# ConTEXt Lua Documents

Hans Hagen

# Contents

# Introduction

Sometimes you hear folks complain about the TEX input language, i.e. the backslashed commands that determine your output. Of course, when alternatives are being discussed everyone has a favorite programming language. In practice coding a document in each of them triggers similar sentiments with regards to coding as TEX itself does. And in then, as the problem doesn't change, one ends up with the same annoyances.

So, just for fun, I added a couple of commands to CONTEXT MKIV that permit coding a document in LUA. In retrospect it has been surprisingly easy to implement a feature like this using metatables. Of course it's a bit slower than using TEX as input language but sometimes the LUA interface is more readable given the problem at hand.

After a while I decided to use that interface in non-critical core CONTEXT code and in styles (modules) and solutions for projects. Using the LUA approach is sometimes more convenient, especially if the code mostly manipulates data. For instance, if you process XML files of database output you can use the interface that is available at the TEX end, or you can use LUA code to do the work, or you can use a combination. So, from that moment on, in CONTEXT you can code your style and document source in (a mixture of) TEX, XML, METAPOST and LUA.

Because CONTEXT is a TEX macro package quite some action happens at the TEX end. Although we do a lot in LUA, there is no need to abandon the macro language. If I wanted a typesetting system written in procedural language I'd already come up with one but I have no need for it and TEX is more fun anyway. But the fact that we mix these conceptual different languages means that we need ways to communicate between them. The possibilities to communicate and switch between TEX, LUA, and of course METAPOST have evolved over time. It started with just `\directlua` but we now can use tokens scanners that permit nice interfaces. We have access to scanners in TEX and METAPOST and can pipe back data to both in efficient ways. Actually, we seldom use that direct command, at most we use an equivalent `\ctxlua`, which permits a future overload. As a consequence the interfaces in CONTEXT also evolved, although mostly deep down out of sight of users: CONTEXT always tries to offer abstract interfaces that as designed in a way that permits upward compatible upgrades.

Going from LUA to TEX is mostly done with the `context` command. The most low level approach would be to use the `tex.print` functions but using the `context` interface is often better. Directly going from TEX to LUA is done with `\ctxlua` and friends. There is no need to use the low level 'print' interfaces and it can even be counter-productive when you want tracing. We keep the low-level (primitive) interfaces unblocked but best first try what CONTEXT provides on top of them.

In these internet times it is no problem to find comments on code and coding practices (of course not seldom by people who themselves produce code that deserves a lot of comment but get away with it). The same is true for complaints about bugs that not always are put in the perspective of the total amount of well working code or experiments; on the average TEX systems behave quite well and have not that many bugs, but some low level features they can be confusing. We don't impose a coding style in user files but we do so in the core, but that only hampers ourselves.

Another issue is complaints about manuals or lack of documentation, again, often by people who themselves never produced something that those who they complain about can use, but, that said, lack of documentation has the benefit that one can learn by playing around. There are plenty of examples to learn from, also in the over ten thousand pages of CONTEXT documentation and articles.

And of course there is the test suite which has examples. Some users even like to look into the source code to learn some tricks. Now, the internet is not always the friendly and tolerant environment that its marketing suggests and one can run into patronizing comments on how to do thing, but: we don't (and cannot) enforce best practices.

However, when you define LUA functions you should *not* overload existing functions without knowing what you deal with, simply because the whole CONTEXT ecosystem could be affected. Don't expect support when you do so. The same is true for TEX macros and even more so for primitives. Although there is some protection going on we think the system should be as open as possible, so little is hidden for the user.[1]

The reason for mentioning this is that we do see users come up with (often surprising) solutions and we invite them to keep playing around, even with low level code. When writing a document, playing around can be a nice distraction. But, using a high level interface has some benefits too. It gives a bit of protection against bad interactions with other code. It also often involves less coding. Performance might be better although that is not always a real issue as there are plenty of ways to make a TEX system slow anyway.

In the following chapters I will introduce typesetting in LUA, but as we rely on CONTEXT it is unavoidable that some regular CONTEXT code shows up. The fact that you can ignore backslashes does not mean that you can do without knowledge of the underlying system. I expect that the user is somewhat familiar with this macro package. Some chapters are follow ups on articles or earlier publications.

Some information (and mechanism) show up in more than one chapter. This is a side effect of LUA being integrated in many places, so an isolated discussion is a bit hard.

In the meantime most of the code is rather stable and proven. However, this manual will never be complete. You can find examples all over the code base, and duplicating everything here makes no sense. If you find errors, please let me know. If you think that something is missing, you can try to convince me to add it. It's hard to keep up with what gets added so input is welcome.

I started writing this manual in the early days if MKIV using LUATEX but in the meantime we've arrived at MKXL using LUAMETATEX. This means that there can be features discussed here that are only available in the latest greatest. In that perspective it is important to notice that we no longer expect LUAJITTEX to be used, and assume LUA 5.5 or later, which comes with some additional functionality and parsing rules.

Hans Hagen
Hasselt NL
2009 — 2025+

[1] In the core code and modules that come with CONTEXT we are rather rigorous. There we have patterns that need to be followed. Hacking around and patching at will is not an option there. In the rather open system that a macro package provides it is no problem to quickly mess up the lot with bad code. There is a strong relationship between the coding and way syntax highlighting evolved, as offered in the SCITE module that comes with CONTEXT.

# 1 A bit of Lua

## 1.1 The language

Small is beautiful and this is definitely true for the programming language LUA (moon in Portuguese). We had good reasons for using this language in LUATEX: simplicity, speed, syntax and size to mention a few. Of course personal taste also played a role and after using a couple of scripting languages extensively the switch to LUA was rather pleasant.

As the LUA reference manual is an excellent book there is no reason to discuss the language in great detail: just buy 'Programming in LUA' by the LUA team. Nevertheless I will give a short summary of the important concepts but consult the book if you want more details.

## 1.2 Data types

The most basic data type is `nil`. When we define a variable, we don't need to give it a value:

```
local v
```

Here the variable `v` can get any value but till that happens it equals `nil`. There are simple data types like `numbers`, `booleans` and `strings`. Here are some numbers:

```
local n = 1 + 2 * 3
local x = 2.3
```

Numbers are always floats[2] and you can use the normal arithmetic operators on them as well as functions defined in the math library. Inside TEX we have only integers, although for instance dimensions can be specified in points using floats but that's more syntactic sugar. One reason for using integers in TEX has been that this was the only way to guarantee portability across platforms. However, we're 30 years along the road and in LUA the floats are implemented identical across platforms, so we don't need to worry about compatibility.

Strings in LUA can be given between quotes or can be so called long strings forced by square brackets.

```
local s = "Whatever"
local t = s .. ' you want'
local u = t .. [[ to know]] .. [[--[ about Lua!]--]]
```

The two periods indicate a concatenation. Strings are hashed, so when you say:

```
local s = "Whatever"
local t = "Whatever"
local u = t
```

only one instance of `Whatever` is present in memory and this fact makes LUA very efficient with respect to strings. Strings are constants and therefore when you change variable `s`, variable `t` keeps its value. When you compare strings, in fact you compare pointers, a method that is really fast. This compensates the time spent on hashing pretty well.

[2] This is true for all versions upto 5.2 but following version can have a more hybrid model.

Booleans are normally used to keep a state or the result from an expression.

```
local b = false
local c = n > 10 and s == "whatever"
```

The other value is `true`. There is something that you need to keep in mind when you do testing on variables that are yet unset.

```
local b = false
local n
```

The following applies when `b` and `n` are defined this way:

```
b == false  true
n == false  false
n == nil    true
b == nil    false
b == n      false
n == nil    true
```

Often a test looks like:

```
if somevar then
    ...
else
    ...
end
```

In this case we enter the else branch when `somevar` is either `nil` or `false`. It also means that by looking at the code we cannot beforehand conclude that `somevar` equals `true` or something else. If you want to really distinguish between the two cases you can be more explicit:

```
if somevar == nil then
    ...
elseif somevar == false then
    ...
else
    ...
end
```

or

```
if somevar == true then
    ...
else
    ...
end
```

but such an explicit test is seldom needed.

There are a few more data types: tables and functions. Tables are very important and you can recognize them by the same curly braces that make TEX famous:

```
local t = { 1, 2, 3 }
local u = { a = 4, b = 9, c = 16 }
local v = { [1] = "a", [3] = "2", [4] = false }
local w = { 1, 2, 3, a = 4, b = 9, c = 16 }
```

The `t` is an indexed table and `u` a hashed table. Because the second slot is empty, table `v` is partially indexed (slot 1) and partially hashed (the others). There is a gray area there, for instance, what happens when you nil a slot in an indexed table? In practice you will not run into problems as you will either use a hashed table, or an indexed table (with no holes), so table `w` is not uncommon.

We mentioned that strings are in fact shared (hashed) but that an assignment of a string to a variable makes that variable behave like a constant. Contrary to that, when you assign a table, and then copy that variable, both variables can be used to change the table. Take this:

```
local t = { 1, 2, 3 }
local u = t
```

We can change the content of the table as follows:

```
t[1], t[3] = t[3], t[1]
```

Here we swap two cells. This is an example of a parallel assigment. However, the following does the same:

```
t[1], t[3] = u[3], u[1]
```

After this, both `t` and `u` still share the same table. This kind of behaviour is quite natural. Keep in mind that expressions are evaluated first, so

```
t[#t+1], t[#t+1] = 23, 45
```

Makes no sense, as the values end up in the same slot. There is no gain in speed so using parallel assignments is mostly a convenience feature.

There are a few specialized data types in LUA, like `coroutines` (built in), `file` (when opened), `lpeg` (only when this library is linked in or loaded). These are called 'userdata' objects and in LUATEX we have more userdata objects as we will see in later chapters. Of them nodes are the most noticeable: they are the core data type of the TEX machinery. Other libraries, like `math` and `bit32` are just collections of functions operating on numbers.

Functions look like this:

```
function sum(a,b)
  print(a, b, a + b)
end
```

or this:

```
function sum(a,b)
  return a + b
end
```

There can be many arguments of all kind of types and there can be multiple return values. A function is a real type, so you can say:

```
local f = function(s) print("the value is: " .. s) end
```

In all these examples we defined variables as `local`. This is a good practice and avoids clashes. Now watch the following:

```
local n = 1

function sum(a,b)
  n = n + 1
  return a + b
end

function report()
  print("number of summations: " .. n)
end
```

Here the variable `n` is visible after its definition and accessible for the two global functions. Actually the variable is visible to all the code following, unless of course we define a new variable with the same name. We can hide `n` as follows:

```
do
  local n = 1

  sum = function(a,b)
    n = n + 1
    return a + b
  end

  report = function()
    print("number of summations: " .. n)
  end
end
```

This example also shows another way of defining the function: by assignment.

The `do ... end` creates a so called closure. There are many places where such closures are created, for instance in function bodies or branches like `if ... then ... else`. This means that in the following snippet, variable `b` is not seen after the end:

```
if a > 10 then
  local b = a + 10
  print(b*b)
end
```

When you process a blob of LUA code in TEX (using `\directlua` or `\latelua`) it happens in a closure with an implied `do ... end`. So, `local` defined variables are really local.

## 1.3 TEX's data types

We mentioned `numbers`. At the TEX end we have counters as well as dimensions. Both are numbers but dimensions are specified differently

```
local n = tex.count[0]
local m = tex.dimen.lineheight
local o = tex.sp("10.3pt") -- sp or 'scaled point' is the smallest unit
```

The unit of dimension is 'scaled point' and this is a pretty small unit: 10 points equals to 655360 such units.

Another accessible data type is tokens. They are automatically converted to strings and vice versa.

```
tex.toks[0] = "message"
print(tex.toks[0])
```

Be aware of the fact that the tokens are letters so the following will come out as text and not issue a message:

```
tex.toks[0] = "\message{just text}"
print(tex.toks[0])
```

## 1.4 Control structures

Loops are not much different from other languages: we have `for ... do`, `while ... do` and `repeat ... until`. We start with the simplest case:

```
for index=1,10 do
  print(index)
end
```

You can specify a step and go downward as well:

```
for index=22,2,-2 do
  print(index)
end
```

Indexed tables can be traversed this way:

```
for index=1,#list do
  print(index, list[index])
end
```

Hashed tables on the other hand are dealt with as follows:

```
for key, value in next, list do
  print(key, value)
end
```

Here `next` is a built in function. There is more to say about this mechanism but the average user will use only this variant. Slightly less efficient is the following, more readable variant:

```
for key, value in pairs(list) do
  print(key, value)
end
```

and for an indexed table:

```
for index, value in ipairs(list) do
  print(index, value)
end
```

The function call to `pairs(list)` returns `next, list` so there is an (often neglectable) extra overhead of one function call.

The other two loop variants, `while` and `repeat`, are similar.

```
i = 0
while i < 10  do
  i = i + 1
  print(i)
end
```

This can also be written as:

```
i = 0
repeat
  i = i + 1
  print(i)
until i = 10
```

Or:

```
i = 0
while true do
  i = i + 1
  print(i)
  if i = 10 then
    break
  end
end
```

Of course you can use more complex expressions in such constructs.

## 1.5 Conditions

Conditions have the following form:

```
if a == b or c > d or e then
  ...
elseif f == g then
  ...
else
  ...
end
```

Watch the double `==`. The complement of this is `~=`. Precedence is similar to other languages. In practice, as strings are hashed. Tests like

```
if key == "first" then
  ...
end
```

and

```
if n == 1 then
  ...
end
```

are equally efficient. There is really no need to use numbers to identify states instead of more verbose strings.

## 1.6 Namespaces

Functionality can be grouped in libraries. There are a few default libraries, like `string`, `table`, `lpeg`, `math`, `io` and `os` and LuaTeX adds some more, like `node`, `tex` and `texio`.

A library is in fact nothing more than a bunch of functionality organized using a table, where the table provides a namespace as well as place to store public variables. Of course there can be local (hidden) variables used in defining functions.

```
do
  mylib = { }

  local n = 1

  function mylib.sum(a,b)
    n = n + 1
    return a + b
  end

  function mylib.report()
    print("number of summations: " .. n)
  end
end
```

The defined function can be called like:

```
mylib.report()
```

You can also create a shortcut, This speeds up the process because there are less lookups then. In the following code multiple calls take place:

```
local sum = mylib.sum

for i=1,10 do
  for j=1,10 do
    print(i, j, sum(i,j))
  end
end
```

```
mylib.report()
```

As LUA is pretty fast you should not overestimate the speedup, especially not when a function is called seldom. There is an important side effect here: in the case of:

```
print(i, j, sum(i,j))
```

the meaning of `sum` is frozen. But in the case of

```
print(i, j, mylib.sum(i,j))
```

The current meaning is taken, that is: each time the interpreter will access `mylib` and get the current meaning of `sum`. And there can be a good reason for this, for instance when the meaning is adapted to different situations.

In CONTEXT we have quite some code organized this way. Although much is exposed (if only because it is used all over the place) you should be careful in using functions (and data) that are still experimental. There are a couple of general libraries and some extend the core LUA libraries. You might want to take a look at the files in the distribution that start with `l-`, like `l-table.lua`. These files are preloaded.[3] For instance, if you want to inspect a table, you can say:

```
local t = { "aap", "noot", "mies" }
table.print(t)
```

You can get an overview of what is implemented by running the following command:

```
context s-tra-02 --mode=tablet
```

*todo: add nice synonym for this module and also add helpinfo at the to so that we can do* `context --styles`

## 1.7 Comment

You can add comments to your LUA code. There are basically two methods: one liners and multi line comments.

```
local option = "test" -- use this option with care

local method = "unknown" --[[comments can be very long and when entered
                             this way they and span multiple lines]]
```

The so called long comments look like long strings preceded by `--` and there can be more complex boundary sequences.

## 1.8 Pitfalls

Sometimes `nil` can bite you, especially in tables, as they have a dual nature: indexed as well as hashed.

[3] In fact, if you write scripts that need their functionality, you can use `mtxrun` to process the script, as `mtxrun` has the core libraries preloaded as well.

```
\startluacode
local n1 = # { nil, 1, 2, nil }      -- 3
local n2 = # { nil, nil, 1, 2, nil } -- 0

context("n1 = %s and n2 = %s",n1,n2)
\stopluacode
```

results in: n1 = 3 and n2 = 0. So, you cannot really depend on the length operator here. On the other hand, with:

```
\startluacode
local function check(...)
    return select("#",...)
end

local n1 = check ( nil, 1, 2, nil )      -- 4
local n2 = check ( nil, nil, 1, 2, nil ) -- 5

context("n1 = %s and n2 = %s",n1,n2)
\stopluacode
```

we get: n1 = 4 and n2 = 5, so the `select` is quite usable. However, that function also has its specialities. The following example needs some close reading:

```
\startluacode
local function filter(n,...)
    return select(n,...)
end

local v1 = { filter ( 1, 1, 2, 3 ) }
local v2 = { filter ( 2, 1, 2, 3 ) }
local v3 = { filter ( 3, 1, 2, 3 ) }

context("v1 = %+t and v2 = %+t and v3 = %+t",v1,v2,v3)
\stopluacode
```

We collect the result in a table and show the concatination:

v1 = 1+2+3 and v2 = 2+3 and v3 = 3

So, what you effectively get is the whole list starting with the given offset.

```
\startluacode
local function filter(n,...)
    return (select(n,...))
end

local v1 = { filter ( 1, 1, 2, 3 ) }
local v2 = { filter ( 2, 1, 2, 3 ) }
local v3 = { filter ( 3, 1, 2, 3 ) }
```

```
context("v1 = %+t and v2 = %+t and v3 = %+t",v1,v2,v3)
\stopluacode
```

Now we get: v1 = 1 and v2 = 2 and v3 = 3. The extra `()` around the result makes sure that we only get one return value.

Of course the same effect can be achieved as follows:

```
local function filter(n,...)
    return select(n,...)
end

local v1 = filter ( 1, 1, 2, 3 )
local v2 = filter ( 2, 1, 2, 3 )
local v3 = filter ( 3, 1, 2, 3 )

context("v1 = %s and v2 = %s and v3 = %s",v1,v2,v3)
```

It can be that you want this so called 'vararg' as a table, which is easy to accomplish: just wrap the `...` in curly braces (a table constructor):

```
\startluacode
local function grab(v,...)
    context("%s = % + t",v,{ ... })
end

grab("x",1,2,3) context(" and ") grab("y",4,5,6)
\stopluacode
```

And indeed we get: x = 1 + 2 + 3 and y = 4 + 5 + 6. There is an extension in the working that permits the second variant of:

```
\startluacode
local function grab(v,...)
    local temp = { ... }
    context("%s = % + t",v,temp)
end

local function grab(v,... temp)
    context("%s = % + t",v,temp)
end
\stopluacode
```

At the moment of writing this, it works but it is a bit less performing than the first variant. Of course this might change: we're talking 'beta versions' here.

## 1.9 A few suggestions

You can wrap all kind of functionality in functions but sometimes it makes no sense to add the overhead of a call as the same can be done with hardly any code.

If you want a slice of a table, you can copy the range needed to a new table. A simple version with no bounds checking is:

```
local new = { } for i=a,b do new[#new+1] = old[i] end
```

Another, much faster, variant is the following.

```
local new = { unpack(old,a,b) }
```

You can use this variant for slices that are not extremely large. The function `table.sub` is an equivalent:

```
local new = table.sub(old,a,b)
```

An indexed table is empty when its size equals zero:

```
if #indexed == 0 then ... else ... end
```

Sometimes this is better:

```
if indexed and #indexed == 0 then ... else ... end
```

So how do we test if a hashed table is empty? We can use the `next` function as in:

```
if hashed and next(indexed) then ... else ... end
```

Say that we have the following table:

```
local t = { a=1, b=2, c=3 }
```

The call `next(t)` returns the first key and value:

```
local k, v = next(t)   -- "a", 1
```

The second argument to `next` can be a key in which case the following key and value in the hash table is returned. The result is not predictable as a hash is unordered. The generic for loop uses this to loop over a hashed table:

```
for k, v in next, t do
    ...
end
```

Anyway, when `next(t)` returns zero you can be sure that the table is empty. This is how you can test for exactly one entry:

```
if t and not next(t,next(t)) then ... else ... end
```

Here it starts making sense to wrap it into a function.

```
function table.has_one_entry(t)
    t and not next(t,next(t))
end
```

On the other hand, this is not that usefull, unless you can spent the runtime on it:

```
function table.is_empty(t)
    return not t or not next(t)
end
```

## 1.10 Interfacing

We have already seen that you can embed LUA code using commands like:

```
\startluacode
    print("this works")
\stopluacode
```

This command should not be confused with:

```
\startlua
    print("this works")
\stoplua
```

The first variant has its own catcode regime which means that tokens between the start and stop command are treated as LUA tokens, with the exception of TEX commands. The second variant operates under the regular TEX catcode regime.

Their short variants are `\ctxluacode` and `\ctxlua` as in:

```
\ctxluacode{print("this works")}
\ctxlua{print("this works")}
```

In practice you will probably use `\startluacode` when using or defining a blob of LUA and `\ctxlua` for inline code. Keep in mind that the longer versions need more initialization and have more overhead.

There are some more commands. For instance `\ctxcommand` can be used as an efficient way to access functions in the `commands` namespace. The following two calls are equivalent:

```
\ctxlua    {commands.thisorthat("...")}
\ctxcommand         {thisorthat("...")}
```

There are a few shortcuts to the `context` namespace. Their use can best be seen from their meaning:

```
\cldprocessfile#1{\directlua{context.runfile("#1")}}
\cldloadfile   #1{\directlua{context.loadfile("#1")}}
\cldcontext    #1{\directlua{context(#1)}}
\cldcommand    #1{\directlua{context.#1}}
```

The `\directlua{}` command can also be implemented using the token parser and LUA itself. A variant is therefore `\luascript{}` which can be considered an alias but with a bit different error reporting. A variant on this is the `\luathread{name} {code}` command. Here is an example of their usage:

```
\luascript        {        context("foo 1:") context(i) } \par
\luathread {test} { i = 10 context("bar 1:") context(i) } \par
\luathread {test} {        context("bar 2:") context(i) } \par
```

```
\luathread {test} {} % resets
\luathread {test} {       context("bar 3:") context(i) } \par
\luascript        {       context("foo 2:") context(i) } \par
```

These commands result in:

foo 1:
bar 1:10
bar 2:10
bar 3:
foo 2:

The variable `i` is local to the thread (which is not really a thread in LUA but more a named piece of code that provides an environment which is shared over the calls with the same name. You will probably never need these.

Each time a call out to LUA happens the argument eventually gets parsed, converted into tokens, then back into a string, compiled to bytecode and executed. The next example code shows a mechanism that avoids this:

```
\startctxfunction MyFunctionA
    context(" A1 ")
\stopctxfunction

\startctxfunctiondefinition MyFunctionB
    context(" B2 ")
\stopctxfunctiondefinition
```

The first command associates a name with some LUA code and that code can be executed using:

```
\ctxfunction{MyFunctionA}
```

The second definition creates a command, so there we do:

```
\MyFunctionB
```

There are some more helpers but for use in document sources they make less sense. You can always browse the source code for examples.

## 1.11 LUA versions

The LUA language is rather stable. The main change over the LUATEX life cycle has been the numbers going from doubles to either integers or doubles. This has a side effect of numbers being printed as `10.0` or `10` depending on their history. For instance, multiplication or division can convert an integer into a double even if the result is an integer.

Some changes have little consequence, for instance the addition of bitwise operators. Of course it means that we can ditch the function variant but old code still works. The change of the (default) garbage collector of course can have impact but there has been no change needed in coding. This is less true when a new keyword is introduced. The `<constant>` prefix is harmless, but `global` actually forced us to no longer use that as equivalent for `_G`.

A more significant, but for the (rather large) CONTEXT code base relatively harmless change was that iterators use local constant variables. This means that:

```
\startluacode
for i=1,100 do
    i = 2 * i
    ...
end
\stopluacode
```

is no longer valid, although

```
\startluacode
for i=1,100 do
    local i = 2 * i
    ...
end
\stopluacode
```

is. Lucky us that it is intercepted by the LUA parser, so we immediately know that we need to fix it.

One of the reasons that we keep up with (even beta) versions is that because we have a large code base, we're able to detect issues. Consider it our way to contribute to the testing. We're currently at the development variant of 5.5 as published in the not official repository.

# 2 Getting started

## 2.1 Some basics

I assume that you have either the so called CONTEXT standalone (formerly known as minimals) installed or TEXLIVE. You only need LUATEX and can forget about installing PDFTEX or XETEX, which saves you some megabytes and hassle. Now, from the users perspective a CONTEXT run goes like:

```
context yourfile
```

and by default a file with suffix `tex`, `mkvi` or `mkvi` will be processed. There are however a few other options:

```
context yourfile.xml
context yourfile.rlx --forcexml
context yourfile.lua
context yourfile.pqr --forcelua
context yourfile.cld
context yourfile.xyz --forcecld
context yourfile.mp
context yourfile.xyz --forcemp
```

When processing a LUA file the given file is loaded and just processed. This options will seldom be used as it is way more efficient to let `mtxrun` process that file. However, the last two variants are what we will discuss here. The suffix `cld` is a shortcut for CONTEXT LUA Document.

A simple `cld` file looks like this:

```
context.starttext()
context.chapter("Hello There!")
context.stoptext()
```

So yes, you need to know the CONTEXT commands in order to use this mechanism. In spite of what you might expect, the codebase involved in this interface is not that large. If you know CONTEXT, and if you know how to call commands, you basically can use this LUA method.

The examples that I will give are either (sort of) standalone, i.e. they are dealt with from LUA, or they are run within this document. Therefore you will see two patterns. If you want to make your own documentation, then you can use this variant:

```
\startbuffer
context("See this!")
\stopbuffer

\typebuffer \ctxluabuffer
```

I use anonymous buffers here but you can also use named ones. The other variant is:

```
\startluacode
context("See this!")
```

```
\stopluacode
```

This will process the code directly. Of course we could have encoded this document completely in LUA but that is not much fun for a manual.

## 2.2 The main command

There are a few rules that you need to be aware of. First of all no syntax checking is done. Second you need to know what the given commands expects in terms of arguments. Third, the type of your arguments matters:

`nothing` : just the command, no arguments
`string` : an argument with curly braces
`array` : a list between square backets (sometimes optional)
`hash` : an assignment list between square brackets
`boolean` : when `true` a newline is inserted
: when `false`, omit braces for the next argument

In the code above you have seen examples of this but here are some more:

```
context.chapter("Some title")
context.chapter({ "first" }, "Some title")
context.startchapter({ title = "Some title", label = "first" })
```

This blob of code is equivalent to:

```
\chapter{Some title}
\chapter[first]{Some title}
\startchapter[title={Some title},label=first]
```

You can simplify the third line of the LUA code to:

```
context.startchapter { title = "Some title", label = "first" }
```

In case you wonder what the distinction is between square brackets and curly braces: the first category of arguments concerns settings or lists of options or names of instances while the second category normally concerns some text to be typeset.

Strings are interpreted as TEX input, so:

```
context.mathematics("\\sqrt{2^3}")
```

and if you don't want to escape:

```
context.mathematics([[\sqrt{2^3}]])
```

are both correct. As TEX math is a language in its own and a de-facto standard way of inputting math this is quite natural, even at the LUA end.

## 2.3 Spaces and Lines

In a regular TEX file, spaces and newline characters are collapsed into one space. At the LUA end the same happens. Compare the following examples. First we omit spaces:

```
context("left")
context("middle")
context("right")
```

leftmiddleright

Next we add spaces:

```
context("left")
context(" middle ")
context("right")
```

left middle right

We can also add more spaces:

```
context("left ")
context(" middle ")
context(" right")
```

left middle right

In principle all content becomes a stream and after that the TEX parser will do its normal work: collapse spaces unless configured to do otherwise. Now take the following code:

```
context("before")
context("word 1")
context("word 2")
context("word 3")
context("after")
```

beforeword 1word 2word 3after

Here we get no spaces between the words at all, which is what we expect. So, how do we get lines (or paragraphs)?

```
context("before")
context.startlines()
context("line 1")
context("line 2")
context("line 3")
context.stoplines()
context("after")
```

before

line 1line 2line 3

after

This does not work out well, as again there are no lines seen at the TEX end. Newline tokens are injected by passing `true` to the `context` command:

```
context("before")
```

```
context.startlines()
context("line 1") context(true)
context("line 2") context(true)
context("line 3") context(true)
context.stoplines()
context("after")
```

before

line 1
line 2
line 3

after

Don't confuse this with:

```
context("before") context.par()
context("line 1") context.par()
context("line 2") context.par()
context("line 3") context.par()
context("after")  context.par()
```

before

line 1

line 2

line 3

after

There we use the regular \par command to finish the current paragraph and normally you will use that method. In that case, when set, whitespace will be added between paragraphs.

This newline issue is a somewhat unfortunate inheritance of traditional TEX, where \n and \r mean something different. I'm still not sure if the CLD do the right thing as dealing with these tokens also depends on the intended effect. Catcodes as well as the LUATEX input parser also play a role. Anyway, the following also works:

```
context.startlines()
context("line 1\n")
context("line 2\n")
context("line 3\n")
context.stoplines()
```

## 2.4 Direct output

The CONTEXT user interface is rather consistent and the use of special input syntaxes is discouraged. Therefore, the LUA interface using tables and strings works quite well. However, imagine that you need to support some weird macro (or a primitive) that does not expect its argument between curly

braces or brackets. The way out is to precede an argument by another one with the value `false`. We call this the direct interface. This is demonstrated in the following example.

```
\unexpanded\def\bla#1{[#1]}

\startluacode
context.bla(false,"***")
context.par()
context.bla("***")
\stopluacode
```

This results in:

[*]**

[***]

Here, the first call results in three `*` being passed, and `#1` picks up the first token. The second call to `bla` gets `{***}` passed so here `#1` gets the triplet. In practice you will seldom need the direct interface.

In CONTEXT for historical reasons, combinations accept the following syntax:

```
\startcombination % optional specification, like [2*3]
  {\framed{content one}} {caption one}
  {\framed{content two}} {caption two}
\stopcombination
```

You can also say:

```
\startcombination
  \combination {\framed{content one}} {caption one}
  \combination {\framed{content two}} {caption two}
\stopcombination
```

When coded in LUA, we can feed the first variant as follows:

```
context.startcombination()
  context.direct("one","two")
  context.direct("one","two")
context.stopcombination()
```

To give you an idea what this looks like, we render it:

one one
two two

So, the `direct` function is basically a no-op and results in nothing by itself. Only arguments are passed. An equivalent but bit more ugly looking is:

```
context.startcombination()
  context(false,"one","two")
  context(false,"one","two")
```

```
context.stopcombination()
```

## 2.5 Catcodes

If you are familiar with the inner working of TEX, you will know that characters can have special meanings. This meaning is determined by their catcodes.

```
context("$x=1$")
```

This gives: $x = 1$ because the dollar tokens trigger inline math mode. If you think that this is annoying, you can do the following:

```
context.pushcatcodes("text")
context("$x=1$")
context.popcatcodes()
```

Now we get: $x=1$. There are several catcode regimes of which only a few make sense in the perspective of the cld interface.

| | |
|---|---|
| `ctx, ctxcatcodes, context` | the normal CONTEXT catcode regime |
| `prt, prtcatcodes, protect` | the CONTEXT protected regime, used for modules |
| `tex, texcatcodes, plain` | the traditional (plain) TEX regime |
| `txt, txtcatcodes, text` | the CONTEXT regime but with less special characters |
| `vrb, vrbcatcodes, verbatim` | a regime specially meant for verbatim |
| `xml, xmlcatcodes` | a regime specially meant for XML processing |

In the second case you can still get math:

```
context.pushcatcodes("text")
context.mathematics("x=1")
context.popcatcodes()
```

When entering a lot of math you can also consider this:

```
context.startimath()
context("x")
context("=")
context("1")
context.stopimath()
```

Module writers of course can use `unprotect` and `protect` as they do at the TEX end.

As we've seen, a function call to `context` acts like a print, as in:

```
context("test ")
context.bold("me")
context(" first")
```

test **me** first

When more than one argument is given, the first argument is considered a format conforming the `string.format` function.

```
context.startimath()
context("%s = %0.5f",utf.char(0x03C0),math.pi)
context.stopimath()
```

$\pi = 3.14159$

This means that when you say:

```
context(a,b,c,d,e,f)
```

the variables `b` till `f` are passed to the format and when the format does not use them, they will not end up in your output.

```
context("%s %s %s",1,2,3)
context(1,2,3)
```

The first line results in the three numbers being typeset, but in the second case only the number 1 is typeset.

# 3 More on functions

## 3.1 Why we need them

In a previous chapter we introduced functions as arguments. At first sight this feature looks strange but you need to keep in mind that a call to a `context` function has no direct consequences. It generates TEX code that is executed after the current LUA chunk ends and control is passed back to TEX. Take the following code:

```
context.framed( {
    frame = "on",
    offset = "5mm",
    align = "middle"
  },
  context.input("knuth")
)
```

We call the function `framed` but before the function body is executed, the arguments get evaluated. This means that `input` gets processed before `framed` gets done. As a result there is no second argument to `framed` and no content gets passed: an error is reported. This is why we need the indirect call:

```
context.framed( {
    frame = "on",
    align = "middle"
  },
  function() context.input("knuth") end
)
```

This way we get what we want:

> Thus, I came to the conclusion that the designer of a new system must not only be the implementer and first large–scale user; the designer should also write the first user manual. The separation of any of these four components would have hurt TEX significantly. If I had not participated fully in all these activities, literally hundreds of improvements would never have been made, because I would never have thought of them or perceived why they were important. But a system cannot be successful if it is too strongly influenced by a single person. Once the initial design is complete and fairly robust, the real test begins as people with many different viewpoints undertake their own experiments.

The function is delayed till the `framed` command is executed. If your applications use such calls a lot, you can of course encapsulate this ugliness:

```
mycommands = mycommands or { }

function mycommands.framed_input(filename)
  context.framed( {
    frame = "on",
    align = "middle"
```

```
  },
  function() context.input(filename) end
end

mycommands.framed_input("knuth")
```

Of course you can nest function calls:

```
context.placefigure(
  "caption",
  function()
    context.framed( {
      frame = "on",
      align = "middle"
    },
      function() context.input("knuth") end
    )
  end
)
```

Or you can use a more indirect method:

```
function text()
  context.framed( {
      frame = "on",
      align = "middle"
    },
    function() context.input("knuth") end
  )
end

context.placefigure(
  "none",
  function() text() end
)
```

You can develop your own style and libraries just like you do with regular LUA code. Browsing the already written code can give you some ideas.

## 3.2 How we can avoid them

As many nested functions can obscure the code rather quickly, there is an alternative. In the following examples we use `test`:

```
\def\test#1{[#1]}

context.test("test 1 ",context("test 2a")," test 3")
```

This gives: test 2a[test 1 ] test 3. As you can see, the second argument is executed before the encapsulating call to `test`. So, we should have packed it into a function but here is an alternative:

```
context.test("test 1 ",context.delayed("test 2a")," test 3")
```

Now we get: [test 1 ]test 2a test 3. We can also delay functions themselves, look at this:

```
context.test("test 1 ",context.delayed.test("test 2b")," test 3")
```

The result is: [test 1 ][test 2b] test 3. This feature also conveniently permits the use of temporary variables, as in:

```
local f = context.delayed.test("test 2c")
context("before ",f," after")
```

Of course you can limit the amount of keystrokes even more by creating a shortcut:

```
local delayed = context.delayed

context.test("test 1 ",delayed.test("test 2")," test 3")
context.test("test 4 ",delayed.test("test 5")," test 6")
```

So, if you want you can produce rather readable code and readability of code is one of the reasons why LUA was chosen in the first place. This is a good example of why coding in TEX makes sense as it looks more intuitive:

```
\test{test 1 \test{test 2} test 3}
\test{test 4 \test{test 5} test 6}
```

The `context.nested` variant is now an alias to `context.delayed` and no longer builds a string representation.

## 3.3 Trial typesetting

Some typesetting mechanisms demand a preroll. For instance, when determining the most optimal way to analyse and therefore typeset a table, it is necessary to typeset the content of cells first. Inside CONTEXT there is a state tagged 'trial typesetting' which signals other mechanisms that for instance counters should not be incremented more than once.

Normally you don't need to worry about these issues, but when writing the code that implements the LUA interface to CONTEXT, it definitely had to be taken into account as we either or not can free cached (nested) functions.

You can influence this caching to some extend. If you say

```
function()
  context("whatever")
end
```

the function will be removed from the cache when CONTEXT is not in the trial typesetting state. You can prevent removal of a function by returning `true`, as in:

```
function()
  context("whatever")
  return true
```

```
end
```

Whenever you run into a situation that you don't get the outcome that you expect, you can consider returning `true`. However, keep in mind that it will take more memory, something that only matters on big runs. You can force flushing the whole cache by:

```
context.restart()
```

An example of an occasion where you need to keep the function available is in repeated content, for instance in headers and footers.

```
context.setupheadertexts {
  function()
    context.pagenumber()
    return true
  end
}
```

Of course it is not needed when you use the following method:

```
context.pagenumber("pagenumber")
```

Because here CONTEXT itself deals with the content driven by the keyword `pagenumber`.

## 3.4 Steppers

The `context` commands are accumulated within a `\ctxlua` call and only after the call is finished, control is back at the TEX end. Sometimes you want (in your LUA code) to go on and pretend that you jump out to TEX for a moment, but come back to where you left. The stepper mechanism permits this.

A not so practical but nevertheless illustrative example is the following:

```
\startluacode
  context.stepwise (function()
    context.startitemize()
       context.startitem()
         context.step("BEFORE 1")
       context.stopitem()
       context.step("\\setbox0\\hbox{!!!!}")
       context.startitem()
         context.step("%p",tex.getbox(0).width)
       context.stopitem()
       context.startitem()
         context.step("BEFORE 2")
       context.stopitem()
       context.step("\\setbox2\\hbox{????}")
       context.startitem()
         context.step("%p",tex.getbox(2).width)
       context.startitem()
```

```
      context.step("BEFORE 3")
    context.stopitem()
    context.startitem()
      context.step("\\copy0\\copy2")
    context.stopitem()
    context.startitem()
      context.step("BEFORE 4")
      context.startitemize()
        context.stepwise (function()
          context.step("\\bgroup")
          context.step("\\setbox0\\hbox{>>>>}")
          context.startitem()
            context.step("%p",tex.getbox(0).width)
          context.stopitem()
          context.step("\\setbox2\\hbox{<<<<}")
          context.startitem()
            context.step("%p",tex.getbox(2).width)
          context.stopitem()
          context.startitem()
            context.step("\\copy0\\copy2")
          context.stopitem()
          context.startitem()
            context.step("\\copy0\\copy2")
          context.stopitem()
          context.step("\\egroup")
        end)
      context.stopitemize()
    context.stopitem()
    context.startitem()
      context.step("AFTER 1\\par")
    context.stopitem()
    context.startitem()
      context.step("\\copy0\\copy2\\par")
    context.stopitem()
    context.startitem()
      context.step("\\copy0\\copy2\\par")
    context.stopitem()
    context.startitem()
      context.step("AFTER 2\\par")
    context.stopitem()
    context.startitem()
      context.step("\\copy0\\copy2\\par")
    context.stopitem()
    context.startitem()
      context.step("\\copy0\\copy2\\par")
    context.stopitem()
  context.stopitemize()
end)
```

`\stopluacode`

This gives an (ugly) itemize with a nested one:

- BEFORE 1
- 12.22528pt
- BEFORE 2
- 19.52527pt
- BEFORE 3
- !!!!????
- BEFORE 4
  - 33.72943pt
  - 33.72943pt
  - >>>><<<<
  - >>>><<<<
- AFTER 1
- !!!!????
- !!!!????
- AFTER 2
- !!!!????
- !!!!????

As you can see in the code, the `step` call accepts multiple arguments, but when more than one argument is given the first one is treated as a formatter.

# 4 A few Details

## 4.1 Variables

Normally it makes most sense to use the English version of CONTEXT. The advantage is that you can use English keywords, as in:

```
context.framed( {
    frame = "on",
  },
  "some text"
)
```

If you use the Dutch interface it looks like this:

```
context.omlijnd( {
    kader = "aan",
  },
  "wat tekst"
)
```

A rather neutral way is:

```
context.framed( {
    frame = interfaces.variables.on,
  },
  "some text"
)
```

But as said, normally you will use the English user interface so you can forget about these matters. However, in the CONTEXT core code you will often see the variables being used this way because there we need to support all user interfaces.

