

low level

TEX

lowlevel

Contents

2	Conditionals	6
2.1	Preamble	7
2.2	\TeX primitives	12
2.3	$\varepsilon\text{-}\text{\TeX}$ primitives	19
2.4	\LuaTeX primitives	22
2.5	\LuaMetaTeX primitives	26
2.6	For the brave	30
2.7	Relaxing	32
3	Boxes	35
3.1	Introduction	36
3.2	Boxes	36
3.3	\TeX primitives	37
3.4	$\varepsilon\text{-}\text{\TeX}$ primitives	40
3.5	\LuaTeX primitives	40
3.6	\LuaMetaTeX primitives	41
3.7	Splitting	49
4	Expansion	55
4.1	Preamble	56
4.2	\TeX primitives	56
4.3	$\varepsilon\text{-}\text{\TeX}$ primitives	61
4.4	\LuaTeX primitives	63
4.5	\LuaMetaTeX primitives	64
4.6	Dirty tricks	74
5	Registers	77
5.1	Preamble	78
5.2	\TeX primitives	78
5.3	$\varepsilon\text{-}\text{\TeX}$ primitives	81
5.4	\LuaTeX primitives	81
5.5	\LuaMetaTeX primitives	82
6	Macros	84
6.1	Preamble	85
6.2	Definitions	85
6.3	Runaway arguments	95
6.4	Introspection	96

6.5	nesting	97
6.6	Prefixes	100
6.7	Arguments	102
6.8	Constants	103
7	Grouping	105
7.1	Introduction	106
7.2	Pascal	106
7.3	T _E X	106
7.4	MetaPost	107
7.5	Lua	108
7.6	C	108
8	Security	110
8.1	Preamble	111
8.2	Flags	111
8.3	Complications	114
8.4	Introspection	115
9	Characters	116
9.1	Introduction	117
9.2	History	117
9.3	The heritage	118
9.4	The LMTX approach	119
10	Scope	123
10.1	Introduction	124
10.2	Registers	124
10.3	Allocation	126
10.4	Files	129
11	Paragraphs	132
11.1	Introduction	133
11.2	Paragraphs	133
11.3	Properties	137
11.4	Wrapping up	139
11.5	Hanging	140
11.6	Shapes	140
11.7	Modes	158
11.8	Leaders	158
11.9	Prevdepth	165
11.10	Normalization	167

11.11 Dirty tricks	167
12 Alignments	168
12.1 Introduction	169
12.2 Between the lines	171
12.3 Pre-, inter- and post-tab skips	173
12.4 Cell widths	176
12.5 Plugins	177
12.6 Pitfalls and tricks	180
12.7 Rows	183
12.8 Remark	186
13 Marks	187
13.1 Introduction	188
13.2 The basics	189
13.3 Migration	190
13.4 Tracing	192
13.5 High level commands	193
13.6 Pitfalls	195
14 Inserts	197
14.1 Introduction	198
14.2 The page builder	198
14.3 Inserts	200
14.4 Storing	201
14.5 Synchronizing	201
14.6 Migration	201
14.7 Callbacks	202
15 Localboxes	203
15.1 Introduction	204
15.2 The basics	204
15.3 The interface	208
15.4 The helpers	213
16 Loops	215
16.1 Introduction	216
16.2 Primitives	216
16.3 Wrappers	220
17 Tokens	223
17.1 Introduction	224

17.2	What are tokens	224
17.3	Some implementation details	228
17.4	Other data management	229
17.5	Macros	230
17.6	Looking at tokens	230
18	Buffers	238
18.1	Preamble	239
18.2	Encoding	239
18.3	Performance	241
18.4	Files	241
18.5	Macros	242
18.6	Token lists	243
18.7	Buffers	243
18.8	Setups	246
18.9	xml	247
18.10	Lua	248
18.11	Protection	248

Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

2 Conditionals

low level

TEX

conditionals

Contents

2.1	Preamble	7
2.2	T _E X primitives	12
2.3	ε -T _E X primitives	19
2.4	LuaT _E X primitives	22
2.5	LuaMetaT _E X primitives	26
2.6	For the brave	30
2.7	Relaxing	32

2.1 Preamble

2.1.1 Introduction

You seldom need the low level conditionals because there are quite some so called support macros available in ConT_EXt. For instance, when you want to compare two values (or more accurate: sequences of tokens), you can do this:

```
\doifelse {foo} {bar} {
  the same
} {
  different
}
```

But if you look in the ConT_EXt code, you will see that often we use primitives that start with `\if` in low level macros. There are good reasons for this. First of all, it looks familiar when you also code in other languages. Another reason is performance but that is only true in cases where the snippet of code is expanded very often, because T_EX is already pretty fast. Using low level T_EX can also be more verbose, which is not always nice in a document source. But, the most important reason (for me) is the layout of the code. I often let the look and feel of code determine the kind of coding. This also relates to the syntax highlighting that I am using, which is consistent for T_EX, MetaPost, Lua, etc. and evolved over decades. If code looks bad, it probably is bad. Of course this doesn't mean all my code looks good; you're warned. In general we can say that I often use `\if...` when coding core macros, and `\doifelse...` macros in (document) styles and modules.

In the sections below I will discuss the low level conditions in T_EX. For the often more convenient ConT_EXt wrappers you can consult the source of the system and support modules, the wiki and/or manuals.

Some of the primitives shown here are only available in Lua \TeX , and some only in LuaMeta \TeX . We could do without them for decades but they were added to these engines because of convenience and, more important, because then made for nicer code. Of course there's also the fun aspect. This manual is not an invitation to use these very low level primitives in your document source. The ones that probably make most sense are `\ifnum`, `\ifdim` and `\ifcase`. The others are often wrapped into support macros that are more convenient.

In due time I might add more examples and explanations. Also, maybe some more tests will show up as part of the LuaMeta \TeX project.

2.1.2 Number and dimensions

Numbers and dimensions are basic data types in \TeX . When you enter one, a number is just that but a dimension gets a unit. Compare:

```
1234
1234pt
```

If you also use MetaPost, you need to be aware of the fact that in that language there are not really dimensions. The `post` part of the name implies that eventually a number becomes a PostScript unit which represents a base point (bp) in \TeX . When in MetaPost you entry `1234pt` you actually multiply 1234 by the variable `pt`. In \TeX on the other hand, a unit like `pt` is one of the keywords that gets parsed. Internally dimensions are also numbers and the unit (keyword) tells the scanner what multiplier to use. When that multiplier is one, we're talking of scaled points, with the unit `sp`.

```
\the\dimexpr 12.34pt \relax
\the\dimexpr 12.34sp \relax
\the\dimexpr 12.99sp \relax
\the\dimexpr 1234sp  \relax
\the\numexpr 1234   \relax
```

```
12.34pt
0.00018pt
0.00018pt
0.01883pt
1234
```

When we serialize a dimension it always shows the dimension in points, unless we serialize it as number.

```
\scratchdimen1234sp
\number\scratchdimen
\the\scratchdimen
```

```
1234
0.01883pt
```

When a number is scanned, the first thing that is taken care of is the sign. In many cases, when T_EX scans for something specific it will ignore spaces. It will happily accept multiple signs:

```
\number +123
\number +++123
\number + + + 123
\number +--+123
\number --123
\number ---123
```

```
123
123
123
123
123
123
-123
```

Watch how the negation accumulates. The scanner can handle decimal, hexadecimal and octal numbers:

```
\number -123
\number -"123
\number -'123
```

```
-123
-291
-83
```

A dimension is scanned like a number but this time the scanner checks for upto three parts: an either or not signed number, a period and a fraction. Here no number means zero, so the next is valid:

```
\the\dimexpr . pt \relax
\the\dimexpr 1. pt \relax
\the\dimexpr .1pt \relax
```

```
\the\dimexpr 1.1pt \relax
```

```
0.0pt
```

```
1.0pt
```

```
0.1pt
```

```
1.1pt
```

Again we can use hexadecimal and octal numbers but when these are entered, there can be no fractional part.

```
\the\dimexpr 16 pt \relax
```

```
\the\dimexpr "10 pt \relax
```

```
\the\dimexpr '20 pt \relax
```

```
16.0pt
```

```
16.0pt
```

```
16.0pt
```

The reason for discussing numbers and dimensions here is that there are cases where when $\text{T}_{\text{E}}\text{X}$ expects a number it will also accept a dimension. It is good to know that for instance a macro defined with `\chardef` or `\mathchardef` also is treated as a number. Even normal characters can be numbers, when prefixed by a ``` (backtick).

The maximum number in $\text{T}_{\text{E}}\text{X}$ is 2147483647 so we can do this:

```
\scratchcounter2147483647
```

but not this

```
\scratchcounter2147483648
```

as it will trigger an error. A dimension can be positive and negative so there we can do at most:

```
\scratchdimen 1073741823sp
```

```
\scratchdimen1073741823sp
```

```
\number\scratchdimen
```

```
\the\scratchdimen
```

```
\scratchdimen16383.99998pt
```

```
\number\scratchdimen
```

```
\the\scratchdimen
```

```
1073741823
```

16383.99998pt
 1073741823
 16383.99998pt

We can also do this:

```
\scratchdimen16383.99999pt
\number\scratchdimen
\the\scratchdimen
```

1073741823
 16383.99998pt

but the next one will fail:

```
\scratchdimen16383.9999999pt
```

Just keep in mind that T_EX scans both parts as number so the error comes from checking if those numbers combine well.

```
\ifdim 16383.99999 pt = 16383.99998 pt the same \else different \fi
\ifdim 16383.999979 pt = 16383.999980 pt the same \else different \fi
\ifdim 16383.999987 pt = 16383.999991 pt the same \else different \fi
```

Watch the difference in dividing, the / rounds, while the : truncates.

the same
 the same
 the same

You need to be aware of border cases, although in practice they never really are a problem:

```
\ifdim \dimexpr16383.99997 pt/2\relax = \dimexpr 16383.99998 pt/2\relax
  the same \else different
\fi
\ifdim \dimexpr16383.99997 pt:2\relax = \dimexpr 16383.99998 pt:2\relax
  the same \else different
\fi
```

different
 the same

```
\ifdim \dimexpr1.99997 pt/2\relax = \dimexpr 1.99998 pt/2\relax
```

the same `\else` different

`\fi`

`\ifdim \dimexpr 1.99997 pt:2\relax = \dimexpr 1.99998 pt:2\relax`

the same `\else` different

`\fi`

different

the same

`\ifdim \dimexpr 1.99999 pt/2\relax = \dimexpr 1.999995 pt/2\relax`

the same `\else` different

`\fi`

`\ifdim \dimexpr 1.99999 pt:2\relax = \dimexpr 1.999995 pt:2\relax`

the same `\else` different

`\fi`

the same

the same

This last case demonstrates that at some point the digits get dropped (still assuming that the fraction is within the maximum permitted) so these numbers then are the same. Anyway, this is not different in other programming languages and just something you need to be aware of.

2.2 T_EX primitives

2.2.1 `\if`

I seldom use this one. Internally T_EX stores (and thinks) in terms of tokens. If you see for instance `\def` or `\dimen` or `\hbox` these all become tokens. But characters like A or @ also become tokens. In this test primitive all non-characters are considered to be the same. In the next examples this is demonstrated.

`[\if AB yes\else nop\fi]`

`[\if AA yes\else nop\fi]`

`[\if CDyes\else nop\fi]`

`[\if CCyes\else nop\fi]`

`[\if\dimen\font yes\else nop\fi]`

`[\if\dimen\font yes\else nop\fi]`

Watch how spaces after the two characters are kept: [nop] [yes] [nop] [yes] [yes] [yes]. This primitive looks at the next two tokens but when doing so it expands. Just look at the following:

```
\def\AA{AA}%
\def\AB{AB}%
[\if\AA yes\else nop\fi]
[\if\AB yes\else nop\fi]
```

We get: [yes] [nop].

2.2.2 \ifcat

In T_EX characters (in the input) get interpreted according to their so called catcodes. The most common are letters (alphabetic) and other (symbols) but for instance the backslash has the property that it starts a command, the dollar signs trigger math mode, while the curly braced deal with grouping. If for instance either or not the ampersand is special (for instance as column separator in tables) depends on the macro package.

```
[\ifcat AB yes\else nop\fi]
[\ifcat AA yes\else nop\fi]
[\ifcat CDyes\else nop\fi]
[\ifcat CCyes\else nop\fi]
[\ifcat C1yes\else nop\fi]
[\ifcat\dimen\font yes\else nop\fi]
[\ifcat\dimen\font yes\else nop\fi]
```

This time we also compare a letter with a number: [yes] [yes] [yes] [yes] [nop] [yes] [yes]. In that case the category codes differ (letter vs other) but in this test comparing the letters result in a match. This is a test that is used only once in ConT_EXt and even that occasion is dubious and will go away.

You can use \noexpand to prevent expansion:

```
\def\A{A}%
\let\B B%
\def\C{D}%
\let\D D%
[\ifcat\noexpand\A Ayes\else nop\fi]
[\ifcat\noexpand\B Byes\else nop\fi]
[\ifcat\noexpand\C Cyes\else nop\fi]
[\ifcat\noexpand\C Dyes\else nop\fi]
```

```
[\ifcat\noexpand\D Dyes\else nop\fi]
```

We get: [nop] [yes] [nop] [nop] [yes], so who still thinks that T_EX is easy to understand for a novice user?

2.2.3 `\ifnum`

This condition compares its argument with another one, separated by an <, = or > character.

```
\ifnum\scratchcounter<0
  less than
\else\ifnum\scratchcounter>0
  more than
\else
  equal to
\fi zero
```

This is one of these situations where a dimension can be used instead. In that case the dimension is in scaled points.

```
\ifnum\scratchdimen<0
  less than
\else\ifnum\scratchdimen>0
  more than
\else
  equal to
\fi zero
```

Of course this equal treatment of a dimension and number is only true when the dimension is a register or box property.

2.2.4 `\ifdim`

This condition compares one dimension with another one, separated by an <, = or > sign.

```
\ifdim\scratchdimen<0pt
  less than
\else\ifdim\scratchdimen>0pt
  more than
\else
```

equal to
`\fi` zero

While when comparing numbers a dimension is a valid quantity but here you cannot mix them: something with a unit is expected.

2.2.5 `\ifodd`

This one can come in handy, although in ConT_EXt it is only used in checking for an odd of even page number.

```
\scratchdimen 3sp
\scratchcounter4

\ifodd\scratchdimen very \else not so \fi odd
\ifodd\scratchcounter very \else not so \fi odd
```

As with the previously discussed `\ifnum` you can use a dimension variable too, which is then interpreted as representing scaled points. Here we get:

```
very odd
not so odd
```

2.2.6 `\ifvmode`

This is a rather trivial check. It takes no arguments and just is true when we're in vertical mode. Here is an example:

```
\hbox{\ifvmode\else\par\fi\ifvmode v\else h\fi mode}
```

We're always in horizontal mode and issuing a `\par` inside a horizontal box doesn't change that, so we get: `hmode`.

2.2.7 `\ifhmode`

As with `\ifvmode` this one has no argument and just tells if we're in vertical mode.

```
\vbox {
  \noindent \ifhmode h\else v\fi mode
  \par
  \ifhmode h\else \noindent v\fi mode
}
```

You can use it for instance to trigger injection of code, or prevent that some content (or command) is done more than once:

```
hmode
vmode
```

2.2.8 `\ifmmode`

Math is something very \TeX so naturally you can check if you're in math mode. here is an example of using this test:

```
\def\enforcemath#1{\ifmmode#1\else$ #1 $\fi}
```

Of course in reality macros that do such things are more advanced than this one.

2.2.9 `\ifinner`

```
\def\ShowMode
  {\ifhmode    \ifinner inner \fi hmode
  \else\ifvmode \ifinner inner \fi vmode
  \else\ifmmode \ifinner inner \fi mmode
  \else       \ifinner inner \fi unset
  \fi\fi\fi}
```

```
\ShowMode \ShowMode
```

```
\vbox{\ShowMode}
```

```
\hbox{\ShowMode}
```

```
$_ShowMode$
```

```
$$\ShowMode$$
```

The first line has two tests, where the first one changes the mode to horizontal simply because a text has been typeset. Watch how display math is not inner.

```
vmode hmode
inner vmode
inner hmode
innermmode
innermmode
```

By the way, moving the `\ifinner` test outside the branches (to the top of the macro) won't work because once the word `inner` is typeset we're no longer in vertical mode, if we were at all.

2.2.10 `\ifvoid`

A box is one of the basic concepts in \TeX . In order to understand this primitive we present four cases:

```
\setbox0\hbox{}           \ifvoid0 void \else content \fi
\setbox0\hbox{123}       \ifvoid0 void \else content \fi
\setbox0\hbox{} \box0    \ifvoid0 void \else content \fi
\setbox0\hbox to 10pt{} \ifvoid0 void \else content \fi
```

In the first case, we have a box which is empty but it's not void. It helps to know that internally an `hbox` is actually an object with a pointer to a linked list of nodes. So, the first two can be seen as:

```
hlist -> [nothing]
hlist -> 1 -> 2 -> 3 -> [nothing]
```

but in any case there is a `hlist`. The third case puts something in a `hlist` but then flushes it. Now we have not even the `hlist` any more; the box register has become void. The last case is a variant on the first. It is an empty box with a given width. The outcome of the four lines (with a box flushed in between) is:

```
content
content
```

```
void
content
```

So, when you want to test if a box is really empty, you need to test also its dimensions, which can be up to three tests, depending on your needs.

```
\setbox0\emptybox           \ifvoid0 void\else content\fi
\setbox0\emptybox          \wd0=10pt \ifvoid0 void\else content\fi
\setbox0\hbox to 10pt {}    \ifvoid0 void\else content\fi
\setbox0\hbox               {} \wd0=10pt \ifvoid0 void\else content\fi
```

Setting a dimension of a void box (empty) doesn't make it less void:

```
void
```

```
void
content
content
```

2.2.11 `\ifhbox`

This test takes a box number and gives true when it is an hbox.

2.2.12 `\ifvbox`

This test takes a box number and gives true when it is an vbox. Both a `\vbox` and `\vtop` are vboxes, the difference is in the height and depth and the baseline. In a `\vbox` the last line determines the baseline

```
vbox or vtop
vtop or vbox
```

And in a `\vtop` the first line takes control:

```
vbox or vtop
vtop or vbox
```

but, once wrapped, both internally are just vlists.

2.2.13 `\ifx`

This test is actually used a lot in `ConTEXt`: it compares two token(list)s:

```

\ifx a b Y\else N\fi
\ifx ab Y\else N\fi
\def\A {a}\def\B{b}\ifx \A\B Y\else N\fi
\def\A{aa}\def\B{a}\ifx \A\B Y\else N\fi
\def\A {a}\def\B{a}\ifx \A\B Y\else N\fi
```

Here the result is: “NNNNY”. It does not expand the content, if you want that you need to use an `\edef` to create two (temporary) macros that get compared, like in:

```
\edef\TempA{...}\edef\TempB{...}\ifx\TempA\TempB ...\else ... \fi
```

2.2.14 `\ifeof`

This test checks if a the pointer in a given input channel has reached its end. It is also true when the file is not present. The argument is a number which relates to the `\openin` primitive that is used to open files for reading.

2.2.15 `\iftrue`

It does what it says: always true.

2.2.16 `\iffalse`

It does what it says: always false.

2.2.17 `\ifcase`

The general layout of an `\ifcase` tests is as follows:

```
\ifcase<number>
  when zero
\or
  when one
\or
  when two
\or
  ...
\else
  when something else
\fi
```

As in other places a number is a sequence of signs followed by one of more digits

2.3 ϵ -**T_EX** primitives

2.3.1 `\ifdefined`

This primitive was introduced for checking the existence of a macro (or primitive) and with good reason. Say that you want to know if `\MyMacro` is defined? One way to do that is:

```
\ifx\MyMacro\undefined
  {\bf undefined indeed}
\fi
```

This results in: **undefined indeed**, but is this macro really undefined? When T_EX scans your source and sees a the escape character (the forward slash) it will grab the next characters and construct a control sequence from it. Then it finds out that there is nothing with that name and it will create a hash entry for a macro with that name but with no meaning. Because `\undefined` is also not defined, these two macros have the same meaning and therefore the `\ifx` is true. Imagine that you do this many times, with different macro names, then your hash can fill up. Also, when a user defined `\undefined` you're suddenly get a different outcome.

In order to catch the last problem there is the option to test directly:

```
\ifdefined\MyOtherMacro \else
  {\bf also undefined}
\fi
```

This (or course) results in: **also undefined**, but the macro is still sort of defined (with no meaning). The next section shows how to get around this.

2.3.2 `\ifcurname`

A macro is often defined using a ready made name, as in:

```
\def\OhYes{yes}
```

The name is made from characters with catcode letter which means that you cannot use for instance digits or underscores unless you also give these characters that catcode, which is not that handy in a document. You can however use `\curname` to define a control sequence with any character in the name, like:

```
\expandafter\def\curname Oh Yes : 1\endcurname{yes}
```

Later on you can get this one with `\curname`:

```
\curname Oh Yes : 1\endcurname
```

However, if you say:

```
\curname Oh Yes : 2\endcurname
```

you won't get some result, nor a message about an undefined control sequence, but the name triggers a define anyway, this time not with no meaning (undefined) but as equivalent to `\relax`, which is why

```
\expandafter\ifx\cename Oh Yes : 2\endcename\relax
  {\bf relaxed indeed}
\fi
```

is the way to test its existence. As with the test in the previous section, this can deplete the hash when you do lots of such tests. The way out of this is:

```
\ifcename Oh Yes : 2\endcename \else
  {\bf unknown indeed}
\fi
```

This time there is no hash entry created and therefore there is not even an undefined control sequence.

In LuaTeX there is an option to return false in case of a messy expansion during this test, and in LuaMetaTeX that is default. This means that tests can be made quite robust as it is pretty safe to assume that names that make sense are constructed from regular characters and not boxes, font switches, etc.

2.3.3 `\iffontchar`

This test was also part of the ε -TeX extensions and it can be used to see if a font has a character.

```
\iffontchar\font `A
  {\em This font has an A!}
\fi
```

And, as expected, the outcome is: “*This font has an A!*”. The test takes two arguments, the first being a font identifier and the second a character number, so the next checks are all valid:

```
\iffontchar\font `A yes\else nop\fi\par
\iffontchar\nullfont `A yes\else nop\fi\par
\iffontchar\textfont0 `A yes\else nop\fi\par
```

In the perspective of LuaMetaTeX I considered also supporting `\fontid` but it got a bit messy due to the fact that this primitive expands in a different way so this extension was rejected.

2.3.4 `\unless`

You can negate the results of a test by using the `\unless` prefix, so for instance you can replace:

```
\ifdim\scratchdimen=10pt
  \dosomething
\else\ifdim\scratchdimen<10pt
  \dosomething
\fi\fi
```

by:

```
\unless\ifdim\scratchdimen>10pt
  \dosomething
\fi
```

2.4 Lua_TE_X primitives

2.4.1 `\ifincsname`

As it had no real practical usage it might get dropped in LuaMeta_TE_X, so it will not be discussed here.

2.4.2 `\ifprimitive`

As it had no real practical usage due to limitations, this one is not available in LuaMeta_TE_X so it will not be discussed here.

2.4.3 `\ifabsnum`

This test is inherited from pdf_TE_X and behaves like `\ifnum` but first turns a negative number into a positive one.

2.4.4 `\ifabsdim`

This test is inherited from pdf_TE_X and behaves like `\ifdim` but first turns a negative dimension into a positive one.

2.4.5 `\ifcondition`

This is not really a test but in order to understand that you need to know how \TeX internally deals with tests.

```
\ifdim\scratchdimen>10pt
  \ifdim\scratchdimen<20pt
    result a
  \else
    result b
  \fi
\else
  result c
\fi
```

When we end up in the branch of “result a” we need to skip two `\else` branches after we're done. The `\if..` commands increment a level while the `\fi` decrements a level. The `\else` needs to be skipped here. In other cases the true branch needs to be skipped till we end up at the right `\else`. When doing this skipping, \TeX is not interested in what it encounters beyond these tokens and this skipping (therefore) goes real fast but it does see nested conditions and doesn't interpret grouping related tokens.

A side effect of this is that the next is not working as expected:

```
\def\ifmorethan{\ifdim\scratchdimen>}
\def\iflessthan{\ifdim\scratchdimen<}

\ifmorethan10pt
  \iflessthan20pt
    result a
  \else
    result b
  \fi
\else
  result c
\fi
```

The `\iflessthan` macro is not seen as an `\if...` so the nesting gets messed up. The solution is to fool the scanner in thinking that it is. Say we have:

```
\scratchdimen=25pt

\def\ifmorethan{\ifdim\scratchdimen>}
```

```
\def\iflessthan{\ifdim\scratchdimen<}
```

and:

```
\ifcondition\ifmorethan10pt
  \ifcondition\iflessthan20pt
    result a
  \else
    result b
  \fi
\else
  result c
\fi
```

When we expand this snippet we get: “result b” and no error concerning a failure in locating the right `\fi`'s. So, when scanning the `\ifcondition` is seen as a valid `\if...` but when the condition is really expanded it gets ignored and the `\ifmorethan` has better come up with a match or not.

In this perspective it is also worth mentioning that nesting problems can be avoided this way:

```
\def\WhenTrue {something \iftrue ...}
\def\WhenFalse{something \iffalse ...}

\ifnum\scratchcounter>123
  \let\next\WhenTrue
\else
  \let\next\WhenFalse
\fi
\next
```

This trick is mentioned in The \TeX book and can also be found in the plain \TeX format. A variant is this:

```
\ifnum\scratchcounter>123
  \expandafter\WhenTrue
\else
  \expandafter\WhenFalse
\fi
```

but using `\expandafter` can be quite intimidating especially when there are multiple in a row. It can also be confusing. Take this: an `\ifcondition` expects the code that follows to produce a test. So:

```

\def\ifwhatever#1%
  {\ifdim#1>10pt
   \expandafter\iftrue
  \else
   \expandafter\iffalse
  \fi}

\ifcondition\ifwhatever{10pt}
  result a
\else
  result b
\fi

```

This will not work! The reason is in the already mentioned fact that when we end up in the greater than 10pt case, the scanner will happily push the `\iftrue` after the `\fi`, which is okay, but when skipping over the `\else` it sees a nested condition without matching `\fi`, which makes it fail. I will spare you a solution with lots of nasty tricks, so here is the clean solution using `\ifcondition`:

```

\def\truecondition {\iftrue}
\def>falsecondition{\iffalse}

\def\ifwhatever#1%
  {\ifdim#1>10pt
   \expandafter\truecondition
  \else
   \expandafter>falsecondition
  \fi}

\ifcondition\ifwhatever{10pt}
  result a
\else
  result b
\fi

```

It will be no surprise that the two macros at the top are predefined in `ConTeXt`. It might be more of a surprise that at the time of this writing the usage in `ConTeXt` of this `\ifcondition` primitive is rather minimal. But that might change.

As a further teaser I'll show another simple one,

```

\def\HowOdd#1{\unless\ifnum\numexpr ((#1):2)*2\relax=\numexpr#1\relax}

```

```

\ifcondition\HowOdd{1}very \else not so \fi odd
\ifcondition\HowOdd{2}very \else not so \fi odd
\ifcondition\HowOdd{3}very \else not so \fi odd

```

This renders:

```

very odd
not so odd
very odd

```

The code demonstrates several tricks. First of all we use `\numexpr` which permits more complex arguments, like:

```

\ifcondition\HowOdd{4+1}very \else not so \fi odd
\ifcondition\HowOdd{2\scratchcounter+9}very \else not so \fi odd

```

Another trick is that we use an integer division (the `:`) which is an operator supported by `LuaMetaTeX`.

2.5 LuaMetaTeX primitives

2.5.1 `\ifcmpnum`

This one is part of a set of three tests that all are a variant of a `\ifcase` test. A simple example of the first test is this:

```

\ifcmpnum 123 345 less \or equal \else more \fi

```

The test scans for two numbers, which of course can be registers or expressions, and sets the case value to 0, 1 or 2, which means that you then use the normal `\or` and `\else` primitives for follow up on the test.

2.5.2 `\ifchknum`

This test scans a number and when it's okay sets the case value to 1, and otherwise to 2. So you can do the next:

```

\ifchknum 123\or good \else bad \fi
\ifchknum bad\or good \else bad \fi

```

An error message is suppressed and the first `\or` can be seen as a sort of recovery token, although in fact we just use the fast scanner mode that comes with the `\ifcase`: because the result is 1 or 2, we never see invalid tokens.

2.5.3 `\ifnumval`

A sort of combination of the previous two is `\ifnumval` which checks a number but also if it's less, equal or more than zero:

```
\ifnumval 123\or less \or equal \or more \else error \fi
\ifnumval bad\or less \or equal \or more \else error \fi
```

You can decide to ignore the bad number or do something that makes more sense. Often the to be checked value will be the content of a macro or an argument like `#1`.

2.5.4 `\ifcmpdim`

This test is like `\ifcmpnum` but for dimensions.

2.5.5 `\ifchkdim`

This test is like `\ifchknum` but for dimensions. The last checked value is available as `\lastchknum`.

2.5.6 `\ifdimval`

This test is like `\ifnumval` but for dimensions. The last checked value is available as `\lastchkdim`

2.5.7 `\iftok`

Although this test is still experimental it can be used. What happens is that two to be compared ‘things’ get scanned for. For each we first gobble spaces and `\relax` tokens. Then we can have several cases:

1. When we see a left brace, a list of tokens is scanned up to the matching right brace.
2. When a reference to a token register is seen, that register is taken as value.
3. When a reference to an internal token register is seen, that register is taken as value.
4. When a macro is seen, its definition becomes the to be compared value.
5. When a number is seen, the value of the corresponding register is taken

An example of the first case is:

```
\iftok {abc} {def}%
...

```

```
\else
...
\fi
```

The second case goes like this:

```
\iftok\scratchtoksone\scratchtokstwo
...
\else
...
\fi
```

Case one and four mixed:

```
\iftok{123}\TempX
...
\else
...
\fi
```

The last case is more a catch: it will issue an error when no number is given. Eventually that might become a bit more clever (depending on our needs.)

2.5.8 `\ifcstok`

There is a subtle difference between this one and `iftok`: spaces and `\relax` tokens are skipped but nothing gets expanded. So, when we arrive at the to be compared ‘things’ we look at what is there, as-is.

2.5.9 `\iffrozen`

This is an experimental test. Commands can be defined with the `\frozen` prefix and this test can be used to check if that has been the case.

2.5.10 `\ifprotected`

Commands can be defined with the `\protected` prefix (or in `ConTeXt`, for historic reasons, with `\unexpanded`) and this test can be used to check if that has been the case.

2.5.11 `\ifusercmd`

This is an experimental test. It can be used to see if the command is defined at the user level or is a build in one. This one might evolve.

2.5.12 `\ifarguments`

This conditional can be used to check how many arguments were matched. It only makes sense when used with macros defined with the `\tolerant` prefix and/or when the sentinel `\ignorearguments` after the arguments is used. More details can be found in the lowlevel macros manual.

2.5.13 `\orelse`

This is not really a test primitive but it does act that way. Say that we have this:

```
\ifdim\scratchdimen>10pt
  case 1
\else\ifdim\scratchdimen<20pt
  case 2
\else\ifcount\scratchcounter>10
  case 3
\else\ifcount\scratchcounter<20
  case 4
\fi\fi\fi\fi
```

A bit nicer looks this:

```
\ifdim\scratchdimen>10pt
  case 1
\orelse\ifdim\scratchdimen<20pt
  case 2
\orelse\ifcount\scratchcounter>10
  case 3
\orelse\ifcount\scratchcounter<20
  case 4
\fi
```

We stay at the same level. Sometimes a more flat test tree had advantages but if you think that it gives better performance then you will be disappointed. The fact that we stay at the same level is compensated by a bit more parsing, so unless you have millions such cases (or expansions) it might make a bit of a difference. As mentioned, I'm a bit sensitive for how code looks so that was the main motivation for introducing it. There is a companion `\orunless` continuation primitive.

A rather neat trick is the definition of `\quitcondition`:

```
\def\quitcondition{\orelse\iffalse}
```

This permits:

```
\ifdim\scratchdimen>10pt
  case 1a
  \quitcondition
  case 4b
\fi
```

where, of course, the quitting normally is the result of some intermediate extra test. But let me play safe here: beware of side effects.

2.6 For the brave

2.6.1 Full expansion

If you don't understand the following code, don't worry. There is seldom much reason to go this complex but obscure T_EX code attracts some users so ...

When you have a macro that has for instance assignments, and when you expand that macro inside an `\edef`, these assignments are not actually expanded but tokenized. In LuaMetaT_EX there is a way to apply these assignments without side effects and that feature can be used to write a fully expandable user test. For instance:

```
\def>truecondition {\iftrue}
\def>falsecondition{\iffalse}

\def\fontwithidhaschar#1#2%
  {\beginlocalcontrol
    \scratchcounter\numexpr\fontid\font\relax
    \setfontid\numexpr#1\relax
    \endlocalcontrol
    \iffontchar\font\numexpr#2\relax
      \beginlocalcontrol
        \setfontid\scratchcounter
        \endlocalcontrol
        \expandafter>truecondition
    \else
      \expandafter>falsecondition
    \fi}
```

The `\iffontchar` test doesn't handle numeric font id, simply because at the time it was added to ε -TeX, there was no access to these id's. Now we can do:

```
\edef\foo{\fontwithidhaschar{1} {75}yes\else nop\fi} \meaning\foo
\edef\foo{\fontwithidhaschar{1}{999}yes\else nop\fi} \meaning\foo

[\ifcondition\fontwithidhaschar{1} {75}yes\else nop\fi]
[\ifcondition\fontwithidhaschar{1}{999}yes\else nop\fi]
```

These result in:

```
macro:yes
macro:nop
```

```
[yes]
[nop]
```

If you remove the `\immediateassignment` in the definition above then the typeset results are still the same but the meanings of `\foo` look different: they contain the assignments and the test for the character is actually done when constructing the content of the `\edef`, but for the current font. So, basically that test is now useless.

2.6.2 User defined if's

There is a `\newif` macro that defines three other macros:

```
\newif\ifOnMyOwnTerms
```

After this, not only `\ifOnMyOwnTerms` is defined, but also:

```
\OnMyOwnTermstrue
\OnMyOwnTermsfalse
```

These two actually are macros that redefine `\ifOnMyOwnTerms` to be either equivalent to `\iftrue` and `\iffalse`. The (often derived from plain TeX) definition of `\newif` is a bit of a challenge as it has to deal with removing the `if` in order to create the two extra macros and also make sure that it doesn't get mixed up in a catcode jungle.

In ConTeXt we have a variant:

```
\newconditional\MyConditional
```

that can be used with:

```
\settrue\MyConditional
\setfalse\MyConditional
```

and tested like:

```
\ifconditional\MyConditional
  ...
\else
  ...
\fi
```

This one is cheaper on the hash and doesn't need the two extra macros per test. The price is the use of `\ifconditional`, which is *not* to confused with `\ifcondition` (it has bitten me already a few times).

2.7 Relaxing

When \TeX scans for a number or dimension it has to check tokens one by one. On the case of a number, the scanning stops when there is no digit, in the case of a dimension the unit determine the end of scanning. In the case of a number, when a token is not a digit that token gets pushed back. When digits are scanned a trailing space or `\relax` is pushed back. Instead of a number of dimension made from digits, periods and units, the scanner also accepts registers, both the direct accessors like `\count` and `\dimen` and those represented by one token. Take these definitions:

```
\newdimen\MyDimenA \MyDimenA=1pt \dimen0=\MyDimenA
\newdimen\MyDimenB \MyDimenB=2pt \dimen2=\MyDimenB
```

I will use these to illustrate the side effects of scanning. Watch the spaces in the result.

First I show what effect we want to avoid. When second argument contains a number (digits) the zero will become part of it so we actually check `\dimen00` here.

```
\def\whatever#1#2%
  {\ifdim#1=#20\else1\fi}
```

```
\whatever{1pt}{2pt}           [macro:1]
\whatever{1pt}{1pt}           [macro:0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:]
\whatever\MyDimenA\MyDimenB   [macro:1]
\whatever\MyDimenA\MyDimenB   [macro:1]
```

The solution is to add a space but watch how that one can end up in the result:

```
\def\whatever#1#2%
  {\ifdim#1=#2 0\else1\fi}

\whatever{1pt}{2pt}      [macro:1]
\whatever{1pt}{1pt}      [macro:0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:0]
\whatever\MyDimenA\MyDimenB [macro:1]
\whatever\MyDimenA\MyDimenB [macro:1]
```

A variant is using `\relax` and this time we get this token retained in the output.

```
\def\whatever#1#2%
  {\ifdim#1=#2\relax0\else1\fi}

\whatever{1pt}{2pt}      [macro:1]
\whatever{1pt}{1pt}      [macro:\relax 0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:\relax 0]
\whatever\MyDimenA\MyDimenB [macro:1]
\whatever\MyDimenA\MyDimenB [macro:1]
```

A solution that doesn't have side effects of forcing the end of a number (using a space or `\relax` is one where we use expressions. The added overhead of scanning expressions is taken for granted because the effect is what we like:

```
\def\whatever#1#2%
  {\ifdim\dimexpr#1\relax=\dimexpr#2\relax0\else1\fi}

\whatever{1pt}{2pt}      [macro:1]
\whatever{1pt}{1pt}      [macro:0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:0]
\whatever\MyDimenA\MyDimenB [macro:1]
\whatever\MyDimenA\MyDimenB [macro:1]
```

Just for completeness we show a more obscure trick: this one hides assignments to temporary variables. Although performance is okay, it is the least efficient one so far.

```
\def\whatever#1#2%
  {\beginlocalcontrol
```

```

\MyDimenA#1\relax
\MyDimenB#2\relax
\endlocalcontrol
\ifdim\MyDimenA=\MyDimenB0\else1\fi}

```

```

\whatever{1pt}{2pt}          [macro:1]
\whatever{1pt}{1pt}         [macro:0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:0]
\whatever\MyDimenA\MyDimenB [macro:1]
\whatever\MyDimenA\MyDimenB [macro:1]

```

It is kind of a game to come up with alternatives but for sure those involve dirty tricks and more tokens (and runtime). The next can be considered a dirty trick too: we use a special variant of `\relax`. When a number is scanned it acts as `relax`, but otherwise it just is ignored and disappears.

```

\def\whatever#1#2%
  {\ifdim#1=#2\norelax0\else1\fi}

\whatever{1pt}{2pt}          [macro:1]
\whatever{1pt}{1pt}         [macro:0]
\whatever{\dimen 0}{\dimen 2} [macro:1]
\whatever{\dimen 0}{\dimen 0} [macro:0]
\whatever\MyDimenA\MyDimenB [macro:1]
\whatever\MyDimenA\MyDimenB [macro:1]

```

2.7 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

3 Boxes

low level

TEX

boxes

Contents

3.1	Introduction	36
3.2	Boxes	36
3.3	\TeX primitives	37
3.4	$\varepsilon\text{-}\TeX$ primitives	40
3.5	Lua \TeX primitives	40
3.6	LuaMeta \TeX primitives	41
3.7	Splitting	49

3.1 Introduction

An average Con \TeX t user will not use the low level box primitives but a basic understanding of how \TeX works doesn't hurt. In fact, occasionally using a box command might bring a solution not easily achieved otherwise, simply because a more high level interface can also be in the way.

The best reference is of course The \TeX book so if you're really interested in the details you should get a copy of that book. Below I will not go into details about all kind of glues, kerns and penalties, just boxes it is.

This explanation will be extended when I feel the need (or users have questions that can be answered here).

3.2 Boxes

This paragraph of text is made from lines that contain words that themselves contain symbolic representations of characters. Each line is wrapped in a so called horizontal box and eventually those lines themselves get wrapped in what we call a vertical box.

When we expose some details of a paragraph it looks like this:

H	This is a rather narrow	LH:0.000 LS:0.000	RS:13.888	RH:0.000
H	paragraph blown up a	LH:0.000 LS:0.000	RS:19.008	RH:0.000
H	bit. Here we use a flush	LH:0.000 LS:0.000	RS:7.557	RH:0.000
H	left, aka ragged right,	LH:0.000 LS:0.000	RS:19.706	RH:0.000
H	approach.	LH:0.000 LS:0.000	RS:0.000	RH:0.000

The left only shows the boxes, the variant at the right shows (font) kerns and glue too. Because we flush left, there is rather strong right skip glue at the right boundary of the box. If font kerns show up depends on the font, not all fonts have them (or have only a few). The glyphs themselves are also kind of boxed, as their dimensions determine the area that they occupy:

This is a rather ...

But, internally they are not really boxed, as they already are a single quantity. The same is true for rules: they are just blobs with dimensions. A box on the other hand wraps a linked list of so called nodes: glyphs, kerns, glue, penalties, rules, boxes, etc. It is a container with properties like width, height, depth and shift.

3.3 T_EX primitives

The box model is reflected in T_EX's user interface but not by that many commands, most noticeably `\hbox`, `\vbox` and `\vtop`. Here is an example of the first one:

```
\hbox width 10cm{text}
\hbox width 10cm height 1cm depth 5mm{text}
text \raise5mm\hbox{text} text
```

The `\raise` and `\lower` commands behave the same but in opposite directions. One could as well have been defined in terms of the other.

```
text \raise 5mm \hbox to 2cm {text}
text \lower -5mm \hbox to 2cm {text}
text \raise -5mm \hbox to 2cm {text}
text \lower 5mm \hbox to 2cm {text}
```

```

      .text.      .text.
text.      text.      text.      text.
              .text.      .text.
```

A box can be moved to the left or right but, believe it or not, in ConT_EXt we never use that feature, probably because the consequences for the width are such that we can as well use kerns. Here are some examples:

```
text \vbox{\moveleft 5mm \hbox {left}}text !
text \vbox{\moveright 5mm \hbox{right}}text !
```

lefttext ! text righttext !

```
text \vbox{\moveleft 25mm \hbox {left}}text !
text \vbox{\moveright 25mm \hbox{right}}text !
```

left text text ! text righttext !

Code like this will produce a complaint about an underfull box but we can easily get around that:

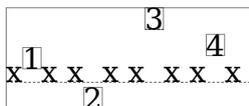
```
text \raise 5mm \hbox to 2cm {\hss text}
text \lower -5mm \hbox to 2cm {text\hss}
text \raise -5mm \hbox to 2cm {\hss text}
text \lower 5mm \hbox to 2cm {text\hss}
```

The `\hss` primitive injects a glue that when needed will fill up the available space. So, here we force the text to the right or left.

