

tagpdf – A package to experiment with pdf tagging*

Ulrike Fischer[†]

Released 2022-12-22

Contents

1	Initialization and test if pdfmanagement is active.	7
2	Package options	7
3	Packages	8
3.1	a LastPage label	8
4	Variables	9
5	Variants of l3 commands	10
6	Setup label attributes	11
7	Label commands	11
8	Commands to fill seq and prop	12
9	General tagging commands	12
10	Keys for tagpdfsetup	13
11	loading of engine/more dependent code	14
I	The tagpdf-checks module	
	Messages and check code	
	Part of the tagpdf package	16
1	Commands	16

*This file describes v0.98, last revised 2022-12-22.

[†]E-mail: fischer@troubleshooting-tex.de

2	Description of log messages	16
2.1	\ShowTagging command	16
2.2	Messages in checks and commands	16
2.3	Messages from the ptagging code	17
2.4	Warning messages from the lua-code	17
2.5	Info messages from the lua-code	17
2.6	Debug mode messages and code	18
2.7	Messages	18
3	Messages	19
3.1	Messages related to mc-chunks	19
3.2	Messages related to structures	20
3.3	Attributes	21
3.4	Roles	21
3.5	Miscellaneous	22
4	Retrieving data	22
5	User conditionals	23
6	Internal checks	23
6.1	checks for active tagging	23
6.2	Checks related to structures	24
6.3	Checks related to roles	25
6.4	Check related to mc-chunks	26
6.5	Checks related to the state of MC on a page or in a split stream	28
II The tagpdf-user module		
Code related to L ^A T _E X2e user commands and document commands		
Part of the tagpdf package		32
1	Setup commands	32
2	Commands related to mc-chunks	32
3	Commands related to structures	33
4	Debugging	33
5	Extension commands	33
5.1	Fake space	34
5.2	Paratagging	34
5.3	Header and footer	34
5.4	Link tagging	35
6	User commands and extensions of document commands	35
7	Setup and preamble commands	35
8	Commands for the mc-chunks	35

9	Commands for the structure	36
10	Debugging	36
11	Commands to extend document commands	39
11.1	Document structure	39
11.2	Structure destinations	40
11.3	Fake space	40
11.4	Paratagging	41
11.5	Header and footer	43
11.6	Links	45
III	The <code>tagpdf-tree</code> module	
	Commands trees and main dictionaries	
	Part of the <code>tagpdf</code> package	47
1	Trees, pdfmanagement and finalization code	47
1.1	Check structure	47
1.2	Catalog: MarkInfo and StructTreeRoot	48
1.3	Writing structure elements	48
1.4	ParentTree	49
1.5	Rolemap dictionary	51
1.6	Classmap dictionary	52
1.7	Namespaces	53
1.8	Finishing the structure	54
1.9	StructParents entry for Page	54
IV	The <code>tagpdf-mc-shared</code> module	
	Code related to Marked Content (mc-chunks), code shared by	
	all modes	
	Part of the <code>tagpdf</code> package	55
1	Public Commands	55
2	Public keys	56
3	Marked content code – shared	56
3.1	Variables and counters	57
3.2	Functions	58
3.3	Keys	60
V	The <code>tagpdf-mc-generic</code> module	
	Code related to Marked Content (mc-chunks), generic mode	
	Part of the <code>tagpdf</code> package	62

1	Marked content code – generic mode	62
1.1	Variables	62
1.2	Functions	63
1.3	Looking at MC marks in boxes	66
1.4	Keys	73
VI	The <code>tagpdf-mc-luacode</code> module	
	Code related to Marked Content (mc-chunks), luamode-specific	
	Part of the tagpdf package	75
1	Marked content code – luamode code	75
1.1	Commands	76
1.2	Key definitions	80
VII	The <code>tagpdf-struct</code> module	
	Commands to create the structure	
	Part of the tagpdf package	83
1	Public Commands	83
2	Public keys	84
2.1	Keys for the structure commands	84
2.2	Setup keys	86
3	Variables	86
3.1	Variables used by the keys	88
3.2	Variables used by tagging code of basic elements	89
4	Commands	89
4.1	Initialization of the StructTreeRoot	90
4.2	Filling in the tag info	90
4.3	Handlings kids	91
4.4	Output of the object	94
5	Keys	97
6	User commands	102
7	Attributes and attribute classes	107
7.1	Variables	107
7.2	Commands and keys	107
VIII	The <code>tagpdf-luatex.def</code>	
	Driver for luatex	
	Part of the tagpdf package	110
1	Loading the lua	110
2	Logging functions	114

3	Helper functions	116
3.1	Retrieve data functions	116
3.2	Functions to insert the pdf literals	118
4	Function for the real space chars	119
5	Function for the tagging	122
6	Parenttree	127
IX	The tagpdf-roles module	
	Tags, roles and namespace code	
	Part of the tagpdf package	129
1	Code related to roles and structure names	129
1.1	Variables	129
1.2	Namespaces	132
1.3	Adding a new tag	133
1.3.1	pdf 1.7 and earlier	134
1.3.2	The pdf 2.0 version	135
1.4	Helper command to read the data from files	136
1.5	Reading the default data	138
1.6	Parent-child rules	139
1.6.1	Reading in the csv-files	139
1.6.2	Retrieving the parent-child rule	141
1.7	Remapping of tags	146
1.8	Key-val user interface	147
X	The tagpdf-space module	
	Code related to real space chars	
	Part of the tagpdf package	149
1	Code for interword spaces	149
	Index	152

<code>\ref_value:nnn</code>	<code>\ref_value:nnn{<label>}{<attribute>}{<fallback default>}</code>
-----------------------------	---

This is a temporary definition which will have to move to l3ref. It allows to locally set a default value if the label or the attribute doesn't exist. See issue #4 in Accessible-xref.

<code>\tag_stop_group_begin:</code>	We need commands to stop tagging in some places. There simply switches the two local booleans. The grouping commands can be used to group the effect.
<code>\tag_stop_group_end:</code>	
<code>\tag_stop:</code>	
<code>\tag_start:</code>	
<code>\tag_stop:n</code>	
<code>\tag_start:n</code>	

<code>\tag_stop:n</code>	<code>\tag_stop:n{<label>}</code>
<code>\tag_start:n</code>	<code>\tag_start:n{<label>}</code>

This commands are intended as a pair. The start command will only restart tagging if the previous stop command with the same label actually stopped tagging.