## 4.2 Modes

Context carries a concept of modes. You can use modes to create conditional sections in your style (and/or content). You can control modes in your styles or you can set them at the command line or in job control files. When a mode test has to be done at processing time, then you need constructs like the following:

```
context.doifmodeelse( "screen",
  function()
      ... -- mode == screen
  end,
  function()
      ... -- mode ~= screen
  end
)
```

However, often a mode does not change during a run, and then we can use the following method:

```
if tex.modes["screen"] then
  ...
else
  ...
end
```

Watch how the `modes` table lives in the `tex` namespace. We also have `systemmodes`. At the TEX end these are mode names preceded by a `*`, so the following code is similar:

```
if tex.modes["*mymode"] then
  -- this is the same
elseif tex.systemmodes["mymode"] then
  -- test as this
else
  -- but not this
end
```

Inside CONTEXT we also have so called constants, and again these can be consulted at the LUA end:

```
if tex.constants["someconstant'] then
  ...
else
  ...
end
```

But you will hardly need these and, as they are often not public, their meaning can change, unless of course they *are* documented as public.

## 4.3 Token lists

There is normally no need to mess around with nodes and tokens at the LUA end yourself. However, if you do, then you might want to flush them as well. Say that at the TEX end we have said:

```
\toks0 = {Don't get \inframed{framed}!}
```

Then at the LUA end you can say:

```
context(tex.toks[0])
```

and get: Don't get framed! In fact, token registers are exposed as strings so here, register zero has type `string` and is treated as such.

```
context("< %s >",tex.toks[0])
```

This gives: < Don't get framed! >. But beware, if you go the reverse way, you don't get what you might expect:

```
tex.toks[0] = [[\framed{oeps}]]
```

If we now say `\the\toks0` we will get `\framed{oeps}` as all tokens are considered to be letters.

## 4.4 Node lists

If you're not deep into TEX you will never feel the need to manipulate node lists yourself, but you might want to flush boxes. As an example we put something in box zero (one of the scratch boxes).

```
\setbox0 = \hbox{Don't get \inframed{framed}!}
```

At the TEX end you can flush this box (`\box0`) or take a copy (`\copy0`). At the LUA end you would do:

```
context.copy()
context.direct(0)
```

or:

```
context.copy(false,0)
```

but this works as well:

```
context(node.copylist(tex.box[0]))
```

So we get: Don't get framed! If you do:

```
context(tex.box[0])
```

you also need to make sure that the box is freed but let's not go into those details now.

Here is an example if messing around with node lists that get seen before a paragraph gets broken into lines, i.e. when hyphenation, font manipulation etc take place. First we define some colors:

```
\definecolor[mynesting:0][r=.6]
\definecolor[mynesting:1][g=.6]
\definecolor[mynesting:2][r=.6,g=.6]
```

Next we define a function that colors nodes in such a way that we can see the different processing stages.

```
\startluacode
local enabled  = false
local count    = 0
local setcolor = nodes.tracers.colors.set

function userdata.processmystuff(head)
    if enabled then
        local color = "mynesting:" .. (count % 3)
     -- for n in node.traverse(head) do
        for n in node.traverseid(nodes.nodecodes.glyph,head) do
            setcolor(n,color)
        end
        count = count + 1
        return head, true
    end
```

```
    return head, false
end

function userdata.enablemystuff()
    enabled = true
end

function userdata.disablemystuff()
    enabled = false
end
\stopluacode
```

We hook this function into the normalizers category of the processor callbacks:

```
\startluacode
nodes.tasks.appendaction("processors", "normalizers", "userdata.processmystuff")
\stopluacode
```

We now can enable this mechanism and show an example:

```
\startbuffer
Node lists are processed \hbox {nested from \hbox{inside} out} which is not
what you might expect. But, \hbox{coloring} does not \hbox {happen} really
nested here, more \hbox {in} \hbox {the} \hbox {order} \hbox {of} \hbox
{processing}.
\stopbuffer

\ctxlua{userdata.enablemystuff()}
\par \getbuffer \par
\ctxlua{userdata.disablemystuff()}
```

The \par is needed because otherwise the processing is already disabled before the paragraph gets seen by TEX.

Node lists are processed nested from inside out which is not what you might expect. But, coloring does not happen really nested here, more in the order of processing.

```
\startluacode
nodes.tasks.disableaction("processors", "userdata.processmystuff")
\stopluacode
```

Instead of using an boolean to control the state, we can also do this:

```
\startluacode
local count    = 0
local setcolor = nodes.tracers.colors.set

function userdata.processmystuff(head)
    count = count + 1
    local color = "mynesting:" .. (count % 3)
    for n in node.traverseid(nodes.nodecodes.glyph,head) do
```

```
        setcolor(n,color)
    end
    return head, true
end

nodes.tasks.appendaction("processors", "after", "userdata.processmystuff")
\stopluacode
```

Disabling now happens with:

```
\startluacode
nodes.tasks.disableaction("processors", "userdata.processmystuff")
\stopluacode
```

As you might want to control these things in more details, a simple helper mechanism was made: markers. The following example code shows the way:

```
\definemarker[mymarker]
```

Again we define some colors:

```
\definecolor[mymarker:1][r=.6]
\definecolor[mymarker:2][g=.6]
\definecolor[mymarker:3][r=.6,g=.6]
```

The LUA code like similar to the code presented before:

```
\startluacode
local setcolor   = nodes.tracers.colors.setlist
local getmarker  = nodes.markers.get
local hlist_code = nodes.nodecodes.hlist
local traverseid = node.traverseid

function userdata.processmystuff(head)
    for n in traverseid(hlist_code,head) do
        local m = getmarker(n,"mymarker")
        if m then
            setcolor(n.list,"mymarker:" .. m)
        end
    end
    return head, true
end

nodes.tasks.appendaction("processors", "after", "userdata.processmystuff")
nodes.tasks.disableaction("processors", "userdata.processmystuff")
\stopluacode
```

This time we disabled the processor (if only because in this document we don't want the overhead.

```
\startluacode
nodes.tasks.enableaction("processors", "userdata.processmystuff")
```

```
\stopluacode

Node lists are processed \hbox \boxmarker{mymarker}{1} {nested from \hbox{inside}
out} which is not what you might expect. But, \hbox {coloring} does not \hbox
{happen} really nested here, more \hbox {in} \hbox \boxmarker{mymarker}{2} {the}
\hbox {order} \hbox {of} \hbox \boxmarker{mymarker}{3} {processing}.

\startluacode
nodes.tasks.disableaction("processors", "userdata.processmystuff")
\stopluacode
```

The result looks familiar:

Node lists are processed nested from inside out which is not what you might expect. But, coloring does not happen really nested here, more in the order of processing.

# 5 Some more examples

## 5.1 Appetizer

Before we give some more examples, we will have a look at the way the title page is made. This way you get an idea what more is coming.

```
local todimen, random = number.todimen, math.random

context.startTEXpage()

local paperwidth  = tex.dimen.paperwidth
local paperheight = tex.dimen.paperheight
local nofsteps    = 25
local firstcolor  = "darkblue"
local secondcolor = "white"

context.definelayer(
    { "titlepage" }
)

context.setuplayer(
    { "titlepage" },
    {
        width  = todimen(paperwidth),
        height = todimen(paperheight),
    }
)

context.setlayerframed(
    { "titlepage" },
    { offset = "-5pt" },
    {
        width  = todimen(paperwidth),
        height = todimen(paperheight),
        background = "color",
        backgroundcolor = firstcolor,
        backgroundoffset = "10pt",
        frame = "off",
    },
    ""
)

local settings = {
    frame = "off",
    background = "color",
    backgroundcolor = secondcolor,
    foregroundcolor = firstcolor,
```

```
    foregroundstyle = "type",
}

for i=1, nofsteps do
    for j=1, nofsteps do
        context.setlayerframed(
            { "titlepage" },
            {
                x = todimen((i-1) * paperwidth /nofsteps),
                y = todimen((j-1) * paperheight/nofsteps),
                rotation = random(360),
            },
            settings,
            "CLD"
        )
    end
end

context.tightlayer(
    { "titlepage" }
)

context.stopTEXpage()

return true
```

This does not look that bad, does it? Of course in pure TEX code it looks mostly the same but loops and calculations feel a bit more natural in LUA then in TEX. The result is shown in **figure 5.1**. The actual cover page was derived from this.

## 5.2 A few examples

As it makes most sense to use the LUA interface for generated text, here is another example with a loop:

```
context.startitemize { "a", "packed", "two" }
  for i=1,10 do
    context.startitem()
      context("this is item %i",i)
    context.stopitem()
  end
context.stopitemize()
```

a. this is item 1
b. this is item 2
c. this is item 3
d. this is item 4
e. this is item 5
f. this is item 6

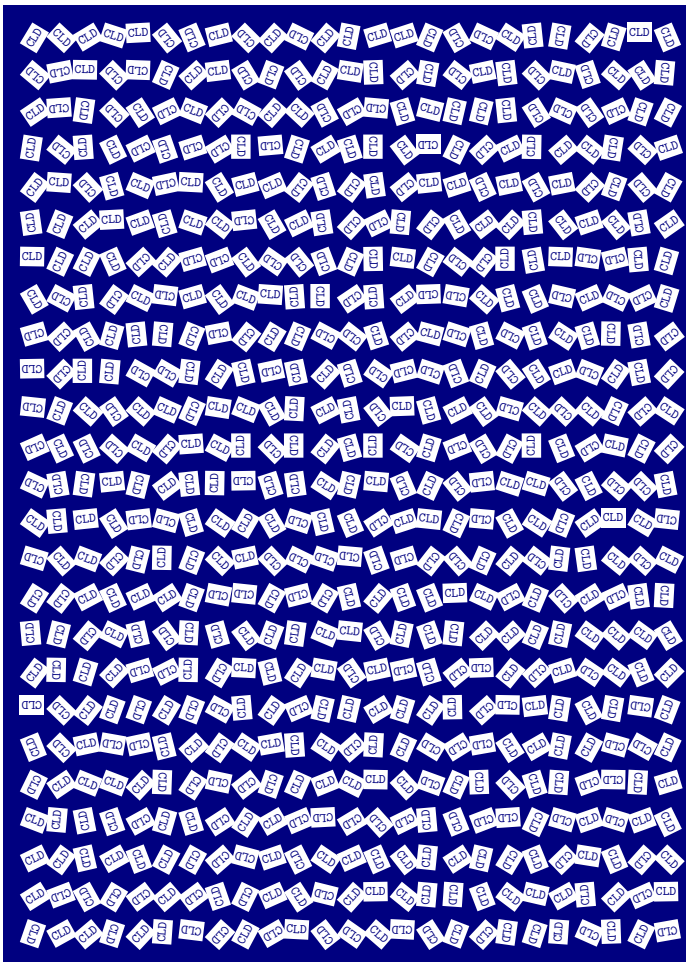


**Figure 5.1** The simplified cover page.

g. this is item 7
h. this is item 8
i. this is item 9
j. this is item 10

Just as you can mix TEX with XML and METAPOST, you can define bits and pieces of a document in LUA. Tables are good candidates:

```
local one = {
  align = "middle",
  style = "type",
}
local two = {
  align = "middle",
  style = "type",
  background = "color",
  backgroundcolor = "darkblue",
  foregroundcolor = "white",
}
local random = math.random
context.bTABLE { framecolor = "darkblue" }
    for i=1,10 do
      context.bTR()
```

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 52 | 75 | 75 | 6 | 49 | 26 | 92 | 10 | 43 | 39 | 57 | 80 | 15 | 21 | 23 | 52 | 24 | 62 | 55 | 59 |
| 86 | 55 | 31 | 77 | 89 | 35 | 98 | 26 | 93 | 63 | 51 | 66 | 65 | 35 | 35 | 53 | 75 | 15 | 98 | 9 |
| 91 | 47 | 20 | 17 | 67 | 20 | 82 | 92 | 42 | 72 | 41 | 85 | 80 | 44 | 88 | 22 | 57 | 8 | 58 | 72 |
| 96 | 40 | 8 | 89 | 65 | 68 | 67 | 65 | 92 | 41 | 46 | 22 | 22 | 70 | 59 | 69 | 17 | 73 | 7 | 33 |
| 2 | 72 | 10 | 40 | 59 | 77 | 27 | 2 | 85 | 21 | 98 | 51 | 93 | 67 | 58 | 88 | 45 | 30 | 22 | 94 |
| 39 | 43 | 13 | 58 | 43 | 55 | 63 | 13 | 83 | 91 | 42 | 14 | 99 | 66 | 20 | 7 | 54 | 82 | 54 | 4 |
| 52 | 83 | 93 | 2 | 41 | 25 | 56 | 6 | 2 | 47 | 40 | 91 | 41 | 60 | 81 | 88 | 28 | 97 | 70 | 51 |
| 19 | 29 | 26 | 83 | 96 | 66 | 48 | 35 | 28 | 79 | 96 | 65 | 30 | 20 | 52 | 42 | 32 | 65 | 53 | 42 |
| 50 | 36 | 89 | 83 | 70 | 74 | 67 | 37 | 93 | 24 | 24 | 51 | 22 | 85 | 76 | 54 | 69 | 93 | 34 | 92 |
| 67 | 34 | 7 | 94 | 85 | 51 | 28 | 81 | 55 | 54 | 51 | 77 | 42 | 7 | 70 | 15 | 8 | 25 | 4 | 13 |

**Table 5.1** A table generated by LUA.

```
      for i=1,20 do
          local r = random(99)
          context.bTD(r < 50 and one or two)
          context("%2i",r)
          context.eTD()
      end
      context.eTR()
    end
context.eTABLE()
```

Here we see a function call to `context` in the most indented line. The first argument is a format that makes sure that we get two digits and the random number is substituted into this format. The result is shown in **table 5.1**. The line correction is ignored when we use this table as a float, otherwise it assures proper vertical spacing around the table. Watch how we define the tables `one` and `two` beforehand. This saves 198 redundant table constructions.

Not all code will look as simple as this. Consider the following:

```
context.placefigure(
  "caption",
  function() context.externalfigure( { "cow.pdf" } ) end
)
```

Here we pass an argument wrapped in a function. If we would not do that, the external figure would end up wrong, as arguments to functions are evaluated before the function that gets them (we already showed some alternative approaches in previous chapters). A function argument is treated as special and in this case the external figure ends up right. Here is another example:

```
context.placefigure("Two cows!",function()
  context.bTABLE()
    context.bTR()
      context.bTD()
        context.externalfigure(
            { "cow.pdf" },
            { width = "3cm", height = "3cm" }
```

```
        )
      context.eTD()
      context.bTD { align = "{lohi,middle}" }
        context("and")
      context.eTD()
      context.bTD()
        context.externalfigure(
            { "cow.pdf" },
            { width = "4cm", height = "3cm" }
        )
      context.eTD()
    context.eTR()
  context.eTABLE()
end)
```

In this case the figure is not an argument so it gets flushed sequentially with the rest.

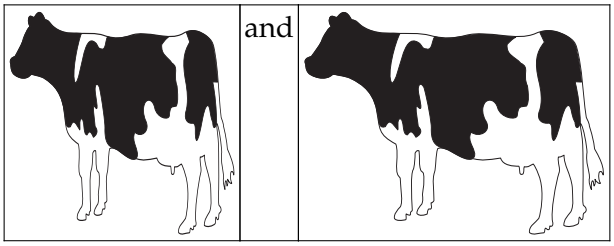



**Figure 5.2** Two cows!

## 5.3 Styles

Say that you want to typeset a word in a bold font. You can do that this way:

```
context("This is ")
context.bold("important")
context("!")
```

Now imagine that you want this important word to be in red too. As we have a nested command, we end up with a nested call:

```
context("This is ")
context.bold(function() context.color( { "red" }, "important") end)
context("!")
```

or

```
context("This is ")
context.bold(context.delayed.color( { "red" }, "important"))
context("!")
```

In that case it's good to know that there is a command that combines both features:

```
context("This is ")
context.style( { style = "bold", color = "red" }, "important")
```

```
context("!")
```

But that is still not convenient when we have to do that often. So, you can wrap the style switch in a function.

```
local function mycommands.important(str)
    context.style( { style = "bold", color = "red" }, str )
end

context("This is ")
mycommands.important( "important")
context(", and ")
mycommands.important( "this")
context(" too !")
```

Or you can setup a named style:

```
context.setupstyle( { "important" }, { style = "bold", color = "red" } )

context("This is ")
context.style( { "important" }, "important")
context(", and ")
context.style( { "important" }, "this")
context(" too !")
```

Or even define one:

```
context.definestyle( { "important" }, { style = "bold", color = "red" } )

context("This is ")
context.important("important")
context(", and ")
context.important("this")
context(" too !")
```

This last solution is especially handy for more complex cases:

```
context.definestyle( { "important" }, { style = "bold", color = "red" } )

context("This is ")
context.startimportant()
context.inframed("important")
context.stopimportant()
context(", and ")
context.important("this")
context(" too !")
```

This is **important**, and **this** too !

## 5.4 A complete example

One day my 6 year old niece Lorien was at the office and wanted to know what I was doing. As I knew she was practicing arithmetic at school I wrote a quick and dirty script to generate sheets with exercises. The most impressive part was that the answers were included. It was a rather braindead bit of LUA, written in a few minutes, but the weeks after I ended up running it a few more times, for her and her friends, every time a bit more difficult and also using different arithmetic. It was that script that made me decide to extend the basic cld manual into this more extensive document.

We generate three columns of exercises. Each exercise is a row in a table. The last argument to the function determines if answers are shown.

```
local random = math.random

local function ForLorien(n,maxa,maxb,answers)
  context.startcolumns { n = 3 }
  context.starttabulate { "|r|c|r|c|r|" }
  for i=1,n do
    local sign = random(0,1) > 0.5
    local a, b = random(1,maxa or 99), random(1,max or maxb or 99)
    if b > a and not sign then a, b = b, a end
    context.NC()
    context(a)
    context.NC()
    context.mathematics(sign and "+" or "-")
    context.NC()
    context(b)
    context.NC()
    context("=")
    context.NC()
    context(answers and (sign and a+b or a-b))
    context.NC()
    context.NR()
  end
  context.stoptabulate()
  context.stopcolumns()
  context.page()
end
```

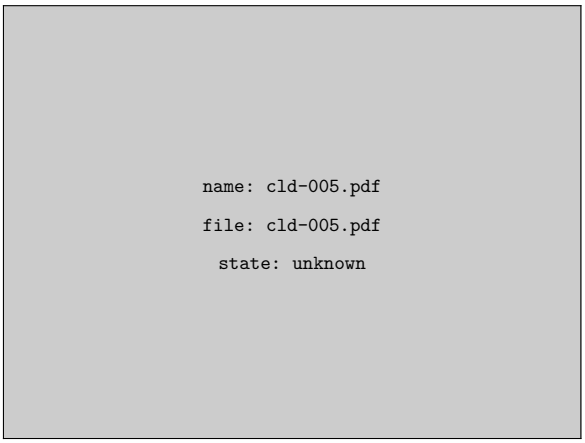
This is a typical example of where it's more convenient to write the code in LUA that in TEX's macro language. As a consequence setting up the page also happens in LUA:

```
context.setupbodyfont {
  "palatino",
  "14pt"
}

context.setuplayout {
  backspace = "2cm",
  topspace  = "2cm",
```

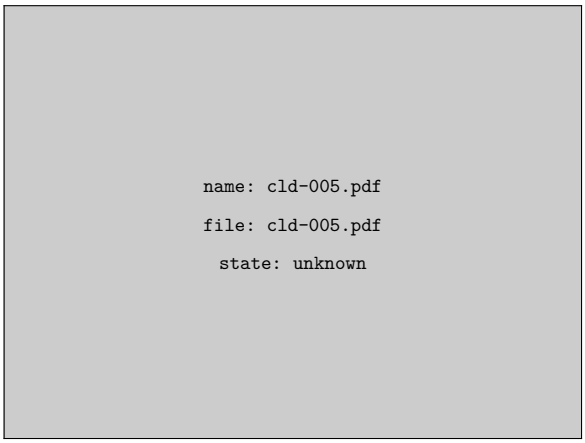
```
  header    = "1cm",
  footer    = "0cm",
  height    = "middle",
  width     = "middle",
}
```

This leave us to generate the document. There is a pitfall here: we need to use the same random number for the exercises and the answers, so we freeze and defrost it. Functions in the `commands` namespace implement functionality that is used at the TEX end but better can be done in LUA than in TEX macro code. Of course these functions can also be used at the LUA end.

```
context.starttext()

  local n = 120

  commands.freezerandomseed()

  ForLorien(n,10,10)
  ForLorien(n,20,20)
  ForLorien(n,30,30)
  ForLorien(n,40,40)
  ForLorien(n,50,50)

  commands.defrostrandomseed()

  ForLorien(n,10,10,true)
  ForLorien(n,20,20,true)
  ForLorien(n,30,30,true)
  ForLorien(n,40,40,true)
  ForLorien(n,50,50,true)

context.stoptext()
```




exercises




answers

**Figure 5.3** Lorien's challenge.

A few pages of the result are shown in **figure 5.3**. In the CONTEXT distribution a more advanced version can be found in `s-edu-01.cld` as I was also asked to generate multiplication and table exercises.

In the process I had to make sure that there were no duplicates on a page as she complained that was not good. There a set of sheets is generated with:

```
moduledata.educational.arithematic.generate {
  name     = "Bram Otten",
  fontsize = "12pt",
  columns  = 2,
  run      = {
    { method = "bin_add_and_subtract", maxa =   8, maxb =   8 },
    { method = "bin_add_and_subtract", maxa =  16, maxb =  16 },
    { method = "bin_add_and_subtract", maxa =  32, maxb =  32 },
    { method = "bin_add_and_subtract", maxa =  64, maxb =  64 },
    { method = "bin_add_and_subtract", maxa = 128, maxb = 128 },
  },
}
```

## 5.5 Interfacing

The fact that we can define functionality using LUA code does not mean that we should abandon the TEX interface. As an example of this we use a relatively simple module for typesetting morse code.[4] First we create a proper namespace:

```
moduledata.morse = moduledata.morse or { }
local morse      = moduledata.morse
```

We will use a few helpers and create shortcuts for them. The first helper loops over each UTF character in a string. The other two helpers map a character onto an uppercase (because morse only deals with uppercase) or onto an similar shaped character (because morse only has a limited character set).

```
local utfcharacters = string.utfcharacters
local ucchars, shchars = characters.ucchars, characters.shchars
```

The morse codes are stored in a table.

```
local codes = {

    ["A"] = "·−",     ["B"] = "−···",
    ["C"] = "−·−·",   ["D"] = "−··",
    ["E"] = "·",      ["F"] = "··−·",
    ["G"] = "−−·",    ["H"] = "····",
    ["I"] = "··",     ["J"] = "·−−−",
    ["K"] = "−·−",    ["L"] = "·−··",
    ["M"] = "−−",     ["N"] = "−·",
    ["O"] = "−−−",    ["P"] = "·−−·",
    ["Q"] = "−−·−",   ["R"] = "·−·",
    ["S"] = "···",    ["T"] = "−",
    ["U"] = "··−",    ["V"] = "···−",
```

[4] The real module is a bit larger and can format verbose morse.

```
    ["W"] = "·--",    ["X"] = "-··-",
    ["Y"] = "-·--",   ["Z"] = "--··",

    ["0"] = "-----",  ["1"] = "·----",
    ["2"] = "··---",  ["3"] = "···--",
    ["4"] = "····-",  ["5"] = "·····",
    ["6"] = "-····",  ["7"] = "--···",
    ["8"] = "---··",  ["9"] = "----·",

    ["."] = "·-·-·-", [","] = "--··--",
    [":"] = "---···", [";"] = "-·-·-",
    ["?"] = "··--··", ["!"] = "-·-·--",
    ["-"] = "-····-", ["/"] = "-··-· ",
    ["("] = "-·--·",  [")"] = "-·--·-",
    ["="] = "-···-",  ["@"] = "·--·-·",
    ["'"] = "·----·", ['""'] = "·-··-·",

    ["À"] = "·--·-",
    ["Å"] = "·--·-",
    ["Ä"] = "·-·-",
    ["Æ"] = "·-·-",
    ["Ç"] = "-·-··",
    ["É"] = "··-··",
    ["È"] = "·-··-",
    ["Ñ"] = "--·--",
    ["Ö"] = "---·",
    ["Ø"] = "---·",
    ["Ü"] = "··--",
    ["ß"] = "··· ···",

}

morse.codes = codes
```

As you can see, there are a few non ASCII characters supported as well. There will never be full UNICODE support simply because morse is sort of obsolete. Also, in order to support UNICODE one could as well use the bits of UTF characters, although . . . memorizing the whole UNICODE table is not much fun.

We associate a metatable index function with this mapping. That way we can not only conveniently deal with the casing, but also provide a fallback based on the shape. Once found, we store the representation so that only one lookup is needed per character.

```
local function resolvemorse(t,k)
    if k then
        local u = ucchars[k]
        local v = rawget(t,u) or rawget(t,shchars[u]) or false
        t[k] = v
        return v
    else
```

```
        return false
    end
end

setmetatable(codes, { __index = resolvemorse })
```

Next comes some rendering code. As we can best do rendering at the TEX end we just use macros.

```
local MorseBetweenWords      = context.MorseBetweenWords
local MorseBetweenCharacters = context.MorseBetweenCharacters
local MorseLong              = context.MorseLong
local MorseShort             = context.MorseShort
local MorseSpace             = context.MorseSpace
local MorseUnknown           = context.MorseUnknown
```

The main function is not that complex. We need to keep track of spaces and newlines. We have a nested loop because a fallback to shape can result in multiple characters.

```
function morse.tomorse(str)
    local inmorse = false
    for s in utfcharacters(str) do
        local m = codes[s]
        if m then
            if inmorse then
                MorseBetweenWords()
            else
                inmorse = true
            end
            local done = false
            for m in utfcharacters(m) do
                if done then
                    MorseBetweenCharacters()
                else
                    done = true
                end
                if m == "·" then
                    MorseShort()
                elseif m == "-" then
                    MorseLong()
                elseif m == " " then
                    MorseBetweenCharacters()
                end
            end
            inmorse = true
        elseif s == "\n" or s == " " then
            MorseSpace()
            inmorse = false
        else
            if inmorse then
                MorseBetweenWords()
```

```
            else
                inmorse = true
            end
            MorseUnknown(s)
        end
    end
end
```

We use this function in two additional functions. One typesets a file, the other a table of available codes.

```
function morse.filetomorse(name,verbose)
    morse.tomorse(resolvers.loadtexfile(name),verbose)
end

function morse.showtable()
    context.starttabulate { "|l|l|" }
    for k, v in table.sortedpairs(codes) do
        context.NC() context(k)
        context.NC() morse.tomorse(v,true)
        context.NC() context.NR()
    end
    context.stoptabulate()
end
```

We're done with the LUA code that we can either put in an external file or put in the module file. The TEX file has two parts. The typesetting macros that we use at the LUA end are defined first. These can be overloaded.

```
\def\MorseShort
  {\dontleavehmode
   \vrule
      width  \MorseWidth
      height \MorseHeight
      depth  \zeropoint
   \relax}

\def\MorseLong
  {\dontleavehmode
   \vrule
      width 3\dimexpr\MorseWidth
      height         \MorseHeight
      depth          \zeropoint
   \relax}

\def\MorseBetweenCharacters
  {\kern\MorseWidth}

\def\MorseBetweenWords
  {\hskip3\dimexpr\MorseWidth\relax}
```

```
\def\MorseSpace
  {\hskip7\dimexpr\MorseWidth\relax}

\def\MorseUnknown#1
  {[\detokenize{#1}]}
```

The dimensions are stored in macros as well. Of course we could provide a proper setup command, but it hardly makes sense.

```
\def\MorseWidth {0.4em}
\def\MorseHeight{0.2em}
```

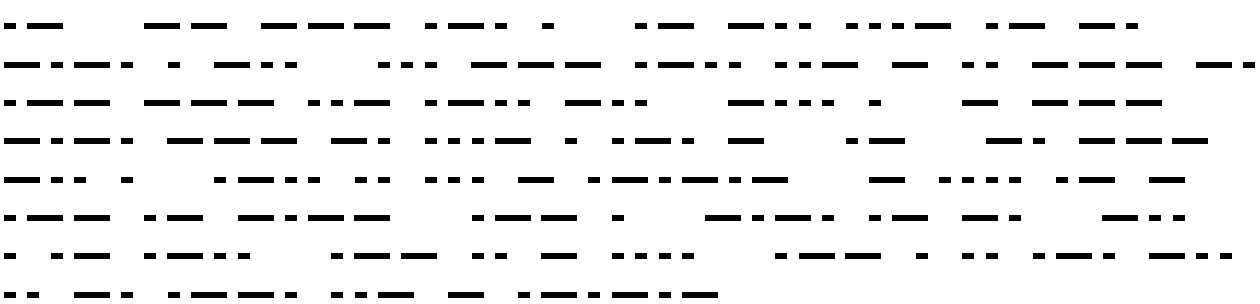
Finally we have arrived at the macros that interface to the LUA functions.

```
\def\MorseString#1{\ctxlua{moduledata.morse.tomorse(\!!bs#1\!!es)}}
\def\MorseFile  #1{\ctxlua{moduledata.morse.filetomorse("#1")}}
\def\MorseTable   {\ctxlua{moduledata.morse.showtable()}}
```

A string is converted to morse with the first command.

```
\Morse{A more advanced solution would be to convert a node list. That
way we can deal with weird input.}
```

This shows up as:



Reduction and uppercasing is demonstrated in the next example:

```
\MorseString{ÀÁÂÃÄÅàáâãäå}
```

This gives:

## 5.6 Using helpers

The next example shows a bit of LPEG. On top of the standard functionality a few additional functions are provided. Let's start with a pure TEX example:

```
\defineframed
  [coloredframed]
  [foregroundcolor=red,
   foregroundstyle=\underbar,
```

```
   offset=.1ex,
   location=low]

\processisolatedwords {\input ward \relax} \coloredframed
```

The Earth, as a habitat for animal life, is in old age and has a fatal illness. Several, in fact. It would be happening whether humans had ever evolved or not. But our presence is like the effect of an old-age patient who smokes many packs of cigarettes per day—and we humans are the cigarettes.

Because this processor macro operates at the TEX end it has some limitations. The content is collected in a very narrow box and from that a regular paragraph is constructed. It is for this reason that no color is applied: the snippets that end up in the box are already typeset.

An alternative is to delegate the task to LUA:

```
\startluacode
local function process(data)

  local words = lpeg.split(lpeg.patterns.spacer,data or "")

  for i=1,#words do
    if i == 1 then
        context.dontleavehmode()
    else
        context.space()
    end
    context.coloredframed(words[i])
  end

end

process(io.loaddata(resolvers.findfile("ward.tex")))
\stopluacode
```

The Earth, as a habitat for animal life, is in old age and has a fatal illness. Several, in fact. It would be happening whether humans had ever evolved or not. But our presence is like the effect of an old-age patient who smokes many packs of cigarettes per day — and we humans are the cigarettes.

The function splits the loaded data into a table with individual words. We use a splitter that splits on spacing tokens. The special case for `i = 1` makes sure that we end up in horizontal mode (read: properly start a paragraph). This time we do get color because the typesetting is done directly. Here is an alternative implementation:

```
local done = false

local function reset()
    done = false
    return true
end
```

```
local function apply(s)
  if done then
    context.space()
  else
    done = true
    context.dontleavehmode()
  end
  context.coloredframed(s)
end

local splitter = lpeg.P(reset)
               * lpeg.splitter(lpeg.patterns.spacer,apply)

local function process(data)
  lpeg.match(splitter,data)
end
```

This version is more efficient as it does not create an intermediate table. The next one is comaprable:

```
local function apply(s)
  context.coloredframed("%s ",s)
end

local splitter lpeg.splitter(lpeg.patterns.spacer,apply)

local function process(data)
  context.dontleavevmode()
  lpeg.match(splitter,data)
  context.removeunwantedspaces()
end
```

## 5.7 Formatters

Sometimes can save a bit of work by using formatters. By default, the `context` command, when called directly, applies a given formatter. But when called as table this feature is lost because then we want to process non-strings as well. The next example shows a way out:

The last one is the most interesting one here: in the subnamespace `formatted` (watch the `d`) a format specification with extra arguments is expected.

# 6 Graphics

## 6.1 The regular interface

If you are familiar with CONTEXT, which by now probably is the case, you will have noticed that it integrates the METAPOST graphic subsystem. Drawing a graphic is not that complex:

```
context.startMPcode()
context [[
  draw
    fullcircle scaled 1cm
    withpen pencircle scaled 1mm
    withcolor .5white
    dashed dashpattern (on 2mm off 2mm) ;
 ]]
context.stopMPcode()
```

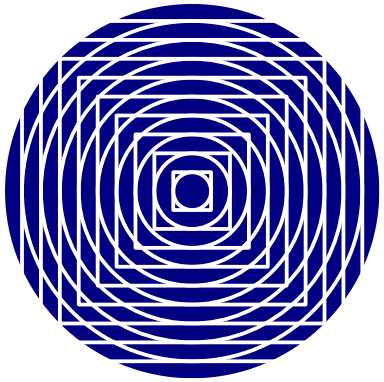
We get a gray dashed circle rendered with an one millimeter thick line:

So, we just use the regular commands and pass the drawing code as strings. Although METAPOST is a rather normal language and therefore offers loops and conditions and the lot, you might want to use LUA for anything else than the drawing commands. Of course this is much less efficient, but it could be that you don't care about speed. The next example demonstrates the interface for building graphics piecewise.

```
context.resetMPdrawing()

context.startMPdrawing()
context([[fill fullcircle scaled 5cm withcolor (0,0,.5) ;]])
context.stopMPdrawing()

context.MPdrawing("pickup pencircle scaled .5mm ;")
context.MPdrawing("drawoptions(withcolor white) ;")

for i=0,50,5 do
  context.startMPdrawing()
  context("draw fullcircle scaled %smm ;",i)
  context.stopMPdrawing()
end

for i=0,50,5 do
  context.MPdrawing("draw fullsquare scaled " .. i .. "mm ;")
end

context.MPdrawingdonetrue()
```

```
context.getMPdrawing()
```

This gives:



I the first loop we can use the format options associated with the simple `context` call. This will not work in the second case. Even worse, passing more than one argument will definitely give a faulty graphic definition. This is why we have a special interface for MetaFun. The code above can also be written as:

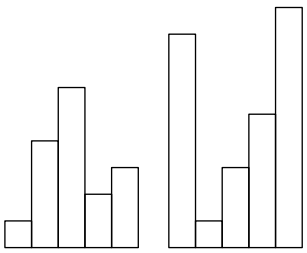
```
local metafun = context.metafun

metafun.start()

metafun("fill fullcircle scaled 5cm withcolor %s ;",
    metafun.color("darkblue"))

metafun("pickup pencircle scaled .5mm ;")
metafun("drawoptions(withcolor white) ;")

for i=0,50,5 do
  metafun("draw fullcircle scaled %smm ;",i)
end

for i=0,50,5 do
  metafun("draw fullsquare scaled %smm ;",i)
end

metafun.stop()
```

Watch the call to `color`, this will pass definitions at the TEX end to MetaPost. Of course you really need to ask yourself "Do I want to use MetaPost this way?". Using Lua loops instead of MetaPost ones makes much more sense in the following case:

```
local metafun = context.metafun

function metafun.barchart(t)
  metafun.start()
  local t = t.data
  for i=1,#t do
```

```
    metafun("draw unitsquare xyscaled(%s,%s) shifted (%s,0);",
      10, t[i]*10, i*10)
  end
  metafun.stop()
end

local one = { 1, 4, 6, 2, 3, }
local two = { 8, 1, 3, 5, 9, }

context.startcombination()
  context.combination(metafun.delayed.barchart { data = one }, "one")
  context.combination(metafun.delayed.barchart { data = two }, "two")
context.stopcombination()
```

We get two barcharts alongside:



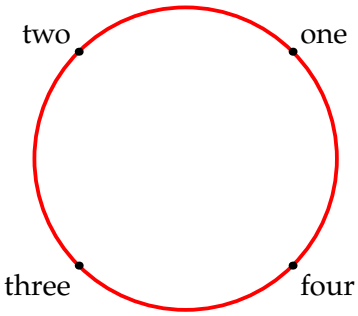

```
local template = [[
  path p, q ; color c[] ;
  c1 := \MPcolor{darkblue} ;
  c2 := \MPcolor{darkred} ;
  p := fullcircle scaled 50 ;
  l := length p ;
  n := %s ;
  q := subpath (0,%s/n*l) of p ;
  draw q withcolor c2 withpen pencircle scaled 1 ;
  fill fullcircle scaled 5 shifted point length q of q withcolor c1 ;
  setbounds currentpicture to unitsquare shifted (-0.5,-0.5) scaled 60 ;
  draw boundingbox currentpicture withcolor c1 ;
  currentpicture := currentpicture xsized(1cm) ;
]]

local function steps(n)
  for i=0,n do
    context.metafun.start()
      context.metafun(template,n,i)
    context.metafun.stop()
    if i < n then
        context.quad()
    end
  end
end
```

```
context.hbox(function() steps(10) end)
```

Using a template is quite convenient but at some point you can loose track of the replacement values. Also, adding an extra value can force you to adapt the following ones which enlarges the change for making an error. An alternative is to use the template mechanism. Although this mechanism was originally made for other purposes, you can use it for whatever you like.

```
local template = [[
    path p ; p := fullcircle scaled 4cm ;
    draw p withpen pencircle scaled .5mm withcolor red ;
    freedotlabel ("%lefttop%",    point 1 of p,origin) ;
    freedotlabel ("%righttop%",   point 3 of p,origin) ;
    freedotlabel ("%leftbottom%", point 5 of p,origin) ;
    freedotlabel ("%rightbottom%",point 7 of p,origin) ;
]]

local variables = {
  lefttop     = "one",
  righttop    = "two",
  leftbottom  = "three",
  rightbottom = "four" ,
}

context.metafun.start()
  context.metafun(utilities.templates.replace(template,variables))
context.metafun.stop()
```

Here we use named placeholders and pass a table with associated values to the replacement function. Apart from convenience it's also more readable. And the overhead is rather minimal.




To some extent we fool ourselves with this kind of LUAfication of METAPOST code. Of course we can make a nice METAPOST library and put the code in a macro instead. In that sense, doing this in CONTEXT directly often gives better and more efficient code.

Of course you can use all relevant commands in the LUA interface, like:

```
context.startMPpage()
  context("draw origin")
```

```
  for i=0,100,10 do
    context("..{down}(%d,0)",i)
  end
  context(" withcolor \\MPcolor{darkred} ;")
context.stopMPpage()
```

to get a graphic that has its own page. Don't use the `metafun` namespace here, as it will not work here. This drawing looks like:



## 6.2 The LUA interface

Messing around with graphics is normally not needed and if you do it, you'd better know what you're doing. For TEX a graphic is just a black box: a rectangle with dimensions. You specify a graphic, in a format that the backend can deal with, either or not apply some scaling and from then on a reference to that graphic, normally wrapped in a normal TEX box, enters the typesetting machinery. Because the backend, the part that is responsible for translating typeset content onto a viewable or printable format like PDF, is built into LUATEX, at some point the real image has to be injected and the backend can only handle a few image formats: PNG, JPG, JBIG and PDF.

In CONTEXT some more image formats are supported but in practice this boils down to converting the image to a format that the backend can handle. Such a conversion depends on an external programs and in order not to redo the conversion each run CONTEXT keeps track of the need to redo it.

