and parser with common identifier $\langle id \rangle$ and search pattern $\langle find \rangle$. Each occurrence of $\langle find \rangle$ will be replaced by $\langle replace \rangle$. When $\langle find \rangle$ is not found, the sandbox builder for $\langle id-next \rangle$ is called. This may be another replacing macro of the same kind. To return the result without further transformations, call job LEAVE (section 5.3). Optional argument $\langle chng \rangle$ changes category codes locally for reading $\langle find \rangle$ and $\langle replace \rangle$.

```
150 \newcommand*\MakeExpandableAllReplacer{
151     {\StartFDsetup\mk_setup_xpdbl_all_repl}
152 \newcommand*\mk_setup_xpdbl_all_repl}[3]{%
153     %% #1 pattern, #2 substitute, #3 pass to
154     \endgroup
```

We take pains to call next jobs by single command strings and store them this way, not by $\backslash csname$, as $\backslash ProcessStringWith$ would do it. $\backslash edef\@tempa$ is used for this purpose, but ...

```
155     \edef\@tempa{%
156         \noexpand\mk_setup_substr_cond{#1}{%
157             }{#2}%
158         \noexpand\noexpand
```

That $\backslash edef\@tempa$ must *not expand* the controll words after they have been computed from $\backslash csname$ etc. Moreover, expansion of the parser commands must be avoided another time, when $\backslash @tempa$ is executed.

```
159         \make_not_expanding_cs{\substr_cond\fdParserId}%
160         \noexpand\noexpand
161         \make_not_expanding_cs{\setup_substr_cond #3}}}%
```

Those internal setup commands start with $\backslash endgroup$ to switch back to standard category codes. We must match them here by $\backslash begingroup$.

```
162     \begingroup \@tempa
163     \begingroup \mk_substr_cond{#1}{%
164         \repl_all_chain_expandable{##1}{##2}}}
```

The final command is the one that we explained first.

```
\PrependExpandableAllReplacer{<id>}[<cat>]{<find>}{<replace>}
```

is hoped to be a slight relief in composing replacement chains. It does something like invoking $\backslash MakeExpandableAllReplacer$ with $\langle prev-setup-id \rangle$ for the last $\langle next-id \rangle$ argument where $\langle prev-setup-id \rangle$ is the $\langle id \rangle$ of the job that was set up most recently. If you have adjacent lines

```

\MakeExpandableAllReplacer{<id-0>}{<find-0>}{<subst-0>}{LEAVE}
\PrependExpandableAllReplacer{<id-1>}{<find-1>}{<subst-1>}
\PrependExpandableAllReplacer{<id-2>}{<find-2>}{<subst-2>}

```

and call $\langle id-2 \rangle$, it will call $\langle id-1 \rangle$, and the latter will call $\langle id-0 \rangle$. So you can reorder the chain by moving `\Prepend...` lines.

```

165 \newcommand*{\PrependExpandableAllReplacer}{%
166     \let\fdParserId_before\fdParserId
167     \StartFDsetup\prep_xpdbl_all_repl}
168 \newcommand*{\prep_xpdbl_all_repl}[2]{%
169     \mk_setup_xpdbl_all_repl{#1}{#2}{\fdParserId_before}}%

```

`\MakeDocCorrectHook{<string>}` belongs to `makedoc`, but in the meantime (nicetext release 0.3) I have proposed to use it with `fifinddo` only as well (running files `fdxtxtex.tpl`, `fdxtxtex.tex`). Therefore I offer some simplification `\SetCorrectHookJob{<job-id>}` for defining `\MakeDocCorrectHook` *here*.

```

170 \newcommand*{\SetCorrectHookJob}[1]{%
171     \def\MakeDocCorrectHook##1{\ProcessStringWith{##1}{#1}}

```

`\SetCorrectHookJobLast` just uses the job that was set up most recently.

```

172 \newcommand*{\SetCorrectHookJobLast}
173     {\SetCorrectHookJob\fdParserId}

```

7 Leave package mode

We restore the underscore `_` for math subscripts. (This might better depend on something ...)

```

174 \catcode'\_ =8    %% restores underscore use for subscripts
175 \endinput

```

\TeX ignores the rest of the file when it is *input* “in the sense of `\input`”, as opposed to just reading the file line by line to a macro like `\fdInputLine`.