<code>activate-space_<setup-key></code>	<code>activate-space</code> activates the additional parsing needed for interword spaces. It is not documented, the parsing is currently implicitly activated by the known key <code>interwordspace</code> , as the code will perhaps move to some other place, now that it is better separated.
---	--

<code>activate-mc_<setup-key></code>
<code>activate-tree_<setup-key></code>
<code>activate-struct_<setup-key></code>
<code>activate-all_<setup-key></code>

Keys to activate the various tagging steps

<code>no-struct-dest_<setup-key></code>	The key allows to suppress the creation of structure destinations
---	---

<code>log_<setup-key></code>	The <code>log</code> takes currently the values <code>none</code> , <code>v</code> , <code>vv</code> , <code>vvv</code> , <code>all</code> . More details are in <code>tagpdf-checks</code> .
------------------------------------	---

<code>tagunmarked_<setup-key></code>	This key allows to set if (in luamode) unmarked text should be marked up as artifact. The initial value is true.
--	--

<code>tabsorder_<setup-key></code>	This sets the tabsorder on a page. The values are <code>row</code> , <code>column</code> , <code>structure</code> (default) or <code>none</code> . Currently this is set more or less globally. More finer control can be added if needed.
--	--

<code>tagstruct</code>	These are attributes used by the label/ref system.
<code>tagstructobj</code>	
<code>tagabspage</code>	
<code>tagmcabs</code>	
<code>tagmcid</code>	

1 Initialization and test if pdfmanagement is active.

```

1 <@@=tag>
2 <*package>
3 \ProvidesExplPackage {tagpdf} {2022-12-22} {0.98}
4 { A package to experiment with pdf tagging }
5
6 \bool_if:nF
7 {
8   \bool_lazy_and_p:nn
9     {\cs_if_exist_p:N \pdfmanagement_if_active_p:}
10    {\pdfmanagement_if_active_p: }
11 }
12 { %error for now, perhaps warning later.
13   \PackageError{tagpdf}
14   {
15     PDF~resource~management~is~no~active!\MessageBreak
16     tagpdf~will~no~work.
17   }
18   {
19     Activate~it~with \MessageBreak
20     \string\RequirePackage{pdfmanagement-testphase}\MessageBreak
21     \string\DocumentMetadata{<options>}\MessageBreak
22     before~\string\documentclass
23   }
24 }
25 </package>
26
27 <*debug>
28 \ProvidesExplPackage {tagpdf-debug} {2022-12-22} {0.98}
29 { debug code for tagpdf }
30 \Atifpackageloaded{tagpdf}{\PackageWarning{tagpdf-debug}{tagpdf~not~loaded,~quitting}}\endinput
31 </debug> We map the internal module name “tag” to “tagpdf” in messages.
32 <*package>
33 \prop_gput:Nnn \g_msg_module_name_prop { tag }{ tagpdf }
34 </package>
35
36 Debug mode has its special mapping:
37 <*debug>
38 \prop_gput:Nnn \g_msg_module_type_prop { tag / debug} {}
39 \prop_gput:Nnn \g_msg_module_name_prop { tag / debug }{tagpdf~DEBUG}
40 </debug>

```

2 Package options

There are only two options to switch for luatex between generic and luamode, TODO try to get rid of them.

```

36 <*package>
37 \bool_new:N\g__tag_mode_lua_bool
38 \DeclareOption {luamode} { \sys_if_engine luatex:T { \bool_gset_true:N \g__tag_mode_lua_bo
39 \DeclareOption {genericmode}{ \bool_gset_false:N\g__tag_mode_lua_bool }
40 \ExecuteOptions{luamode}
41 \ProcessOptions

```

3 Packages

We need the temporary version of l3ref until this is in the kernel.

```
42 \RequirePackage{l3ref-tmp}
```

To be on the safe side for now, load also the base definitions

```
43 \RequirePackage{tagpdf-base}
```

```
44 \</package>
```

```
45 \<*base>
```

```
46 \ProvidesExplPackage {tagpdf-base} {2022-12-22} {0.98}
```

```
47 {part of tagpdf - provide base, no-op versions of the user commands }
```

```
48 \</base>
```

The no-op version should behave as near enough to the real code as possible, so we define a command which is a special in the relevant backends:

```
49 \<*base>
```

```
50 \AddToHook{begindocument}
```

```
51 {
```

```
52   \str_case:NnF \c_sys_backend_str
```

```
53   {
```

```
54     { luatex } { \cs_new_protected:Npn \__tag_whatsits: {} }
```

```
55     { dvisvgm } { \cs_new_protected:Npn \__tag_whatsits: {} }
```

```
56   }
```

```
57   {
```

```
58     \cs_new_protected:Npn \__tag_whatsits: {\tex_special:D {}} }
```

```
59   }
```

```
60 }
```

```
61 \</base>
```

3.1 a LastPage label

See also issue #2 in Accessible-xref

```
\__tag_lastpagelabel:
```

```
62 \<*package>
```

```
63 \cs_new_protected:Npn \__tag_lastpagelabel:
```

```
64 {
```

```
65   \legacy_if:nT { @files } {
```

```
66     {
```

```
67       \exp_args:NNnx \exp_args:NNx\iow_now:Nn \@auxout
```

```
68       {
```

```
69         \token_to_str:N \newlabeldata
```

```
70         {__tag_LastPage}
```

```
71         {
```

```
72           {abspage} { \int_use:N \g_shipout_readonly_int }
```

```
73           {tagmcabs}{ \int_use:N \c@g__tag_MCID_abs_int }
```

```
74         }
```

```
75       }
```

```
76     }
```

```
77   }
```

```
78
```

```
79 \AddToHook{enddocument/afterlastpage}
```

```
80 {\__tag_lastpagelabel:}
```


(End definition for `_tag_lastpagelabel:`.)

`\ref_value:nnn` This allows to locally set a default value if the label or the attribute doesn't exist.

```

81 \cs_if_exist:NF \ref_value:nnn
82 {
83   \cs_new:Npn \ref_value:nnn #1#2#3
84   {
85     \exp_args:Nee
86     \__ref_value:nnn
87     { \tl_to_str:n {#1} } { \tl_to_str:n {#2} } {#3}
88   }
89   \cs_new:Npn \__ref_value:nnn #1#2#3
90   {
91     \tl_if_exist:cTF { g__ref_label_ #1 _ #2 _tl }
92     { \tl_use:c { g__ref_label_ #1 _ #2 _tl } }
93     {
94       #3
95     }
96   }
97 }
```

(End definition for `\ref_value:nnn`. This function is documented on page 6.)