Instead of `\raise` you can also provide the shift (up or down) with a keyword:

```
\ruledhbox\bggroup
x\raise 5pt\ruledhbox                      {1}x
x\raise -10pt\ruledhbox                    {2}x
x\raise -5pt\ruledhbox shift -20pt{3}x
x\ruledhbox                                shift -10pt{4}x
```

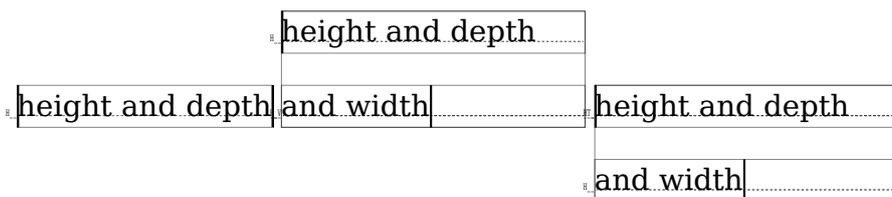
`\egroup`



We have three kind of boxes: `\hbox`, `\vbox` and `\vtop`. Actually we have a fourth type `\dbox` but that is a variant on `\vbox` to which we come back later.

```
\hbox{\strut height and depth\strut}
\vbox{\hsize 4cm \strut height and depth\par and width\strut}
\vtop{\hsize 4cm \strut height and depth\par and width\strut}
```

A `\vbox` aligns at the bottom and a `\vtop` at the top. I have added some so called struts to enforce a consistent height and depth. A strut is an invisible quantity (consider it a black box) that enforces consistent line dimensions: height and depth.



You can store a box in a register but you need to be careful not to use a predefined one. If you need a lot of boxes you can reserve some for your own:

```
\newbox\MySpecialBox
```

but normally you can do with one of the scratch registers, like 0, 2, 4, 6 or 8, for local boxes, and 1, 3, 5, 7 and 9 for global ones. Registers are used like:

```
\setbox0\hbox{here}
\global\setbox1\hbox{there}
```

In ConT_EXt you can also use

```
\setbox\scratchbox \hbox{here}
\setbox\scratchboxone\hbox{here}
\setbox\scratchboxtwo\hbox{here}
```

and some more. In fact, there are quite some predefined scratch registers (boxes, dimensions, counters, etc). Feel free to investigate further.

When a box is stored, you can consult its dimensions with `\wd`, `\ht` and `\dp`. You can of course store them for later use.

```
\scratchwidth \wd\scratchbox
\scratchheight\ht\scratchbox
\scratchdepth \dp\scratchbox
\scratchtotal \dimexpr\ht\scratchbox+\dp\scratchbox\relax
\scratchtotal \htdp\scratchbox
```

The last line is ConT_EXt specific. You can also set the dimensions

```
\wd\scratchbox 10cm
\ht\scratchbox 10mm
\dp\scratchbox 5mm
```

So you can cheat! A box is placed with `\copy`, which keeps the original intact or `\box` which just inserts the box and then wipes the register. In practice you seldom need a copy, which is more expensive in runtime anyway. Here we use copy because it serves the examples.

```
\copy\scratchbox
\box \scratchbox
```

3.4 ε -T_EX primitives

The ε -T_EX extensions don't add something relevant for boxes, apart from that you can use the expressions mechanism to mess around with their dimensions. There is a mechanism for typesetting r2l within a paragraph but that has limited capabilities and doesn't change much as it's mostly a way to trick the backend into outputting a stretch of text in the other direction. This feature is not available in LuaT_EX because it has an alternative direction mechanism.

3.5 LuaT_EX primitives

The concept of boxes is the same in LuaT_EX as in its predecessors but there are some aspects to keep in mind. When a box is typeset this happens in LuaT_EX:

1. A list of nodes is constructed. In LuaT_EX this is a double linked list (so that it can easily be manipulated in Lua) but T_EX itself only uses the forward links.
2. That list is hyphenated, that is: so called discretionary nodes are injected. This depends on the language properties of the glyph (character) nodes.
3. Then ligatures are constructed, if the font has such combinations. When this built-in mechanism is used, in ConT_EXt we speak of base mode.
4. After that inter-character kerns are applied, if the font provides them. Again this is a base mode action.
5. Finally the box gets packaged:
 - In the case of a horizontal box, the list is packaged in a hlist node, basically one liner, and its dimensions are calculated and set.
 - In the case of a vertical box, the paragraph is broken into one or more lines, without hyphenation, with optimal hyphenation or in the worst case with so called emergency stretch applied, and the result becomes a vlist node with its dimensions set.

In traditional T_EX the first four steps are interwoven but in LuaT_EX we need them split because the step 5 can be overloaded by a callback. In that case steps 3 and 4 (and

maybe 2) are probably also overloaded, especially when you bring handling of fonts under Lua control.

New in LuaTeX are three packers: `\hpack`, `\vpack` and `\tpack`, which are companions to `\hbox`, `\vbox` and `\vtop` but without the callbacks applied. Using them is a bit tricky as you never know if a callback should be applied, which, because users can often add their own Lua code, is not something predictable.

Another box related extension is `direction`. There are four possible directions but because in LuaMetaTeX there are only two. Because this model has been upgraded, it will be discussed in the next section. A ConTeXt user is supposed to use the official ConTeXt interfaces in order to be downward compatible.

3.6 LuaMetaTeX primitives

There are two possible directions: left to right (the default) and right to left for Hebrew and Arabic. Here is an example that shows how it'd done with low level directives:

```
\hbox direction 0 {from left to right}
\hbox direction 1 {from right to left}
```

from left to right
tfel ot thgir morf

A low level direction switch is done with:

```
\hbox direction 0
  {from left to right \textdirection 1 from right to left}
\hbox direction 1
  {from right to left \textdirection 1 from left to right}
```

from left to right tfel ot thgir morf
thgir ot tfel morf tfel ot thgir morf

but actually this is kind of *not done* in ConTeXt, because there you are supposed to use the proper direction switches:

```
\naturalhbox {from left to right}
\reversehbox {from right to left}
\naturalhbox {from left to right \righttoleft from right to left}
\reversehbox {from right to left \lefttoright from left to right}
```

from left to right
 tfel ot thgir morf
 from left to right tfel ot thgir morf
 from left to right tfel ot thgir morf

Often more is needed to properly support right to left typesetting so using the ConT_EXt commands is more robust.

In LuaMetaT_EX the box model has been extended a bit, this as a consequence of dropping the vertical directional typesetting, which never worked well. In previous sections we discussed the properties width, height and depth and the shift resulting from a `\raise`, `\lower`, `\moveleft` and `\moveright`. Actually, the shift is also used in for instance positioning math elements.

The way shifting influences dimensions can be somewhat puzzling. Internally, when T_EX packages content in a box there are two cases:

- When a horizontal box is made, and height - shift is larger than the maximum height so far, that delta is taken. When depth + shift is larger than the current depth, then that depth is adapted. So, a shift up influences the height and a shift down influences the depth.
- In the case of vertical packaging, when width + shift is larger than the maximum box (line) width so far, that maximum gets bumped. So, a shift to the right can contribute, but a shift to the left cannot result in a negative width. This is also why vertical typesetting, where height and depth are swapped with width, goes wrong: we somehow need to map two properties onto one and conceptually T_EX is really set up for horizontal typesetting. (And it's why I decided to just remove it from the engine.)

This is one of these cases where T_EX behaves as expected but it also means that there is some limitation to what can be manipulated. Setting the shift using one of the four commands has a direct consequence when a box gets packaged which happens immediately because the box is an argument to the foursome.

There is in traditional T_EX, probably for good reason, no way to set the shift of a box, if only because the effect would normally be none. But in LuaT_EX we can cheat, and therefore, for educational purposed ConT_EXt has implements some cheats.

We use this sample box:

```
\setbox\scratchbox\hbox\bgroup
  \middlegray\vrule width 20mm depth -.5mm height 10mm
```

```

\hskip-20mm
\darkgray \vrule width 20mm height -.5mm depth 5mm
\egroup

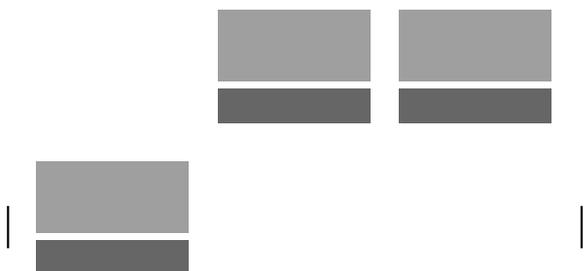
```

When we mess with the shift using the ConT_EXt `\shiftbox` helper, we see no immediate effect. We only get the shift applied when we use another helper, `\hpackbox`.

```

\hbox\bgroup
  \showstruts \strut
  \quad \copy\scratchbox
  \quad \shiftbox\scratchbox -20mm \copy\scratchbox
  \quad \hpackbox\scratchbox \box \scratchbox
  \quad \strut
\egroup

```

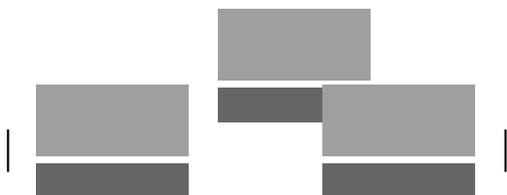


When instead we use `\vpackbox` we get a different result. This time we move left.

```

\hbox\bgroup
  \showstruts \strut
  \quad \copy\scratchbox
  \quad \shiftbox\scratchbox -10mm \copy\scratchbox
  \quad \vpackbox\scratchbox \copy\scratchbox
  \quad \strut
\egroup

```



The shift is set via Lua and the repackaging is also done in Lua, using the low level `hpack` and `vpack` helpers and these just happen to look at the shift when doing their job. At the T_EX end this never happens.

This long exploration of shifting serves a purpose: it demonstrates that there is not that much direct control over boxes apart from their three dimensions. However this was never a real problem as one can just wrap a box in another one and use kerns to move the embedded box around. But nevertheless I decided to see if the engine can be a bit more helpful, if only because all that extra wrapping gives some overhead and complications when we want to manipulate boxes. And of course it is also a nice playground.

We start with changing the direction. Changing this property doesn't require repackaging because directions are not really dealt with in the frontend. When a box is converted to (for instance pdf) the reversion happens.

```
\setbox\scratchbox\hbox{whatever}
\the\boxdirection\scratchbox: \copy\scratchbox \crlf
\boxdirection\scratchbox 1
\the\boxdirection\scratchbox: \copy\scratchbox
```

0: whatever

1: revetahw

Another property that can be queried and set is an attribute. In order to get a private attribute we define one.

```
\newattribute\MyAt
\setbox\scratchbox\hbox attr \MyAt 123 {whatever}
[\the\boxattribute\scratchbox\MyAt]
\boxattribute\scratchbox\MyAt 456
[\the\boxattribute\scratchbox\MyAt]
[\ifnum\boxattribute\scratchbox\MyAt>400 okay\fi]
```

[123] [456] [okay]

The sum of the height and depth is available too. Because for practical reasons setting that property is also needed then, the choice was made to distribute the value equally over height and depth.

```
\setbox\scratchbox\hbox {height and depth}
[\the\ht\scratchbox]
[\the\dp\scratchbox]
[\the\boxtotal\scratchbox]
\boxtotal\scratchbox=20pt
[\the\ht\scratchbox]
[\the\dp\scratchbox]
```

```
[\the\boxtotal\scratchbox]
```

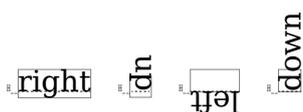
```
[8.35742pt] [2.44385pt] [10.80127pt] [10.0pt] [10.0pt] [20.0pt]
```

We've now arrived to a set of properties that relate to each other. They are a bit complex and given the number of possibilities one might need to revert to some trial and error: orientations and offsets. As with the dimensions, directions and attributes, they are passed as box specification. We start with the orientation.

```
\hbox \bgroup \showboxes
```

```
    \hbox orientation 0 {right}
\quad \hbox orientation 1 {up}
\quad \hbox orientation 2 {left}
\quad \hbox orientation 3 {down}
```

```
\egroup
```

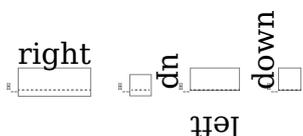


When the orientation is set, you can also set an offset. Where shifting around a box can have consequences for the dimensions, an offset is virtual. It gets effective in the backend, when the contents is converted to some output format.

```
\hbox \bgroup \showboxes
```

```
    \hbox orientation 0 yoffset 10pt {right}
\quad \hbox orientation 1 xoffset 10pt {up}
\quad \hbox orientation 2 yoffset -10pt {left}
\quad \hbox orientation 3 xoffset -10pt {down}
```

```
\egroup
```



The reason that offsets are related to orientation is that we need to know in what direction the offsets have to be applied and this binding forces the user to think about it. You can also set the offsets using commands.

```
\setbox\scratchbox\hbox{whatever}%
```

```
1 \copy\scratchbox
2 \boxorientation\scratchbox 2 \copy\scratchbox
3 \boxxoffset \scratchbox -15pt \copy\scratchbox
```

```
4 \boxyoffset \scratchbox -15pt \copy\scratchbox
5
```

```
1 whatever2 3 4 5
```

```
\setbox\scratchboxone\hbox{whatever}%
```

```
\setbox\scratchboxtwo\hbox{whatever}%
```

```
1 \boxxoffset \scratchboxone -15pt \copy\scratchboxone
```

```
2 \boxyoffset \scratchboxone -15pt \copy\scratchboxone
```

```
3 \boxxoffset \scratchboxone -15pt \copy\scratchboxone
```

```
4 \boxyoffset \scratchboxone -15pt \copy\scratchboxone
```

```
5 \boxxmove \scratchboxtwo -15pt \copy\scratchboxtwo
```

```
6 \boxymove \scratchboxtwo -15pt \copy\scratchboxtwo
```

```
7 \boxxmove \scratchboxtwo -15pt \copy\scratchboxtwo
```

```
8 \boxymove \scratchboxtwo -15pt \copy\scratchboxtwo
```

```
whatever 2 3 4 whatever6 7 8
  whatever whatever whatever           whatever
                                             whatever
```

The move commands are provided as convenience and contrary to the offsets they do adapt the dimensions. Internally, with the box, we register the orientation and the offsets and when you apply these commands multiple times the current values get overwritten. But ... because an orientation can be more complex you might not get the effects you expect when the options we discuss next are used. The reason is that we store the original dimensions too and these come into play when these other options are used: anchoring. So, normally you will apply an orientation and offsets once only.

The orientation specifier is actually a three byte number that best can be seen hexadecimal (although we stay within the decimal domain). There are three components: x-anchoring, y-anchoring and orientation:

```
0x<X><Y><O>
```

or in TeX speak:

```
"<X><Y><O>
```

The landscape and seascape variants both sit on top of the baseline while the flipped variant has its depth swapped with the height. Although this would be enough a bit more control is possible.

The vertical options of the horizontal variants anchor on the baseline, lower corner, upper corner or center.

```
\ruledhbox orientation "002 {\TEX} and
\ruledhbox orientation "012 {\TEX} and
\ruledhbox orientation "022 {\TEX} and
\ruledhbox orientation "032 {\TEX}
```

The horizontal options of the horizontal variants anchor in the center, left, right, halfway left and halfway right.

```
\ruledhbox orientation "002 {\TEX} and
\ruledhbox orientation "102 {\TEX} and
\ruledhbox orientation "202 {\TEX} and
\ruledhbox orientation "302 {\TEX} and
\ruledhbox orientation "402 {\TEX}
```

The orientation has consequences for the dimensions so they are dealt with in the expected way in constructing lines, paragraphs and pages, but the anchoring is virtual, like the offsets. There are two extra variants for orientation zero: on top of baseline or below, with dimensions taken into account.

```
\ruledhbox orientation "000 {\TEX} and
\ruledhbox orientation "004 {\TEX} and
\ruledhbox orientation "005 {\TEX}
```

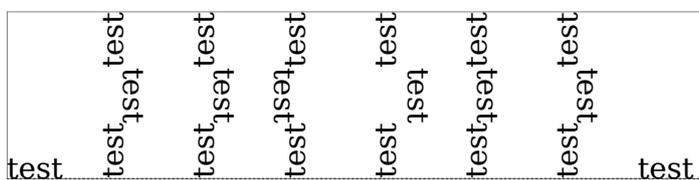
The anchoring can look somewhat confusing but you need to keep in mind that it is normally only used in very controlled circumstances and not in running text. Wrapped in macros users don't see the details. We're talking boxes here, so for instance:

```
test\quad
\hbox orientation 3 \bgroup
  \strut test\hbox orientation "002 \bgroup\strut test\egroup test%
\egroup \quad
\hbox orientation 3 \bgroup
  \strut test\hbox orientation "002 \bgroup\strut test\egroup test%
\egroup \quad
\hbox orientation 3 \bgroup
```

```

\strut test\hbox orientation "012 \bgroup\strut test\egroup test%
\egroup \quad
\hbox orientation 3 \bgroup
\strut test\hbox orientation "022 \bgroup\strut test\egroup test%
\egroup \quad
\hbox orientation 3 \bgroup
\strut test\hbox orientation "032 \bgroup\strut test\egroup test%
\egroup \quad
\hbox orientation 3 \bgroup
\strut test\hbox orientation "042 \bgroup\strut test\egroup test%
\egroup
\quad test

```



Where a `\vtop` has the baseline at the top, a `\vbox` has it at the bottom. In LuaMeta- \TeX we also have a `\dbox`, which is a `\vbox` with that behaves like a `\vtop` when it's appended to a vertical list: the height of the first box or rule determines the (base)line correction that gets applied. The following example demonstrates this:

<pre> XXXXXXXXXXXXXXXXXX We thrive in information--thick worlds be- cause of our marvelous and everyday ca- pacity to select, edit, single out, struc- ture, highlight, group, pair, merge, har- monize, synthesize, focus, organize, con- dense, reduce, boil down, choose, cate- gorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discrim- inate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, av- erage, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsis, winnow the wheat from the chaff and separate the sheep from the goats. XXXXXXXXXXXXXXXXXX </pre> <p style="text-align: center;"><code>\vbox</code></p>	<pre> XXXXXXXXXXXXXXXXXX We thrive in information--thick worlds be- cause of our marvelous and everyday ca- pacity to select, edit, single out, struc- ture, highlight, group, pair, merge, har- monize, synthesize, focus, organize, con- dense, reduce, boil down, choose, cate- gorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discrim- inate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, av- erage, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsis, winnow the wheat from the chaff and separate the sheep from the goats. XXXXXXXXXXXXXXXXXX </pre> <p style="text-align: center;"><code>\vtop</code></p>	<pre> XXXXXXXXXXXXXXXXXX We thrive in information--thick worlds be- cause of our marvelous and everyday ca- pacity to select, edit, single out, struc- ture, highlight, group, pair, merge, har- monize, synthesize, focus, organize, con- dense, reduce, boil down, choose, cate- gorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discrim- inate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, av- erage, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsis, winnow the wheat from the chaff and separate the sheep from the goats. XXXXXXXXXXXXXXXXXX </pre> <p style="text-align: center;"><code>\dbox</code></p>
--	--	--

The `d` stands for ‘dual’ because we (sort of) have two baselines. The regular height and depth are those of a `\vbox`.

3.7 Splitting

When you feed T_EX a paragraph of text it will accumulate the content in a list of nodes. When the paragraph is finished by `\par` or an empty line it will be fed into the par builder that will try to break the lines as good as possible. Normally that paragraph will be added to the page and at some point there can be breaks between lines in order not to overflow the page. When you collect the paragraph in a box you can use `\vsplit` to emulate this.

```
\setbox\scratchbox\vbox{\samplefile{tufte}}
```

```
\startlinecorrection
```

```
\ruledhbox{\vsplit\scratchbox to 2\lineheight}
```

```
\stoplinecorrection
```

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe-

The split off box is given the specified height, but in LuaMetaT_EX you can also get the natural dimensions:

```
\setbox\scratchbox\vbox{\samplefile{tufte}}
```

```
\startlinecorrection
```

```
\ruledhbox{\vsplit\scratchbox upto 2\lineheight}
```

```
\stoplinecorrection
```

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe-

We can force a resulting box type by using `\vsplit`, `\tsplit` and `\dsplit` (here we use the visualized variants):

```
\setbox\scratchbox\vbox{\samplefile{tufte}}
```

```
\startlinecorrection
```

```
\ruledtsplit \scratchbox upto 2\lineheight
```

```
\stoplinecorrection
```

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe-

```
\setbox\scratchbox\vbox{\samplefile{tufte}}
```

```

\startlinecorrection
\ruledvsplit \scratchbox upto 2\lineheight
\stoplinecorrection

```

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe-

```
\setbox\scratchbox\vbox{\samplefile{tufte}}
```

```

\startlinecorrection
\ruleddsplit \scratchbox upto 2\lineheight
\stoplinecorrection

```

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe-

The engine provides vertical splitters but ConT_EXt itself also has a horizontal one.¹

```

\starttexdefinition Test #1#2#3
  \par
  \dontleavehmode
  \strut
  \llap{{\infofont #2}\quad}
  \blackrule[width=#2,color=darkblue]
  \par
  \setbox\scratchbox\hbox{\samplefile{#1}}
  \hsplit\scratchbox
    to          #2
    depth       \strutdp
    height      \strutht
    shrinkcriterium #3 % badness
  \par
\stoptexdefinition

\dostepwiserecurse {100} {120} {2} {
  \Test{tufte}{#1mm}{1000}
  \Test{tufte}{#1mm}{-100}
}

```

100mm

We thrive in information-thick worlds because of

¹ At some point I might turn that one into a native engine primitive.

100mm We thrive in information-thick worlds because of

102mm We thrive in information-thick worlds because of our

102mm We thrive in information-thick worlds because of our

104mm We thrive in information-thick worlds because of our

104mm We thrive in information-thick worlds because of our

106mm We thrive in information-thick worlds because of our

106mm We thrive in information-thick worlds because of our

108mm We thrive in information-thick worlds because of our

108mm We thrive in information-thick worlds because of our

110mm We thrive in information-thick worlds because of our

110mm We thrive in information-thick worlds because of our

112mm We thrive in information-thick worlds because of our mar-

112mm We thrive in information-thick worlds because of our mar-

114mm We thrive in information-thick worlds because of our mar-

114mm We thrive in information-thick worlds because of our mar-

116mm We thrive in information-thick worlds because of our mar-

116mm We thrive in information-thick worlds because of our mar-

118mm We thrive in information-thick worlds because of our mar-

118mm We thrive in information-thick worlds because of our mar-

120mm We thrive in information-thick worlds because of our mar-

120mm

We thrive in information–thick worlds because of our mar-

A split off box gets packed at its natural size and a badness as well as overshoot amount is calculated. When the overshoot is positive and the the badness is larger than the stretch criterium, the box gets repacked to the natural size. The same happens when the overshoot is negative and the badness exceeds the shrink criterium. When the overshoot is zero (basically we have a fit) but the badness still exceeds the stretch or shrink we also repack. Indeed this is a bit fuzzy, but so is badness.

```
\starttexdefinition Test #1#2#3
  \par
  \dontleavehmode
  \strut
  \llap{{\infofont #2}\quad}
  \blackrule[width=#2,color=darkblue]
  \par
  \setbox\scratchbox\hbox{\samplefile{#1}}
  \doloop {
    \ifvoid\scratchbox
      \exitloop
    \else
      \hsplit\scratchbox
        to      #2
        depth  \strutdp
        height \strutht
        #3
      \par
      \allowbreak
    \fi
  }
\stoptexdefinition

\Test{tufte}{100mm}{shrinkcriterium 1000}
\Test{tufte}{100mm}{shrinkcriterium 0}
\Test{tufte}{100mm}{}
```

100mm

We thrive in information–thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense,

reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate the sheep from the goats.

100mm

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate the sheep from the goats.

100mm

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate

BBB
the sheep from the goats.

Watch how the last line get stretched when we set the criterium to zero. I'm sure that users will find reasons to abuse this effect.

3.7 Colofon

Author Hans Hagen
ConT_EXt 2023.04.27 17:04
LuaMetaT_EX 2.1008
Support www.pragma-ade.com
 contextgarden.net

111111

XXX

222222

AAA

4 Expansion

low level

TEX

expansion

Contents

4.1	Preamble	56
4.2	T _E X primitives	56
4.3	ε-T _E X primitives	61
4.4	LuaT _E X primitives	63
4.5	LuaMetaT _E X primitives	64
4.6	Dirty tricks	74

4.1 Preamble

This short manual demonstrates a couple of properties of the macro language. It is not an in-depth philosophical expose about macro languages, tokens, expansion and such that some T_EXies like. I prefer to stick to the practical aspects. Occasionally it will be technical but you can just skip those paragraphs (or later return to them) when you can't follow the explanation. It's often not that relevant. I won't talk in terms of mouth, stomach and gut the way the T_EXbook does and although there is no way to avoid the word 'token' I will do my best to not complicate matters by too much token speak. Examples show best what we mean.

4.2 T_EX primitives

The T_EX language provides quite some commands and those built in are called primitives. User defined commands are called macros. A macro is a shortcut to a list of primitives and/or macro calls. All can be mixed with characters that are to be typeset somehow.

```
\def\MyMacro{b}
```

```
a\MyMacro c
```

When T_EX reads this input the a gets turned into a glyph node with a reference to the current font set and the character a. Then the parser sees a macro call, and it will enter another input level where it expands this macro. In this case it sees just an b and it will give this the same treatment as the a. The macro ends, the input level decrements and the c gets its treatment.

Before we move on to more examples and differences between engines, it is good to stress that `\MyMacro` is not a primitive command: we made our command here. The b actually can be seen as a sort of primitive because in this macro it gets stored as so

called token with a primitive property. That primitive property can later on be used to determine what to do. More explicit examples of primitives are `\hbox`, `\advance` and `\relax`. It will be clear that ConT_EXt extends the repertoire of primitive commands with a lot of macro commands. When we typeset a source using module `m-scite` the primitives come out dark blue.

The amount of primitives differs per engine. It all starts with T_EX as written by Don Knuth. Later ε -T_EX added some more primitives and these became official extensions adopted by other variants of T_EX. The pdfT_EX engine added quite some and as follow up on that LuaT_EX added more but didn't add all of pdfT_EX. A few new primitives came from Omega (Aleph). The LuaMetaT_EX engine drops a set of primitives that comes with LuaT_EX and adds plenty new ones. The nature of this engine (no backend and less frontend) makes that we need to implement some primitives as macros. But the basic set is what good old T_EX comes with.

Internally these so called primitives are grouped in categories that relate to their nature. They can be directly expanded (a way of saying that they get immediately interpreted) or delayed (maybe stored for later usage). They can involve definitions, calculations, setting properties and values or they can result in some typesetting. This is what makes T_EX confusing to new users: it is a macro programming language, an interpreter but at the same time an executor of typesetting instructions.

A group of primitives is internally identified as a command (they have a `cmd` code) and the sub commands are flagged by their `chr` code. This sounds confusing but just thing of the fact that most of what we input are characters and therefore they make up most sub commands. For instance the 'letter `cmd`' is used for characters that are seen as letters that can be used in the name of user commands, can be typeset, are valid for hyphenation etc. The letter related `cmd` can have many `chr` codes (all of Unicode). I'd like to remark that the grouping is to a large extend functional, so sometimes primitives that you expect to be similar in nature are in different groups. This has to do with the fact that T_EX needs to be able to determine efficiently if a primitive is operating (or forbidden) in horizontal, vertical and/or math mode.

There are more than 150 internal `cmd` groups. if we forget about the mentioned character related ones, some, have only a few sub commands (`chr`) and others many more (just consider all the OpenType math spacing related parameters). A handful of these commands deal with what we call macros: user defined combinations of primitives and other macros, consider them little programs. The `\MyMacro` example above is an example. There are differences between engines. In standard T_EX there are `\outer` and `\long` commands, and most engines have these. However, in LuaMetaT_EX the later to be discussed `\protected` macros have their own specific 'call `cmd`'. Users don't need to bother about this.

So, when from now on we talk about primitives, we mean the built in, hard coded commands, and when we talk about macros we mean user commands. Although internally there are less cmd categories than primitives, from the perspective of the user they are all unique. Users won't consult the source anyway but when they do they are warned. Also, when in LuaMetaTeX you use the low level interfacing to TeX you have to figure out these subtle aspects because there this grouping does matter.

Before we continue I want to make clear that expansion (as discussed in this document) can refer to a macro being expanded (read: its meaning gets injected into the input, so the engine kind of sidetracks from what it was doing) but also to direct consequences of running into a primitive. However, users only need to consider expansion in the perspective of macros. If a user has `\advance` in the input it immediately gets done. But when it's part of a macro definition it only is executed when the macro expands. A good check in (traditional) TeX is to compare what happens in `\def` and `\edef` which is why we will use these two in the upcoming examples. You put something in a macro and then check what `\meaning` or `\show` reports.

Now back to user defined macros. A macro can contain references to macros so in practice the input can go several levels up and some applications push back a lot so this is why your TeX input stack can be configured to be huge.

```
\def\MyMacroA{ and }
\def\MyMacroB{1\MyMacroA 2}

a\MyMacroA b
```

When `\MyMacroB` is defined, its body gets three so called tokens: the character token `1` with property 'other', a token that is a reference to the macro `\MyMacroB`, and a character token `2`, also with property 'other' The meaning of `\MyMacroA` is five tokens: a reference to a space token, then three character tokens with property 'letter', and finally a space token.

```
\def \MyMacroA{ and }
\edef\MyMacroB{1\MyMacroA 2}

a\MyMacroA b
```

In the second definition an `\edef` is used, where the `e` indicates expansion. This time the meaning gets expanded immediately. So we get effectively the same as in:

```
\def\MyMacroB{1 and 2}
```

Characters are easy: they just expand to themselves or trigger adding a glyph node, but not all primitives expand to their meaning or effect.

```
\def\MyMacroA{\scratchcounter = 1 }
\def\MyMacroB{\advance\scratchcounter by 1}
\def\MyMacroC{\the\scratchcounter}
```

```
\MyMacroA a
\MyMacroB b
\MyMacroB c
\MyMacroB d
\MyMacroC
```

a b c d 4

```
macro:\scratchcounter = 1
macro:\advance \scratchcounter by 1
macro:\the \scratchcounter
```

Let's assume that `\scratchcounter` is zero to start with and use `\edef`'s:

```
\edef\MyMacroA{\scratchcounter = 1 }
\edef\MyMacroB{\advance\scratchcounter by 1}
\edef\MyMacroC{\the\scratchcounter}
```

```
\MyMacroA a
\MyMacroB b
\MyMacroB c
\MyMacroB d
\MyMacroC
```

a b c d 0

```
macro:\scratchcounter = 1
macro:\advance \scratchcounter by 1
macro:0
```

So, this time the third macro has its meaning frozen, but we can prevent this by applying a `\noexpand` when we do this:

```
\edef\MyMacroA{\scratchcounter = 1 }
\edef\MyMacroB{\advance\scratchcounter by 1}
\edef\MyMacroC{\noexpand\the\scratchcounter}
```

```
\MyMacroA a
\MyMacroB b
```

```
\MyMacroB c
\MyMacroB d
\MyMacroC
```

```
a b c d 4
```

```
macro:\scratchcounter = 1
macro:\advance \scratchcounter by 1
macro:\the \scratchcounter
```

Of course this is a rather useless example but it serves its purpose: you'd better be aware what gets expanded immediately in an `\edef`. In most cases you only need to worry about `\the` and embedded macros (and then of course their meanings).

You can also store tokens in a so-called token register. Here we use a predefined scratch register:

```
\def\MyMacroA{ and }
\def\MyMacroB{1\MyMacroA 2}
\scratchtoks {\MyMacroA}
```

The content of `\scratchtoks` is: “`\MyMacroA`”, so no expansion has happened here.

```
\def\MyMacroA{ and }
\def\MyMacroB{1\MyMacroA 2}
\scratchtoks \expandafter {\MyMacroA}
```

Now the content of `\scratchtoks` is: “`and`”, so this time expansion has happened.

```
\def\MyMacroA{ and }
\def\MyMacroB{1\MyMacroA 2}
\scratchtoks \expandafter {\MyMacroB}
```

Indeed the macro gets expanded but only one level: “`1\MyMacroA 2`”. Compare this with:

```
\def\MyMacroA{ and }
\edef\MyMacroB{1\MyMacroA 2}
\scratchtoks \expandafter {\MyMacroB}
```

The trick is to expand in two steps with an intermediate `\edef`: “`1 and 2`”. Later we will see that other engines provide some more expansion tricks. The only way to get some grip on expansion is to just play with it.

The `\expandafter` primitive expands the token (which can be a macro) standing after the next next one and then injects its meaning into the stream. So:

```
\expandafter \MyMacroA \MyMacroB
```

works okay. In a normal document you will never need this kind of hackery: it only happens in a bit more complex macros. Here is an example:

```
\scratchcounter 1
\bgroup
\advance\scratchcounter 1
\egroup
\the\scratchcounter
```

```
\scratchcounter 1
\bgroup
\advance\scratchcounter 1
\expandafter
\egroup
\the\scratchcounter
```

The first one gives 1, while the second gives 2.

4.3 ε - $\text{T}_{\text{E}}\text{X}$ primitives

In this engine a couple of extensions were added and later on pdf $\text{T}_{\text{E}}\text{X}$ added some more. We only discuss a few that relate to expansion. There is however a pitfall here. Before ε - $\text{T}_{\text{E}}\text{X}$ showed up, Con $\text{T}_{\text{E}}\text{X}$ t already had a few mechanism that also related to expansion and it used some names for macros that clash with those in ε - $\text{T}_{\text{E}}\text{X}$. This is why we will use the `\normal` prefix here to indicate the primitive.²

```
\def\MyMacroA{a}
\def\MyMacroB{b}
\normalprotected\def\MyMacroC{c}
\edef\MyMacroABC{\MyMacroA\MyMacroB\MyMacroC}
```

These macros have the following meanings:

```
macro:a
macro:b
```

² In the meantime we no longer have a low level `\protected` macro so one can use the primitive

```
protected macro:c
macro:ab\MyMacroC
```

In ConTEXt you will use the `\unexpanded` prefix instead, because that one did something similar in older versions of ConTEXt. As we were early adopters of ε -TEX, this later became a synonym to the ε -TEX primitive.

```
\def\MyMacroA{a}
\def\MyMacroB{b}
\normalprotected\def\MyMacroC{c}
\normalexpanded{\scratchtoks{\MyMacroA\MyMacroB\MyMacroC}}
```

Here the wrapper around the token register assignment will expand the three macros, unless they are protected, so its content becomes “ab\MyMacroC”. This saves either a lot of more complex `\expandafter` usage or the need to use an intermediate `\edef`. In ConTEXt the `\expanded` macro does something simpler but it doesn't expand the first token as this is meant as a wrapper around a command, like:

```
\expanded{\chapter{...}} % a ConTeXt command
```

where we do want to expand the title but not the `\chapter` command (not that this would happen actually because `\chapter` is a protected command.)

The counterpart of `\normalexpanded` is `\normalunexpanded`, as in:

```
\def\MyMacroA{a}
\def\MyMacroB{b}
\normalprotected\def\MyMacroC{c}
\normalexpanded {\scratchtoks
  {\MyMacroA\normalunexpanded {\MyMacroB}\MyMacroC}}
```

The register now holds “a\MyMacroB\MyMacroC”: three tokens, one character token and two macro references.

Tokens can represent characters, primitives, macros or be special entities like starting math mode, beginning a group, assigning a dimension to a register, etc. Although you can never really get back to the original input, you can come pretty close, with:

```
\detokenize{this can $ be anything \bgroup}
```

This (when typeset monospaced) is: this can \$ be anything \bgroup. The detokenizer is like `\string` applied to each token in its argument. Compare this to:

```
\normalexpanded {
```

```
\normaldetokenize{10pt}
}
```

We get four tokens: 10pt.

```
\normalexpanded {
  \string 1\string 0\string p\string t
}
```

So that was the same operation: 10pt, but in both cases there is a subtle thing going on: characters have a catcode which distinguishes them. The parser needs to know what makes up a command name and normally that's only letters. The next snippet shows these catcodes:

```
\normalexpanded {
  \noexpand\the\catcode` \string 1 \noexpand\enspace
  \noexpand\the\catcode` \string 0 \noexpand\enspace
  \noexpand\the\catcode` \string p \noexpand\enspace
  \noexpand\the\catcode` \string t \noexpand
}
```

The result is “12 12 11 11”: two characters are marked as ‘letter’ and two fall in the ‘other’ category.

4.4 Lua_T_EX primitives

This engine adds a little to the expansion repertoire. First of all it offers a way to extend token lists registers:

```
\def\MyMacroA{a}
\def\MyMacroB{b}
\normalprotected\def\MyMacroC{b}
\scratchtoks{\MyMacroA\MyMacroB}
```

The result is: “\MyMacroA\MyMacroB”.

```
\toksapp\scratchtoks{\MyMacroA\MyMacroB}
```

We're now at: “\MyMacroA\MyMacroB\MyMacroA\MyMacroB\MyMacroA\MyMacroB”.

```
\etoksapp\scratchtoks{\MyMacroA\space\MyMacroB\space\MyMacroC}
```

The register has this content: “\MyMacroA\MyMacroB\MyMacroA\MyMacroB a b \MyMacroC a b \MyMacroC”, so the additional context got expanded in the process, except of course the protected macro \MyMacroC.

There is a bunch of these combinators: \toksapp and \tokspre for local appending and prepending, with global companions: \gtoksapp and \gtokspre, as well as expanding variant: \etoksapp, \etokspre, \xtoksapp and \xtokspre.

These are not beforehand more efficient than using intermediate expanded macros or token lists, simply because in the process $\text{T}_{\text{E}}\text{X}$ has to create token lists too, but sometimes they're just more convenient to use. In $\text{ConT}_{\text{E}}\text{Xt}$ we actually do benefit from these.

4.5 LuaMeta $\text{T}_{\text{E}}\text{X}$ primitives

We already saw that macro's can be defined protected which means that

```
\def\TestA{A}
\protected \def\TestB{B}
\edef\TestC{\TestA\TestB}
```

gives this:

```
\TestC : A\TestB
```

One way to get \TestB expanded is to prefix it with \expand:

```
\def\TestA{A}
\protected \def\TestB{B}
\edef\TestC{\TestA\TestB}
\edef\TestD{\TestA\expand\TestB}
```

We now get:

```
\TestC : A\TestB
\TestD : AB
```

There are however cases where one wishes this to happen automatically, but that will also make protected macros expand which will create havoc, like switching fonts.

```
\def\TestA{A}
\protected \def\TestB{B}
\semiprotected \def\TestC{C}
\edef\TestD{\TestA\TestB\TestC}
```

```

\edef\TestE{\normalexpanded{\TestA\TestB\TestC}}
\edef\TestF{\semiexpanded {\TestA\TestB\TestC}}

```

This time `\TestC` loses its protection:

```

\TestA : A
\TestB : B
\TestC : C
\TestD : A\TestB \TestC
\TestE : A\TestB \TestC
\TestF : A\TestB C

```

Actually adding `\fullyexpanded` would be trivial but it makes not much sense to add the overhead (at least not now). This feature is experimental anyway so it might go away when I see no real advantage from it.

When you store something in a macro or token register you always need to keep an eye on category codes. A dollar in the input is normally treated as math shift, a hash indicates a macro parameter or preamble entry. Characters like ‘A’ are letters but ‘[’ and ‘]’ are tagged as ‘other’. The \TeX scanner acts according to these codes. If you ever find yourself in a situation that changing catcodes is no option or cumbersome, you can do this:

```

\edef\TestOA{\expandtoken\othercatcode `A}
\edef\TestLA{\expandtoken\lettercatcode `A}

```

In both cases the meaning is A but in the first case it's not a letter but a character flagged as ‘other’.

A whole new category of commands has to do with so called local control. When \TeX scans and interprets the input, a process takes place that is called tokenizing: (sequences of) characters get a symbolic representation and travel through the system as tokens. Often they immediately get interpreted and are then discarded. But when for instance you define a macro they end up as a linked list of tokens in the macro body. We already saw that expansion plays a role. In most cases, unless \TeX is collecting tokens, the main action is dealt with in the so-called main loop. Something gets picked up from the input but can also be pushed back, for instance because of some lookahead that didn't result in an action. Quite some time is spent in pushing and popping from the so-called input stack.

When we are in Lua, we can pipe back into the engine but all is collected till we're back in \TeX where the collected result is pushed into the input. Because \TeX is a mix of programming and action there basically is only that main loop. There is no real way

to start a sub run in Lua and do all kind of things independent of the current one. This makes sense when you consider the mix: it would get too confusing.

However, in LuaTeX and even better in LuaMetaTeX, we can enter a sort of local state and this is called ‘local control’. When we are in local control a new main loop is entered and the current state is temporarily forgotten: we can for instance expand where one level up expansion was not done. It sounds complicated and indeed it is complicated so examples have to clarify it.

```
1 \setbox0\hbox to 10pt{2} \count0=3 \the\count0 \multiply\count0 by 4
```

This snippet of code is not that useful but illustrates what we're dealing with:

- The 1 gets typeset. So, characters like that are seen as text.
- The \setbox primitive triggers picking up a register number, then goes on scanning for a box specification and that itself will typeset a sequence of whatever until the group ends.
- The count primitive triggers scanning for a register number (or reference) and then scans for a number; the equal sign is optional.
- The the primitive injects some value into the current input stream and it does so by entering a new input level.
- The multiply primitive picks up a register specification and multiplies that by the next scanned number. The by is optional.