Some converters are built in, for example one that deals with GIF images. This is normally not a preferred format, but it happens that we have to deal with it in cases where organizations use that format (if only because they use the web). Here is how this works at the LUA end:

```
figures.converters.gif = {
  pdf = function(oldname,newname)
    os.execute(string.format("gm convert %s %s",oldname,newname))
  end
}
```

We use `gm` (Graphic Magic) for the conversion and pass the old and new names. Given this definition at the TEX end we can say:

```
\externalfigure[whatever.gif][width=4cm]
```

Here is a another one:

```
figures.converters.bmp = {
  pdf = function(oldname,newname)
      os.execute(string.format("gm convert %s %s",oldname,newname))
  end
}
```

In both examples we convert to PDF because including this filetype is quite fast. But you can also go to other formats:

```
figures.converters.png = {
```

```
  png = function(oldname,newname,resolution)
    local command = string.format('gm convert -depth 1 "%s" "%s"',oldname,newname)
    logs.report(string.format("running command %s",command))
    os.execute(command)
  end
}
```

Instead of directly defining such a table, you can better do this:

```
figures.converters.png = figures.converters.png or { }

figures.converters.png.png = function(oldname,newname,resolution)
  local command = string.format('gm convert -depth 1 "%s" "%s"',oldname,newname)
  logs.report(string.format("running command %s",command))
  os.execute(command)
end
```

Here we check if a table exists and if not we extend the table. Such converters work out of the box if you specify the suffix, but you can also opt for a simple:

```
\externalfigure[whatever][width=4cm]
```

In this case CONTEXT will check for all known supported formats, which is not that efficient when no graphic can be found. In order to let for instance files with suffix `bmp` can be included you have to register it as follows. The second argument is the target.

```
figures.registersuffix("bmp","bmp")
```

At some point more of the graphic inclusion helpers will be opened up for general use but for now this is what you have available.

# 7 Macros

## 7.1 Introduction

You can skip this chapter if you're not interested in defining macros or are quite content with defining them in TEX. It's just an example of possible future interface definitions and it's not the fastest mechanism around.

## 7.2 Parameters

Right from the start CONTEXT came with several user interfaces. As a consequence you need to take this into account when you write code that is supposed to work with interfaces other than the English one. The TEX command:

```
\setupsomething[key=value]
```

and the LUA call:

```
context.setupsomething { key = value }
```

are equivalent. However, all keys at the TEX end eventually become English, but the values are unchanged. This means that when you code in LUA you should use English keys and when dealing with assigned values later on, you need to translate them of compare with translations (which is easier). This is why in the CONTEXT code you will see:

```
if somevalue == interfaces.variables.yes then
  ...
end
```

instead of:

```
if somevalue == "yes" then
  ...
end
```

## 7.3 User interfacing

Unless this is somehow inhibited, users can write their own macros and this is done in the TEX language. Passing data to macros is possible and looks like this:

```
\def\test#1#2{.. #1 .. #2 .. }      \test{a}{b}
\def\test[#1]#2{.. #1 .. #2 .. }    \test[a]{b}
```

Here `#1` and `#2` represent an argument and there can be at most 9 of them. The `[]` are delimiters and you can delimit in many ways so the following is also right:

```
\def\test(#1><#2){.. #1 .. #2 .. }  \test(a><b)
```

Macro packages might provide helper macros that for instance take care of optional arguments, so that we can use calls like:

```
\test[1,2,3][a=1,b=2,c=3]{whatever}
```

and alike. If you are familiar with the CONTEXT syntax you know that we use this syntax all over the place.

If you want to write a macro that calls out to LUA and handles things at that end, you might want to avoid defining the macro itself and this is possible.

An example of a definition and usage at the LUA end is:

```
\startluacode
function document.test(opt_1, opt_2, arg_1)
    context.startnarrower()
    context("options 1: %s",interfaces.tolist(opt_1))
    context.par()
    context("options 2: %s",interfaces.tolist(opt_2))
    context.par()
    context("argument 1: %s",arg_1)
    context.stopnarrower()
end

interfaces.definecommand {
    name = "test",
    arguments = {
        { "option", "list" },
        { "option", "hash" },
        { "content", "string" },
    },
    macro = document.test,
}
\stopluacode

test: \test[1][a=3]{whatever}
```

The call gives:

test:
options 1: 1
options 2: a=3
argument 1: whatever

If you want to to define an environment (i.e. a `start-stop` pair, it looks as follows:

```
\startluacode
local function startmore(opt_1)
    context.startnarrower()
    context("start more, options: %s",interfaces.tolist(opt_1))
    context.startnarrower()
end

local function stopmore()
```

```
    context.stopnarrower()
    context("stop more")
    context.stopnarrower()
end

interfaces.definecommand ( "more", {
    environment = true,
    arguments = {
        { "option", "list" },
    },
    starter = startmore,
    stopper = stopmore,
} )
\stopluacode

more: \startmore[1] one \startmore[2] two \stopmore one \stopmore
```

This gives:

more:
start more, options: 1
one
start more, options: 2
two
stop more
one
stop more

The arguments are know in both `startmore` and `stopmore` and nesting is handled automatically.

## 7.4 Looking inside

If needed you can access the body of a macro. Take for instance:

```
\def\TestA{A}
\def\TestB{\def\TestC{c}}
\def\TestC{C}
```

The following example demonstrates how we can look inside these macros. You need to be aware of the fact that the whole blob of LUA codes is finished before we return to TEX, so when we pipe the meaning of `TestB` back to TEX it only gets expanded afterwards. We can use a function to get back to LUA. It's only then that the meaning of `testC` is changed by the (piped) expansion of `TestB`.

```
\startluacode
context(tokens.getters.macro("TestA"))
context(tokens.getters.macro("TestB"))
context(tokens.getters.macro("TestC"))
tokens.setters.macro("TestA","a")
context(tokens.getters.macro("TestA"))
context(function()
```

```
    context(tokens.getters.macro("TestA"))
    context(tokens.getters.macro("TestB"))
    context(tokens.getters.macro("TestC"))
end)
\stopluacode
```

ACaac

Here is another example:

```
\startluacode
if tokens.getters.macro("fontstyle") == "rm" then
    context("serif")
else
    context("unknown")
end
\stopluacode
```

Of course this assumes that you have some knowledge of the CONTEXT internals.

serif

## 7.5 Deep down

Some commands in the previous sections use a more fundamental interfacing feature: implementers. These are used a lot deep down in CONTEXT and use high performance scanners to pick up data from the TEX end. As with other definitions you have to be careful using them and choose names wisely; you never know when we add something that will clash with your command name and what we define in the core wins.

```
\startluacode
interfaces.implement {
    name      = "foo",
    arguments = { "string", "integer" },
    actions   = function(s,i)
        -- something
    end,
}
\stopluacode
```

This defines a command `\clf_foo` so it is hidden from the user due to the use of an underscore. You can instead say:

```
\startluacode
interfaces.implement {
    name      = "foo",
    public    = true,
    arguments = { "string", "integer" },
    actions   = function(s,i)
        -- something
```

```
    end,
}
\stopluacode
```

If you want a protected macro you say:

```
\startluacode
interfaces.implement {
    name      = "foo",
    public    = true,
    protected = true,
    arguments = { "string", "integer" },
    actions   = function(s,i)
        -- something
    end,
}
\stopluacode
```

Other prefix keys are `untraced` and `noaligned`. The prefix permanent is always used so that implementers are protected against overloading what than mechanism is active. A special directive is `usage`:

```
\startluacode
interfaces.implement {
    name      = "iffoo",
    public    = true,
    usage     = "condition",
    arguments = { "string", "integer" },
    actions   = function(s,i)
        -- return something
    end,
}
\stopluacode
```

which gives you a native `\if`-test. A `value` usage is also special and you can for instance use it after `\the` in which case the (here) `what` signals that some value has to be returned. Otherwise you can for instance scan for a value.

```
\startluacode
interfaces.implement {
    name      = "foo",
    public    = true,
    usage     = "value",
    actions   = function(what)
        if what == "value" then
            -- return something
        else
            -- something
        end
    end,
```

```
}
\stopluacode
```

When returning a value you need to specify the type (a number), and a value. Valid types are:

```
 0  none          3  dimension     6  float          9  direct
 1  integer       4  skip          7  string        10  conditional
 2  cardinal      5  boolean       8  node
```

When scanning for something you have a lot of choice, like

```
\startluacode
local i = tokens.scanners.integer()
\stopluacode
```

Valid scanners are:

`argument`, `argumentasis`, `array`, `boolean`, `box`, `bracketed`, `bracketedasis`, `cardinal`, `char`, `cmdchr`, `cmdchrexpanded`, `code`, `conditional`, `count`, `csname`, `csnameunchecked`, `delimited`, `detokened`, `dimen`, `dimension`, `dimensionargument`, `float`, `glue`, `gluespec`, `gluevalues`, `hash`, `hbox`, `integer`, `integerargument`, `ischar`, `key`, `keyword`, `keywordcs`, `letters`, `list`, `lua`, `luacardinal`, `luainteger`, `luanumber`, `lxmlid`, `next`, `nextchar`, `nextexpanded`, `number`, `optional`, `peek`, `peekchar`, `peekexpanded`, `posit`, `real`, `scanclose`, `scanopen`, `skip`, `skipexpanded`, `string`, `table`, `token`, `tokencode`, `tokenlist`, `tokens`, `tokenstring`, `toks`, `value`, `vbox`, `verbatim`, `vtop`, `whd`, `word`

but normally users will stick to `string`, `integer` and `dimension` if they use this interface at all, so for now we will not go into details about the less obvious ones. Here it is enough to know that the implement feature is the one we use to extend the repertoire of low level, sometimes primitive-like commands.

Keep in mind that when you see these mechanisms in the code base that they are CONTEXT specific. Right from the start we had such interfaces and they are the reason why we always had a well integrated system. But they are not generic!

Here are some examples of `usage`:

```
\startluacode
interfaces.implement {
    name      = "IfFoo",
    public    = true,
    usage     = "condition",
    arguments = { "integer", "integer" },
    actions   = function(what)
        return tokens.values.boolean, one == two
    end,
}
\stopluacode
```

With:

```
[\IfFoo {123 + 10} {789 / 3}  yes\else no\fi]
[\IfFoo {123 + 10} {10 + 123} yes\else no\fi]
```

we get: [ yes] [ yes]. A value variant is:

```
\startluacode
local value = 123

interfaces.implement {
    name    = "MyFoo",
    public  = true,
    usage   = "value",
    actions = function(what)
        if what == "value" then
            return tokens.values.integer, value
        else
            value = tokens.scanners.integer(true)
        end
    end,
}
\stopluacode
```

With:

```
[\the\MyFoo]\MyFoo = 456
[\the\MyFoo]\MyFoo = { 456 + 123 }%
[\the\MyFoo]
```

we get: [123][456][579]. The `true` in the scanner call makes that we can use a `=` in the assignment. When scanning for an integer or dimension you actually use the expression scanner.

# 8 Verbatim

## 8.1 Introduction

If you are familiar with traditional TEX, you know that some characters have special meanings. For instance a `$` starts and ends inline math mode:

```
$e=mc^2$
```

If we want to typeset math from the LUA end, we can say:

```
context.mathematics("e=mc^2")
```

This is in fact:

```
\mathematics{e=mc^2}
```

However, if we want to typeset a dollar and use the `ctxcatcodes` regime, we need to explicitly access that character using `\char` or use a command that expands into the character with catcode other.

One step further is that we typeset all characters as they are and this is called verbatim. In that mode all characters are tokens without any special meaning.

## 8.2 Special treatment

The formula in the introduction can be typeset verbatim as follows:

```
context.verbatim("$e=mc^2$")
```

This gives:

$e=mc^2$

You can also do things like this:

```
context.verbatim.bold("$e=mc^2$")
```

Which gives:

**$e=mc^2$**

So, within the `verbatim` namespace, each command gets its arguments verbatim.

```
context.verbatim.inframed({ offset = "0pt" }, "$e=mc^2$")
```

Here we get: $e=mc^2$. So, settings and alike are processed as if the user had used a regular `context.inframed` but the content comes out verbose.

If you wonder why verbatim is needed as we also have the `type` function (macro) the answer is that it is faster, easier to key in, and sometimes the only way to get the desired result.

## 8.3 Multiple lines

Currently we have to deal with linebreaks in a special way. This is due to the way TEX deals with linebreaks. In fact, when we print something to TEX, the text after a `\n` is simply ignored.

For this reason we have a few helpers. If you want to put something in a buffer, you cannot use the regular buffer functions unless you make sure that they are not overwritten while you're still at the LUA end.

```
context.tobuffer("temp",str)
context.getbuffer("temp")
```

Another helper is the following. It splits the string into lines and feeds them piecewise using the `context` function and in the process adds a space at the end of the line (as this is what TEX normally does.

```
context.tolines(str)
```

Catcodes can get in the way when you pipe something to TEX that itself changes the catcodes. This happens for instance when you write buffers that themselves have buffers or have code that changes the line endings as with `startlines`. In that case you need to feed back the content as if it were a file. This is done with:

```
context.viafile(str)
```

The string can contain newlines. The string is written to a virtual file that is input. Currently names looks like `virtual://virtualfile.1` but future versions might have a different name part, so best use the variable instead. After all, you don't know the current number in advance anyway.

## 8.4 Pretty printing

In CONTEXT MKII there have always been pretty printing options. We needed it for manuals and it was also handy to print sources in the same colors as the editor uses. Most of those pretty printers work in a line-by-line basis, but some are more complex, especially when comments or strings can span multiple lines.

When the first versions of LUATEX showed up, rewriting the MKII code to use LUA was a nice exercise and the code was not that bad, but when LPEG showed up, I put it on the agenda to reimplement them again.

We only ship a few pretty printers. Users normally have their own preferences and it's not easy to make general purpose pretty printers. This is why the new framework is a bit more flexible and permits users to kick in their own code.

Pretty printing involves more than coloring some characters or words:

- spaces should honoured and can be visualized
- newlines and empty lins need to be honoured as well
- optionally lines have to be numbered but
- wrapped around lines should not be numbered

It's not much fun to deal with these matters each time that you write a pretty printer. This is why we can start with an existing one like the default pretty printer. We show several variants of doing the same. We start with a simple clone of the default parser.[5]

```
local P, V = lpeg.P, lpeg.V

local grammar = visualizers.newgrammar("default", {
  pattern    = V("default:pattern"),
  visualizer = V("pattern")^1
} )

local parser = P(grammar)

visualizers.register("test-0", { parser = parser })
```

We distinguish between grammars (tables with rules), parsers (a grammar turned into an LPEG expression), and handlers (collections of functions that can be applied. All three are registered under a name and the verbatim commands can refer to that name.

```
\starttyping[option=test-0,color=]
Test 123,
test 456 and
test 789!
\stoptyping
```

Nothing special happens here. We just get straightforward verbatim.

```
Test 123,
test 456 and
test 789!
```

Next we are going to color digits. We collect as many as possible in a row, so that we minimize the calls to the colorizer.

```
local patterns, P, V = lpeg.patterns, lpeg.P, lpeg.V

local function colorize(s)
  context.color{"darkred"}
  visualizers.writeargument(s)
end

local grammar = visualizers.newgrammar("default", {
  digit      = patterns.digit^1 / colorize,
  pattern    = V("digit") + V("default:pattern"),
  visualizer = V("pattern")^1
} )
```

[5] In the meantime the lexer of the SciTE editor that I used also provides a mechanism for using LPEG based lexers. Although in the pretty printing code we need a more liberal one I might backport the lexers I wrote for editing TEX, METAPOST, LUA, CLD, XML and PDF as a variant for the ones we use in MKIV now. That way we get similar colorschemes which might be handy sometimes.

```
local parser = P(grammar)

visualizers.register("test-1", { parser = parser })
```

Watch how we define a new rule for the digits and overload the pattern rule. We can refer to the default rule by using a prefix. This is needed when we define a rule with the same name.

```
\starttyping[option=test-1,color=]
Test 123,
test 456 and
test 789!
\stoptyping
```

This time the digits get colored.

```
Test 123,
test 456 and
test 789!
```

In a similar way we can colorize letters. As with the previous example, we use CONTEXT commands at the LUA end.

```
\starttyping[option=test-2,color=]
Test 123,
test 456 and
test 789!
\stoptyping
```

Again we get some coloring.

```
Test 123,
test 456 and
test 789!
```

It will be clear that the amount of rules and functions is larger when we use a more complex parser. It is for this reason that we can group functions in handlers. We can also make a pretty printer configurable by defining handlers at the TEX end.

```
\definestartstop
  [MyDigit]
  [style=bold,color=darkred]

\definestartstop
  [MyLowercase]
  [style=bold,color=darkgreen]

\definestartstop
  [MyUppercase]
  [style=bold,color=darkblue]
```

The LUA code now looks different. Watch out: we need an indirect call to for instance `MyDigit` because a second argument can be passed: the settings for this environment and you don't want that get passed to `MyDigit` and friends.

```
\starttyping[option=test-3,color=]
Test 123,
test 456 and
test 789!
\stoptyping
```

We get digits, upper- and lowercase characters colored:

```
Test 123,
test 456 and
test 789!
```

You can also use parsers that don't use LPEG:

```
local function parser(s)
  visualizers.write("["..s.."]")
end

visualizers.register("test-4", { parser = parser })

\starttyping[option=test-4,space=on,color=darkred]
Test 123,
test 456 and
test 789!
\stoptyping
```

The function `visualizer.write` takes care of spaces and newlines.

```
[Test␣123,
test␣456␣and
test␣789!]
```

We have a few more helpers:

| | |
|---|---|
| `visualizers.write` | interprets the argument and applies methods |
| `visualizers.writenewline` | goes to the next line (similar to `\par` |
| `visualizers.writeemptyline` | inserts an empty line (similer to `\blank` |
| `visualizers.writespace` | inserts a (visible) space |
| `visualizers.writedefault` | writes the argument verbatim without interpretation |

These mechanism have quite some overhead in terms of function calls. In the worst case each token needs a (nested) call. However, doing all this at the TEX end also comes at a price. So, in practice this approach is more flexible but without too large a penalty.

In all these examples we typeset the text verbose: what is keyed in normally comes out (either or not with colors), so spaces stay spaces and linebreaks are kept.

```
local function parser(s)
  local s = string.gsub(s,"show","demonstrate")
  local s = string.gsub(s,"'re"," are")
  context(s)
end
```

```
visualizers.register("test-5", { parser = parser })
```

We can apply this visualizer as follows:

```
\starttyping[option=test-5,color=darkred,style=]
This is just some text to show what we can do with this mechanism. In
spite of what you might think we're not bound to verbose text.
\stoptyping
```

This time the text gets properly aligned:

This is just some text to demonstrate what we can do with this mechanism. In spite of what you might think we are not bound to verbose text.

It often makes sense to use a buffer:

```
\startbuffer[demo]
This is just some text to show what we can do with this mechanism. In
spite of what you might think we're not bound to verbose text.
\stopbuffer
```

Instead of processing the buffer in verbatim mode you can then process it directly:

```
\setuptyping[file][option=test-5,color=darkred,style=]
\ctxluabuffer[demo]
```

Which gives:

In this case, the space is a normal space and not the fixed verbatim space, which looks better.

# 9 Logging

Logging and localized messages have always been rather standardized in CONTEXT, so upgrading the related mechanism had been quite doable. In MKIV for a while we had two systems in parallel: the old one, mostly targeted at messages at the TEX end, and a new one used at the LUA end. But when more and more hybrid code showed up, integrating both systems made sense.

Most logging concerns tracing and can be turned on and off on demand. This kind of control is now possible for all messages. Given that the right interfaces are used, you can turn off all messages:

```
context --silent
```

This was already possible in MKII, but there TEX's own messages still were visible. More important is that we have control:

```
context --silent=structure*,resolve*,font*
```

This will disable all reporting for these three categories. It is also possible to only disable messages to the console:

```
context --noconsole
```

In CONTEXT you can use directives:

```
\enabledirectives[logs.blocked=structure*,resolve*,font*]
\enabledirectives[logs.target=file]
```

As all logging is under LUA control and because this (and other) kind of control has to kick in early in the initialization the code might look somewhat tricky. Users won't notice this because they only deal with the formal interface. Here we will only discuss the LUA interfaces.

Messages related to tracing are done as follows:

```
local report_whatever = logs.reporter("modules","whatever")

report_whatever("not found: %s","this or that")
```

The first line defined a logger in the category `modules`. You can give a second argument as well, the subcategory. Both will be shown as part of the message, of which an example is given in the second line.

These messages are shown directly, that is, when the function is called. However, when you generate TEX code, as we discuss in this document, you need to make sure that the message is synchronized with that code. This can be done with a messenger instead of a reporter.

```
local report_numbers = logs.reporter("numbers","check")
local status_numbers = logs.messenger("numbers","check")

status_numbers("number 1: %s, number 2: %s",123,456)
report_numbers("number 1: %s, number 2: %s",456,123)
```

Both reporters and messages are localized when the pattern given as first argument can be found in the `patterns` subtable of the interface messages. Categories and subcategories are also translated, but these are looked up in the `translations` subtable. So in the case of

```
report_whatever("found: %s",filename)
report_whatever("not found: %s",filename)
```

you should not be surprised if it gets translated. Of course the category and subcategory provide some contextual information.

# 10 Lua Functions

## 10.1 Introduction

When you run CONTEXT you have some libraries preloaded. If you look into the LUA files you will find more than is discussed here, but keep in mind that what is not documented, might be gone or done different one day. Some extensions live in the same namespace as those provided by stock LUA and LUATEX, others have their own. There are many more functions and the more obscure (or never being used) ones will go away.

The LUA code in CONTEXT is organized in quite some modules. Those with names like `l-*.lua` are rather generic and are automatically available when you use `mtxrun` to run a LUA file. These are discusses in this chapter. A few more modules have generic properties, like some in the categories `util-*.lua`, `trac-*.lua`, `luat-*.lua`, `data-*.lua` and `lxml-*.lua`. They contain more specialized functions and are discussed elsewhere, if at all.

Before we move on the the real code, let's introduce a handy helper:

```
inspect(somevar)
```

Whenever you feel the need to see what value a variable has you can insert this function to get some insight. It knows how to deal with several data types.

Some of the functions here are rather specific to CONTEXT, even if we load them in our plain TEX reference format. Don't depend on them outside the perspective of CONTEXT.

There are more modules than discussed here. You can find more about them in the LUAMETATEX manual. Some have additional functions that are implemented on top of low level ones. There are also modules that are (kind of) specific to for instance LUAMETAFUN; these are wrapped into METAPOST interfaces.

## 10.2 Tables

### [lua] concat

These functions come with LUA itself and are discussed in detail in the LUA reference manual so we stick to some examples. The `concat` function stitches table entries in an indexed table into one string, with an optional separator in between. If can also handle a slice of the table

```
local str = table.concat(t)
local str = table.concat(t,separator)
local str = table.concat(t,separator,first)
local str = table.concat(t,separator,first,last)
```

Only strings and numbers can be concatenated.

```
table.concat({"a","b","c","d","e"})
```

```
abcde
```

```
table.concat({"a","b","c","d","e"},"+")
```

```
a+b+c+d+e
```

```
table.concat({"a","b","c","d","e"},"+",2,3)
```

```
b+c
```

## [lua] insert remove

You can use `insert` and `remove` for adding or replacing entries in an indexed table.

```
table.insert(t,position,value)
value = table.remove(t,position)
```

The position is optional and defaults to the last entry in the table. For instance a stack is built this way:

```
table.insert(stack,"top")
local top = table.remove(stack)
```

Beware, the `insert` function returns nothing. You can provide an additional position:

```
table.insert(list,"injected in slot 2",2)
local thiswastwo = table.remove(list,2)
```

## [lua] pack unpack

You can access entries in an indexed table as follows:

```
local a, b, c = t[1], t[2], t[3]
```

but this does the same:

```
local a, b, c = table.unpack(t)
```

```
table.unpack({"a","b","c"})
```

```
a b c
```

```
table.unpack({"a",nil,"b"})
```

```
a
```

```
table.unpack({"a","b","c"},2)
```

```
b c
```

```
table.unpack({"a","b","c"},1,2)
```

```
a b
```

This is less efficient but there are situations where `unpack` comes in handy. The reverse is done with `pack`:

```
local t = table.pack(1,2,3)
```

You'll notice that the resulting table has a field `n` that holds the number of entries (the last index).

```
table.pack("a","b","c")
```

```
t={
 "a",
 "b",
 "c",
 ["n"]=3,
}
```

```
table.pack("a",nil,"c")
```

```
t={
 "a",
 [3]="c",
 ["n"]=3,
}
```

## [lua] sort

Sorting is done with `sort`, a function that does not return a value but operates on the given table.

```
table.sort(t)
table.sort(t,comparefunction)
```

The compare function has to return a consistent equivalent of `true` or `false`. For sorting more complex data structures there is a specialized sort module available.

```
t={"a","b","c"} table.sort(t)
```

```
t={ "a", "b", "c" }
```

```
t={"a","b","c"} table.sort(t,function(x,y) return x > y end)
```

```
t={ "c", "b", "a" }
```

```
t={"a","b","c"} table.sort(t,function(x,y) return x < y end)
```

```
t={ "a", "b", "c" }
```

## sorted

The built-in `sort` function does not return a value but sometimes it can be if the (sorted) table is returned. This is why we have:

```
local a = table.sorted(b)
```

## keys sortedkeys sortedhashkeys sortedhash

The keys function returns an indexed list of keys. The order is undefined as it depends on how the table was constructed. A sorted list is provided by sortedkeys. This function is rather liberal with respect to the keys. If the keys are strings you can use the faster alternative sortedhashkeys.

```
local s = table.keys (t)
local s = table.sortedkeys (t)
local s = table.sortedhashkeys (t)
```

Because a sorted list is often processed there is also an iterator:

```
for key, value in table.sortedhash(t) do
    print(key,value)
end
```

There is also a synonym sortedpairs which sometimes looks more natural when used alongside the pairs and ipairs iterators.

```
table.keys({ [1] = 2, c = 3, [true] = 1 })

t={ 1, true, "c" }

table.sortedkeys({ [1] = 2, c = 3, [true] = 1 })

t={ 1, "c", true }

table.sortedhashkeys({ a = 2, c = 3, b = 1 })

t={ "a", "b", "c" }
```

## sortedhashonly sortedindexonly

These are more selective variants of the previous discussed collectors.

```
table.sortedhashonly({ [1] = 1, a = 2, c = 3})

t={ 1 }

table.sortedindexonly({ [1] = 1, a = 2, c = 1})

t={ "a", "c" }
```

## serialize print tohandle tofile

The serialize function converts a table into a verbose representation. The print function does the same but prints the result to the console which is handy for tracing. The tofile function writes the table to a file, using reasonable chunks so that less memory is used. The fourth variant tohandle takes a handle so that you can do whatever you like with the result.

```
table.serialize (root, name, reduce, noquotes, hexify)
table.print (root, name, reduce, noquotes, hexify)
```

```
table.tofile (filename, root, name, reduce, noquotes, hexify)
table.tohandle (handle, root, name, reduce, noquotes, hexify)
```

The serialization can be controlled in several ways. Often only the first two options makes sense:

```
table.serialize({ a = 2 })
```

```
t={
 ["a"]=2,
}
```

```
table.serialize({ a = 2 }, "name")
```

```
name={
 ["a"]=2,
}
```

```
table.serialize({ a = 2 }, true)
```

```
return {
 ["a"]=2,
}
```

```
table.serialize({ a = 2 }, false)
```

```
{
 ["a"]=2,
}
```

```
table.serialize({ a = 2 }, "return")
```

```
return {
 ["a"]=2,
}
```

```
table.serialize({ a = 2 }, 12)
```

```
["12"]={
 ["a"]=2,
}
```

```
table.serialize({ a = 2, [3] = "b", [true] = "6" }, nil, true)
```

```
t={
 [3]="b",
 ["a"]=2,
 [true]="6",
}
```

```
table.serialize({ a = 2, [3] = "b", [true] = "6" }, nil, true, true)
```

```
t={
```

```
 [3]="b",
 ["a"]=2,
 [true]="6",
}
```

```
table.serialize({ a = 2, [3] = "b", [true] = "6" }, nil, true, true, true)
```

```
t={
 [3]="b",
 ["a"]=2,
 [true]="6",
}
```

In CONTEXT there is also a `tocontext` function that typesets the table verbose. This is handy for manuals and tracing.

## identical are_equal

These two function compare two tables that have a similar structure. The `identical` variant operates on a hash while `are_equal` assumes an indexed table.

```
local b = table.identical (one, two)
local b = table.are_equal (one, two)
```

```
table.identical({ a = { x = 2 } }, { a = { x = 3 } })
```

```
false
```

```
table.identical({ a = { x = 2 } }, { a = { x = 2 } })
```

```
true
```

```
table.are_equal({ a = { x = 2 } }, { a = { x = 3 } })
```

```
true
```

```
table.are_equal({ a = { x = 2 } }, { a = { x = 2 } })
```

```
true
```

```
table.identical({ "one", "two" }, { "one", "two" })
```

```
true
```

```
table.identical({ "one", "two" }, { "two", "one" })
```

```
false
```

```
table.are_equal({ "one", "two" }, { "one", "two" })
```

```
true
```

```
table.are_equal({ "one", "two" }, { "two", "one" })
```

```
false
```

### tohash fromhash swapped swaphash reversed reverse mirrored

We use `tohash` quite a lot in CONTEXT. It converts a list into a hash so that we can easily check if (a string) is in a given set. The `fromhash` function does the opposite: it creates a list of keys from a hashed table where each value that is not `false` or `nil` is present.

```
local hashed  = table.tohash  (indexed)
local indexed = table.fromhash(hashed)
```

The function `swapped` turns keys into values vise versa while the `reversed` and `reverse` reverses the values in an indexed table. The last one reverses the table itself (in-place).

```
local swapped  = table.swapped  (indexedtable)
local reversed = table.reversed (indexedtable)
local reverse  = table.reverse  (indexedtable)
local mirrored = table.mirrored (hashedtable)

table.tohash({ "a", "b", "c" })

t={
 ["a"]=true,
 ["b"]=true,
 ["c"]=true,
}

table.fromhash({ a = true, b = false, c = true })

t={ "c", "a" }

table.swapped({ "a", "b", "c" })

t={
 ["a"]=1,
 ["b"]=2,
 ["c"]=3,
}

table.reversed({ "a", "b", "c" })

t={ "c", "b", "a" }

table.reverse({ 1, 2, 3, 4 })

t={ 4, 3, 2, 1 }

table.mirrored({ a = "x", b = "y", c = "z" })

t={
 ["a"]="x",
 ["b"]="y",
 ["c"]="z",
 ["x"]="a",
```

```
 ["y"]="b",
 ["z"]="c",
}
```

### append prepend

These two functions operate on a pair of indexed tables. The first table gets appended or prepended by the second. The first table is returned as well.

```
table.append (one, two)
table.prepend(one, two)
```

The functions are similar to loops using `insert`.

```
table.append({ "a", "b", "c" }, { "d", "e" })
```

```
t={ "a", "b", "c", "d", "e" }
```

```
table.prepend({ "a", "b", "c" }, { "d", "e" })
```

```
t={ "d", "e", "a", "b", "c" }
```

### merge merged imerge imerged

You can merge multiple hashes with `merge` and indexed tables with `imerge`. The first table is the target and is returned.

```
table.merge   (one, two, ...)
table.imerge  (one, two, ...)
```

The variants ending with a `d` merge the given list of tables and return the result leaving the first argument untouched.

```
local merged = table.merged  (one, two, ...)
local merged = table.imerged (one, two, ...)
```

```
table.merge({ a = 1, b = 2, c = 3 }, { d = 1 }, { a = 0 })
```

```
t={
 ["a"]=0,
 ["b"]=2,
 ["c"]=3,
 ["d"]=1,
}
```

```
table.imerge({ "a", "b", "c" }, { "d", "e" }, { "f", "g" })
```

```
t={ "a", "b", "c", "d", "e", "f", "g" }
```

## copy fastcopy

When copying a table we need to make a real and deep copy. The `copy` function is an adapted version from the LUA wiki. The `fastcopy` is faster because it does not check for circular references and does not share tables when possible. In practice using the fast variant is okay.

```
local copy = table.copy    (t)
local copy = table.fastcopy(t)
```

## flattened

A nested table can be unnested using `flattened`. Normally you will only use this function if the content is somewhat predictable. Often using one of the merge functions does a similar job.

```
local flattened = table.flatten(t)

table.flattened({ a = 1, b = 2, { c = 3 }, d = 4})

t={ 4, 2, 1, 3 }

table.flattened({ 1, 2, { 3, { 4 } }, 5})

t={ 1, 2, 3, 4, 5 }

table.flattened({ 1, 2, { 3, { 4 } }, 5}, 1)

t={
 1,
 2,
 3,
 { 4 },
 5,
}

table.flattened({ a = 1, b = 2, { c = 3 }, d = 4})

t={ 4, 2, 1, 3 }

table.flattened({ 1, 2, { 3, { c = 4 } }, 5})

t={ 1, 2, 3, 4, 5 }

table.flattened({ 1, 2, { 3, { c = 4 } }, 5}, 1)

t={
 1,
 2,
 3,
 {
  ["c"]=4,
 },
 5,
```

```
}
```

### loweredkeys

The name says it all: this function returns a new table with the keys being lower case. This is handy in cases where the keys have a change to be inconsistent, as can be the case when users input keys and values in less controlled ways.

```
local normalized = table.loweredkeys { a = "a", A = "b", b = "c" }

table.loweredkeys({ a = 1, b = 2, C = 3})
```

```
t={
 ["a"]=1,
 ["b"]=2,
 ["c"]=3,
}
```

### contains

This function works with indexed tables. Watch out, when you look for a match, the number `1` is not the same as string `"1"`. The function returns the index or `false`.

```
if table.contains(t, 5 ) then ... else ... end
if table.contains(t,"5") then ... else ... end

table.contains({ "a", 2, true, "1"}, 1)
```

```
false
```

```
table.contains({ "a", 2, true, "1"}, "1")
```

```
4
```

### unique

When a table (can) contain duplicate entries you can get rid of them by using the `unique` helper:

```
local t = table.unique { 1, 2, 3, 4, 3, 2, 5, 6 }

table.unique( { "a", "b", "c", "a", "d" } )
```

```
t={ "a", "b", "c", "d" }
```

### count

The name speaks for itself: this function counts the number of entries in the given table. For an indexed table `#t` is faster.

```
local n = table.count(t)
```

```
table.count({ 1, 2, [4] = 4, a = "a" })
```

```
4
```

### sequenced

Normally, when you trace a table, printing the serialized version is quite convenient. However, when it concerns a simple table, a more compact variant is:

```
print(table.sequenced(t, separator))
```

```
table.sequenced({ 1, 2, 3, 4})
```

```
1 | 2 | 3 | 4
```

```
table.sequenced({ 1, 2, [4] = 4, a = "a" }, ", ")
```

```
1, 2
```

### [lua] create

This is a relative new native LUA table function that does what the name says:

```
local indexed = table.create(10)
local hashed  = table.create(0,10)
local hybrid  = table.create(10,10)
```

In LUAMETATEX we already introduced this creator that expects two arguments:

```
local mytable = lua.newtable(10,0)
```

which has a variant:

```
local myarray = lua.newindex(10,2)
```

where the values are preset (in this case to the integer value 2)..

### [lua] move

This is now a built-in helper:

```
local result = table.move(input,first,last,target)
local result = table.move(input,first,last,target,output)
```

The last argument is optional.

```
table.move({ 1, 2, 3, 4, 5, 6 }, 1, 2, 3)
```

```
t={ 1, 2, 1, 2, 5, 6 }
```

```
table.move({ 1, 2, 3, 4, 5, 6 }, 1, 2, 5)
```

```
t={ 1, 2, 3, 4, 1, 2 }
```

```
table.move({ 1, 2, 3, 4, 5, 6 }, 4, 5, 2)
```

```
t={ 1, 4, 5, 4, 5, 6 }
```

```
table.move({ 1, 2, 3, 4 }, 1, 2, 3, { "a", "b", "c", "d" })
```

```
t={ "a", "b", 1, 2 }
```

### sub

This helper does the same as `move` with an empty target table starting at one. Apart from the table the arguments are optional.

```
table.sub({ 1, 2, 3, 4, 5, 6 }, 1, 2 )
```

```
t={ 1, 2 }
```

```
table.sub({ 1, 2, 3, 4, 5, 6 }, 3, 5 )
```

```
t={ 3, 4, 5 }
```

```
table.sub({ 1, 2, 3, 4, 5, 6 }, 3 )
```

```
t={ 3, 4, 5, 6 }
```

```
table.sub({ 1, 2, 3, 4, 5, 6 } )
```

```
t={ 1, 2, 3, 4, 5, 6 }
```

### save load

You can use these helpers to save data and later (maybe in a second run) load it.

```
local t = { a = 10, c = 30, d = { b = 20 } }
table.save("mytable.lua",t)

local t = table.load("mytable.lua")
```

In CONTEXT we have various ways to store data for two-pass usage which has the advantage that one can trigger on a change. This is just a convenient variant that you need to manage yourself.

### combine

This is a resolver: we feed it a source and a target

```
\startluacode
local list  = {
    ["one"] = { preset = "alpha" },
    ["two"] = { preset = "beta"  },
}
```

```
local presets  = {
    ["alpha"] = { name = "this", slot = 123 },
    ["beta"]  = { name = "that", slot = 345 },
}

list.one = table.combine(presets[list.one.preset],list.one)

table.tocontext(list.one,false)
\stopluacode
```

Here the preset, given as string, gets replaced by a table. We don't make a deep copy, so you should not rewrite the result.

```
{
 ["name"]="this",
 ["preset"]="alpha",
 ["slot"]=123,
}
```

As a side effect of the implementation you can actually do this:

```
\startluacode
local target = { }

table.combine(target, { 4,  9,    6,   2 })
table.combine(target, { 3, nil,   1, nil })
table.combine(target, { 8, nil, nil,   7 })

table.tocontext(target,false)
\stopluacode
```

Which gives you the set of values most recently seen:

```
{ 8, 9, 1, 7 }
```

## fastserialize deserialize

These are for instance used when we store a LUA table in an SQL database field.

```
local t = { "a", "a", "", "b", "c", e = "f" }
local s = table.fastserialize(t)

local s = [[return{[1]="a",[2]="a",[3]="",[4]="b",[5]="c",["e"]="f",}]]
local t = table.deserialize(s)
```

Of course 'fast' is subjective because we have to construct the string piecewise.

```
table.fastserialize({ "a", "a", "", "b", "c", e = "f" })
```

```
return{[1]="a",[2]="a",[3]="",[4]="b",[5]="c",["e"]="f",}
```

```
table.deserialize([[return{[1]="a",[2]="a",[3]="",[4]="b",[5]="c",["e"]="f",}]])
```

```
t={
 "a",
 "a",
 "",
 "b",
 "c",
 ["e"]="f",
}
```

### has_one_entry is_empty

Often one will inline these but we keep them around:

```
table.has_one_entry({ 1, 2, 3 })
```

```
false
```

```
table.has_one_entry({ 1 })
```

```
true
```

```
table.has_one_entry({ k = 1, v = 2 })
```

```
false
```

```
table.has_one_entry({ k = 1 })
```

```
true
```

```
table.is_empty({ 1, 2, 3 })
```

```
false
```

```
table.is_empty({ })
```

```
true
```

### is_simple_table

This is a helper that happens to be exposed so let's see what it does.

```
table.is_simple_table({ 1, 2, 3 })
```

```
t={ 1, 2, 3 }
```

```
table.is_simple_table({ 1, "a", true })
```

```
t={ 1, "\"a\"", "true" }
```

```
table.is_simple_table({ 1, "a", { 1 } })
```

```
nil
```

The return value is either a table with the elements or `nil`. We use this helper in the serializer.

### hashed

This helper adds values as keys to a table, which can be handy if you want to do a reverse lookup. It mostly makes sense for a list of strings:

```
table.hashed({ "foo", "oof", "ofo" })
```

```
t={
 "foo",
 "oof",
 "ofo",
 ["foo"]=1,
 ["ofo"]=3,
 ["oof"]=2,
}
```

```
table.hashed({ 1, "a", true })
```

```
t={
 1,
 "a",
 true,
 ["a"]=2,
 [true]=3,
}
```

### makeweak

Say that we have this:

```
local t = { }
for i=1,9 do
    t[i] = function() end
end
```

The table has nine entries and it will keep them until the whole table is collected or a value is set to `nil`. When we say:

```
local t = table.makeweak()
for i=1,9 do
    t[i] = function() print (i) end
end
```

you can never be sure what the entries are but you can at least test if a value is set. If so, you can reuse that value, if not, you can redefine the function and reassign it. It is a way to prevent the same (defined by index) function again and again. The value has to be an collectable object.