4 Variables

`\l__tag_tmpa_tl` A few temporary variables

```

\l__tag_tmpb_tl 98 \tl_new:N \l__tag_tmpa_tl
\l__tag_get_tmpc_tl 99 \tl_new:N \l__tag_tmpb_tl
\tag_get_parent_tmpb_tl\l__tag_tmpa_str 100 \tl_new:N \l__tag_get_tmpc_tl
\l__tag_tmpa_prop 101 \tl_new:N \l__tag_get_parent_tmpa_tl
\l__tag_tmpa_seq 102 \tl_new:N \l__tag_get_parent_tmpb_tl
\l__tag_tmpb_seq 103 \str_new:N \l__tag_tmpa_str
\l__tag_tmpa_clist 104 \prop_new:N \l__tag_tmpa_prop
\l__tag_tmpa_int 105 \seq_new:N \l__tag_tmpa_seq
\l__tag_tmpa_box 106 \seq_new:N \l__tag_tmpb_seq
\l__tag_tmpb_box 107 \clist_new:N \l__tag_tmpa_clist
108 \int_new:N \l__tag_tmpa_int
109 \box_new:N \l__tag_tmpa_box
110 \box_new:N \l__tag_tmpb_box
```

(End definition for `\l__tag_tmpa_tl` and others.)

Attribute lists for the label command. We have a list for mc-related labels, and one for structures.

```

\c__tag_refmc_clist
\c__tag_refstruct_clist 111 \clist_const:Nn \c__tag_refmc_clist {tagabspage,tagmcabs,tagmcid}
112 \clist_const:Nn \c__tag_refstruct_clist {tagstruct,tagstructobj}
```

(End definition for `\c__tag_refmc_clist` and `\c__tag_refstruct_clist`.)

`\l__tag_loglevel_int` This integer hold the log-level and so allows to control the messages. TODO: a list which log-level shows what is needed. The current behaviour is quite ad-hoc.

```

113 \int_new:N \l__tag_loglevel_int
```

(End definition for \l__tag_loglevel_int.)

\g__tag_active_space_bool These booleans should help to control the global behaviour of tagpdf. Ideally it should more or less do nothing if all are false. The space-boolean controls the interword space code, the mc-boolean activates \tag_mc_begin:n, the tree-boolean activates writing the finish code and the pdfmanagement related commands, the struct-boolean activates the storing of the structure data. In a normal document all should be active, the split is only there for debugging purpose. Structure destination will be activated automatically if pdf version 2.0 is detected, but with the boolean struct-dest-boolean one can suppress them. Also we assume currently that they are set only at begin document. But if some control passing over groups are needed they could be perhaps used in a document too. TODO: check if they are used everywhere as needed and as wanted.

```
114 \bool_new:N \g__tag_active_space_bool
115 \bool_new:N \g__tag_active_mc_bool
116 \bool_new:N \g__tag_active_tree_bool
117 \bool_new:N \g__tag_active_struct_bool
118 \bool_new:N \g__tag_active_struct_dest_bool
119 \bool_gset_true:N \g__tag_active_struct_dest_bool
```

(End definition for \g__tag_active_space_bool and others.)

\l__tag_active_mc_bool These booleans should help to control the *local* behaviour of tagpdf. In some cases it could e.g. be necessary to stop tagging completely. As local booleans they respect groups. TODO: check if they are used everywhere as needed and as wanted.

```
120 \bool_new:N \l__tag_active_mc_bool
121 \bool_set_true:N \l__tag_active_mc_bool
122 \bool_new:N \l__tag_active_struct_bool
123 \bool_set_true:N \l__tag_active_struct_bool
```

(End definition for \l__tag_active_mc_bool and \l__tag_active_struct_bool.)

\g__tag_tagunmarked_bool This boolean controls if the code should try to automatically tag parts not in mc-chunk. It is currently only used in luamode. It would be possible to use it in generic mode, but this would create quite a lot empty artifact mc-chunks.

```
124 \bool_new:N \g__tag_tagunmarked_bool
```

(End definition for \g__tag_tagunmarked_bool.)

5 Variants of l3 commands

```
125 \prg_generate_conditional_variant:Nnn \pdf_object_if_exist:n {e}{T,F}
126 \cs_generate_variant:Nn \pdf_object_ref:n {e}
127 \cs_generate_variant:Nn \pdfannot_dict_put:nnn {nnx}
128 \cs_generate_variant:Nn \pdffile_embed_stream:nnn {nxx,oxx}
129 \cs_generate_variant:Nn \prop_gput:Nnn {Nxx,Nen}
130 \cs_generate_variant:Nn \prop_put:Nnn {Nxx}
131 \cs_generate_variant:Nn \prop_item:Nn {No,Ne}
132 \cs_generate_variant:Nn \ref_label:nn { nv }
133 \cs_generate_variant:Nn \seq_set_split:Nnn{Nne}
134 \cs_generate_variant:Nn \str_set_convert:Nnnn {Nonn, Noon, Nnon }
135 \cs_generate_variant:Nn \clist_map_inline:nn {on}
```

6 Setup label attributes

tagstruct This are attributes used by the label/ref system. With structures we store the structure number **tagstruct** and the object reference **tagstructobj**. The second is needed to be able to reference a structure which hasn't been created yet. The alternative would be to create the object in such cases, but then we would have to check the object existence all the time.

With mc-chunks we store the absolute page number **tagabspage**, the absolute id **tagmcabc**, and the id on the page **tagmcid**.

```

136 \ref_attribute_gset:nnnn { tagstruct } {0} { now }
137 { \int_use:N \c@g__tag_struct_abs_int }
138 \ref_attribute_gset:nnnn { tagstructobj } {} { now }
139 {
140   \pdf_object_if_exist:eT {__tag/struct/\int_use:N \c@g__tag_struct_abs_int}
141   {
142     \pdf_object_ref:e{__tag/struct/\int_use:N \c@g__tag_struct_abs_int}
143   }
144 }
145 \ref_attribute_gset:nnnn { tagabspage } {0} { shipout }
146 { \int_use:N \g_shipout_readonly_int }
147 \ref_attribute_gset:nnnn { tagmcabs } {0} { now }
148 { \int_use:N \c@g__tag_MCID_abs_int }
149 \ref_attribute_gset:nnnn {tagmcid } {0} { now }
150 { \int_use:N \g__tag_MCID_tmp_bypage_int }
```

(End definition for *tagstruct* and others. These functions are documented on page 6.)

7 Label commands

__tag_ref_label:nn A version of **\ref_label:nn** to set a label which takes a keyword **mc** or **struct** to call the relevant lists. TODO: check if **\@bsphack** and **\@esphack** make sense here.

```

151 \cs_new_protected:Npn \__tag_ref_label:nn #1 #2 %#1 label, #2 name of list mc or struct
152 {
153   \@bsphack
154   \ref_label:nv {#1}{c__tag_ref#2_clist}
155   \@esphack
156 }
157 \cs_generate_variant:Nn \__tag_ref_label:nn {en}
```

(End definition for *__tag_ref_label:nn*.)