8 Pondered

```

176 %% TODO abbreviated commands (aliases) \MkSubstrCond...
177 %% TODO \@onlypreamble!?
178 \newcommand*{\ApplySubstringConditional}[1]{%
179     %% #1 identifier; text to be searched expected next
180     \csname setup_substr_cond:#1\endcsname}
181 \newcommand*{\ApplySubstringConditionalToExpanded}[1]{% 2009/03/31+
182     \csname setup_substr_cond:#1\expandafter \endcsname \expandafter}
183 \newcommand*{\ApplySubstringConditionalToInputString}[1]{% 2009/03/31+
184     \csname setup_substr_cond:#1\expandafter \endcsname
185     \expandafter {\fdInputLine}}
186 %% TODO or '\OutputString', even '\read' to '\OutputString'?

```

```

187 % \newcommand*\ApplySubstringConditionalToExpanded}[2]{%
188 %   %% note: without assignments, robust!
189 %   %% BUT the '\csname ... \expandafter \endcsname' method is faster
190 %   \expandafter \reversed_apply_substr_cond
191 %     \expandafter {#2}{#1}}
192 % \newcommand*\reversed_apply_substr_cond}[2]{%
193 %   \ApplySubstringConditional{#2}{#1}}
194 %   %% ODER:
195 % \newcommand*\expand_attach_arg}[2]{%% 2009/03/31
196 %   %% #1 command with previous args, TODO cf. LaTeX3
197 %   \expandafter \attach_arg \expandafter {#1}{#2}}
198 %   %% actually #1 may contain more than one token,
199 %   %% only first expanded
200 % \newcommand*\attach_arg}[2]{#2{#1}}
201 % \newcommand*\ApplySubstringConditionalToExpanded}[2]{%
202 %   \expandafter \attach_arg \expandafter
203 %     {#2}{\ApplySubstringConditional{#1}}}

```

9 VERSION HISTORY

```

204 v0.1  2009/04/03  very first version, tested on morgan.sty
205 v0.2  2009/04/05  counter fdInputLine, \ProvidesFile moved from
206                \ProcessFile to \ResultFile, \CopyFD...,
207                category section first, more sectioning,
208                suppressing empty code lines before section
209                titles; discussion, \Delimiters
210                2009/04/06  more discussion
211                2009/04/07  more discussion, factored \WriteProvides out from
212                \ResultFile, \ProcessExpandedWith corrected
213                2009/04/08  \InputString -> \fdInputline;
214                removed \ignorespaces
215                2009/04/09  \WhenInputLine[2] -> \IfInputline[3],
216                \ProcessInputWith, typos,
217                \WriteProvides message 'with'
218                2009/04/10  \make_not_expanding_cs
219                DISCOVERED 'IF SUBSTRING" ALGORITHM WRONG
220                (<str1><str2><str1> in <str1><str2>)
221 v0.3  2009/04/11  SOME THINGS GIVEN UP EARLIER WILL BE REMOVED,
222                TO BE STORED IN THE COPY AS OF 2009/04/10
223                mainly: sandbox setup (tilde/dollar)
224                REAL ADDITION: setup for expandable replacing
225                2009/04/12  played with 'chain' vs. 'sequence';
226                plain '...', 'cf.', 'etc.' for 'mdcorr.cfg'
227                2009/04/13  \RemoveTilde...
228                2009/04/15  reworked text, same mistake \in@
229 v0.31 2009/04/21f. comments on ted, stringstrings
230                2009/12/28  "onwards)" !? "safer", not "more safe"
231                2010/03/10  the loop starts
232                2010/03/17  corr. t^ete; set up -> setup for

```

233 2010/03/18 TODO EOF, ctan.org/pkg/newfile; non-ASCII
 234 2010/03/19 extended description of \MakeExpandableAll...;
 235 '' -> "
 236 2010/03/20 \ctanpkgref
 237 2010/03/22 \StartFDsetup, \Prepend...
 238 2010/03/23 URL for 'substr.sty'
 239 SENT TO CTAN
 240
 241 v0.4 2010/03/24 removed \pagebreak before "substrings";
 242 <relation> with \IfInputLine precisely
 243 2010/03/25 todo \ProcessExp... more precisely, etc.
 244 2010/03/26 ... was wrong, removed
 245 2010/03/29 \SetPatternCodes, \ResetPatternCodes,
 246 \SetCorrectHookJob, \SetCorrectHookJobLast;
 247 <relation> with \HardNVerb;
 248 don't mention \begingroup with
 249 \mk_setup_substr_cond; renamed v0.4
 250 v0.4a 2010/04/04 copyright 2010
 251
 252 TODO: cleveref 2010/03/18
 253