We now look at this snippet again but with an expansion context:

```
\def \TestA{1 \setbox0\hbox{2} \count0=3 \the\count0}
```

```
\edef\TestB{1 \setbox0\hbox{2} \count0=3 \the\count0}
```

These two macros have a slightly different body. Make sure you see the difference before reading on.

control sequence: TestA

542601	12	49	other char	1	U+00031	
226982	10	32	spacer			
540394	122	0	set box			setbox
543515	12	48	other char	0	U+00030	
543731	30	14	make box			hbox
543549	1	123	left brace			

32720	12	50	other char	2	U+00032	
543827	2	125	right brace			
543498	10	32	spacer			
539317	115	1	register			count
542847	12	48	other char	0	U+00030	
542835	12	61	other char	=	U+0003D	
543043	12	51	other char	3	U+00033	
560622	10	32	spacer			
543247	135	0	the			the
538520	115	1	register			count
543526	12	48	other char	0	U+00030	

control sequence: TestB

542622	12	49	other char	1	U+00031	
543073	10	32	spacer			
543488	122	0	set box			setbox
543478	12	48	other char	0	U+00030	
538395	30	14	make box			hbox
543743	1	123	left brace			
543391	12	50	other char	2	U+00032	
543797	2	125	right brace			
543803	10	32	spacer			
543817	115	1	register			count
543471	12	48	other char	0	U+00030	
540503	12	61	other char	=	U+0003D	
261633	12	51	other char	3	U+00033	
543165	10	32	spacer			
542704	12	49	other char	1	U+00031	

We now introduce a new primitive `\localcontrolled`:

```
\edef\TestB{1 \setbox0\hbox{2} \count0=3 \the\count0}
```

```
\edef\TestC{1 \setbox0\hbox{2} \localcontrolled{\count0=3} \the\count0}
```

Again, watch the subtle differences:

control sequence: TestB

543823	12	49	other char	1	U+00031	
542809	10	32	spacer			

543234	122	0	set box			setbox
543056	12	48	other char	0	U+00030	
543051	30	14	make box			hbox
28960	1	123	left brace			
540454	12	50	other char	2	U+00032	
543117	2	125	right brace			
543253	10	32	spacer			
543057	115	1	register			count
413052	12	48	other char	0	U+00030	
261632	12	61	other char	=	U+0003D	
540295	12	51	other char	3	U+00033	
543903	10	32	spacer			
548177	12	49	other char	1	U+00031	

control sequence: TestC

540206	12	49	other char	1	U+00031	
539343	10	32	spacer			
539112	122	0	set box			setbox
540544	12	48	other char	0	U+00030	
540262	30	14	make box			hbox
226994	1	123	left brace			
119162	12	50	other char	2	U+00032	
543932	2	125	right brace			
543581	10	32	spacer			
542587	10	32	spacer			
543924	12	51	other char	3	U+00033	

Another example:

```
\edef\TestB{1 \setbox0\hbox{2} \count0=3 \the\count0}
```

```
\edef\TestD{\localcontrolled{1 \setbox0\hbox{2} \count0=3 \the\count0}}
```

1 3 ← Watch how the results end up here!

control sequence: TestB

32744	12	49	other char	1	U+00031	
413154	10	32	spacer			
543012	122	0	set box			setbox
543807	12	48	other char	0	U+00030	

540485	30	14	make box			hbox
542616	1	123	left brace			
543825	12	50	other char	2	U+00032	
543302	2	125	right brace			
539126	10	32	spacer			
413208	115	1	register			count
542738	12	48	other char	0	U+00030	
538607	12	61	other char	=	U+0003D	
543018	12	51	other char	3	U+00033	
543891	10	32	spacer			
211385	12	51	other char	3	U+00033	

control sequence: TestD

<no tokens>

We can use this mechanism to define so called fully expandable macros:

```
\def\WidthOf#1%
{\beginlocalcontrol
\setbox0\hbox{#1}%
\endlocalcontrol
\wd0 }
```

```
\scratchdimen\WidthOf{The Rite Of Spring}
```

```
\the\scratchdimen
```

104.72021pt

When you want to add some grouping, it quickly can become less pretty:

```
\def\WidthOf#1%
{\dimexpr
\beginlocalcontrol
\begingroup
\setbox0\hbox{#1}%
\expandafter
\endgroup
\expandafter
\endlocalcontrol
\the\wd0
```

```
\relax}
```

```
\scratchdimen\WidthOf{The Rite Of Spring}
```

```
\the\scratchdimen
```

```
104.72021pt
```

A single token alternative is available too and its usage is like this:

```
\def\TestA{\scratchcounter=100 }
\edef\TestB{\localcontrol\TestA \the\scratchcounter}
\edef\TestC{\localcontrolled{\TestA} \the\scratchcounter}
```

The content of `\TestB` is '100' and of course the `\TestC` macro gives ' 100'.

We now move to the Lua end. Right from the start the way to get something into \TeX from Lua has been the print functions. But we can also go local (immediate). There are several methods:

- via a set token register
- via a defined macro
- via a string

Among the things to keep in mind are catcodes, scope and expansion (especially in when the result itself ends up in macros). We start with an example where we go via a token register:

```
\toks0={\setbox0\hbox{The Rite Of Spring}}
\toks2={\setbox0\hbox{The Rite Of Spring!}}
```

```
\startluacode
```

```
tex.runlocal(0) context("[1: %p]",tex.box[0].width)
```

```
tex.runlocal(2) context("[2: %p]",tex.box[0].width)
```

```
\stopluacode
```

```
[1: 104.72021pt][2: 109.14062pt]
```

We can also use a macro:

```
\def\TestA{\setbox0\hbox{The Rite Of Spring}}
\def\TestB{\setbox0\hbox{The Rite Of Spring!}}
```

```
\startluacode
```

```

tex.runlocal("TestA") context("[3: %p]",tex.box[0].width)
tex.runlocal("TestB") context("[4: %p]",tex.box[0].width)
\stoptluacode

```

[3: 104.72021pt][4: 109.14062pt]

A third variant is more direct and uses a (Lua) string:

```

\startluacode
tex.runstring([[ \setbox0\hbox{The Rite Of Spring} ]])
context("[5: %p]",tex.box[0].width)
tex.runstring([[ \setbox0\hbox{The Rite Of Spring!} ]])
context("[6: %p]",tex.box[0].width)
\stoptluacode

```

[5: 104.72021pt][6: 109.14062pt]

A bit more high level:

```

context.runstring([[ \setbox0\hbox{(Here \bf 1.2345)} ]])
context.runstring([[ \setbox0\hbox{(Here \bf %.3f)} ]],1.2345)

```

Before we had runstring this was the way to do it when staying in Lua was needed:

```

\startluacode
token.setmacro("TestX",[[ \setbox0\hbox{The Rite Of Spring} ]])
tex.runlocal("TestX")
context("[7: %p]",tex.box[0].width)
\stoptluacode

```

[7: 104.72021pt]

```

\startluacode
tex.scantoks(0,tex.ctxcatcodes,[[ \setbox0\hbox{The Rite Of Spring!} ]])
tex.runlocal(0)
context("[8: %p]",tex.box[0].width)
\stoptluacode

```

[8: 109.14062pt]

The order of flushing matters because as soon as something is not stored in a token list or macro body, \TeX will typeset it. And as said, a lot of this relates to pushing stuff into the input which is stacked. Compare:

```
\startluacode
context("[HERE 1]")
context("[HERE 2]")
\stopluacode
```

[HERE 1][HERE 2]

with this:

```
\startluacode
tex.pushlocal() context("[HERE 1]") tex.poplocal()
tex.pushlocal() context("[HERE 2]") tex.poplocal()
\stopluacode
```

[HERE 1][HERE 2]

You can expand a macro at the Lua end with `token.expandmacro` which has a peculiar interface. The first argument has to be a string (the name of a macro) or a userdata (a valid macro token). This macro can be fed with parameters by passing more arguments:

string	serialized to tokens
true	wrap the next string in curly braces
table	each entry will become an argument wrapped in braces
token	inject the token directly
number	change control to the given catcode table

There are more scanner related primitives, like the ε -TeX primitive `\detokenize`:

```
[\detokenize {test \relax}]
```

This gives: [test \relax] . In LuaMetaTeX we also have complementary primitive(s):

```
[\tokenized catcodetable \vrbcatcodes {test {\bf test} test}]
[\tokenized {test {\bf test} test}]
[\retokenized \vrbcatcodes {test {\bf test} test}]
```

The `\tokenized` takes an optional keyword and the examples above give: [test {\bf test} test] [test **test** test] [test {\bf test} test] . The LuaTeX primitive `\scantextokens` which is a variant of ε -TeX's `\scantokens` operates under the current catcode regime (the last one honors `\everyeof`). The difference with `\tokenized` is that this one first serializes the given token list (just like `\detokenize`).³

³ The `\scan *tokens` primitives now share the same helpers as Lua, but they should behave the same as in LuaTeX.

With `\retokenized` the catcode table index is mandatory (it saves a bit of scanning and is easier on intermixed `\expandafter` usage. There often are several ways to accomplish the same:

```
\def\MyTitle{test {\bf test} test}
\detokenize          \expandafter{\MyTitle}: 0.46\crlf
\meaningless        \MyTitle : 0.47\crlf
\retokenized         \notcatcodes{\MyTitle}: 0.87\crlf
\tokenized   catcodetable \notcatcodes{\MyTitle}: 0.93\crlf
```

```
test {\bf test} test: 0.46
test {\bf test} test: 0.47
test {\bf test} test: 0.87
test {\bf test} test: 0.93
```

Here the numbers show the relative performance of these methods. The `\detokenize` and `\meaningless` win because they already know that a verbose serialization is needed. The last two first serialize and then reinterpret the resulting token list using the given catcode regime. The last one is slowest because it has to scan the keyword.

There is however a pitfall here:

```
\def\MyText {test}
\def\MyTitle{test \MyText\space test}
\detokenize          \expandafter{\MyTitle}\crlf
\meaningless        \MyTitle \crlf
\retokenized         \notcatcodes{\MyTitle}\crlf
\tokenized   catcodetable \notcatcodes{\MyTitle}\crlf
```

The outcome is different now because we have an expandable embedded macro call. The fact that we expand in the last two primitives is also the reason why they are ‘slower’.

```
test \MyText \space test
test \MyText \space test
test test test
test test test
```

To complete this picture, we show a variant than combines much of what has been introduced in this section:

```

\semiprotected\def\MyTextA {test}
\def\MyTextB {test}
\def\MyTitle{test \MyTextA\space \MyTextB\space test}
\detokenize          \expandafter{\MyTitle}\crlf
\meaningless        \MyTitle \crlf
\retokenized         \notcatcodes{\MyTitle}\crlf
\retokenized         \notcatcodes{\semiexpanded{\MyTitle}}\crlf
\tokenized   catcodetable \notcatcodes{\MyTitle}\crlf
\tokenized   catcodetable \notcatcodes{\semiexpanded{\MyTitle}}

```

This time compare the last four lines:

```

test \MyTextA \space \MyTextB \space test
test \MyTextA \space \MyTextB \space test
test \MyTextA test test
test test test test
test \MyTextA test test
test test test test

```

Of course the question remains to what extent we need this and eventually will apply in ConT_EXt. The `\detokenize` is used already. History shows that eventually there is a use for everything and given the way LuaMetaT_EX is structured it was not that hard to provide the alternatives without sacrificing performance or bloating the source.

4.6 Dirty tricks

When I was updating this manual Hans vd Meer and I had some discussions about expansion and tokenization related issues when combining of xml processing with T_EX macros where he did some manipulations in Lua. In these mixed cases you can run into catcode related problems because in xml you want for instance a `#` to be a hash mark (other character) and not an parameter identifier. Normally this is handled well in ConT_EXt but of course there are complex cases where you need to adapt.

Say that you want to compare two strings (officially we should say token lists) with mixed catcodes. Let's also assume that you want to use the normal `\if` construct (which was part of the discussion). We start with defining a test set. The reason that we present this example here is that we use commands discussed in previous sections:

```

\def\abc{abc}
\semiprotected \def\xyz{xyz}
\edef\pqr{\expandtoken\notcatcodes`p%

```

```
\expandtoken\notcatcodes`q%
\expandtoken\notcatcodes`r}
```

```
1: \ifcondition\similartokens{abc} {def}YES\else NOP\fi (NOP) \quad
2: \ifcondition\similartokens{abc}{\abc}YES\else NOP\fi (YES)

3: \ifcondition\similartokens{xyz} {pqr}YES\else NOP\fi (NOP) \quad
4: \ifcondition\similartokens{xyz}{\xyz}YES\else NOP\fi (YES)

5: \ifcondition\similartokens{pqr} {pqr}YES\else NOP\fi (YES) \quad
6: \ifcondition\similartokens{pqr}{\pqr}YES\else NOP\fi (YES)
```

So, we have a mix of expandable and semi expandable macros, and also a mix of cat-codes. A naive approach would be:

```
\permanent\protected\def\similartokens#1#2%
  {\iftok{#1}{#2}}
```

but that will fail on some cases:

```
1: NOP(NOP)   2: YES(YES)
3: NOP(NOP)   4: NOP(YES)
5: YES(YES)   6: NOP(YES)
```

So how about:

```
\permanent\protected\def\similartokens#1#2%
  {\iftok{\detokenize{#1}}{\detokenize{#2}}}
```

That one is even worse:

```
1: NOP(NOP)   2: NOP(YES)
3: NOP(NOP)   4: NOP(YES)
5: YES(YES)   6: NOP(YES)
```

We need to expand so we end up with this:

```
\permanent\protected\def\similartokens#1#2%
  {\normalexpanded{\noexpand\iftok
    {\noexpand\detokenize{#1}}
    {\noexpand\detokenize{#2}}}}
```

Better:

```
1: NOP(NOP)   2: YES(YES)
```

3: NOP(NOP) 4: NOP(YES)
 5: YES(YES) 6: YES(YES)

But that will still not deal with the mildly protected macro so in the end we have:

```
\permanent\protected\def\similartokens#1#2%
  {\semiexpanded{\noexpand\iftok
    {\noexpand\detokenize{#1}}
    {\noexpand\detokenize{#2}}}}
```

Now we're good:

1: NOP(NOP) 2: YES(YES)
 3: NOP(NOP) 4: YES(YES)
 5: YES(YES) 6: YES(YES)

Finally we wrap this one in the usual `\doifelse...` macro:

```
\permanent\protected\def\doifelsesimilartokens#1#2%
  {\ifcondition\similartokens{#1}{#2}%
    \expandafter\firstoftwoarguments
  \else
    \expandafter\secondoftwoarguments
  \fi}
```

so that we can do:

```
\doifelsesimilartokens{pqr}{\pqr}{YES}{NOP}
```

A companion macro of this is `\wipetoken` but for that one you need to look into the source.

4.6 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

5 Registers

low level

TEX

registers

Contents

5.1	Preamble	78
5.2	$\text{T}_{\text{E}}\text{X}$ primitives	78
5.3	$\varepsilon\text{-T}_{\text{E}}\text{X}$ primitives	81
5.4	Lua $\text{T}_{\text{E}}\text{X}$ primitives	81
5.5	LuaMeta $\text{T}_{\text{E}}\text{X}$ primitives	82

5.1 Preamble

Registers are sets of variables that are accessed by index and as such resemble registers in a processing unit. You can store a quantity in a register, retrieve it, and also manipulate it.

There is hardly any need to use them in Con $\text{T}_{\text{E}}\text{X}$ t so we keep it simple.

5.2 $\text{T}_{\text{E}}\text{X}$ primitives

There are several categories:

- Integers (int): in order to be portable (at the time it surfaced) there are only integers and no floats. The only place where $\text{T}_{\text{E}}\text{X}$ uses floats internally is when glue gets effective which happens in the backend.
- Dimensions (dimen): internally these are just integers but when they are entered they are sliced into two parts so that we have a fractional part. The internal representation is called a scaled point.
- Glue (skip): these are dimensions with a few additional properties: stretch and shrink. Being a compound entity they are stored differently and thereby a bit less efficient than numbers and dimensions.
- Math glue (muskip): this is the same as glue but with a unit that adapts to the current math style properties. It's best to think about them as being relative measures.
- Token lists (toks): these contain a list of tokens coming from the input or coming from a place where they already have been converted.

The original $\text{T}_{\text{E}}\text{X}$ engine had 256 entries per set. The first ten of each set are normally reserved for scratch purposes: the even ones for local use, and the odd ones for global usage. On top of that macro packages can reserve some for its own use. It was quite

easy to reach the maximum but there were tricks around that. This limitation is no longer present in the variants in use today.

Let's set a few dimension registers:

```
\dimen 0 = 10 pt
\dimen2=10pt
\dimen4 10pt
\scratchdimen 10pt
```

We can serialize them with:

```
\the \dimen0
\number \dimen2
\meaning\dimen4
\meaning\scratchdimen
```

The results of these operations are:

```
10.0pt
655360
\dimen4
dimension 10.0pt
```

The last two is not really useful but it is what you see when tracing options are set. Here `\scratchdimen` is a shortcut for a register. This is *not* a macro but a defined register. The low level `\dimendef` is used for this but in a macro package you should not use that one but the higher level `\newdimen` macro that uses it.

```
\newdimen\MyDimenA
\def \MyDimenB{\dimen999}
\dimendef\MyDimenC998

\meaning\MyDimenA
\meaning\MyDimenB
\meaning\MyDimenC
```

Watch the difference:

```
\dimen264
macro:\dimen 999
\dimen998
```

The first definition uses a yet free register so you won't get a clash. The second one is just a shortcut using a macro and the third one too but again direct shortcut. Try to imagine how the second line gets interpreted:

```
\MyDimenA10pt \MyDimenA10.5pt
\MyDimenB10pt \MyDimenB10.5pt
\MyDimenC10pt \MyDimenC10.5pt
```

Also try to imagine what messing around with `\MyDimenC` will do when we also have defined a few hundred extra dimensions with `\newdimen`.

In the case of dimensions the `\number` primitive will make the register serialize as scaled points without unit `sp`.

Next we see some of the other registers being assigned:

```
\count 0 = 100
\skip 0 = 10pt plus 3pt minus 2pt
\skip 0 = 10pt plus 1fill
\muskip 0 = 10mu plus 3mu minus 2mu
\muskip 0 = 10mu minus 1 fil
\toks 0 = {hundred}
```

When a number is expected, you can use for instance this:

```
\scratchcounter\scratchcounterone
```

Here we use a few predefined scratch registers. You can also do this:

```
\scratchcounter\numexpr\scratchcounterone+\scratchcountertwo\relax
```

There are some quantities that also qualify as number:

```
\chardef\MyChar=123 % refers to character 123 (if present)
\scratchcounter\MyChar
```

In the past using `\chardef` was a way to get around the limited number of registers, but it still had (in traditional \TeX) a limitation: you could not go beyond 255. The `\mathchardef` could go higher as it also encodes a family number and class. This limitation has been lifted in \LuaTeX .

A character itself can also be interpreted as number, in which case it has to be prefixed with a reverse quote: ```, so:

```
\scratchcounter\numexpr`0+5\relax
```

```
\char\scratchcounter
```

produces “5” because the `\0` expands into the (ascii and utf8) slot 48 which represents the character zero. In this case the next makes more sense:

```
\char\numexpr`0+5\relax
```

If you want to know more about all these quantities, “TeX By Topic” provides a good summary of what TeX has to offer, and there is no need to repeat it here.

5.3 ε -TeX primitives

Apart from the ability to use expressions, the contribution to registers that ε -TeX brought was that suddenly we could use upto 65K of them, which is more than enough. The extra registers were not as efficient as the first 256 because they were stored in the hash table, but that was not really a problem. In Omega and later LuaTeX regular arrays were used, at the cost of more memory which in the meantime has become cheap. As ConTeXt moved to ε -TeX rather early its users never had to worry about it.

5.4 LuaTeX primitives

The LuaTeX engine introduced attributes. These are numeric properties that are bound to the nodes that are the result of typesetting operations. They are basically like integer registers but when set their values get bound and when unset they are kind of invisible.

- Attribute (attribute): a numeric property that when set becomes part of the current attribute list that gets assigned to nodes.

Attributes can be used to communicate properties to Lua callbacks. There are several functions available for setting them and querying them.

```
\attribute999 = 123
```

Using attributes this way is dangerous (of course I can only speak for ConTeXt) because an attribute value might trigger some action in a callback that gives unwanted side effects. For convenience ConTeXt provides:

```
\newattribute\MyAttribute
```

Which currently defines `\MyAttribute` as integer 1026 and is meant to be used as:⁴

⁴ The low level `\attributedef` command is rather useless in the perspective of ConTeXt.

```
\attribute\MyAttribute = 123
```

Just be aware that defining attributes can have an impact on performance. As you cannot access them at the T_EX end you seldom need them. If you do you can better use the proper more high level definers (not discussed here).

5.5 LuaMetaT_EX primitives

The fact that scanning stops at a non-number or `\relax` can be sort of unpredictable which is why in LuaMetaT_EX we also support the following variant:

```
\scratchdimen\dimexpr 10pt + 3pt \relax  
\scratchdimen\dimexpr {10pt + 3pt}
```

At the cost of one more token braces can be used as boundaries instead of the single `\relax` boundary.

An important property of registers is that they can be accessed by a number. This has big consequences for the implementation: they are part of the big memory store and consume dedicated ranges. If we had only named access T_EX's memory layout could be a bit leaner. In principle we could make the number of registers smaller because any limit on the amount at some point can be an obstacle. It is for that reason that we also have name-only variants:

```
\dimensiondef \MyDimenA 12pt  
\integerdef \MyIntegerA 12  
\gluespecdef \MyGlueA 12pt + 3pt minus 4pt  
\mugluespecdef\MyMuA 12mu + 3mu minus 4mu
```

These are as efficient but not accessible by number but they behave like registers which means that you (can) use `\the`, `\advance`, `\multiply` and `\divide` with them.⁵ In case you wonder why there is no alternative for `\toksdef`, there actually are multiple: they are called macros.

todo: expressions

⁵ There are also the slightly more efficient `\advanceby`, `\multiplyby` and `\divideby` that don't check for the by keyword.

5.5 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

6 Macros

low level

TEX

macros

Contents

6.1	Preamble	85
6.2	Definitions	85
6.3	Runaway arguments	95
6.4	Introspection	96
6.5	nesting	97
6.6	Prefixes	100
6.7	Arguments	102
6.8	Constants	103

6.1 Preamble

This chapter overlaps with other chapters but brings together some extensions to the macro definition and expansion parts. As these mechanisms were stepwise extended, the other chapters describe intermediate steps in the development.

Now, in spite of the extensions discussed here the main idea is still that we have \TeX act like before. We keep the charm of the macro language but these additions make for easier definitions, but (at least initially) none that could not be done before using more code.

6.2 Definitions

A macro definition normally looks like like this:⁶

```
\def\macro#1#2%
  {\dontleavehmode\hbox to 6em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

Such a macro can be used as:

```
\macro {1}{2}
\macro {1} {2} middle space gobbled
\macro 1 {2} middle space gobbled
\macro {1} 2 middle space gobbled
\macro 1 2 middle space gobbled
```

We show the result with some comments about how spaces are handled:

⁶ The `\dontleavehmode` command make the examples stay on one line.

```

12
12      middle space gobbled
12      middle space gobbled
12      middle space gobbled
12      middle space gobbled

```

A definition with delimited parameters looks like this:

```

\def\macro[#1]%
  {\dontleavehmode\hbox to 6em{\vl\type{#1}\vl\hss}}

```

When we use this we get:

```

\macro [1]
\macro [ 1]   leading space kept
\macro [1 ]   trailing space kept
\macro [ 1 ]   both spaces kept

```

Again, watch the handling of spaces:

```

1
1 |   leading space kept
1 |   trailing space kept
1 |   both spaces kept

```

Just for the record we show a combination:

```

\def\macro[#1]#2%
  {\dontleavehmode\hbox to 6em{\vl\type{#1}\vl\type{#2}\vl\hss}}

```

With this:

```

\macro [1]{2}
\macro [1] {2}
\macro [1] 2

```

we can again see the spaces go away:

```

12
12
12

```

A definition with two separately delimited parameters is given next:

```

\def\macro[#1#2]%

```

```
{\dontleavehmode\hbox to 6em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

When used:

```
\macro [12]
\macro [ 12]    leading space gobbled
\macro [12 ]   trailing space kept
\macro [ 12 ]   leading space gobbled, trailing space kept
\macro [1 2]   middle space kept
\macro [ 1 2 ] leading space gobbled, middle and trailing space kept
```

We get ourselves:

```
|12|
|12|    leading space gobbled
|12|    trailing space kept
|12|    leading space gobbled, trailing space kept
|1 2|   middle space kept
|1 2|   leading space gobbled, middle and trailing space kept
```

These examples demonstrate that the engine does some magic with spaces before (and therefore also between multiple) parameters.

We will now go a bit beyond what traditional T_EX engines do and enter the domain of LuaMetaT_EX specific parameter specifiers. We start with one that deals with this hard coded space behavior:

```
\def\macro[#^#^]%
  {\dontleavehmode\hbox to 6em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

The #[^] specifier will count the parameter, so here we expect again two arguments but the space is kept when parsing for them.

```
\macro [12]
\macro [ 12]
\macro [12 ]
\macro [ 12 ]
\macro [1 2]
\macro [ 1 2 ]
```

Now keep in mind that we could deal well with all kind of parameter handling in ConT_EXt for decades, so this is not really something we missed, but it complements the to be discussed other ones and it makes sense to have that level of control. Also, availability triggers usage. Nevertheless, some day the #[^] specifier will come in handy.

```

12|
|12|
12|
|12|
1 2|
|1 2|

```

We now come back to an earlier example:

```

\def\macro[#1]%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\hss}}

```

When we use this we see that the braces in the second call are removed:

```

\macro [1]
\macro [{1}]

```

```

|1| |1|

```

This can be prohibited by the #+ specifier, as in:

```

\def\macro[#+]%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\hss}}

```

As we see, the braces are kept:

```

\macro [1]
\macro [{1}]

```

Again, we could easily get around that (for sure intended) side effect but it just makes nicer code when we have a feature like this.

```

|1| |{1}|

```

Sometimes you want to grab an argument but are not interested in the results. For this we have two specifiers: one that just ignores the argument, and another one that keeps counting but discards it, i.e. the related parameter is empty.

```

\def\macro[#1][#0][#3][#-][#4]%
  {\dontleavehmode\hbox spread 1em
   {\vl\type{#1}\vl\type{#2}\vl\type{#3}\vl\type{#4}\vl\hss}}

```

The second argument is empty and the fourth argument is simply ignored which is why we need #4 for the fifth entry.

```
\macro [1][2][3][4][5]
```

Here is proof that it works:

```
|1|3|5|
```

The reasoning behind dropping arguments is that for some cases we get around the nine argument limitation, but more important is that we don't construct token lists that are not used, which is more memory (and maybe even cpu cache) friendly.

Spaces are always kind of special in T_EX, so it will be no surprise that we have another specifier that relates to spaces.

```
\def\macro[#1]#* [#2]%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

This permits usage like the following:

```
\macro [1][2]
\macro [1] [2]
```

```
|1|2| |1|2|
```

Without the optional 'grab spaces' specifier the second line would possibly throw an error. This because T_EX then tries to match][so the] [in the input is simply added to the first argument and the next occurrence of][will be used. That one can be someplace further in your source and if not T_EX complains about a premature end of file. But, with the #* option it works out okay (unless of course you don't have that second argument [2]).

Now, you might wonder if there is a way to deal with that second delimited argument being optional and of course that can be programmed quite well in traditional macro code. In fact, ConT_EXt does that a lot because it is set up as a parameter driven system with optional arguments. That subsystem has been optimized to the max over years and it works quite well and performance wise there is very little to gain. However, as soon as you enable tracing you end up in an avalanche of expansions and that is no fun.

This time the solution is not in some special specifier but in the way a macro gets defined.

```
\tolerant\def\macro[#1]#* [#2]%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

The magic \tolerant prefix with delimited arguments and just quits when there is no match. So, this is acceptable:

```
\macro [1][2]
\macro [1] [2]
\macro [1]
\macro
```

```
|12| |12| |1| |
```

We can check how many arguments have been processed with a dedicated conditional:

```
\tolerant\def\macro[#1]#*[#2]%
  {\ifarguments 0\or 1\or 2\or ?\fi: \vl\type{#1}\vl\type{#2}\vl}
```

We use this test:

```
\macro [1][2] \macro [1] [2] \macro [1] \macro
```

The result is: 2: |12| 2: |12| 1: |1|0: | which is what we expect because we flush inline and there is no change of mode. When the following definition is used in display mode, the leading n= can for instance start a new paragraph and when code in `\everypar` you can loose the right number when macros get expanded before the n gets injected.

```
\tolerant\def\macro[#1]#*[#2]%
  {n=\ifarguments 0\or 1\or 2\or ?\fi: \vl\type{#1}\vl\type{#2}\vl}
```

In addition to the `\ifarguments` test primitive there is also a related internal counter `\lastarguments` set that you can consult, so the `\ifarguments` is actually just a shortcut for `\ifcase\lastarguments`.

We now continue with the argument specifiers and the next two relate to this optional grabbing. Consider the next definition:

```
\tolerant\def\macro#1#*#2%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

With this test:

```
\macro {1} {2}
\macro {1}
\macro
```

We get:

```
|12| |1|\macro|
```

This is okay because the last `\macro` is a valid (single token) argument. But, we can make the braces mandate:

```
\tolerant\def\macro#=#*#=#%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

Here the `#=` forces a check for braces, so:

```
\macro {1} {2}
\macro {1}
\macro
```

gives this:

```
|1|2| |1| |
```

However, we do lose these braces and sometimes you don't want that. Of course when you pass the results downstream to another macro you can always add them, but it was cheap to add a related specifier:

```
\tolerant\def\macro#_#*#_%
  {\dontleavehmode\hbox spread 1em{\vl\type{#1}\vl\type{#2}\vl\hss}}
```

Again, the magic `\tolerant` prefix works will quit scanning when there is no match. So:

```
\macro {1} {2}
\macro {1}
\macro
```

leads to:

```
|{1}|{2}| |{1}| |
```

When you're tolerant it can be that you still want to pick up some argument later on. This is why we have a continuation option.

```
\tolerant\def\foo      [#1]*[#2]#:#3{!#1!#2!#3!}
\tolerant\def\oof[#1]*[#2]#:(#3)#:#4{!#1!#2!#3!#4!}
\tolerant\def\ofo      [#1]#:(#2)#:#3{!#1!#2!#3!}
```

Hopefully the next example demonstrates how it works:

```
\foo{3} \foo[1]{3} \foo[1][2]{3}
```

```

\oof{4} \oof[1]{4} \oof[1][2]{4}
\oof[1][2](3){4} \oof[1](3){4} \oof(3){4}
\ofo{3} \ofo[1]{3}
\ofo[1](2){3} \ofo(2){3}

```

As you can see we can have multiple continuations using the #: directive:

```

!!!3! !1!!3! !1!2!3!
!!!!4! !1!!!!4! !1!2!!!4!
!1!2!3!4! !1!!3!4! !!!3!4!
!!!3! !1!!3!
!1!2!3! !!2!3!

```

The last specifier doesn't work well with the \ifarguments state because we no longer know what arguments were skipped. This is why we have another test for arguments. A zero value means that the next token is not a parameter reference, a value of one means that a parameter has been set and a value of two signals an empty parameter. So, it reports the state of the given parameter as a kind of \ifcase.

```

\def\foo#1#2{ [\ifparameter#1\or(ONE)\fi\ifparameter#2\or(TWO)\fi] }

```

Of course the test has to be followed by a valid parameter specifier:

```

\foo{1}{2} \foo{1}{} \foo{}{2} \foo{}{}

```

The previous code gives this:

```

[(ONE)(TWO)] [(ONE)] [(TWO)] []

```

A combination check \ifparameters, again a case, matches the first parameter that has a value set.

We could add plenty of specifiers but we need to keep in mind that we're not talking of an expression scanner. We need to keep performance in mind, so nesting and backtracking are no option. We also have a limited set of useable single characters, but here's one that uses a symbol that we had left:

```

\def\startfoo[#/#/#/\stopfoo{ [#1](#2) }

```

The slash directive removes leading and trailing so called spacers as well as tokens that represent a paragraph end:

```

\startfoo [x ] x \stopfoo
\startfoo [ x ] x \stopfoo

```

```
\startfoo [ x] x \stopfoo
\startfoo [ x] \par x \par \par \stopfoo
```

So we get this:

```
[x](x) [x](x) [x](x) [x](x)
```

The next directive, the quitter #;, is demonstrated with an example. When no match has occurred, scanning picks up after this signal, otherwise we just quit.

```
\tolerant\def\foo[#1]#;(#2){/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par
\foo(1)\quad\foo(2)\quad\foo(3)\par
```

```
\tolerant\def\foo[#1]#;#=#{/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```
\tolerant\def\foo[#1]#;#2{/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```
\tolerant\def\foo[#1]#;(#2)#;#=#{/#1/#2/#3/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par
\foo(1)\quad\foo(2)\quad\foo(3)\par
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```
/1// /2// /3//
```

```
//1/ //2/ //3/
```

```
/1// /2// /3//
```

```
//1/ //2/ //3/
```

```
/1// /2// /3//
```

```
//1/ //2/ //3/
```

```
/1/// /2/// /3///
```

```
//1// //2// //3//
```

```
///1/ ///2/ ///3/
```

I have to admit that I don't really need it but it made some macros that I was redefining behave better, so there is some self-interest here. Anyway, I considered some other features, like picking up a detokenized argument but I don't expect that to be of much

use. In the meantime we ran out of reasonable characters, but some day #? and #! might show up, or maybe I find a use for #< and #>. A summary of all this is given here:

+	keep the braces
-	discard and don't count the argument
/	remove leading and trailing spaces and pars
=	braces are mandate
_	braces are mandate and kept
^	keep leading spaces

1-9	an argument
0	discard but count the argument

*	ignore spaces
:	pick up scanning here
;	quit scanning

.	ignore pars and spaces
,	push back space when quit

The last two have not been discussed and were added later. The period directive gobbles space and par tokens and discards them in the process. The comma directive is like * but it pushes back a space when the matching quits.

```
\tolerant\def\foo[#1]#; (#2) {/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par  
\foo(1)\quad\foo(2)\quad\foo(3)\par
```

```
\tolerant\def\foo[#1]#;#={/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par  
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```
\tolerant\def\foo[#1]#;#2{/#1/#2/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par  
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```
\tolerant\def\foo[#1]#; (#2)#;#={/#1/#2/#3/}
```

```
\foo[1]\quad\foo[2]\quad\foo[3]\par  
\foo(1)\quad\foo(2)\quad\foo(3)\par  
\foo{1}\quad\foo{2}\quad\foo{3}\par
```

```

/1// /2// /3//
//1/ //2/ //3/
/1// /2// /3//
//1/ //2/ //3/
/1// /2// /3//
//1/ //2/ //3/
/1/// /2/// /3///
//1// //2// //3//
///1/ ///2/ ///3/

```

Gobbling spaces versus pushing back is an interface design decision because it has to do with consistency.

6.3 Runaway arguments

There is a particular troublesome case left: a runaway argument. The solution is not pretty but it's the only way: we need to tell the parser that it can quit.

```

\tolerant\def\foo[#1=#2]%
  {\ifarguments 0\or 1\or 2\or 3\or 4\fi:\v\type{#1}\v\type{#2}\v\}

```

The outcome demonstrates that one still has to do some additional checking for sane results and there are alternative way to (ab)use this mechanism. It all boils down to a clever combination of delimiters and `\ignorearguments`.

```

\dontleavehmode \foo[a=1]
\dontleavehmode \foo[b=]
\dontleavehmode \foo[=]
\dontleavehmode \foo[x]\ignorearguments

```

All calls are accepted:

```

2:a|
2:b|
2:|
1:x]|

```

Just in case you wonder about performance: don't expect miracles here. On the one hand there is some extra overhead in the engine (when defining macros as well as when collecting arguments during a macro call) and maybe using these new features can sort of compensate that. As mentioned: the gain is mostly in cleaner macro code and less clutter in tracing. And I just want the ConT_EXt code to look nice: that way users

can look in the source to see what happens and not drown in all these show-off tricks, special characters like underscores, at signs, question marks and exclamation marks.

For the record: I normally run tests to see if there are performance side effects and as long as processing the test suite that has thousands of files of all kind doesn't take more time it's okay. Actually, there is a little gain in ConT_EXt but that is to be expected, but I bet users won't notice it, because it's easily offset by some inefficient styling. Of course another gain of loosing some indirectness is that error messages point to the macro that the user called for and not to some follow up.

6.4 Introspection

A macro has a meaning. You can serialize that meaning as follows:

```
\tolerant\protected\def\foo#1[#2]#* [#3]%
  {(1=#1) (2=#3) (3=#3)}
```

```
\meaning\foo
```

The meaning of \foo comes out as:

```
tolerant protected macro:#1[#2]#* [#3]->(1=#1) (2=#3) (3=#3)
```

When you load the module system-tokens you can also say:

```
\luatokenable\foo
```

This produces a table of tokens specifications:

```
tolerant protected macro:#1[#2]#* [#3]->(1=#1) (2=#3) (3=#3)
```

tolerant protected control sequence: foo

543628	19	49	match	argument 1
543398	12	91	other char	[U+0005B
540301	19	50	match	argument 2
543764	12	93	other char] U+0005D
543787	19	42	match	argument *
543814	12	91	other char	[U+0005B
543226	19	51	match	argument 3
543231	12	93	other char] U+0005D
543989	20	0	end match	
<hr/>				
540278	12	40	other char	(U+00028

543734	12	49	other char	1	U+00031
461605	12	61	other char	=	U+0003D
539284	21	1	parameter reference		
544622	12	41	other char)	U+00029
544576	10	32	spacer		
542952	12	40	other char	(U+00028
540508	12	50	other char	2	U+00032
544645	12	61	other char	=	U+0003D
540329	21	3	parameter reference		
543682	12	41	other char)	U+00029
543079	10	32	spacer		
543592	12	40	other char	(U+00028
542823	12	51	other char	3	U+00033
544639	12	61	other char	=	U+0003D
543685	21	3	parameter reference		
543864	12	41	other char)	U+00029

A token list is a linked list of tokens. The magic numbers in the first column are the token memory pointers. and because macros (and token lists) get recycled at some point the available tokens get scattered, which is reflected in the order of these numbers. Normally macros defined in the macro package are more sequential because they stay around from the start. The second and third row show the so called command code and the specifier. The command code groups primitives in categories, the specifier is an indicator of what specific action will follow, a register number a reference, etc. Users don't need to know these details. This macro is a special version of the online variant:

```
\showluatokens\foo
```

That one is always available and shows a similar list on the console. Again, users normally don't want to know such details.

6.5 nesting

You can nest macros, as in:

```
\def\foo#1#2{\def\oof##1{<#1>##1<#2>}}
```

At first sight the duplication of # looks strange but this is what happens. When T_EX scans the definition of `\foo` it sees two arguments. Their specification ends up in the preamble that defines the matching. When the body is scanned, the `#1` and `#2` are turned into a parameter reference. In order to make nested macros with arguments

possible a # followed by another # becomes just one #. Keep in mind that the definition of \oof is delayed till the macro \foo gets expanded. That definition is just stored and the only thing that get's replaced are the two references to a macro parameter

control sequence: foo

543973	19	49	match		argument 1
539345	19	50	match		argument 2
543765	20	0	end match		
<hr/>					
544064	121	1	def		def
543087	140	0	tolerant call		oof
542586	6	35	parameter		
542737	12	49	other char	1	U+00031
544000	1	123	left brace		
540363	12	60	other char	<	U+0003C
179409	21	1	parameter reference		
463017	12	62	other char	>	U+0003E
544834	6	35	parameter		
543299	12	49	other char	1	U+00031
373598	12	60	other char	<	U+0003C
540351	21	2	parameter reference		
544036	12	62	other char	>	U+0003E
543503	2	125	right brace		

Now, when we look at these details, it might become clear why for instance we have ‘variable’ names like #4 and not #whatever (with or without hash). Macros are essentially token lists and token lists can be seen as a sequence of numbers. This is not that different from other programming environments. When you run into buzzwords like ‘bytecode’ and ‘virtual machines’ there is actually nothing special about it: some high level programming (using whatever concept, and in the case of T_EX it's macros) eventually ends up as a sequence of instructions, say bytecodes. Then you need some machinery to run over that and act upon those numbers. It's something you arrive at naturally when you play with interpreting languages.⁷

So, internally a #4 is just one token, a operator-operand combination where the operator is “grab a parameter” and the operand tells “where to store” it. Using names is of course

⁷ I actually did when I wrote an interpreter for some computer assisted learning system, think of a kind of interpreted Pascal, but later realized that it was a a bytecode plus virtual machine thing. I'd just applied what I learned when playing with eight bit processors that took bytes, and interpreted opcodes and such. There's nothing spectacular about all this and I only realized decades later that the buzzwords describes old natural concepts.

an option but then one has to do more parsing and turn the name into a number⁸, add additional checking in the macro body, figure out some way to retain the name for the purpose of reporting (which then uses more token memory or strings). It is simply not worth the trouble, let alone the fact that we loose performance, and when T_EX showed up those things really mattered.

It is also important to realize that a # becomes either a preamble token (grab an argument) or a reference token (inject the passed tokens into a new input level). Therefore the duplication of hash tokens ## that you see in macro nested bodies also makes sense: it makes it possible for the parser to distinguish between levels. Take:

```
\def\foo#1{\def\oof##1{#1##1#1}}
```

Of course one can think of this:

```
\def\foo#fence{\def\oof#text{#fence#text#fence}}
```

But such names really have to be unique then! Actually ConT_EXt does have an input method that supports such names, but discussing it here is a bit out of scope. Now, imagine that in the above case we use this:

```
\def\foo[#1][#2]{\def\oof##1{#1##1#2}}
```

If you're a bit familiar with the fact that T_EX has a model of category codes you can imagine that a predictable “hash followed by a number” is way more robust than enforcing the user to ensure that catcodes of ‘names’ are in the right category (read: is a bracket part of the name or not). So, say that we go completely arbitrary names, we then suddenly needs some escaping, like:

```
\def\foo[#{left}][#{right}]{\def\oof#{text}#{left}#{text}#{right}}
```

And, if you ever looked into macro packages, you will notice that they differ in the way they assign category codes. Asking users to take that into account when defining macros makes not that much sense.

So, before one complains about T_EX being obscure (the hash thing), think twice. Your demand for simplicity for your coding demand will make coding more cumbersome for the complex cases that macro packages have to deal with. It's comparable using T_EX for input or using (say) mark down. For simple documents the later is fine, but when things become complex, you end up with similar complexity (or even worse because you lost

⁸ This is kind of what MetaPost does with parameters to macros. The side effect is that in reporting you get text0, expr2 and such reported which doesn't make things more clear.

the enforced detailed structure). So, just accept the unavoidable: any language has its peculiar properties (and for sure I do know why I dislike some languages for it). The \TeX system is not the only one where dollars, percent signs, ampersands and hashes have special meaning.

6.6 Prefixes

Traditional \TeX has three prefixes that can be used with macros: `\global`, `\outer` and `\long`. The last two are no-op's in $\text{LuaMeta}\TeX$ and if you want to know what they do (did) you can look it up in the \TeX book. The ε - \TeX extension gave us `\protected`.

In $\text{LuaMeta}\TeX$ we have `\global`, `\protected`, `\tolerant` and overload related prefixes like `\frozen`. A protected macro is one that doesn't expand in an expandable context, so for instance inside an `\edef`. You can force expansion by using the `\expand` primitive in front which is also something $\text{LuaMeta}\TeX$.

Frozen macros cannot be redefined without some effort. This feature can to some extent be used to prevent a user from overloading, but it also makes it harder for the macro package itself to redefine on the fly. You can remove the lock with `\unletfrozen` and add a lock with `\letfrozen` so in the end users still have all the freedoms that \TeX normally provides.