__tag_ref_value:nnn A local version to retrieve the value. It is a direct wrapper, but to keep naming consistent It uses the variant defined temporarily above.

```

158 \cs_new:Npn \__tag_ref_value:nnn #1 #2 #3 %#1 label, #2 attribute, #3 default
159 {
160   \ref_value:nnn {#1}{#2}{#3}
161 }
162 \cs_generate_variant:Nn \__tag_ref_value:nnn {enn}
```

(End definition for *__tag_ref_value:nnn*.)

`__tag_ref_value_lastpage:nn` A command to retrieve the lastpage label, this will be adapted when there is a proper, kernel lastpage label.

```

163 \cs_new:Npn \__tag_ref_value_lastpage:nn #1 #2
164 {
165   \ref_value:nnn {\__tag_LastPage}{#1}{#2}
166 }

```

(End definition for `__tag_ref_value_lastpage:nn`.)

8 Commands to fill seq and prop

With most engines these are simply copies of the expl3 commands, but luatex will overwrite them, to store the data also in lua tables.

```

\__tag_prop_new:N
\__tag_seq_new:N
\__tag_prop_gput:Nnn
\__tag_seq_gput_right:Nn
\__tag_seq_item:cn
\__tag_prop_item:cn
\__tag_seq_show:N
\__tag_prop_show:N
167 \cs_set_eq:NN \__tag_prop_new:N \prop_new:N
168 \cs_set_eq:NN \__tag_seq_new:N \seq_new:N
169 \cs_set_eq:NN \__tag_prop_gput:Nnn \prop_gput:Nnn
170 \cs_set_eq:NN \__tag_seq_gput_right:Nn \seq_gput_right:Nn
171 \cs_set_eq:NN \__tag_seq_item:cn \seq_item:cn
172 \cs_set_eq:NN \__tag_prop_item:cn \prop_item:cn
173 \cs_set_eq:NN \__tag_seq_show:N \seq_show:N
174 \cs_set_eq:NN \__tag_prop_show:N \prop_show:N
175
176 \cs_generate_variant:Nn \__tag_prop_gput:Nnn { Nxn , Nxx, Nnx , cnn, cxn, cnx, cno }
177 \cs_generate_variant:Nn \__tag_seq_gput_right:Nn { Nx , No, cn, cx }
178 \cs_generate_variant:Nn \__tag_prop_new:N { c }
179 \cs_generate_variant:Nn \__tag_seq_new:N { c }
180 \cs_generate_variant:Nn \__tag_seq_show:N { c }
181 \cs_generate_variant:Nn \__tag_prop_show:N { c }

```

(End definition for `__tag_prop_new:N` and others.)

9 General tagging commands

`\tag_stop_group_begin:` We need commands to stop tagging in some places. This simply switches the two local booleans. In some cases tagging should only restart, if it actually was stopped before.

`\tag_stop_group_end:` For this it is possible to label a stop.

```

\tag_stop:
\tag_start:
\tag_stop:n
\tag_start:n
182 \cs_new_protected:Npn \tag_stop_group_begin:
183 {
184   \group_begin:
185   \bool_set_false:N \l__tag_active_struct_bool
186   \bool_set_false:N \l__tag_active_mc_bool
187 }
188 \cs_set_eq:NN \tag_stop_group_end: \group_end:
189 \cs_set_protected:Npn \tag_stop:
190 {
191   \bool_set_false:N \l__tag_active_struct_bool
192   \bool_set_false:N \l__tag_active_mc_bool
193 }
194 \cs_set_protected:Npn \tag_start:
195 {

```

```

196     \bool_set_true:N \l__tag_active_struct_bool
197     \bool_set_true:N \l__tag_active_mc_bool
198   }
199   \prop_new:N\g__tag_state_prop
200   \cs_set_protected:Npn \tag_stop:n #1
201   {
202     \tag_if_active:TF
203     {
204       \bool_set_false:N \l__tag_active_struct_bool
205       \bool_set_false:N \l__tag_active_mc_bool
206       \prop_gput:Nnn \g__tag_state_prop { #1 }{ 1 }
207     }
208     {
209       \prop_gremove:Nn \g__tag_state_prop { #1 }
210     }
211   }
212   \cs_set_protected:Npn \tag_start:n #1
213   {
214     \prop_gpop:NnN \g__tag_state_prop {#1}\l__tag_tmpa_tl
215     \quark_if_no_value:NF \l__tag_tmpa_tl
216     {
217       \bool_set_true:N \l__tag_active_struct_bool
218       \bool_set_true:N \l__tag_active_mc_bool
219     }
220   }
221   </package>
222   <*base>
223   \cs_new_protected:Npn \tag_stop:{}
224   \cs_new_protected:Npn \tag_start:{}
225   \cs_new_protected:Npn \tag_stop:n{}
226   \cs_new_protected:Npn \tag_start:n{}
227   </base>

```

(End definition for `\tag_stop_group_begin:` and others. These functions are documented on page 6.)

10 Keys for tagpdfsetup

TODO: the log-levels must be sorted

`activate-space□(setup-key)` Keys to (globally) activate tagging. `activate-space` activates the additional parsing needed for interword spaces. It is not documented, the parsing is currently implicitly activated by the known key `interwordspace`, as the code will perhaps move to some other place, now that it is better separated. `no-struct-dest` allows to suppress structure destinations.

```

228   <*package>
229   \keys_define:nn { __tag / setup }
230   {
231     activate-space .bool_gset:N = \g__tag_active_space_bool,
232     activate-mc .bool_gset:N = \g__tag_active_mc_bool,
233     activate-tree .bool_gset:N = \g__tag_active_tree_bool,
234     activate-struct .bool_gset:N = \g__tag_active_struct_bool,
235     activate-all .meta:n =
236     {activate-mc={#1},activate-tree={#1},activate-struct={#1}},

```

```

237     activate-all .default:n = true,
238     no-struct-dest .bool_gset_inverse:N = \g__tag_active_struct_dest_bool,
239

```

(End definition for activate-space (setup-key) and others. These functions are documented on page 6.)

log_␣(setup-key) The log takes currently the values none, v, vv, vvv, all. The description of the log levels is in tagpdf-checks.

```

240     log .choice:,
241     log / none .code:n = {\int_set:Nn \l__tag_loglevel_int { 0 }},
242     log / v .code:n =
243     {
244         \int_set:Nn \l__tag_loglevel_int { 1 }
245         \cs_set_protected:Nn \__tag_check_typeout_v:n { \iow_term:x {##1} }
246     },
247     log / vv .code:n = {\int_set:Nn \l__tag_loglevel_int { 2 }},
248     log / vvv .code:n = {\int_set:Nn \l__tag_loglevel_int { 3 }},
249     log / all .code:n = {\int_set:Nn \l__tag_loglevel_int { 10 }},