### toxml

This feature can best be shown with examples. The second argument is a specification. When a `result` table is specified the code gets appended to that table.

```
table.toxml({ "foo", "oof", { "ofo" } }, { spaces = 4 })
```

```
<?xml version='1.0' standalone='yes' ?>
<data>
    <entry n='1'>foo</entry>
    <entry n='2'>oof</entry>
    <entry n='3'>
        <entry n='1'>ofo</entry>
    </entry>
</data>
```

```
table.toxml({ "foo", "oof", { "ofo" } }, { indent = 3 })
```

```
<?xml version='1.0' standalone='yes' ?>
   <data>
    <entry n='1'>foo</entry>
    <entry n='2'>oof</entry>
    <entry n='3'>
     <entry n='1'>ofo</entry>
    </entry>
   </data>
```

```
table.toxml({ "foo", "oof", { "ofo" } }, { name = "root" })
```

```
<?xml version='1.0' standalone='yes' ?>
<root>
 <entry n='1'>foo</entry>
 <entry n='2'>oof</entry>
 <entry n='3'>
  <entry n='1'>ofo</entry>
 </entry>
</root>
```

```
table.toxml({ "foo", "oof", { "ofo" } }, { name = false })
```

```
<entry n='1'>foo</entry>
<entry n='2'>oof</entry>
<entry n='3'>
 <entry n='1'>ofo</entry>
</entry>
```

### tocsv

This is a quick-and-dirty converter from a table of tables to a (normally) comma separated list:

```
table.tocsv({ { "foo", "oof" } }, { preamble = true, fields = { "a", "b" } })
```

```
"a","b"
```

```
"foo","foo"
```

The table is either index or a hash, in the later case a field list is mandate:

```
table.tocsv({ { b = "X", a = "Y" } }, { preamble = true, fields = { "a", "b" } })
```

```
"a","b"
"Y","X"
```

You can set the separator:

```
table.tocsv({ { "foo", "oof", "ofo" } }, { separator = ";" })
```

```
"foo";"foo";"foo"
```

### setmetatablecall setmetatableindex setmetatablenewindex

These functions define (or patch) a table with metatable driven behavior; the table is optional.

```
local t = table.setmetatableindex({ }, function(t,k) end)
local t = table.setmetatableindex({ }, sometable)
local t = table.setmetatableindex({ }, "empty")
local t = table.setmetatableindex({ }, "self")
local t = table.setmetatableindex({ }, "table")
local t = table.setmetatableindex({ }, "number")
```

When a table slot is accessed and no value has been set, `empty` makes sure that an empty string is returned, `self` returns the key, `table` an empty table and `number` a zero. These values are also assigned. When a function is passed that one has to do the assignment and return a value.

```
local t = table.setmetatablenewindex({ }, function(t,k,v) end)
local t = table.setmetatablenewindex({ }, "ignore")
```

Here setting a table goes via the function call. Setting a value then has to be done with `rawset`.

```
local t = table.setmetatablecall({ }, function(t,...) end)
```

Calling a table triggers the set function where optional arguments are passed.

### setmetatablenewindices

You can set all three plugins in one go, so you pass three functions.

```
local t = table.setmetatableindices  ({ }, f_index, f_newindex, f_call)
```

### derive

A shortcut to define a table that has a table as metatable index is the following:

```
local t = table.derive { a = 1, b = 2, c = 3 }
```

### sparse compact

You can remove entries that store an empty strings or `false` value from a table:

```
table.sparse({ a = 1, b = 2, c = false, d = "" })
```

```
t={
 ["a"]=1,
 ["b"]=2,
}
```

The `compact` variant calls sparse with two extra `true` values, one that signals that we want to nest, and one that ensures that we keep tables.

### filtered

This helper started as an experiment, was never used but for now we keep it around.

```
\startluacode
for k, v in table.filtered({ foo = 1, bar = 2, oof = 3 },"f") do
    context("[%s = %s]",k,v)
end

for k, v in table.filtered({ foo = 1, bar = 2, oof = 3 },"f",true) do
    context("(%s = %s)",k,v)
end
\stopluacode
```

The results are: [oof = 3][foo = 1](foo = 1)(oof = 3). As fourth optional argument a sort compare function can be passed.

### strip

This function removes leading and trailing spaces from string entries and omits empty strings.

```
table.strip({ "a", "b", "", " d e ", "", "f" })
```

```
t={ "a", "b", "d e", "f" }
```

## 10.3 Math

The LUA interpreter comes with a default `math` library. In LUAMETATEX we also have `xmath`, `xcomplex`, `xdecimal`, `posit` and `vector`. The math library has a few more helpers but if we use a LUA version or a built-in one depends on the engine.

In addition to the built-in math function we provide: `round`, `odd`, `even`, `div`, `mod`, `sind`, `cosd` and `tand`, `cosh`, `sinh`, `tanh`, `pow`, `atan2` (an alias), `ceiling` (an alias), `ldexp`, `log10`, `type`, `tointeger` and `ult`.

At the TEX end we have a helper `luaexpr` that you can use to do calculations:

```
\luaexpr{1 + 2.3 * 4.5 + math.pi} = \cldcontext{1 + 2.3 * 4.5 + math.pi}
```

Both calls return the same result, but the first one is normally faster than the `context` command which has quite some overhead.

14.491592653589793 = 14.491592653589793

The `\luaexpr` command can also better deal with for instance conditions, where it returns `true` or `false`, while `\cldcontext` would interpret the boolean value as a special signal.

## 10.4 Booleans

### tonumber

This function returns the number one or zero. You will seldom need this function.

```
local state = boolean.tonumber(str)
```

```
boolean.tonumber(true)
```

```
1
```

### toboolean

When dealing with configuration files or tables a bit flexibility in setting a state makes sense, if only because in some cases it's better to say `yes` than `true`.

```
local b = toboolean(str)
local b = toboolean(str,tolerant)
```

When the second argument is true, the strings `true`, `yes`, `on`, `1`, `t` and the number `1` all turn into `true`. Otherwise only `true` is honoured. This function is also defined in the global namespace.

```
string.toboolean("true")
```

```
true
```

```
string.toboolean("yes")
```

```
false
```

```
string.toboolean("yes",true)
```

```
true
```

### is_boolean

This function is somewhat similar to the previous one. It interprets the strings `true`, `yes`, `on` and `t` as `true` and `false`, `no`, `off` and `f` as `false`. Otherwise `nil` is returned, unless a default value is given, in which case that is returned.

```
if is_boolean(str)                then ... end
if is_boolean(str,default)        then ... end
if is_boolean(str,default,struct) then ... end

string.is_boolean("true")

true

string.is_boolean("off")

false

string.is_boolean("crap",true)

true
```

Like the other boolean converters this one related pretty much to the way we interface in CONTEXT. For instance the type strict options makes that we don't test zero and one as string.

### booleanstring

This function takes one argument and returns a boolean. Relevant values are:

| | | | | | |
|---|---|---|---|---|---|
| "0" | false | "true" | true | on | false |
| "1" | true | 0 | false | t | false |
| "" | false | 1 | true | f | false |
| "false" | false | yes | false | | |

## 10.5 Strings

LUA strings are simply sequences of bytes. Of course in some places special treatment takes place. For instance \n expands to one or more characters representing a newline, depending on the operating system, but normally, as long as you manipulate strings in the perspective of LUATEX, you don't need to worry about such issues too much. As LUATEX is a UTF-8 engine, strings normally are in that encoding but again, it does not matter much as LUA is quite agnostic about the content of strings: it does not care about three characters reflecting one UNICODE character or not. This means that when you use for instance the functions discussed here, or use libraries like lpeg behave as you expect.

Although we now live in LUAMETATEX times a bit of history doesn't hurt. When we started with LUATEX, we included the slunicode library but that one never got updated. That was no big deal because we (needed and therefore) wrote UNICODE helpers for CONTEXT that were more flexible. We also didn't need some of its features. So, instead we used string manipulators, some provided by the engine, as well as LPEG magic. Over time the repertoire evolved. At some point LUA itself got some support for UTF8 built in. As a consequence string functions are implemented on various modules, often depending on where other helpers reside. In LUAMETATEX a few critical converters were added to the engine and as we never used slunicode it was just removed.

Not all of the helpers discussed here are of general use because after all they were often made for CONTEXT. This means that we can adapt details as we go forward. It's not like CONTEXT users have

to deal with conversions when they use the system, most is hidden. The repertoire presented here is a mix of functions that come with LUA, functions that are implemented in LUA, and functions that were added to LUATEX and later LUAMETATEX. When the later are available, we use them instead of the LUA variants.

## [lua] byte char

As long as we're dealing with ASCII characters we can use these two functions to go from numbers to characters and vise versa.

```
string.byte("luatex")
```

```
108
```

```
string.byte("luatex",1,3)
```

```
108 117 97
```

```
string.byte("luatex",-3,-1)
```

```
116 101 120
```

```
string.char(65)
```

```
A
```

```
string.char(65,66,67)
```

```
ABC
```

## [lua] sub

You cannot directly access a character in a string but you can take any slice you want using `sub`. You need to provide a start position and negative values will count backwards from the end.

```
local slice = string.sub(str,first,last)
```

```
string.sub("abcdef",2)
```

```
bcdef
```

```
string.sub("abcdef",2,3)
```

```
bc
```

```
string.sub("abcdef",-3,-2)
```

```
de
```

## [lua] gsub

There are two ways of analyzing the content of a string. The more modern and flexible approach is to use `lpeg`. The other one uses some functions in the `string` namespace that accept so called patterns

for matching. While lpeg is more powerfull than regular expressions, the pattern matching is less powerfull but sometimes faster and also easier to specify. In many cases it can do the job quite well.

```
local new, count = string.gsub(old,pattern,replacement)
```

The replacement can be a function. Often you don't want the number of matches, and the way to avoid this is either to store the result in a variable:

```
local new = string.gsub(old,"lua","LUA")
print(new)
```

or to use parentheses to signal the interpreter that only one value is return.

```
print((string.gsub(old,"lua","LUA"))
```

Patterns can be more complex so you'd better read the LUA manual if you want to know more about them.

```
string.gsub("abcdef","b","B")
```

```
aBcdef
```

```
string.gsub("abcdef","[bc]",string.upper)
```

```
aBCdef
```

An optional fourth argument specifies how often the replacement has to happen

```
string.gsub("textextextex","tex","abc")
```

```
abcabcabcabc
```

```
string.gsub("textextextex","tex","abc",1)
```

```
abctextextex
```

```
string.gsub("textextextex","tex","abc",2)
```

```
abcabctextex
```

You can also pass a table as replacement, as in:

```
string.gsub("we use tex",".",{t="T",x="X"})
```

```
we use TeX
```

So to summarize, we can use strings, tables and functions as replacement values, but in the later two cases when there is no return value or match, no replacement takes place.

See the match section below for more about pattern matching in gsub.

## [lua] find

The find function returns the first and last position of the match:

```
local first, last = find(str,pattern)
```

If you're only interested if there is a match at all, it's enough to know that there is a first position. No match returns `nil`. So,

```
if find("luatex","tex") then ... end
```

works out okay. You can pass an extra argument to `find` that indicates the start position. So you can use this function to loop over all matches: just start again at the end of the last match.

A fourth optional argument is a boolean that signals not to interpret the pattern but use it as-is.

```
string.find("abc.def","c%.d",1,false)
```

```
3
```

```
string.find("abc.def","c%.d",1,true)
```

```
nil
```

```
string.find("abc%.def","c%.d",1,false)
```

```
nil
```

```
string.find("abc%.def","c%.d",1,true)
```

```
3
```

See the `match` section below for more about pattern matching in `find`.

## [lua] match gmatch

With `match` you can split of bits and pieces of a string. The parenthesis indicate the captures.

```
local a, b, c, ... = string.match(str,pattern)
```

The `gmatch` function is used to loop over a string, for instance the following code prints the elements in a comma separated list, ignoring spaces after commas.

```
for s in string.gmatch(str,"([^,%s])+") do
  print(s)
end
```

A more detailed description of patterns can be found in the LUA reference manual, so we only mention the special directives. Characters are grouped in classes:

| | |
|---|---|
| %a | letters |
| %l | lowercase letters |
| %u | uppercase letters |
| %d | digits |
| %w | letters and digits |
| %c | control characters |
| %p | punctuation |

| | |
|---|---|
| %x | hexadecimal characters |
| %s | space related characters |

You can create sets too:

| | |
|---|---|
| [%l%d] | lowercase letters and digits |
| [^%d%p] | all characters except digits and punctuation |
| [p-z] | all characters in the range p upto z |
| [pqr] | all characters p, q and r |

There are some characters with special meanings:

| | |
|---|---|
| ^ | the beginning of a string |
| $ | end of a string |
| . | any character |
| * | zero or more of the preceding specifier, greedy |
| - | zero or more of the preceding specifier, least possible |
| + | one or more of the preceding specifier |
| ? | zero or one of the preceding specifier |
| ( ) | encapsulate capture |
| %b | capture all between the following two characters |

You can use whatever you like to be matched:

| | |
|---|---|
| pqr | the sequence pqr |
| my name is (%w) | the word following my name is |

If you want to specify such a token as it is, then you can precede it with a percent sign, so to get a percent, you need two in a row.

```
string.match("before:after","^(.-):")
```

```
before
```

```
string.match("before:after","^([^:])")
```

```
b
```

```
string.match("before:after","bef(.*)ter")
```

```
ore:af
```

```
string.match("abcdef","[b-e]+")
```

```
bcde
```

```
string.match("abcdef","[b-e]*")
```

```
string.match("abcdef","b-e+")
```

```
e
```

```
string.match("abcdef","b-e*")
```

Such patterns should not be confused with regular expressions, although to some extent they can do the same. If you really want to do complex matches, you should look into LPEG.

One trick is worth mentioning: matching a previous match. Here is an exmaple

```
string.match("AAoneBBtwoBBthreeAA","^(AA)(.*)(BB)(.*)%3(.*)%1$")
```

```
AA one BB two three
```

```
string.gsub("AAtestAA","^(AA)(.*)%1$","%2")
```

```
test 1
```

It is easy to forget that such features exist, although the situations where you have fenced that are the same are limited (single and double quotes as well as vertical bars come to mind).

## [lua] lower upper

These two function spreak for themselves.

```
string.lower("LOW")
```

```
low
```

```
string.upper("upper")
```

```
UPPER
```

## [lua] format

The `format` function takes a template as first argument and one or more additional arguments depending on the format. The template is similar to the one used in C but it has some extensions.

```
local s = format(format, str, ...)
```

The following table gives an overview of the possible format directives. The `s` is the most probably candidate and can handle numbers well as strings. Watch how the minus sign influences the alignment.[6]

| | | | |
|---|---|---|---|
| **integer** | %i | 12345 | 12345 |
| **integer** | %d | 12345 | 12345 |
| **unsigned** | %u | -12345 | 12345 |
| **character** | %c | 123 | Y |
| **hexadecimal** | %x | 123 | 7b |
| | %X | 123 | 7B |

[6] There can be differences between platforms although so far we haven't run into problems. Also, LUA 5.2 does a bit more checking on correct arguments and LUA 5.3 is more picky on integers.

| | | | |
|---|---|---|---|
| **octal** | %o | 12345 | 30071 |
| **string** | %s | abc | abcd |
| | %-8s | 123 | 123 |
| | %8s | 123 | 123 |
| **float** | %0.2f | 12.345 | 12.35 |
| **exponential** | %0.2e | 12.345 | 1.23e+01 |
| | %0.2E | 12.345 | 1.23E+01 |
| **autofloat** | %0.2g | 12.345 | 12 |
| | %0.2G | 12.345 | 12 |

```
string.format("U+% 05X",2010)
```

```
U+007DA
```

## striplines

The `striplines` function can strip leading and trailing empty lines, collapse or delete intermediate empty lines and strips leading and trailing spaces. We will demonstrate this with string `str`:

```
<sp><sp><lf><sp><sp><sp><sp>aap<lf><sp><sp>noot<sp>mies<lf><sp><sp><lf><sp>
<sp><sp><sp><lf><sp>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun<sp><sp><sp>
<sp>vuur<sp>gijs<lf><sp><sp><sp><sp><sp><sp><sp>lam<sp><sp><sp><sp>kees<sp>
bok<sp>weide<lf><sp><sp><sp><sp><lf>does<sp>hok<sp>duif<sp>schapen<sp><sp>
<lf><sp><sp>
```

```
1     aap
2   noot mies
3
4
5  wim    zus   jet
6 teun    vuur gijs
7        lam    kees bok weide
8
9 does hok duif schapen
```

The different options for stripping are demonstrated below, We use verbose descriptions instead of vague boolean flags.

```
utilities.strings.striplines(str,"collapse")
```

```
<lf><sp>aap<lf><sp>noot<sp>mies<lf><sp><lf><sp><lf><sp>wim<sp>zus<sp>jet<lf>
teun<sp>vuur<sp>gijs<lf><sp>lam<sp>kees<sp>bok<sp>weide<lf><sp><lf>does<sp>
hok<sp>duif<sp>schapen<sp><lf>
```

```
1     aap
2   noot mies
3
4
```

```
 wim    zus   jet
teun    vuur gijs
       lam    kees bok weide

does hok duif schapen
```

```
utilities.strings.striplines(str,"collapse all")
```

```
<sp>aap<sp>noot<sp>mies<sp>wim<sp>zus<sp>jet<sp>teun<sp>vuur<sp>gijs<sp>lam
<sp>kees<sp>bok<sp>weide<sp>does<sp>hok<sp>duif<sp>schapen
```

```
    aap
  noot mies


 wim    zus   jet
teun    vuur gijs
       lam    kees bok weide

does hok duif schapen
```

```
utilities.strings.striplines(str,"prune")
```

```
aap<lf>noot<sp>mies<lf><lf><lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun
<sp><sp><sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide<lf>
<lf>does<sp>hok<sp>duif<sp>schapen
```

```
    aap
  noot mies


 wim    zus   jet
teun    vuur gijs
       lam    kees bok weide

does hok duif schapen
```

```
utilities.strings.striplines(str,"prune and collapse")
```

```
aap<lf>noot<sp>mies<lf><lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun<sp>
<sp><sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide<lf><lf>
does<sp>hok<sp>duif<sp>schapen
```

```
    aap
  noot mies


 wim    zus   jet
teun    vuur gijs
       lam    kees bok weide
```

```
8
9  does hok duif schapen
```

```
utilities.strings.striplines(str,"prune and no empty")
```

```
aap<lf>noot<sp>mies<lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun<sp><sp>
<sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide<lf>does<sp>
hok<sp>duif<sp>schapen
```

```
1      aap
2    noot mies
3
4
5   wim     zus    jet
6  teun     vuur gijs
7         lam     kees bok weide
8
9  does hok duif schapen
```

```
utilities.strings.striplines(str,"prune and to space")
```

```
aap<sp>noot<sp>mies<sp>wim<sp>zus<sp>jet<sp>teun<sp>vuur<sp>gijs<sp>lam<sp>
kees<sp>bok<sp>weide<sp>does<sp>hok<sp>duif<sp>schapen
```

```
1      aap
2    noot mies
3
4
5   wim     zus    jet
6  teun     vuur gijs
7         lam     kees bok weide
8
9  does hok duif schapen
```

```
utilities.strings.striplines(str,"retain")
```

```
<lf>aap<lf>noot<sp>mies<lf><lf><lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>
teun<sp><sp><sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide
<lf><lf>does<sp>hok<sp>duif<sp>schapen<lf>
```

```
1      aap
2    noot mies
3
4
5   wim     zus    jet
6  teun     vuur gijs
7         lam     kees bok weide
8
9  does hok duif schapen
```

```
utilities.strings.striplines(str,"retain and collapse")
```

```
<lf>aap<lf>noot<sp>mies<lf><lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun
<sp><sp><sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide<lf>
<lf>does<sp>hok<sp>duif<sp>schapen<lf>
```

```
1      aap
2    noot mies
3
4
5   wim     zus    jet
6  teun     vuur gijs
7         lam     kees bok weide
8
9  does hok duif schapen
```

```
utilities.strings.striplines(str,"retain and no empty")
```

```
<lf>aap<lf>noot<sp>mies<lf>wim<sp><sp><sp><sp>zus<sp><sp><sp>jet<lf>teun<sp>
<sp><sp><sp>vuur<sp>gijs<lf>lam<sp><sp><sp><sp>kees<sp>bok<sp>weide<lf>does
<sp>hok<sp>duif<sp>schapen<lf>
```

```
1      aap
2    noot mies
3
4
5   wim     zus    jet
6  teun     vuur gijs
7         lam     kees bok weide
8
9  does hok duif schapen
```

You can of course mix usage with the normal `context` helper commands, for instance put them in buffers. Buffers normally will prune leading and trailing empty lines anyway.

```
context.tobuffer("dummy",utilities.strings.striplines(str))
context.typebuffer( { "dummy" }, { numbering = "line" })
```

## formatters

The `format` function discussed before is the built-in. As an alternative CONTEXT provides an additional formatter that has some extensions. Interesting is that that one is often more efficient, although there are cases where the speed is comparable. As we run out of keys, some extra ones are a bit counter intuitive, like `l` for booleans (logical).

| **utf character** | %c | 322 | ł |
|---|---|---|---|
| **string** | %s | foo | foo |
| **force tostring** | %S | nil | |
| **quoted string** | %q | foo | "foo" |
| **force quoted string** | %Q | nil | |
| | %N | 0123 | 123 |

| | | | |
|---|---|---|---|
| **automatic quoted** | %a | true | 'true' |
| | %A | true | "true" |
| **left aligned utf** | %30< | xx½xx | xx½xx |
| **right aligned utf** | %30> | xx½xx | xx½xx |
| **integer** | %i | 1234 | 1234 |
| **integer** | %d | 1234 | 1234 |
| **signed number** | %I | 1234 | +1234 |
| **rounded number** | %r | 1234.56 | 1235 |
| **stripped number** | %N | 000123 | 123 |
| **comma/period float** | %m | 12.34 | 12.34 |
| **period/comma float** | %M | 12.34 | 12,34 |
| **hexadecimal** | %x | 1234 | 4d2 |
| | %X | 1234 | 4D2 |
| **octal** | %o | 1234 | 2322 |
| **float** | %0.2f | 12.345 | 12.35 |
| **formatted float** | %2.3k | 12.3456 | 12.346 |
| **checked float** | %0.2F | 12.30 | 12.30 |
| **exponential** | %.2e | 12.345e120 | 1.23e121 |
| | %.2E | 12.345e120 | 1.23E121 |
| **sparse exp** | %0.2j | 12.345e120 | 1.23e121 |
| | %0.2J | 12.345e120 | 1.23E121 |
| **autofloat** | %g | 12.345 | 1.23E1 |
| | %G | 12.345 | 1.23E1 |
| **unicode value 0x** | %h | ł 1234 | |
| | %H | ł 1234 | |
| **unicode value U+** | %u | ł 1234 | u+00142 u+004d2 |
| | %U | ł 1234 | U+00142 U+004D2 |
| **points** | %p | 1234567 | 18.838pt |
| **points no unit** | %P | 1234567 | 18.838pt |
| **basepoints** | %b | 1234567 | 18.76762bp |
| **basepoints no unit** | %B | 1234567 | 18.76762bp |
| **table concat** | %t | {1,2,3} | 123 |
| | %*t | {1,2,3} | 1*2*3 |
| | %{ AND }t | {a=1,b=3} | |
| **table serialize** | %T | {1,2,3} | 1*2*3 |
| | %T | {a=1,b=3} | a=1  b=2 |
| | %+T | {a=1,b=3} | a=1[+b=2] |
| **boolean (logic)** | %l | "a" == "b" | |
| | %L | "a" == "b" | |
| **whitespace** | %w | 3 | \| \| |
| | %2w | 3 | \| \| |
| **(fixed)** | %4W | | \| |

| | | | |
|---|---|---|---|
| **skip** | %2z | 1,2,3,4 | 14 |
| **reference argument** | %2Z | 1,2,3,4 | 122 |

The generic formatters `a` and `A` convert the argument into a string and deals with strings, number, booleans, tables and whatever. We mostly use these in tracing. The lowercase variant uses single quotes, and the uppercase variant uses double quotes.

A special one is the alignment formatter, which is a variant on the `s` one that also takes an optional positive of negative number:

```
\startluacode
context.start()
context.tttf()
context.verbatim("[[% 30<]]","xxaxx") context.par()
context.verbatim("[[% 30<]]","xx½xx") context.par()
context.verbatim("[[% 30>]]","xxaxx") context.par()
context.verbatim("[[% 30>]]","xx½xx") context.par()
context.verbatim("[[%-30<]]","xxaxx") context.par()
context.verbatim("[[%-30<]]","xx½xx") context.par()
context.verbatim("[[%-30>]]","xxaxx") context.par()
context.verbatim("[[%-30>]]","xx½xx") context.par()
context.stop()
\stopluacode
```

```
[[xxaxx                         ]]

[[xx½xx                         ]]

[[                         xxaxx]]

[[                         xx½xx]]

[[                         xxaxx]]

[[                         xx½xx]]

[[xxaxx                         ]]

[[xx½xx                         ]]
```

There are two more formatters plugged in: `!xml!` and `!tex!`. These are best demonstrated with an example:

```
local xf = formatter["xml escaped: %!xml!"]
local xr = formatter["tex escaped: %!tex!"]

print(xf("x > 1 && x < 10"))
print(xt("this will cost me $123.00 at least"))
```

weird, this fails when cld-verbatim is there as part of the big thing: catcodetable 4 suddenly lacks the comment being a other

The `context` command uses the formatter so one can say:

```
\startluacode
context("first some xml: %!xml!, and now some %!tex!",
    "x > 1 && x < 10", "this will cost me $123.00 at least")
\stopluacode
```

This renders as follows:

first some xml: x > 1 && x < 10, and now some this will cost me $123.00 at least

You can extend the formatter but we advise you not to do that unless you're sure what you're doing. You never know what CONTEXT itself might add for its own benefit.

However, you can define your own formatter and add to that without interference. In fact, the main formatter is just defined that way. This is how it works:

```
local MyFormatter = utilities.strings.formatters.new()

utilities.strings.formatters.add (
    MyFormatter,
    "upper",
    "globaldata.string.upper(%s)"
)
```

Now you can use this one as:

```
context.bold(MyFormatter["It's %s or %!upper!."]("this","that"))
```

**It's this or THAT.**

Because we're running inside CONTEXT, a better definition would be this:

```
local MyFormatter = utilities.strings.formatters.new()

utilities.strings.formatters.add (
    MyFormatter,
    "uc",
    "myupper(%s)",
 -- "local myupper = globaldata.characters.upper"
    { myupper = globaldata.characters.upper }
)

utilities.strings.formatters.add (
    MyFormatter,
    "lc",
    "mylower(%s)",
 -- "local mylower = globaldata.characters.lower"
    { mylower = globaldata.characters.lower }
)

utilities.strings.formatters.add (
```

```
    MyFormatter,
    "sh",
    "myshaped(%s)",
 -- "local myshaped = globaldata.characters.shaped"
    { myshaped = globaldata.characters.shaped }
)

context(MyFormatter["Uppercased: %!uc!"]("ÀÁÂÃÄÅàáâãäå"))
context.par()
context(MyFormatter["Lowercased: %!lc!"]("ÀÁÂÃÄÅàáâãäå"))
context.par()
context(MyFormatter["Reduced: %!sh!"]("ÀÁÂÃÄÅàáâãäå"))
```

The last arguments creates shortcuts. As expected we get:

Uppercased: ÀÁÂÃÄÅÀÁÂÃÄÅ

Lowercased: àáâãäåàáâãäå

Reduced: AAAAAAaaaaaa

Of course you can also apply the casing functions directly so in practice you shouldn't use formatters without need. Among the advantages of using formatters are:

- They provide a level of abstraction.
- They can replace multiple calls to `\context`.
- Sometimes they make source code look better.
- Using them is often more efficient and faster.

The last argument might sound strange but considering the overhead involved in the `context` (related) functions, doing more in one step has benefits. Also, formatters are implemented quite efficiently, so their overhead can be neglected.

In the examples you see that a formatter extension is itself a template.

```
local FakeXML = utilities.strings.formatters.new()

utilities.strings.formatters.add(FakeXML,"b",[["<" ..%s..">" ]])
utilities.strings.formatters.add(FakeXML,"e",[["</"..%s..">" ]])
utilities.strings.formatters.add(FakeXML,"n",[["<" ..%s.."/>"]])

context(FakeXML["It looks like %!b!xml%!e! doesn't it?"]("it","it"))
```

This gives: It looks like <it>xml</it> doesn't it?. Of course we could go over the top here:

```
local FakeXML = utilities.strings.formatters.new()

local stack = { }

function document.f_b(s)
    table.insert(stack,s)
    return "<" .. s .. ">"
```

```
end

function document.f_e()
    return "</" .. table.remove(stack) .. ">"
end

utilities.strings.formatters.add(FakeXML,"b",[[globaldata.document.f_b(%s)]])
utilities.strings.formatters.add(FakeXML,"e",[[globaldata.document.f_e()]])

context(FakeXML["It looks like %1!b!xml%0!e! doesn't it?"]("it"))
```

This gives: It looks like <it>xml</it> doesn't it?. Such a template look horrible, although it's not too far from the regular format syntax: just compare `%1f` with `%1!e!`. The zero trick permits us to inject information that we've put on the stack. As this kind of duplicate usage might occur most, a better solution is available:

```
local FakeXML = utilities.strings.formatters.new()

utilities.strings.formatters.add(FakeXML,"b",[["<"  .. %s .. ">"]])
utilities.strings.formatters.add(FakeXML,"e",[["</" .. %s .. ">"]])

context(FakeXML["It looks like %!b!xml%-1!e! doesn't it?"]("it"))
```

We get: It looks like <it>xml</it> doesn't it?. Anyhow, in most cases you will never feel the need for such hackery and the regular formatter works fine. Adding this extension mechanism was rather trivial and it doesn't influence the performance.

In CONTEXT we have a few more extensions:

```
utilities.strings.formatters.add (
    strings.formatters, "unichr",
    [["U+" .. format("%%05X",%s) .. " (" .. utfchar(%s) .. ")"]]
)

utilities.strings.formatters.add (
    strings.formatters, "chruni",
    [[utfchar(%s) .. " (U+" .. format("%%05X",%s) .. ")"]]
)
```

This one is used in messages:

```
context("Missing character %!chruni! in font.",234) context.par()
context("Missing character %!unichr! in font.",234)
```

This shows up as:

context("Missing character context("Missing character

If you look closely to the definition, you will notice that we use `%s` twice. This is a feature of the definer function: if only one argument is picked up (which is default) then the replacement format can use that two times. Because we use a format in the constructor, we need to escape the percent sign there.

### strip

This function removes any leading and trailing whitespace characters.

```
local s = string.strip(str)
```

```
string.strip(" lua + tex = luatex ")
```

```
lua + tex = luatex
```

### split splitlines checkedsplit

The line splitter is a special case of the generic splitter. The `split` function can get a string as well an `lpeg` pattern. The `checkedsplit` function removes empty substrings.

```
local t = string.split        (str, pattern)
local t = string.split        (str, lpeg)
local t = string.checkedsplit (str, lpeg)
local t = string.splitlines   (str)
```

```
string.split("a, b,c, d", ",")
```

```
t={ "a", " b", "c", " d" }
```

```
string.split("p.q,r", lpeg.S(",."))
```

```
t={ "p", "q", "r" }
```

```
string.checkedsplit(";one;;two", ";")
```

```
t={ "one", "two" }
```

```
string.splitlines("lua\ntex  nic")
```

```
t={ "lua", "tex nic" }
```

### quoted unquoted

You will hardly need these functions. The `quoted` function can normally be avoided using the `format` pattern `%q`. The `unquoted` function removes single or double quotes but only when the string starts and ends with the same quote.

```
local q = string.quoted  (str)
local u = string.unquoted(str)
```

```
string.quoted([[test]])
```

```
"test"
```

```
string.quoted([[test"test]])
```

```
"test\"test"
```

```
string.unquoted([["test]])
```

```
"test
```

```
string.unquoted([["t\"est"]])
```

```
t\"est
```

```
string.unquoted([["t\"est"x]])
```

```
"t\"est"x
```

```
string.unquoted("\'test\'")
```

```
test
```

### count

The function `count` returns the number of times that a given pattern occurs. Beware: if you want to deal with UTF strings, you need the variant that sits in the `lpeg` namespace.

```
local n = count(str,pattern)
```

```
string.count("test me", "e")
```

```
2
```

### limit

This function can be handy when you need to print messages that can be rather long. By default, three periods are appended when the string is chopped.

```
print(limit(str,max,sentinel)
```

```
string.limit("too long", 6)
```

```
too...
```

```
string.limit("too long", 6, " (etc)")
```

```
(etc)
```

### is_empty

A string considered empty by this function when its length is zero or when it only contains spaces.

```
if is_empty(str) then ... end
```

```
string.is_empty("")
```

```
true
```

```
string.is_empty(" ")
```

```
true
```

```
string.is_empty(" ? ")
```

```
false
```

### escapedpattern topattern

These two functions are rather specialized. They come in handy when you need to escape a pattern, i.e. prefix characters with a special meaning by a %.

```
local e = escapedpattern(str, simple)
local p = topattern     (str, lowercase, strict)
```

The simple variant does less escaping (only `-.?*` and is for instance used in wildcard patterns when globbing directories. The `topattern` function always does the simple escape. A strict pattern gets anchored to the beginning and end. If you want to see what these functions do you can best look at their implementation.

### autosingle autodouble

These are helpers that we use in formatters: they wrap their content in single or double quotes. For practical reasons we keep them in the `string` but they could as well be private because we don't use them in other cases.

```
string.autodouble('123"456')
```

```
"123"456"
```

```
string.autosingle('123"456')
```

```
'123"456'
```

```
string.autodouble({ 1, 2, 3 })
```

```
"1,2,3"
```

```
string.autosingle({ 4, 5, 6 },"+")
```

```
'4+5+6'
```

### tohex toHEX todec tobytes hextocharacters

Here are a few of the public converters. In an ecosystem like CONTEXT hexadecimal strings are all over the place deep down, especially in the backend.

```
string.tohex("A")
```

```
41
```

```
string.tohex("A,B:C.D")
```

```
412c423a432e44
```

```
string.toHEX("A,B:C.D")
```

```
412C423A432E44
```

```
string.todec("A")
```

```
065
```

```
string.todec("A,B:C.D")
```

```
065044066058067046068
```

```
string.tobytes("6C7561")
```

```
lua
```

```
string.tobytes("4c7561")
```

```
Lua
```

The `tobytes` function is also available as `hextocharacters`, due to the history of the implementation.

### dectointeger hextointeger octtointeger chrtointeger

These convert from various (milti-)byte 'encodings' to an integer.

```
string.dectointeger("123")
```

```
123
```

```
string.hextointeger("1B3")
```

```
435
```

```
string.octtointeger("17")
```

```
15
```

```
string.chrtointeger("1")
```

```
49
```

### tracedchar

This is typically a function that we only provide because it is used in other functions.

```
string.tracedchar("\0")
```

```
[null]
```

```
string.tracedchar("\6")
```

```
[ack]
```

Only the range upto space (0x20) returns a string, otherwise you get nil. This permits usage in ASCII as well as UTF tracing because we might want to do something with the visible cases.

### nospaces

If spaces bother you the next function can be of help:

```
string.nospaces(" test test test ")
```

```
testtesttest
```

### [lua] rep

There are not that many helpers in the standard string library but rep (repeat would clash with the keyword) is one of them:

```
string.rep("twice",2)
```

```
twicetwice
```

### [lua] reverse

This is a more recent addition which only makes sense if we talk bytes; why reverse a string anyway?

```
string.reverse("hello dlrow")
```

```
world olleh
```

### [lua] pack unpack packsize

These are basically scanners driven by a syntax specification pattern. There is not much to tell about them here as we don't use them in CONTEXT. If you need them you ave a good reason to buy the official LUA manual.

### [lua] dump

This function dumps a function as bytecode that later can be loaded.

### toboolean

See the boolean section.

### bytes bytepairs

The byte iterators originally were meant for manipulating an eight bit (single byte encodings) or sixteen bit (UTF16) input stream in order to get UTF8.

```
\startluacode
for a in string.bytes("foof") do
    context("<%i> ",a)
end

for a, b in string.bytepairs("foof") do
    context("[%i,%i] ",a,b)
end

\stopluacode
```

Both give: <102> <111> <111> <102> [102,111] [111,102].

### characters characterpairs

These two iterators grab single byte character which limits their application:

```
\startluacode
for a in string.characters("foof") do
    context("<%s> ",a)
end

for a, b in string.characterpairs("foof") do -- originally meant for utf16
    context("[%s,%s] ",a,b)
end
\stopluacode
```

The two examples return: <f> <o> <o> <f> [f,o] [o,f].

### utfcharacters utfvalues

These two iterators grab UTF character:

```
\startluacode
for a in string.utfcharacters("foof") do -- originally meant for utf16
    context("<%s> ",a)
end

for a in string.utfvalues("foof") do
    context("[%i] ",a)
end
\stopluacode
```

They return: <f> <o> <o> <f> [102] [111] [111] [102].

### packrowscolumns

This is mostly a helper for low level mechanisms like bitmap graphics. In that case the entries in the passed table are rows.

```
\startluacode
context.type (
    string.packrowscolumns {
        { 76, 85, 65 },
        { 84, 69, 88 },
    }
)
\stopluacode
```

We get a string with bytes: LUATEX. Here the horizontal resolution is 3 and the vertical resolution is 2 and of course in practice a 8-bit bitmap is not that readable.

### len

Instead of the # operators you can use the len function, for instance when a hash is interpreted in a special way.

```
string.len("In a TeX source a hash can be a bit of a problem!")
```

```
49
```

### itself valid

In case you need a a function that returns its argument, you can use:

```
string.itself("indeed")
```

```
indeed
```

The valid checks if we have a string and if it is not empty. The default value is returned if these criteria are not met.

```
string.valid("whatever")
```

```
whatever
```

```
string.valid("","default")
```

```
default
```

### quote

When you output a value that might need to be read in again, strings need to be between quotes and non integer numbers need to be output as exact as possible. The format function has %q for that. Here is that one as function. Watch the hexadecimal double:

```
string.quote(('That was "well" done!'))
```

```
"That was \"well\" done!"
```

```
string.quote((123))
```

```
123
```

```
string.quote((123.456))
```

```
0x1.edd2f1a9fbe77p+6
```

```
string.quote((true))
```

```
true
```

## optionalquoted

This rather special function is used in command-line escaping. It is probably of little use to the average user. So:

```
\startluacode
local s = '"foo"bar \"and " whatever"'
context.formatted.type("%s : %s",s,string.optionalquoted(s))
\stopluacode
```

escapes to:

```
"foo"bar "and " whatever" : "foo\"bar\"and\" whatever"
```

while:

```
\startluacode
local s = 'foo"bar \"and " whatever'
context.formatted.type("%s : %s",s,string.optionalquoted(s))
\stopluacode
```

becomes:

```
foo"bar "and " whatever : "foo\"bar\"and\" whatever"
```

## splitup

This splits a string in single byte characters. The pattern can be a string or LPEG specification.

```
string.splitup("a b c d"," ")
```

```
a b c d
```

```
string.splitup("a,b,c,d",",")
```

```
a b c d
```

```
string.splitup("a b c d",lpeg.patterns.whitespace^1)
```

```
a b c d
```

```
string.splitup("a b c d",lpeg.patterns.whitespace^1)
```

```
a b c d
```

### bytetable

This is yet another splitter, this time single byte characters become integers:

```
string.bytetable"just bytes"
```

```
t={ 106, 117, 115, 116, 32, 98, 121, 116, 101, 115 }
```

### linetable

When splitting a string into lines we look at ASCII character 10 and 13:

```
string.linetable("a\013b\010c\013\010d")
```

```
t={ "a", "b", "c", "d" }
```

### wordsplitter

This function returns a splitter function, so it's an indirect feature:

```
\startluacode
local split = string.wordsplitter(",")
local list  = split("words,of,wisdom")
\stopluacode
```

```
string.wordsplitter(",")("words,of,wisdom")
```

```
t={ "words", "of", "wisdom" }
```

```
string.wordsplitter(",")(" words, of, wisdom ")
```

```
t={ "words", "of", "wisdom" }
```

```
string.wordsplitter(" ")("words of wisdom")
```

```
t={ "words", "of", "wisdom" }
```

```
string.wordsplitter(" ")(" words of wisdom ")
```

```
t={ "words", "of", "wisdom" }
```

### containsws

This is basically an option checker, where options are separated by spaces, like:

```
string.containsws("red green blue","green")
```

```
true
```

```
string.containsws("high left bold","italic")
```

```
false
```

## texnewlines replacenewlines

The texnewlines function normalizes newlines to those that LuaTEX expects (ASCII 13 or \r in C or Lua speak). It's not that useful outside that context.