```

                \def\foo{foo} 1: \meaning\foo
        \frozen\def\foo{foo} 2: \meaning\foo
    \unletfrozen  \foo        3: \meaning\foo
\protected\frozen\def\foo{foo} 4: \meaning\foo
    \unletfrozen  \foo        5: \meaning\foo

```

```

1: macro:foo
2: macro:foo
3: macro:foo
4: protected macro:foo
5: protected macro:foo

```

This actually only works when you have set `\overloadmode` to a value that permits redefining a frozen macro, so for the purpose of this example we set it to zero.

A `\tolerant` macro is one that will quit scanning arguments when a delimiter cannot be matched. We saw examples of that in a previous section.

These prefixes can be chained (in arbitrary order):

```
\frozen\tolerant\protected\global\def\foo[#1]#*[#2]{...}
```

There is actually an additional prefix, `\immediate` but that one is there as signal for a macro that is defined in and handled by Lua. This prefix can then perform the same function as the one in traditional `TEX`, where it is used for backend related tasks like `\write`.

Now, the question is of course, to what extent will `ConTEXt` use these new features. One important argument in favor of using `\tolerant` is that it gives (hopefully) better error messages. It also needs less code due to lack of indirectness. Using `\frozen` adds some safeguards although in some places where `ConTEXt` itself overloads commands, we need to defrost. Adapting the code is a tedious process and it can introduce errors due to mistypings, although these can easily be fixed. So, it will be used but it will take a while to adapt the code base.

One problem with frozen macros is that they don't play nice with for instance `\futurelet`. Also, there are places in `ConTEXt` where we actually do redefine some core macro that we also want to protect from redefinition by a user. One can of course `\unletfrozen` such a command first but as a bonus we have a prefix `\overloaded` that can be used as prefix. So, one can easily redefine a frozen macro but it takes a little effort. After all, this feature is mainly meant to protect a user for side effects of definitions, and not as final blocker.⁹

A frozen macro can still be overloaded, so what if we want to prevent that? For this we have the `\permanent` prefix. Internally we also create primitives but we don't have a prefix for that. But we do have one for a very special case which we demonstrate with an example:

```
\def\F00 % trickery needed to pick up an optional argument
  {\noalign{\vskip10pt}}

\noaligned\protected\tolerant\def\00F[#1]%
  {\noalign{\vskip\iftok{#1}\emptytoks10pt\else#1\fi}}

\starttabulate[|l|l|]
  \NC test \NC test \NC \NR
  \NC test \NC test \NC \NR
  \F00
  \NC test \NC test \NC \NR
  \00F[30pt]
```

⁹ As usual adding features like this takes some experimenting and we're now at the third variant of the implementation, so we're getting there. The fact that we can apply such features in large macro package like `ConTEXt` helps figuring out the needs and best approaches.

```

\NC test \NC test \NC \NR
\OOF
\NC test \NC test \NC \NR
\stoptabulate

```

When $\text{T}_{\text{E}}\text{X}$ scans input (from a file or token list) and starts an alignment, it will pick up rows. When a row is finished it will look ahead for a `\noalign` and it expands the next token. However, when that token is protected, the scanner will not see a `\noalign` in that macro so it will likely start complaining when that next macro does get expanded and produces a `\noalign` when a cell is built. The `\noaligned` prefix flags a macro as being one that will do some `\noalign` as part of its expansion. This trick permits clean macros that pick up arguments. Of course it can be done with traditional means but this whole exercise is about making the code look nice.

The table comes out as:

```

test test
test test

test test

test test

test test

```

One can check the flags with `\ifflags` which takes a control sequence and a number, where valid numbers are:

1 frozen	2 permanent	4 immutable	8 primitive
16 mutable	32 noaligned	64 instance	

The level of checking is controlled with the `\overloadmode` but I'm still not sure about how many levels we need there. A zero value disables checking, the values 1 and 3 give warnings and the values 2 and 4 trigger an error.

6.7 Arguments

The number of arguments that a macro takes is traditionally limited to nine (or ten if one takes the trailing `#` into account). That this is enough for most cases is demonstrated by the fact that $\text{ConT}_{\text{E}}\text{Xt}$ has only a handful of macros that use `#9`. The reason for this limitation is in part a side effect of the way the macro preamble and arguments are parsed. However, because in $\text{LuaMetaT}_{\text{E}}\text{X}$ we use a different implementation, it was not

csname construction. The gain is less than one expects although there are a few cases in ConT_EXt where extreme usage of parameters benefits from it. Users are unlikely to use these two primitives.

Another example of a constant usage is this:

```
\letnothing\foo
```

which gives `\foo` an empty body. That one is used in the core, if only because it gives a bit smaller code. Performance is not that different from

```
\let\foo\empty
```

but it saves one token (8 bytes) when used in a macro. The assignment itself is not that different because `\foo` is made an alias to `\empty` which in turn only needs incrementing a reference counter.

6.8 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

7 Grouping

low level

TEX

grouping

Contents

7.1	Introduction	106
7.2	Pascal	106
7.3	T _E X	106
7.4	MetaPost	107
7.5	Lua	108
7.6	C	108

7.1 Introduction

This is a rather short explanation. I decided to write it after presenting the other topics at the 2019 ConT_EXt meeting where there was a question about grouping.

7.2 Pascal

In a language like Pascal, the language that T_EX has been written in, or Modula, its successor, there is no concept of grouping like in T_EX. But we can find keywords that suggests this:

```
for i := 1 to 10 do begin ... end
```

This language probably inspired some of the syntax of T_EX and MetaPost. For instance an assignment in MetaPost uses := too. However, the begin and end don't really group but define a block of statements. You can have local variables in a procedure or function but the block is just a way to pack a sequence of statements.

7.3 T_EX

In T_EX macros (or source code) the following can occur:

```
\begingroup
  ...
\endgroup
```

as well as:

```
\bgroup
  ...
\egroup
```

Here we really group in the sense that assignments to variables inside a group are forgotten afterwards. All assignments are local to the group unless they are explicitly done global:

```
\scratchcounter=1
\def\foo{foo}
\begingroup
  \scratchcounter=2
  \global\globalscratchcounter=2
  \gdef\foo{F00}
\endgroup
```

Here `\scratchcounter` is still one after the group is left but its global counterpart is now two. The `\foo` macro is also changed globally.

Although you can use both sets of commands to group, you cannot mix them, so this will trigger an error:

```
\bgroup
\endgroup
```

The bottomline is: if you want a value to persist after the group, you need to explicitly change its value globally. This makes a lot of sense in the perspective of $\text{T}_{\text{E}}\text{X}$.

7.4 MetaPost

The MetaPost language also has a concept of grouping but in this case it's more like a programming language.

```
begingroup ;
  n := 123 ;
engroup ;
```

In this case the value of `n` is 123 after the group is left, unless you do this (for numerics there is actually no need to declare them):

```
begingroup ;
  save n ; numeric n ; n := 123 ;
engroup ;
```

Given the use of MetaPost (read: MetaFont) this makes a lot of sense: often you use macros to simplify code and you do want variables to change. Grouping in this language

serves other purposes, like hiding what is between these commands and let the last expression become the result. In a `vardef` grouping is implicit.

So, in MetaPost all assignments are global, unless a variable is explicitly saved inside a group.

7.5 Lua

In Lua all assignments are global unless a variable is defines local:

```
local x = 1
local y = 1
for i = 1, 10 do
    local x = 2
    y = 2
end
```

Here the value of `x` after the loop is still one but `y` is now two. As in Lua_{TeX} we mix _{TeX}, MetaPost and Lua you can mix up these concepts. Another mixup is using `:=`, `endfor`, `fi` in Lua after done some MetaPost coding or using `end` instead of `endfor` in MetaPost which can make the library wait for more without triggering an error. Proper syntax highlighting in an editor clearly helps.

7.6 C

The Lua language is a mix between Pascal (which is one reason why I like it) and C.

```
int x = 1 ;
int y = 1 ;
for (i=1; i<=10;i++) {
    int x = 2 ;
    y = 2 ;
}
```

The semicolon is also used in Pascal but there it is a separator and not a statement end, while in MetaPost it does end a statement (expression).

7.6 Colofon

Author Hans Hagen
ConT_EXt 2023.04.27 17:04
LuaMetaT_EX 2.1008
Support www.pragma-ade.com
 contextgarden.net

8 Security

low level

TEX

security

Contents

8.1	Preamble	111
8.2	Flags	111
8.3	Complications	114
8.4	Introspection	115

8.1 Preamble

Here I will discuss a moderate security subsystem of LuaMetaT_EX and therefore ConT_EXt LMTX. This is not about security in the sense of the typesetting machinery doing harm to your environment, but more about making sure that a user doesn't change the behavior of the macro package in ways that introduce interference and thereby unwanted side effect. It's all about protecting macros.

This is all very experimental and we need to adapt the ConT_EXt source code to this. Actually that will happen a few times because experiments trigger that. It might take a few years before the security model is finalized and all files are updated accordingly. There are lots of files and macros involved. In the process the underlying features in the engine might evolve.

8.2 Flags

Before we go into the security levels we see what flags can be set. The T_EX language has a couple of so called prefixes that can be used when setting values and defining macros. Any engine that has traditional T_EX with ε -T_EX extensions can do this:

```

                \def\foo{foo}
\global        \def\foo{foo}
\global\protected\def\foo{foo}

```

And LuaMetaT_EX adds another one:

```

                \tolerant      \def\foo{foo}
\global\tolerant      \def\foo{foo}
\global\tolerant\protected\def\foo{foo}

```

What these prefixes do is discussed elsewhere. For now it is enough to know that the two optional prefixes `\protected` and `\tolerant` make for four distinctive cases of macro calls.

But there are more prefixes:

frozen	a macro that has to be redefined in a managed way
permanent	a macro that had better not be redefined
primitive	a primitive that normally will not be adapted
immutable	a macro or quantity that cannot be changed, it is a constant
mutable	a macro that can be changed no matter how well protected it is
instance	a macro marked as (for instance) be generated by an interface
noaligned	the macro becomes acceptable as <code>\noalign</code> alias
overloaded	when permitted the flags will be adapted
enforced	all is permitted (but only in zero mode or ini mode)
aliased	the macro gets the same flags as the original

These prefixed set flags to the command at hand which can be a macro but basically any control sequence.

To what extent the engine will complain when a property is changed in a way that violates the above depends on the parameter `\overloadmode`. When this parameter is set to zero no checking takes place. More interesting are values larger than zero. If that is the case, when a control sequence is flagged as mutable, it is always permitted to change. When it is set to immutable one can never change it. The other flags determine the kind of checking done. Currently the following overload values are used:

		immutable	permanent	primitive	frozen	instance
1	warning	*	*	*		
2	error	*	*	*		
3	warning	*	*	*	*	
4	error	*	*	*	*	
5	warning	*	*	*	*	*
6	error	*	*	*	*	*

The even values (except zero) will abort the run. In `ConTeXt` we plug in a callback that deals with the messages. A value of 255 will freeze this parameter. At level five and above the instance flag is also checked but no drastic action takes place. We use this to signal to the user that a specific instance is redefined (of course the definition macros can check for that too).

So, how does it work. The following is okay:

```
\def\MacroA{A}
\def\MacroB{B}
```

```
\let\MyMacro\MacroA
\let\MyMacro\MacroB
```

The first two macros are ordinary ones, and the last two lines just create an alias. Such an alias shares the definition, but when for instance `\MacroA` is redefined, its new meaning will not be reflected in the alias.

```
\permanent\protected\def\MacroA{A}
\permanent\protected\def\MacroB{B}
\let\MyMacro\MacroA
\let\MyMacro\MacroB
```

This also works, because the `\let` will create an alias with the protected property but it will not take the permanent property along. For that we need to say:

```
\permanent\protected\def\MacroA{A}
\permanent\protected\def\MacroB{B}
\permanent\let\MyMacro\MacroA
\permanent\let\MyMacro\MacroB
```

or, when we want to copy all properties:

```
\permanent\protected\def\MacroA{A}
\permanent\protected\def\MacroB{B}
\aliased\let\MyMacro\MacroA
\aliased\let\MyMacro\MacroB
```

However, in `ConTEXt` we have commands that we like to protect against overloading but at the same time have a different meaning depending on the use case. An example is the `\NC` (next column) command that has a different implementation in each of the table mechanisms.

```
\permanent\protected\def\NC_in_table {...}
\permanent\protected\def\NC_in_tabulate{...}
\aliased\let\NC\NC_in_table
\aliased\let\NC\NC_in_tabulate
```

Here the second aliasing of `\NC` fails (assuming of course that we enabled overload checking). One can argue that grouping can be used but often no grouping takes place when we redefine on the fly. Because `frozen` is less restrictive than `primitive` or `permanent`, and of course `immutable`, the next variant works:

```
\frozen\protected\def\NC_in_table {...}
```

```
\frozen\protected\def\NC_in_tabulate{...}
\overloaded\let\NC\NC_in_table
\overloaded\let\NC\NC_in_tabulate
```

However, in practice, as we want to keep the overload checking, we have to do:

```
\frozen\protected\def\NC_in_table  {...}
\frozen\protected\def\NC_in_tabulate{...}
\overloaded\frozen\let\NC\NC_in_table
\overloaded\frozen\let\NC\NC_in_tabulate
```

or use `\aliased`, but there might be conflicting permissions. This is not that nice, so there is a kind of dirty trick possible. Consider this:

```
\frozen\protected\def\NC_in_table  {...}
\frozen\protected\def\NC_in_tabulate{...}
\def\setNCintable  {\enforced\let\frozen\let\NC\NC_in_table}
\def\setNCintabulate{\enforced\let\frozen\let\NC\NC_in_tabulate}
```

When we're in so called `initex` mode or when the overload mode is zero, the `\enforced` prefix is internalized in a way that signals that the follow up is not limited by the overload mode and permissions. This definition time binding mechanism makes it possible to use permanent macros that users cannot redefine, but existing macros can, unless of course they tweak the mode parameter.

Now keep in mind that users can always cheat but that is intentional. If you really want to avoid that you can set the overload mode to 255 after which it cannot be set any more. However, it can be useful to set the mode to zero (or some warning level) when foreign macro packages are used.

8.3 Complications

One side effect of all this is that all those prefixes can lead to more code. On the other hand we save some due to the extended macro argument handling features. When you take the size of the format file as reference, in the end we get a somewhat smaller file. Every token that you add or remove gives a 8 bytes difference. The extra overhead that got added to the engine is compensated by the fact that some macro implementations can be more efficient. In the end, in spite of these new features and the more extensive testing of flags performance is about the same.¹⁰

¹⁰ And if you wonder about memory, by compacting the used (often scattered) token memory before dumping I manages to save some 512K on the format file, so often the loss and gain are somewhere else.

8.4 Introspection

In case you want to get some details about the properties of a macro, you can check its meaning. The full variant shows all of them.

`% a macro with two optional arguments with optional spacing in between:`

```
\permanent\tolerant\protected\def\MyFoo[#1]#*[#2]{(#1)(#2)}
```

```
\meaningless\MyFoo\par
```

```
\meaning \MyFoo\par
```

```
\meaningfull\MyFoo\par
```

```
[#1]#*[#2]->(#1)(#2)
```

```
tolerant protected macro:[#1]#*[#2]->(#1)(#2)
```

```
permanent tolerant protected macro:[#1]#*[#2]->(#1)(#2)
```

8.4 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

9 Characters

low level

TEX

characters

Contents

9.1	Introduction	117
9.2	History	117
9.3	The heritage	118
9.4	The LMTX approach	119

9.1 Introduction

This explanation is part of the low level manuals because in practice users will not have to deal with these matters in MkIV and even less in LMTX. You can skip to the last section for commands.

9.2 History

If we travel back in time to when $\text{T}_{\text{E}}\text{X}$ was written we end up in eight bit character universe. In fact, the first versions assumed seven bits, but for comfortable use with languages other than English that was not sufficient. Support for eight bits permits the usage of so called code pages as supported by operating systems. Although ascii input became kind of the standard soon afterwards, the engine can be set up for different encodings. This is not only true for $\text{T}_{\text{E}}\text{X}$, but for many of its companions, like MetaFont and therefore MetaPost.¹¹

Core components of a $\text{T}_{\text{E}}\text{X}$ engine are hyphenation of words, applying inter-character kerns and build ligatures. In traditional $\text{T}_{\text{E}}\text{X}$ engines those processes are interwoven into the par builder but in Lua $\text{T}_{\text{E}}\text{X}$ these are separate stages. The original approach is the reason that there is a relation between the input encoding and the font encoding: the character in the input is the slot used in a reference to a glyph. When producing the final result (e.g. pdf) there can also be a mapping to an index in a font resource.

```
input A [tex ->] font slot A [backend ->] glyph index A
```

The mapping that $\text{T}_{\text{E}}\text{X}$ does is normally one-to-one but an input character can undergo some transformation. For instance a character beyond ascii 126 can be made active and expand to some character number that then becomes the font slot. So, it is the expansion (or meaning) of a character that end up as numeric reference in the glyph

¹¹ This remapping to an internal representation (e.g. ebcdic) is not present in Lua $\text{T}_{\text{E}}\text{X}$ where we assume utf8 to be the input encoding. The MetaPost library that comes with Lua $\text{T}_{\text{E}}\text{X}$ still has that code but in LuaMeta $\text{T}_{\text{E}}\text{X}$ it's gone. There one can set up the machinery to be utf8 aware too.

node. Virtual fonts can introduce yet another remapping but that's only visible in the backend.

Actually, in Lua \TeX the same happens but in practice there is no need to go active because (at least in Con \TeX t) we assume a Unicode path so there the font slot is the Unicode got from the utf8 input.

In the eight bit universe macro packages (have to) provide all kind of means to deal with (in the perspective of English) special characters. For instance, `\"a` would put a diaeresis on top of the a or even better, refer to a character in the encoding that the chosen font provides. Because there are some limitations of what can go in an eight bit font, and because in different countries the used \TeX fonts evolved kind of independent, we ended up with quite some different variants of fonts. It was only with the Latin Modern project that this became better. Interesting is that when we consider the fact that such a font has often also hardly used symbols (like registered or copyright) coming up with an encoding vector that covers most (latin based) European languages (scripts) is not impossible¹² Special symbols could simply go into a dedicated font, also because these are always accessed via a macro so who cares about the input. It never happened.

Keep in mind that when utf8 is used with eight bit engines, Con \TeX t will convert sequences of characters into a slot in a font (depending on the font encoding used which itself depends on the coverage needed). For this every first (possible) byte of a multi-byte utf sequence is an active character, which is no big deal because these are outside the ascii range. Normal ascii characters are single byte utf sequences and fall through without treatment.

Anyway, in Con \TeX t MkII we dealt with this by supporting mixed encodings, depending on the (local) language, referencing the relevant font. It permits users to enter the text in their preferred input encoding and also get the words properly hyphenated. But we can leave these MkII details behind.

9.3 The heritage

In MkIV we got rid of input and font encodings, although one can still load files in a specific code page.¹³ We also kept the means to enter special characters, if only because text editors seldom support(ed) a wide range of visual editing of those. This is why we still have

¹² And indeed in the Latin Modern project we came up with one but it was already to late for it to become popular.

¹³ I'm not sure if users ever depend on an input encoding different from utf8.

```
\u \^a \v{s} \AE \ij \eacute \oslash
```

and many more. The ones with one character names are rather common in the T_EX community but it is definitely a weird mix of symbols. The next two are kind of outdated: in these days you delegate that to the font handler, where turning them into ‘single’ character references depends on what the font offers, how it is set up with respect to (for instance) ligatures, and even might depend on language or script.

The ones with the long names partly are tradition, but as we have a lot of them, in MkII they actually serve a purpose. These verbose names are used in the so called encoding vectors and are part of the utf expansion vectors. They are also used in labels so that we have a good indication if what goes in there: remember that in those times editors often didn't show characters, unless the font for display had them, or the operating system somehow provided them from another font. These verbose names are used for latin, greek and cyrillic and for some other scripts and symbols. They take up quite a bit of hash space and the format file.¹⁴

9.4 The LMTX approach

In the process of tagging all (public) macros in LMTX (which happened in 2020-2021) I wondered if we should keep these one character macros, the references to special characters and the verbose ones. When asked on the mailing list it became clear that users still expect the short ones to be present, often just because old bibT_EX files are used that might need them. However, in MkIV and LMTX we load bibT_EX files in a way that turn these special character references into proper utf8 input so it makes a weak argument. Anyway, although they could go, for now we keep them because users expect them. However, in LMTX the implementation is somewhat different now, a bit more efficient in terms of hash and memory, potentially a bit less efficient in runtime, but no one will notice that.

A new command has been introduced, the very short `\chr`.

```
\chr {a`} \chr {a´} \chr {a¨}
\chr {`a} \chr {'a} \chr {"a}
\chr {a acute} \chr {a grave} \chr {a umlaut}
\chr {aacute} \chr {agrave} \chr {aumlaut}
```

In the first line the composed character using two characters, a base and a so called mark. Actually, one doesn't have to use `\chr` in that case because ConT_EXt does already

¹⁴ In MkII we have an abstract front-end with respect to encodings and also an abstract backend with respect to supported drivers but both approaches no longer make sense today.

collapse characters for you. The second line looks like the shortcuts `\``, `\'` and `\"`. The third and fourth lines could eventually replace the more symbolic long names, if we feel the need. Watch out: in Unicode input the marks come *after*.

```
à á ä
à á ä
á à a~mla~t
á à a~mla~t
```

Currently the repertoire is somewhat limited but it can be easily be extended. It all depends on user needs (doing Greek and Cyrillic for instance). The reason why we actually save code deep down is that the helpers for this have always been there.¹⁵

The `\` commands are now just aliases to more verbose and less hackery looking macros:

```
\withgrave      à  \`  à
\withacute      á  \'  á
\withcircumflex â  \^  â
\withtilde      ã  \~  ã
\withmacron     ā  \=  ā
\withbreve      ě  \u  ě
\withdotaccent  ċ  \.  .c
\withdiaeresis  ë  \"  ë
\withring       ũ  \r  ũ
\withhungarumlaut  ű  \H  ű
\withcaron      ě  \v  ě
\withcedilla    ç  \c  ç
\withogonek     ę  \k  ę
```

Not all fonts have these special characters. Most natural is to have them available as precomposed single glyphs, but it can be that they are just two shapes with the marks anchored to the base. It can even be that the font somehow overlays them, assuming (roughly) equal widths. The compose font feature in ConT_EXt normally can handle most well.

An occasional ugly rendering doesn't matter that much: better have something than nothing. But when it's the main language (script) that needs them you'd better look for a font that handles them. When in doubt, in ConT_EXt you can enable checking:

¹⁵ So if needed I can port this approach back to MkIV, but for now we keep it as is because we then have a reference.

command	equivalent to
<code>\checkmissingcharacters</code>	<code>\enabletrackers[fonts.missing]</code>
<code>\removemissingcharacters</code>	<code>\enabletrackers[fonts.missing=remove]</code>
<code>\replacemissingcharacters</code>	<code>\enabletrackers[fonts.missing=replace]</code>
<code>\handlemissingcharacters</code>	<code>\enabletrackers[fonts.missing={decompose,replace}]</code>

The decompose variant will try to turn a composed character into its components so that at least you get something. If that fails it will inject a replacement symbol that stands out so that you can check it. The console also mentions missing glyphs. You don't need to enable this by default¹⁶ but you might occasionally do it when you use a font for the first time.

In LMTX this mechanism has been upgraded so that replacements follow the shape and are actually real characters. The decomposition has not yet been ported back to MkIV.

The full list of commands can be queried when a tracing module is loaded:

```
\usemodule[s][characters-combinations]
```

```
\showcharactercombinations
```

We get this list:

acute	U+00301	´	<code>\withacute</code>
breve	U+00306	˘	<code>\withbreve</code>
caron	U+0030C	ˇ	<code>\withcaron</code>
caron below	U+0032C	ˇ̣	<code>\withcaronbelow</code>
cedilla	U+00327	¸	<code>\withcedilla</code>
circumflex	U+00302	ˆ	<code>\withcircumflex</code>
circumflex below	U+0032D	ˆ̣	<code>\withcircumflexbelow</code>
comma below	U+00327	¸	<code>\withcommabelow</code>
diaeresis	U+00308	¨	<code>\withdiaeresis</code>
dieresis	U+00308	¨	<code>\withdieresis</code>
dot	U+00307	·	<code>\withdot</code>
dot below	U+00323	·̣	<code>\withdotbelow</code>
double acute	U+0030B	˚	<code>\withdoubleacute</code>
double grave	U+0030F	˘̣	<code>\withdoublegrave</code>
double vertical line	U+0030E	="	<code>\withdoubleverticalline</code>
grave	U+00300	˘	<code>\withgrave</code>
hook	U+00309	ˆ̣	<code>\withhook</code>

¹⁶ There is some overhead involved here.

hook below	U+1FA9D		\withhookbelow
hungarumlaut	U+0030B	“	\withhungarumlaut
inverted breve	U+00311	^	\withinvertedbreve
line	U+00304	-	\withline
line below	U+00331	_	\withlinebelow
macron	U+00304	-	\withmacron
macron below	U+00331	_	\withmacronbelow
middle dot	U+000B7	·	\withmiddledot
ogonek	U+00328	˘	\withogonek
overline	U+00305	—	
ring	U+0030A	°	\withring
ring below	U+00325	◦	\withringbelow
slash	U+0002F	/	\withslash
stroke	U+0002F	/	\withstroke
tilde	U+00303	~	\withtilde
tilde below	U+00330	˜	\withtildebelow
vertical line	U+0030D		\withverticalline

Some combinations are special for ConT_EXt because Unicode doesn't specify decomposition for all composed characters.

9.4 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

10 Scope

low level

TEX

scope

Contents

10.1 Introduction	124
10.2 Registers	124
10.3 Allocation	126
10.4 Files	129

10.1 Introduction

When I visited the file where register allocations are implemented I wondered to what extent it made sense to limit allocation to global instances only. This chapter deals with this phenomena.

10.2 Registers

In \TeX definitions can be local or global. Most assignments are local within a group. Registers and definitions can be assigned global by using the `\global` prefix. There are also some properties that are global by design, like `\prevdepth`. A mixed breed are boxes. When you tweak its dimensions you actually tweak the current box, which can be an outer level. Compare:

```
\scratchcounter = 1
here the counter has value 1
\begingroup
  \scratchcounter = 2
  here the counter has value 2
\endgroup
here the counter has value 1
```

with:

```
\setbox\scratchbox=\hbox{}
here the box has zero width
\begingroup
  \wd\scratchbox=10pt
  here the box is 10pt wide
\endgroup
here the box is 10pt wide
```

It all makes sense so a remark like “Assignments to box dimensions are always global” are sort of confusing. Just look at this:

```

\setbox\scratchbox=\hbox to 20pt{}
here the box is \the\wd\scratchbox\ wide\par
\begingroup
  \setbox\scratchbox=\hbox{}
  here the box is \the\wd\scratchbox\ wide\par
  \begingroup
    \wd\scratchbox=15pt
    here the box is \the\wd\scratchbox\ wide\par
  \endgroup
  here the box is \the\wd\scratchbox\ wide\par
\endgroup
here the box is \the\wd\scratchbox\ wide\par

here the box is 20.0pt wide
here the box is 0.0pt wide
here the box is 15.0pt wide
here the box is 15.0pt wide
here the box is 20.0pt wide

```

If you don't think about it, what happens is what you expect. Now watch the next variant:

The `\global` is only effective for the current box. It is good to realize that when we talk registers, the box register behaves just like any other register but the manipulations happen to the current one.

```

\setbox\scratchbox=\hbox to 20pt{}
here the box is \the\wd\scratchbox\ wide\par
\begingroup
  \setbox\scratchbox=\hbox{}
  here the box is \the\wd\scratchbox\ wide\par
  \begingroup
    \global\wd\scratchbox=15pt
    here the box is \the\wd\scratchbox\ wide\par
  \endgroup
  here the box is \the\wd\scratchbox\ wide\par
\endgroup
here the box is \the\wd\scratchbox\ wide\par

here the box is 20.0pt wide
here the box is 0.0pt wide
here the box is 15.0pt wide

```

here the box is 15.0pt wide

here the box is 20.0pt wide

```
\scratchdimen=20pt
```

here the dimension is `\the\scratchdimen\par`

```
\begingroup
```

```
\scratchdimen=0pt
```

here the dimension is `\the\scratchdimen\par`

```
\begingroup
```

```
\global\scratchdimen=15pt
```

here the dimension is `\the\scratchdimen\par`

```
\endgroup
```

here the dimension is `\the\scratchdimen\par`

```
\endgroup
```

here the dimension is `\the\scratchdimen\par`

here the dimension is 20.0pt

here the dimension is 0.0pt

here the dimension is 15.0pt

here the dimension is 15.0pt

here the dimension is 15.0pt

10.3 Allocation

The plain \TeX format has set some standards and one of them is that registers are allocated with `\new...` commands. So we can say:

```
\newcount\mycounta
```

```
\newdimen\mydimena
```

These commands take a register from the pool and relate the given name to that entry. In $\text{Con}\text{\TeX}$ t we have a bunch of predefined scratch registers for general use, like:

```
scratchcounter      : \meaningfull\scratchcounter
scratchcounterone  : \meaningfull\scratchcounterone
scratchcountertwo  : \meaningfull\scratchcountertwo
scratchdimen       : \meaningfull\scratchdimen
scratchdimenone    : \meaningfull\scratchdimenone
scratchdimentwo    : \meaningfull\scratchdimentwo
```

The meaning reveals what these are:

Allocation

```

scratchcounter : integer 1026
scratchcounterone : integer 0
scratchcountertwo : integer 0
scratchdimen : dimension 15.0pt
scratchdimenone : dimension 0.0pt
scratchdimentwo : dimension 0.0pt

```

You can use the numbers directly but that is a bad idea because they can clash! In the original $\text{T}_{\text{E}}\text{X}$ engine there are only 256 registers and some are used by the engine and the core of a macro package itself, so that leaves a little amount for users. The $\varepsilon\text{-T}_{\text{E}}\text{X}$ extension lifted that limitation and bumped to 32K and $\text{LuaT}_{\text{E}}\text{X}$ upped that to 64K. One could go higher but what makes sense? These registers are taking part of the fixed memory slots because that makes nested (grouped) usage efficient and access fast. The number you see above is deduced from the so called command code (here indicated by `\count`) and an index encoded in the same token. So, `\scratchcounter` takes a single token contrary to the verbose `\count257` that takes four tokens where the number gets parsed every time it is needed. But those are details that a user can forget.

As mentioned, commands like `\newcount\foo` create a global control sequence `\foo` referencing a counter. You can locally redefine that control sequence unless in $\text{LuaMeta-}\text{T}_{\text{E}}\text{X}$ you have so called overload mode enabled. You can do local or global assignments to these registers.

```

\scratchcounter = 123
\begingroup
  \scratchcounter = 456
  \begingroup
    \global\scratchcounter = 789
  \endgroup
\endgroup

```

And in both cases count register 257 is set. When an assignment is global, all current values to that register get the same value. Normally this is all quite transparent: you get what you ask for. However the drawback is that as a user you cannot know what variables are already defined, which means that this will fail (that is: it will issue a message):

```
\newcount\scratchcounter
```

as will the second line in:

```
\newcount\myscratchcounter
```

```
\newcount\myscratchcounter
```

In ConT_EXt the scratch registers are visible but there are lots of internally used ones are protected from the user by more obscure names. So what if you want to use your own register names without ConT_EXt barking to you about not being able to define it. This is why in LMTX (and maybe some day in MkIV) we now have local definitions:

```
\begingroup
  \newlocaldimen\mydimena    \mydimena1\onpoint
  \newlocaldimen\mydimenb    \mydimenb2\onpoint
  (\the\mydimena,\the\mydimenb)
  \begingroup
    \newlocaldimen\mydimena    \mydimena3\onpoint
    \newlocaldimen\mydimenb    \mydimenb4\onpoint
    \newlocaldimen\mydimenc    \mydimenc5\onpoint
    (\the\mydimena,\the\mydimenb,\the\mydimenc)
    \begingroup
      \newlocaldimen\mydimena    \mydimena6\onpoint
      \newlocaldimen\mydimenb    \mydimenb7\onpoint
      (\the\mydimena,\the\mydimenb)
    \endgroup
    \newlocaldimen\mydimend    \mydimend8\onpoint
    (\the\mydimena,\the\mydimenb,\the\mydimenc,\the\mydimend)
  \endgroup
  (\the\mydimena,\the\mydimenb)
\endgroup
```

The allocated registers get zero values but you can of course set them to any value that fits their nature:

```
(1.0pt,2.0pt)
(3.0pt,4.0pt,5.0pt)
(6.0pt,7.0pt)
(3.0pt,4.0pt,5.0pt,8.0pt)
(1.0pt,2.0pt)
```

You can also use the next variant where you also pass the initial value:

```
\begingroup
  \setnewlocaldimen\mydimena    1\onpoint
  \setnewlocaldimen\mydimenb    2\onpoint
  (\the\mydimena,\the\mydimenb)
```

```

\begingroup
  \setnewlocaldimen\mydimena 3\onepoint
  \setnewlocaldimen\mydimenb 4\onepoint
  \setnewlocaldimen\mydimenc 5\onepoint
  (\the\mydimena,\the\mydimenb,\the\mydimenc)
\begingroup
  \setnewlocaldimen\mydimena 6\onepoint
  \setnewlocaldimen\mydimenb 7\onepoint
  (\the\mydimena,\the\mydimenb)
\endgroup
\setnewlocaldimen\mydimend 8\onepoint
(\the\mydimena,\the\mydimenb,\the\mydimenc,\the\mydimend)
\endgroup
(\the\mydimena,\the\mydimenb)
\endgroup

```

So, again we get:

```

(1.0pt,2.0pt)
(3.0pt,4.0pt,5.0pt)
(6.0pt,7.0pt)
(3.0pt,4.0pt,5.0pt,8.0pt)
(1.0pt,2.0pt)

```

When used in the body of the macro there is of course a little overhead involved in the repetitive allocation but normally that can be neglected.

10.4 Files

When adding these new allocators I also wondered about the read and write allocators. We don't use them in ConT_EXt but maybe users like them, so let's give an example and see what more demands they have:

```

\integerdef\StartHere\numexpr\inputlineno+2\relax
\starthiding
SOME LINE 1
SOME LINE 2
SOME LINE 3
SOME LINE 4
\stophiding
\integerdef\StopHere\numexpr\inputlineno-2\relax

```

```

\begingroup
  \newlocalread\myreada
  \immediate\openin\myreada {lowlevel-scope.tex}
  \dostepwiserecurse{\StopHere}{\StartHere}{-1}{
    \readline\myreada line #1 to \scratchstring #1 : \scratchstring \par
  }
  \blank
  \dostepwiserecurse{\StartHere}{\StopHere}{1}{
    \read    \myreada line #1 to \scratchstring #1 : \scratchstring \par
  }
  \immediate\closein\myreada
\endgroup

```

Here, instead of hard coded line numbers we used the stored values. The optional line keyword is a LMTX speciality.

```

281 : SOME LINE 4
280 : SOME LINE 3
279 : SOME LINE 2
278 : SOME LINE 1

```

```

278 : SOME LINE 1
279 : SOME LINE 2
280 : SOME LINE 3
281 : SOME LINE 4

```

Actually an application can be found in a small (demonstration) module:

```

\usemodule[system-readers]

```

This provides the code for doing this:

```

\startmarkedlines[test]
SOME LINE 1
SOME LINE 2
SOME LINE 3
\stopmarkedlines

```

```

\begingroup
  \newlocalread\myreada
  \immediate\openin\myreada {\markedfilename{test}}
  \dostepwiserecurse{\lastmarkedline{test}}{\firstmarkedline{test}}{-1}{
    \readline\myreada line #1 to \scratchstring #1 : \scratchstring \par
  }

```

```
}  
\immediate\closein\myreada  
\endgroup
```

As you see in these examples, we can locally define a read channel without getting a message about it already being defined.

10.4 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

11 Paragraphs

low level

TEX

paragraphs

Contents

11.1	Introduction	133
11.2	Paragraphs	133
11.3	Properties	137
11.4	Wrapping up	139
11.5	Hanging	140
11.6	Shapes	140
11.7	Modes	158
11.8	Leaders	158
11.9	Prevdepth	165
11.10	Normalization	167
11.11	Dirty tricks	167

11.1 Introduction

This manual is mostly discussing a few low level wrappers around low level $\text{T}_{\text{E}}\text{X}$ features. Its writing is triggered by an update to the MetaFun and LuaMetaFun manuals where we mess a bit with shapes. It gave a good reason to also cover some more paragraph related topics but it might take a while to complete. Remind me if you feel that takes too much time.

Because paragraphs and their construction are rather central to $\text{T}_{\text{E}}\text{X}$, you can imagine that the engine exposes dealing with them. This happens via commands (primitives) but only when it's robust. Then there are callbacks, and some provide detailed information about what we're dealing with. However, intercepting node lists can already be hairy and we do that a lot in $\text{ConT}_{\text{E}}\text{Xt}$. Intercepting and tweaking paragraph properties is even more tricky, which is why we try to avoid that in the core. But ... in the following sections you will see that there are actually a couple of mechanism that do so. Often new features like this are built in stepwise and enabled locally for a while and when they seem okay they get enabled by default.¹⁷

11.2 Paragraphs

Before we demonstrate some trickery, let's see what a paragraph is. Normally a document source is formatted like this:

¹⁷ For this we have `\enableexperiments` which one can use in `cont-loc.mkxl` or `cont-exp.mkxl`, files that are loaded runtime when on the system. When you use them, make sure they don't interfere; they are not part of the updates, contrary to `cont-new.mkxl`.

```
some text (line 1)
some text (line 2)
```

```
some more test (line 1)
some more test (line 2)
```

There are two blocks of text here separated by an empty line and they become two paragraphs. Unless configured otherwise an empty line is an indication that we end a paragraph. You can also explicitly do that:

```
some text (line 1)
some text (line 2)
\par
some more test (line 1)
some more test (line 2)
```

When $\text{T}_{\text{E}}\text{X}$ starts a paragraph, it actually also does something think of:

```
[\the\everypar]some text      (line 1) some text      (line 2) \par
[\the\everypar]some more test (line 1) some more test (line 2) \par
```

or more accurate:

```
[\the\everypar]some text      some text      \par
[\the\everypar]some more test some more test \par
```

because the end-of-line character has become a space. As mentioned, an empty line is actually the end of a paragraph. But in $\text{LuaMetaT}_{\text{E}}\text{X}$ we can cheat a bit. If we have this:

```
line 1
```

```
line 2
```

We can do this (watch how we need to permit overloading a primitive when we have enabled `\overloadmode`):

```
\pushoverloadmode
\def\linepar{\removeunwantedspaces !\ignorespaces}
\popoverloadmode
line 1

line 2
```

This comes out as:

line 1

line 2

I admit that since it got added (as part of some cleanup halfway the overhaul of the engine) I never saw a reason to use it, but it is a cheap feature. The `\linepar` primitive is undefined (`\undefined`) by default so no user sees it anyway. Just don't use it unless maybe for some pseudo database trickery (I considered using it for the database module but it is not needed). In a similar fashion, just don't redefine `\par`: it's asking for troubles and 'not done' in ConT_EXt anyway.

Back to reality. In LuaT_EX we get a node list that starts with a so called `localpar` node and ends with a `\parfillskip`. The first node is prepended automatically. That list travels through the system: hyphenation, applying font properties, break the effectively one line into lines, wrap them and add them to a vertical list, etc. Each stage can be intercepted via callbacks.

When the paragraph is broken into lines hanging indentation or a so called par shape can be applied, and we will see more of that later, here we talk `\par` and show another LuaMetaT_EX trick:

```
\def\foo{{\bf test:} \ignorepars}
```

```
\foo
```

line

The macro typesets some text and then skips to the next paragraph:

test: line

Think of this primitive as being a more powerful variant of `\ignorespaces`. This leaves one aspect: how do we start a paragraph. Technically we need to force T_EX into so called horizontal mode. When you look at plain T_EX documents you will notice commands like `\noindent` and `\indent`. In ConT_EXt we have more high level variants, for instance we have `\noindentation`.

A robust way to make sure that you get in horizontal mode is using `\dontleavehmode` which is a wink to `\leavevmode`, a command that you should never use in ConT_EXt, so when you come from plain or L^AT_EX, it's one of the commands you should wipe from your memory.

When T_EX starts with a paragraph the `\everypar` token list is expanded and again this is a primitive you should not mess with yourself unless in very controlled situations.

If you change its content, you're on your own with respect to interferences and side effects.

One of the things that \TeX does in injecting the indentation. Even when there is none, it gets added, not as skip but as an empty horizontal box of a certain width. This is easier on the engine when it constructs the paragraph from the one liner: starting with a skip demands a bit more testing in the process (a nice trick so to say). However, in $\text{Con}\TeX$ we enable the $\text{LuaMeta}\TeX$ feature that does use a skip instead of a box. It's part of the normalization that is discussed later. Instead of checking for a box with property `indent`, we check for a skip with such property. This is often easier and cleaner.

A bit off topic is the fact that in traditional \TeX empty lines or `\par` primitives can trigger an error. This has to do with the fact that the program evolved in a time where paper terminals were used and runtime could be excessive. So, in order to catch a possible missing brace, a concept of `\long` macros, permitting `\par` or equivalents in arguments, was introduced as well as not permitting them in for instance `display math`. In $\text{Con}\TeX$ MkII most macros that could be sensitive for this were defined as `\long` so that users never had to bother about it and probably were not even aware of it. Right from the start in $\text{Lua}\TeX$ these error-triggers could be disabled which of course we enable in $\text{Con}\TeX$ and in $\text{LuaMeta}\TeX$ these features have been removed altogether. I don't think users will complain about this.

If you want to enforce a newline but not a new paragraph you can use the `\crlf` command. When used on its own it will produce an empty line. Don't use this to create whitespace between lines.

If you want to do something after so called par tokens are seen you can do this:

```
\def\foo{{\bf >>>> }}
\expandafterpars\foo
```

this is a new paragraph ...

```
\expandafterpars\foo
\par\par\par\par
```

this is a new paragraph ...

This not to be confused with `\everypar` which is a token list that \TeX itself injects before each paragraph (also nested ones).