```

(End definition for log (setup-key). This function is documented on page 6.)

tagunmarked_␣(setup-key) This key allows to set if (in luamode) unmarked text should be marked up as artifact. The initial value is true.

```

250     tagunmarked .bool_gset:N = \g__tag_tagunmarked_bool,
251     tagunmarked .initial:n = true,

```

(End definition for tagunmarked (setup-key). This function is documented on page 6.)

tabsorder_␣(setup-key) This sets the tabsorder on a page. The values are row, column, structure (default) or none. Currently this is set more or less globally. More finer control can be added if needed.

```

252     tabsorder .choice:,
253     tabsorder / row .code:n =
254         \pdfmanagement_add:nnn { Page } {Tabs}{/R},
255     tabsorder / column .code:n =
256         \pdfmanagement_add:nnn { Page } {Tabs}{/C},
257     tabsorder / structure .code:n =
258         \pdfmanagement_add:nnn { Page } {Tabs}{/S},
259     tabsorder / none .code:n =
260         \pdfmanagement_remove:nn {Page} {Tabs},
261     tabsorder .initial:n = structure,
262     uncompress .code:n = { \pdf_uncompress: },
263 }

```

(End definition for tabsorder (setup-key). This function is documented on page 6.)

11 loading of engine/more dependent code

```

264 \sys_if_engine luatex:T
265 {
266     \file_input:n {tagpdf-luatex.def}
267 }
268 \</package>

```

```

269 <*mcloding>
270 \bool_if:NTF \g__tag_mode_lua_bool
271 {
272   \RequirePackage {tagpdf-mc-code-lua}
273 }
274 {
275   \RequirePackage {tagpdf-mc-code-generic} %
276 }
277 </mcloding>
278 <*debug>
279 \bool_if:NTF \g__tag_mode_lua_bool
280 {
281   \RequirePackage {tagpdf-debug-lua}
282 }
283 {
284   \RequirePackage {tagpdf-debug-generic} %
285 }
286 </debug>

```

Part I

The tagpdf-checks module

Messages and check code

Part of the tagpdf package

1 Commands

`\tag_if_active_p:` ★ This command tests if tagging is active. It only gives true if all tagging has been activated, `\tag_if_active:` *TF* ★ *and* if tagging hasn't been stopped locally.

`\tag_get:n` ★ `\tag_get:n{<keyword>}`

This is a generic command to retrieve data for the current structure or mc-chunk. Currently the only sensible values for the argument *<keyword>* are `mc_tag` and `struct_tag` and `struct_num`.

2 Description of log messages

2.1 \ShowTagging command

Argument	type	note
<code>\ShowTaggingmc-data = num</code>	log+term	lua-only
<code>\ShowTaggingmc-current</code>	log+term	
<code>\ShowTaggingstruck-stack= [log show]</code>	log or term+stop	

2.2 Messages in checks and commands

command	message	action
<code>\@@_check_structure_has_tag:n</code>	struct-missing-tag	error
<code>\@@_check_structure_tag:N</code>	role-unknown-tag	warning
<code>\@@_check_info_closing_struct:n</code>	struct-show-closing	info
<code>\@@_check_no_open_struct:</code>	struct-faulty-nesting	error
<code>\@@_check_struct_used:n</code>	struct-used-twice	warning
<code>\@@_check_add_tag_role:nn</code>	role-missing, role-tag, role-unknown	warning, info (>0), warning
<code>\@@_check_mc_if_nested:</code>	mc-nested	warning
<code>\@@_check_mc_if_open:</code>	mc-not-open	warning
<code>\@@_check_mc_pushed_popped:nn</code>	mc-pushed, mc-popped	info (2), info+seq_log (>2)
<code>\@@_check_mc_tag:N</code>	mc-tag-missing, role-unknown-tag	error (missing), warning (unknown).
<code>\@@_check_mc_used:n</code>	mc-used-twice	warning
<code>\@@_check_show_MCID_by_page:</code>		
<code>\tag_mc_use:n</code>	mc-label-unknown, mc-used-twice	warning
<code>\role_add_tag:nn</code>	new-tag	info (>0)
	sys-no-interwordspace	warning
<code>\@@_struct_write_obj:n</code>	struct-no-objnum	error
<code>\tag_struct_begin:n</code>	struct-faulty-nesting	error
<code>\@@_struct_insert_annot:nn</code>	struct-faulty-nesting	error
<code>tag_struct_use:n</code>	struct-label-unknown	warning
<code>attribute-class, attribute</code>	attr-unknown	error
<code>\@@_tree_fill_parenttree:</code>	tree-mcid-index-wrong	warning TODO: should trigger a standard rerun m
<code>in enddocument/info-hook</code>	para-hook-count-wrong	error (warning?)

2.3 Messages from the ptagging code

A few messages are issued in generic mode from the code which reinserts missing TMB/TME. This is currently done if log-level is larger than zero. TODO: reconsider log-level and messages when this code settles down.

2.4 Warning messages from the lua-code

The messages are triggered if the log-level is at least equal to the number.

message	log-level	remark
WARN TAG-NOT-TAGGED:	1	
WARN TAG-OPEN-MC:	1	
WARN SHIPOUT-MC-OPEN:	1	
WARN SHIPOUT-UPS:	0	shouldn't happen
WARN TEX-MC-INSERT-MISSING:	0	shouldn't happen
WARN TEX-MC-INSERT-NO-KIDS:	2	e.g. from empty hbox

2.5 Info messages from the lua-code

The messages are triggered if the log-level is at least equal to the number. TAG messages are from the traversing function, TEX from code used in the tagpdf-mc module. PARENTREE is the code building the parenttree.

message	log-level	remark
INFO SHIPOUT-INSERT-LAST-EMC	3	finish of shipout code
INFO SPACE-FUNCTION-FONT	3	interwordspace code
INFO TAG-ABSPAGE	3	
INFO TAG-ARGS	4	
INFO TAG-ENDHEAD	4	
INFO TAG-ENDHEAD	4	
INFO TAG-HEAD	3	
INFO TAG-INSERT-ARTIFACT	3	
INFO TAG-INSERT-BDC	3	
INFO TAG-INSERT-EMC	3	
INFO TAG-INSERT-TAG	3	
INFO TAG-KERN-SUBTYPE	4	
INFO TAG-MATH-SUBTYPE	4	
INFO TAG-MC-COMPARE	4	
INFO TAG-MC-INTO-PAGE	3	
INFO TAG-NEW-MC-NODE	4	
INFO TAG-NODE	3	
INFO TAG-NO-HEAD	3	
INFO TAG-NOT-TAGGED	2	replaced by artifact
INFO TAG-QUITTING-BOX	4	
INFO TAG-STORE-MC-KID	4	
INFO TAG-TRAVERSING-BOX	3	
INFO TAG-USE-ACTUALTEXT	3	
INFO TAG-USE-ALT	3	
INFO TAG-USE-RAW	3	
INFO TEX-MC-INSERT-KID	3	

message	log-level	remark
INFO TEX-MC-INSERT-KID-TEST	4	
INFO TEX-MC-INTO-STRUCT	3	
INFO TEX-STORE-MC-DATA	3	
INFO TEX-STORE-MC-KID	3	
INFO PARENTTREE-CHUNKS	3	
INFO PARENTTREE-NO-DATA	3	
INFO PARENTTREE-NUM	3	
INFO PARENTTREE-NUMENTRY	3	
INFO PARENTTREE-STRUCT-OBJREF	4	