The replacenewlines normalizes newlines to the ones the operating system expects. Normally in ConTeXt you don't need to worry about these issues.

## unquoted escapedquotes unescapedquotes

Quotes are always a potential hazard. In general, all languages have special characters which means that no matter what you prefer (some dislike TEX's backslashes but then like hashes and dots) you have to somehow escape them when they have their normal meaning. Here is a helper. We show quite some examples so that you can see the various catches.

```
string.unquoted("test")
```

```
test
```

```
string.unquoted([["t\"est"]])
```

```
t\"est
```

```
string.unquoted([["t\"est"x]])
```

```
"t\"est"x
```

```
string.unquoted("\'test\'")
```

```
test
```

```
string.unquoted('"test"')
```

```
test
```

```
string.unquoted('"test"')
```

```
test
```

```
string.unquoted("'what'ever'")
```

```
what'ever
```

```
string.unquoted("'whatever'")
```

```
whatever
```

```
string.unquoted('"what"ever"')
```

```
what"ever
```

```
string.unquoted('"whatever"')
```

```
whatever
```

You can also use `string.unquote` which is just an alias. The other two functions also deal with quotes; they complement each other.

```
string.escapedquotes("test\"test")
```

```
"test\"test"
```

```
string.unescapedquotes("test\"test")
```

```
test"test
```

### collapsespaces fullstrip

These two fit in the repertoire is space manipulators. Multiple spaces are seen as one.

```
string.collapsespaces("a b c d ")
```

```
a b c d
```

```
string.fullstrip(" test test ")
```

```
test test
```

### tformat

Because in a TEX source a percent normally starts a comment, we provide a variant where the `@` performs its function ion a pattern. Of course this only makes sense when you run some LUA from the TEX end and pass an argument.

```
string.tformat("[%i]",123)
```

```
[123]
```

```
string.tformat("[@i]",123)
```

```
[123]
```

### utflength utfvalue utfvaluetable utfcharactertable

We have an UTF library but the string library also has some UTF related functions.

```
string.utflength("TèX")
```

```
3
```

```
string.utfvalue("TèX")
```

```
84
```

```
string.utfvaluetable("TèX")
```

```
t={ 84, 232, 88 }
```

```
string.utfcharactertable("TèX")
```

```
t={ "T", "è", "X" }
```

### utfcharacter utftabletostring

The first one can handle a mix, the second just numbers:

```
string.utfcharacter(232)
```

```
è
```

```
string.utfcharacter(84, 232, 88)
```

```
TèX
```

```
string.utfcharacter{ 84, 232, 88 }
```

```
TèX
```

```
string.utftabletostring{ 84, 232, 88 }
```

```
TèX
```

### utfpadd utfpadding

For `utfpadd` the default padding character is a space, but here we use a dash to show the effect:

```
string.utfpadd("TèX", 20,"-")
```

```
----------------TèX
```

```
string.utfpadd("TèXTèX", 20,"-")
```

```
--------------TèXTèX
```

```
string.utfpadd("TèX", -20,"-")
```

```
TèX----------------
```

```
string.utfpadd("TèXTèX",-20,"-")
```

```
TèXTèX--------------
```

### toutf8 toutf16 toutf32

These three take a table and produce a byte string. We can't show the 16 and 32 bit variants so we just show a call:

```
\startluacode
local b = string.toutf8  { 84, 232, 88 }
local b = string.toutf16 { 84, 232, 88 }
local b = string.toutf32 { 84, 232, 88 }
\stopluacode
```

The last two can also be called with a second argument:

```
\startluacode
local b = string.toutf16(t, true)
local b = string.toutf32(t, true)
\stopluacode
```

in which case two or four zero characters will be added.

### tounicode16

A hexadecimal 16 bit UNICODE string (as for instance needed in the backend code) can be generated with:

```
string.tounicode16(232)
```

```
00E8
```

## 10.6 UTF

We used to have the `slunicode` library available but as most of it is not used and because it has a somewhat fuzzy state, we will no longer rely on it. In fact we only used a few functions in the `utf` namespace so as CONTEXT user you'd better stick to what is presented here. You don't have to worry how they are implemented. Depending on the version of LUATEX it can be that a library, a native function, or LPEGis used.

### char byte

As UTF is a multibyte encoding the term char in fact refers to a LUA string of one upto four 8-bit characters.

```
local b = utf.byte("å")
local c = utf.char(0xE5)
```

The number of places in CONTEXT where do such conversion is not that large: it happens mostly in tracing messages.

```
logs.report("panic","the character U+%05X is used",utf.byte("æ"))
```

```
utf.byte("æ")
```

```
230
```

```
utf.char(0xE6)
```

```
æ
```

### sub

If you need to take a slice of an UTF encoded string the `sub` function can come in handy. This function takes a string and a range defined by two numbers. Negative numbers count from the end of the string.

```
utf.sub("123456àáâãäå",1,7)
```

```
123456à
```

```
utf.sub("123456àáâãäå",0,7)
```

```
123456à
```

```
utf.sub("123456àáâãäå",0,9)
```

```
123456àáâ
```

```
utf.sub("123456àáâãäå",4)
```

```
456àáâãäå
```

```
utf.sub("123456àáâãäå",0)
```

```
123456àáâãäå
```

```
utf.sub("123456àáâãäå",0,0)
```

```
utf.sub("123456àáâãäå",4,4)
```

```
4
```

```
utf.sub("123456àáâãäå",4,0)
```

```
utf.sub("123456àáâãäå",-3,0)
```

```
utf.sub("123456àáâãäå",0,-3)
```

```
123456àáâã
```

```
utf.sub("123456àáâãäå",-5,-3)
```

```
áâã
```

```
utf.sub("123456àáâãäå",-3)
```

```
ãäå
```

## len

There are probably not that many people that can instantly see how many bytes the string in the following example takes:

```
local l = utf.len("ÀÁÂÃÄÅàáâãäå")
```

Programming languages use ASCII mostly so there each characters takes one byte. In CJK scripts however, you end up with much longer sequences. If you ever did some typesetting of such scripts you

have noticed that the number of characters on a page is less than in the case of a Latin script. As information is coded in less characters, effectively the source of a Latin or CJK document will not differ that much.

```
utf.len("ÒÓÔÕÖòóôõö")
```

```
10
```

### values characters

There are two iterators that deal with UTF. In LuaTEX these are extensions to the `string` library but for consistency we've move them to the `utf` namespace.

The following function loops over the UTF characters in a string and returns the UNICODE number in `u`:

```
for u in utf.values(str) do
    ... -- u is a number
end
```

The next one returns a string `c` that has one or more characters as UTF characters can have upto 4 bytes.

```
for c in utf.characters(str) do
    ... -- c is a string
end
```

### ustring xstring tocodes

These functions are mostly useful for logging where we want to see the UNICODE number.

```
utf.ustring(0xE6)
```

```
U+000E6
```

```
utf.ustring("ù")
```

```
U+000F9
```

```
utf.xstring(0xE6)
```

```
0x000E6
```

```
utf.xstring("à")
```

```
0x000E0
```

```
utf.tocodes("ùúü")
```

```
0x00F9 0x00FA 0x00FC
```

```
utf.tocodes("àáä","")
```

```
0x00E00x00E10x00E4
```

```
utf.tocodes("òóö","+")
```

```
0x00F2+0x00F3+0x00F6
```

### split splitlines totable

The `split` function splits a sequence of UTF characters into a table which one character per slot. The `splitlines` does the same but each slot has a line instead. The `totable` function is similar to `split`, but the later strips an optionally present UTF bom.

```
utf.split("òóö")
```

```
table: 00000412324092c0
```

### count

This function counts the number of times that a given substring occurs in a string. The patterns can be a string or an LPEG pattern.

```
utf.count("òóöòóöòóö","ö")
```

```
3
```

```
utf.count("äáàa",lpeg.P("á") + lpeg.P("à"))
```

```
2
```

### remapper replacer substituter

With `remapper` you can create a remapping function that remaps a given string using a (hash) table.

```
local remap = utf.remapper { a = 'd', b = "c", c = "b", d = "a" }

print(remap("abcd 1234 abcd"))
```

A remapper checks each character against the given mapping table. Its cousin `replacer` is more efficient and skips non matches. The `substituter` function only does a quick check first and avoids building a string with no replacements. That one is much faster when you expect not that many replacements.

The `replacer` and `substituter` functions take table as argument and an indexed as well as hashed one are acceptable. In fact you can even do things like this:

```
local rep = utf.replacer { [lpeg.patterns.digit] = "!" }
```

### is_valid

This function returns false if the argument is no valid UTF string. As LuaTeX is pretty strict with respect to the input, this function is only useful when dealing with external files.

```
function checkfile(filename)
```

```
  local data = io.loaddata(filename)
  if data and data ~= "" and not utf.is_valid(data) then
    logs.report("error","file %q contains invalid utf",filename)
  end
end
```

## 10.7 Numbers and bits

In the `number` namespace we collect some helpers that deal with numbers as well as bits. Starting with LUA 5.2 a library `bit32` is but the language itself doesn't provide for them via operators: the library uses functions to manipulate numbers upto $2^{32}$. In the latest LUATEX you can use the new bit related operators.

### tobitstring

There is no format option to go from number to bits in terms of zeros and ones so we provide a helper: `tobitsting`.

```
number.tobitstring(2013)
```

```
00000000000000000000011111011101
```

```
number.tobitstring(2013,3)
```

```
00000000000000000000011111011101
```

```
number.tobitstring(2013,1)
```

```
00000000000000000000011111011101
```

### valid

This function can be used to check or convert a number, for instance in user interfaces.

```
number.valid(12)
```

```
12
```

```
number.valid("34")
```

```
34
```

```
number.valid("ab",56)
```

```
56
```

## 10.8 LPEG patterns

For LUATEX and CONTEXT MKIV the `lpeg` library came at the right moment as we can use it in lots of places. An in-depth discussion makes no sense as it's easier to look into `l-lpeg.lua`, so we stick

to an overview.[7] Most functions return an `lpeg` object that can be used in a match. In time critical situations it's more efficient to use the match on a predefined pattern that to create the pattern new each time. Patterns are cached so there is no penalty in predefining a pattern. So, in the following example, the `splitter` that splits at the asterisk will only be created once.

```
local splitter_1 = lpeg.splitat("*")
local splitter_2 = lpeg.splitat("*")

local n, m = lpeg.match(splitter_1,"2*4")
local n, m = lpeg.match(splitter_2,"2*4")
```

## [lua] match print P R S V C Cc Cs ...

The `match` function does the real work. Its first argument is a `lpeg` object that is created using the functions with the short uppercase names.

```
local P, R, C, Ct = lpeg.P, lpeg.R, lpeg.C, lpeg.Ct

local pattern = Ct((P("[") * C(R("az")^0) * P(']') + P(1))^0)

local words = lpeg.match(pattern,"a [first] and [second] word")
```

In this example the words between square brackets are collected in a table. There are lots of examples of `lpeg` in the CONTEXT code base.

## anywhere

```
local p = anywhere(pattern)

lpeg.match(lpeg.Ct((lpeg.anywhere("->")/"!")^0), "oeps->what->more")

t={ "!", "!" }
```

## splitter splitat firstofsplit secondofsplit

The `splitter` function returns a pattern where each match gets an action applied. The action can be a function, table or string.

```
local p = splitter(pattern, action)
```

The `splitat` function returns a pattern that will return the split off parts. Unless the second argument is `true` the splitter keeps splitting

```
local p = splitat(separator,single)
```

When you need to split off a prefix (for instance in a label) you can use:

```
local p = firstofsplit(separator)
local p = secondofsplit(separator)
```

[7] If you search the web for `lua lpeg` you will end up at the official documentation and tutorial.

The first function returns the original when there is no match but the second function returns `nil` instead.

```
lpeg.match(lpeg.Ct(lpeg.splitat("->",false)), "oeps->what->more")
```

```
t={ "oeps", "what", "more" }
```

```
lpeg.match(lpeg.Ct(lpeg.splitat("->",false)), "oeps")
```

```
t={ "oeps" }
```

```
lpeg.match(lpeg.Ct(lpeg.splitat("->",true)), "oeps->what->more")
```

```
t={ "oeps", "what->more" }
```

```
lpeg.match(lpeg.Ct(lpeg.splitat("->",true)), "oeps")
```

```
t={ "oeps" }
```

```
lpeg.match(lpeg.firstofsplit(":"), "before:after")
```

```
before
```

```
lpeg.match(lpeg.firstofsplit(":"), "whatever")
```

```
whatever
```

```
lpeg.match(lpeg.secondofsplit(":"), "before:after")
```

```
after
```

```
lpeg.match(lpeg.secondofsplit(":"), "whatever")
```

```
nil
```

## split checkedsplit

The next two functions have counterparts in the `string` namespace. They return a table with the split parts. The second function omits empty parts.

```
local t = split       (separator,str)
local t = checkedsplit(separator,str)
```

```
lpeg.split(",","a,b,c")
```

```
t={ "a", "b", "c" }
```

```
lpeg.split(",",",a,,b,c,")
```

```
t={ "", "a", "", "b", "c", "" }
```

```
lpeg.checkedsplit(",",",a,,b,c,")
```

```
t={ "a", "b", "c" }
```

### stripper keeper replacer

These three functions return patterns that manipulate a string. The `replacer` gets a mapping table passed.

```
local p = stripper(str or pattern)
local p = keeper  (str or pattern)
local p = replacer(mapping)
```

```
lpeg.match(lpeg.stripper(lpeg.R("az")), "[-a-b-c-d-]")
```

```
[-----]
```

```
lpeg.match(lpeg.stripper("ab"), "[-a-b-c-d-]")
```

```
[---c-d-]
```

```
lpeg.match(lpeg.keeper(lpeg.R("az")), "[-a-b-c-d-]")
```

```
abcd
```

```
lpeg.match(lpeg.keeper("ab"), "[-a-b-c-d-]")
```

```
ab
```

```
lpeg.match(lpeg.replacer{{"a","p"},{"b","q"}}, "[-a-b-c-d-]")
```

```
[-p-q-c-d-]
```

### balancer

One of the nice things about `lpeg` is that it can handle all kind of balanced input. So, a function is provided that returns a balancer pattern:

```
local p = balancer(left,right)
```

```
lpeg.match(lpeg.Ct((lpeg.C(lpeg.balancer("{","}"))+1)^0),"{a} {b{c}}")
```

```
t={ "{a}", "{b{c}}" }
```

```
lpeg.match(lpeg.Ct((lpeg.C(lpeg.balancer("((","]"))+1)^0),"((a] ((b((c]]")
```

```
t={ "((a]", "((b((c]]" }
```

### counter

The `counter` function returns a function that returns the length of a given string. The `count` function differs from its counterpart living in the `string` namespace in that it deals with UTF and accepts strings as well as patterns.

```
local fnc = counter(lpeg.P("á") + lpeg.P("à"))
local len = fnc("äáàa")
```

### UP US UR

In order to make working with UTF-8 input somewhat more convenient a few helpers are provided.

```
local p = lpeg.UP(utfstring)
local p = lpeg.US(utfstring)
local p = lpeg.UR(utfpair)
local p = lpeg.UR(first,last)
```

```
utf.count("äáàa",lpeg.UP("áà"))
```

```
1
```

```
utf.count("äáàa",lpeg.US("àá"))
```

```
2
```

```
utf.count("äáàa",lpeg.UR("aá"))
```

```
4
```

```
utf.count("äáàa",lpeg.UR("àá"))
```

```
2
```

```
utf.count("äáàa",lpeg.UR(0x0000,0xFFFF))
```

```
4
```

### patterns

The following patterns are available in the `patterns` table in the `lpeg` namespace:

HEX alwaysmatched anything argument b_collapser b_stripper balanced beginline beginofstring bytestoHEX bytestodec bytestohex bytetoHEX bytetodec bytetohex cardinal cfloat chartonumber cnumber collapser colon comma commaspacer containseol content context cpfloat cpnumber cpunsigned csletter ctxescape cunsigned decafloat digit digits dimenpair doublequoted dquote e_collapser e_stripper emptyline endofstring eol equal escaped escapedquotes float formattednumber fullstripper hex hexadecimal hexafloat hexdigit hexdigits hextobyte hextobytes integer letter linesplitter longtostring lowercase luaescape luaquoted m_collapser m_stripper nested nestedbraces nestedbrackets nestedparents newline nodquote nonspacer nonwhitespace nospacer nosquote number oct octal octdigit octdigits paragraphs pdffromeight period propername qualified quoted rootbased semicolon sentences sign singlequoted somecontent space spaceortab spacer splitthousands splittime sqlescape sqlquoted squote stripper stripzero stripzeros tab texescape textline toentities tolower toshape toupper underscore undouble unescapedquotes unquoted unsigned unsingle unspacer uppercase url urldecoder urlencoder urlescaper urlgetcleaner urlsplitter urlunescaped urlunescaper utf16_to_utf8_be utf16_to_utf8_le utf32_to_utf8_be utf32_to_utf8_le utf8 utf8byte utf8char utf8character utf8four utf8lower utf8lowercharacter utf8one utf8shape utf8shapecharacter utf8three utf8two utf8upper utf8uppercharacter utf_16_be_nl utf_16_le_nl utf_32_be_nl utf_32_le_nl utfbom utfbom_16_be utfbom_16_le

utfbom_32_be utfbom_32_le utfbom_8 utflinesplitter utfoffset utfstricttype utftohigh utftolow utftype validatedutf validdimen validutf8 validutf8char whitespace words x_collapser x_stripper xml xmlescape

There will probably be more of them in the future.

## 10.9 IO

The io library is extended with a couple of functions as well and variables but first we mention a few predefined functions.

### [lua] open popen...

The IO library deals with in- and output from the console and files.

```
local f  = io.open(filename)
```

When the call succeeds f is a file object. You close this file with:

```
f:close()
```

Reading from a file is done with f:read(...) and writing to a file with f:write(...). In order to write to a file, when opening a second argument has to be given, often wb for writing (binary) data. Although there are more efficient ways, you can use the f:lines() iterator to process a file line by line.

You can open a process with io.popen but dealing with this one depends a bit on the operating system.

### fileseparator pathseparator

The value of the following two strings depends on the operating system that is used.

```
io.fileseparator
io.pathseparator
```

io.fileseparator

\

io.pathseparator

;

### loaddata savedata

These two functions save you some programming. The first function loads a whole file in a string. By default the file is loaded in binary mode, but when the second argument is true, some interpretation takes place (for instance line endings). In practice the second argument can best be left alone.

```
io.loaddata(filename,textmode)
```

Saving the data is done with:

```
io.savedata(filename,str)
io.savedata(filename,tab,joiner)
```

When a table is given, you can optionally specify a string that ends up between the elements that make the table.

### exists size noflines

These three function don't need much comment.

```
io.exists(filename)
io.size(filename)
io.noflines(fileobject)
io.noflines(filename)
```

### characters bytes readnumber readstring

When I wrote the icc profile loader, I needed a few helpers for reading strings of a certain length and numbers of a given width. Both accept five values of `n`: `-4`, `-2`, `1`, `2` and `4` where the negative values swap the characters or bytes.

```
io.characters(f,n) --
io.bytes(f,n)
```

The function `readnumber` accepts five sizes: `1`, `2`, `4`, `8`, `12`. The string function handles any size and strings zero bytes from the string.

```
io.readnumber(f,size)
io.readstring(f,size)
```

Optionally you can give the position where the reading has to start:

```
io.readnumber(f,position,size)
io.readstring(f,position,size)
```

### ask

In practice you will probably make your own variant of the following function, but at least a template is there:

```
io.ask(question,default,options)
```

For example:

```
local answer = io.ask("choice", "two", { "one", "two" })
```

## 10.10 File

The file library is one of the larger core libraries that comes with CONTEXT.

### pathpart basename suffix suffixes nameonly

We start with a few filename manipulators.

```
local path   = file.pathpart(name,default) -- or file.dirname
local base   = file.basename(name)
local suffix = file.suffix(name,default)   -- or file.extname or file.suffixonly
local name   = file.nameonly(name)
```

```
file.pathpart("/data/temp/whatever.cld")
```

```
/data/temp
```

```
file.pathpart("c:/data/temp/whatever.cld")
```

```
c:/data/temp
```

```
file.basename("/data/temp/whatever.cld")
```

```
whatever.cld
```

```
file.suffix("c:/data/temp/whatever.cld")
```

```
cld
```

```
file.nameonly("/data/temp/whatever.cld")
```

```
whatever
```

```
file.suffixes("c:/data/temp/whatever.a.b")
```

```
a b
```

### addsuffix replacesuffix removesuffix

These functions are used quite often:

```
local filename = file.addsuffix(filename, suffix, criterium)
local filename = file.replacesuffix(filename, suffix)
local filename = file.removesuffix(filename)
```

The first one adds a suffix unless one is present. When `criterium` is `true` no checking is done and the suffix is always appended. The second function replaces the current suffix or add one when there is none.

```
file.addsuffix("whatever","cld")
```

```
whatever.cld
```

```
file.addsuffix("whatever.tex","cld")
```

```
whatever.tex
```

```
file.addsuffix("whatever.tex","cld",true)
```

```
whatever.tex.cld
```

```
file.replacesuffix("whatever","cld")
```

```
whatever.cld
```

```
file.replacesuffix("whatever.tex","cld")
```

```
whatever.cld
```

```
file.removesuffix("whatever.tex")
```

```
whatever
```

### size

As the name suggests, this function returns the file size (here from a file that we create runtime elsewhere):

```
file.size("temp.log")
```

```
9
```

### iswritable isreadable

These two test the nature of a file:

```
file.iswritable(name) -- file.is_writable lfs.writablefile
file.isreadable(name) -- file.is_readable lfs.readablefile
```

### splitname nametotable join collapsepath

Instead of splitting off individual components you can get them all in one go:

```
local drive, path, base, suffix = file.splitname(name)
```

The `drive` variable is empty on operating systems other than MS WINDOWS. Such components are joined with the function:

```
file.join(...)
```

The given snippets are joined using the `/` as this is rather platform independent. Some checking takes place in order to make sure that nu funny paths result from this. There is also `collapsepath` that does some cleanup on a path with relative components, like `...`

```
file.splitname("a:/b/c/d.e")
```

```
a:/b/c/ d e
```

```
file.nametotable(resolvers.findfile("context.mkxl"))
```

```
t={
 ["base"]="context",
 ["name"]="context.mkxl",
 ["path"]="c:/data/develop/context/sources/",
 ["suffix"]="mkxl",
}
```

```
file.join("a","b","c.d")
```

```
a/b/c.d
```

```
file.collapsepath("a/b/../c.d")
```

```
a/c.d
```

```
file.collapsepath("a/b/../c.d",true)
```

```
t:/manuals/mkiv/external/cld/a/c.d
```

## splitpath joinpath splitbase

By default splitting a execution path specification is done using the operating system dependent separator, but you can force one as well:

```
file.splitpath(str,separator)
```

The reverse operation is done with:

```
file.joinpath(tab,separator)
```

Beware: in the following examples the separator is system dependent so the outcome depends on the platform you run on.

```
file.splitpath("a:b:c")
```

```
t={ "a:b:c" }
```

```
file.splitpath("a;b;c")
```

```
t={ "a", "b", "c" }
```

```
file.joinpath({"a","b","c"})
```

```
a;b;c
```

As bonus we have:

```
file.splitbase("temp.log")
```

```
file.splitbase("foo/temp.log")
```

```
foo/
```

### collapsepath expandname reslash

Sometimes you need the full path, without embedded `.` or `..` placeholders:

```
file.collapsepath("a/b/c/d/../e.tex")
```

```
a/b/c/e.tex
```

The next variant assembles a name from the current path and the given filename:

```
file.expandname("e.tex")
```

```
t:/manuals/mkiv/external/cld/e.tex
```

You can normalize slashes to forward slashes, which are also understood in MS WINDOWS when used in function calls:

```
file.reslash("a\\b/c.tex")
```

```
a/b/c.tex
```

### robustname

In workflows filenames with special characters can be a pain so the following function replaces characters other than letters, digits, periods, slashes and hyphens by hyphens.

```
file.robustname(str,strict)
```

```
file.robustname("We don't like this!")
```

```
We-don-t-like-this-
```

```
file.robustname("We don't like this!",true)
```

```
We-don-t-like-this
```

### savedata readdata

A fast way to save data to a file and read if back is to use the following, where the load function is equivalent to `io.loaddata`.

```
file.savedata("temp.log","demo line")
```

```
true
```

```
file.readdata("temp.log")
```

```
demo line
```

The boolean result permits checking:

```
if file.savedata("temp.log","demo line") then
    local data = file.readdata("temp.log"))
else
    print("something went wrong")
end
```

### copy

There is not much to comment on this one:

```
file.copy(oldname,newname)
```

### is_qualified_path is_rootbased_path

A qualified path has at least one directory component while a rootbased path is anchored to the root of a filesystem or drive.

```
file.is_qualified_path(filename)
file.is_rootbased_path(filename)

file.is_qualified_path("a")

false

file.is_qualified_path("a/b")

true

file.is_rootbased_path("a/b")

false

file.is_rootbased_path("/a/b")

true
```

### checksum savechecksum loadchecksum

The next two functions manage a MD5 checksum for a given file. Although there is always a change that two different files give the same checksum, in practice that seldom happens.

```
file.savechecksum("temp.log")

93A7BCF9BF7EFF92B3EFF9227CE723F6

file.loadchecksum("temp.log")

93A7BCF9BF7EFF92B3EFF9227CE723F6
```

The checksum function returns the checksum:

```
file.checksum("temp.log")
```

```
93A7BCF9BF7EFF92B3EFF9227CE723F6
```

### syncmtimes needsupdating

Instead of comparing checksums you can also check timestamps, for instance when one file is derived from another.

```
if file.needsupdating(oldname,newname,threshold) then -- default threshold 1
    -- create file "newname" from "oldname"
    file.syncmtimes(oldname,newname)
end
```

## 10.11 Dir

The `dir` library uses functions of the `lfs` library that is linked into LuaTeX.

### current

This returns the current directory:

```
dir.current()
```

### glob globpattern globfiles globdirs

The `glob` function collects files with names that match a given pattern. The pattern can have wildcards: `*` (oen of more characters), `?` (one character) or `**` (one or more directories). You can pass the function a string or a table with strings. Optionally a second argument can be passed, a table that the results are appended to.

```
local files = dir.glob(pattern,target)
local files = dir.glob({pattern,...},target)
```

The target is optional and often you end up with simple calls like:

```
local files = dir.glob("*.tex")
```

There is a more extensive version where you start at a path, and applies an action to each file that matches the pattern. You can either or not force recursion.

```
dir.globpattern(path,patt,recurse,action)
```

The `globfiles` function collects matches in a table that is returned at the end. You can pass an existing table as last argument. The first argument is the starting path, the second arguments controls analyzing directories and the third argument has to be a function that gets a name passed and is supposed to return `true` or `false`. This function determines what gets collected.

```
dir.globfiles(path,recurse,func,files)
```

The globdirs function collects directories below the given directory and also takes up to four arguments.

### push pop

You can (temporary) jump to another directory, assuming that this is permitted. After you've done the work there you can return to where you came from.

```
if dir.push("foo") then
    -- whatever you need to do
    dir.pop()
end
```

### ls

The list files function takes one argument, when no argument is given an empty string is assumed. The return value is a string where every match is on its own line.

```
local s = dir.ls()
local s = dir.ls("*.tex")
```

### makedirs

With makedirs (or its alias mkdirs) you can create the given directory. If more than one name is given they are concatinated.

```
dir.makedirs(name,...)
```

### expandname

This function tries to resolve the given path, including relative paths.

```
local fullname = dir.expandname(name)
```

```
dir.expandname(".")
```

```
t:/manuals/mkiv/external/cld
```

### found

This function takes a list of potential path names and returns the first one that exists like:

```
local found = dir.found("here","there")
```

### expandlink

This function is a bit special and depends on the operating system. Users normally don't need to call this one.

```
local realname = dir.expandlink(usedname)
```

### collectpattern

We don't use this one often, it returns a table of paths where we can find files that match the given pattern.

```
local locations = dir.collectpattern("t:/sources","*.lua",true) -- result
```

## 10.12 URL

### split hashed construct

This is a specialized library. You can split an `url` into its components. An URL is constructed like this:

```
foo://example.com:2010/alpha/beta?gamma=delta#epsilon
```

```
scheme     foo://
authority  example.com:2010
path       /alpha/beta
query      gamma=delta
fragment   epsilon
```

A string is split into a hash table with these keys using the following function:

```
url.hashed(str)
```

or in strings with:

```
url.split(str)
```

The hash variant is more tolerant than the split. In the hash there is also a key `original` that holds the original URL and and the boolean `noscheme` indicates if there is a scheme at all.

The reverse operation is done with:

```
url.construct(hash)
```

```
url.hashed("foo://example.com:2010/alpha/beta?gamma=delta#epsilon")
```

```
t={
 ["authority"]="example.com:2010",
 ["filename"]="example.com:2010/alpha/beta",
 ["fragment"]="epsilon",
 ["host"]="example.com",
 ["noscheme"]=false,
 ["original"]="foo://example.com:2010/alpha/beta?gamma=delta#epsilon",
 ["path"]="alpha/beta",
 ["port"]=2010,
 ["queries"]={
  ["gamma"]="delta",
 },
 ["query"]="gamma=delta",
```

```
 ["scheme"]="foo",
}
```

```
url.hashed("alpha/beta")
```

```
t={
 ["authority"]="",
 ["filename"]="alpha/beta",
 ["fragment"]="",
 ["noscheme"]=true,
 ["original"]="alpha/beta",
 ["path"]="alpha/beta",
 ["query"]="",
 ["scheme"]="file",
}
```

```
url.split("foo://example.com:2010/alpha/beta?gamma=delta#epsilon")
```

```
t={ "foo", "example.com:2010", "alpha/beta", "gamma=delta", "epsilon" }
```

```
url.split("alpha/beta")
```

```
t={ "", "", "", "", "" }
```

### hasscheme addscheme filename query

There are a couple of helpers and their names speaks for themselves:

```
url.hasscheme(str)
url.addscheme(str,scheme)
url.filename(filename)
url.query(str)
```

```
url.hasscheme("http://www.pragma-ade.com/cow.png")
```

```
http
```

```
url.hasscheme("www.pragma-ade.com/cow.png")
```

```
false
```

```
url.addscheme("www.pragma-ade.com/cow.png","http://")
```

```
http://:///www.pragma-ade.com/cow.png
```

```
url.addscheme("www.pragma-ade.com/cow.png")
```

```
file:///www.pragma-ade.com/cow.png
```

```
url.filename("http://www.pragma-ade.com/cow.png")
```

```
http://www.pragma-ade.com/cow.png
```

```
url.query("a=b&c=d")
```

```
t={
 ["a"]="b",
 ["c"]="d",
}
```

## 10.13 OS

### [lua luatex] env setenv getenv

In CONTEXT normally you will use the resolver functions to deal with the environment and files. However, a more low level interface is still available. You can query and set environment variables with two functions. In addition there is the `env` table as interface to the environment. This threesome replaces the built in functions.

```
os.setenv(key,value)
os.getenv(key)
os.env[key]
```

### [lua] execute

There are several functions for running programs. One comes directly from LUA, the otheres come with LUATEX. All of them are are overloaded in CONTEXT in order to get more control.

```
os.execute(...)
```

### resultof launch

The following function runs the command and returns the result as string. Multiple lines are combined.

```
os.resultof(command)
```

The next one launches a file assuming that the operating system knows what application to use.

```
os.launch(str)
```

### type name platform libsuffix binsuffix binsuffixes

There are a couple of strings that reflect the current machinery: `type` returns either `windows` or `unix`. The variable `name` is more detailed: `windows`, `msdos`, `linux`, `macosx`, etc. If you also want the architecture you can consult `platform`.

```
local t = os.type
local n = os.name
local p = os.platform
```

These three variables as well as the next two are used internally and normally they are not needed in your applications as most functions that matter are aware of what platform specific things they have to deal with.

```
local s = os.libsuffix
local b = os.binsuffix
```

These are string, not functions.

```
os.type
```

```
windows
```

```
os.name
```

```
windows
```

```
os.platform
```

```
win64
```

```
os.libsuffix
```

```
dll
```

```
os.binsuffix
```

```
exe
```

The `binsuffixes` array contains a list of suffixes specific for the operating system in use.

## [lua] time [lua] difftime

The built in time function returns a number. The accuracy is implementation dependent and not that large.

```
os.time()
```

```
1763584627
```

The `difftime` function returns the difference (a double) between two given times (integers). It is kind of inaccurate so we don't use this in CONTEXT.

## [lua] sleep

You can freeze the program by passing a number, but real small values can be unreliable as this also depends on the operating system and workload.

## [lua] date

This command returns a formatted date. You can pass a template; consult the official LUA manual for the possibilities.

```
os.date()
```

```
2025-11-19 21:37
```

## [lua] remove [lua] rename [lua] tmpname

The names tell what these commands do. Of course you have to be careful in using them:

```
if os.remove(filename) then
    -- success
end
if os.rename(oldname,newname) then
    -- success
end
```

The `os.tmpname` function returns a temporary filename but we don't rely on that in CONTEXT. Use it at your own risk.

## [lua] clock

This function returns the time elapsed since we started in milliseconds. The accuracy is okay for most case but next we will discuss a more precise variant.

```
os.clock()
```

```
3.591
```

## [luatex] times gettimeofday

Although LUA has a built in type os.time function, we normally will use the one provided by LUATEX as it is more precise:

```
os.gettimeofday()
```

```
os.gettimeofday()
```

```
1763584627.2308829
```

## runtime

More interesting is:

```
os.runtime()
```

which returns the time spent in the application so far.

```
os.runtime()
```

```
3.5280230045318604
```

### timezone

Sometimes you need to add the timezone to a verbose time and the following function does that for you.

```
os.timezone(delta)
```

```
os.timezone()
```

```
+01:00
```

```
os.timezone(1)
```

```
1
```

```
os.timezone(-1)
```

```
1
```

### startuptime localtime fulltime now

Here are a few more time functions that were introduced in LuaMetaTeX:

```
os.startuptime()
```

```
1763584623.704865
```

```
os.localtime()
```

```
2025-11-19 21:37:07
```

```
os.fulltime()
```

```
2025-11-19 21:37:07+01:00
```

```
os.now()
```

```
2025-11-19 20:37:07
```

### today

Here is a function that returns a table with various date and time related values as of today:

```
os.today()
```

```
t={
 ["day"]=19,
 ["hour"]=20,
 ["isdst"]=false,
 ["min"]=37,
 ["month"]=11,
 ["sec"]=7,
 ["wday"]=4,
 ["yday"]=323,
```

```
 ["year"]=2025,
}
```

### nofdays isleapyear weekday

Here we need to pass a year:

```
os.nofdays(2021)
```

```
364
```

```
os.isleapyear(2020)
```

```
true
```

Here we pass a day, month and year (in that order):

```
os.weekday(18,11,2025)
```

```
3
```

### uuid

A version 4 UUID can be generated with:

```
os.uuid()
```

The generator is good enough for our purpose.

```
os.uuid()
```

```
7d930ae7-4b4a-b19e-1693-12846e0b0612
```

### self*

```
os.selfbin
```

```
luametatex
```

```
os.selfcore
```

```
luametatex
```

```
os.selfdir
```

```
c:/data/develop/tex-context/tex/texmf-win64/bin
```

```
os.selflink
```

```
c:/data/develop/tex-context/tex/texmf-win64/bin
```

```
os.selfname
```

```
luametatex
```

```
os.selfpath
```

```
c:/data/develop/tex-context/tex/texmf-win64/bin
```

The `os.selfarg` table holds all the command line arguments as passed to the runner or engine. The zero entry has the program name. The `os.bits` variable has the value 32 (unlikely) or 64 (likely).

### enableansi

This is a rather specific command: it checks if a console supports ANSI escape sequences and enables that when possible. Here is an example of usage. It is rather CONTEXT specific:

```
if os.enableansi() then
    logs.setformatters("ansi")
else
    logs.setformatters("ansilog")
end

logs.report("system","OEPS") -- now system comes out in green

logs.setformatters()

logs.report("system","OEPS") -- we're back in black and white mode
```

In CONTEXT you can pass `--ansi` to the `mtxrun` and `context` runners, so normally users never have to deal with this themselves.

### where

This function returns the location where a program resides. You can also `which` which is the command on UNIX.

```
os.where("luametatex")
```

```
c:/data/develop/tex-context/tex/texmf-win64/bin/luametatex.exe
```

### uname

Here we get some information about the running system, for as far as known:

```
os.uname()
```

```
t={
 ["machine"]="x86_64",
 ["nodename"]="LAPTOP-8",
 ["release"]="",
 ["sysname"]="Windows",
 ["version"]="",
}
```

## getunamefields getnamevalues gettypevalues

As with many LuaMetaTeX features that have a Lua interface, we can get some information about the to-be-expected keys and/or values:

```
os.getunamefields()
```

```
t={ "sysname", "machine", "release", "version", "nodename" }
```

```
os.getnamevalues()
```

```
t={ "windows", "linux", "macosx", "freebsd", "bsd", "gnu", "generic" }
```

```
os.gettypevalues()
```

```
t={ "windows", "unix" }
```

## [lua] exit setexitcode

You can exit the running program. Setting the exit code is mostly relevant from within a TeX run. A set exit code wins over a passed one, so in the next case the 123 is used instead of 456:

```
os.setexitcode(123)
os.exit()
os.exit(456)
```

The set exit code is not used when the program exits otherwise. You can actually change that with `lua.setexitcode` which has a companion `lua.getexitcode`.

# 11 The LUA interface code

## 11.1 Introduction

There is a lot of LUA code in MKIV. Much is not exposed and a lot of what is exposed is not meant to be used directly at the LUA end. But there is also functionality and data that can be accessed without side effects.

In the following sections a subset of the built in functionality is discussed. There are often more functions alongside those presented but they might change or disappear. So, if you use undocumented features, be sure to tag them somehow in your source code so that you can check them out when there is an update. Best would be to have more functionality defined local so that it is sort of hidden but that would be unpractical as for instance functions are often used in other modules and or have to be available at the TEX end.

It might be tempting to add your own functions to namespaces created by CONTEXT or maybe overload some existing ones. Don't do this. First of all, there is no guarantee that your code will not interfere, nor that it overloads future functionality. Just use your own namespace. Also, future versions of CONTEXT might have a couple of protection mechanisms built in. Without doubt the following sections will be extended as soon as interfaces become more stable.