```
>>>> this is a new paragraph ...
```

```
>>>> this is a new paragraph ...
```

This is typically a primitive that will only be used in macros. You can actually program it using macros: pickup a token, check and push it back when it's not a par equivalent token. The primitive is just nicer (and easier on the log when tracing is enabled).

11.3 Properties

A paragraph is just a collection of lines that result from one input line that got broken. This process of breaking into lines is influenced by quite some parameters. In traditional \TeX and also in \LuaMetaTeX by default the values that are in effect when the end of the paragraph is met are used. So, when you change them in a group and then ends the paragraph after the group, the values you've set in the group are not used.

However, in \LuaMetaTeX we can optionally store them with the paragraph. When that happens the values current at the start are frozen. You can still overload them but that has to be done explicitly then. The advantage is that grouping no longer interferes with the line break algorithm. The magic primitive is `\snapshotpar` which takes a number made from categories mentioned below:

variable	category	code
<code>\hsize</code>	hsize	0x01
<code>\leftskip</code>	skip	0x02
<code>\rightskip</code>	skip	0x02
<code>\hangindent</code>	hang	0x04
<code>\hangafter</code>	hang	0x04
<code>\parindent</code>	indent	0x08
<code>\parfillleftskip</code>	par fill	0x10
<code>\parfillrightskip</code>	par fill	0x10
<code>\adjustspacing</code>	adjust	0x20
<code>\adjustspacingstep</code>	adjust	0x20
<code>\adjustspacingshrink</code>	adjust	0x20
<code>\adjustspacingstretch</code>	adjust	0x20
<code>\protrudechars</code>	protrude	0x40
<code>\pretolerance</code>	tolerance	0x80
<code>\tolerance</code>	tolerance	0x80
<code>\emergencystretch</code>	stretch	0x100
<code>\looseness</code>	looseness	0x200
<code>\lastlinefit</code>	last line	0x400
<code>\linepenalty</code>	line penalty	0x800
<code>\interlinepenalty</code>	line penalty	0x800
<code>\interlinepenalties</code>	line penalty	0x800
<code>\clubpenalty</code>	club penalty	0x1000

<code>\clubpenalties</code>	club penalty	0x1000
<code>\widowpenalty</code>	widow penalty	0x2000
<code>\widowpenalties</code>	widow penalty	0x2000
<code>\displaywidowpenalty</code>	display penalty	0x4000
<code>\displaywidowpenalties</code>	display penalty	0x4000
<code>\brokenpenalty</code>	broken penalty	0x8000
<code>\adjdemerits</code>	demerits	0x10000
<code>\doublehyphendemerits</code>	demerits	0x10000
<code>\finalhyphendemerits</code>	demerits	0x10000
<code>\parshape</code>	shape	0x20000
<code>\baselineskip</code>	line	0x40000
<code>\lineskip</code>	line	0x40000
<code>\lineskiplimit</code>	line	0x40000
<code>\hyphenationmode</code>	hyphenation	0x80000

As you can see here, there are more paragraph related parameters than in for instance pdf \TeX and Lua \TeX and these are (to be) explained in the LuaMeta \TeX manual. You can imagine that keeping this around with the paragraph adds some extra overhead to the machinery but most users won't notice that because it is compensated by gains elsewhere.

This is pretty low level and there are a bunch of helpers that support this but these are not really user level macros. As with everything \TeX you can mess around as much as you like, and the code gives plenty of examples but when you do this, you're on your own because it can interfere with Con \TeX t core functionality.

In LMTX taking these snapshots is turned on by default and because it thereby fundamentally influences the par builder, users can run into compatibility issues but in practice there has been no complaints (and this feature has been in use quite a while before this document was written). One reason for users not noticing is that one of the big benefits is probably handled by tricks mentioned on the mailing list. Imagine that you have this:

```
{\bf watch out:} here is some text
```

In this small example the result will be as expected. But what if something magic with the start of a paragraph is done? Like this:

```
\placefigure[left]{A cow!}{\externalfigure[cow.pdf]}
```

```
{\bf watch out:} here is some text ... of course much more is needed to  
get a flow around the figure!
```

Properties

The figure will hang at the left side of the paragraph but it is put there when the text starts and that happens inside the bold group. It means that the properties we set in order to get the shape around the figure are lost as soon as we're at 'here is some text' and definitely is wrong when the paragraph ends and the par builder has to use them to get the shape right. We get text overlapping the figure. A trick to overcome this is:

```
\dontleavehmode {\bf watch out:} here is some text ... of course much
    more is needed to get a flow around the figure!
```

where the first macro makes sure we already start a paragraph before the group is entered (using a `\strut` also works). It's not nice and I bet users have been bitten by this and by now know the tricks. But, with snapshots such fuzzy hacks are not needed any more! The same is true with this:

```
{\leftskip 1em some text \par}
```

where we had to explicitly end the paragraph inside the group in order to retain the skip. I suppose that users normally use the high level environments so they never had to worry about this. It's also why users probably won't notice that this new mechanism has been active for a while. Actually, when you now change a parameter inside the paragraph its new value will not be applied (unless you prefix it with `\frozen` or `snapshot` it) but no one did that anyway.

11.4 Wrapping up

In `ConTEXt LMTX` we have a mechanism to exercise macros (or content) before a paragraph ends. This is implemented using the `\wrapuppar` primitive. The to be wrapped up material is bound to the current paragraph which in order to get this done has to be started when this primitive is used.

Although the high level interface has been around for a while it still needs a bit more testing (read: use cases are needed). In the few cases where we already use it application can be different because again it relates to snapshots. This because in the past we had to use tricks that also influenced the user interface of some macros (which made them less natural as one would expect). So the question is: where do we apply it in old mechanisms and where not.

todo: accumulation, interference, where applied, limitations

11.5 Hanging

There are two mechanisms for getting a specific paragraph shape: rectangular hanging and arbitrary shapes. Both mechanisms work top-down. The first mechanism uses a combination of `\hangafter` and `\hangindent`, and the second one depends on `\parshape`. In this section we discuss the rectangular one.

```
\hangafter 4 \hangindent 4cm \samplefile{tufte} \page
\hangafter -4 \hangindent 4cm \samplefile{tufte} \page
\hangafter 4 \hangindent -4cm \samplefile{tufte} \page
\hangafter -4 \hangindent -4cm \samplefile{tufte} \page
```

As you can see in figure 11.1, the four cases are driven by the sign of the values. If you want to hang into the margin you need to use different tricks, like messing with the `\leftskip`, `\rightskip` or `\parindent` parameters (which then of course can interfere with other mechanisms uses at the same time).

11.6 Shapes

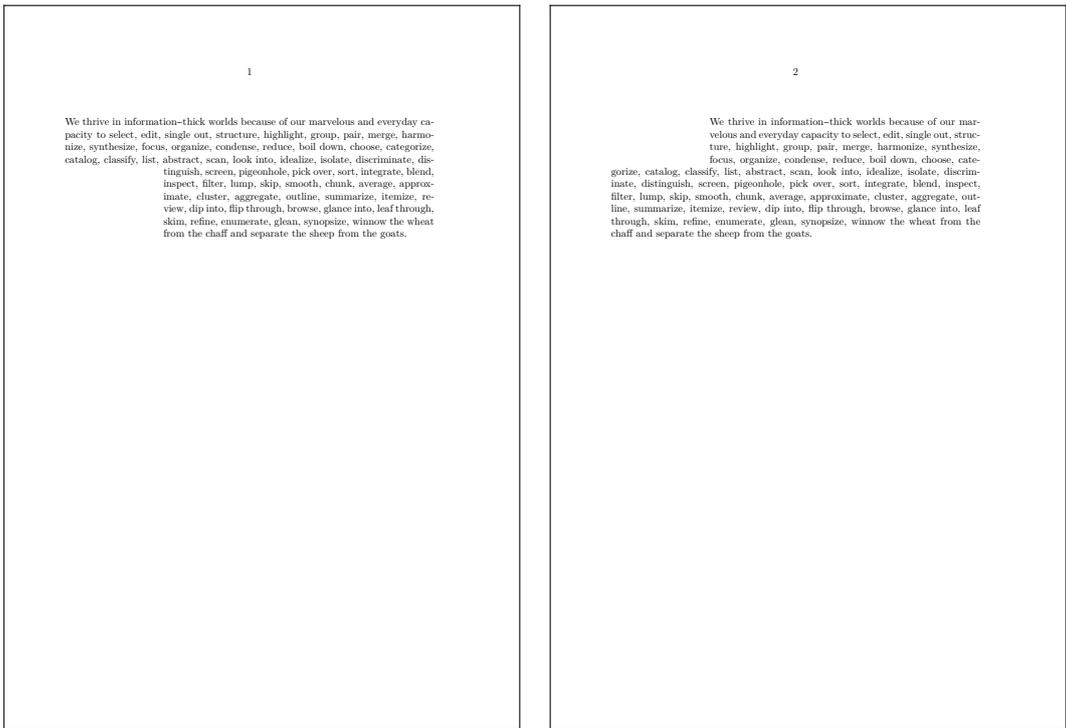
In `ConTEXt` we don't use `\parshape` a lot. It is used in for instance side floats but even there not in all cases. It's more meant for special applications. This means that in `MkII` and `MkIV` we don't have some high level interface. However, when `MetaFun` got upgraded to `LuaMetaFun`, and the manual also needed an update, one of the examples in that manual that used shapes also got done differently (read: nicer). And that triggered the arrival of a new low level shape mechanism.

One important property of the `\parshape` mechanism is that it works per paragraph. You define a shape in terms of a left margin and width of a line. The shape has a fixed number of such pairs and when there is more content, the last one is used for the rest of the lines. When the paragraph is finished, the shape is forgotten.¹⁸

The high level interface is a follow up on the example in the `MetaFun` manual and uses shapes that carry over to the next paragraph. In addition we can cycle over a shape. In this interface shapes are defined using keyword. Here are some examples:

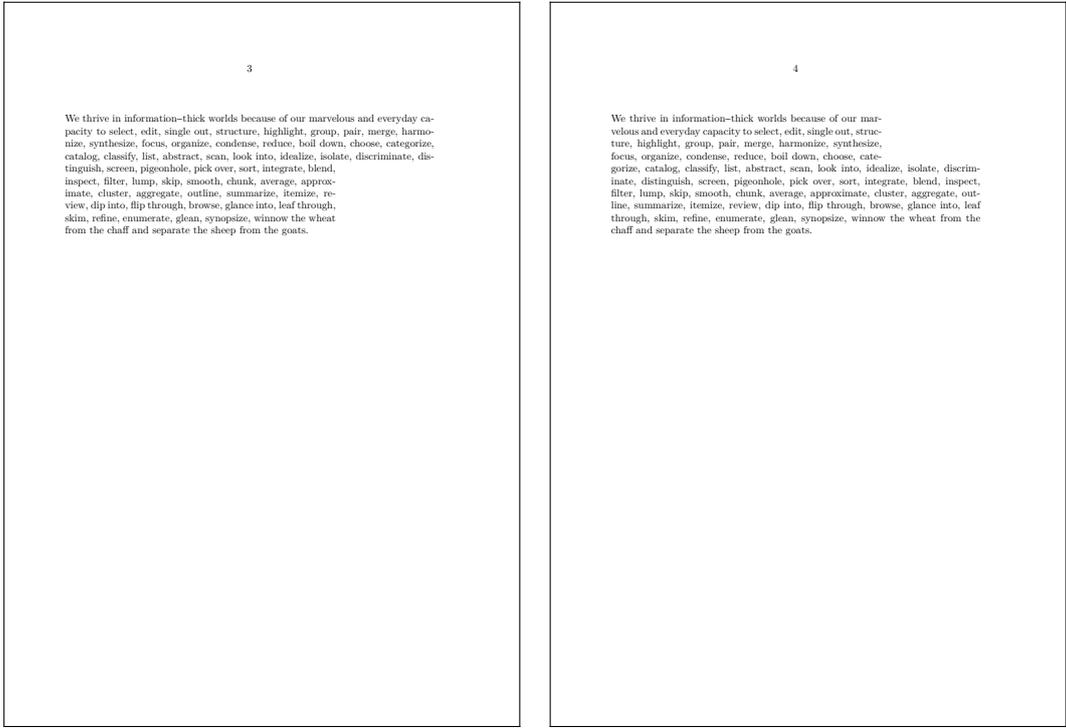
```
\startparagraphshape[test]
  left 1mm right 1mm
  left 5mm right 5mm
```

¹⁸ Not discussed here is a variant that might end up in `LuaMetaTEX` that works with the progression, i.e. takes the height of the content so far into account. This is somewhat tricky because for that to work vertical skips need to be frozen, which is no real big deal but has to be done careful in the code.



\hangafter +4
 \hangindent +4cm

\hangafter -4
 \hangindent +4cm



\hangafter +4
 \hangindent -4cm

\hangafter -4
 \hangindent -4cm

Figure 11.1 Hanging indentation

`\stopparagraphshape`

This shape has only two entries so the first line will have a 1mm margin while later lines will get 5mm margins. This translates into a `\parshape` like:

```
\parshape 2
  1mm \dimexpr\hsize-1mm\relax
  5mm \dimexpr\hsize-5mm\relax
```

Watch the number 2: it tells how many specification lines follow. As you see, we need to calculate the width.

`\startparagraphshape[test]`

```
left 1mm right 1mm
left 5mm right 5mm
repeat
```

`\stopparagraphshape`

This variant will alternate between 1mm and 5mm margins. The repeating feature is translated as follows. Maybe at some point I will introduce a few more options.

```
\parshape 2 options 1
  1mm \dimexpr\hsize-1mm\relax
  5mm \dimexpr\hsize-5mm\relax
```

A shape can have some repetition, and we can save keystrokes by copying the last entry. The resulting `\parshape` becomes rather long.

`\startparagraphshape[test]`

```
left 1mm right 1mm
left 2mm right 2mm
left 3mm right 3mm
copy 8
left 4mm right 4mm
left 5mm right 5mm
left 5mm hsize 10cm
```

`\stopparagraphshape`

Also watch the `hsize` keyword: we don't calculate the `hsize` from the left and right values but explicitly set it.

`\startparagraphshape[test]`

```
left 1mm right 1mm
```

```

right 3mm
left 5mm right 5mm
repeat
\stopparagraphshape

```

When a right keywords comes first the left is assumed to be zero. In the examples that follow we will use a couple of definitions:

```

\startparagraphshape[test]
  both 1mm both 2mm both 3mm both 4mm both 5mm both 6mm
  both 7mm both 6mm both 5mm both 4mm both 3mm both 2mm
\stopparagraphshape

```

```

\startparagraphshape[test-repeat]
  both 1mm both 2mm both 3mm both 4mm both 5mm both 6mm
  both 7mm both 6mm both 5mm both 4mm both 3mm both 2mm
  repeat
\stopparagraphshape

```

The last one could also be defines as:

```

\startparagraphshape[test-repeat]
  \rawparagraphshape{test} repeat
\stopparagraphshape

```

In the previous code we already introduced the repeat option. This will make the shape repeat at the engine level when the shape runs out of specified lines. In the application of a shape definition we can specify a method to be used and that determine if the next paragraph will start where we left off and discard afterwards (shift) or that we move the discarded lines up front so that we never run out of lines (cycle). It sounds complicated but just keep in mind that repeat is part of the \parshape and act within a paragraph while shift and cycle are applied when a new paragraph is started.

In figure 11.2 you see the following applied:

```

\startshapedparagraph[list=test]
  \dorecurse{8}{\showparagraphshape\samplefile{tufte} \par}
\stopshapedparagraph

\startshapedparagraph[list=test-repeat]
  \dorecurse{8}{\showparagraphshape\samplefile{tufte} \par}
\stopshapedparagraph

```



```
\startshapedparagraph[list=test,method=shift]
  \dorecurse{8}{\showparagraphshape\samplefile{tufte} \par}
\stopshapedparagraph
```

Finally, in figure 11.4 we use:

```
\startshapedparagraph[list=test,method=cycle]
  \dorecurse{8}{\showparagraphshape\samplefile{tufte} \par}
\stopshapedparagraph
```

These examples are probably too small to see the details but you can run them yourself or zoom in on the details. In the margin we show the values used. Here is a simple example of (non) poetry. There are other environments that can be used instead but this makes a good example anyway.

```
\startparagraphshape[test]
  left 0em right 0em
  left 1em right 0em
  repeat
\stopparagraphshape

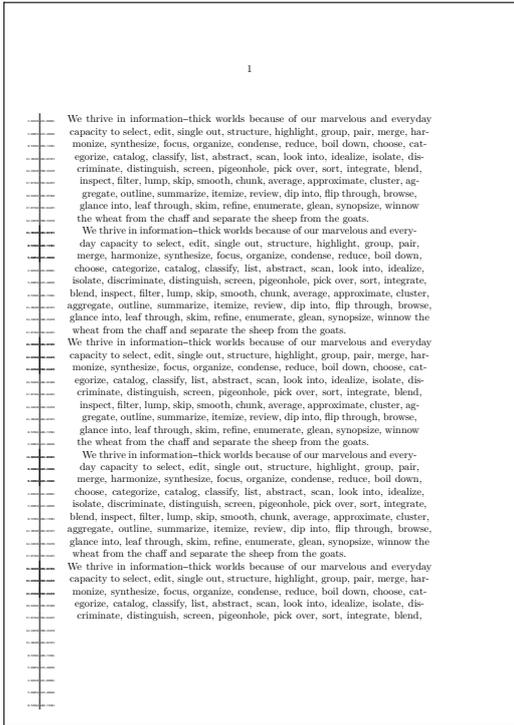
\startshapedparagraph[list=test,method=cycle]
  verse line 1.1\crlf verse line 2.1\crlf
  verse line 3.1\crlf verse line 4.1\par
  verse line 1.2\crlf verse line 2.2\crlf
  verse line 3.2\crlf verse line 4.2\crlf
  verse line 5.2\crlf verse line 6.2\par
\stopshapedparagraph
```

```
verse line 1.1
  verse line 2.1
verse line 3.1
  verse line 4.1

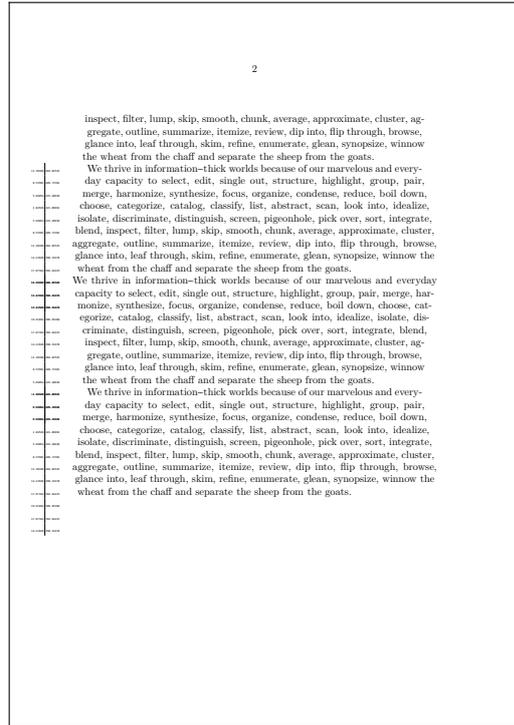
verse line 1.2
  verse line 2.2
verse line 3.2
  verse line 4.2
verse line 5.2
  verse line 6.2
```

Because the idea for this feature originates in MetaFun, we will now kick in some Meta-Post. The following code creates a shape for a circle. We use a 2mm offset here:

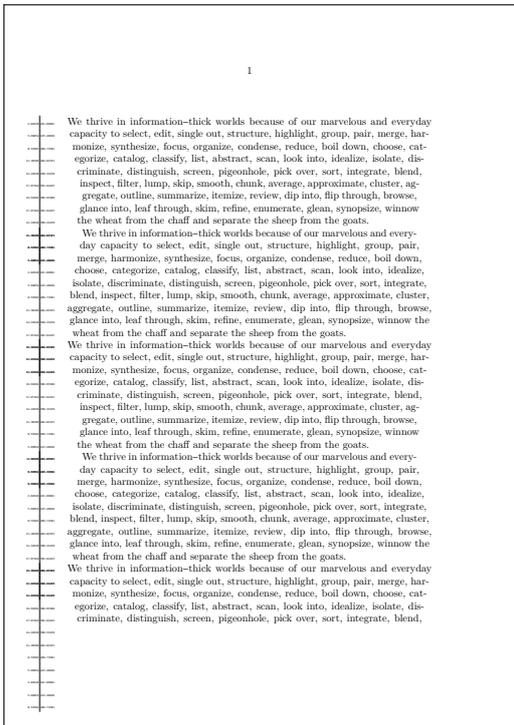
Shapes



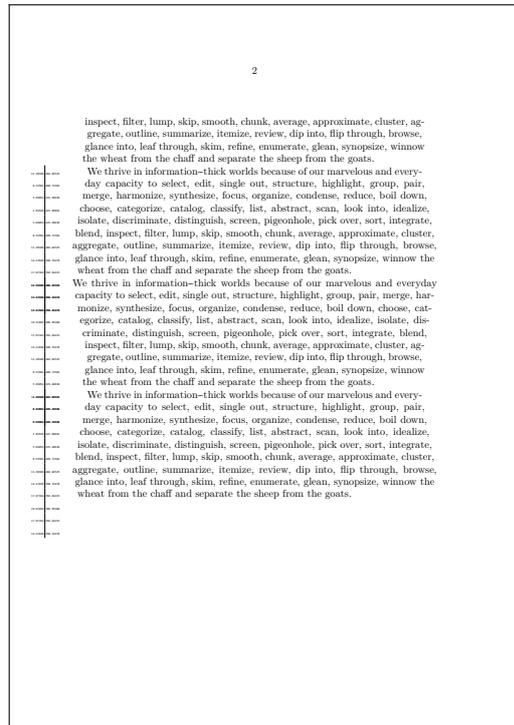
cycle, finite shape, page 1



cycle, finite shape, page 2



cycle, repeat in shape, page 1



cycle, repeat in shape, page 2

Figure 11.4 Cycled shaping

```

\startuseMPgraphic{circle}
  path p ; p := fullcircle scaled TextWidth ;
  build_parshape(p,
    2mm, 0, 0,
    LineHeight, StrutHeight, StrutDepth, StrutHeight
  ) ;
\stopuseMPgraphic

```

We plug this into the already described macros:

```

\startshapedparagraph[mp=circle]%
  \setupalign[verytolerant,stretch,last]%
  \samplefile{tufte}
  \samplefile{tufte}
\stopshapedparagraph

```

And get ourself a circular shape. Watch out, at this moment the shape environment does not add grouping so when as in this case you change the alignment it can influence the document.

We thrive in information-thick
worlds because of our marvelous and every-
day capacity to select, edit, single out, structure,
highlight, group, pair, merge, harmonize, synthesize, focus,
organize, condense, reduce, boil down, choose, categorize, cat-
alog, classify, list, abstract, scan, look into, idealize, isolate, discrimi-
nate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, in-
spect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggre-
gate, outline, summarize, itemize, review, dip into, flip through, browse, glance
into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the wheat
from the chaff and separate the sheep from the goats. We thrive in information-
thick worlds because of our marvelous and everyday capacity to select, edit, single
out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize,
condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan,
look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort,
integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, clus-
ter, aggregate, outline, summarize, itemize, review, dip into, flip through, browse,
glance into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the
wheat from the chaff and separate the sheep from the goats.

Assuming that the shape definition above is in a buffer we can do this:

```

\startshapedparagraph[mp=circle]%

```

```

\setupalign[verytolerant,stretch,last]%
\samplefile{tufte}
\samplefile{tufte}
\stopshapedparagraph

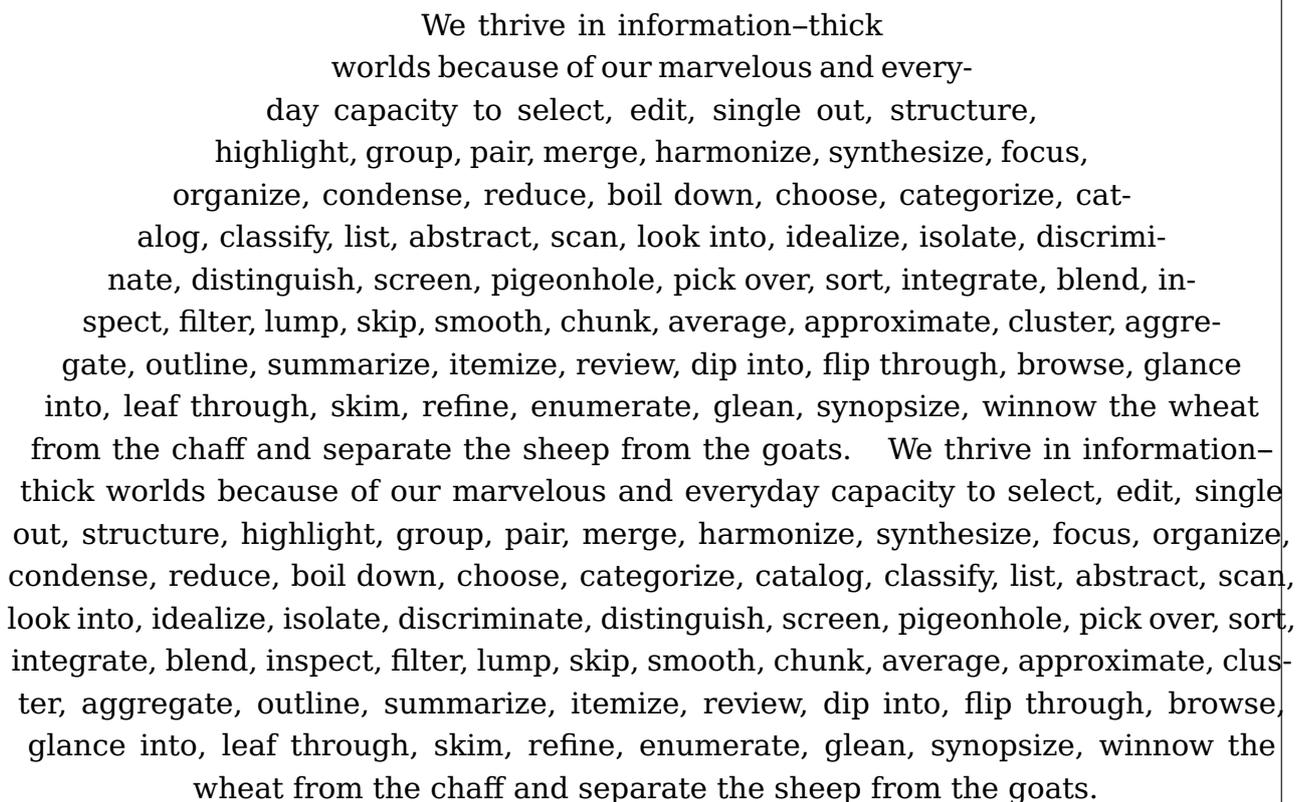
```

The result is shown in figure 11.5. Because all action happens in the framed environment, we can also use this definition:

```

\startuseMPgraphic{circle}
  path p ; p := fullcircle scaled \the\dimexpr\framedwidth+\framedoffset
    *2\relax ;
  build_parshape(p,
    \framedoffset, 0, 0,
    LineHeight, StrutHeight, StrutDepth, StrutHeight
  ) ;
  draw p ;
\stopuseMPgraphic

```



We thrive in information-thick
 worlds because of our marvelous and every-
 day capacity to select, edit, single out, structure,
 highlight, group, pair, merge, harmonize, synthesize, focus,
 organize, condense, reduce, boil down, choose, categorize, cat-
 alog, classify, list, abstract, scan, look into, idealize, isolate, discrimi-
 nate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, in-
 spect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggre-
 gate, outline, summarize, itemize, review, dip into, flip through, browse, glance
 into, leaf through, skim, refine, enumerate, glean, synthesize, winnow the wheat
 from the chaff and separate the sheep from the goats. We thrive in information-
 thick worlds because of our marvelous and everyday capacity to select, edit, single
 out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize,
 condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan,
 look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort,
 integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, clus-
 ter, aggregate, outline, summarize, itemize, review, dip into, flip through, browse,
 glance into, leaf through, skim, refine, enumerate, glean, synthesize, winnow the
 wheat from the chaff and separate the sheep from the goats.

Figure 11.5 A framed circular shape

A mechanism like this is often never completely automatic in the sense that you need to keep an eye on the results. Depending on user demands more features can be added.

With weird shapes you might want to set up the alignment to be tolerant and have some stretch.

The interface described in the MetaFun manual is pretty old, the time stamp of the original code is mid 2000, but the principles didn't change. The examples in `meta-imp-txt.mkx` can now be written as:

```
\startshapetext[test 1,test 2,test 3,test 4]
  \setupalign[verytolerant,stretch,normal]%
  \samplefile{douglas} % Douglas R. Hofstadter
\stopshapetext
\startcombination[2*2]
  {\framed[offset=overlay,frame=off,background=test 1]{\getshapetext}}
  {test 1}
  {\framed[offset=overlay,frame=off,background=test 2]{\getshapetext}}
  {test 2}
  {\framed[offset=overlay,frame=off,background=test 3]{\getshapetext}}
  {test 3}
  {\framed[offset=overlay,frame=off,background=test 4]{\getshapetext}}
  {test 4}
\stopcombination
```

In figure 11.6 we see the result. Watch how for two shapes we have enabled tracing. Of course you need to tweak till all fits well but we're talking of special situations anyway.

Here is a bit more extreme example. Again we use a circle:

```
\startuseMPgraphic{circle}
  lmt_parshape [
    path      = fullcircle scaled 136mm,
    offset    = 2mm,
    bottomskip = - 1.5LineHeight,
  ] ;
\stopuseMPgraphic
```

But we output a longer text:

```
\startshapedparagraph[mp=circle,repeat=yes,method=cycle]%
  \setupalign[verytolerant,stretch,last]\dontcomplain
  {\darkred      \samplefile{tufte}}\par
  {\darkgreen    \samplefile{tufte}}\par
  {\darkblue     \samplefile{tufte}}\par
```

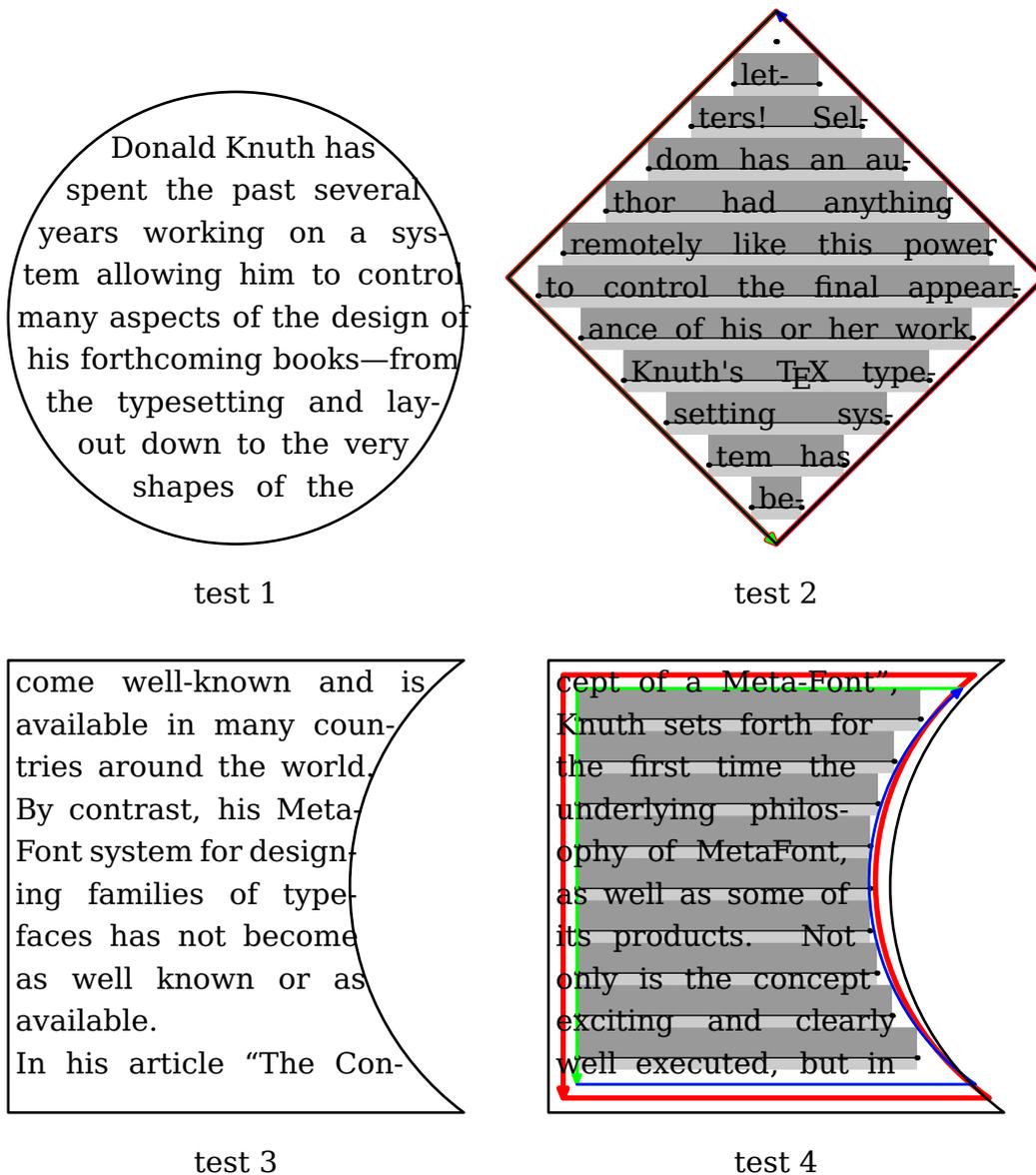


Figure 11.6

```

{\darkcyan \samplefile{tufte}}\par
{\darkmagenta \samplefile{tufte}}\par
\stopshapedparagraph

```

We get a multi-page shape:

We thrive in information-thick
worlds because of our marvelous and every-
day capacity to select, edit, single out, structure,
highlight, group, pair, merge, harmonize, synthesize, focus,
organize, condense, reduce, boil down, choose, categorize, cat-

alog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the wheat from the chaff and separate the sheep from the goats.

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the wheat from the chaff and separate the sheep from the goats.

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the wheat from the chaff and separate the sheep from the goats.

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsisize, winnow the wheat from the chaff and separate the sheep from the goats.

We thrive in information-thick worlds because of our marvelous and everyday capacity

to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate the sheep from the goats.

Compare this with:

```
\startshapedparagraph[mp=circle,repeat=yes,method=cycle]%
  \setupalign[verytolerant,stretch,last]\dontcomplain
  {\darkred      \samplefile{tufte}}
  {\darkgreen    \samplefile{tufte}}
  {\darkblue     \samplefile{tufte}}
  {\darkcyan     \samplefile{tufte}}
  {\darkmagenta  \samplefile{tufte}}
\stopshapedparagraph
```

Which gives:

We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate the sheep from the goats. We thrive in information-thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate the sheep from the goats. We thrive in information-thick worlds because of our marvelous and

everyday capacity to select, edit, single out, structure, highlight, group, pair,
 merge, harmonize, synthesize, focus, organize, condense, reduce, boil
 down, choose, categorize, catalog, classify, list, abstract, scan, look
 into, idealize, isolate, discriminate, distinguish, screen, pigeon-
 hole, pick over, sort, integrate, blend, inspect, filter, lump,
 skip, smooth, chunk, average, approximate, cluster,
 aggregate, outline, summarize, itemize, re-
 view, dip into, flip through, browse,
 glance into, leaf
 through, skim, refine, enumer-
 ate, glean, synopsise, winnow the wheat
 from the chaff and separate the sheep from the
 goats. We thrive in information-thick worlds because of
 our marvelous and everyday capacity to select, edit, single out,
 structure, highlight, group, pair, merge, harmonize, synthesize, fo-
 cus, organize, condense, reduce, boil down, choose, categorize, catalog,
 classify, list, abstract, scan, look into, idealize, isolate, discrimi-
 nate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump,
 skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summa-
 rize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim,
 refine, enumerate, glean, synopsise, winnow the wheat from the chaff and separate
 the sheep from the goats. We thrive in information-thick worlds because of our mar-
 velous and everyday capacity to select, edit, single out, structure, highlight, group, pair,
 merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose,
 categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discrimi-
 nate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter,
 lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summa-
 rize, itemize, review, dip into, flip through, browse, glance into, leaf through,
 skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff and
 separate the sheep from the goats.

Here the `bottomskip` takes care of subtle rounding issues as well as discarding the last line in the shape so that we get nicer continuation. There is no full automated solution for all you can come up with.

Mixing a MetaPost specification into a regular one is also possible. The next example demonstrates this as well as the option to remove some lines from a specification:

```

\startparagaphshape[test]
  left 0em right 0em
  left 1em right 0em

```

```

metapost {circle}
delete 3
metapost {circle,circle,circle}
delete 7
metapost {circle}
repeat
\stopparagraphshape

```

You can combine a shape with narrowing a paragraph. Watch the `absolute` keyword in the next code. The result is shown in figure 11.7.

```

\startuseMPgraphic{circle}
  \lmt_parshape [
    path      = fullcircle scaled TextWidth,
    bottomskip = - 1.5LineHeight,
  ] ;
\stopuseMPgraphic

\startparagraphshape[test-1]
  metapost {circle} repeat
\stopparagraphshape

\startparagraphshape[test-2]
  absolute left metapost {circle} repeat
\stopparagraphshape

\startparagraphshape[test-3]
  absolute right metapost {circle} repeat
\stopparagraphshape

\startparagraphshape[test-4]
  absolute both metapost {circle} repeat
\stopparagraphshape

\showframe

\startnarrower[4*left,2*right]
  \startshapedparagraph[list=test-1,repeat=yes,method=repeat]%
    \setupalign[verytolerant,stretch,last]\dontcomplain
    \dorecurse{3}{\samplefile{thuan}}
  \stopshapedparagraph
  \page
  \startshapedparagraph[list=test-2,repeat=yes,method=repeat]%

```

```

    \setupalign[verytolerant,stretch,last]\dontcomplain
    \dorecurse{3}{\samplefile{thuan}}
\stopshapedparagraph
\page
\startshapedparagraph[list=test-3,repeat=yes,method=repeat]%
    \setupalign[verytolerant,stretch,last]\dontcomplain
    \dorecurse{3}{\samplefile{thuan}}
\stopshapedparagraph
\page
\startshapedparagraph[list=test-4,repeat=yes,method=repeat]%
    \setupalign[verytolerant,stretch,last]\dontcomplain
    \dorecurse{3}{\samplefile{thuan}}
\stopshapedparagraph
\stopnarrower

```

The shape mechanism has a few more tricks but these are really meant for usage in specific situations, where one knows what one deals with. The following examples are visualized in figure 11.8.

```

\useMPlibrary[dum]
\usemodule[article-basics]

\startbuffer
    \externalfigure[dummy][width=6cm]
\stopbuffer

\startshapedparagraph[text=\getbuffer]
    \dorecurse{3}{\samplefile{ward}\par}
\stopshapedparagraph

\page

\startshapedparagraph[text=\getbuffer,distance=1em]
    \dorecurse{3}{\samplefile{ward}\par}
\stopshapedparagraph

\page

\startshapedparagraph[text=\getbuffer,distance=1em,
    hoffset=-2em]
    \dorecurse{3}{\samplefile{ward}\par}

```



```
\stopshapedparagraph
```

```
\page
```

```
\startshapedparagraph[text=\getbuffer,distance=1em,
  voffset=-2ex,hoffset=-2em]
  \dorecurse{3}{\samplefile{ward}\par}
\stopshapedparagraph
```

```
\page
```

```
\startshapedparagraph[text=\getbuffer,distance=1em,
  voffset=-2ex,hoffset=-2em,lines=1]
  \dorecurse{3}{\samplefile{ward}\par}
\stopshapedparagraph
```

```
\page
```

```
\startshapedparagraph[width=4cm,lines=4]
  \dorecurse{3}{\samplefile{ward}\par}
\stopshapedparagraph
```

11.7 Modes

todo: some of the side effects of so called modes

11.8 Leaders

Leaders are a basic feature that users probably never run into directly. They repeat content till it fits the specified width which can be stretched out. The content is typeset once and it is the backend that does the real work of repetition.

```
\strut\leaders \hbox{!}\hfill\strut
\strut\xleaders\hbox{!}\hfill\strut
\strut\cleaders\hbox{!}\hfill\strut
\strut\gleaders\hbox{!}\hfill\strut
```

Here `\leaders` starts at the left edge and are repeats the box as long as it fits, `\xleaders` spreads till the edges and `\cleaders` centers the lot. The `\gleaders` primitive (which is not in original T_EX) takes the outer box as reference and further behaves like `\cleaders`.

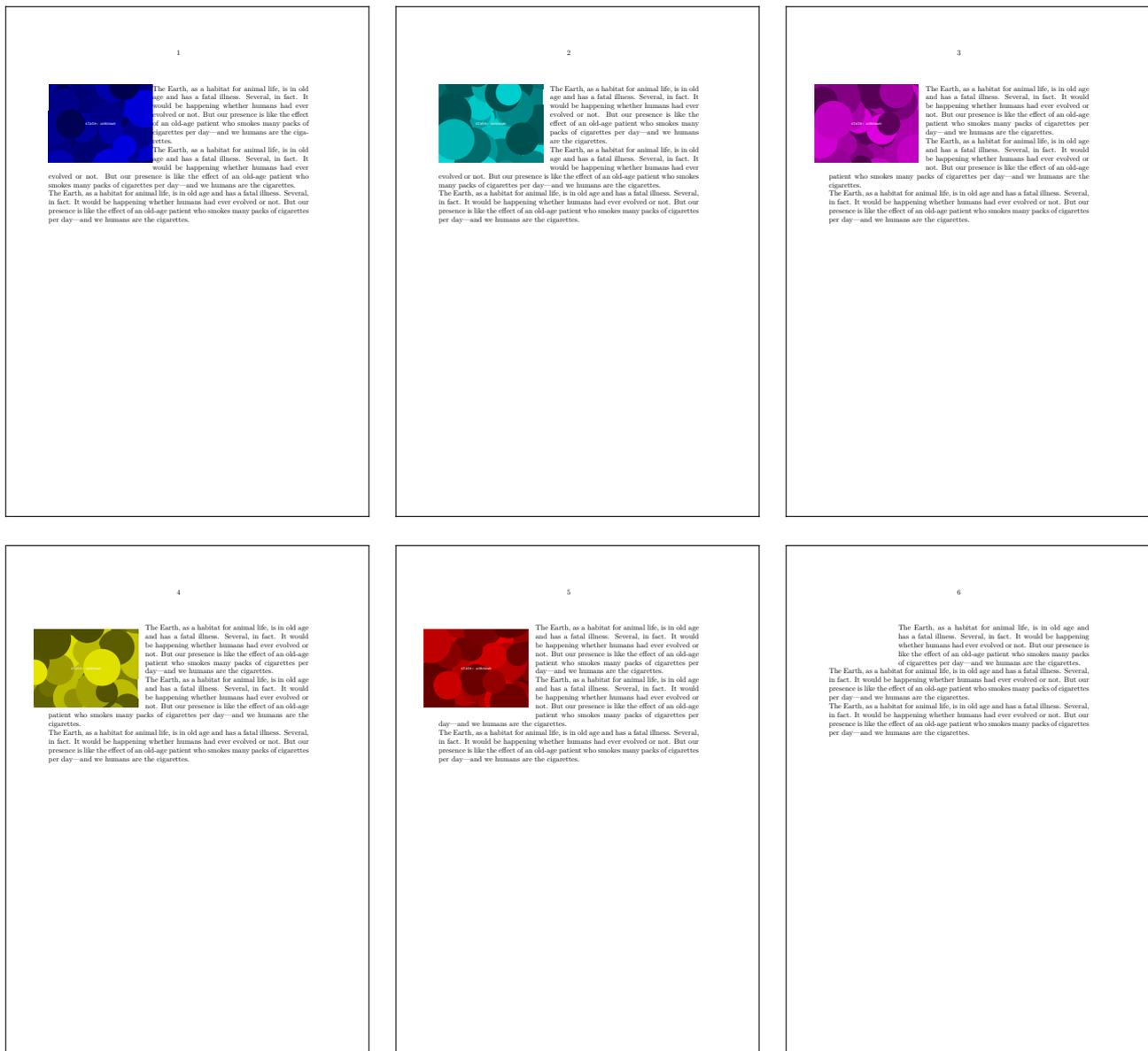


Figure 11.8 Flow around something



The leader primitives take box or rule but in LuaMetaTeX a glyph can also be specified, which saves wrapping in a box.