2.6 Debug mode messages and code

If the package tagpdf-debug is loaded a number of commands are redefined and enhanced with additional commands which can be used to output debug messages or collect statistics. The commands are present but do nothing if the log-level is zero.

command	name	action	remark
\tag_mc_begin:n	mc-begin-insert	msg	
	mc-begin-ignore	msg	if inactive

2.7 Messages

mc-nested
mc-tag-missing
mc-label-unknown
mc-used-twice
mc-not-open
mc-pushed
mc-popped
mc-current

Various messages related to mc-chunks. TODO document their meaning.

struct-no-objnum
struct-faulty-nesting
struct-missing-tag
struct-used-twice
struct-label-unknown
struct-show-closing

Various messages related to structure. TODO document their meaning.

attr-unknown

Message if an attribute is unknown.

role-missing
role-unknown
role-unknown-tag
role-tag
new-tag

Messages related to role mapping.

<hr/> <hr/> tree-mcid-index-wrong	Used in the tree code, typically indicates the document must be rerun.
<hr/> <hr/> sys-no-interwordspace	Message if an engine doesn't support inter word spaces
<hr/> <hr/> para-hook-count-wrong	Message if the number of begin paragraph and end paragraph differ. This normally means faulty structure.
	<pre> 1 <@@=tag> 2 <*header> 3 \ProvidesExplPackage {tagpdf-checks-code} {2022-12-22} {0.98} 4 {part of tagpdf - code related to checks, conditionals, debugging and messages} 5 </header> </pre>

3 Messages

3.1 Messages related to mc-chunks

mc-nested This message is issue is a mc is opened before the previous has been closed. This is not relevant for luamode, as the attributes don't care about this. It is used in the `\@@_check_mc_if_nested`: test.

```

6 <*package>
7 \msg_new:nnn { tag } {mc-nested} { nested-marked-content-found~~mcid~#1 }

```

(End definition for mc-nested. This function is documented on page 18.)

mc-tag-missing If the tag is missing

```

8 \msg_new:nnn { tag } {mc-tag-missing} { required-tag-missing~~mcid~#1 }

```

(End definition for mc-tag-missing. This function is documented on page 18.)

mc-label-unknown If the label of a mc that is used in another place is not known (yet) or has been undefined as the mc was already used.

```

9 \msg_new:nnn { tag } {mc-label-unknown}
10 { label~#1~unknown~or~has~been~already~used.\\
11   Either~rerun~or~remove~one~of~the~uses. }

```

(End definition for mc-label-unknown. This function is documented on page 18.)

mc-used-twice An mc-chunk can be inserted only in one structure. This indicates wrong coding and so should at least give a warning.

```

12 \msg_new:nnn { tag } {mc-used-twice} { mc~#1~has~been~already~used }

```

(End definition for mc-used-twice. This function is documented on page 18.)

mc-not-open This is issued if a `\tag_mc_end`: is issued wrongly, wrong coding.

```

13 \msg_new:nnn { tag } {mc-not-open} { there-is-no-mc-to-end-at~#1 }

```

(End definition for mc-not-open. This function is documented on page 18.)

mc-pushed Informational messages about mc-pushing.
mc-popped 14 \msg_new:nnn { tag } {mc-pushed} { #1~has~been~pushed~to~the~mc~stack}
 15 \msg_new:nnn { tag } {mc-popped} { #1~has~been~removed~from~the~mc~stack }
 (End definition for mc-pushed and mc-popped. These functions are documented on page 18.)

mc-current Informational messages about current mc state.
 16 \msg_new:nnn { tag } {mc-current}
 17 { current~MC:~
 18 \bool_if:NTF\g__tag_in_mc_bool
 19 {abscnt=__tag_get_mc_abs_cnt:,~tag=\g__tag_mc_key_tag_tl}
 20 {no~MC~open,~current~abs cnt=__tag_get_mc_abs_cnt:"}
 21 }
 (End definition for mc-current. This function is documented on page 18.)

3.2 Messages related to structures

struct-unknown if for example a parent key value points to structure that doesn't exist (yet)
 22 \msg_new:nnn { tag } {struct-unknown}
 23 { structure~with~number~#1~doesn't~exist\\ #2 }
 (End definition for struct-unknown. This function is documented on page ??.)

struct-no-objnum Should not happen ...
 24 \msg_new:nnn { tag } {struct-no-objnum} { objnum~missing~for~structure~#1 }
 (End definition for struct-no-objnum. This function is documented on page 18.)

struct-faulty-nesting This indicates that there is somewhere one \tag_struct_end: too much. This should be normally an error.
 25 \msg_new:nnn { tag }
 26 {struct-faulty-nesting}
 27 { there~is~no~open~structure~on~the~stack }
 (End definition for struct-faulty-nesting. This function is documented on page 18.)

struct-missing-tag A structure must have a tag.
 28 \msg_new:nnn { tag } {struct-missing-tag} { a~structure~must~have~a~tag! }
 (End definition for struct-missing-tag. This function is documented on page 18.)

struct-used-twice
 29 \msg_new:nnn { tag } {struct-used-twice}
 30 { structure~with~label~#1~has~already~been~used }
 (End definition for struct-used-twice. This function is documented on page 18.)

struct-label-unknown label is unknown, typically needs a rerun.
 31 \msg_new:nnn { tag } {struct-label-unknown}
 32 { structure~with~label~#1~is~unknown~rerun }
 (End definition for struct-label-unknown. This function is documented on page 18.)

struct-show-closing Informational message shown if log-mode is high enough

```
33 \msg_new:nnn { tag } {struct-show-closing}
34 { closing-structure-#1~tagged~\use:e{\prop_item:cn{g__tag_struct_#1_prop}{S}} }
```

(End definition for struct-show-closing. This function is documented on page 18.)

tree-struct-still-open Message issued at the end if there are beside Root other open structures on the stack.

```
35 \msg_new:nnn { tag } {tree-struct-still-open}
36 {
37   There~are~still~open~structures~on~the~stack!\\
38   The~stack~contains~\seq_use:Nn\g__tag_struct_tag_stack_seq{,}.\\
39   The~structures~are~automatically~closed,\\
40   but~their~nesting~can~be~wrong.
41 }
```

(End definition for tree-struct-still-open. This function is documented on page ??.)