## 11.2 Characters

There are quite some data tables defined but the largest is the character database. You can consult this table any time you want but you're not supposed to add or change its content if only because changes will be overwritten when you update CONTEXT. Future versions may carry more information. The table can be accessed using an unicode number. A relative simple entry looks as follows:

```
characters.data[0x00C1]
```

```
{
 ["category"]="lu",
 ["contextname"]="Aacute",
 ["description"]="LATIN CAPITAL LETTER A WITH ACUTE",
 ["direction"]="l",
 ["lccode"]=225,
 ["linebreak"]="al",
 ["shcode"]=65,
 ["specials"]={ "char", 65, 769 },
 ["unicodeslot"]=193,
}
```

Much of this is rather common information but some of it is specific for use with CONTEXT. Some characters have even more information, for instance those that deal with mathematics:

```
characters.data[0x2190]
```

```
{
 ["category"]="sm",
```

```
 ["cjkwd"]="a",
 ["description"]="LEFTWARDS ARROW",
 ["direction"]="on",
 ["linebreak"]="ai",
 ["mathextensible"]="l",
 ["mathfiller"]="leftarrowfill",
 ["mathgroup"]="binary relation",
 ["mathlist"]={ 60, 8722 },
 ["mathmeaning"]="gets",
 ["mathspec"]={
  {
   ["class"]="relation",
   ["name"]="leftarrow",
  },
  {
   ["class"]="relation",
   ["name"]="gets",
  },
  {
   ["class"]="under",
   ["name"]="underleftarrow",
  },
  {
   ["class"]="over",
   ["name"]="overleftarrow",
  },
 },
 ["mathstretch"]="h",
 ["unicodeslot"]=8592,
}
```

Not all characters have a real entry. For instance most CJK characters are virtual and share the same data:

```
characters.data[0x3456]
```

```
{
 ["unicodeslot"]=13398,
}
```

You can also access the table using UTF characters:

```
characters.data["ä"]
```

```
{
 ["category"]="ll",
 ["contextname"]="adiaeresis",
 ["description"]="LATIN SMALL LETTER A WITH DIAERESIS",
 ["direction"]="l",
 ["linebreak"]="al",
```

```
 ["shcode"]=97,
 ["specials"]={ "char", 97, 776 },
 ["uccode"]=196,
 ["unicodeslot"]=228,
}
```

A more verbose string access is also supported:

```
characters.data["U+0070"]
```

```
{
 ["category"]="ll",
 ["cjkwd"]="na",
 ["description"]="LATIN SMALL LETTER P",
 ["direction"]="l",
 ["linebreak"]="al",
 ["mathclass"]="variable",
 ["uccode"]=80,
 ["unicodeslot"]=112,
}
```

Another (less usefull) table contains information about ranges in this character table. You can access this table using rather verbose names, or you can use collapsed lowercase variants.

```
characters.blocks["CJK Compatibility Ideographs"]
```

```
{
 ["description"]="CJK Compatibility Ideographs",
 ["first"]=63744,
 ["last"]=64255,
 ["otf"]="hang",
}
```

```
characters.blocks["hebrew"]
```

```
{
 ["description"]="Hebrew",
 ["first"]=1424,
 ["last"]=1535,
 ["otf"]="hebr",
}
```

```
characters.blocks["combiningdiacriticalmarks"]
```

```
{
 ["description"]="Combining Diacritical Marks",
 ["first"]=768,
 ["last"]=879,
}
```

Some fields can be accessed using functions. This can be handy when you need that information for tracing purposes or overviews. There is some overhead in the function call, but you get some extra testing for free. You can use characters as well as numbers as index.

```
characters.contextname("ä")
```

```
adiaeresis
```

```
characters.adobename(228)
```

```
characters.description("ä")
```

```
LATIN SMALL LETTER A WITH DIAERESIS
```

The category is normally a two character tag, but you can also ask for a more verbose variant:

```
characters.category(228)
```

```
ll
```

```
characters.category(228,true)
```

```
Letter Lowercase
```

The more verbose category tags are available in a table:

```
characters.categorytags["lu"]
```

```
Letter Uppercase
```

There are several fields in a character entry that help us to remap a character. The `lccode` indicates the lowercase code point and the `uccode` to the uppercase code point. The `shcode` refers to one or more characters that have a similar shape.

```
characters.shape ("ä")
```

```
97
```

```
characters.uccode("ä")
```

```
196
```

```
characters.lccode("ä")
```

```
228
```

```
characters.shape (100)
```

```
100
```

```
characters.uccode(100)
```

```
68
```

```
characters.lccode(100)
```

```
100
```

You can use these function or access these fields directly in an entry, but we also provide a few virtual tables that avoid accessing the whole entry. This method is rather efficient.

```
characters.lccodes["ä"]
```

```
228
```

```
characters.uccodes["ä"]
```

```
196
```

```
characters.shcodes["ä"]
```

```
97
```

```
characters.lcchars["ä"]
```

```
ä
```

```
characters.ucchars["ä"]
```

```
Ä
```

```
characters.shchars["ä"]
```

```
a
```

As with other tables, you can use a number instead of an UTF character. Watch how we get a table for multiple shape codes but a string for multiple shape characters.

```
characters.lcchars[0x00C6]
```

```
æ
```

```
characters.ucchars[0x00C6]
```

```
Æ
```

```
characters.shchars[0x00C6]
```

```
AE
```

```
characters.shcodes[0x00C6]
```

```
{ 65, 69 }
```

These codes are used when we manipulate strings. Although there are `upper` and `lower` functions in the `string` namespace, the following ones are the real ones to be used in critical situations.

```
characters.lower("ÀÁÂÃÄÅàáâãäå")
```

```
àáâãäåàáâãäå
```

```
characters.upper("ÀÁÂÃÄÅàáâãäå")
```

```
ÀÁÂÃÄÅÀÁÂÃÄÅ
```

```
characters.shaped("ÀÁÂÃÄÅàáâãäå")
```

```
AAAAAAaaaaaa
```

A rather special one is the following:

```
characters.lettered("Only 123 letters + count!")
```

```
Onlyletterscount
```

With the second argument is true, spaces are kept and collapsed. Leading and trailing spaces are stripped.

```
characters.lettered("Only 123 letters + count!",true)
```

```
Only letters count
```

Access to tables can happen by number or by string, although there are some limitations when it gets too confusing. Take for instance the number `8` and string `"8"`: if we would interpret the string as number we could never access the entry for the character eight. However, using more verbose hexadecimal strings works okay. The remappers are also available as functions:

```
characters.tonumber("a")
```

```
97
```

```
characters.fromnumber(100)
```

```
d
```

```
characters.fromnumber(0x0100)
```

```
Ā
```

```
characters.fromnumber("0x0100")
```

```
Ā
```

```
characters.fromnumber("U+0100")
```

```
Ā
```

In addition to the already mentioned category information you can also use a more direct table approach:

```
characters.categories["ä"]
```

```
ll
```

```
characters.categories[100]
```

```
ll
```

In a similar fashion you can test if a given character is in a specific category. This can save a lot of tests.

```
characters.is_character[characters.categories[67]]
```

```
true
```

```
characters.is_character[67]
```

```
true
```

```
characters.is_character[characters.data[67].category]
```

```
true
```

```
characters.is_letter[characters.data[67].category]
```

```
true
```

```
characters.is_command[characters.data[67].category]
```

```
nil
```

Another virtual table is the one that provides access to special information, for instance about how a composed character is made up of components.

```
characters.specialchars["ä"]
```

```
a
```

```
characters.specialchars[100]
```

```
d
```

The outcome is often similar to output that uses the shapecode information.

Although not all the code deep down in CONTEXT is meant for use at the user level, it sometimes can eb tempting to use data and helpers that are available as part of the general housekeeping. The next table was used when looking into sorting Korean. For practical reasons we limit the table to ten entries; otherwise we would have ended up with hundreds of pages.

```
가   ㄱ   ㅏ       HANGUL SYLLABLE GA
각   ㄱ   ㅏ   ㄱ   HANGUL SYLLABLE GAG
갂   ㄱ   ㅏ   ㄲ   HANGUL SYLLABLE GAGG
갃   ㄱ   ㅏ   ㄳ   HANGUL SYLLABLE GAGS
간   ㄱ   ㅏ   ㄴ   HANGUL SYLLABLE GAN
갅   ㄱ   ㅏ   ㄵ   HANGUL SYLLABLE GANJ
갆   ㄱ   ㅏ   ㄶ   HANGUL SYLLABLE GANH
갇   ㄱ   ㅏ   ㄷ   HANGUL SYLLABLE GAD
갈   ㄱ   ㅏ   ㄹ   HANGUL SYLLABLE GAL
갉   ㄱ   ㅏ   ㄺ   HANGUL SYLLABLE GALG
```

```
\startluacode
local data = characters.data
```

```
local map  = characters.hangul.remapped

local first, last = characters.getrange("hangulsyllables")

last = first + 9 -- for now

context.start()

context.definedfont { "file:unbatang" }

context.starttabulate { "|T||T||T||T||T|" }
for unicode = first, last do
    local character = data[unicode]
    local specials = character.specials
    if specials then
        context.NC()
        context.formatted("%04V",unicode)
        context.NC()
        context.formatted("%c",unicode)
        for i=2,4 do
            local chr = specials[i]
            if chr then
                chr = map[chr] or chr
                context.NC()
                context.formatted("%04V",chr)
                context.NC()
                context.formatted("%c",chr)
            else
                context.NC()
                context.NC()
            end
        end
        context.NC()
        context(character.description)
        context.NC()
        context.NR()
    end
end
context.stoptabulate()

context.stop()
\stopluacode
```

## 11.3 Fonts

There is a lot of code that deals with fonts but most is considered to be a black box. When a font is defined, its data is collected and turned into a form that TEX likes. We keep most of that data available at the LUA end so that we can later use it when needed. In this chapter we discuss some of

the possibilities. More details can be found in the font manual(s) so we don't aim for completeness here.

A font instance is identified by its id, which is a number where zero is reserved for the so called `nullfont`. The current font id can be requested by the following function.

```
fonts.currentid()
```

```
8
```

The `fonts.current()` call returns the table with data related to the current id. You can access the data related to any id as follows:

```
local tfmdata = fonts.identifiers[number]
```

Not all entries in the table make sense for the user as some are just meant to drive the font initialization at the TEX end or the backend. The next table lists the most important ones. Some of the tables are just shortcuts to en entry in one of the `shared` subtables.

| | | |
|---|---|---|
| `ascender` | `number` | the height of a line conforming the font |
| `descender` | `number` | the depth of a line conforming the font |
| `italicangle` | `number` | the angle of the italic shapes (if present) |
| `designsize` | `number` | the design size of the font (if known) |
| `size` | `number` | the size in scaled points if the font instance |
| `factor` | `number` | the multiplication factor for unscaled dimensions |
| `hfactor` | `number` | the horizontal multiplication factor |
| `vfactor` | `number` | the vertical multiplication factor |
| `extend` | `number` | the horizontal scaling to be used by the backend |
| `slant` | `number` | the slanting to be applied by the backend |
| `characters` | `table` | the scaled character (glyph) information (tfm) |
| `descriptions` | `table` | the original unscaled glyph information (otf, afm, tfm) |
| `indices` | `table` | the mapping from unicode slot to glyph index |
| `unicodes` | `table` | the mapoing from glyph names to unicode |
| `marks` | `table` | a hash table with glyphs that are marks as entry |
| `parameters` | `table` | the font parameters as TEX likes them |
| `mathconstants` | `table` | the OPENTYPE math parameters |
| `mathparameters` | `table` | a reference to the `MathConstants` table |
| `shared` | `table` | a table with information shared between instances |
| `unique` | `table` | a table with information unique for this instance |
| `unscaled` | `table` | the unscaled (intermediate) table |
| `goodies` | `table` | the CONTEXT specific extra font information |
| `fonts` | `table` | the table with references to other fonts |
| `cidinfo` | `table` | a table with special information for the backend |
| `filename` | `string` | the full path of the loaded font |
| `fontname` | `string` | the font name as specified in the font (limited in size) |
| `fullname` | `string` | the complete font name as specified in the font |
| `name` | `string` | the (short) name of the font |
| `psname` | `string` | the (unique) name of the font as used by the backend |
| `hash` | `string` | the hash that makes this instance unique |

| | | |
|---|---|---|
| id | number | the id (number) that TEX will use for this instance |
| type | string | an idicator if the font is virtual or real |
| format | string | a qualification for this font, e.g. opentype |
| mode | string | the CONTEXT processing mode, node or base |

The parameters table contains variables that are used by TEX itself. You can use numbers as index and these are equivalent to the so called \fontdimen variables. More convenient is is to access by name:

| | |
|---|---|
| slant | the slant per point (seldom used) |
| space | the interword space |
| spacestretch | the interword stretch |
| spaceshrink | the interword shrink |
| xheight | the x-height (not per se the heigth of an x) |
| quad | the so called em-width (often the width of an emdash) |
| extraspace | additional space added in specific situations |

The math parameters are rather special and explained in the LUATEX manual. Quite certainly you never have to touch these parameters at the LUA end.

En entry in the characters table describes a character if we have entries within the UNICODE range. There can be entries in the private area but these are normally variants of a shape or special math glyphs.

| | |
|---|---|
| name | the name of the character |
| index | the index in the raw font table |
| height | the scaled height of the character |
| depth | the scaled depth of the character |
| width | the scaled height of the character |
| tounicode | a UTF-16 string representing the conversion back to unicode |
| expansion_factor | a multiplication factor for (horizontal) font expansion |
| left_protruding | a multiplication factor for left side protrusion |
| right_protruding | a multiplication factor for right side protrusion |
| italic | the italic correction |
| next | a pointer to the next character in a math size chain |
| vert_variants | a pointer to vertical variants conforming OPENTYPE math |
| horiz_variants | a pointer to horizontal variants conforming OPENTYPE math |
| top_accent | information with regards to math top accents |
| mathkern | a table describing stepwise math kerning (following the shape) |
| kerns | a table with intercharacter kerning dimensions |
| ligatures | a (nested) table describing ligatures that start with this character |
| commands | a table with commands that drive the backend code for a virtual shape |

Not all entries are present for each character. Also, in so called node mode, the ligatures and kerns tables are empty because in that case they are dealt with at the LUA end and not by TEX.

Say that you run into a glyph node and want to access the data related to that glyph. Given that variable n points to the node, the most verbose way of doing that is:

```
local g = fonts.identifiers[n.id].characters[n.char]
```

Given the speed of LUATEX this is quite fast. Another method is the following:

```
local g = fonts.characters[n.id][n.char]
```

For some applications you might want faster access to critical parameters, like:

```
local quad    = fonts.quads   [n.id][n.char]
local xheight = fonts.xheights[n.id][n.char]
```

but that only makes sense when you don't access more than one such variable at the same time.

Among the shared tables is the feature specification:

```
fonts.current().shared.features
```

```
{
 ["analyze"]=true,
 ["autolanguage"]="position",
 ["autoscript"]="position",
 ["checkautoeffect"]=true,
 ["checkmarks"]=true,
 ["checkmissing"]=true,
 ["compoundhyphen"]=true,
 ["curs"]=true,
 ["devanagari"]=true,
 ["dummies"]=true,
 ["extensions"]=true,
 ["extrafeatures"]=true,
 ["extraprivates"]=true,
 ["features"]=true,
 ["fixdot"]=true,
 ["goodies"]="texgyre-text",
 ["indic"]="auto",
 ["itlc"]=true,
 ["kern"]=true,
 ["liga"]=true,
 ["mark"]=true,
 ["mathrules"]=true,
 ["mkmk"]=true,
 ["mode"]="node",
 ["number"]=35,
 ["script"]="dflt",
 ["spacekern"]=true,
 ["textcontrol"]="collapsehyphens,replaceapostrophe",
 ["visualspace"]=true,
 ["wipemath"]=true,
}
```

As features are a prominent property of OPENTYPE fonts, there are a few datatables that can be used to get their meaning.

```
fonts.handlers.otf.tables.features['liga']
```

```
standard ligatures
```

```
fonts.handlers.otf.tables.languages['nld']
```

```
dutch
```

```
fonts.handlers.otf.tables.scripts['arab']
```

```
arabic
```

There is a rather extensive font database built in but discussing its interface does not make much sense. Most usage happens automatically when you use the `name:` and `spec:` methods of defining fonts and the `mtx-fonts` script is built on top of it.

```
table.sortedkeys(fonts.names.data)
```

```
{ "cache_format", "cache_uuid", "cache_version", "datastate", "fallbacks", "families",
"files", "indices", "mappings", "names", "rejected", "sorted_fallbacks", "sorted_famili
"sorted_mappings", "specifications", "statistics", "version" }
```

You can load the database (if it's not yet loaded) with:

```
names.load(reload,verbose)
```

When the first argument is true, the database will be rebuild. The second arguments controls verbosity.

Defining a font normally happens at the TEX end but you can also do it in LUA.

```
local id, fontdata = fonts.definers.define {
    lookup = "file",             -- use the filename (file spec name)
    name   = "pagella-regular",  -- in this case the filename
    size   = 10*65535,           -- scaled points
    global = false,              -- define the font globally
    cs     = "MyFont",           -- associate the name \MyFont
    method = "featureset",       -- featureset or virtual (* or @)
    sub    = nil,                -- no subfont specifier
    detail = "whatever",         -- the featureset (or whatever method applies)
}
```

In this case the `detail` variable defines what featureset has to be applied. You can define such sets at the LUA end too:

```
fonts.definers.specifiers.presetcontext (
    "whatever",
    "default",
    {
        mode = "node",
        dlig = "yes",
    }
)
```

The first argument is the name of the featureset. The second argument can be an empty string or a reference to an existing featureset that will be taken as starting point. The final argument is the featureset. This can be a table or a string with a comma separated list of key/value pairs.

## 11.4 Nodes

Nodes are the building blocks that make a document reality. Nodes are linked into lists and at various moments in the typesetting process you can manipulate them. Deep down in CONTEXT we use quite some LUA magic to manipulate lists of nodes. Therefore it is no surprise that we have some tracing available. Take the following box.

```
\setbox0\hbox{It's in \hbox{\bf all} those nodes.}
```

This box contains characters and glue between the words. The box is already constructed. There can also be kerns between characters, but of course only if the font provides such a feature. Let's inspect this box:

```
nodes.toutf(tex.box[0])
```

```
It's in all those nodes.
```

```
nodes.toutf(tex.box[0].list)
```

```
It's in all those nodes.
```

This tracer returns the text and spacing and recurses into nested lists. The next tracer does not do this and marks non glyph nodes as `[-]`:

```
nodes.listtoutf(tex.box[0])
```

```
[-]
```

```
nodes.listtoutf(tex.box[0].list)
```

```
It'[-]s[-]in[-][-][-]t[-]hose[-]nodes.
```

A more verbose tracer is the next one. It does show a bit more detailed information about the glyphs nodes.

```
nodes.tosequence(tex.box[0])
```

```
hlist
```

```
nodes.tosequence(tex.box[0].list)
```

```
U+0049:I U+0074:t U+2019:' kern U+0073:s glue U+0069:i U+006E:n glue hlist glue
U+0074:t kern U+0068:h U+006F:o U+0073:s U+0065:e glue U+006E:n U+006F:o U+0064:d
U+0065:e U+0073:s U+002E:.
```

The fourth tracer does not show that detail and collapses sequences of similar node types.

```
nodes.idstostring(tex.box[0])
```

```
[hlist]
```

```
nodes.idstostring(tex.box[0].list)
```

```
[3*glyph] [kern] [glyph] [glue] [2*glyph] [glue] [hlist] [glue] [glyph] [kern]
[4*glyph] [glue] [6*glyph]
```

The number of nodes in a list is identified with the `countall` function. Nested nodes are counted too.

```
nodes.countall(tex.box[0])
```

```
28
```

```
nodes.countall(tex.box[0].list)
```

```
27
```

There are a lot of helpers in the `nodes` namespace. In fact, we map all the helpers provided by the engine itself under `nodes` too. These are described in the LUATEX manual. There are for instance functions to check node types and node id's:

```
local str = node.type(1)
local num = node.id("vlist")
```

These are basic LUATEX functions. In addition to those we also provide a few more helpers as well as mapping tables. There are two tables that map node id's to strings and backwards:

| | |
|---|---|
| `nodes.nodecodes` | regular nodes, some fo them are sort of private to the engine |
| `nodes.noadcodes` | math nodes that later on are converted into regular nodes |

Nodes can have subtypes. Again we have tables that map the subtype numbers onto meaningfull names and reverse.

| | |
|---|---|
| `nodes.listcodes` | subtypes of `hlist` and `vlist` nodes |
| `nodes.kerncodes` | subtypes of `kern` nodes |
| `nodes.gluecodes` | subtypes of `glue` nodes (skips) |
| `nodes.glyphcodes` | subtypes of `glyph` nodes, the subtype can change |
| `nodes.mathcodes` | math specific subtypes |
| `nodes.fillcodes` | these are not really subtypes but indicate the strength of the filler |
| `nodes.whatsitcodes` | subtypes of a rather large group of extension nodes |

Some of the names of types and subtypes have underscores but you can omit them when you use these tables. You can use tables like this as follows:

```
local glyph_code = nodes.nodecodes.glyph
local kern_code  = nodes.nodecodes.kern
local glue_code  = nodes.nodecodes.glue

for n in nodes.traverse(list) do
    local id == n.id
    if id == glyph_code then
        ...
    elseif id == kern_code then
        ...
```

```
    elseif id == glue_code then
        ...
    else
        ...
    end
end
```

You only need to use such temporary variables in time critical code. In spite of what you might think, lists are not that long and given the speed of LUA (and successive optimizations in LUATEX) looping over a paragraphs is rather fast.

Nodes are created using `node.new`. If you study the CONTEXT code you will notice that there are quite some functions in the `nodes.pool` namespace, like:

```
local g = nodes.pool.glyph(fnt,chr)
```

Of course you need to make sure that the font id is valid and that the referred glyph in in the font. You can use the allocators but don't mess with the code in the `pool` namespace as this might interfere with its usage all over CONTEXT.

The `nodes` namespace provides a couple of helpers and some of them are similar to ones provided in the `node` namespace. This has practical as well as historic reasons. For instance some were prototypes functions that were later built in.

```
local head, current      = nodes.before (head, current, new)
local head, current      = nodes.after  (head, current, new)
local head, current      = nodes.delete (head, current)
local head, current      = nodes.replace(head, current, new)
local head, current, old = nodes.remove (head, current)
```

Another category deals with attributes:

```
nodes.setattribute      (head, attribute, value)
nodes.unsetattribute    (head, attribute)
nodes.setunsetattribute (head, attribute, value)
nodes.setattributes     (head, attribute, value)
nodes.unsetattributes   (head, attribute)
nodes.setunsetattributes(head, attribute, value)
nodes.hasattribute      (head, attribute, value)
```

## 11.5 Resolvers

All IO is handled by functions in the `resolvers` namespace. Most of the code that you find in the `data-*.lua` files is of litle relevance for users, especially at the LUA end, so we won't discuss it here in great detail.

The resolver code is modelled after the KPSE library that itself implements the TEX Directory Structure in combination with a configuration file. However, we go a bit beyond this structure, for instance in integrating support for other resources that file systems. We also have our own configuration file. But important is that we still support a similar logic too so that regular configurations are dealt with.

During a run LUATEX needs files of a different kind: source files, font files, images, etc. In practice you will probably only deal with source files. The most fundamental function is `findfile`. The first argument is the filename to be found. A second optional argument indicates the file type.

The following table relates so called formats to suffixes and variables in the configuration file.

| variable | format | suffix |
|---|---|---|
| AFMFONTS | afm | afm |
| | adobe font metric | |
| | adobe font | |
| | metrics | |
| BIBINPUTS | bib | bib |
| BSTINPUTS | bst | bst |
| FONTCIDMAPS | cid | cid cidmap |
| | cid map | |
| | cid maps | |
| | cid file | |
| | cid files | |
| FONTFEATURES | fea | fea |
| | font feature | |
| | font features | |
| | font feature file | |
| | font feature | |
| | files | |
| TEXFORMATS | fmt | fmt |
| | format | |
| | tex format | |
| FONTCONFIG_PATH | fontconfig | |
| | fontconfig file | |
| | fontconfig files | |
| ICCPROFILES | icc | icc |
| | icc profile | |
| | icc profiles | |
| CLUAINPUTS | lib | dll |
| LUAINPUTS | lua | lmt lua lud tma tmd |
| MPMEMS | mem | mem |
| | metapost format | |
| MPINPUTS | mp | mpxl mpvi mpiv mpii mp |
| OFMFONTS | ofm | ofm tfm |
| | omega font metric | |
| | omega font | |
| | metrics | |
| OPENTYPEFONTS | otf | otf |
| | opentype | |
| | opentype font | |
| | opentype fonts | |
| OVFFONTS | ovf | ovf vf |
| | omega virtual | |
| | font | |

```
                  omega virtual
                  fonts
T1FONTS           pfb               pfb pfa
                  type1
                  type 1
                  type1 font
                  type 1 font
                  type1 fonts
                  type 1 fonts
PKFONTS           pk                pk
TEXINPUTS         tex               tex mkiv mkvi mkxl mklx mkii cld lfg xml
TEXMFSCRIPTS      texmfscript       lua rb pl py
                  texmfscripts
                  script
                  scripts
TFMFONTS          tfm               tfm
                  tex font metric
                  tex font metrics
TTFONTS           ttf               ttf ttc dfont
                  truetype
                  truetype font
                  truetype fonts
                  truetype collection
                  truetype collections
                  truetype dictionary
                  truetype dictionaries
VFFONTS           vf                vf
                  virtual font
                  virtual fonts
```

There are a couple of more formats but these are not that relevant in the perspective of CONTEXT.

When a lookup takes place, spaces are ignored and formats are normalized to lowercase.[8]

```
resolvers.findfile("context.tex")
```

```
resolvers.findfile("context.mkiv")

c:/data/develop/context/sources/context.mkiv

resolvers.findfile("context")

c:/data/develop/tex-context/tex/texmf-context/scripts/context/stubs/unix/context

resolvers.findfile("data-res.lua")

c:/data/develop/context/sources/data-res.lua
```

[8] Here the text might run off the page depending on where the files are located.

```
resolvers.findfile("lmsans10-bold")
```

```
resolvers.findfile("lmsans10-bold.otf")
```

```
c:/data/develop/tex-context/tex/texmf/fonts/opentype/public/lm/lmsans10-bold.otf
```

```
resolvers.findfile("lmsans10-bold","otf")
```

```
c:/data/develop/tex-context/tex/texmf/fonts/opentype/public/lm/lmsans10-bold.otf
```

```
resolvers.findfile("lmsans10-bold","opentype")
```

```
c:/data/develop/tex-context/tex/texmf/fonts/opentype/public/lm/lmsans10-bold.otf
```

```
resolvers.findfile("lmsans10-bold","opentypefonts")
```

```
c:/data/develop/tex-context/tex/texmf/fonts/opentype/public/lm/lmsans10-bold.otf
```

```
resolvers.findfile("lmsans10-bold","opentype fonts")
```

```
c:/data/develop/tex-context/tex/texmf/fonts/opentype/public/lm/lmsans10-bold.otf
```

The plural variant of this function returns one or more matches.[9]

```
resolvers.findfiles("texmfcnf.lua","cnf")
```

```
{ "c:/data/develop/tex-context/tex/texmf-local/web2c/texmfcnf.lua" }
```

```
resolvers.findfiles("context.mkiv","")
```

```
{ "c:/data/develop/context/sources/context.mkiv", "c:/data/develop/tex-context/tex/texm
}
```

[9] Again we can run off the page.

# 12 Scanners

## 12.1 Introduction

Here we discuss methods to define macros that directly interface with the LUA side. It involves all kind of scanners. There are actually more than we discuss here but some are meant for low level usage. What is describe here has been used for ages and works quite well.

*We don't discuss some of the more obscure options here. Some are there just because we need them as part of bootstrapping or initializing code and are of no real use to users.*

## 12.2 A teaser first

Most of this chapter is examples and you learn TEX (and LUA) best by just playing around. The nice thing about TEX is that it's all about visual output, so that's why in the next examples we just typeset some of what we just scanned. Of course in practice the `actions` will be more complex.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroA",
        public    = true,
        permanent = false,
        arguments = "string",
        actions   = function(s)
            context("(%s)",s)
        end,
    }
\stopluacode
```

By default a macro gets defined in the `\clf_` namespace but the `public` option makes it visible. This default indicates that it is actually a low level mechanism in origin. More often than not these interfaces are used like this:

```
\def\MyMacro#1{... \clf_MyMacroA{#1} ...}
```

When we look at the meaning of `\MyMacroA` we get:

```
\MyMacroA luacall \MyMacroA
```

And when we apply this macro as:

```
\MyMacroA{123}
\MyMacroA{abc}
\edef\temp{\MyMacroA{abc}}
```

We get

(123) (abc)

The meaning of `\temp` is:

```
\temp macro:(abc)
```

We can also define the macro to be protected (`\unexpanded`) in CONTEXT speak). We can overload existing scanners but unless we specify the `overload` option, we get a warning on the console. However, in LMTX there is catch. Implementers by default define macros as permanent unless one explicitly disables this so this is why in the previous definition we have done so. The overload flag below only makes sense in special cases, when for instance format file is made. (Of course overload protection only kicks in when it has been enabled.)

```
\startluacode
    interfaces.implement {
        name      = "MyMacroA",
        public    = true,
     -- overload  = true,
        protected = true,
        arguments = "string",
        actions   = function(s)
            context("[%s]",s)
        end,
    }
\stopluacode
```

This time we get:

[123] [abc]

The meaning of `\temp` is:

```
\temp macro:\MyMacroA {abc}
```

## 12.3 Basic data types

It is actually possible to write very advanced scanners but unless you're in for obscurity the limited subset discussed here is normally enough. The CONTEXT user interface is rather predictable, unless you want to show off with weird additional interfaces, for instance by using delimiters other than curly braces and brackets, or by using separators other than commas.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroB",
        public    = true,
        arguments = { "string", "integer", "boolean", "dimen" },
        actions   = function(s,i,b,d)
            context("<%s> <%i> <%l> <%p>",s,i,b,d)
        end,
    }
\stopluacode
```

This time we grab four arguments, each of a different type:

```
\MyMacroB{foo} 123 true 45.67pt
```

```
\def\temp       {oof}
\scratchcounter 321
\scratchdimen   76.54pt

\MyMacroB\temp \scratchcounter false \scratchdimen
```

The above usage gives:

<foo> <123> <true> <45.67pt>

<oof> <321> <false> <76.53999pt>

As you can see, registers can be used as well, and the `\temp` macro is also accepted as argument. The integer and dimen arguments scan standard TEX values. If you want a LUA number you can specify that as well. As our first example showed, when there is one argument you don't need an array to specify it.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroC",
        public    = true,
        arguments = "number",
        actions   = function(f)
            context("<%.2f>",f)
        end,
    }
\stopluacode
```

As you can see, hexadecimal numbers are also accepted:

```
\MyMacroC 1.23
\MyMacroC 1.23E4
\MyMacroC -1.23E4
\MyMacroC 0x1234
```

The above usage gives:

<1.23><12300.00><-12300.00><4660.00>

## 12.4 Tables

A list can be grabbed too. The individual items are separated by spaces and items can be bound by braces.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroD",
        public    = true,
        arguments = "list",
        actions   = function(t)
            context("< % + t >",t)
```

```
        end,
    }
\stopluacode
```

The macro call:

```
\MyMacroD { 1 2 3 4 {5 6} }
```

results in:

< 1 + 2 + 3 + 4 + 5 6 >

Often in LUA scripts tables are uses all over the place. Picking up a table is also supported by the implementer.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroE",
        public    = true,
        arguments = {
            {
                { "bar", "integer" },
                { "foo", "dimen" },
            }
        },
        actions   = function(t)
            context("<foo : %p> <bar : %i>",t.foo,t.bar)
        end,
    }
\stopluacode
```

Watch out, we don't use equal signs and commas here:

```
\MyMacroE {
    foo 12pt
    bar 34
}
```

We get:

<foo : 12pt> <bar : 34>

All the above can be combined:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroF",
        public    = true,
        arguments = {
            "string",
            {
                { "bar", "integer" },
```

```
                { "foo", "dimen" },
            },
            {
                { "one", "string" },
                { "two", "string" },
            },
        },
        actions   = function(s,t1,t2)
            context("<%s> <%p> <%i> <%s> <%s>",s,t1.foo,t1.bar,t2.one,t2.two)
        end,
    }
\stopluacode
```

The following call:

```
\MyMacroF
    {oeps}
    { foo 12pt bar 34 }
    { one {x} two {y} }
```

Results in one string and two table arguments.

<oeps> <12pt> <34> <x> <y>

You can nest tables, as in:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroG",
        public    = true,
        arguments = {
            "string",
            "string",
            {
                { "data", "string" },
                { "tab", "string" },
                { "method", "string" },
                { "foo", {
                    { "method", "integer" },
                    { "compact", "number" },
                    { "nature" },
                    { "*" }, -- any key
                } },
                { "compact", "string", "tonumber" },
                { "nature", "boolean" },
                { "escape" },
            },
            "boolean",
        },
        actions   = function(s1, s2, t, b)
```

```
            context("<%s> <%s>",s1,s2)
            context("<%s> <%s> <%s>",t.data,t.tab,t.compact)
            context("<%i> <%s> <%s>",t.foo.method,t.foo.nature,t.foo.whatever)
            context("<%l>",b)
        end,
    }
\stopluacode
```

Although the `\relax` is not really needed in the next calls, I often use it to indicate that we're done:

```
\MyMacroG
    {s1}
    {s2}
    {
        data    {d}
        tab     {t}
        compact {12.34}
        foo     { method 1 nature {n} whatever {w} }
    }
    true
\relax
```

This typesets:

<s1> <s2><d> <t> <12.34><1> <n> <w><true>

## 12.5 Expansion

When working with scanners it is important to realize that we have to do with an expansion engine. When TEX picks up a token, it can be done as-is, that is the raw token, but it can also expand that token first (which can be recursive) and then pick up the first token that results from that. Sometimes you want that expansion, for instance when you pick up keywords, sometimes you don't.

Expansion effects are most noticeable when we pickup a 'string' kind of value. In the implementor we have two methods for that: `string` and `argument`. The argument method has an expandable form (the default) and one that doesn't expand. Take this:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroH",
        public    = true,
        arguments = {
            "string",
            "argument",
            "argumentasis",
        },
        actions   = function(a,b,c)
            context.type(a or "-") context.quad()
            context.type(b or "-") context.quad()
            context.type(c or "-") context.crlf()
```

```
        end,
    }
\stopluacode
```

Now take this input:

```
\def\Ca{A} \def\Cb{B} \def\Cc{C}
\MyMacroH{a}{b}{c}
\MyMacroH{a\Ca}{b\Cb}{c\Cc}
\MyMacroH\Ca\Cb\Cc\relax
\MyMacroH\Ca xx\relax
```

We we use the string method we need a `\relax` (or some spacer) to end scanning of the string when we don't use curly braces. The last line is actually kind of tricky because the macro expects two arguments after scanning the first string.

```
a  b  c
aA  bB  c\Cc
ABC  -  -
Axx  -  -
```

Here is a variant:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroI",
        public    = true,
        arguments = {
            "argument",
            "argumentasis",
        },
        actions   = function(a,b,c)
            context.type(a or "-") context.quad()
            context.type(b or "-") context.quad()
            context.type(c or "-") context.crlf()
        end,
    }
\stopluacode
```

With:

```
\def\a{A} \def\b{B}
\MyMacroI{a}{b}
\MyMacroI{a\a}{b\b}
\MyMacroI\a\b\relax
```

we get:

```
a  b  -
aA  b\b  -
```

```
A  B  -
```

## 12.6 Boxes

You can pick up a box too. The value returned is a list node:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroJ",
        public    = true,
        arguments = "box",
        actions   = function(b)
            context(b)
        end,
    }
\stopluacode
```

The usual box specifiers are supported:

So, with:

```
\MyMacroJ \hbox        {\strut Test 1}
\MyMacroJ \hbox to 4cm {\strut Test 2}
```

we get:

Test 1
Test 2

There are three variants that don't need the box operator `hbox`, `vbox` and `vtop`:

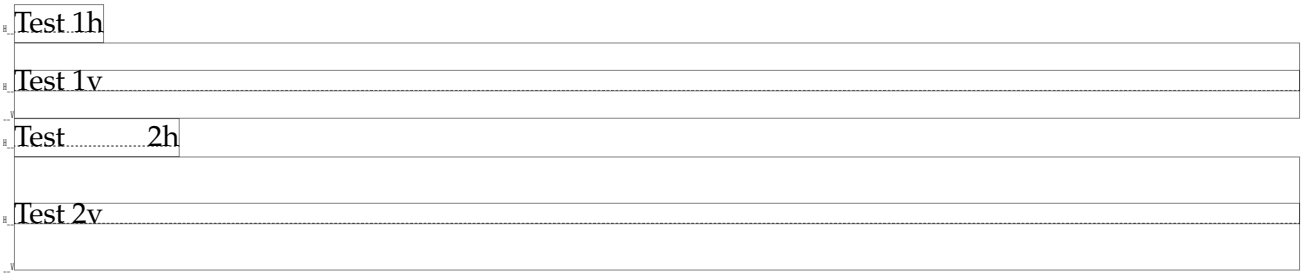
```
\startluacode
    interfaces.implement {
        name      = "MyMacroL",
        public    = true,
        arguments = {
            "hbox",
            "vbox",
        },
        actions   = function(h,v)
            context(h)
            context(v)
        end,
    }
\stopluacode
```

Again, the usual box specifiers are supported:

This:

```
\MyMacroL            {\strut Test 1h} to 10mm {\vfill Test 1v\vfill}
\MyMacroL spread 1cm {\strut Test 2h} to 15mm {\vfill Test 2v\vfill}
```

gives:




## 12.7 Like CONTEXT

The previously discussed scanners don't use equal signs and commas as separators, but you can enforce that regime in the following way:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroN",
        public    = true,
        arguments = {
            "hash",
            "array",
        },
        actions   = function(h, a)
            context.totable(h)
            context.quad()
            context.totable(a)
        end,
    }
\stopluacode
```

This:

```
\MyMacroN
    [ a = 1,  b = 2 ]
    [ 3, 4, 5, {6 7} ]
```

gives:

[b=2 ,a=1] [3,4,5,6 7]

## 12.8 Verbatim

There are a couple of rarely used scanners (there are more of course but these are pretty low level and not really used directly using implementors).

```
\startluacode
```

```
    interfaces.implement {
        name      = "MyMacroO",
        public    = true,
        arguments = "verbatim",
        actions   = function(v)
            context.type(v)
        end,
    }
\stopluacode
```

There is no expansion applied in:

```
\MyMacroO{this is \something verbatim}
```

so we get what we input:

```
this is \something verbatim
```

## 12.9 Macros

We can pick up a control sequence without bothering what it actually represents:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroP",
        public    = true,
        arguments = "csname",
        actions   = function(c)
            context("{\\ttbf name:} {\\tttf %s}",c)
        end,
    }
\stopluacode
```

The next control sequence is picked up and its name without the leading escape character is returned:

```
\MyMacroP\framed
```

So here we get:

**name:** framed

## 12.10 Token lists

If you have no clue what tokens are in the perspective of TEX, you can skip this section. We can grab a token list in two ways. The most LUAish way is to grab it as a table:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroQ",
        public    = true,
        arguments = "toks",
```

```
        actions   = function(t)
            context("%S : ",t)
            context.sprint(t)
            context.crlf()
        end,
    }
\stopluacode

\MyMacroQ{this is a {\bf token} list}
\MyMacroQ{this is a \inframed{token} list}
```

The above sample code gives us:

table: 000004123807fcb0 : this is a **token** list
table: 0000041235facc80 : this is a token list

An alternative is to keep the list a user data object:

```
\startluacode
    interfaces.implement {
        name      = "MyMacroR",
        public    = true,
        arguments = "tokenlist",
        actions   = function(t)
            context("%S : ",t)
            context.sprint(t)
            context.crlf()
        end,
    }
\stopluacode

\MyMacroR{this is a {\bf token} list}
\MyMacroR{this is a \inframed{token} list}
```

Now we get:

<lua token : 620502 => 620487 : refcount> : this is a **token** list
<lua token : 623512 => 622396 : refcount> : this is a token list

## 12.11 Actions

The plural `actions` suggests that there can be more than one and indeed that is the case. The next example shows a sequence of actions that are applied. The first one gets the arguments passes.

```
\startluacode
    interfaces.implement {
        name      = "MyMacroS",
        public    = true,
        arguments = "string",
```

```
        actions   = { characters.upper, context },
    }
\stopluacode

\MyMacroS{uppercase}
```

Gives: UPPERCASE

You can pass default arguments too. That way you can have multiple macros using the same action. Here's how to do that:

```
\startluacode
    local function MyMacro(a,b,sign)
        if sign then
            context("$%i + %i = %i$",a,b,a+b)
        else
            context("$%i - %i = %i$",a,b,a-b)
        end
    end

    interfaces.implement {
        name      = "MyMacroPlus",
        public    = true,
        arguments = { "integer", "integer", true },
        actions   = MyMacro,
    }

    interfaces.implement {
        name      = "MyMacroMinus",
        public    = true,
        arguments = { "integer", "integer", false },
        actions   = MyMacro,
    }
\stopluacode
```

So,

```
\MyMacroPlus  654 321 \crlf
\MyMacroMinus 654 321 \crlf
```

Gives:

$654 + 321 = 975$
$654 - 321 = 333$

If you need to pass a string, you pass it as `"'preset'"`, so single quotes inside the double ones. Otherwise strings are interpreted as scanner types.

## 12.12 Embedded LUA code

When you mix TEX and LUA, you can put the LUA code in a TEX file, for instance a style. In the previous sections we used this approach:

```
\startluacode
    -- lua code
\stopluacode
```

This method is both reliable and efficient but you need to keep into mind that macros get expanded. In the next code, the second line will give an error when you have not defined `\foo` as expandable macro. When it is unexpandable it will get passed as it is and LUA will see a `\f` as an escaped character. So, when you want a macro be passes as macro, you need to do it as in the third line. The fact that there is a comment trigger (`--`) doesn't help here.