```
\ruledvbox \bgroup \hsize 10cm
\strut\clleaders\hbox{!}\hfill\strut
\egroup
```

```
\ruledvbox \bgroup \hsize 10cm
  \strut\cleaders\hrule\hfill\strut
\egroup
```

```
\ruledvbox \bgroup \hsize 10cm
  \strut\cleaders\glyph`!\hfill\strut
\egroup
```



The LuaMetaTeX engine also introduced `\uleaders`

We show three boxes, a regular one first (red):

```
x xx xxx xxxx
\ruledhbox{L\hss R}\space
x xx xxx xxxx
```

The second one (blue) is also a box but one that stretches upto 100pt and is in a later stage, when the paragraph has been built, is repackaged to the effective width. The third example (green) leaves out the background.

```
x xx xxx xxxx LR x xx xxx xxxx x xx xxx xxxx LR x xx xxx xxxx x xx xxx xxxx
LR xxx xxx xxx xxx xxx LR x xx xxx xxxx x xx xxx xxxx LR x xx xxx xxxx
xx xxx xxx LR x xx xxx xxxx x xx xxx xxxx LR x xx xxx xxxx x xx xxx xxxx
x xx xxx xxx LR xxx xxx
xx xxx xxx LR x xx xxx xxxx x xx xxx xxxx LR xxx xxx LR xxx xxx LR xxx xxx
xxx xxx LR xxx xxx xxxx LR xxx xxx LR xxx xxx LR xxx xxx LR xxx xxx LR xxx xxx
```

In ConTeXt we have wrapped this feature in the adaptive box mechanism, so here a few a few examples:

```
\startsetups adaptive:test:a
  \setbox\usedadaptivebox\vbox to \usedadaptivetotal \bgroup
  \externalfigure
  [cow.pdf]
  [width=\framedmaxwidth,
  frame=on,
  height=\usedadaptivetotal]%
\egroup
\stopsetups
```

```

\startsetups adaptive:test:b
  \setbox\usedadaptivebox\vbox to \usedadaptivetotal \bgroup
  \externalfigure
    [cow.pdf]
    [width=\usedadaptivewidth,
     frame=on,
     height=\usedadaptivetotal]%
  \egroup
\stopsetups

```

We use this as follows (see figure 11.9 for the result):

```

\framed[height=18cm,align=middle,adaptive=yes,top=,bottom=] {%
  \begstrut \samplefile{tufte} \endstrut
  \par
  \adaptivevbox
    [strut=yes,setup=adaptive:test:a]
    {\showstruts\strut\hsize5cm\hss}%
  \par
  \adaptivevbox
    [strut=yes,setup=adaptive:test:b]
    {\showstruts\strut\hsize5cm\hss}%
  \par
  \begstrut \samplefile{tufte} \endstrut
}

```

Here is one that you can test yourself:

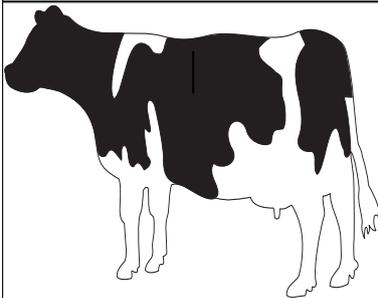
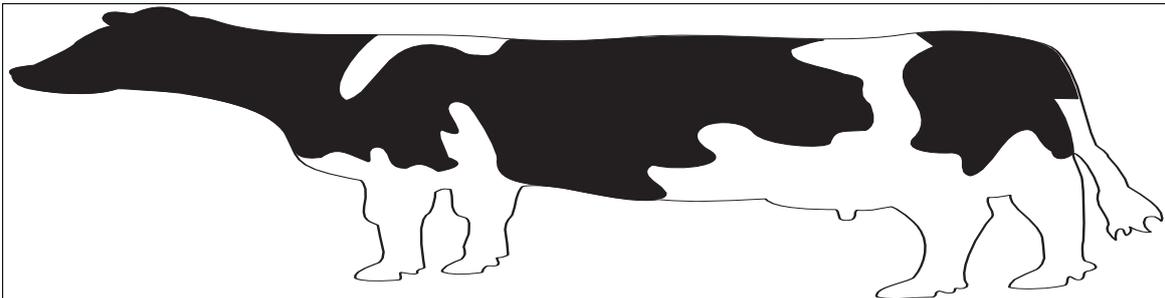
```

\startsetups adaptive:test
  \setbox\usedadaptivebox\vbox to \usedadaptivetotal \bgroup
  \externalfigure
    [cow.pdf]
    [width=\usedadaptivewidth,
     height=\usedadaptivetotal]%
  \egroup
\stopsetups

\ruledvbox to \textheight {
  \par \begstrut \samplefile{tufte} \endstrut \par
  \adaptivevbox[strut=yes,setup=adaptive:test]{\hsize\textwidth\hss}
  \par \begstrut \samplefile{tufte} \endstrut
}

```

We thrive in information–thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synthesize, winnow the wheat from the chaff and separate the sheep from the goats.



We thrive in information–thick worlds because of our marvelous and everyday capacity to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeonhole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, refine, enumerate, glean, synthesize, winnow the wheat from the chaff and separate the sheep from the goats.

Figure 11.9

The next example comes from the test suite (where it runs over many pages in order to illustrate the idea):

```
\startMPdefinitions
def TickTock =
  interim linecap := squared;
```

```

save p ; path p ;
p := fullsquare xysized(AdaptiveWidth,.9(AdaptiveHeight+AdaptiveDepth))
;
fill p withcolor AdaptiveColor ;
draw bottomboundary (p enlarged (-AdaptiveThickness) )
    withdashes (3*AdaptiveThickness)
    withpen pencircle scaled AdaptiveThickness
    withcolor white ;
enddef ;
\stopMPdefinitions

\startsetups adaptive:test
  \setbox\usedadaptivebox\hbox
    to      \usedadaptivewidth
    yoffset -.9\usedadaptivedepth
  \bgroup
    \hss
    \startMPcode
      TickTock ;
    \stopMPcode
    \hss
  \egroup
\stopsetups

\definecolor[adaptive:tick][.25(blue,green)]
\definecolor[adaptive:tock][.75(blue,green)]

\defineadaptive
  [tick]
  [setups=adaptive:test,
  color=adaptive:tick,
  foregroundcolor=white,
  foregroundstyle=\infofont,
  strut=yes]

\defineadaptive
  [tock]
  [tick]
  [color=adaptive:tock]

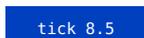
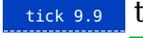
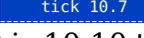
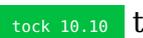
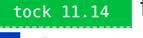
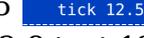
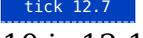
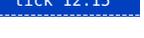
\dostepwiserecurse{8}{12}{1}{%

```

```

\dostepwiserecurse{5}{15}{1}{%
  this~#1.##1 is~#1.##1 test~#1.##1
  \ifodd##1\relax
    \adaptivebox[tick]{\hss tick #1.##1\hss}
  \else
    \adaptivebox[tock]{\hss tock #1.##1\hss}
  \fi
}
}

```

this 8.5 is 8.5 test 8.5  this 8.6 is 8.6 test 8.6  this 8.7 is 8.7 test 8.7
 this 8.8 is 8.8 test 8.8  this 8.9 is 8.9 test 8.9  this 8.10 is 8.10 test 8.10
 this 8.11 is 8.11 test 8.11  this 8.12 is 8.12 test 8.12  this 8.13 is 8.13
 test 8.13  this 8.14 is 8.14 test 8.14  this 8.15 is 8.15 test 8.15
 this 9.5 is 9.5 test 9.5  this 9.6 is 9.6 test 9.6  this 9.7 is 9.7 test 9.7
 this 9.8 is 9.8 test 9.8  this 9.9 is 9.9 test 9.9  this 9.10 is 9.10 test 9.10
 this 9.11 is 9.11 test 9.11  this 9.12 is 9.12 test 9.12  this 9.13 is 9.13
 test 9.13  this 9.14 is 9.14 test 9.14  this 9.15 is 9.15 test 9.15
 this 10.5 is 10.5 test 10.5  this 10.6 is 10.6 test 10.6  this 10.7 is 10.7
 test 10.7  this 10.8 is 10.8 test 10.8  this 10.9 is 10.9 test 10.9
 this 10.10 is 10.10 test 10.10  this 10.11 is 10.11 test 10.11  this 10.12
 is 10.12 test 10.12  this 10.13 is 10.13 test 10.13  this 10.14 is 10.14
 test 10.14  this 10.15 is 10.15 test 10.15  this 11.5 is 11.5 test 11.5
 this 11.6 is 11.6 test 11.6  this 11.7 is 11.7 test 11.7  this 11.8 is 11.8
 test 11.8  this 11.9 is 11.9 test 11.9  this 11.10 is 11.10 test 11.10
 this 11.11 is 11.11 test 11.11  this 11.12 is 11.12 test 11.12  this 11.13
 is 11.13 test 11.13  this 11.14 is 11.14 test 11.14  this 11.15 is 11.15
 test 11.15  this 12.5 is 12.5 test 12.5  this 12.6 is 12.6 test 12.6
 this 12.7 is 12.7 test 12.7  this 12.8 is 12.8 test 12.8  this 12.9 is 12.9
 test 12.9  this 12.10 is 12.10 test 12.10  this 12.11 is 12.11 test 12.11
 this 12.12 is 12.12 test 12.12  this 12.13 is 12.13 test 12.13  this 12.14
 is 12.14 test 12.14  this 12.15 is 12.15 test 12.15 

In the next example the graphics adapt to the available space:

```

\startsetups adaptive:test
  \setbox\usedadaptivebox\hbox
  to      \usedadaptivewidth
  yoffset -\usedadaptivedepth
\bgroup
  \externalfigure

```

```

[cow.pdf]
[width=\usedadaptivewidth,
height=\dimexpr\usedadaptivetotal\relax]%
\egroup
\stopsetups

\dostepwiserecurse{1}{50}{1}{%
  this~#1 is~#1 test~#1
  {\adaptivebox[strut=yes,setup=adaptive:test]{} }
}

```

this 1 is 1 test 1  this 2 is 2 test 2  this 3 is 3 test 3  this 4 is 4 test 4  this 5
 is 5 test 5  this 6 is 6 test 6  this 7 is 7 test 7  this 8 is 8 test 8  this 9 is 9
 test 9  this 10 is 10 test 10  this 11 is 11 test 11  this 12 is 12 test 12  this 13
 is 13 test 13  this 14 is 14 test 14  this 15 is 15 test 15  this 16 is 16 test 16
 this 17 is 17 test 17  this 18 is 18 test 18  this 19 is 19 test 19  this 20 is 20
 test 20  this 21 is 21 test 21  this 22 is 22 test 22  this 23 is 23 test 23
 this 24 is 24 test 24  this 25 is 25 test 25  this 26 is 26 test 26  this 27 is 27
 test 27  this 28 is 28 test 28  this 29 is 29 test 29  this 30 is 30 test 30
 this 31 is 31 test 31  this 32 is 32 test 32  this 33 is 33 test 33  this 34 is 34
 test 34  this 35 is 35 test 35  this 36 is 36 test 36  this 37 is 37 test 37
 this 38 is 38 test 38  this 39 is 39 test 39  this 40 is 40 test 40  this 41 is 41
 test 41  this 42 is 42 test 42  this 43 is 43 test 43  this 44 is 44 test 44
 this 45 is 45 test 45  this 46 is 46 test 46  this 47 is 47 test 47  this 48 is 48
 test 48  this 49 is 49 test 49  this 50 is 50 test 50 

11.9 Prevdepth

The depth of a box is normally positive but rules can have a negative depth in order to get a rule above the baseline. When T_EX was written the assumption was that a negative depth of more than 1000 point made no sense at all. The last depth on a vertical list is registered in the `\prevdepth` variable. This is basically a reference into the current list. In order to illustrate some interesting side effects of setting this `\prevdepth` and especially when we set it to `-1000pt`. In order to illustrate this this special value can be set to a different value in LuaMetaT_EX. However, as dealing with the property is somewhat special in the engine you should not set it unless you know that the macro package is ware of it.

```
line 1\par line 2 \par \nointerlineskip line 3 \par
```

Assuming that we haven't set any inter paragraph spacing this gives:

Prevdepth

line 1
line 2
line 3

Here `\nointerlineskip` is (normally) defined as:

```
\prevdepth-1000pt
```

although in ConT_EXt we use `\ignoredepthcriterium` instead of the hard coded dimension. We now give a more extensive example:

In this example we set `\ignoredepthcriterium` to `-50.0pt` instead of the normal `-1000pt`. The helper is defined as:

or

The result of the following example is shown in figures 11.10 and 11.11. The first case is what we normally have in text and we haven't set `prevdepth` explicitly between lines so T_EX will just look at the depth of the lines. In the second case the depth is ignored when less than the criterium which is why, when we set the depth of the box to a negative value we get somewhat interesting skips.

FIRST						
-10.0pt	-20.0pt	-49.9pt	-50.0pt	-50.1pt	-60.0pt	-80.0pt
LAST						

Figure 11.10

FIRST	FIRST	FIRST
-10.0pt	-20.0pt	-49.9pt
LAST	LAST	LAST

FIRST	FIRST	FIRST	FIRST
-50.0pt	-50.1pt	-60.0pt	-80.0pt

Figure 11.11

I'm sure one can use this effect otherwise than intended but I doubt is any user is willing to do this but the fact that we can lower the criterium makes for nice experiments. Just for the record, in figure 11.12 you see what we get with positive values:

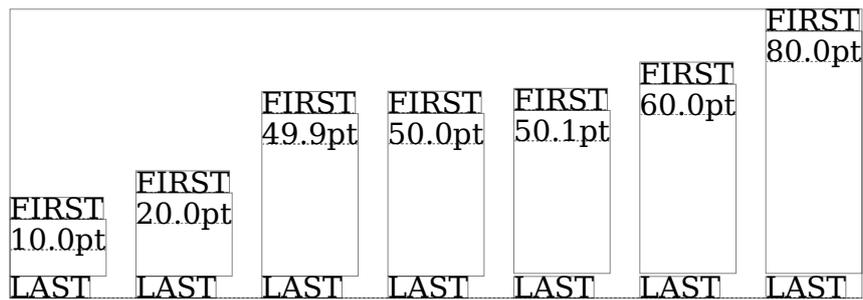


Figure 11.12

Watch the interline skip kicking in when we make the depth larger than in `\ignoredepthcriterion` being 50pt.

11.10 Normalization

todo: users don't need to bother about this but it might be interesting anyway

11.11 Dirty tricks

todo: explain example for combining paragraphs

11.11 Colofon

Author Hans Hagen
 ConT_EXt 2023.04.27 17:04
 LuaMetaT_EX 2.1008
 Support www.pragma-ade.com
 contextgarden.net

12 Alignments

low level

TEX

alignments

Contents

12.1	Introduction	169
12.2	Between the lines	171
12.3	Pre-, inter- and post-tab skips	173
12.4	Cell widths	176
12.5	Plugins	177
12.6	Pitfalls and tricks	180
12.7	Rows	183
12.8	Remark	186

12.1 Introduction

\TeX has a couple of subsystems and alignments is one of them. This mechanism is used to construct tables or alike. Because alignments use low level primitives to set up and construct a table, and because such a setup can be rather extensive, in most cases users will rely on macros that hide this.

```
\halign {
    \alignmark\hss \aligntab
  \hss\alignmark\hss \aligntab
  \hss\alignmark      \cr
  1.1      \aligntab 2,2      \aligntab 3=3      \cr
  11.11    \aligntab 22,22   \aligntab 33=33   \cr
  111.111  \aligntab 222,222 \aligntab 333=333 \cr
}
```

That one doesn't look too complex and comes out as:

```
1.1      2,2      3=3
11.11    22,22   33=33
111.111  222,222 333=333
```

This is how the previous code comes out when we use one of the $\text{Con}\text{\TeX}t$ table mechanism.

```
\starttabulate[|l|c|r|]
  \NC 1.1      \NC 2,2      \NC 3=3      \NC \NR
  \NC 11.11    \NC 22,22   \NC 33=33   \NC \NR
  \NC 111.111  \NC 222,222 \NC 333=333 \NC \NR
\stoptabulate
```

```
1.1      2,2      3=3
11.11    22,22    33=33
111.111  222,222  333=333
```

That one looks a bit different with respect to spaces, so let's go back to the low level variant:

```
\halign {
    \alignmark\hss \aligntab
    \hss\alignmark\hss \aligntab
    \hss\alignmark    \cr
  1.1\aligntab      2,2\aligntab      3=3\cr
  11.11\aligntab    22,22\aligntab    33=33\cr
  111.111\aligntab 222,222\aligntab 333=333\cr
}
```

Here we don't have spaces in the content part and therefore also no spaces in the result:

```
1.1      2,2      3=3
11.11    22,22    33=33
111.111222,222333=333
```

You can automate dealing with unwanted spacing:

```
\halign {
    \ignorespaces\alignmark\unskip\hss \aligntab
    \hss\ignorespaces\alignmark\unskip\hss \aligntab
    \hss\ignorespaces\alignmark\unskip    \cr
  1.1      \aligntab 2,2      \aligntab 3=3      \cr
  11.11    \aligntab 22,22    \aligntab 33=33    \cr
  111.111  \aligntab 222,222  \aligntab 333=333  \cr
}
```

We get:

```
1.1      2,2      3=3
11.11    22,22    33=33
111.111222,222333=333
```

By moving the space skipping and cleanup to the so called preamble we don't need to deal with it in the content part. We can also deal with inter-column spacing there:

```
\halign {
```

```

\ignorespaces\alignmark\unskip\hss \tabskip 1em \aligntab
\hss\ignorespaces\alignmark\unskip\hss \tabskip 1em \aligntab
\hss\ignorespaces\alignmark\unskip \tabskip 0pt \cr
1.1 \aligntab 2,2 \aligntab 3=3 \cr
11.11 \aligntab 22,22 \aligntab 33=33 \cr
111.111 \aligntab 222,222 \aligntab 333=333 \cr
}

1.1 2,2 3=3
11.11 22,22 33=33
111.111 222,222 333=333

```

If for the moment we forget about spanning columns (`\span`) and locally ignoring preamble entries (`\omit`) these basic commands are not that complex to deal with. Here we use `\alignmark` but that is just a primitive that we use instead of `#` while `\aligntab` is the same as `&`, but using the characters instead also assumes that they have the catcode that relates to a parameter and alignment tab (and in ConT_EXt that is not the case). The T_EXbook has plenty alignment examples so if you really want to learn about them, consult that must-have-book.

12.2 Between the lines

The individual rows of a horizontal alignment are treated as lines. This means that, as we see in the previous section, the interline spacing is okay. However, that also means that when we mix the lines with rules, the normal T_EX habits kick in. Take this:

```

\halign {
\ignorespaces\alignmark\unskip\hss \tabskip 1em \aligntab
\hss\ignorespaces\alignmark\unskip\hss \tabskip 1em \aligntab
\hss\ignorespaces\alignmark\unskip \tabskip 0pt \cr
\noalign{\hrule}
1.1 \aligntab 2,2 \aligntab 3=3 \cr
\noalign{\hrule}
11.11 \aligntab 22,22 \aligntab 33=33 \cr
\noalign{\hrule}
111.111 \aligntab 222,222 \aligntab 333=333 \cr
\noalign{\hrule}
}

```

The result doesn't look pretty and actually, when you see documents produced by T_EX using alignments you should not be surprised to notice rather ugly spacing. The user (or the macropackage) should deal with that explicitly, and this is not always the case.

1.1	2,2	3=3
11.11	22,22	33=33
111.111	222,222	333=333

The solution is often easy:

```
\halign {
  \ignorespaces\strut\alignmark\unskip\hss \tabskip 1em \aligntab
  \hss\ignorespaces\strut\alignmark\unskip\hss \tabskip 1em \aligntab
  \hss\ignorespaces\strut\alignmark\unskip \tabskip 0pt \cr
  \noalign{\hrule}
  1.1 \aligntab 2,2 \aligntab 3=3 \cr
  \noalign{\hrule}
  11.11 \aligntab 22,22 \aligntab 33=33 \cr
  \noalign{\hrule}
  111.111 \aligntab 222,222 \aligntab 333=333 \cr
  \noalign{\hrule}
}
```

1.1	2,2	3=3
11.11	22,22	33=33
111.111	222,222	333=333

The user will not notice it but alignments put some pressure on the general T_EX scanner. Actually, the scanner is either scanning an alignment or it expects regular text (including math). When you look at the previous example you see `\noalign`. When the preamble is read, T_EX will pick up rows till it finds the final brace. Each row is added to a temporary list and the `\noalign` will enter a mode where other stuff gets added to that list. It all involves subtle look ahead but with minimal overhead. When the whole alignment is collected a final pass over that list will package the cells and rows (lines) in the appropriate way using information collected (like the maximum width of a cell and width of the current cell. It will also deal with spanning cells then.

So let's summarize what happens:

1. scan the preamble that defines the cells (where the last one is repeated when needed)
2. check for `\cr`, `\noalign` or a right brace; when a row is entered scan for cells in parallel the preamble so that cell specifications can be applied (then start again)
3. package the preamble based on information with regards to the cells in a column
4. apply the preamble packaging information to the columns and also deal with pending cell spans
5. flush the result to the current list, unless packages in a box a `\halign` is seen as paragraph and rows as lines (such a table can split)

The second (repeated) step is complicated by the fact that the scanner has to look ahead for a `\noalign`, `\cr`, `\omit` or `\span` and when doing that it has to expand what comes. This can give side effects and often results in obscure error messages. When for instance an `\if` is seen and expanded, the wrong branch can be entered. And when you use protected macros embedded alignment commands are not seen at all; of course they still need to produce valid operations in the current context.

All these side effects are to be handled in a macro package when it wraps alignments in a high level interface and ConT_EXt does that for you. But because the code doesn't always look pretty then, in LuaMetaT_EX the alignment mechanism has been extended a bit over time.

Nesting `\noalign` is normally not permitted (but one can redefine this primitive such that a macro package nevertheless handles it). The first extension permits nested usage of `\noalign`. This has resulted of a little reorganization of the code. A next extension showed up when overload protection was introduced and extra prefixes were added. We can signal the scanner that a macro is actually a `\noalign` variant.¹⁹

```
\noaligned\protected\def\InBetween{\noalign{...}}
```

Here the `\InBetween` macro will get the same treatment as `\noalign` and it will not trigger an error. This extension resulted in a second bit of reorganization (think of internal command codes and such) but still the original processing of alignments was there.

A third overhaul of the code actually did lead to some adaptations in the way alignments are constructed so let's move on to that.

12.3 Pre-, inter- and post-tab skips

The basic structure of a preamble and row is actually not that complex: it is a mix of tab skip glue and cells (that are just boxes):

```
\tabskip 10pt
\halign {
  \strut\alignmark\tabskip 12pt\aligntab
  \strut\alignmark\tabskip 14pt\aligntab
  \strut\alignmark\tabskip 16pt\cr
  \noalign{\hrule}
```

¹⁹ One can argue for using the name `\peekaligned` because in the meantime other alignment primitives also can use this property.

```
cell 1.1\aligntab cell 1.2\aligntab cell 1.3\cr
\noalign{\hrule}
cell 2.1\aligntab cell 2.2\aligntab cell 2.3\cr
\noalign{\hrule}
}
```

The tab skips are set in advance and apply to the next cell (or after the last one).

cell 1.1	cell 1.2	cell 1.3
cell 2.1	cell 2.2	cell 2.3

In the ConT_EXt table mechanisms the value of `\tabskip` is zero in most cases. As in:

```
\tabskip 0pt
\halign {
  \strut\alignmark\aligntab
  \strut\alignmark\aligntab
  \strut\alignmark\cr
  \noalign{\hrule}
  cell 1.1\aligntab cell 1.2\aligntab cell 1.3\cr
  \noalign{\hrule}
  cell 2.1\aligntab cell 2.2\aligntab cell 2.3\cr
  \noalign{\hrule}
}
```

When these skips are zero, they still show up in the end:

cell 1.1	cell 1.2	cell 1.3	
cell 2.1	cell 2.2	cell 2.3	

Normally, in order to achieve certain effects there will be more align entries in the preamble than cells in the table, for instance because you want vertical lines between

cells. When these are not used, you can get quite a bit of empty boxes and zero skips. Now, of course this is seldom a problem, but when you have a test document where you want to show font properties in a table and that font supports a script with some ten thousand glyphs, you can imagine that it accumulates and in LuaT_EX (and LuaMetaT_EX) nodes are larger so it is one of these cases where in ConT_EXt we get messages on the console that node memory is bumped.²⁰

After playing a bit with stripping zero tab skips I found that the code would not really benefit from such a feature: lots of extra tests made it quite ugly. As a result a first alternative was to just strip zero skips before an alignment got flushed. At least we're then a bit leaner in the processes that come after it. This feature is now available as one of the normalizer bits.

But, as we moved on, a more natural approach was to keep the skips in the preamble, because that is where a guaranteed alternating skip/box is assumed. It also makes that the original documentation is still valid. However, in the rows construction we can be lean. This is driven by a keyword to `\halign`:

```
\tabskip 0pt
\halign noskips {
  \strut\alignmark\aligntab
  \strut\alignmark\aligntab
  \strut\alignmark\cr
  \noalign{\hrule}
  cell 1.1\aligntab cell 1.2\aligntab cell 1.3\cr
  \noalign{\hrule}
  cell 2.1\aligntab cell 2.2\aligntab cell 2.3\cr
  \noalign{\hrule}
}
```

No zero tab skips show up here:

²⁰ I suppose it was a coincidence that a few weeks after these features came available a user consulted the mailing list about a few thousand page table that made the engine run out of memory, something that could be cured by enabling these new features.

cell_{SP:3 497}1.1 cell_{SP:3 497}1.2 cell_{SP:3 497}1.3

cell_{SP:2 467}2.1 cell_{SP:2 467}2.2 cell_{SP:2 467}2.3

When playing with all this the LuaMetaTeX engine also got a tracing option for alignments. We already had one that showed some of the `\noalign` side effects, but showing the preamble was not yet there. This is what `\tracingalignments = 2` results in:

```
<preamble>
\glue[ignored][...] 0.0pt
\alignrecord
..{\strut }
..<content>
..\endtemplate }
\glue[ignored][...] 0.0pt
\alignrecord
..{\strut }
..<content>
..\endtemplate }
\glue[ignored][...] 0.0pt
\alignrecord
..{\strut }
..<content>
..\endtemplate }
\glue[ignored][...] 0.0pt
```

The ignored subtype is (currently) only used for these alignment tab skips and it triggers a check later on when the rows are constructed. The `<content>` is what get injected in the cell (represented by `\alignmark`). The pseudo primitives are internal and not public.

12.4 Cell widths

Imagine this:

```
\halign {
```

```
x\hbox to 3cm{\strut \alignmark\hss}\aligntab
x\hbox to 3cm{\strut\hss\alignmark\hss}\aligntab
x\hbox to 3cm{\strut\hss\alignmark } \cr
cell 1.1\aligntab cell 1.2\aligntab cell 1.3\cr
cell 2.1\aligntab cell 2.2\aligntab cell 2.3\cr
}
```

which renders as:

xcell 1.1	x	cell 1.2	x	cell 1.3
xcell 2.1	x	cell 2.2	x	cell 2.3

A reason to have boxes here is that it enforces a cell width but that is done at the cost of an extra wrapper. In LuaMetaT_EX the hlist nodes are rather large because we have more options than in original T_EX, for instance offsets and orientation. In a table with 10K rows of 4 cells yet get 40K extra hlist nodes allocated. Now, one can argue that we have plenty of memory but being lazy is not really a sign of proper programming.

```
\halign {
  x\tabsize 3cm\strut \alignmark\hss\aligntab
  x\tabsize 3cm\strut\hss\alignmark\aligntab
  x\tabsize 3cm\strut\hss\alignmark\hss\cr
  cell 1.1\aligntab cell 1.2\aligntab cell 1.3\cr
  cell 2.1\aligntab cell 2.2\aligntab cell 2.3\cr
}
```

If you look carefully you will see that this time we don't have the embedded boxes:

xcell 1.1	x	cell 1.2	x	cell 1.3
xcell 2.1	x	cell 2.2	x	cell 2.3

So, both the sparse skip and new \tabsize feature help to make these extreme tables (spanning hundreds of pages) not consume irrelevant memory and also make that later on we don't have to consult useless nodes.

12.5 Plugins

Yet another LuaMetaT_EX extension is a callback that kicks in between the preamble pre-roll and finalizing the alignment. Initially as test and demonstration a basic character

alignment feature was written but that works so well that in some places it can replace (or compliment) the already existing features in the ConT_EXt table mechanisms.

```
\starttabulate[|lG{.}|cG{,}|rG{=}|cG{x}||
\NC 1.1      \NC 2,2      \NC 3=3      \NC a 0xFF    \NC \NR
\NC 11.11   \NC 22,22   \NC 33=33   \NC b 0xFFF   \NC \NR
\NC 111.111 \NC 222,222 \NC 333=333 \NC c 0xFFFF \NC \NR
\stoptabulate
```

The tabulate mechanism in ConT_EXt is rather old and stable and it is the preferred way to deal with tabular content in the text flow. However, adding the G specifier (as variant of the g one) could be done without interference or drop in performance. This new G specifier tells the tabulate mechanism that in that column the given character is used to vertically align the content that has this character.

```
1.1      2,2      3=3      a 0xFF
11.11   22,22   33=33   b 0xFFF
111.111 222,222 333=333 c 0xFFFF
```

Let's make clear that this is *not* an engine feature but a ConT_EXt one. It is however made easy by this callback mechanism. We can of course use this feature with the low level alignment primitives, assuming that you tell the machinery that the plugin is to be kicked in.

```
\halign noskips \alignmentcharactertrigger \bgroup
\tabskip2em
  \setalignmentcharacter.\ignorespaces\alignmark\unskip\hss \aligntab
\hss\setalignmentcharacter.\ignorespaces\alignmark\unskip\hss \aligntab
\hss\setalignmentcharacter=\ignorespaces\alignmark\unskip \aligntab
\hss \ignorespaces\alignmark\unskip\hss \cr
1.1 \aligntab 2,2 \aligntab 3=3 \aligntab \setalignmentcharacter{.}\relax 4.4\cr
11.11 \aligntab 22,22 \aligntab 33=33 \aligntab \setalignmentcharacter{,}\relax 44,44\cr
111.111 \aligntab 222,222 \aligntab 333=333 \aligntab \setalignmentcharacter{!}\relax 444!444\cr
x \aligntab x \aligntab x \aligntab \setalignmentcharacter{/}\relax /\cr
.1 \aligntab ,2 \aligntab =3 \aligntab \setalignmentcharacter{?}\relax ?4\cr
.111 \aligntab ,222 \aligntab =333 \aligntab \setalignmentcharacter{=}\relax 44=444\cr
\egroup
```

This rather verbose setup renders as:

```
1.1      2,2      3=3      4 . 4
11.11   22,22   33=33   44 , 44
111.111 222,222 333=333 444 ! 444
x        x        x        /
.1       ,2       =3       ?4
.111    ,222     =333     44=444
```

Using a high level interface makes sense but local control over such alignment too, so here follow some more examples. Here we use different alignment characters:

```
\starttabulate[|lG{.}|cG{,}|rG{=}|cG{x}||
\NC 1.1      \NC 2,2      \NC 3=3      \NC a 0xFF    \NC \NR
\NC 11.11   \NC 22,22   \NC 33=33   \NC b 0xFFF   \NC \NR
\NC 111.111 \NC 222,222 \NC 333=333 \NC c 0xFFFF \NC \NR
\stoptabulate
```

```
1.1      2,2      3=3      a 0xFF
11.11   22,22   33=33   b 0xFFF
111.111 222,222 333=333 c 0xFFFF
```

In this example we specify the characters in the cells. We still need to add a specifier in the preamble definition because that will trigger the plugin.

```
\starttabulate[|lG{}|rG{}|]
\NC left      \NC right      \NC \NR
\NC \showglyphs \setalignmentcharacter{.}1.1 \NC \setalignmentcharacter{.}1.1 \NC \NR
\NC \showglyphs \setalignmentcharacter{,}11,11 \NC \setalignmentcharacter{,}11,11 \NC \NR
\NC \showglyphs \setalignmentcharacter{=}111=111 \NC \setalignmentcharacter{=}111=111 \NC \NR
\stoptabulate
```

```
left      right
 1.1      1.1
11,11     11,11
111=111   111=111
```

You can mix these approaches:

```
\starttabulate[|lG{.}|rG{}|]
\NC left      \NC right      \NC \NR
\NC 1.1      \NC \setalignmentcharacter{.}1.1 \NC \NR
\NC 11.11   \NC \setalignmentcharacter{.}11.11 \NC \NR
\NC 111.111 \NC \setalignmentcharacter{.}111.111 \NC \NR
\stoptabulate
```

```
left      right
 1.1      1.1
11.11     11.11
111.111   111.111
```

Here the already present alignment feature, that at some point in tabulate might use this new feature, is meant for numbers, but here we can go wild with words, although

of course you need to keep in mind that we deal with typeset text, so there may be no match.

```
\starttabulate[|lG{.}|rG{.}|]
\NC foo.bar \NC foo.bar \NC \NR
\NC oo.ba \NC oo.ba \NC \NR
\NC o.b \NC o.b \NC \NR
\stoptabulate
```

```
foo.bar foo.bar
oo.ba oo.ba
o.b o.b
```

This feature will only be used in know situations and those seldom involve advanced typesetting. However, the following does work:²¹

```
\starttabulate[|cG{d}|]
\NC \smallcaps abcdefgh \NC \NR
\NC xdy \NC \NR
\NC \sl xdy \NC \NR
\NC \tttf xdy \NC \NR
\NC \tfd d \NC \NR
\stoptabulate
```

```
abc d efgh
x d y
x d y
x d y
d
```

As always with such mechanisms, the question is “Where to stop?” But it makes for nice demos and as long as little code is needed it doesn't hurt.

12.6 Pitfalls and tricks

The next example mixes bidirectional typesetting. It might look weird at first sight but the result conforms to what we discussed in previous paragraphs.

```
\starttabulate[|lG{.}|lG{}|]
\NC \righttoleft 1.1 \NC \righttoleft \setalignmentcharacter{.}1.1 \NC\NR
```

²¹ Should this be an option instead?

```

\NC          1.1 \NC          \setalignmentcharacter{.}1.1 \NC\NR
\NC \righttoleft 1.11 \NC \righttoleft \setalignmentcharacter{.}1.11 \NC\NR
\NC          1.11 \NC          \setalignmentcharacter{.}1.11 \NC\NR
\NC \righttoleft 1.111 \NC \righttoleft \setalignmentcharacter{.}1.111 \NC\NR
\NC          1.111 \NC          \setalignmentcharacter{.}1.111 \NC\NR
\stoptabulate

```

```

  1.1    1.1
1.1    1.1
  11.1   11.1
1.11   1.11
111.1  111.1
1.111  1.111

```

In case of doubt, look at this:

```

\starttabulate[|lG{.}|lG{.}|lG{.}|lG{.}|]
\NC \righttoleft 1.1 \NC \righttoleft \setalignmentcharacter{.}1.1 \NC
      1.1 \NC          \setalignmentcharacter{.}1.1 \NC\NR
\NC \righttoleft 1.11 \NC \righttoleft \setalignmentcharacter{.}1.11 \NC
      1.11 \NC          \setalignmentcharacter{.}1.11 \NC\NR
\NC \righttoleft 1.111 \NC \righttoleft \setalignmentcharacter{.}1.111 \NC
      1.111 \NC          \setalignmentcharacter{.}1.111 \NC\NR
\stoptabulate

```

```

  1.1    1.1  1.1    1.1
  11.1   11.1  1.11   1.11
111.1  111.1  1.111  1.111

```

The next example shows the effect of `\omit` and `\span`. The first one makes that in this cell the preamble template is ignored.

`\halign \bgroup`

```

\tabsize 2cm\relax [\alignmark]\hss \aligntab
\tabsize 2cm\relax \hss[\alignmark]\hss \aligntab
\tabsize 2cm\relax \hss[\alignmark]\cr
      1\aligntab      2\aligntab      3\cr
\omit 1\aligntab \omit 2\aligntab \omit 3\cr
      1\aligntab      2\span          3\cr
      1\span          2\aligntab      3\cr
      1\span          2\span          3\cr
      1\span \omit    2\span \omit    3\cr
\omit 1\span \omit    2\span \omit    3\cr

```

`\egroup`

Spans are applied at the end so you see a mix of templates applied.


```

1\aligntab 2\aligntab 3\cr
A\aligntab B\aligntab C\cr
\egroup

```

Where the helpers are defined as:

```

\noaligned\protected\def\aligncontentleft
  {\ignorespaces\aligncontent\unskip\hss}

\noaligned\protected\def\aligncontentmiddle
  {\hss\ignorespaces\aligncontent\unskip\hss}

\noaligned\protected\def\aligncontentright
  {\hss\ignorespaces\aligncontent\unskip}

```

The preamble scanner see such macros as candidates for a single level expansion so it will inject the meaning and see the `\aligncontent` eventually.