3.3 Attributes

Not much yet, as attributes aren't used so much.

attr-unknown

```
42 \msg_new:nnn { tag } {attr-unknown} { attribute-#1~is~unknown}
```

(End definition for attr-unknown. This function is documented on page 18.)

3.4 Roles

role-missing Warning message if either the tag or the role is missing

```
43 \msg_new:nnn { tag } {role-missing} { tag-#1~has~no~role~assigned }
```

```
44 \msg_new:nnn { tag } {role-unknown} { role-#1~is~not~known }
```

```
45 \msg_new:nnn { tag } {role-unknown-tag} { tag-#1~is~not~known }
```

(End definition for role-missing, role-unknown, and role-unknown-tag. These functions are documented on page 18.)

role-parent-child This is info and warning message about the containment rules between child and parent tags.

```
46 \msg_new:nnn { tag } {role-parent-child}
47 { The~rule~between~parent~'#1'~\and~child~'#2'~is~#3}
```

(End definition for role-parent-child. This function is documented on page ??.)

role-parent-child This is info and warning message about the containment rules between child and parent tags.

```
48 \msg_new:nnn { tag } {role-remapping}
49 { remapping~tag~to~#1 }
```

(End definition for role-parent-child. This function is documented on page ??.)

role-tag Info messages.

```
new-tag 50 \msg_new:nnn { tag } {role-tag}          { mapping~tag~#1~to~role~#2  }
51 \msg_new:nnn { tag } {new-tag}             { adding~new~tag~#1  }
52 \msg_new:nnn { tag } {read-namespace}      { reading~namespace~definitions~tagpdf~ns~
#1.def  }
53 \msg_new:nnn { tag } {namespace-missing}{ namespace~definitions~tagpdf~ns~#1.def~not~found }
54 \msg_new:nnn { tag } {namespace-unknown}{ namespace~#1~is~not~declared }
```

(End definition for role-tag and new-tag. These functions are documented on page 18.)

3.5 Miscellaneous

tree-mcid-index-wrong Used in the tree code, typically indicates the document must be rerun.

```
55 \msg_new:nnn { tag } {tree-mcid-index-wrong}
56   {something-is-wrong-with-the-mcid--rerun}
```

(End definition for tree-mcid-index-wrong. This function is documented on page 19.)

sys-no-interwordspace Currently only pdf_latex and lua_latex have some support for real spaces.

```
57 \msg_new:nnn { tag } {sys-no-interwordspace}
58   {engine/output-mode~#1~doesn't~support~the~interword~spaces}
```

(End definition for sys-no-interwordspace. This function is documented on page 19.)

__tag_check_typeout_v:n A simple logging function. By default it gobbles its argument, but the log-keys sets it to typeout.

```
59 \cs_set_eq:NN \__tag_check_typeout_v:n \use_none:n
```

(End definition for __tag_check_typeout_v:n.)

para-hook-count-wrong At the end of the document we check if the count of para-begin and para-end is identical. If not we issue a warning: this is normally a coding error and breaks the structure.

```
60 \msg_new:nnnn { tag } {para-hook-count-wrong}
61   {The~number~of~automatic~begin~(#1)~and~end~(#2)~para~hooks~differ!}
62   {This~quite~probably~a~coding~error~and~the~structure~will~be~wrong!}
63 \package
```

(End definition for para-hook-count-wrong. This function is documented on page 19.)

4 Retrieving data

\tag_get:n This retrieves some data. This is a generic command to retrieve data. Currently the only sensible values for the argument are mc_tag, struct_tag and struct_num.

```
64 \base\cs_new:Npn \tag_get:n #1 { \use:c {__tag_get_data_#1: } }
```

(End definition for \tag_get:n. This function is documented on page 16.)

5 User conditionals

`\tag_if_active_p:` This is a test if tagging is active. This allows packages to add conditional code. The test is true if all booleans, the global and the two local one are true.
`\tag_if_active:TF`

```

65 <*base>
66 \prg_new_conditional:Npnn \tag_if_active: { p , T , TF, F }
67 { \prg_return_false: }
68 </base>
69 <*package>
70 \prg_set_conditional:Npnn \tag_if_active: { p , T , TF, F }
71 {
72   \bool_lazy_all:nTF
73   {
74     {\g__tag_active_struct_bool}
75     {\g__tag_active_mc_bool}
76     {\g__tag_active_tree_bool}
77     {\l__tag_active_struct_bool}
78     {\l__tag_active_mc_bool}
79   }
80   {
81     \prg_return_true:
82   }
83   {
84     \prg_return_false:
85   }
86 }
```

(End definition for `\tag_if_active:TF`. This function is documented on page 16.)

6 Internal checks

These are checks used in various places in the code.

6.1 checks for active tagging

```

\__tag_check_if_active_mc:TF Structures must have a tag, so we check if the S entry is in the property. It is an error if
\__tag_check_if_active_struct:TF this is missing. The argument is a number.
87 \prg_new_conditional:Npnn \__tag_check_if_active_mc: {T,F,TF}
88 {
89   \bool_lazy_and:nnTF { \g__tag_active_mc_bool } { \l__tag_active_mc_bool }
90   {
91     \prg_return_true:
92   }
93   {
94     \prg_return_false:
95   }
96 }
97 \prg_new_conditional:Npnn \__tag_check_if_active_struct: {T,F,TF}
98 {
99   \bool_lazy_and:nnTF { \g__tag_active_struct_bool } { \l__tag_active_struct_bool }
100   {
101     \prg_return_true:

```

```

102     }
103     {
104         \prg_return_false:
105     }
106 }

```

(End definition for `__tag_check_if_active_mc:TF` and `__tag_check_if_active_struct:TF`.)

6.2 Checks related to structures

`__tag_check_structure_has_tag:n`

Structures must have a tag, so we check if the S entry is in the property. It is an error if this is missing. The argument is a number. The tests for existence and type is split in structures, as the tags are stored differently to the mc case.

```

107 \cs_new_protected:Npn \__tag_check_structure_has_tag:n #1 %#1 struct num
108 {
109     \prop_if_in:cnF { g__tag_struct_#1_prop }
110     {S}
111     {
112         \msg_error:nn { tag } {struct-missing-tag}
113     }
114 }

```

(End definition for `__tag_check_structure_has_tag:n`.)