```
\startluacode
    context("foo")
 -- context("\foo")
    context("\\bar")
\stopluacode
```

When you use `\ctxlua` the same is true but there you also need to keep an eye on special characters. For instance a percent sign is then interpreted in the TEX way and because all becomes one line, a `--` somewhere in the middle will make the rest of the line comment:

```
\ctxlua {
    context("foo")
 %  context("\foo")
 -- context("\\bar")
}
```

Here, the second line goes away (TEX comment) and the third line obscures all that follows. You can use `\letterpercent` to smuggle a percent sign in a `\ctxlua` call. Special characters like a hash symbol also need to be passed by name. Normally curly braces are no problem because LUA also likes them properly nested.

When things become too messy and complex you can always put the code in an external file and load that one (e.g. with `require`.

In the examples in this chapter we put the function in the table, but for long ones you might want to do this:

```
\startluacode
    local function MyMacro(s)
        -- lots of code
    end

    interfaces.implement {
        name      = "MyMacro",
        public    = true,
        arguments = "string",
```

```
        actions   = MyMacro,
    }
\stopluacode
```

It is a common mistake not to define variables and functions as local. If you define them global for sure there will become a time when this bites you.

# 13 Variables

## 13.1 Introduction

Sometimes a bit of experimenting and exploring the boundaries of the TEX-LUA interfaces results in new mechanisms. To what extent they are really useful is hard to say but we just keep them around. Some are described here. This is, at least for the moment, a LMTX specific chapter.

## 13.2 Simple tables

The basic TEX data types are counters (integers), dimensions (kind of floating point variables with typographic dimensions), token lists, node lists (boxes), fonts, and so on, but no data organized in tables. The first mechanism that we discuss is one that obeys grouping. In that respect is behaves like regular registers.

```
\newhashedtable\somehashtable

\somehashtable{ foo = 123, bar = "foo" }
[123 = \the\somehashtable{foo}]
[foo = \the\somehashtable{bar}]

\bgroup
\somehashtable{ foo = 456, bar = "oof" }
[456 = \the\somehashtable{foo}]
[oof = \the\somehashtable{bar}]
\egroup

[123 = \the\somehashtable{foo}]
[foo = \the\somehashtable{bar}]

\bgroup
\global\somehashtable{ foo = 456 }
       \somehashtable{ bar = "oof" }
[456 = \the\somehashtable{foo}]
[oof = \the\somehashtable{bar}]
\egroup

[456 = \the\somehashtable{foo}]
[foo = \the\somehashtable{bar}]
```

We define a hashed table, one with keys and values. The definition is global. We can then assign values to this table where the assignments themselves are hash tables. The above code generates:

[123 = 123] [foo = foo]

[456 = 456] [oof = oof]

[123 = 456] [foo = oof]

[456 = 456] [oof = oof]

[456 = 456] [foo = oof]

As you can see, the \global prefix makes the value persistent outside the group. You can mix local and global assignments.

Instead of a hashed table, you can have an indexed table. This time the keys are numbers.

```
\newindexedtable\someindextable

\someindextable{ 123, "foo" }
[123 = \the\someindextable 1]
[foo = \the\someindextable 2]

\bgroup
\someindextable{ 456, "oof" }
[456 = \the\someindextable 1]
[oof = \the\someindextable 2]
\egroup

[123 = \the\someindextable 1]
[foo = \the\someindextable 2]

\bgroup
\global\someindextable{ [1] = 456 }
       \someindextable{ [2] = "oof" }
[456 = \the\someindextable 1]
[oof = \the\someindextable 2]
\egroup

[456 = \the\someindextable 1]
[foo = \the\someindextable 2]
```

The outcome is the same as before:

[123 = 123] [foo = foo]

[456 = 456] [oof = oof]

[123 = 456] [foo = oof]

[456 = 456] [oof = oof]

[456 = 456] [foo = oof]

At the LUA end you can access these tables too:

```
\startluacode
context("[hashed : bar = %s]",context.hashedtables.somehashtable.bar)
context("[indexed : 2 = %s]", context.indexedtables.someindextable[2])
\stopluacode
```

Indeed we get:

[hashed : bar = oof][indexed : 2 = oof]

## 13.3 Data tables

In LUA, tables can be more complex than in the previous section. When you need more complex tables, it is likely that your grouping needs are also different, which is why we have another mechanism. This mechanism is build on top of another model: data values. There are 64K integer registers in any modern TEX engine and normally that is more than enough. However, in addition in LUAMETATEX we have a variant integer storage unit, one that is lightweight and where the amount is limited by the size of the hash table. It was added as part of the low level cleanup up of the LUA token interface (a bit more abstraction). Here is an example of its usage:

```
\dorecurse {100} {
    \setdatavalue{#1}{#1}
}

\start \tttf \darkred \raggedright \dorecurse {100} {
    #1=\scratchcounter\getdatavalue{#1}\the\scratchcounter
} \par \stop \blank

\start \tttf \darkgreen \raggedright \dorecurse {100} {
    #1=\thedatavalue{#1}%
} \par \stop \blank
```

```
1=1 2=2 3=3 4=4 5=5 6=6 7=7 8=8 9=9 10=10 11=11 12=12 13=13 14=14 15=15 16=16
17=17 18=18 19=19 20=20 21=21 22=22 23=23 24=24 25=25 26=26 27=27 28=28 29=29
30=30 31=31 32=32 33=33 34=34 35=35 36=36 37=37 38=38 39=39 40=40 41=41 42=42
43=43 44=44 45=45 46=46 47=47 48=48 49=49 50=50 51=51 52=52 53=53 54=54 55=55
56=56 57=57 58=58 59=59 60=60 61=61 62=62 63=63 64=64 65=65 66=66 67=67 68=68
69=69 70=70 71=71 72=72 73=73 74=74 75=75 76=76 77=77 78=78 79=79 80=80 81=81
82=82 83=83 84=84 85=85 86=86 87=87 88=88 89=89 90=90 91=91 92=92 93=93 94=94
95=95 96=96 97=97 98=98 99=99 100=100

1=1 2=2 3=3 4=4 5=5 6=6 7=7 8=8 9=9 10=10 11=11 12=12 13=13 14=14 15=15 16=16
17=17 18=18 19=19 20=20 21=21 22=22 23=23 24=24 25=25 26=26 27=27 28=28 29=29
30=30 31=31 32=32 33=33 34=34 35=35 36=36 37=37 38=38 39=39 40=40 41=41 42=42
43=43 44=44 45=45 46=46 47=47 48=48 49=49 50=50 51=51 52=52 53=53 54=54 55=55
56=56 57=57 58=58 59=59 60=60 61=61 62=62 63=63 64=64 65=65 66=66 67=67 68=68
69=69 70=70 71=71 72=72 73=73 74=74 75=75 76=76 77=77 78=78 79=79 80=80 81=81
82=82 83=83 84=84 85=85 86=86 87=87 88=88 89=89 90=90 91=91 92=92 93=93 94=94
95=95 96=96 97=97 98=98 99=99 100=100
```

We define hundred values which are a simple numeric macros. Here we use two auxiliary macros because we prefer to use a dedicated namespace for these variables. However, there are also primitives that deal with these data values:

```
\setdatavalue{my-data}{12345}%
\integerdef   \MyDataI 67890
```

```
\dimensiondef \MyDataD 12345pt
\thedatavalue{my-data}, \the\MyDataI, \the\MyDataD.
```

gives: 12345, 67890, 12345.0pt.

But, when more is needed than simple integers, tables come into view. This interface is different from the simple one and involves more commands. The next examples show about all:

```
\newluatable\testtable
\setluatable\testtable{ foo = 123, bar = "456", oof = "rab" }
% \inspectluatable\testtable
\darkcyan
foo = \getfromluatable\testtable{foo}\par
bar = \getfromluatable\testtable{bar}\par
oof = \getfromluatable\testtable{oof}\par
\bgroup
    \useluatable\testtable
    \setluatable\testtable{ foo = 123123, bar = "456456" }
  % \inspectluatable\testtable
    \darkmagenta
    foo = \getfromluatable\testtable{foo}\par
    bar = \getfromluatable\testtable{bar}\par
    oof = \getfromluatable\testtable{oof}\par
    \startluacode
        local t = context.luatables.get("testtable")
        context("<%s %s %s>",t.foo,t.bar,t.oof)
    \stopluacode \par
\egroup
\darkyellow
foo = \getfromluatable\testtable{foo}\par
bar = \getfromluatable\testtable{bar}\par
oof = \getfromluatable\testtable{oof}\par
\startluacode
    local t = context.luatables.get("testtable")
    context("<%s %s %s>",t.foo,t.bar,t.oof)
\stopluacode \par
% \inspectluatable\testtable
```

The tables are semi-local: `\newluatable` creates a table and `\useluatable` will create a local copy that is discarded when the group ends.

```
foo = 123
bar = 456
oof = rab
foo = 123123
bar = 456456
oof = rab
<123123 456456 rab>
foo = 123
bar = 456
```

```
oof = rab
<123 456 rab>
```

Hashed and indexed tables can be used mixed, but there are additional accessors for indexed tables because there we expect numbers.

```
\newluatable\moretable
\setluatable\moretable{ 1, "foo" }
\darkcyan
[1] = \getfromluatable\moretable{1}\par
[2] = \idxfromluatable\moretable 2 \par
\bgroup
    \useluatable\moretable
    \setluatable\moretable{ foo = 123123, bar = "456456" }
    \darkmagenta
    [1] = \getfromluatable\moretable{1}\par
    [2] = \idxfromluatable\moretable 2 \par
    \startluacode
        local t = context.luatables.get("moretable")
        context("<%s %s>",t[1],t[2])
    \stopluacode \par
\egroup
\darkyellow
[1] = \getfromluatable\moretable{1}\par
[2] = \idxfromluatable\moretable 2 \par
\startluacode
    local t = context.luatables.get("moretable")
    context("<%s %s>",t[1],t[2])
\stopluacode \par
```

```
[1] = 1
[2] = foo
[1] = 1
[2] = foo
<1 foo>
[1] = 1
[2] = foo
<1 foo>
```

You can create more complex (nested) tables that you can handle at the LUA end in the usual way. Basically any LUA that makes a table can go between the curly braces.

## 13.4 Named variables

Just because it can be done, and maybe it even has it use, we offer additional numbers, stored and accessible by name. The different types all live in their own namespace:

```
\luainteger bar 123456
\luafloat   bar 123.456e12
```

```
\luainteger gnu = 0xFFFF

{\darkcyan   \the\luainteger bar}
{\darkmagenta\the\luafloat   bar}
{\darkyellow \the\luainteger gnu}
```

These serialize like: 123456 123456000000000 65535, and when not set they default to zero. There is an extra type, cardinal:

```
\luainteger  a  0x7FFFFFFFFFFFFFFF
\luainteger  b -0x7FFFFFFFFFFFFFFF
\luacardinal c  0xFFFFFFFFFFFFFFFF
```

x7FFFFFFFFFFFFFFF

We cheat a bit behind the screens because it is actually a double but we scan it as positive integer and serialize it as such.

```
[\the \luainteger  a\relax]\par
[\the \luainteger  b\relax]\par
[\the \luacardinal c\relax]\par
```

[9223372036854775807]

[0]

[-1]

You have access to the numbers at the LUA end, as in:

```
\luainteger one 123 \luafloat two 456.678
\luaexpr{interfaces.numbers.one/1000 + interfaces.numbers.two/10000}
```

There are `integers`, `cardinals` and `floats` but the `numbers` one is the most generic as it tries to resolve it from one of these namespaces, so we get: [ 0.1686678 ]. There is also a related expression command:

```
(?: \luaexpression          {n.one/1000 + n.two/10000})
(f: \luaexpression float    {n.one/1000 + n.two/10000})
(i: \luaexpression integer  {n.one/1000 + n.two/10000})
(c: \luaexpression cardinal {n.one/1000 + n.two/10000})
(l: \luaexpression lua      {n.one/1000 + n.two/10000})
```

It typesets this (watch the `lua` variant, which is a precise roundtrip serialization method):

```
(?: 0.1686678)
(f: 0.16866780000000000661053434214409207925200462341308593750)
(i: 0)
(c: 0)
(l: 0x1.596e80e71b2cbp-3)
```

The `n` table is just a shortcut to `interfaces.numbers`.

# 14 Callbacks

## 14.1 Introduction

The LuaTeX engine provides the usual basic TeX functionality plus a bit more. It is a deliberate choice not to extend the core engine too much. Instead all relevant processes can be overloaded by new functionality written in Lua. In ConTeXt callbacks are wrapped in a protective layer: on the one hand there is extra functionality (usually interfaced through macros) and on the other hand users can pop in their own handlers using hooks. Of course a plugged in function has to do the right thing and not mess up the data structures. In this chapter the layer on top of callbacks is described.

## 14.2 Actions

Nearly all callbacks in LuaTeX are used in ConTeXt. In the following list the callbacks tagged with `enabled` are used and frozen, the ones tagged `disabled` are blocked and never used, while the ones tagged `undefined` are yet unused.

| id | name | [state] comment |
|---|---|---|
| 1 | test_only | |
| 2 | find_log_file | [set frozen touched fundamental] provide the log file name |
| 3 | find_format_file | [set frozen touched fundamental] locate the format file |
| 4 | open_data_file | [set frozen touched fundamental] open the given file for reading |
| 5 | process_jobname | [frozen touched fundamental] manipulate jobname |
| 6 | start_run | [set frozen touched fundamental] actions performed at the beginning of a run |
| 7 | stop_run | [set frozen touched fundamental] actions performed at the end of a run |
| 8 | define_font | [set frozen touched fundamental] define and/or load a font |
| 9 | quality_font | [set frozen touched selective] initialize expansion and protrusion |
| 10 | pre_output | [set frozen touched selective] preparing output box |
| 11 | buildpage | [set frozen touched] vertical spacing etc (mvl) |
| 12 | hpack | [disabled frozen touched selective] hlist processing (not used, replaced) |
| 13 | vpack | [set frozen touched selective] vertical spacing etc |
| 14 | hyphenate | [disabled frozen touched selective] normal hyphenation routine, called elsewhere |
| 15 | ligaturing | [disabled frozen touched selective] normal ligaturing routine, called elsewhere |

```
16  kerning               [disabled frozen touched selective]
                          normal kerning routine, called elsewhere
17  glyph_run             [set frozen touched selective]
                          glyph processing
18  pre_linebreak         [set frozen touched]
                          horizontal manipulations (before par break)
19  linebreak             [set disabled frozen touched selective]
                          breaking paragraps into lines
20  post_linebreak        [set frozen touched]
                          horizontal manipulations (after par break)
21  append_to_vlist       [disabled frozen touched selective]
                          vlist processing (not used, replaced)
22  alignment             [set frozen touched selective]
                          things done with alignments
23  local_box             [set frozen touched selective]
                          process local boxes
24  packed_vbox           [set frozen touched selective]
                          packed vbox treatments
25  mlist_to_hlist        [set frozen touched selective]
                          convert a noad list into a node list
26  pre_dump              [set frozen touched fundamental]
                          lua related finalizers called before we dump the format
27  start_file            [set frozen touched fundamental]
                          report opening of a file
28  stop_file             [set frozen touched fundamental]
                          report closing of a file
29  intercept_tex_error   [set frozen touched tracing]
                          intercept_tex_error
30  intercept_lua_error   [set frozen touched tracing]
                          intercept_lua_error
31  show_error_message    [disabled frozen touched tracing]
                          show_error_message
32  show_warning_message  [set frozen touched tracing]
                          show_warning_message
33  hpack_quality         [set frozen touched tracing]
                          report horizontal packing quality
34  vpack_quality         [set frozen touched tracing]
                          report vertical packing quality
35  linebreak_check       [set touched tracing]
                          linebreak_check
36  balance_check         [set touched tracing]
                          balance_check
37  show_vsplit           [tracing]
38  show_build            [tracing]
39  insert_par            [set disabled frozen touched]
                          after paragraph start
40  append_adjust         [set frozen touched selective]
                          things done with vertical adjusts
```

```
41  append_migrate        [set disabled frozen touched selective]
                          things done with migrated material
42  append_line           [set disabled frozen touched selective]
                          things done with lines
43  insert_distance       [set frozen touched]
                          check spacing around inserts
44  wrapup_run            [set frozen touched fundamental]
                          actions performed after closing files
45  begin_paragraph       [set disabled frozen touched]
                          before paragraph start
46  paragraph_context     [set disabled frozen touched]
                          when the context is dealt with
47  math_rule
48  make_extensible       [set frozen touched]
                          construct an extensible glyph
49  register_extensible   [set frozen touched]
                          register math extensible construct
50  show_whatsit          [set frozen touched tracing]
                          provide whatsit details
51  get_attribute         [set frozen touched tracing]
                          provide verbose attribute name
52  get_noad_class        [set frozen touched tracing]
                          provide math class name
53  get_math_dictionary   [set frozen touched tracing]
                          provide math dictionary details
54  show_lua_call         [set frozen touched tracing]
                          provide lua call details
55  trace_memory          [set frozen touched tracing]
56  handle_overload       [set frozen touched tracing]
                          handle primitive and macro overload protection
57  missing_character     [set frozen touched tracing]
                          report details about a missing character
58  process_character     [set frozen touched selective]
                          apply an action to a character in a font
59  linebreak_quality     [tracing]
60  paragraph_pass        [set touched selective]
                          paragraph_pass
61  handle_uleader        [set frozen touched selective]
                          resolve user leaders
62  handle_uinsert        [set frozen touched selective]
                          process user inserts in balancer
63  italic_correction     [set frozen touched]
                          insert italic correction
64  show_loners           [tracing]
65  tail_append           [set frozen touched selective]
                          process tail related action
66  balance_boundary      [set frozen touched]
                          process balance specific boundary
```

```
67  balance_insert          [set frozen touched]
                            collect inserts from balance slot
```

Eventually all callbacks will be used so don't rely on undefined callbacks not being protected. Some callbacks are only set when certain functionality is enabled.

It may sound somewhat harsh but if users kick in their own code, we cannot guarantee CONTEXT's behaviour any more and support becomes a pain. If you really need to use a callback yourself, you should use one of the hooks and make sure that you return the right values.

All callbacks related to file handling, font definition and housekeeping are frozen and cannot be overloaded. A reason for this are that we need some kind of protection against misuse. Another reason is that we operate in a well defined environment, the so called TEX directory structure, and we don't want to mess with that. And of course, the overloading permits CONTEXT to provide extensions beyond regular engine functionality.

So as a fact we only open up some of the node list related callbacks and these are grouped as follows:

| category | callback | usage |
|---|---|---|
| processors | pre_linebreak_filter | called just before the paragraph is broken into lines |
| | hpack_filter | called just before a horizontal box is constructed |
| finalizers | post_linebreak_filter | called just after the paragraph has been broken into lines |
| shipouts | no callback yet | applied to the box (or xform) that is to be shipped out |
| mvlbuilders | buildpage_filter | called after some material has been added to the main vertical list |
| vboxbuilders | vpack_filter | called when some material is added to a vertical box |
| math | mlist_to_hlist | called just after the math list is created, before it is turned into an horizontal list |

Each category has several subcategories but for users only two make sense: `before` and `after`. Say that you want to hook some tracing into the `mvlbuilder`. This is how it's done:

```
function third.mymodule.myfunction(where)
    nodes.show_simple_list(tex.lists.contrib_head)
end

nodes.tasks.appendaction("processors", "before", "third.mymodule.myfunction")
```

As you can see, in this case the function gets no `head` passed (at least not currently). This example also assumes that you know how to access the right items. The arguments and return values are given below.[10]

| category | arguments | return value |
|---|---|---|
| processors | head, ... | head, done |
| finalizers | head, ... | head, done |

[10] This interface might change a bit in future versions of CONTEXT. Therefore we will not discuss the few more optional arguments that are possible.

| | | |
|---|---|---|
| shipouts | head | head, done |
| mvlbuilders | | done |
| vboxbuilders | head, ... | head, done |
| parbuilders | head, ... | head, done |
| pagebuilders | head, ... | head, done |
| math | head, ... | head, done |

## 14.3 Tasks

In the previous section we already saw that the actions are in fact tasks and that we can append (and therefore also prepend) to a list of tasks. The before and after task lists are valid hooks for users contrary to the other tasks that can make up an action. However, the task builder is generic enough for users to be used for individual tasks that are plugged into the user hooks.

Of course at some point, too many nested tasks bring a performance penalty with them. At the end of a run MkIV reports some statistics and timings and these can give you an idea how much time is spent in Lua.

The following tables list all the registered tasks for the processors actions:

| category | function |
|---|---|
| before | unset |
| normalizers | languages.transliteration.handler |
| | builders.kernel.collapsing |
| | typesetters.periodkerns.handler |
| | languages.replacements.handler |
| | typesetters.characters.handler |
| | fonts.collections.process |
| | fonts.checkers.missing |
| | userdata.processmystuff |
| characters | scripts.autofontfeature.handler |
| | scripts.splitters.handler |
| | typesetters.cleaners.handler |
| | typesetters.directions.handler |
| | typesetters.cases.handler |
| | typesetters.breakpoints.handler |
| | scripts.injectors.handler |
| words | typesetters.compare.handler |
| | languages.words.check |
| | languages.hyphenators.handler |
| | typesetters.initials.handler |
| | typesetters.firstlines.handler |
| fonts | builders.paragraphs.solutions.splitters.split |
| | nodes.handlers.characters |
| | nodes.injections.handler |
| | typesetters.fontkerns.handler |
| | nodes.handlers.protectglyphs |

| | |
|---|---|
| | builders.kernel.ligaturing |
| | builders.kernel.kerning |
| | nodes.handlers.show |
| | nodes.handlers.stripping |
| | nodes.handlers.flatten |
| | fonts.goodies.colorschemes.coloring |
| lists | typesetters.rubies.check |
| | typesetters.characteralign.handler |
| | typesetters.spacings.handler |
| | typesetters.kerns.handler |
| | typesetters.tighten.handler |
| | typesetters.digits.handler |
| | typesetters.corrections.handler |
| | typesetters.italics.handler |
| | languages.visualizediscretionaries |
| after | typesetters.marksuspects |
| | userdata.processmystuff |

Some of these do have subtasks and some of these even more, so you can imagine that quite some action is going on there.

The finalizer tasks are:

| **category** | **function** |
|---|---|
| before | unset |
| normalizers | unset |
| fonts | builders.paragraphs.solutions.splitters.optimize |
| lists | nodes.handlers.showhyphenation |
| | nodes.handlers.visualizehyphenation |
| | typesetters.margins.localhandler |
| | builders.paragraphs.keeptogether |
| | builders.paragraphs.tag |
| | nodes.linefillers.handler |
| after | unset |

Shipouts concern:

| **category** | **function** |
|---|---|
| before | unset |
| normalizers | nodes.handlers.cleanuppage |
| | typesetters.swapping.handler |
| | typesetters.stacking.handler |
| | typesetters.showsuspects |
| | typesetters.showinvisibles |
| | typesetters.margins.finalhandler |
| | builders.paragraphs.expansion.trace |
| | typesetters.alignments.handler |

| | |
|---|---|
| | nodes.references.handler |
| | nodes.destinations.handler |
| | nodes.rules.handler |
| | nodes.shifts.handler |
| | nodes.shadows.handler |
| | nodes.handlers.backgrounds |
| | typesetters.rubies.attach |
| | nodes.tracers.directions |
| finishers | nodes.visualizers.handler |
| | attributes.colors.handler |
| | attributes.transparencies.handler |
| | attributes.colorintents.handler |
| | attributes.negatives.handler |
| | attributes.effects.handler |
| | attributes.alternates.handler |
| | attributes.viewerlayers.handler |
| after | unset |
| wrapup | structures.tags.handler |
| | nodes.handlers.export |
| | luatex.synctex.collect |

There are not that many mvlbuilder tasks currently:

| **category** | **function** |
|---|---|
| before | unset |
| normalizers | streams.collect |
| | nodes.handlers.backgroundspage |
| | typesetters.margins.globalhandler |
| | nodes.handlers.migrate |
| | builders.vspacing.pagehandler |
| | builders.profiling.pagehandler |
| | builders.profiling.pagelinehandler |
| | typesetters.checkers.handler |
| | typesetters.synchronize.handler |
| after | unset |

The vboxbuilder perform similar tasks:

| **category** | **function** |
|---|---|
| before | unset |
| normalizers | nodes.handlers.backgroundsvbox |
| | builders.vspacing.vboxhandler |
| | builders.profiling.boxlinehandler |
| | typesetters.checkers.handler |
| | typesetters.synchronize.handler |
| after | unset |

In the future we expect to have more parbuilder tasks. Here again there are subtasks that depend on the current typesetting environment, so this is the right spot for language specific treatments.

The following actions are applied just before the list is passed on the the output routine. The return value is a vlist.

| category | function |
|---|---|
| before | unset |
| normalizers | unset |
| after | unset |

*Both the parbuilders and pagebuilder tasks are unofficial and not yet meant for users.*

Finally, we have tasks related to the math list:

| category | function |
|---|---|
| before | unset |
| normalizers | noads.handlers.showtree<br>noads.handlers.numbers<br>noads.handlers.spacing<br>noads.handlers.fencing<br>noads.handlers.unscript<br>noads.handlers.unstack<br>noads.handlers.relocate<br>noads.handlers.variants<br>noads.handlers.families<br>noads.handlers.render<br>noads.handlers.collapse<br>noads.handlers.autofences<br>noads.handlers.alternates<br>noads.handlers.intervals<br>noads.handlers.tags<br>noads.handlers.kernpairs<br>noads.handlers.classes<br>noads.handlers.dictionaries<br>noads.handlers.suspicious |
| builders | builders.kernel.mlisttohlist<br>noads.handlers.makeup |
| finalizers | noads.handlers.export<br>noads.handlers.normalize<br>noads.handlers.snap |
| after | unset |

As MkIV is developed in sync with LuaTeX and code changes from experimental to more final and reverse, you should not be too surprised if the registered function names change.

You can create your own task list with:

```
nodes.tasks.new("mytasks",{ "one", "two" })
```

After that you can register functions. You can append as well as prepend them either or not at a specific position.

```
nodes.tasks.appendaction ("mytask","one","bla.alpha")
nodes.tasks.appendaction ("mytask","one","bla.beta")

nodes.tasks.prependaction("mytask","two","bla.gamma")
nodes.tasks.prependaction("mytask","two","bla.delta")

nodes.tasks.appendaction ("mytask","one","bla.whatever","bla.alpha")
```

Functions can also be removed:

```
nodes.tasks.removeaction("mytask","one","bla.whatever")
```

As removal is somewhat drastic, it is also possible to enable and disable functions. From the fact that with these two functions you don't specify a category (like `one` or `two`) you can conclude that the function names need to be unique within the task list or else all with the same name within this task will be disabled.

```
nodes.tasks.enableaction ("mytask","bla.whatever")
nodes.tasks.disableaction("mytask","bla.whatever")
```

The same can be done with a complete category:

```
nodes.tasks.enablegroup ("mytask","one")
nodes.tasks.disablegroup("mytask","one")
```

There is one function left:

```
nodes.tasks.actions("mytask",2)
```

This function returns a function that when called will perform the tasks. In this case the function takes two extra arguments in addition to `head`.[11]

Tasks themselves are implemented on top of sequences but we won't discuss them here.

## 14.4 Paragraph and page builders

Building paragraphs and pages is implemented differently and has no user hooks. There is a mechanism for plugins but the interface is quite experimental.

## 14.5 Some examples

*todo*

[11] Specifying this number permits for some optimization but is not really needed

# 15 Backend code

## 15.1 Introduction

In CONTEXT we've always separated the backend code in so called driver files. This means that in the code related to typesetting only calls to the API take place, and no backend specific code is to be used. Currently a PDF backend is supported as well as an XML export.[12]

Some CONTEXT users like to add their own PDF specific code to their styles or modules. However, such extensions can interfere with existing code, especially when resources are involved. Therefore the construction of PDF data structures and resources is rather controlled and has to be done via the official helper macros.

## 15.2 Structure

A PDF file is a tree of indirect objects. Each object has a number and the file contains a table (or multiple tables) that relates these numbers to positions in a file (or position in a compressed object stream). That way a file can be viewed without reading all data: a viewer only loads what is needed.

```
1 0 obj <<
    /Name (test) /Address 2 0 R
>>
2 0 obj [
   (Main Street) (24) (postal code) (MyPlace)
]
```

For the sake of the discussion we consider strings like `(test)` also to be objects. In the next table we list what we can encounter in a PDF file. There can be indirect objects in which case a reference is used (`2 0 R`) and direct ones.

It all starts in the document's root object. From there we access the page tree and resources. Each page carries its own resource information which makes random access easier. A page has a page stream and there we find the to be rendered content as a mixture of (UNICODE) strings and special drawing and rendering operators. Here we will not discuss them as they are mostly generated by the engine itself or dedicated subsystems like the METAPOST converter. There we use literal or `\latelua` whatsits to inject code into the current stream.

## 15.3 Data types

There are several datatypes in PDF and we support all of them one way or the other.

| type | form | meaning |
|---|---|---|
| constant | `/...` | A symbol (prescribed string). |
| string | `(...)` | A sequence of characters in pdfdoc encoding |
| unicode | `<...>` | A sequence of characters in utf16 encoding |

[12] This chapter is derived from an article on these matters. You can find nore information in `hybrid.pdf`.

| | | |
|---|---|---|
| number | `3.1415` | A number constant. |
| boolean | `true/false` | A boolean constant. |
| reference | `N 0 R` | A reference to an object |
| dictionary | `<< ... >>` | A collection of key value pairs where the value itself is an (indirect) object. |
| array | `[ ... ]` | A list of objects or references to objects. |
| stream | | A sequence of bytes either or not packaged with a dictionary that contains descriptive data. |
| xform | | A special kind of object containing an reusable blob of data, for example an image. |

While writing additional backend code, we mostly create dictionaries.

```
<< /Name (test) /Address 2 0 R >>
```

In this case the indirect object can look like:

```
[ (Main Street) (24) (postal code) (MyPlace) ]
```

The LuaTeX manual mentions primitives like `\pdfobj`, `\pdfannot`, `\pdfcatalog`, etc. However, in MkIV no such primitives are used. You can still use many of them but those that push data into document or page related resources are overloaded to do nothing at all.

In the Lua backend code you will find function calls like:

```
local d = lpdf.dictionary {
    Name    = lpdf.string("test"),
    Address = lpdf.array {
        "Main Street", "24", "postal code", "MyPlace",
    }
}
```

Equaly valid is:

```
local d = lpdf.dictionary()
d.Name = "test"
```

Eventually the object will end up in the file using calls like:

```
local r = lpdf.immediateobject(tostring(d))
```

or using the wrapper (which permits tracing):

```
local r = lpdf.flushobject(d)
```

The object content will be serialized according to the formal specification so the proper `<< >>` etc. are added. If you want the content instead you can use a function call:

```
local dict = d()
```

An example of using references is:

```
local a = lpdf.array {
```

```
    "Main Street", "24", "postal code", "MyPlace",
}
local d = lpdf.dictionary {
    Name    = lpdf.string("test"),
    Address = lpdf.reference(a),
}
local r = lpdf.flushobject(d)
```

We have the following creators. Their arguments are optional.

| function | optional parameter |
|---|---|
| lpdf.null | |
| lpdf.number | number |
| lpdf.constant | string |
| lpdf.string | string |
| lpdf.unicode | string |
| lpdf.boolean | boolean |
| lpdf.array | indexed table of objects |
| lpdf.dictionary | hash with key/values |
| lpdf.reference | string |
| lpdf.verbose | indexed table of strings |

```
tostring(lpdf.null())
```

```
null
```

```
tostring(lpdf.number(123))
```

```
123
```

```
tostring(lpdf.constant("whatever"))
```

```
/whatever
```

```
tostring(lpdf.string("just a string"))
```

```
(just a string)
```

```
tostring(lpdf.unicode("just a string"))
```

```
(just a string)
```

```
tostring(lpdf.boolean(true))
```

```
true
```

```
tostring(lpdf.array { 1, lpdf.constant("c"), true, "str" })
```

```
[ 1 /c true (str) ]
```

```
tostring(lpdf.dictionary { a=1, b=lpdf.constant("c"), d=true, e="str" })
```

```
<< /a 1 /b /c /d true /e (str) >>
```

```
tostring(lpdf.reference(123))
```

```
123 0 R
```

```
tostring(lpdf.verbose("whatever"))
```

```
whatever
```

## 15.4 Managing objects

Flushing objects is done with:

```
lpdf.flushobject(obj)
```

Reserving object is or course possible and done with:

```
local r = lpdf.reserveobject()
```

Such an object is flushed with:

```
lpdf.flushobject(r,obj)
```

We also support named objects:

```
lpdf.reserveobject("myobject")

lpdf.flushobject("myobject",obj)
```

A delayed object is created with:

```
local ref = pdf.delayedobject(data)
```

The data will be flushed later using the object number that is returned (`ref`). When you expect that many object with the same content are used, you can use:

```
local obj = lpdf.shareobject(data)
local ref = lpdf.shareobjectreference(data)
```

This one flushes the object and returns the object number. Already defined objects are reused. In addition to this code driven optimization, some other optimization and reuse takes place but all that happens without user intervention. Only use this when it's really needed as it might consume more memory and needs more processing time.

## 15.5 Resources

While LuaTeX itself will embed all resources related to regular typesetting, MkIV has to take care of embedding those related to special tricks, like annotations, spot colors, layers, shades, transparencies, metadata, etc. Because third party modules (like tikz) also can add resources we provide some macros that makes sure that no interference takes place:

```
\pdfbackendsetcatalog       {key}{string}
\pdfbackendsetinfo          {key}{string}
```

```
\pdfbackendsetname          {key}{string}

\pdfbackendsetpageattribute {key}{string}
\pdfbackendsetpagesattribute{key}{string}
\pdfbackendsetpageresource  {key}{string}

\pdfbackendsetextgstate     {key}{pdfdata}
\pdfbackendsetcolorspace    {key}{pdfdata}
\pdfbackendsetpattern       {key}{pdfdata}
\pdfbackendsetshade         {key}{pdfdata}
```

One is free to use the LUA interface instead, as there one has more possibilities but when code is shared with other macro packages the macro interface makes more sense. The names of the LUA functions are similar, like:

```
lpdf.addtoinfo(key,anything_valid_pdf)
```

Currently we expose a bit more of the backend code than we like and future versions will have a more restricted access. The following function will stay public:

```
lpdf.addtopageresources  (key,value)
lpdf.addtopageattributes (key,value)
lpdf.addtopagesattributes(key,value)

lpdf.adddocumentextgstate(key,value)
lpdf.adddocumentcolorspac(key,value)
lpdf.adddocumentpattern  (key,value)
lpdf.adddocumentshade    (key,value)

lpdf.addtocatalog        (key,value)
lpdf.addtoinfo           (key,value)
lpdf.addtonames          (key,value)
```

## 15.6 Annotations

You can use the LUA functions that relate to annotations etc. but normally you will use the regular CONTEXT user interface. You can look into some of the `lpdf-*` modules to see how special annotations can be dealt with.

## 15.7 Tracing

There are several tracing options built in and some more will be added in due time:

```
\enabletrackers
  [backend.finalizers,
   backend.resources,
   backend.objects,
   backend.detail]
```

As with all trackers you can also pass them on the command line, for example:

```
context --trackers=backend.* yourfile
```

The reference related backend mechanisms have their own trackers. When you write code that generates PDF, it also helps to look in the PDF file so see if things are done right. In that case you need to disable compression:

```
\nopdfcompression
```

## 15.8 Analyzing

The epdf library that comes with LuaTEX offers a userdata interface to PDF files. On top of that ConTEXt provides a more Lua-ish access, using tables. You can open a PDF file with:

```
local mypdf = lpdf.epdf.load(filename)
```

When opening is successful, you have access to a couple of tables:

```
\NC \type{pages}         \NC indexed \NC \NR
\NC \type{destinations}  \NC hashed  \NC \NR
\NC \type{javascripts}   \NC hashed  \NC \NR
\NC \type{widgets}       \NC hashed  \NC \NR
\NC \type{embeddedfiles} \NC hashed  \NC \NR
\NC \type{layers}        \NC indexed \NC \NR
```

These provide efficient access to some data that otherwise would take a bit of code to deal with. Another top level table is the for PDF characteristic `Catalog`. Watch the capitalization: as with other native PDF data structures, keys are case sensitive and match the standard.

Here is an example of usage:

```
local MyDocument = lpdf.epdf.load("somefile.pdf")

context.starttext()

  local pages    = MyDocument.pages
  local nofpages = pages.n

  context.starttabulate { "|c|c|c|" }

    context.NC() context("page")
    context.NC() context("width")
    context.NC() context("height") context.NR()

    for i=1, nofpages do
      local page = pages[i]
      local bbox = page.CropBox or page.MediaBox
      context.NC() context(i)
      context.NC() context(bbox[4]-bbox[2])
      context.NC() context(bbox[3]-bbox[1]) context.NR()
    end

  context.stoptabulate()

context.stoptext()
```

# 16 Font goodies

## 16.1 Introduction

One of the interesting aspects of TEX is that it provides control over fonts and LUATEX provides quite some. In CONTEXT we support basic functionality, like OPENTYPE features, as well as some extra functionality. We also have a mechanism for making virtual fonts which is mostly used for the transition from TYPE1 math fonts to OPENTYPE math fonts. Instead of hard coding specific details in the core LUA code, we use so called LUA Font Goodies to control them. These goodies are collected in tables and live in files. When a font is loaded, one or more such goodie files can be loaded alongside.

In the following typescript we load a goodies file that defines a virtual Lucida math font. The goodie file is loaded immediately and some information in the table is turned into a form that permits access later on: the virtual font id `lucida-math` that is used as part of the font specification.

```
\starttypescript [math] [lucida]
    \loadfontgoodies[lucida-math]
    \definefontsynonym[MathRoman][lucidamath@lucida-math]
\stoptypescript
```

Not all information is to be used directly. Some can be accessed when needed. In the following case the file `dingbats.lfg` gets loaded (only once) when the font is actually used. In that file, there is information that is used by the `unicoding` feature.

```
\definefontfeature
  [dingbats]
  [mode=base,
   goodies=dingbats,
   unicoding=yes]

\definefont[dingbats][file:dingbats][features=dingbats]
```

In the following sections some aspects of goodies are discussed. We don't go into details of what these goodies are, but just stick to the LUA side of the specification.

## 16.2 Virtual math fonts

A virtual font is defined using the `virtuals` entry in the `mathematics` subtable. As TYPE1 fonts are used, an additional table `mapfiles` is needed to specify the files that map filenames onto real files.

```
return {
  name = "px-math",
  version = "1.00",
  comment = "Goodies that complement px math.",
  author = "Hans Hagen",
  copyright = "ConTeXt development team",
  mathematics = {
    mapfiles = {
        "mkiv-px.map",
    },
```

```
    virtuals = {
      ["px-math"] = {
        { name = "texgyrepagella-regular.otf", features = "virtualmath", main = true },
        { name = "rpxr.tfm", vector = "tex-mr" } ,
        { name = "rpxmi.tfm", vector = "tex-mi", skewchar=0x7F },
        { name = "rpxpplri.tfm", vector = "tex-it", skewchar=0x7F },
        { name = "pxsy.tfm", vector = "tex-sy", skewchar=0x30, parameters = true } ,
        { name = "pxex.tfm", vector = "tex-ex", extension = true } ,
        { name = "pxsya.tfm", vector = "tex-ma" },
        { name = "pxsyb.tfm", vector = "tex-mb" },
        { name = "texgyrepagella-bold.otf", vector = "tex-bf" } ,
        { name = "texgyrepagella-bolditalic.otf", vector = "tex-bi" } ,
        { name = "lmsans10-regular.otf", vector = "tex-ss", optional=true },
        { name = "lmmono10-regular.otf", vector = "tex-tt", optional=true },
      },
    }
  }
}
```

Here the `px-math` virtual font is defined. A series of fonts is loaded and combined into one. The `vector` entry is used to tell the builder how to map the glyphs onto UNICODE. Additional vectors can be defined, for instance:

```
fonts.encodings.math["mine"] = {
    [0x1234] = 0x56,
}
```

Eventually these specifications wil be replaced by real OPENTYPE fonts, but even then we will keep the virtual definitions around.

## 16.3 Math alternates

In addition to the official `ssty` feature for enforcing usage of script and scriptscript glyphs, some stylistic alternates can be present.