```

1          2          3
A          B          C

```

The same effect could be achieved by using the `\span` prefix:

```

\def\aligncontentleft{\ignorespaces\aligncontent\unskip\hss}

\halign { ... \span\aligncontentleft ...}

```

One of the reasons for not directly using the low level `\halign` command is that it's a lot of work but by providing a set of helpers like here might change that a bit. Keep in mind that much of the above is not new in the sense that we could not achieve the same already, it's just a bit programmer friendly.

12.7 Rows

Alignment support is what the documented source calls ‘interwoven’. When the engine scans for input it processing text, math or alignment content. While doing alignments it collects rows, and inside these cells but also deals with material that ends up in between. In LuaMetaTeX I tried to isolate the bits and pieces as good as possible but it remains complicated (for all good reasons). Cells as well as rows are finalized after the whole alignment has been collected and processed. In the end cells and rows are boxes but till we're done they are in an ‘unset’ state.

Scanning starts with interpreting the preamble, and then grabbing rows. There is some nasty lookahead involved for `\noalign`, `\span`, `\omit`, `\cr` and `\crr` and that is not code

one wants to tweak too much (although we did in LuaMetaTeX). This means for instance that adding ‘let's start a row here’ primitive is sort of tricky (but it might happen some day) which in turn means that it is not really possible to set row properties. As an experiment we can set some properties now by hijacking `\noalign` and storing them on the alignment stack (indeed: at the cost of some extra overhead and memory). This permits the following:

```

\halign {
  \hss
  \ignorespaces \alignmark \removeunwantedspaces
  \hss
  \quad \aligntab \quad
  \hss
  \ignorespaces \alignmark \removeunwantedspaces
  \hss
  \cr
  \noalign xoffset 40pt {}
  {\darkred cell one} \aligntab {\darkgray cell one} \cr
  \noalign orientation "002 {}
  {\darkgreen cell one} \aligntab {\darkblue cell one} \cr
  \noalign xoffset 40pt {}
  {\darkred cell two} \aligntab {\darkgray cell two} \cr
  \noalign orientation "002 {}
  {\darkgreen cell two} \aligntab {\darkblue cell two} \cr
  \noalign xoffset 40pt {}
  {\darkred cell three} \aligntab {\darkgray cell three} \cr
  \noalign orientation "002 {}
  {\darkgreen cell three} \aligntab {\darkblue cell three} \cr
  \noalign xoffset 40pt {}
  {\darkred cell four} \aligntab {\darkgray cell four} \cr
  \noalign orientation "002 {}
  {\darkgreen cell four} \aligntab {\darkblue cell four} \cr
}

```

cell one	cell one
cell one	cell one
cell two	cell two
cell two	cell two
cell three	cell three
cell three	cell three
cell four	cell four
cell four	cell four

The supported keywords are similar to those for boxes: source, target, anchor, orientation, shift, xoffset, yoffset, xmove and ymove. The dimensions can be prefixed by add and reset wipes all. Here is another example:

```
\halign {
  \hss
  \ignorespaces \alignmark \removeunwantedspaces
  \hss
  \quad \aligntab \quad
  \hss
  \ignorespaces \alignmark \removeunwantedspaces
  \hss
  \cr
  \noalign xmove 40pt {}
  {\darkred cell one} \aligntab {\darkgray cell one} \cr
  {\darkgreen cell one} \aligntab {\darkblue cell one} \cr
  \noalign xmove 20pt {}
  {\darkred cell two} \aligntab {\darkgray cell two} \cr
  {\darkgreen cell two} \aligntab {\darkblue cell two} \cr
  \noalign xmove 40pt {}
  {\darkred cell three} \aligntab {\darkgray cell three} \cr
  {\darkgreen cell three} \aligntab {\darkblue cell three} \cr
  \noalign xmove 20pt {}
  {\darkred cell four} \aligntab {\darkgray cell four} \cr
  {\darkgreen cell four} \aligntab {\darkblue cell four} \cr
}
```

cell one	cell one
cell one	cell one
cell two	cell two
cell two	cell two
cell three	cell three
cell three	cell three
cell four	cell four
cell four	cell four

Some more features might be added in the future as is it an interesting playground. It is to be seen how this ends up in ConT_EXt high level interfaces like tabulate.

12.8 Remark

It can be that the way alignments are interfaced with respect to attributes is a bit different between LuaT_EX and LuaMetaT_EX but because the former is frozen (in order not to interfere with current usage patterns) this is something that we will deal with deep down in ConT_EXt LMTX.

In principle we can have hooks into the rows for pre and post material but it doesn't really pay of as grouping will still interfere. So for now I decided not to add these.

12.8 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

13 Marks

low level

TEX

marks

Contents

13.1	Introduction	188
13.2	The basics	189
13.3	Migration	190
13.4	Tracing	192
13.5	High level commands	193
13.6	Pitfalls	195

13.1 Introduction

Marks are one of the subsystems of $\text{T}_{\text{E}}\text{X}$, as are for instance alignments and math as well as inserts which they share some properties with. Both inserts and marks put signals in the list that later on get intercepted and can be used to access stored information. In the case of inserts this is typeset materials, like footnotes, and in the case of marks it's token lists. Inserts are taken into account when breaking pages, and marks show up when a page has been broken and is presented to the output routine. Marks are used for running headers but other applications are possible.

In MkII marks are used to keep track of colors, transparencies and more properties that work across page boundaries. It permits picking up at the top of a page from where one left at the bottom of the preceding one. When MkII was written there was only one mark so on top of that a multiple mark mechanism was implemented that filtered specific marks from a collection. Later, $\varepsilon\text{-T}_{\text{E}}\text{X}$ provided mark classes so that mechanism could be simplified. Although it is not that hard to do, this extension to $\text{T}_{\text{E}}\text{X}$ didn't add any further features, so we can assume that there was no real demand for that.²²

But, marks have some nasty limitations, so from the $\text{ConT}_{\text{E}}\text{Xt}$ perspective there always was something to wish for. When you hide marks in boxes they will not be seen (the same is true for inserts). You cannot really reset them either. Okay, you can set them to nothing, but even then already present marks are still there. The $\text{LuaT}_{\text{E}}\text{X}$ engine has a `\clearmarks` primitive but that works global. In $\text{LuaMetaT}_{\text{E}}\text{X}$ a proper mark flusher is available. That engine also can work around the deeply nested disappearing marks. In addition, the current state of a mark can be queried and we have some tracing facilities.

In MkIV the engine's marks were not used at all and an alternative mechanism was written using Lua. It actually is one of the older MkIV features. It doesn't have the side

²² This is probably true for most $\text{LuaT}_{\text{E}}\text{X}$ and $\text{LuaMetaT}_{\text{E}}\text{X}$ extensions, maybe example usage create retrospective demand. But one reason for picking up on engine development is that in the $\text{ConT}_{\text{E}}\text{Xt}$ perspective we actually had some demands.

effects that native marks have but it comes at the price of more overhead, although that is bearable.

In this document we discuss marks but assume that LuaMetaT_EX is used with ConT_EXt LMTX. There we experiment with using the native marks, complemented by a few Lua mechanisms, but it is to be seen if that will be either a replacement or an alternative.

13.2 The basics

Although the original T_EX primitives are there, the plural ε -T_EX mark commands are to be used. Marks, signals with token lists, are set with:

```
\marks0{This is mark 0} % equivalent to: \mark{This is mark 0}
\marks4{This is mark 4}
```

When a page has been split off, you can (normally this only makes sense in the output routine) access marks with:

```
\topmarks 4
\firstmarks4
\botmarks 4
```

A ‘top’ mark is the last one on the previous page(s), the ‘first’ and ‘bottom’ refer to the current page. A mark is a so called node, something that ends up in the current list and the token list is stored with it. The accessors are just commands and they fetch the token list from a separately managed storage. When you set or access a mark that has not yet been used, the storage is bumped to the right size, so it doesn't make sense to use e.g. `\marks 999` when there are no 998 ones too: it not only takes memory, it also makes T_EX run over all these mark stores when synchronization happens. The best way to make sure that you are sparse is:

```
\newmarks\MyMark
```

Currently the first 16 marks are skipped so this makes `\MyMark` become mark 17. The reason is that we want to make sure that users who experiment with marks have some scratch marks available and don't overload system defined ones. Future versions of ConT_EXt might become more restrictive.

Marks can be cleared with:

```
\clearmarks 4
```

which clears the storage that keeps the top, first and bot marks. This happens immediately. You can delay this by putting a signal in the list:

`\flushmarks` 4

This (LuaMetaT_EX) feature makes it for instance easy to reset marks that keep track of section (and lower) titles when a new chapter starts. Of course it still means that one has to implement some mechanism that deals with this but ConT_EXt always had that.

The current, latest assigned, value of a mark is available too:

`\currentmarks` 4

Using this value in for instance headers and footers makes no sense because the last node set can be on a following page.

13.3 Migration

In the introduction we mentioned that LuaMetaT_EX has migration built in. In MkIV we have this as option too, but there it is delegated to Lua. It permits deeply nested inserts (notes) and marks (but we don't use native marks in MkIV).

Migrated marks end up in the postmigrated sublist of a box. In other lowlevel manuals we discuss these pre- and postmigrated sublists. As example we use this definition:

```
\setbox0\ vbox\bgroup
test \marks 4 {mark 4.1}\par
test \marks 4 {mark 4.1}\par
test \marks 4 {mark 4.1}\par
\egroup
```

When we turn migration on (officially the second bit):

```
\automigrationmode"FF \showbox0
```

we get this:

```
> \box0=
2:4: \vbox[normal][...], width 483.69687, height 63.43475, depth 0.15576, direction l2r
2:4: .\list
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...list
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphendemerits 10000,
  finalhyphendemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519
2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
```

```

2:4: ..\glue[par][...] 11.98988pt plus 3.99663pt minus 3.99663pt
2:4: ..\glue[baseline][...] 8.34883pt
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...List
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hangafter 1, hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphendemerits 10000,
finalhyphendemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519
2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
2:4: ..\glue[par][...] 11.98988pt plus 3.99663pt minus 3.99663pt
2:4: ..\glue[baseline][...] 8.34883pt
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...List
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hangafter 1, hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphendemerits 10000,
finalhyphendemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519
2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
2:4: ..\postmigrated
2:4: ..\mark[4][...]
2:4: ..{mark 4.1}

```

When we don't migrate, enforced with:

`\automigrationmode"00 \showbox0`

the result is:

```

> \box0=
2:4: \vbox[normal][...], width 483.69687, height 63.43475, depth 0.15576, direction l2r
2:4: ..List
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...List
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hangafter 1, hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphendemerits 10000,
finalhyphendemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519
2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
2:4: ..\mark[4][...]
2:4: ..{mark 4.1}
2:4: ..\glue[par][...] 11.98988pt plus 3.99663pt minus 3.99663pt
2:4: ..\glue[baseline][...] 8.34883pt
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...List
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hangafter 1, hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphendemerits 10000,
finalhyphendemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519

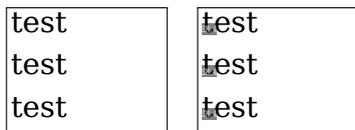
```

```

2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
2:4: ..\mark[4][...]
2:4: ..{mark 4.1}
2:4: ..\glue[par][...] 11.98988pt plus 3.99663pt minus 3.99663pt
2:4: ..\glue[baseline][...] 8.34883pt
2:4: ..\hbox[line][...], width 483.69687, height 7.48193, depth 0.15576, glue 459.20468fil, direction l2r
2:4: ...list
2:4: ....\glue[left hang][...] 0.0pt
2:4: ....\glue[left][...] 0.0pt
2:4: ....\glue[parfillleft][...] 0.0pt
2:4: ....\par[newgraf][...], hangafter 1, hsize 483.69687, pretolerance 100, tolerance 200, adjdemerits 10000, linepenalty 10, doublehyphemerits 10000, finalhyphemerits 5000, clubpenalty 2000, widowpenalty 2000, brokenpenalty 100, parfillskip 0.0pt plus 1.0fil, hyphenationmode 499519
2:4: ....\glue[indent][...] 0.0pt
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000065 e
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000073 s
2:4: ....\glyph[32768][...], language (n=1,l=2,r=3), hyphenationmode "79F3F, options "80, font <8: DejaVuSerif @ 11.0pt>, glyph U+000074 t
2:4: ....\glue[space][...] 3.49658pt plus 1.74829pt minus 1.16553pt, font 8
2:4: ....\penalty[line][...] 10000
2:4: ....\glue[parfill][...] 0.0pt plus 1.0fil
2:4: ....\glue[right][...] 0.0pt
2:4: ....\glue[right hang][...] 0.0pt
2:4: ..\mark[4][...]
2:4: ..{mark 4.1}

```

When you say `\showmakeup` or in this case `\showmakeup[mark]` the marks are visualized:



enabled disabled

Here `sm` means ‘set mark’ while `rm` would indicate a ‘reset mark’. Of course migrated marks don't show up because these are bound to the box and thereby have become a a specific box property as can be seen in the above trace.

13.4 Tracing

The LuaMetaTeX engine has a dedicated tracing option for marks. The fact that the traditional engine doesn't have this can be seen as indication that this is seldom needed.

```

\tracingmarks1
\tracingonline2

```

When tracing is set to 1 we get a list of marks for the just split of page:

```

2:7: <mark class 51, top := bot>
2:7: ..{sample 9.1}
2:7: <mark class 51: first := mark>

```

```

2:7: ..{sample 10.1}
2:7: <mark class 51: bot := mark>
2:7: ..{sample 10.1}
2:7: <mark class 51, page state>
2:7: ..top {sample 9.1}
2:7: ..first {sample 10.1}
2:7: ..bot {sample 10.1}

```

When tracing is set to 2 you also get details we get a list of marks of the analysis:

```

1:9: <mark class 51, top := bot>
1:9: ..{sample 5.1}
1:9: <mark class 51: first := mark>
1:9: ..{sample 6.1}
1:9: <mark class 51: bot := mark>
1:9: ..{sample 6.1}
1:9: <mark class 51: bot := mark>
1:9: ..{sample 7.1}
1:9: <mark class 51: bot := mark>
1:9: ..{sample 8.1}
1:9: <mark class 51: bot := mark>
1:9: ..{sample 9.1}
1:9: <mark class 51, page state>
1:9: ..top {sample 5.1}
1:9: ..first {sample 6.1}
1:9: ..bot {sample 9.1}

```

13.5 High level commands

I think that not that many users define their own marks. They are useful for showing section related titles in headers and footers but the implementation of that is hidden. The native mark references are `top`, `first` and `bottom` but in the `ConTeXt` interface we use different keywords.

ConTeXt	TeX	column	page
previous	top	last before sync	last on previous page
top	first	first in sync	first on page
bottom	bot	last in sync	last on page
first	top	first not top in sync	first on page
last	bot	last not bottom in sync	last on page

default	the same as first
current	the last set value

In order to separate marks in ConTEXt from those in TEX, the term ‘marking’ is used. In MkIV the regular marks mechanism is of course there but, as mentioned, not used. By using a different namespace we could make the transition from MkII to MkIV (the same is true for some more mechanisms).

A marking is defined with

```
\definemarking[MyMark]
```

A defined marking can be set with two equivalent commands:

```
\setmarking[MyMark]{content}
\marking [MyMark]{content}
```

The content is not typeset but stored as token list. In the sectioning mechanism that uses markings we don't even store titles, we store a reference to a title. In order to use that (deep down) we hook in a filter command. By default that command does nothing:

```
\setupmarking[MyMark][filtercommand=\firstofoneargument]
```

The token list does *not* get expanded by default, unless you set it up:

```
\setupmarking[MyMark][expansion=yes]
```

The current state of a marking can be cleared with:

```
\clearmarking[MyMark]
```

but because that en is not synchronized the real deal is:

```
\resetmarking[MyMark]
```

Be aware that it introduces a node in the list. You can test if a marking is defined with (as usual) a test macro. Contrary to (most) other test macros this one is fully expandable:

```
\doifelsemarking {MyMark} {
  defined
} {
  undefined
}
```

Because there can be a chain involved, we can relate markings. Think of sections below chapters and subsections below sections:

```
\relatemarking[MyMark][YourMark]
```

When a marking is set its relatives are also reset, so setting YourMark will reset MyMark. It is this kind of features that made for marks being wrapped into high level commands very early in the ConT_EXt development (and one can even argue that this is why a package like ConT_EXt exists in the first place).

The rest of the (relatively small) repertoire of commands has to do with fetching markings. The general command is `\getmarking` that takes two or three arguments:

```
\getmarking[MyMarking][first]
\getmarking[MyMarking][page][first]
\getmarking[MyMarking][page][first]
\getmarking[MyMarking][column:1][first]
```

There are (normally) three marks that can be fetched so we have three commands that do just that:

```
\fetchonemark [MyMarking][which one]
\fetchtwomarks[MyMarking]
\fetchallmarks[MyMarking]
```

You can setup a separator key which by default is:

```
\setupmarking[MyMarking][separator=\space\emdash\space]
```

Injection is enabled by default due to this default:

```
\setupmarking[MyMarking][state=start]
```

The following three variants are (what is called) fully expandable:

```
\fetchonemarking [MyMarking][which one]
\fetchtwomarkings[MyMarking]
\fetchallmarkings[MyMarking]
```

13.6 Pitfalls

The main pitfall is that a (re)setting a mark will inject a node which in vertical mode can interfere with spacing. In for instance section commands we wrap them with the title so there it should work out okay.

13.6 Colofon

Author Hans Hagen
ConT_EXt 2023.04.27 17:04
LuaMetaT_EX 2.1008
Support www.pragma-ade.com
 contextgarden.net

14 Inserts

low level

TEX

inserts

Contents

14.1	Introduction	198
14.2	The page builder	198
14.3	Inserts	200
14.4	Storing	201
14.5	Synchronizing	201
14.6	Migration	201
14.7	Callbacks	202

14.1 Introduction

This document is a mixed bag. We do discuss inserts but also touch elements of the page builder because inserts and regular page content are handled there. Examples of mechanisms that use inserts are footnotes. These have an anchor in the running text and some content that ends up (normally) at the bottom of the page. When considering a page break the engine tries to make sure that the anchor (reference) and the content end up on the same page. When there is too much, it will distribute (split) the content over pages.

We can discuss page breaks in a (pseudo) scientific way and explore how to optimize this process, taking into accounts also inserts that contain images but it doesn't make much sense to do that because in practice we can encounter all kind of interferences. Theory and practice are too different because a document can contain a wild mix of text, figures, formulas, notes, have backgrounds and location dependent processing. It get seven more complex when we are dealing with columns because $\text{T}_{\text{E}}\text{X}$ doesn't really know that concept.

I will therefore stick to some practical aspects and the main reason for this document is that I sort of document engine features and at the same time give an impression of what we deal with. I will do that in the perspective of $\text{LuaMetaT}_{\text{E}}\text{X}$, which has a few more options and tracing than other engines.

Currently this document is mostly for myself to keep track of the state of inserts and the page builder in $\text{LuaMetaT}_{\text{E}}\text{X}$ and $\text{ConT}_{\text{E}}\text{Xt LMTX}$. The text is not yet corrected and can have errors.

14.2 The page builder

When your document is processed content eventually gets added to the so called main vertical list (mvl). Content first get appended to the list of contributions and at specific

moments it will be handed over to the mvl. This process is called page building. There we can encounter the following elements (nodes):

glue	a vertical skip
penalty	a vertical penalty
kern	a vertical kern
vlist	a a vertical box
hlist	a horizontal box (often a line)
rule	a horizontal rule
boundary	a boundary node
whatsit	a node that is used by user code (often some extension)
mark	a token list (as used for running headers)
insert	a node list (as used for notes)

The engine itself will not insert anything other than this but Lua code can mess up the contribution list and the mvl and that can trigger an error. Handing over the contributions is done by the page builder and that one kicks in in several places:

- When a penalty gets inserted it is part of evaluating if the output routine should be triggered. This triggering can be enforced by values equal or below 10.000 that then can be checked in the set routine.
- The builder is *not* exercised when a glue or kern is injected so there can be multiple of them before another element triggers the builder.
- Adding a box triggers the builder as does the result of an alignment which can be a list of boxes.
- When the output routine is finished the builder is executed because the routine can have pushed back content.
- When a new paragraph is triggered by the `\par` command the builder kicks in but only when the engine was able to enter vertical mode.
- When the job is finished the builder will make sure that pending content is handled.
- An insert and `vadjust can` trigger the builder but only at the nesting level zero which normally is not the case (I need an example).
- At the beginning of a paragraph (like text), before display math is entered, and when display math ends the builder is also activated.

At the \TeX the builder is triggered automatically in the mentioned cases but at the Lua end you can use `tex.triggerbuildpage()` to flush the pending contributions.

The properties that relate to the page look like counter and dimension registers but they are not. These variables are global and managed differently.

<code>\pagegoal</code>	the available space
<code>\pagetotal</code>	the accumulated space

The page builder

<code>\pagestretch</code>	the possible zero order stretch
<code>\pagefilstretch</code>	the possible one order stretch
<code>\pagefillstretch</code>	the possible second order stretch
<code>\pagefilllstretch</code>	the possible third order stretch
<code>\pageshrink</code>	the possible shrink
<code>\pagedepth</code>	the current page depth
<code>\pagevsize</code>	the initial page goal

When the first content is added to an empty page the `\pagegoal` gets the value of `\vsize` and gets frozen but the value is diminished by the space needed by left over inserts. These inserts are managed via a separate list so they don't interfere with the page that itself of course can have additional inserts. The `\pagevsize` is just a (LuaMeta-TeX) status variable that hold the initial `\pagegoal` but it might play a role in future extensions.

Another variable is `\deadcycles` that registers the number of times the output routine is called without returning result.

14.3 Inserts

We now come to inserts. In traditional TeX an insert is a data structure that runs on top of registers: a box, count, dimension and skip. An insert is accessed by a number so for instance insert 123 will use the four registers of that number. Because TeX only offers a command alias mechanism for registers (like `\countdef`) a macro package will implement some allocator management subsystem (like `\newcount`). A `\newinsert` has to be defined in a way that the four registers are not clashing with other allocators. When you start with TeX seeing code that deals with in (in plain TeX) can be puzzling but it follows from the way TeX is set up. But inserts are probably not what you start exploring right away away.

In LuaMetaTeX you can set `\insertmode` to 1 and that is what we do in ConTeXt. In that mode inserts are taken from a pool instead of registers. A side effect is that like the page properties the insert properties are global too but that is normally no problem and can be managed well by a macro package (that probably would assign register the values globally too). The insert pool will grow dynamically on demand so one can just start at 1; in ConTeXt MkIV we use the range 127 upto 255 in order to avoid a clash with registers. In LMTX we start at 1 because there are no clashes.

A consequence of this approach is that we use dedicated commands to set the insert properties:

<code>\insertdistance</code>	glue	the space before the first instance (on a page)
<code>\insertmultiplier</code>	count	a factor that is used to calculate the height used
<code>\insertlimit</code>	dimen	the maximum amount of space on a page to be taken
<code>\insertpenalty</code>	count	the floating penalty (used when set)
<code>\insertmaxdepth</code>	dimen	the maximum split depth (used when set)
<code>\insertstorage</code>	count	signals that the insert has to be stored for later
<code>\insertheight</code>	dimen	the accumulated height of the inserts so far
<code>\insertdepth</code>	dimen	the current depth of the inserts so far
<code>\insertwidth</code>	dimen	the width of the inserts

These commands take a number and an integer, dimension or glue specification. They can be set and queried but setting the dimensions can have side effects. The accumulated height of the inserts is available in `\insertheights` (which can be set too). The `\floatingpenalty` variable determines the penalty applied when a split is needed.

In the output routine the original TeX variable `\insertpenalties` is a counter that keeps the number of insertions that didn't fit on the page while otherwise it has the accumulated penalties of the split insertions. When `\holdinginserts` is non zero the inserts in the list are not collected for output, which permits the list to be fed back for reprocessing.

The LuaMetaTeX specific storage mode `\insertstoring` variable is explained in the next section.

14.4 Storing

This feature is kind of special and still experimental. When `\insertstoring` is set 1, all inserts that have their storage flag set will be saved. Think of a multi column setup where inserts have to end up in the last column. If there are three columns, the first two will store inserts. Then when the last column is dealt with `\insertstoring` can be set to 2 and that will signal the builder that we will inject the inserts. In both cases, the value of this register will be set to zero so that it doesn't influence further processing.

14.5 Synchronizing

The page builder can be triggered by (for instance) a penalty but you can also use `\pageboundary`. This will trigger the page builder but not leave anything behind. (This is experimental.)

14.6 Migration

Todo, nothing new there, so no hurry.

14.7 Callbacks

Todo, nothing new there, so no hurry.

14.7 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

15 Localboxes

low level

TEX

localboxes

Contents

15.1	Introduction	204
15.2	The basics	204
15.3	The interface	208
15.4	The helpers	213

15.1 Introduction

The LuaTeX engine inherited a few features from other engines and adding local boxes to paragraphs is one of them. This concept comes from Omega but over time it has been made a bit more robust, also by using native par nodes instead of whatsit nodes that are used to support TeX's extensions. In another low level manual we discuss paragraph properties and these local boxes are also part of that so you might want to catch up on that. Local boxes are stored in an initial par node with an adequate subtype but users won't notice this (unless they mess around in Lua). The inline par nodes have a different subtype and are injected with the `\localinterlinepenalty`, `\localbrokenpenalty`, `\localleftbox`, `\localrightbox` and LuaMetaTeX specific `\localmiddlebox` primitives. When these primitives are used in vertical mode they just set registers.

The original (Omega) idea was that local boxes are used for repetitive punctuation (like quotes) at the left and/or right end of the lines that make up a paragraph. That means that when these primitives inject nodes they actually introduce states so that a stretch of text can be marked.

When this mechanism was cleaned up in LuaMetaTeX I decided to investigate if other usage made sense. After all, it is a feature that introduces some extra code and it then pays off to use it when possible. Among the extensions are a callback that is triggered when the left and right boxes get added and experiments with that showed some potential but in order to retain performance as well as limit extensive node memory usage (par nodes are large) a system of indices was added. All this will be illustrated below. Warning: the mechanism in LuaMetaTeX is not compatible with LuaTeX.

This is a preliminary, uncorrected manual.

15.2 The basics

This mechanism uses a mix of setting (pseudo horizontal) box registers that get associated with (positions in a) paragraph. When the lines resulting from breaking the list gets packaged into an horizontal (line) box, the local left and right boxes get prepended

and appended to the textual part (inside the left, right and parfills kips and left or right hanging margins). When assigning the current local boxes to the paragraph node(s) references to the pseudo registers are used and the packaging actually copies them. This mix of referencing and copying is somewhat tricky but the engine does it best to hide this for the user.

This mechanism is rather useless when not wrapped into some high level mechanism because by default setting these boxes wipes the existing value. In LuaMetaT_EX you can actually access the boxes so prepending and appending is possible but experiments showed that this could come with a huge performance hit when the lists are not cleaned up during a run. This is why we have introduced indices: when you assign local boxes using the index option that specific index will be replaced and therefore we have a more sparse solution. So, contrary to LuaT_EX, in LuaMetaT_EX the local box registers have a linked lists of local boxes tagged by index. Unless you manipulate in Lua, this is hidden from the user. One can access the boxes from the T_EX the but there can be no confusion with LuaT_EX here because there we don't have access. This is why usage as in LuaT_EX will also work in LuaMetaT_EX.

This mechanism obeys grouping as is demonstrated in the next three examples. The first example is:

```
\start
  \dorecurse{10}{test #1.1 }
  \localleftbox{\blackrule[width=2em,color=darkred] }
  \dorecurse{20}{test #1.2 }
  \removeunwantedspaces
  \localrightbox{ \blackrule[width=3em,color=darkblue]}
  \dorecurse{20}{test #1.3 }
\stop
  \dorecurse{20}{test #1.4 }
  % par ends here
```

The next example differs in a subtle way: watch the keep keyword, it makes the setting retain after the group ends.

```
\start
  \start
    \dorecurse{10}{test #1.1 }
    \localleftbox keep {\blackrule[width=2em,color=darkred] }
    \dorecurse{20}{test #1.2 }
    \removeunwantedspaces
    \localrightbox { \blackrule[width=3em,color=darkblue]}
```

```

    \dorecurse{20}{test #1.3 }
  \stop
    \dorecurse{20}{test #1.4 }
\stop
% par ends here

```

The third example has two times keep. This option is LuaMetaTeX specific.

```

\start
  \start
    \dorecurse{10}{test #1.1 }
    \localleftbox keep { \blackrule[width=2em,color=darkred] }
    \dorecurse{20}{test #1.2 }
    \removeunwantedspaces
    \localrightbox keep { \blackrule[width=3em,color=darkblue] }
    \dorecurse{20}{test #1.3 }
  \stop
    \dorecurse{20}{test #1.4 }
\stop
% par ends here

```

One (nasty) side effect is that when you set these boxes ungrouped they are applied to whatever follows, which is why resetting them is built in the relevant parts of ConTeXt. The next examples are typeset grouped and demonstrate the use of indices:

```

\dorecurse{20}{before #1 }
\localleftbox{\bf \darkred L 1 }%
\localleftbox{\bf \darkred L 2 }%
\dorecurse{20}{after #1 }

```

before 1 before 2 before 3 before 4 before 5 before 6 before 7 before 8 before 9 before 10 before 11 before 12 before 13 before 14 before 15 before 16 before 17 before 18 before 19 before 20 after 1 after 2 after 3 after 4 after 5 after 6 after 7 after 8 after 9 **L 2** after 10 after 11 after 12 after 13 after 14 after 15 after 16 after 17 after 18 after **L 2** 19 after 20

Indices can be set for both sides:

```

\dorecurse{5}{\localrightbox index #1{ \bf \darkgreen R #1}}%
\dorecurse{20}{before #1 }
\dorecurse{5}{\localleftbox index #1{\bf \darkred L #1 }}%
\dorecurse{20}{after #1 }

```

test 1.1 test 2.1 test 3.1 test 4.1 test 5.1 test 6.1 test 7.1 test 8.1 test 9.1 test 10.1 test
 ■ 1.2 test 2.2 test 3.2 test 4.2 test 5.2 test 6.2 test 7.2 test 8.2 test 9.2 test 10.2 test
 ■ 11.2 test 12.2 test 13.2 test 14.2 test 15.2 test 16.2 test 17.2 test 18.2 test 19.2
 ■ test 20.2 test 1.3 test 2.3 test 3.3 test 4.3 test 5.3 test 6.3 test 7.3 test 8.3 ■
 ■ test 9.3 test 10.3 test 11.3 test 12.3 test 13.3 test 14.3 test 15.3 test 16.3 ■
 ■ test 17.3 test 18.3 test 19.3 test 20.3 test 1.4 test 2.4 test 3.4 test 4.4 test 5.4 test
 6.4 test 7.4 test 8.4 test 9.4 test 10.4 test 11.4 test 12.4 test 13.4 test 14.4 test 15.4 test
 16.4 test 17.4 test 18.4 test 19.4 test 20.4

Example 1

test 1.1 test 2.1 test 3.1 test 4.1 test 5.1 test 6.1 test 7.1 test 8.1 test 9.1 test 10.1 test
 ■ 1.2 test 2.2 test 3.2 test 4.2 test 5.2 test 6.2 test 7.2 test 8.2 test 9.2 test 10.2 test
 ■ 11.2 test 12.2 test 13.2 test 14.2 test 15.2 test 16.2 test 17.2 test 18.2 test 19.2
 ■ test 20.2 test 1.3 test 2.3 test 3.3 test 4.3 test 5.3 test 6.3 test 7.3 test 8.3 ■
 ■ test 9.3 test 10.3 test 11.3 test 12.3 test 13.3 test 14.3 test 15.3 test 16.3 ■
 ■ test 17.3 test 18.3 test 19.3 test 20.3 test 1.4 test 2.4 test 3.4 test 4.4 test 5.4 test
 6.4 test 7.4 test 8.4 test 9.4 test 10.4 test 11.4 test 12.4 test 13.4 test 14.4 test 15.4 test
 16.4 test 17.4 test 18.4 test 19.4 test 20.4

Example 2

test 1.1 test 2.1 test 3.1 test 4.1 test 5.1 test 6.1 test 7.1 test 8.1 test 9.1 test 10.1 test
 ■ 1.2 test 2.2 test 3.2 test 4.2 test 5.2 test 6.2 test 7.2 test 8.2 test 9.2 test 10.2 test
 ■ 11.2 test 12.2 test 13.2 test 14.2 test 15.2 test 16.2 test 17.2 test 18.2 test 19.2 test
 ■ 20.2 test 1.3 test 2.3 test 3.3 test 4.3 test 5.3 test 6.3 test 7.3 test 8.3 test ■
 ■ 9.3 test 10.3 test 11.3 test 12.3 test 13.3 test 14.3 test 15.3 test 16.3 test ■
 ■ 17.3 test 18.3 test 19.3 test 20.3 test 1.4 test 2.4 test 3.4 test 4.4 test 5.4 ■
 ■ test 6.4 test 7.4 test 8.4 test 9.4 test 10.4 test 11.4 test 12.4 test 13.4 test ■
 ■ 14.4 test 15.4 test 16.4 test 17.4 test 18.4 test 19.4 test 20.4 ■

Example 3

Figure 15.1

before 1 before 2 before 3 before 4 before 5 before 6 before 7 **R 1 R 2 R 3 R 4 R 5**
 before 8 before 9 before 10 before 11 before 12 before 13 before **R 1 R 2 R 3 R 4 R 5**
 14 before 15 before 16 before 17 before 18 before 19 before 20 **R 1 R 2 R 3 R 4 R 5**
 after 1 after 2 after 3 after 4 after 5 after 6 after 7 after 8 after **R 1 R 2 R 3 R 4 R 5**
L 1 L 2 L 3 L 4 L 5 9 after 10 after 11 after 12 after 13 after 14 **R 1 R 2 R 3 R 4 R 5**
L 1 L 2 L 3 L 4 L 5 after 15 after 16 after 17 after 18 after 19 **R 1 R 2 R 3 R 4 R 5**
L 1 L 2 L 3 L 4 L 5 after 20 **R 1 R 2 R 3 R 4 R 5**

We can instruct this mechanism to hook the local box into the main par node by using the par keyword. Keep in mind that these local boxes only come into play when the

lines are broken, so till then changing them is possible.

```
\dorecurse{3}{\localrightbox index #1{ \bf \darkgreen R #1}}%
\dorecurse{20}{before #1 }
\dorecurse{2}{\localleftbox par index #1{\bf \darkred L #1 }}%
\dorecurse{20}{after #1 }
```

L 1 L 2 before 1 before 2 before 3 before 4 before 5 before 6 before 7 before **R 1 R 2 R 3**
L 1 L 2 8 before 9 before 10 before 11 before 12 before 13 before 14 before **R 1 R 2 R 3**
L 1 L 2 15 before 16 before 17 before 18 before 19 before 20 after 1 after 2 **R 1 R 2 R 3**
L 1 L 2 after 3 after 4 after 5 after 6 after 7 after 8 after 9 after 10 after 11 **R 1 R 2 R 3**
L 1 L 2 after 12 after 13 after 14 after 15 after 16 after 17 after 18 after 19 **R 1 R 2 R 3**
L 1 L 2 after 20 **R 1 R 2 R 3**

15.3 The interface

The interface described here is experimental.

Because it is hard to foresee if this mechanism will be used at all the ConT_EXt interface is somewhat low level: one can build functionality on top of it. In the previous section we saw examples of local boxes being part of the text but one reason for extending the interface was to see if we can also use this engine feature for efficiently placing marginal content.

```
\definelocalboxes
 [lefttext]
 [location=lefttext,width=3em,color=darkblue]
\definelocalboxes
 [lefttextx]
 [location=lefttext,width=3em,color=darkblue]

\definelocalboxes
 [righttext]
 [location=righttext,width=3em,color=darkyellow]
\definelocalboxes
 [righttextx]
 [location=righttext,width=3em,color=darkyellow]
```

The order of definition matters! Here the x variants have a larger index number. There can (currently) be at most 256 indices. The defined local boxes are triggered with `\localbox`:

```

\startnarrower
\dorecurse{20}{before #1 }%
\localbox[lefttext]{[L] }%
\localbox[lefttextx]{[LL] }%
\localbox[righttext]{ [RR]}%
\localbox[righttextx]{ [R]}%
\dorecurse{20}{ after #1}%
\stopnarrower

```

Watch how we obey the margins:

```

before 1 before 2 before 3 before 4 before 5 before 6 before 7 before 8 before 9
before 10 before 11 before 12 before 13 before 14 before 15 before 16 before 17
before 18 before 19 before 20 after 1 after 2 after 3 after 4 after 5 after [RR] [R]
[L] [LL] 6 after 7 after 8 after 9 after 10 after 11 after 12 after 13 after [RR] [R]
[L] [LL] 14 after 15 after 16 after 17 after 18 after 19 after 20 [RR] [R]

```

Here these local boxes have dimensions. The predefined margin variants are virtual. Here we set up the style and color:

```

\setuplocalboxes
[leftmargin]
[style=\bs,
color=darkgreen]
\setuplocalboxes
[rightmargin]
[style=\bs,
color=darkred]

\dorecurse{2}{
\dorecurse{10}{some text #1.##1 }%
KEY#1.1%
\localmargintext[leftmargin]{L #1.1}%
\localmargintext[rightmargin]{R #1.1}%
\dorecurse{10}{some text #1.##1 }%
KEY#1.2%
\localmargintext[leftmargin]{L #1.2}%
\localmargintext[rightmargin]{R #1.2}%
\dorecurse{10}{some text #1.##1 }%
\blank
}

```

You can also use `leftedge` and `rightedge` but using them here would put them outside the page.

L 1.1 some text 1.1 some text 1.2 some text 1.3 some text 1.4 some text 1.5 some text 1.6
 some text 1.7 some text 1.8 some text 1.9 some text 1.10 KEY1.1some text 1.1 some **R 1.1**
 text 1.2 some text 1.3 some text 1.4 some text 1.5 some text 1.6 some text 1.7 some

L 1.2 text 1.8 some text 1.9 some text 1.10 KEY1.2some text 1.1 some text 1.2 some text 1.3 **R 1.2**
 some text 1.4 some text 1.5 some text 1.6 some text 1.7 some text 1.8 some text 1.9
 some text 1.10

L 2.1 some text 2.1 some text 2.2 some text 2.3 some text 2.4 some text 2.5 some text 2.6
 some text 2.7 some text 2.8 some text 2.9 some text 2.10 KEY2.1some text 2.1 some **R 2.1**
 text 2.2 some text 2.3 some text 2.4 some text 2.5 some text 2.6 some text 2.7 some

L 2.2 text 2.8 some text 2.9 some text 2.10 KEY2.2some text 2.1 some text 2.2 some text 2.3 **R 2.2**
 some text 2.4 some text 2.5 some text 2.6 some text 2.7 some text 2.8 some text 2.9
 some text 2.10

In previous examples you can see that setting something at the left will lag behind so deep down we use another trick here: `\localmiddlebox`. When these boxes get placed a callback can be triggered and in `ConTeXt` we use that to move these middle boxes to the margins.

Next we implement line numbers. Watch out: this will not replace the existing mechanisms, it's just an alternative as we have alternative table mechanisms. We have a repertoire of helpers for constructing the result:

```
\definelocalboxes
[linenumberleft]
[command=\LeftNumber,
location=middle,
distance=\leftmargindistance,
width=3em,
style=\bs,
color=darkred]

\definelocalboxes
[linenumberright] % [linenumberleft]
[command=\RightNumber,
location=middle,
distance=\rightmargindistance,
width=3em,
style=\bf,
```

```

color=darkgreen]

\definecounter[MyLineNumberL]
\definecounter[MyLineNumberR]

\setupcounter
  [MyLineNumberL]
  [numberconversion=characters]

\setupcounter
  [MyLineNumberR]
  [numberconversion=romannumerals]

\def\LineNumberL
  {\incrementcounter[MyLineNumberL]%
   \convertedcounter[MyLineNumberL]}

\def\LineNumberR
  {\incrementcounter[MyLineNumberR]%
   \convertedcounter[MyLineNumberR]}

\protected\def\LeftNumber
  {\setbox\localboxcontentbox\hbox
   to \localboxesparameter{width}
   {(\LineNumberL\hss\strut)}%
   \localmarginlefttext\zeropoint}

\protected\def\RightNumber
  {\setbox\localboxcontentbox\hbox
   to \localboxesparameter{width}
   {(\strut\hss\LineNumberR)}%
   \localmarginrighttext\zeropoint}

\localbox[linenumberleft]{}%
\localbox[linenumberright]{}%
\dorecurse{2}{
  \samplefile{tufte}
  \par
}
\resetlocalbox[linenumberleft]%
\resetlocalbox[linenumberright]%

```

We use our tufte example to illustrate the usage:

- (15.d)** We thrive in information–thick worlds because of our marvelous and everyday capacity **(15.i)**
(15.h) to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe- **(15.ii)**
(15.d) size, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, **(15.iii)**
(15.d) list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeon- **(15.iv)**
(15.e) hole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, aver- **(15.v)**
(15.f) age, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip **(15.vi)**
(15.g) through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, **(16.vii)**
(15.h) winnow the wheat from the chaff and separate the sheep from the goats. **(15.viii)**
- (15.i)** We thrive in information–thick worlds because of our marvelous and everyday capacity **(15.ix)**
(15.j) to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe- **(15.x)**
(15.k) size, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, **(15.xi)**
(15.l) list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeon- **(16.xii)**
(15.m) hole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, aver- **(15.xiii)**
(15.n) age, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip **(15.xiv)**
(15.o) through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, **(15.xv)**
(15.p) winnow the wheat from the chaff and separate the sheep from the goats. **(15.xvi)**

For convenience we support ranges like this (we've reset the line number counters here):

```
\startlocalboxrange[linenumberleft]%
\startlocalboxrange[linenumberright]%
\dorecurse{2}{
  \samplefile{tufte}
  \par
}
\stoplocalboxrange
\stoplocalboxrange
```

- (15.d)** We thrive in information–thick worlds because of our marvelous and everyday capacity **(15.i)**
(15.h) to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe- **(15.ii)**
(15.d) size, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, **(15.iii)**
(15.d) list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeon- **(15.iv)**
(15.e) hole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, aver- **(15.v)**
(15.f) age, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip **(15.vi)**
(15.g) through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, **(16.vii)**
(15.h) winnow the wheat from the chaff and separate the sheep from the goats. **(15.viii)**
- (15.i)** We thrive in information–thick worlds because of our marvelous and everyday capacity **(15.ix)**
(15.j) to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthe- **(15.x)**
(15.k) size, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, **(15.xi)**

- (15.l)** list, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, pigeon- **16.xii)**
(15.m) hole, pick over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, aver- **15.xiii)**
(15.n) age, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip **15.xiv)**
(15.o) through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, **15.xv)**
(15.p) winnow the wheat from the chaff and separate the sheep from the goats. **15.xvi)**

15.4 The helpers

For the moment we have these helpers:

<code>\localboxindex</code>	integer
<code>\localboxlinenumber</code>	integer
<code>\localboxlinewidth</code>	dimension
<code>\localboxlocalwidth</code>	dimension
<code>\localboxprogress</code>	dimension
<code>\localboxleftoffset</code>	dimension
<code>\localboxrightoffset</code>	dimension
<code>\localboxleftskip</code>	dimension
<code>\localboxrightskip</code>	dimension
<code>\localboxlefthang</code>	dimension
<code>\localboxrighthang</code>	dimension
<code>\localboxindent</code>	dimension
<code>\localboxparfillleftskip</code>	dimension
<code>\localboxparfillrightskip</code>	dimension
<code>\localboxovershoot</code>	dimension

The progress and offsets are accumulated values of the normalized indent, hangs, skips etc. The line number is the position in the paragraph. In the callback we set the box register `\localboxcontentbox` and use it after the command has been applied. In the line number example you can see how we set its final content, so these boxes are sort of dynamic. Normally in the middle case no content is passed and in the par builder a middle is not taken into account when calculating the line width.

15.4 Colofon

Author Hans Hagen
ConT_EXt 2023.04.27 17:04
LuaMetaT_EX 2.1008
Support www.pragma-ade.com
 contextgarden.net

16 Loops

low level

TEX

loops

Contents

16.1 Introduction	216
16.2 Primitives	216
16.3 Wrappers	220

16.1 Introduction

I have hesitated long before I finally decided to implement native loops in LuaMetaT_EX. Among the reasons against such a feature is that one can define macros that do loops (preferably using tail recursion). When you don't need an expandable loop, counters can be used, otherwise there are dirty and obscure tricks that can be of help. This is often the area where tex programmers can show off but the fact remains that we're using side effects of the expansion machinery and specific primitives like `\romannumeral` magic. In LuaMetaT_EX it is actually possible to use the local control mechanism to hide loop counter advance and checking but that comes with at a performance hit. And, no matter what tricks one used, tracing becomes pretty much cluttered.

In the next sections we describe the new native loop primitives in LuaMetaT_EX as well as the more traditional ConT_EXt loop helpers.

16.2 Primitives

Because MetaPost, which is also a macro language, has native loops, it makes sense to also have native loops in T_EX and in LuaMetaT_EX it was not that hard to add it. One variant uses the local control mechanism which is reflected in its name and two others collect expanded bodies. In the local loop content gets injected as we go, so this one doesn't work well in for instance an `\edef`. The macro takes the usual three loop numbers as well as a token list:

```
\localcontrolledloop 1 100000 1 {%
    % body
}
```

Here is an example of usage:

```
\localcontrolledloop 1 5 1 {%
    [\number\currentloopiterator]
    \localcontrolledloop 1 10 1 {%
        (\number\currentloopiterator)
    }
```

```

    }%
    [\number\currentloopiterator]
  \par
}

```

The `\currentloopiterator` is a numeric token so you need to explicitly serialize it with `\number` or `\the` if you want it to be typeset:

```

[1] (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) [1]
[2] (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) [2]
[3] (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) [3]
[4] (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) [4]
[5] (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) [5]

```

Here is another example. This time we also show the current nesting:

```

\localcontrolledloop 1 100 1 {%
  \ifnum\currentloopiterator>6\relax
    \quitloop
  \else
    [\number\currentloopnesting:\number\currentloopiterator]
    \localcontrolledloop 1 8 1 {%
      (\number\currentloopnesting:\number\currentloopiterator)
    }\par
  \fi
}

```

Watch the `\quitloop`: it will end the loop at the *next* iteration so any content after it will show up. Normally this one will be issued in a condition and we want to end that properly.

```

[1:1] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)
[1:2] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)
[1:3] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)
[1:4] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)
[1:5] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)
[1:6] (2:1) (2:2) (2:3) (2:4) (2:5) (2:6) (2:7) (2:8)

```

The three loop variants all perform differently:

```

l:\testfeatureonce {1000} {\localcontrolledloop 1 2000 1 {\relax}} %
  \elapsedtime
e:\testfeatureonce {1000} {\expandedloop 1 2000 1 {\relax}} %

```

```

\elapsedtime
u:\testfeatureonce {1000} {\unexpandedloop 1 2000 1 {\relax}} %
\elapsedtime

```

An unexpanded loop is (of course) the fastest because it only collects and then feeds back the lot. In an expanded loop each cycle does an expansion of the body and collects the result which is then injected afterwards, and the controlled loop just expands the body each iteration.

```

l: 0.082
e: 0.070
u: 0.023

```

The different behavior is best illustrated with the following example:

```

\edef\TestA{\localcontrolledloop 1 5 1 {A}} % out of order
\edef\TestB{\expandedloop 1 5 1 {B}}
\edef\TestC{\unexpandedloop 1 5 1 {C\relax}}

```

We can show the effective definition:

```

\meaningasis\TestA
\meaningasis\TestB
\meaningasis\TestC

A: \TestA
B: \TestB
C: \TestC

```

Watch how the first test pushes the content in the main input stream:

```

AAAAA
\def \TestA
\def \TestB BBBB
\def \TestC C\relax C\relax C\relax C\relax C\relax

A:
B: BBBB
C: CCCC

```

Here are some examples that show what gets expanded and what not:

```

\edef\whatever
{\expandedloop 1 10 1

```

```
{(\number\currentloopiterator)
\scratchcounter=\number\currentloopiterator\relax}}
```

```
\meaningasis\whatever
```

```
\def \whatever (1) \scratchcounter =1\relax (2) \scratchcounter =2\relax (3) \scratchcounter
=3\relax (4) \scratchcounter =4\relax (5) \scratchcounter =5\relax (6) \scratchcounter
=6\relax (7) \scratchcounter =7\relax (8) \scratchcounter =8\relax (9) \scratchcounter
=9\relax (10) \scratchcounter =10\relax
```

A local control encapsulation hides the assignment:

```
\edef\whatever
{\expandedloop 1 10 1
{(\number\currentloopiterator)
\beginlocalcontrol
\scratchcounter=\number\currentloopiterator\relax
\endlocalcontrol}}
```

```
\meaningasis\whatever
```

```
\def \whatever (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
```

Here we see the assignment being retained but with changing values:

```
\edef\whatever
{\unexpandedloop 1 10 1
{\scratchcounter=1\relax}}
```

```
\meaningasis\whatever
```

```
\def \whatever \scratchcounter =1\relax \scratchcounter =1\relax \scratchcounter =1\relax
\scratchcounter =1\relax \scratchcounter =1\relax \scratchcounter =1\relax \scratchcounter
=1\relax \scratchcounter =1\relax \scratchcounter =1\relax \scratchcounter =1\relax
```

We get no expansion at all:

```
\edef\whatever
{\unexpandedloop 1 10 1
{\scratchcounter=\the\currentloopiterator\relax}}
```

```
\meaningasis\whatever
```

```
\def \whatever \scratchcounter =0\relax \scratchcounter =0\relax \scratchcounter =0\relax
\scratchcounter =0\relax \scratchcounter =0\relax \scratchcounter =0\relax \scratchcounter
=0\relax \scratchcounter =0\relax \scratchcounter =0\relax \scratchcounter =0\relax
```

And here we have a mix:

```
\edef\whatever
  {\expandedloop 1 10 1
   {\scratchcounter=\the\currentloopiterator\relax}}
```

```
\meaningasis\whatever
```

```
\def \whatever \scratchcounter =1\relax \scratchcounter =2\relax \scratchcounter =3\relax
\scratchcounter =4\relax \scratchcounter =5\relax \scratchcounter =6\relax \scratchcounter
=7\relax \scratchcounter =8\relax \scratchcounter =9\relax \scratchcounter =10\relax
```

There is one feature worth noting. When you feed three numbers in a row, like here, there is a danger of them being seen as one:

```
\expandedloop
  \number\dimexpr1pt
  \number\dimexpr2pt
  \number\dimexpr1pt
  {}
```

This gives an error because a too large number is seen. Therefore, these loops permit leading equal signs, as in assignments (we could support keywords but it doesn't make much sense):

```
\expandedloop =\number\dimexpr1pt =\number\dimexpr2pt =\number\dimexpr1pt{}
```

16.3 Wrappers

We always had loop helpers in ConT_EXt and the question is: “What we will gain when we replace the definitions with ones using the above?”. The answer is: “We have little performance but not as much as one expects!”. This has to do with the fact that we support #1 as iterator and #2 as (verbose) nesting values and that comes with some overhead. It is also the reason why these loop macros are protected (unexpandable). However, using the primitives might look somewhat more natural in low level T_EX code.

Also, replacing their definitions can have side effects because the primitives are (and will be) still experimental so it's typically a patch that I will run on my machine for a while.

Here is an example of two loops. The inner state variables have one hash, the outer one extra:

```

\dorecurese{2}{
  \dostepwiserecurese{1}{10}{2}{
    (#1:#2) [##1:##2]
  }\par
}

```

We get this:

```

(1:1) [1:2] (1:1) [3:2] (1:1) [5:2] (1:1) [7:2] (1:1) [9:2]
(2:1) [1:2] (2:1) [3:2] (2:1) [5:2] (2:1) [7:2] (2:1) [9:2]

```

We can also use two state macro but here we would have to store the outer ones:

```

\dorecurese {2} {
  /\recursedepth:\recurselevel/
  \dostepwiserecurese {1} {10} {2} {
    <\recursedepth:\recurselevel>
  }\par
}

```

That gives us:

```

/1:1/ <2:1> <2:3> <2:5> <2:7> <2:9>
/1:2/ <2:1> <2:3> <2:5> <2:7> <2:9>

```

An endless loop works as follows:

```

\doloop {
  ...
  \ifsomeconditionismet
    ...
    \exitloop
  \else
    ...
  \fi
  % \exitloopnow
  ...
}

```

Because of the way we quit there will not be a new implementation in terms of the loop primitives. You need to make sure that you don't leave in the middle of an ongoing condition. The second exit is immediate.

We also have a (simple) expanded variant:

```
\edef\TestX{\doexpandedrecurse{10}{!}} \meaningasis\TestX
```

This helper can be implemented in terms of the loop primitives which makes them a bit faster, but these are not critical:

```
\def\TestX !!!!!!!!!!
```

A variant that supports #1 is the following:

```
\edef\TestX{\doexpandedrecursed{10}{#1}} \meaningasis\TestX
```

So:

```
\def\TestX 12345678910
```

16.3 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net

17 Tokens

low level

TEX

tokens

Contents

17.1	Introduction	224
17.2	What are tokens	224
17.3	Some implementation details	228
17.4	Other data management	229
17.5	Macros	230
17.6	Looking at tokens	230

17.1 Introduction

Most users don't need to know anything about tokens but it happens that when T_EXies meet in person (users group meetings), or online (support platforms) there always seem to pop up folks who love token speak. When you try to explain something to a user it makes sense to talk in terms of characters but then those token speakers can jump in and start correcting you. In the past I have been puzzled by this because, when one can write a decent macro that does the job well, it really doesn't matter if one knows about tokens. Of course one should never make the assumption that token speakers really know T_EX that well or can come up with better solutions than users but that is another matter.²³

That said, because in documents about T_EX the word 'token' does pop up I will try to give a little insight here. But for using T_EX it's mostly irrelevant. The descriptions below for sure won't match the proper token speak criteria which is why at a presentation for the 2020 user meeting I used the title "Tokens as I see them."

17.2 What are tokens

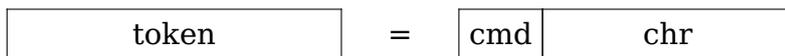
Both the words 'node' and 'token' are quite common in programming and also rather old which is proven by the fact that they also are used in the T_EX source. A node is a storage container that is part of a linked list. When you input the characters tex the three characters become part of the current linked list. They become 'character' nodes (or in LuaT_EX speak 'glyph' nodes) with properties like the font and the character referred to. But before that happens, the three characters in the input t, e and x, are interpreted as in this case being just that: characters. When you enter \TeX the input processors first sees a backslash and because that has a special meaning in T_EX it will

²³ Talking about fashion: it would be more impressive to talk about T_EX and friends as a software stack than calling it a distribution. Today, it's all about marketing.

read following characters and when done does a lookup in it's internal hash table to see what it actually is: a macro that assembled the word \TeX in uppercase with special kerning and a shifted (therefore boxed) 'E'. When you enter $\$ \TeX$ will look ahead for a second one in order to determine display math, push back the found token when there is no match and then enter inline math mode.

A token is internally just a 32 bit number that encodes what \TeX has seen. It is the assembled token that travels through the system, get stored, interpreted and often discarded afterwards. So, the character 'e' in our example gets tagged as such and encoded in this number in a way that the intention can be derived later on.

Now, the way \TeX looks at these tokens can differ. In some cases it will just look at this (32 bit) number, for instance when checking for a specific token, which is fast, but sometimes it needs to know some detail. The mentioned integer actually encodes a command (opcode) and a so called char code (operand). The second name is somewhat confusing because in many cases that code is not representing a character but that is not that relevant here. When you look at the source code of a \TeX engine it is enough to know that a char can also be a sub command.



Back to the three characters: these become tokens where the command code indicates that it is a letter and the char code stores what letter we have at hand and in the case of $\text{Lua}\TeX$ and $\text{LuaMeta}\TeX$ these are Unicode values. Contrary to the traditional 8 bit \TeX engine, in the Unicode engines an utf sequence is read, but these multiple bytes still become one number that will be encoded in the token number. In order to determine that something is a letter the engine has to be told (which is what a macro package does when it sets up the engine). For instance, digits are so called other characters and the backslash is called escape. Every \TeX user knows that curly braces are special and so are dollar symbols and hashes. If this rings a bell, and you relate this to catcodes, you can indeed assume that the command codes of these tokens have the same numbers as the catcodes. Given that Unicode has plenty of characters slots you can imagine that combining 16 catcode commands with all the possible Unicode values makes a large repertoire of tokens.

There are more commands than the 16 basic characters related ones, in $\text{LuaMeta}\TeX$ we have just over 150 command codes ($\text{Lua}\TeX$ has a few more but they are also organized differently). Each of these codes can have a sub command, For instance the primitives $\backslash\text{vbox}$ and $\backslash\text{hbox}$ are both a `make_box_cmd` (we use the symbolic name here) and in $\text{LuaMeta}\TeX$ the first one has sub command code 9 (`vbox_code`) and the second one has code 10 (`hbox_code`). There are twelve primitives that are in the same category.

What are tokens

The many primitives that make up the core of the engine are grouped in a way that permits processing similar ones with one function and also makes it possible to distinguish between the way commands are handled, for instance with respect to expansion.

Now, before we move on it is important to know that all these codes are in fact abstract numbers. Although it is quite likely that engines that are derived from each other have similar numbers (just more) this is not the case for LuaMetaT_EX. Because the internals have been opened up (even more than in LuaT_EX) the command and char codes have been reorganized in a such a way that exposure is consistent. We could not use some of the reuse and remap tricks that the other engines use because it would simply be too confusing (and demand real in depth knowledge of the internals). This is also the reason why development took some time. You probably won't notice it from the current source but it was a very stepwise process. We not only had to make sure that all kept working (ConT_EXt LMTX and LuaMetaT_EX were pretty useable during the process), but also had to (re)consider intermediate choices.

So, input is converted into tokens, in most cases one-by-one. When a token is assembled, it either gets stored (deliberately or as part of some look ahead scanning), or it immediately gets (what is called:) expanded. Depending on what the command is, some action is triggered. For instance, a character gets appended to the node list immediately. An `\hbox` command will start assembling a box which its own node list that then gets some treatment: if this primitive was a follow up on `\setbox` it will get stored, otherwise it might end up in the current node list as so called hlist node. Commands that relate to registers have `0xFFFF` char codes because that is how many registers we have per category.

When a token gets stored for later processing it becomes part of a larger data structure, a so called memory word. These memory words are taken from a large pool of words and they store a token and additional properties. The info field contains the token value, the mentioned command and char. When there is no linked list, the link can actually be used to store a value, something that in LuaMetaT_EX we actually do.

1	info	link
2	info	link
3	info	link
n	info	link

When for instance we say `\toks 0 {tex}` the scanner sees an escape, followed by 4 letters (toks) and the escape triggers a lookup of the primitive (or macro or ...) with that name, in this case a primitive assignment command. The found primitive (its property gets stored in the token) triggers scanning for a number and when that is successful

scanning of a brace delimited token list starts. The three characters become three letter tokens and these are a linked list of the mentioned memory words. This list then gets stored in token register zero. The input sequence `\the\toks 0` will push back a copy of this list into the input.

In addition to the token memory pool, there is also a table of equivalents. That one is part of a larger table of memory words where \TeX stores all it needs to store. The 16 groups of character commands are virtual, storing these makes no sense, so the first real entries are all these registers (count, dimension, skip, box, etc). The rest is taken up by possible hash entries.

main hash	null control sequence
	128K hash entries
	frozen control sequences
	special sequences (undefined)
registers	17 internal & 64K user glues
	4 internal & 64K user mu glues
	12 internal & 64K user tokens
	2 internal & 64K user boxes
	116 internal & 64K user integers
	0 internal & 64K user attribute
	22 internal & 64K user dimensions
specifications	5 internal & 0 user
extra hash	additional entries (grows dynamic)

So, a letter token `t` is just that, a token. A token referring to a register is again just a number, but its char code points to a slot in the equivalents table. A macro, which we haven't discussed yet, is actually just a token list. When a name lookup happens the hash table is consulted and that one runs in parallel to part of the table of equivalents. When there is a match, the corresponding entry in the equivalents table points to a token list.

1	string index	equivalents or (next > n) index
2	string index	equivalents or (next > n) index
n	string index	equivalents or (next > n) index
n + 1	string index	equivalents or (next > n) index
n + 2	string index	equivalents or (next > n) index
n + m	string index	equivalents or (next > n) index

It sounds complex and it actually also is somewhat complex. It is not made easier by the fact that we also track information related to grouping (saving and restoring), need reference counts for copies of macros and token lists, sometimes store information directly instead of via links to token lists, etc. And again one cannot compare LuaMetaTeX with the other engines. Because we did away with some of the limitations of the traditional engine we not only could save some memory but in the end also simplify matters (we're 32/64 bit after all). On the one hand some traditional speedups were removed but these have been compensated by improvements elsewhere, so overall processing is more efficient.

1	level	type	flag	value
2	level	type	flag	value
3	level	type	flag	value
n	level	type	flag	value

So, here LuaMetaTeX differs from other engines because it combines two tables, which is possible because we have at least 32 bits. There are at most 0xFFFF levels but we need at most 0xFF types. In LuaMetaTeX macros can have extra properties (flags) and these also need one byte. Contrary to the other engines, `\protected` macros are native and have their own command code, but `\tolerant` macros duplicate that (so we have four distinct macro commands). All other properties, like the `\permanent` ones are stored in the flags.

Because a macro starts with a reference count we have some room in the info field to store information about it having arguments or not. It is these details that make LuaMetaTeX a bit more efficient in terms of memory usage and performance than its ancestor LuaTeX. But as with the other changes, it was a very stepwise process in order to keep the system compatible and working.

17.3 Some implementation details

Sometimes there is a special head token at the start. This makes for easier appending of extra tokens. In traditional TeX node lists are forward linked, in LuaTeX they are

double linked²⁴. Token lists are always forward linked. Shared token lists use the head node for a reference count.

For various reasons original \TeX uses global variables temporary lists. This is for instance needed when we expand (nested) and need to report issues. But in $\text{Lua}\TeX$ we often just serialize lists and using local variables makes more sense. One of the first things done in $\text{LuaMeta}\TeX$ was to group all global variables in (still global) structures but well isolated. That also made it possible to actually get rid of some globals.

Because \TeX had to run on machines that we nowadays consider rather limited, it had to be sparse and efficient. There are quite some optimizations to limit code and memory consumption. The engine also does its own memory management. Freed token memory words are collected in a cache and reused but they can get scattered which is not that bad, apart from maybe cache hits. In $\text{LuaMeta}\TeX$ we stay as close to original \TeX as possible but there have been some improvements. The Lua interfaces force us to occasionally divert from the original, and that in fact might lead to some retrofit but the original documentation still mostly applies. However, keep in mind that in $\text{Lua}\TeX$ we store much more in nodes (each has a prev pointer and an attribute list pointer and for instance glyph nodes have some 20 extra fields compared to traditional \TeX character nodes).

17.4 Other data management

There is plenty going on in \TeX when it processes your input, just to mention a few:

- Grouping is handled by a nesting stack.
- Nested conditionals (`\if . . .`) have their own stack.
- The values before assignments are saved on the save stack.
- Also other local changes (housekeeping) ends up in the save stack.
- Token lists and macro aliases have references pointers (reuse).
- Attributes, being linked node lists, have their own management.

In all these subsystems tokens or references to tokens can play a role. Reading a single character from the input can trigger a lot of action. A curly brace tagged as begin group command will push the grouping level and from then on registers and some other quantities that are changed will be stored on the save stack so that after the group ends they can be restored. When primitives take keywords, and no match happens, tokens are pushed back into the input which introduces a new input level (also some stack).

²⁴ On the agenda of $\text{LuaMeta}\TeX$ is to use this property in the underlying code, that doesn't yet profit from this and therefore keep previous pointers in store.

When numbers are read a token that represents no digit is pushed back too and macro packages use numbers and dimensions a lot. It is a surprise that T_EX is so fast.

17.5 Macros

There is a distinction between primitives, the build in commands, and macros, the commands defined by users. A primitive relates to a command code and char code but macros are, unless they are made an alias to something else, like a `\countdef` or `\let` does, basically pointers to a token list. There is some additional data stored that makes it possible to parse and grab arguments.

When we have a control sequence (macro) `\controlsequence` the name is looked up in the hash table. When found its value will point to the table of equivalents. As mentioned, that table keeps track of the cmd and points to a token list (the meaning). We saw that this table also stores the current level of grouping and flags.

If we say, in the input, `\hbox to 10pt {x\hss}`, the box is assembled as we go and when it is appended to the current node list there are no tokens left. When scanning this, the engine literally sees a backslash and the four letters `hbox`. However when we have this:

```
\def\MyMacro{\hbox to 10pt {x\hss}}
```

the `\hbox` has become one memory word which has a token representing the `\hbox` primitive plus a link to the next token. The space after a control sequence is gobbled so the next two tokens, again stored in a linked memory word, are letter tokens, followed by two other and two letter tokens for the dimensions. Then we have a space, a brace, a letter, a primitive and a brace. The about 20 characters in the input became a dozen memory words each two times four bytes, so in terms of memory usage we end up with quite a bit more. However, when T_EX runs over that list it only has to interpret the token values because the scanning and conversion already happened. So, the space that a macro takes is more than compensated by efficient reprocessing.

17.6 Looking at tokens

When you say `\tracingall` you will see what the engine does: read input, expand primitives and macros, typesetting etc. You might need to set `\tracingonline` to get a bit more output on the console. One way to look at macros is to use the `\meaning` command, so if we have:

```
\permanent\protected\def\MyMacro#1#2{Do #1 or #2!}
```

we can say this:

```

\meaning    \MyMacro
\meaningless\MyMacro
\meaningfull\MyMacro

```

and get:

```

protected macro:#1#2->Do #1 or #2!
#1#2->Do #1 or #2!
permanent protected macro:#1#2->Do #1 or #2!

```

You get less when you ask for the meaning of a primitive, just its name. The `\meaningfull` primitive gives the most information. In LuaMetaTeX protected macros are first class commands: they have their own command code. In the other engines they are just regular macros with an initial token indicating that they are protected. There are specific command codes for `\outer` and `\long` macros but we dropped that in LuaMetaTeX. Instead we have `\tolerant` macros but that's another story. The flags that were mentioned can mark macros in a way that permits overload protection as well as some special treatment in otherwise tricky cases (like alignments). The overload related flags permits a rather granular way to prevent users from redefining macros and such. They are set via prefixes, and add to that repertoire: we have 14 prefixes but only some eight deal with flags (we can add more if really needed). The probably most well known prefix is `\global` and that one will not become a flag: it has immediate effect.

For the above definition, the `\showluatokens` command will show a meaning on the console.

```
\showluatokens\MyMacro
```

This gives the next list, where the first column is the address of the token, the second one the command code, and the third one the char code. When there are arguments involved, the list of what needs to get matched is shown.

```

permanent protected control sequence: MyMacro
501263  19   49  match                argument 1
501087  19   50  match                argument 2
385528  20    0  end match
-----
501090  11   68  letter              D (U+00044)
 30833  11  111  letter              o (U+0006F)
500776  10   32  spacer
385540  21    1  parameter reference
112057  10   32  spacer

```

```

431886 11 111 letter          o (U+0006F)
 30830 11 114 letter          r (U+00072)
 30805 10  32 spacer
500787 21   2 parameter reference
213412 12  33 other char      ! (U+00021)

```

In the next subsections I will give some examples. This time we use helper defined in a module:

```
\usemodule[system-tokens]
```

17.6.1 Example 1: in the input

```
\luatokenable{1 \bf{2} 3\what {!}}
```

given token list:

547274	12	49	other char	1	U+00031	
547292	10	32	spacer			
199105	138	0	protected call			bf
548230	1	123	left brace			
543836	12	50	other char	2	U+00032	
546118	2	125	right brace			
547402	10	32	spacer			
546131	12	51	other char	3	U+00033	
413323	125	0	undefined cs			what
545819	1	123	left brace			
546134	12	33	other char	!	U+00021	
32660	2	125	right brace			

17.6.2 Example 2: in the input

```
\luatokenable{a \the\scratchcounter b \the\parindent \hbox to 10pt{x}}
```

given token list:

211403	11	97	letter	a	U+00061	
546626	10	32	spacer			
547409	135	0	the			the
540386	108	1026	integer			scratchcounter
543850	11	98	letter	b	U+00062	
546087	10	32	spacer			
546261	135	0	the			the
539360	90	0	internal dimen			parindent
547309	30	14	make box			hbox
544093	11	116	letter	t	U+00074	
545711	11	111	letter	o	U+0006F	
546526	10	32	spacer			

546897	12	49	other char	1	U+00031
545982	12	48	other char	0	U+00030
547047	11	112	letter	p	U+00070
545526	11	116	letter	t	U+00074
547434	1	123	left brace		
544780	11	120	letter	x	U+00078
544305	2	125	right brace		

17.6.3 Example 3: user registers

```
\scratchtoks{foo \framed{\red 123}456}
```

```
\luatokenable\scratchtoks
```

token register: scratchtoks

483734	11	102	letter	f	U+00066	
546873	11	111	letter	o	U+0006F	
545867	11	111	letter	o	U+0006F	
545799	10	32	spacer			
543179	141	0	tolerant protected call			framed
545170	1	123	left brace			
546396	138	0	protected call			red
546436	12	49	other char	1	U+00031	
547499	12	50	other char	2	U+00032	
544840	12	51	other char	3	U+00033	
545370	2	125	right brace			
547446	12	52	other char	4	U+00034	
547070	12	53	other char	5	U+00035	
544114	12	54	other char	6	U+00036	

17.6.4 Example 4: internal variables

```
\luatokenable\everypar
```

internal token variable: everypar

546316	138	0	protected call	dotagsetparcounter	
546330	138	0	protected call	page_otr_command_synchronize_side_floats	
547249	138	0	protected call	checkindentation	
546186	137	0	call	showparagraphnumber	
545791	138	0	protected call	restoreinterlinepenalty	
547294	137	0	call	flushnotes	
546062	138	0	protected call	registerparoptions	
545380	137	0	call	flushpostponednodedata	
543064	137	0	call	typo_delimited_repeat	
547323	137	0	call	spac_paragraphs_flush_intro	
211373	137	0	call	typo_initial_handle	
547356	137	0	call	typo_firstline_handle	

546280	137	0	call	spac_paragraph_wrap
546093	138	0	protected call	spac_paragraph_freeze

17.6.5 Example 5: macro definitions

```
\protected\def\whatever#1[#2](#3)\relax
  {oops #1 and #2 & #3 done ## error}
```

```
\luatokenable\whatever
```

protected control sequence: whatever

545891	19	49	match	argument 1
543474	12	91	other char	[U+0005B
546708	19	50	match	argument 2
546138	12	93	other char] U+0005D
547448	12	40	other char	(U+00028
546975	19	51	match	argument 3
422154	12	41	other char) U+00029
547264	16	0	relax	relax
547410	20	0	end match	
<hr/>				
545175	11	111	letter	o U+0006F
547126	11	101	letter	e U+00065
545698	11	112	letter	p U+00070
546932	11	115	letter	s U+00073
546852	10	32	spacer	
544158	21	1	parameter reference	
545555	10	32	spacer	
547652	11	97	letter	a U+00061
545658	11	110	letter	n U+0006E
546952	11	100	letter	d U+00064
544377	10	32	spacer	
547425	21	2	parameter reference	
546536	10	32	spacer	
546936	12	38	other char	& U+00026
547480	10	32	spacer	
547638	21	3	parameter reference	
544973	10	32	spacer	
546397	11	100	letter	d U+00064
543562	11	111	letter	o U+0006F
544024	11	110	letter	n U+0006E
546776	11	101	letter	e U+00065
413151	10	32	spacer	
547293	6	35	parameter	
545398	10	32	spacer	
547083	11	101	letter	e U+00065
544134	11	114	letter	r U+00072
547547	11	114	letter	r U+00072
546629	11	111	letter	o U+0006F
546020	11	114	letter	r U+00072

17.6.6 Example 6: commands

`\luatokenable\startitemize`

`\luatokenable\stopitemize`

frozen instance protected control sequence: startitemize

546332	141	0	tolerant protected call		startitemgroup
542558	12	91	other char	[U+0005B	
547162	11	105	letter	i U+00069	
543567	11	116	letter	t U+00074	
546107	11	101	letter	e U+00065	
540209	11	109	letter	m U+0006D	
547247	11	105	letter	i U+00069	
544261	11	122	letter	z U+0007A	
543448	11	101	letter	e U+00065	
547284	12	93	other char] U+0005D	

frozen instance protected control sequence: stopitemize

544404	138	0	protected call	stopitemgroup
--------	-----	---	----------------	---------------

17.6.7 Example 7: commands

`\luatokenable\doifelse`

permanent protected control sequence: doifelse

546713	19	49	match	argument 1
545022	19	50	match	argument 2
546094	20	0	end match	
546765	132	26	if test	iftok
547427	1	123	left brace	
546977	21	1	parameter reference	
546666	2	125	right brace	
539231	1	123	left brace	
544490	21	2	parameter reference	
547115	2	125	right brace	
546424	126	0	expand after	expandafter
547382	137	0	call	firstoftwoarguments
540275	132	3	if test	else
547371	126	0	expand after	expandafter
548249	137	0	call	secondoftwoarguments
543611	132	2	if test	fi

17.6.8 Example 8: nothing

`\luatokenable\relax`

primitive control sequence: relax

548251	16	0	relax	relax
--------	----	---	-------	-------

17.6.9 Example 9: hashes

```
\edef\foo#1#2{(#1)(\letterhash)(#2)} \luatokenable\foo
```

control sequence: foo

545963	19	49	match	argument 1
546566	19	50	match	argument 2
539348	20	0	end match	
542910	12	40	other char	(U+00028
545068	21	1	parameter reference	
543782	12	41	other char) U+00029
546678	12	40	other char	(U+00028
89083	12	35	other char	# U+00023
547028	12	41	other char) U+00029
548219	12	40	other char	(U+00028
547585	21	2	parameter reference	
547388	12	41	other char) U+00029

17.6.10 Example 10: nesting

```
\def\foo#1{\def\foo##1{(#1)(##1)}} \luatokenable\foo
```

control sequence: foo

544222	19	49	match	argument 1
546238	20	0	end match	
547604	121	1	def	def
547500	137	0	call	foo
547527	6	35	parameter	
547991	12	49	other char	1 U+00031
549370	1	123	left brace	
545854	12	40	other char	(U+00028
542670	21	1	parameter reference	
547505	12	41	other char) U+00029
547460	12	40	other char	(U+00028
547355	6	35	parameter	
547342	12	49	other char	1 U+00031
546196	12	41	other char) U+00029
545684	2	125	right brace	

17.6.11 Remark

In all these examples the numbers are to be seen as abstractions. Some command codes and sub command codes might change as the engine evolves. This is why the LuaMetaTeX engine has lots of Lua functions that provide information about what number represents what command.

17.6.11 Colofon

Author	Hans Hagen
ConTeXt	2023.04.27 17:04
LuaMetaTeX	2.1008
Support	www.pragma-ade.com contextgarden.net

18 Buffers

low level

TEX

buffers

Contents

18.1 Preamble	239
18.2 Encoding	239
18.3 Performance	241
18.4 Files	241
18.5 Macros	242
18.6 Token lists	243
18.7 Buffers	243
18.8 Setups	246
18.9 xml	247
18.10 Lua	248
18.11 Protection	248

18.1 Preamble

Buffers are not that low level but it makes sense to discuss them in this perspective because it relates to tokenization, internal representation and manipulating.

In due time we can describe some more commands and details here. This is a start. Feel free to tell me what needs to be explained.

18.2 Encoding

Normally processing a document starts with reading from file. In the past we were talking single bytes that were then maps onto a specific input encoding that itself matches the encoding of a font. When you enter an 'a' its (normally ascii) number 97 becomes the index into a font. That same number is also used in the hyphenator which is why font encoding and hyphenation are strongly related. If in an eight bit T_EX engine you need a precomposed 'ä' you have to use an encoding that has that character in some slot with again matching fonts and patterns. The actually used font can have the *shapes* in different slots and remapping is then done in the backend code using encoding and mapping files. When OpenType fonts are used the relationship between characters (input) and glyphs (rendering) also depends on the application of font features.

In eight bit environments all this brings a bit of a resource management nightmare along with complex installation of new fonts. It also puts strain on the macro package, especially when you want to mix different input encodings onto different font encodings and thereby pattern encodings in the same document. You can compare this with code pages in operating system, but imagine them potentially being mixed in one document,

which can happen when you mix multiple languages where the accumulated number of different characters exceeds 256. You end up switching between encodings. One way to deal with it is making special characters active and let their meaning differ per situation. That is for instance how in MkII we handled utf8 and thereby got around distributing multiple pattern files per language as we only needed to encoding them in utf and then remap them to the required encoding when loading patterns. A mental exercise is wondering how to support cjk scripts in an eight bit MkII, something that actually can be done with some effort.

The good news is that when we moved from MkII to MkIV we went exclusively utf8 because that is what the Lua \TeX engine expects. Upto four bytes are read in and translated into one Unicode character. The internal representation is a 32 bit integer (four bytes) instead of a single byte. That also means that in the transition we got rid of quite some encoding related low level font and pattern handling. We still support input encodings (called regimes in Con \TeX t) but I'm pretty sure that nowadays no one uses input other than utf8. While Con \TeX t is normally quite upward compatible this is one area where there were fundamental changes.

There is still some interpretation going on when reading from file: for instance, we need to normalize the Unicode input, and we feed the engine separate lines on demand. Apart from that, some characters like the backslash, dollar sign and curly braces have special meaning so for accessing them as characters we have to use commands that inject those characters. That didn't change when we went from MkII to MkIV. In practice it's never really a problem unless you find yourself in one of the following situations:

- *Example code has to be typeset as-is, so braces etc. are just that.* This means that we have to change the way characters are interpreted. Typesetting code is needed when you want to document \TeX and macros which is why mechanisms for that have to be present right from the start.
- *Content is collected and used later.* A separation of content and usage later on often helps making a source look cleaner. Examples are “wrapping a table in a buffer” and “including that buffer when a table is placed” using the placement macros.
- *Embedded MetaPost and Lua code.* These languages come with different interpretation of some characters and especially MetaPost code is often stored first and used (processed) later.
- *The content comes from a different source.* Examples are xml files where angle brackets are special but for instance braces aren't. The data is interpreted as a stream or as a structured tree.

- *The content is generated.* It can for instance come from Lua, where bytes (representing utf) is just text and no special characters are to be intercepted. Or it can come from a database (using a library).

For these reasons ConT_EXt always had ways to store data in ways that makes this possible. The details on how that is done might have changed over versions, been optimized, extended with additional interfaces and features but given where we come from most has been there from the start.

18.3 Performance

When T_EX came around, the bottlenecks in running T_EX were the processor, memory and disks and depending on the way one used it the speed of the console or terminal; so, basically the whole system. One could sit there and wait for the page counters ([1] [2] . . . to show up. It was possible to run T_EX on a personal computer but it was somewhat resource hungry: one needed a decent disk (a 10 MB hard disk was huge and with today's phone camera snapshots that sounds crazy). One could use memory extenders to get around the 640K limitation (keep in mind that the programs and operating systems also took space). This all meant that one could not afford to store too many tokens in memory but even using files for all kind of (multi-pass) trickery was demanding.

When processors became faster and memory plenty the disk became the bottleneck, but that changed when ssd's showed up. Combined with already present file caching that had some impact. We are now in a situation that cpu cores don't get that much faster (at least not twice as fast per iteration) and with T_EX being a single core byte cruncher we're more or less in a situation where performance has to come from efficient programming. That means that, given enough memory, in some cases storing in tokens wins over storing in files, but it is no rule. In practice there is not much difference so one can even more than yesterday choose for the most convenient method. Just assume that the ConT_EXt code, combined with LuaMetaT_EX will give you what you need with a reasonable performance. When in doubt, test with simple test files and it that works out well compared to the real code, try to figure out where 'mistakes' are made. Inefficient Lua and T_EX code has way more impact than storing a few more tokens or using some files.

18.4 Files

Nearly always files are read once per run. The content (mixed with commands) is scanned and macros are expanded and/or text is typeset as we go. Internally the LuaMetaT_EX engine is in "scanning from file", "scanning from token lists", or "scanning

from Lua output” mode. The first mode is (in principle) the slowest because utf sequences are converted to tokens (numbers) but there is no way around it. The second method is fast because we already have these numbers, but we need to take into account where the linked list of tokens comes from. If it is converted runtime from for instance file input or macro expansion we need to add the involved overhead. But scanning a stored macro body is pretty efficient especially when the macro is part of the loaded macro package (format file). The third method is comparable with reading from file but here we need to add the overhead involved with storing the Lua output into data structures suitable for T_EX's input mechanism, which can involve memory allocation outside the reserved pool of tokens. On modern systems that is not really a problem. It is good to keep in mind that when T_EX was written much attention was paid to optimization and in LuaMetaT_EX we even went a bit further, also because we know what kind of input, processing and output we're dealing with.

When reading from file or Lua output we interpret bytes turned utf numbers and that is when catcode regimes kick in: characters are interpreted according to the catcode properties: escape character (backslash), curly braces (grouping and arguments), dollars (math), etc. While with reading from token lists these catcodes are already taken care of and we're basically interpreting meanings instead of characters. By changing the catcode regime we can for instance typeset content verbatim from files and Lua strings but when reading from token lists we're sort of frozen. There are tricks to reinterpret the token list but that comes with overhead and limitations.

18.5 Macros

A macro can be seen as a named token with a meaning attached. In LuaMetaT_EX macros can take up to 15 arguments (six more than regular T_EX) that can be separated by so called delimiters. A token has a command property (operator) and a value (operand). Because a Unicode character doesn't need all four bytes of an integer and because in the engine numbers, dimensions and pointers are limited in size we can store all of these efficiently with the command code. Here the body of `\foo` is a list of three tokens:

```
\def\foo{abc} \foo \foo \foo
```

When the engine fetches a token from a list it will interpret the command and when it fetches from file it will create tokens on the fly and then interpret those. When a file or list is exhausted the engine pops the stack and continues at the previous level. Because macros are already tokenized they are more efficient than file input. For more about macros you can consult the low level document about them.

The more you use a macro, the more it pays off compared to a file. However don't overestimate this, because in the end the typesetting and expanding all kind of other involved macros might reduce the file overhead to noise.

18.6 Token lists

A token list is like a macro but is part of the variable (register) system. It is just a list (so no arguments) and you can append and prepend to that list.

```
\toks123={abc}    \the\toks123
\scratchtoks{abc} \the\scratchtoks
```

Here `\scratchtoks` is defined with `\newtoks` which creates an efficient reference to a list so that, contrary to the first line, no register number has to be scanned. There are low level manuals about tokens and registers that you can read if you want to know more about this. As with macros the list in this example is three tokens long. Contrary to macros there is no macro overhead as there is no need to check for arguments.²⁵

Because they use more or less the same storage method macros and token list registers perform the same. The power of registers comes from some additional manipulators in Lua_T_E_X (and LuaMeta_T_E_X) and the fact that one can control expansion with `\the`, although that later advantage is compensated with extensions to the macro language (like `\protected macro definitions`).

18.7 Buffers

Buffers are something specific for Con_T_E_Xt and they have always been part of this system. A buffer is defined as follows:

```
\startbuffer[one]
line 1
line 2
\stopbuffer
```

Among the operations on buffers the next two are used most often:

```
\typebuffer[one]
\getbuffer[one]
```

²⁵ In LuaMeta_T_E_X a macro without arguments is also quite efficient.

Scanning a buffer at the $\text{T}_{\text{E}}\text{X}$ end takes a little effort because when we start reading the catcodes are ignored and for instance backslashes and curly braces are retained. Hardly any interpretation takes place. The same is true for spacing, so multiple spaces are not collapsed and newlines stay. The tokenized content of a buffer is converted back to a string and that content is then read in as a pseudo file when we need it. So, basically buffers are files! In MkII they actually were files (in the $\backslash\text{jobname}$ name space and suffix `tmp`), but in MkIV they are stored in and managed by Lua. That also means that you can set them very efficiently at the Lua end:

```
\startluacode
buffers.assign("one",[
line 1
line 2
])
\stopluacode
```

Always keep in mind that buffers eventually are read as files: character by character, and at that time the content gets (as with other files) tokenized. A buffer name is optional. You can nest buffers, with and without names.

Because $\text{ConT}_{\text{E}}\text{Xt}$ is very much about re-use of content and selective processing we have an (already old) subsystem for defining named blocks of text (using $\backslash\text{begin}\dots$ and $\backslash\text{end}\dots$ tagging. These blocks are stored just like buffers but selective flushing is part of the concept. Think of coding an educational document with explanations, questions, answers and then typesetting only the explanations, or the explanation along with some questions. Other components can be typeset later so one can make for instance a special book(let) with answers that either of not repeats the questions. Here we need features like synchronization of numbers so that's why we cannot really use buffers. An alternative is to use xml and filter from that.

The $\backslash\text{definebuffer}$ command defines a new buffer environment. When you set buffers in Lua you don't need to define a buffer because likely you don't need the $\backslash\text{start}$ and $\backslash\text{stop}$ commands. Instead of $\backslash\text{getbuffer}$ you can also use $\backslash\text{getdefinedbuffer}$ with defined buffers. In that case the before and after keys of that specific instance are used.

The $\backslash\text{getinlinebuffer}$ command, which like the getters takes a list of buffer names, ignores leading and trailing spaces. When multiple buffers are flushed this way, spacing between buffers is retained.

The most important aspect of buffers is that the content is *not* interpreted and tokenized: the bytes stay as they are.

```
\definebuffer[MyBuffer]
```

```
\startMyBuffer
\bold{this is
a buffer}
\stopMyBuffer
```

```
\typeMyBuffer \getMyBuffer
```

These commands result in:

```
\bold{this is
a buffer}
```

this is a buffer

There are not that many parameters that can be set: before, after and strip (when set to no leading and trailing spacing will be kept. The `\stop...` command, in our example `\stopMyBuffer`, can be defined independent to do something after the buffer has been read and stored but by default nothing is done.

You can test if a buffer exists with `\doifelsebuffer` (expandable) and `\doifelsebufferempty` (unexpandable). A buffer is kept in memory unless it gets wiped clean with `resetbuffer`.

```
\savebuffer      [MyBuffer][temp]      % gets name: jobname-temp.tmp
\savebufferinfile[MyBuffer][temp.log] % gets name: temp.log
```

You can also stepwise fill such a buffer:

```
\definesavebuffer[slide]
```

```
\startslide
  \starttext
\stopslide
\startslide
  slide 1
\stopslide
text 1 \par
\startslide
  slide 2
\stopslide
text 2 \par
```

```
\startslide
  \stoptext
\stopslide
```

After this you will have a file `\jobname-slide.tex` that has the two lines wrapped as text. You can set up a ‘save buffer’ to use a different filename (with the `file` key), a different prefix using `prefix` and you can set up a directory. A different name is set with the `list` key.

You can assign content to a buffer with a somewhat clumsy interface where we use the delimiter `\endbuffer`. The only restriction is that this delimiter cannot be part of the content:

```
\setbuffer[name]here comes some text\endbuffer
```

For more details and obscure commands that are used in other commands you can peek into the source.

Using buffers in the `cld` interface is tricky because of the `catcode` magick that is involved but there are setters and getters:

function	arguments
<code>buffers.assign</code>	<code>name, content [,catcodes]</code>
<code>buffers.erase</code>	<code>name</code>
<code>buffers.prepend</code>	<code>name, content</code>
<code>buffers.append</code>	<code>name, content</code>
<code>buffers.exists</code>	<code>name</code>
<code>buffers.empty</code>	<code>name</code>
<code>buffers.getcontent</code>	<code>name</code>
<code>buffers.getlines</code>	<code>name</code>

There are a few more helpers that are used in other (low level) commands. Their functionality might adapt to their usage there. The `context.startbuffer` and `context.stopbuffer` are somewhat differently defined than regular `cld` commands.

18.8 Setups

A setup is basically a macro but is stored and accessed in a namespace separated from ordinary macros. One important characteristic is that inside setups newlines are ignored.

```
\startsetups MySetupA
```

```

    This is line 1
    and this is line 2
\stopsetups

```

```
\setup{MySetupA}
```

This is line 1 and this is line 2

A simple way out is to add a comment character preceded by a space. Instead you can also use `\space`:

```

\startsetups [MySetupB]
    This is line 1 %
    and this is line 2\space
    while here we have line 3
\stopsetups

```

```
\setup[MySetupB]
```

This is line 1 and this is line 2 while here we have line 3

You can use square brackets instead of space delimited names in definitions and also in calling up a (list of) setup(s). The `\directsetup` command takes a single setup name and is therefore more efficient.

Setups are basically simple macros although there is some magic involved that comes from their usage in for instance xml where we pass an argument. That means we can do the following:

```

\startsetups MySetupC
    before#1after
\stopsetups

\setupwithargument{MySetupC}{ {\em and} }

```

before *and* after

Because a setup is a macro, the body is a linked list of tokens where each token takes 8 bytes of memory, so MySetupC has 12 tokens that take 96 bytes of memory (plus some overhead related to macro management).

18.9 xml

Discussing xml is outside the scope of this document but it is worth mentioning that once an xml tree is read is, the content is stored in strings and can be filtered into $\text{T}_{\text{E}}\text{X}$,

where it is interpreted as if coming from files (in this case Lua strings). If needed the content can be interpreted as T_EX input.

18.10 Lua

As mentioned already, output from Lua is stored and when a Lua call finishes it ends up on the so called input stack. Every time the engine needs a token it will fetch from the input stack and the top of the stack can represent a file, token list or Lua output. Interpreting bytes from files or Lua strings results in tokens. As a side note: Lua output can also be already tokenized, because we can actually write tokens and nodes from Lua, but that's more an implementation detail that makes the Lua input stack entries a bit more complex. It is normally not something users will do when they use Lua in their documents.

18.11 Protection

When you define macros there is the danger of overloading some defined by the system. Best use CamelCase so that you stay away from clashes. You can enable some checking:

```
\enabledirectives[overloadmode=warning]
```

or when you want to quit on a clash:

```
\enabledirectives[overloadmode=error]
```

When these trackers are enabled you can get around the check with:

```
\pushoverloadmode
...
\popoverloadmode
```

But delay that till you're sure that redefining is okay.

18.11 Colofon

Author	Hans Hagen
ConT _E Xt	2023.04.27 17:04
LuaMetaT _E X	2.1008
Support	www.pragma-ade.com contextgarden.net