```
return {
  name = "xits-math",
  version = "1.00",
  comment = "Goodies that complement xits (by Khaled Hosny).",
  author = "Hans Hagen",
  copyright = "ConTeXt development team",
  mathematics = {
    alternates = {
      cal       = {
        feature = 'ss01',
        value   = 1,
        comment = "Mathematical Calligraphic Alphabet"
      },
      greekssup = {
        feature = 'ss02',
        value   = 1,
        comment = "Mathematical Greek Sans Serif Alphabet"
      },
      greekssit = {
        feature = 'ss03',
        value   = 1,
```

```
      comment = "Mathematical Italic Sans Serif Digits"
    },
    monobfnum = {
      feature = 'ss04',
      value   = 1,
      comment = "Mathematical Bold Monospace Digits"
    },
    mathbbbf  = {
      feature = 'ss05',
      value   = 1,
      comment = "Mathematical Bold Double-Struck Alphabet"
    },
    mathbbit  = {
      feature = 'ss06',
      value   = 1,
      comment = "Mathematical Italic Double-Struck Alphabet"
    },
    mathbbbi  = {
      feature = 'ss07',
      value   = 1,
      comment = "Mathematical Bold Italic Double-Struck Alphabet"
    },
    upint     = {
      feature = 'ss08',
      value   = 1,
      comment = "Upright Integrals"
    },
  }
 }
}
```

These can be activated (in math mode) with the `\mathalternate` command like:

```
$\mathalternate{cal}Z$
```

## 16.4 Math parameters

Another goodie related to math is the overload of some parameters (part of the font itself) and variables (used in making virtual shapes).

```
return {
  name = "lm-math",
  version = "1.00",
  comment = "Goodies that complement latin modern math.",
  author = "Hans Hagen",
  copyright = "ConTeXt development team",
  mathematics = {
    mapfiles = {
      "lm-math.map",
      "lm-rm.map",
      "mkiv-base.map",
    },
    virtuals = {
      ["lmroman5-math"]      = five,
      ["lmroman6-math"]      = six,
      ["lmroman7-math"]      = seven,
      ["lmroman8-math"]      = eight,
```

```
      ["lmroman9-math"]      = nine,
      ["lmroman10-math"]     = ten,
      ["lmroman10-boldmath"] = ten_bold,
      ["lmroman12-math"]     = twelve,
      ["lmroman17-math"]     = seventeen,
    },
    variables = {
      joinrelfactor = 3, -- default anyway
    },
    parameters = { -- test values
  --  FactorA = 123.456,
  --  FactorB = false,
  --  FactorC = function(value,target,original) return 7.89 * target.factor end,
  --  FactorD = "Hi There!",
    },
  }
}
```

In this example you see several virtuals defined which is due to the fact that Latin Modern has design sizes. The values (like `twelve` are tables defined before the return happens and are not shown here. The variables are rather CONTEXT specific, and the parameters are those that come with regular OPENTYPE math fonts (so the example names are invalid).

In the following example we show two wasy to change parameters. In this case we have a regular OPENTYPE math font. First we install a patch to the font itself. That change will be cached. We could also have changed that parameter using the goodies table. The first method is the oldest.

```
local patches = fonts.handlers.otf.enhancers.patches

local function patch(data,filename,threshold)
  local m = data.metadata.math
  if m then
    local d = m.DisplayOperatorMinHeight or 0
    if d < threshold then
      patches.report("DisplayOperatorMinHeight(%s -> %s)",d,threshold)
      m.DisplayOperatorMinHeight = threshold
    end
  end
end

patches.register(
  "after",
  "check math parameters",
  "asana",
  function(data,filename)
    patch(data,filename,1350)
  end
)

local function less(value,target,original)
  return 0.25 * value
end

return {
  name = "asana-math",
  version = "1.00",
  comment = "Goodies that complement asana.",
  author = "Hans Hagen",
```

```
  copyright = "ConTeXt development team",
  mathematics = {
    parameters = {
      StackBottomDisplayStyleShiftDown = less,
      StackBottomShiftDown             = less,
      StackDisplayStyleGapMin          = less,
      StackGapMin                      = less,
      StackTopDisplayStyleShiftUp      = less,
      StackTopShiftUp                  = less,
      StretchStackBottomShiftDown      = less,
      StretchStackGapAboveMin          = less,
      StretchStackGapBelowMin          = less,
      StretchStackTopShiftUp           = less,
    }
  }
}
```

We use a function so that the scaling is taken into account as the values passed are those resulting from the scaling of the font to the requested size.

## 16.5 Unicoding

We still have to deal with existing TYPE1 fonts, and some of them have an encoding that is hard to map onto UNICODE without additional information. The following goodie does that. The keys in the `unicodes` table are the glyph names. Keep in mind that this only works with simple fonts. The CONTEXT code takes care of kerns but that's about it.

```
return {
  name = "dingbats",
  version = "1.00",
  comment = "Goodies that complement dingbats (funny names).",
  author = "Hans Hagen",
  copyright = "ConTeXt development team",
  remapping = {
    tounicode = true,
    unicodes = {
      a1   = 0x2701,
      a10  = 0x2721,
      a100 = 0x275E,
      a101 = 0x2761,
      ............
      a98  = 0x275C,
      a99  = 0x275D,
    },
  },
}
```

The `tounicode` option makes sure that additional information ends ip in the output so that cut-and-paste becomes more trustworthy.

## 16.6 Typescripts

Some font collections, like antykwa, come with so many variants that defining them all in typescripts becomes somewhat of a nuisance. While a regular font has a typescript of a few lines, antykwa needs way more lines. This is why we provide a nother way as well, using goodies.

```
return {
  name = "antykwapoltawskiego",
  version = "1.00",
  comment = "Goodies that complement Antykwa Poltawskiego",
  author = "Hans & Mojca",
  copyright = "ConTeXt development team",
  files = {
    name = "antykwapoltawskiego", -- shared
    list = {
      ["AntPoltLtCond-Regular.otf"] = {
     -- name   = "antykwapoltawskiego",
        weight = "light",
        style  = "regular",
        width  = "condensed",
      },
      ["AntPoltLtCond-Italic.otf"] = {
        weight = "light",
        style  = "italic",
        width  = "condensed",
      },
      ["AntPoltCond-Regular.otf"] = {
        weight = "normal",
        style  = "regular",
        width  = "condensed",
      },

      .......

      ["AntPoltExpd-BoldItalic.otf"] = {
        weight = "bold",
        style  = "italic",
        width  = "expanded",
      },
    },
  },
  typefaces = { -- for Mojca (experiment, names might change)
    ["antykwapoltawskiego-light"] = {
      shortcut     = "rm",
      shape        = "serif",
      fontname     = "antykwapoltawskiego",
      normalweight = "light",
      boldweight   = "medium",
      width        = "normal",
      size         = "default",
      features     = "default",
    },

    .......

  },
}
```

This is a typical example of when a goodies file is loaded directly:

```
\loadfontgoodies[antykwapoltawskiego]
```

A bodyfont is now defined by choosing from the defined combinations:

```
\definetypeface
```

```
  [name=mojcasfavourite,
   preset=antykwapoltawskiego,
   normalweight=light,
   boldweight=bold,
   width=expanded]

\setupbodyfont
  [mojcasfavourite]
```

This mechanism is a follow up on a discussion at a CONTEXT conference, still somewhat experimental, and a playground for Mojca.

## 16.7 Font strategies

This goodie is closely related to the Oriental TEX project where a dedicated paragraph optimizer can be used. A rather advanced font is used (husayni) and its associated goodie file is rather extensive. It defines stylistic features, implements a couple of feature sets, provides colorschemes and most of all, defines some strategies for making paragraphs look better. Some of the goodie file is shown here.

```
local yes = "yes"

local basics = {
  analyze  = yes,
  mode     = "node",
  language = "dflt",
  script   = "arab",
}

local analysis = {
  ccmp = yes,
  init = yes, medi = yes, fina = yes,
}

local regular = {
  rlig = yes, calt = yes, salt = yes, anum = yes,
  ss01 = yes, ss03 = yes, ss07 = yes, ss10 = yes, ss12 = yes, ss15 = yes, ss16 = yes,
  ss19 = yes, ss24 = yes, ss25 = yes, ss26 = yes, ss27 = yes, ss31 = yes, ss34 = yes,
  ss35 = yes, ss36 = yes, ss37 = yes, ss38 = yes, ss41 = yes, ss42 = yes, ss43 = yes,
  js16 = yes,
}

local positioning = {
  kern = yes, curs = yes, mark = yes, mkmk = yes,
}

local minimal_stretching = {
  js11 = yes, js03 = yes,
}

local medium_stretching = {
  js12=yes, js05=yes,
}

local maximal_stretching= {
  js13 = yes, js05 = yes, js09 = yes,
}
```

```
local wide_all = {
  js11 = yes, js12 = yes, js13 = yes, js05 = yes, js09 = yes,
}

local shrink = {
  flts = yes, js17 = yes, ss05 = yes, ss11 = yes, ss06 = yes, ss09 = yes,
}

local default = {
  basics, analysis, regular, positioning, -- xxxx = yes, yyyy = 2,
}

return {
  name = "husayni",
  version = "1.00",
  comment = "Goodies that complement the Husayni font by Idris Samawi Hamid.",
  author = "Idris Samawi Hamid and Hans Hagen",
  featuresets = { -- here we don't have references to featuresets
    default = {
      default,
    },
    minimal_stretching = {
      default,
      js11 = yes, js03 = yes,
    },
    medium_stretching = {
      default,
      js12=yes, js05=yes,
    },
    maximal_stretching= {
      default,
      js13 = yes, js05 = yes, js09 = yes,
    },
    wide_all = {
      default,
      js11 = yes, js12 = yes, js13 = yes, js05 = yes, js09 = yes,
    },
    shrink = {
      default,
      flts = yes, js17 = yes, ss05 = yes, ss11 = yes, ss06 = yes, ss09 = yes,
    },
  },
  solutions = { -- here we have references to featuresets, so we use strings!
    experimental = {
      less = {
        "shrink"
      },
      more = {
        "minimal_stretching",
        "medium_stretching",
        "maximal_stretching",
        "wide_all"
      },
    },
  },
  stylistics = {
    ......
    ss03 = "level-1 stack over Jiim, initial entry only",
    ss04 = "level-1 stack over Jiim, initial/medial entry",
```

```
    ......
    ss54 = "chopped finals",
    ss55 = "idgham-tanwin",
    ......
    js11 = "level-1 stretching",
    js12 = "level-2 stretching",
    ......
    js21 = "Haa.final_alt2",
  },
  colorschemes = {
    default = {
      [1] = {
        "Onedotabove", "Onedotbelow", ...
      },
      [2] = {
        "Fathah", "Dammah", "Kasrah", ...
      },
      [3] = {
        "Ttaa.waqf", "SsLY.waqf", "QLY.waqf", ...
      },
      [4] = {
        "ZeroArabic.ayah", "OneArabic.ayah", "TwoArabic.ayah", ...
      },
      [5] = {
        "Ayah", "Ayah.alt1", "Ayah.alt2", ...
      }
    }
  }
}
\stopmalltyping

Discussion of these goodies is beyond this document and happens elsewhere.

\stopsection

\startsection[title=Composition]

The \type {compose} features extends a font with additional (virtual) shapes.
This is mostly used with \TYPEONE\ fonts that lack support for eastern european
languages. The type {compositions} subtable is used to control placement of
accents. This can be done per font.

\startmalltyping
local defaultunits = 193 - 30

-- local compose = {
--                DY = defaultunits,
--   [0x010C] = { DY = defaultunits }, -- Ccaron
--   [0x02C7] = { DY = defaultunits }, -- textcaron
-- }

-- fractions relative to delta(X_height - x_height)

local defaultfraction = 0.85

local compose = {
  DY = defaultfraction, -- uppercase compensation
}

return {
```

```
  name = "lucida-one",
  version = "1.00",
  comment = "Goodies that complement lucida.",
  author = "Hans and Mojca",
  copyright = "ConTeXt development team",
  compositions = {
    ["lbr"]  = compose,
    ["lbi"]  = compose,
    ["lbd"]  = compose,
    ["lbdi"] = compose,
  }
}
```

## 16.8 Postprocessing

You can hook postprocessors into the scaler. Future versions might provide more control over where this happens.

```
local function statistics(tfmdata)
    commands.showfontparameters(tfmdata)
end

local function squeeze(tfmdata)
  for k, v in next, tfmdata.characters do
    v.height = 0.75 * (v.height or 0)
    v.depth  = 0.75 * (v.depth  or 0)
  end
end

return {
  name = "demo",
  version = "1.00",
  comment = "An example of goodies.",
  author = "Hans Hagen",
  postprocessors = {
    statistics = statistics,
    squeeze    = squeeze,
  },
}
```

# 17 Nice to know

## 17.1 Introduction

As we like to abstract interfaces it is no surprise that CONTEXT and therefore it's LUA libraries come with all kind of helpers. In this chapter I will explain a few of them. Feel free to remind of adding more here.

## 17.2 Templates

*Eventually we will move this to the utilities section.*

When dealing with data from tables or when order matters it can be handy to abstract the actual data from the way it is dealt with. For this we provide a template mechanism. The following example demonstrate its use.

```
require("util-ran") -- needed for this example

local preamble = [[|l|l|c|]]
local template = [[\NC %initials% \NC %surname% \NC %length% \NC \NR]]

context.starttabulate { preamble }
    for i=1,10 do
        local row = utilities.templates.replace(template, {
            surname  = utilities.randomizers.surname(5,10),
            initials = utilities.randomizers.initials(1,3),
            length   = string.format("%0.2f",math.random(140,195)),
        })
        context(row)
    end
context.stoptabulate()
```

This renders a table with random entries:

| | | |
|---|---|---|
| U. | Zisyx | 174.00 |
| O.Z. | Pyximez | 158.00 |
| A.C.I. | Yxoqym | 173.00 |
| I.J.Y. | Udizyduquq | 160.00 |
| Y.G.I. | Umoficevyx | 161.00 |
| O. | Ibyvyd | 164.00 |
| U. | Mudykus | 178.00 |
| I. | Musex | 182.00 |
| O. | Kozofud | 142.00 |
| O.C. | Esajolyx | 175.00 |

The nice thing is that when we change the order of the columns, we don't need to change the table builder.

```
local preamble = [[|c|l|l|]]
```

```
local template = [[\NC %length% \NC %initials% \NC %surname% \NC \NR]]
```

The `replace` function takes a few more arguments. There are also a some more replacement options.

```
replace("test '%[x]%' test",{ x = [[a 'x'  a]] }))
replace("test '%[x]%' test",{ x = true }))
replace("test '%[x]%' test",{ x = [[a 'x'  a]], y = "oeps" },'sql'))
replace("test '%[x]%' test",{ x = [[a '%y%'  a]], y = "oeps" },'sql',true))
replace([[test %[x]% test]],{ x = [[a "x"  a]]}))
replace([[test %(x)% test]],{ x = [[a "x"  a]]}))
```

The first argument is the template and the second one a table with mappings from keys to values. The third argument can be used to inform the replace mechanism what string escaping has to happen. The last argument triggers recursive replacement. The above calls result in the following strings:

```
test 'a 'x' \127 a' test
test 'true' test
test 'a ''x''  a' test
test 'a ''oeps''  a' test
test a \"x\" \127 a test
test "a \"x\" \127 a" test
```

These examples demonstrate that by adding a pair of square brackets we get escaped strings. When using parenthesis the quotes get added automatically. This is somewhat faster in case when LUA is the target, but in practice it is not that noticeable.

## 17.3 Extending

Instead of extending tex endlessly we can also define our own extensions. Here is an example. When you want to manipulate a box at the LUA end you have the problem that the following will not always work out well:

```
local b = tex.getbox(0)
-- mess around with b
tex.setbox(0,b)
```

So we end up with:

```
local b = node.copylist(tex.getbox(0))
-- mess around with b
tex.setbox(0,b)
```

The reason is that at the TEX end grouping plays a role which means that values are saved and restored. However, there is a save way out by defining a function that cheats a bit:

```
function tex.takebox(id)
    local box = tex.getbox(id)
    if box then
        local copy = node.copy(box)
        local list = box.list
        copy.list = list
```

```
        box.list = nil
        tex.setbox(id,nil)
        return copy
    end
end
```

Now we can say:

```
local b = tex.takebox(0)
-- mess around with b
tex.setbox(b)
```

In this case we first get the box content and then let TEX do some housekeeping. But, because we only keep the list node (which we copied) in the register the overhead of copying a whole list is gone.

# 18 A sort of summary

In this chapter we summarize the functionality provided by the `context` namespace. We repeat some of what has been explained in other chapter so that in fact you can start with this summary.

If you have read this manual (or seen code) you know that you can access all the core commands using this namespace:

```
context.somecommand("some argument")
context["somecommand"]("some argument")
```

These calls will eventually expand `\somecommand` with the given argument. This interface has been around form the start and has proven to be quite flexible and robust. In spite of what you might think, the `somecommand` is not really defined in the `context` namespace, but in its own one called `core`, accessible via `context.core`.

Next we describe the commands that are naturally defined in the `context` namespace. Some have counterparts at the macro level (like `bgroup`) but many haven't (for instance `rule`). We tried not to polute the `context` namespace too much but although we could have put the helpers in a separate namespace it would make usage a bit more unnatural.

## 18.1 Access to commands

### `context(".. some text ..")`

The string is flushed as-is:

```
.. some text ..
```

### `context("format",...)`

The first string is a format specification according that is passed to the LUA function `format` in the `string` namespace. Following arguments are passed too.

### `context(123,...)`

The numbers (and following numbers or strings) are flushed without any formatting.

```
123... (concatenated)
```

### `context(true)`

An explicit `endlinechar` is inserted, in TEX speak:

```
^^M
```

### context(false,...)

Strings and numbers are flushed surrounded by curly braces, an indexed table is flushed as option list, and a hashed table is flushed as parameter set.

```
multiple {...} or [...] etc
```

### context(node)

The node (or list of nodes) is injected at the spot. Keep in mind that you need to do the proper memory management yourself.

### context["command"] context.core["command"]

The function that implements `\command`. The `core` table is where these functions realy live.

### context["command"](value,...)

The value (string or number) is flushed as a curly braced (regular) argument.

```
\command {value}...
```

### context["command"]({ value },...)

The table is flushed as value set. This can be an identifier, a list of options, or a directive.

```
\command [value]...
```

### context["command"]({ key = val },...)

The table is flushed as key/value set.

```
\command [key={value}]...
```

### context["command"](true)

An explicit `endlinechar` is inserted.

```
\command ^^M
```

### context["command"](node)

The node(list) is injected at the spot. Keep in mind that you need to do the proper memory management yourself.

```
\command {node(list)}
```

### context["command"](false,value)

The value is flushed without encapsulating tokens.

```
\command value
```

### context["command"]({ value }, { key = val }, val, false, val)

The arguments are flushed accordingly their nature and the order can be any.

```
\command [value][key={value}]{value}value
```

### context.direct(...)

The arguments are interpreted the same as if `direct` was a command, but no `\direct` is injected in front. Braces are added:

```
regular \expandafter \bold \ctxlua{context.direct("bold")} regular
black \expandafter \color \ctxlua{context.direct({"red"})}{red} black
black \expandafter \color \ctxlua{context.direct({"green"},"green")} black
```

The `\expandafter` makes sure that the `\bold` and `\color` macros see the following `{bold}`, `[red]`, and `[green]{green}` arguments.

regular **bold** regular
black red black
black green black

### context.delayed(...)

The arguments are interpreted the same as in a `context` call, but instead of a direct flush, the arguments will be flushed in a next cycle.

### context.delayed["command"](...)

The arguments are interpreted the same as in a `command` call, but instead of a direct flush, the command and arguments will be flushed in a next cycle.

### context.nested["command"]

This command returns the command, including given arguments as a string. No flushing takes place.

### context.nested

This command returns the arguments as a string and treats them the same as a regular `context` call.

### context.formatted["command"]([<regime>,]<format>,<arguments>)

This command returns the command that will pass it's arguments to the string formatter. When the first argument is a number, then it is interpreted as a catcode regime.

### context.formatted([<regime>,]<format>,<arguments>)

This command passes it's arguments to the string formatter. When the first argument is a number, then it is interpreted as a catcode regime.

## 18.2 METAFUN

### context.metafun.start()

This starts a METAFUN (or METAPOST) graphic.

### context.metafun.stop()

This finishes and flushes a METAFUN (or METAPOST) graphic.

### context.metafun("format",...)

The argument is appended to the current graphic data but the string formatter is used on following arguments.

### context.metafun.delayed

This namespace does the same as `context.delayed`: it wraps the code in such a way that it can be used in a function call.

## 18.3 Building blocks

### context.bgroup() context.egroup()

These are just `\bgroup` and `\egroup` equivalents and as these are in fact shortcuts to the curly braced we output these instead.

### context.space()

This one directly maps onto `\space`.

### context.par()

This one directly maps onto `\par`.

## 18.4 Basic Helpers

### context.rule(wd,ht,dp,direction) context.rule(specification)

A rule node is injected with the given properties. A specification is just a table with the four fields. The rule gets the current attributes.

### context.glyph(fontid,n) context.glyph(n)

A glyph node is injected with the given font id. When no id is given, the current font is used. The glyph gets the current attributes.

### context.char(n) context.char(str) context.char(tab)

This will inject one or more copies of \char calls. You can pass a number, a string representing a number, or a table with numbers.

### context.utfchar(n) context.utfchar(str)

This injects is UTF character (one or more bytes). You can pass a number or a string representing a numbers. You need to be aware of special characters in TEX, like #.

## 18.5 Registers

This is a table that hosts a couple of functions. The following new ones are available:

```
local n = newdimen (name)
local n = newskip  (name)
local n = newcount (name)
local n = newmuskip(name)
local n = newtoks  (name)
local n = newbox   (name)
```

These define a register with name name at the LUA end and \name at the TEX end. The registers' number is returned. The next function is like \chardef: it defines \name with value n.

```
local n = newchar(name,n)
```

It's not likely that you will use any of these commands, if only because when you're operating from the LUA end using LUA variables is more convenient.

## 18.6 Catcodes

Normally we operate under the so called context catcode regime. This means that content gets piped to TEX using the same meanings for characters as you normally use in CONTEXT. So, a $ starts math. In **table 18.1** we show the catcode regimes.

### context.catcodes

The context.catcodes tables contains the internal numbers of the catcode tables used. The next table shows the names that can be used.

| name | mnemonic | TEX command |
|---|---|---|
| context | ctx | ctxcatcodes |
| protect | prt | prtcatcodes |
| plain | tex | texcatcodes |

| ascii | context | tex | protect | text | verbatim |
|---|---|---|---|---|---|
| | space | space | space | space | other |
| ! | other | other | letter | other | other |
| " | other | other | other | other | other |
| # | parameter | parameter | parameter | other | other |
| $ | mathshift | mathshift | mathshift | other | other |
| % | comment | comment | comment | other | other |
| & | alignment | other | alignment | other | other |
| ' | other | other | other | other | other |
| ( | other | other | other | other | other |
| ) | other | other | other | other | other |
| * | other | other | other | other | other |
| + | other | other | other | other | other |
| , | other | other | other | other | other |
| - | other | other | other | other | other |
| . | other | other | other | other | other |
| / | other | other | other | other | other |
| 0 .. 9 | other | other | other | other | other |
| : | other | other | other | other | other |
| ; | other | other | other | other | other |
| < | other | other | other | other | other |
| = | other | other | other | other | other |
| > | other | other | other | other | other |
| ? | other | other | letter | other | other |
| @ | other | other | letter | other | other |
| A .. Z | letter | letter | letter | letter | letter |
| [ | other | other | other | other | other |
| \ | escape | escape | escape | escape | other |
| ] | other | other | other | other | other |
| ^ | superscript | other | superscript | other | other |
| _ | subscript | other | letter | other | other |
| ` | other | other | other | other | other |
| a .. z | letter | letter | letter | letter | letter |
| { | begingroup | begingroup | begingroup | begingroup | other |
| \| | other | active | active | other | other |
| } | endgroup | endgroup | endgroup | endgroup | other |
| ~ | other | active | active | other | other |

**Table 18.1** Catcode regimes

```
text         txt     txtcatcodes
verbatim     vrb     vrbcatcodes
```

### context.newindexer(catcodeindex)

This function defines a new indexer. You can think of the context command itself as an indexer. There are two (extra) predefined indexers:

```
context.verbatim = context.newindexer(context.catcodes.verbatim)
context.puretext = context.newindexer(context.catcodes.text)
```

### context.pushcatcodes(n) context.popcatcodes()

These commands switch to another catcode regime and back. They have to be used in pairs. Only the regimes atthe LUA end are set.

### context.unprotect() context.protect()

These commands switch to the protected regime and back. They have to be used in pairs. Beware: contrary to what its name suggests, the `unprotect` enables the protected regime. These functions also issue an `\unprotect` and `\protect` equivalent at the TEX end.

### context.verbatim context.puretext

The differences between these are subtle:

```
\startluacode
    context.verbatim.bold("Why do we use $ for math?") context.par()
    context.verbatim.bold("Why do we use { as start?") context.par()
    context.verbatim.bold("Why do we use } as end?")   context.par()
    context.puretext.bold("Why do we use {\\bi $} at all?")
\stopluacode
```

Verbatim makes all characters letters while pure text leaves the backslash and curly braces special.

**Why do we use $ for math?**
**Why do we use { as start?**
**Why do we use } as end?**
**Why do we use $ at all?**

### context.protected

The protected namespace is only used for commands that are in the CONTEXT private namespace.

### context.escaped(str) context.escape(str)

The first command pipes the escaped string to TEX, while the second one just returns an unescaped string. The characters `# $ % \ \ { }` are escaped.

### context.startcollecting() context.stopcollecting()

These two commands will turn flushing to TEX into collecting. This can be handy when you want to interface commands that grab arguments using delimiters and as such they are used deep down in some table related interfacing. You probably don't need them.

## 18.7 Templates

In addition to the regular template mechanism (part of the utilities) there is a dedicated template feature in the context namespace. An example demonstrates its working:

```
\startluacode
  local MyTable = [[
    \bTABLE
      \bTR
         \bTD \bf %one_first% \eTD
         \bTD %[one_second]% \eTD
      \eTR
      \bTR
        \bTD \bf %two_first% \eTD
        \bTD %[two_second]% \eTD
      \eTR
    \eTABLE
  ]]

  context.templates[MyTable] {
    one_first  = "one",
    two_first  = "two",
    one_second = "just one $",
    two_second = "just two $",
  }
\stopluacode
```

This renders:

| | |
|---|---|
| **one** | just one $ |
| **two** | just two $ |

You can also use more complex tables. Watch the space before and after the keys:

```
\startluacode
  local MyOtherTable = [[
    \bTABLE
      \bTR
        \bTD \bf % ['one']['first'] % \eTD
        \bTD %[ ['one']['second'] ]% \eTD
      \eTR
      \bTR
        \bTD \bf % ['two']['first'] % \eTD
```

```
        \bTD %[ ['two']['second'] ]% \eTD
      \eTR
    \eTABLE
  ]]

  local data = {
    one = { first = "one", second = "only 1$" },
    two = { first = "two", second = "only 2$" },
  }

  context.templates[MyOtherTable](data)

  context.templates(MyOtherTable,data)
\stopluacode
```

We get:

| | | | |
|---|---|---|---|
| **one** | only 1$ | **one** | only 1$ |
| **two** | only 2$ | **two** | only 2$ |

## 18.8 Management

### context.functions

This is private table that hosts managament of functions. You'd better leave this one alone!

### context.nodes

Normally you will just use `context(<somenode>)` to flush a node and this private table is more for internal use.

## 18.9 String handlers

These two functions implement handlers that split a given string into lines and do something with it. We stick to showing their call. They are used for special purpose flushing, like flushing content to TEX in commands discussed here. The XML subsystem also used a couple of dedicated handlers.

```
local foo = newtexthandler {
    content    = function(s) ... end,
    endofline  = function(s) ... end,
    emptyline  = function(s) ... end,
    simpleline = function(s) ... end,
}

local foo = newverbosehandler {
    line    = function(s) ... end,
    space   = function(s) ... end,
```

```
    content = function(s) ... end,
    before  = function() ... end,
    after   = function() ... end,
}
```

### context.printlines(str)

The low level `tex.print` function pipes its content to TEX and thereby terminates at at `\r` (cariage return, ASCII 13), although it depends on the way catcodes and line endings are set up. In fact, a line ending in TEX is not really one, as it gets replaced by a space. Only several lines in succession indicate a new paragraph.

```
\startluacode
    tex.print("line 1\n line 2\r line 3")
\stopluacode
```

This renders only two lines:

line 1 line 2

However, the `context` command gives all three lines:

```
\startluacode
    context("line 1\n line 2\r line 3")
\stopluacode
```

Like:

line 1 line 2 line 3

The `context.printlines` command is a direct way to print a string in a way similar to reading from a file. So,

```
tex.print(io.loaddata(resolvers.findfile("tufte")))
```

Gives one line, while:

```
context.printlines(io.loaddata(resolvers.findfile("tufte")))
```

gives them all, as does:

```
context(io.loaddata(resolvers.findfile("tufte")))
```

as does a naïve:

```
tex.print((string.gsub(io.loaddata(resolvers.findfile("tufte")),"\r","\n")))
```

But, because successive lines need to become paragraph separators as bit more work is needed and that is what `printlines` and `context` do for you. However, a more convenient alternative is presented next.

### context.loadfile(name)

This function locates and loads the file with the given name. The leading and trailing spaces are stripped.

### context.runfile(name)

This function locates and processes the file with the given name. The assumption is that it is a valid LUA file! When no suffix is given, the suffix cld (CONTEXT LUA document) is used.

### context.viafile(data[,tag])

The data is saved to a (pseudo) file with the optional name tag and read in again from that file. This is a robust way to make sure that the data gets processed like any other data read from file. It permits all kind of juggling with catcodes, verbatim and alike.

## 18.10 Helpers

### context.tocontext(variable)

For documentation or tracing it can be handy to serialize a variable. The tocontext function does this:

```
context.tocontext(true)
context.tocontext(123)
context.tocontext("foo")
context.tocontext(tonumber)
context.tocontext(nil)
context.tocontext({ "foo", "bar" },true)
context.tocontext({ this = { foo , "bar" } },true)
```

Beware, tocontext is also a table that you can assign to, but that might spoil serialization. This property makes it possible to extend the serializer.

### context.tobuffer(name,str[,catcodes])

With this function you can put content in a buffer, optionally under a catcode regime.

### context.tolines(str[,true])

This function splits the string in lines and flushes them one by one. When the second argument is true leading and trailing spaces are stripped. Each flushed line always gets one space appended.

### context.fprint([regime,]fmt,...),tex.fprint([regime,]fmt,...)

The tex.fprint is just there to complement the other flushers in the tex namespace and therefore we also have it in the context namespace.

## 18.11 Tracing

### context.settracing(true or false))

You can trace the TEX code that is generated at the TEX end with:

```
\enabletrackers[context.trace]
```

The LUA function sets the tracing from the LUA end. As the `context` command is used a lot in the core, you can expect some more tracing that the code that you're currently checking.

### context.pushlogger(fnc) context.poplogger() context.getlogger()

You can provide your own logger if needed. The pushed function receives one string argument. The getter returns three functions:

```
local flush, writer, flushdirect = context.getlogger()
```

The `flush` function is similar to `tex.sprint` and appends its arguments, while `flushdirect` treats each argument as a line and behaves like `tex.print`. The `flush` function adds braces and paranthesis around its arguments, apartt from the first one, which is considered to be a command. Examples are:

```
flush("one",2,"three") -- catcode, strings|numbers
writer("\\color",{"red"},"this is red")
```

and:

```
flush(context.catcodes.verbatim,"one",2,"three")
writer(context.catcodes.verbatim,"\\color",{"red"},"this is red")
```

## 18.12 States

There are several ways to implement alternative code paths in CONTEXT but modes and conditionals are used mostly. There area few helpers for that.

### context.conditionals context.setconditional(name,value)

Conditionals are used to keep a state. You can set their value using the setter, but their effect is not immediate but part of the current sequence of commands which is delegated to TEX. However, you can easily keep track of your state at the LUA end with an extra boolean. So, after

```
if context.conditionals.whatever then
    context.setconditional("dothis",false)
else
    context.setconditional("dothat",true)
end
```

the value of `dothis` and `dothat` conditions are not yet set in LUA.

### context.modes context.setmode(name,value)

As with conditionals, you can (re)set the modes in LUA but their values get changes as part of the command sequence which is delayed till after the LUA call.

### context.systemmodes context.setsystemmode(name,value)

The same applies as for regular modes.

### context.trialtypesetting()

This function returns `true` if we're in trial typesetting mode (used when for instance prerolling a table).

## 18.13 Steps

The stepper permits stepwise processing of CONTEXT code: after a step contyrol gets delegated to CONTEXT and afterwards back to LUA. There main limitation of this mechanism is that it cannot exceed the number of input levels.

### context.stepwise() context.step([str])

Usage is as follows:

```
context.stepwise (function()
    ...
    context.step(...)
    ...
    context.step(...)
    ...
    context.stepwise (function()
        ...
        context.step(...)
        ...
    context.step(...)
    ...
    end)
    ...
    context.step(...)
    ...
    context.step(...)
    ...
end)
```

# 19 Special commands

## 19.1 Tracing

There are a few functions in the `context` namespace that are no macros at the TEX end.

```
context.runfile("somefile.cld")
```

Another useful command is:

```
context.settracing(true)
```

There are a few tracing options that you can set at the TEX end:

```
\enabletrackers[context.files]
\enabletrackers[context.trace]
```

## 19.2 Overloads

A few macros have special functions (overloads) at the LUA end. One of them is `\char`. The function makes sure that the characters ends up right. The same is true for `\chardef`. So, you don't need to mess around with `\relax` or trailing spaces as you would do at the TEX end in order to tell the scanner to stop looking ahead.

```
context.char(123)
```

Other examples of macros that have optimized functions are `\par`, `\bgroup` and `\egroup`. Or take this:

```
1: \ctxlua{commands.doif(true)}{one}
2: \cldcommand{doif("a","a","two")}
3: \ctxcommand{doif(true)}{three}
```

1: one
2: two
3: three

## 19.3 Steps

We already mentioned the stepper as a very special trick so let's give some more explanation here. When you run the following code:

```
\startluacode
  context.startitemize()
    context.startitem()
      context("BEFORE 1")
    context.stopitem()
    context("\\setbox0\\hbox{!!!!}")
    context.startitem()
```

```
      context("%p",tex.getbox(0).width)
    context.stopitem()
  context.stopitemize()
\stopluacode
```

You get a message like:

```
[ctxlua]:8: attempt to index a nil value
...
10              context("\\setbox0\\hbox{!!!!}")
11              context.startitem()
12 >>                context("%p",tex.getbox(0).width)
...
```

due to the fact that the box is still void. All that the CONTEXT commands feed into TEX happens when the code snippet has finished. You can however run a snippet of code the following way:

```
\startluacode
  context.stepwise (function()
    context.startitemize()
      context.startitem()
        context.step("BEFORE 1")
      context.stopitem()
      context.step("\\setbox0\\hbox{!!!!}")
      context.startitem()
        context.step("%p",tex.getbox(0).width)
      context.stopitem()
    context.stopitemize()
  end)
\stopluacode
```

and get:

- BEFORE 1
- 12.22528pt

A more extensive example is:

```
\startluacode
  context.stepwise (function()
    context.startitemize()
      context.startitem()
        context.step("BEFORE 1")
      context.stopitem()
      context.step("\\setbox0\\hbox{!!!!}")
      context.startitem()
        context.step("%p",tex.getbox(0).width)
      context.stopitem()
      context.startitem()
        context.step("BEFORE 2")
```

```
context.stopitem()
context.step("\\setbox2\\hbox{????}")
context.startitem()
  context.step("%p",tex.getbox(2).width)
context.startitem()
  context.step("BEFORE 3")
context.stopitem()
context.startitem()
  context.step("\\copy0\\copy2")
context.stopitem()
context.startitem()
  context.step("BEFORE 4")
  context.startitemize()
    context.stepwise (function()
      context.step("\\bgroup")
      context.step("\\setbox0\\hbox{>>>>}")
      context.startitem()
        context.step("%p",tex.getbox(0).width)
      context.stopitem()
      context.step("\\setbox2\\hbox{<<<<}")
      context.startitem()
        context.step("%p",tex.getbox(2).width)
      context.stopitem()
      context.startitem()
        context.step("\\copy0\\copy2")
      context.stopitem()
      context.startitem()
        context.step("\\copy0\\copy2")
      context.stopitem()
      context.step("\\egroup")
    end)
  context.stopitemize()
context.stopitem()
context.startitem()
  context.step("AFTER 1\\par")
context.stopitem()
context.startitem()
  context.step("\\copy0\\copy2\\par")
context.stopitem()
context.startitem()
  context.step("\\copy0\\copy2\\par")
context.stopitem()
context.startitem()
  context.step("AFTER 2\\par")
context.stopitem()
context.startitem()
  context.step("\\copy0\\copy2\\par")
context.stopitem()
```

```
      context.startitem()
        context.step("\\copy0\\copy2\\par")
      context.stopitem()
    context.stopitemize()
  end)
\stopluacode
```

which gives:

- BEFORE 1
- 12.22528pt
- BEFORE 2
- 19.52527pt
- BEFORE 3
- !!!!????
- BEFORE 4
  - 33.72943pt
  - 33.72943pt
  - >>>><<<<
  - >>>><<<<
- AFTER 1
- !!!!????
- !!!!????
- AFTER 2
- !!!!????
- !!!!????

A step returns control to TEX immediately and after the TEX code that it feeds back is expanded, returns to LUA. There are some limitations due to the input stack but normally that is no real issue.

You can run the following code:

```
\definenumber[LineCounter][way=bypage]
\starttext
\startluacode
for i=1,2000 do
    context.incrementnumber { "LineCounter" }
    context.getnumber { "LineCounter" }
```

```
    context.par()
end
\stopluacode
\stoptext
```

You will notice however that the number is not right on each page. This is because TEX doesn't know yet that there is no room on the page. The next will work better:

```
\definenumber[LineCounter][way=bypage]
\starttext
\startluacode
context.stepwise(function()
    for i=1,2000 do
        context.testpage { 0 }
        context.incrementnumber { "LineCounter" }
        context.getnumber { "LineCounter" }
        context.par()
        context.step()
    end
end)
\stopluacode
\stoptext
```

Instead of the `testpage` function you can also play directly with registers, like:

```
if tex.pagegtotal + tex.count.lineheight > tex.pagetotal then
```

but often an already defined helper does a better job. Of course you will probably never need this kind of hacks anyway, if only because much more is going on and there are better ways then.

# 20 Files

## 20.1 Preprocessing

Although this option must be used with care, it is possible to preprocess files before they enter TEX. The following example shows this.

```
local function showline(str,filename,linenumber,noflines)
    logs.simple("[lc] file: %s, line: %s of %s, length: %s",
        file.basename(filename),linenumber,noflines,#str)
end

local function showfile(str,filename)
    logs.simple("[fc] file: %s, length: %s",
        file.basename(filename),#str)
end

resolvers.installinputlinehandler(showline)
resolvers.installinputfilehandler(showfile)
```

Preprocessors like this are rather innocent. If you want to manipulate the content you need to be aware of the fact that modules and such also pass your code, and manipulating them can give unexpected side effects. So, the following code will not make CONTEXT happy.

```
local function foo()
    return "bar"
end

resolvers.installinputlinehandler(foo)
```

But, as we pass the filename, you can base your preprocessing on names.

There can be multiple handlers active at the same time, and although more detailed control is possible, the current interface does not provide that, simply because having too many handlers active is asking for trouble anyway. What you can do, is putting your handler in front or after the built in handlers.

```
resolvers.installinputlinehandler("before",showline)
resolvers.installinputfilehandler("after", showfile)
```

Of course you can also preprocess files outside this mechanism, which in most cases might be a better idea. However, the following example code is quite efficient and robust.

```
local function MyHandler(str,filename)
    if file.suffix(filename) == "veryspecial" then
        logs.simple("preprocessing file '%s',filename)
        return MyConverter(str)
    else
        return str
    end
end
```

```
resolvers.installinputfilehandler("before",MyHandler)
```

In this case only files that have a suffix `.veryspecial` will get an extra treatment.

# Index

*needs checking, incomplete*