`__tag_check_structure_tag:N`

This checks if the name of the tag is known, either because it is a standard type or has been rolemapped.

```

115 \cs_new_protected:Npn \__tag_check_structure_tag:N #1
116 {
117     \prop_if_in:NoF \g__tag_role_tags_NS_prop {#1}
118     {
119         \msg_warning:nnx { tag } {role-unknown-tag} {#1}
120     }
121 }

```

(End definition for `__tag_check_structure_tag:N`.)

`__tag_check_info_closing_struct:n`

This info message is issued at a closing structure, the use should be guarded by log-level.

```

122 \cs_new_protected:Npn \__tag_check_info_closing_struct:n #1 %#1 struct num
123 {
124     \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
125     {
126         \msg_info:nnn { tag } {struct-show-closing} {#1}
127     }
128 }
129
130 \cs_generate_variant:Nn \__tag_check_info_closing_struct:n {o,x}

```

(End definition for `__tag_check_info_closing_struct:n`.)

`__tag_check_no_open_struct:`

This checks if there is an open structure. It should be used when trying to close a structure. It errors if false.

```

131 \cs_new_protected:Npn \__tag_check_no_open_struct:
132 {
133     \msg_error:nn { tag } {struct-faulty-nesting}
134 }

```


(End definition for `_tag_check_no_open_struct:`.)

`_tag_check_struct_used:n` This checks if a stashed structure has already been used.

```

135 \cs_new_protected:Npn \_tag\_check\_struct\_used:n #1 %#1 label
136 {
137   \prop\_get:cnNT
138     {g\_tag\_struct\_\_tag\_ref\_value:enn{tagpdfstruct-#1}{tagstruct}{unknown}\_prop}
139     {P}
140     \l\_tmpa\_tl
141     {
142       \msg\_warning:nnn { tag } {struct-used-twice} {#1}
143     }
144 }
```

(End definition for `_tag_check_struct_used:n`.)

6.3 Checks related to roles

`_tag_check_add_tag_role:nn` This check is used when defining a new role mapping.

```

145 \cs_new_protected:Npn \_tag\_check\_add\_tag\_role:nn #1 #2 %#1 tag, #2 role
146 {
147   \tl\_if\_empty:nTF {#2}
148   {
149     \msg\_error:nnn { tag } {role-missing} {#1}
150   }
151   {
152     \prop\_get:NnNTF \g\_tag\_role\_tags\_NS\_prop {#2} \l\_tmpa\_tl
153     {
154       \int\_compare:nNnT {\l\_tag\_loglevel\_int} > { 0 }
155       {
156         \msg\_info:nnnn { tag } {role-tag} {#1} {#2}
157       }
158     }
159     {
160       \msg\_error:nnn { tag } {role-unknown} {#2}
161     }
162   }
163 }
```

Similar with a namespace

```

164 \cs_new_protected:Npn \_tag\_check\_add\_tag\_role:nnn #1 #2 #3 %#1 tag/NS, #2 role #3 namespace
165 {
166   \tl\_if\_empty:nTF {#2}
167   {
168     \msg\_error:nnn { tag } {role-missing} {#1}
169   }
170   {
171     \prop\_get:cnNTF { g\_tag\_role\_NS\_#3\_prop } {#2} \l\_tmpa\_tl
172     {
173       \int\_compare:nNnT {\l\_tag\_loglevel\_int} > { 0 }
174       {
175         \msg\_info:nnnn { tag } {role-tag} {#1} {#2/#3}
176       }
177     }
178   }
179 }
```

```

178         {
179             \msg_error:nnn { tag } {role-unknown} {#2/#3}
180         }
181     }
182 }

```

(End definition for __tag_check_add_tag_role:nn.)

6.4 Check related to mc-chunks

__tag_check_mc_if_nested: Two tests if a mc is currently open. One for the true (for begin code), one for the false part (for end code).

```

183 \cs_new_protected:Npn \__tag_check_mc_if_nested:
184 {
185     \__tag_mc_if_in:T
186     {
187         \msg_warning:nnx { tag } {mc-nested} { \__tag_get_mc_abs_cnt: }
188     }
189 }
190
191 \cs_new_protected:Npn \__tag_check_mc_if_open:
192 {
193     \__tag_mc_if_in:F
194     {
195         \msg_warning:nnx { tag } {mc-not-open} { \__tag_get_mc_abs_cnt: }
196     }
197 }

```

(End definition for __tag_check_mc_if_nested: and __tag_check_mc_if_open:.)

__tag_check_mc_pushed_popped:nn This creates an information message if mc's are pushed or popped. The first argument is a word (pushed or popped), the second the tag name. With larger log-level the stack is shown too.

```

198 \cs_new_protected:Npn \__tag_check_mc_pushed_popped:nn #1 #2
199 {
200     \int_compare:nNnT
201     { \l__tag_loglevel_int } = { 2 }
202     { \msg_info:nnx {tag}{mc-#1}{#2} }
203     \int_compare:nNnT
204     { \l__tag_loglevel_int } > { 2 }
205     {
206         \msg_info:nnx {tag}{mc-#1}{#2}
207         \seq_log:N \g__tag_mc_stack_seq
208     }
209 }

```

(End definition for __tag_check_mc_pushed_popped:nn.)

__tag_check_mc_tag:N This checks if the mc has a (known) tag.

```

210 \cs_new_protected:Npn \__tag_check_mc_tag:N #1 % #1 is var with a tag name in it
211 {
212     \tl_if_empty:NT #1
213     {
214         \msg_error:nnx { tag } {mc-tag-missing} { \__tag_get_mc_abs_cnt: }

```

```

215     }
216     \prop_if_in:NoF \g__tag_role_tags_NS_prop {#1}
217     {
218         \msg_warning:nnx { tag } {role-unknown-tag} {#1}
219     }
220 }

```

(End definition for __tag_check_mc_tag:N.)

\g__tag_check_mc_used_intarray
__tag_check_init_mc_used:

This variable holds the list of used mc numbers. Everytime we store a mc-number we will add one the relevant array index. If everything is right at the end there should be only 1 until the max count of the mcid. 2 indicates that one mcid was used twice, 0 that we lost one. In engines other than luatex the total number of all intarray entries are restricted so we use only a rather small value of 65536, and we initialize the array only at first used, guarded by the log-level. This check is probably only needed for debugging. TODO does this really make sense to check? When can it happen??

```

221 \cs_new_protected:Npn \__tag_check_init_mc_used:
222 {
223     \intarray_new:Nn \g__tag_check_mc_used_intarray { 65536 }
224     \cs_gset_eq:NN \__tag_check_init_mc_used: \prg_do_nothing:
225 }

```

(End definition for \g__tag_check_mc_used_intarray and __tag_check_init_mc_used:.)

__tag_check_mc_used:n

This checks if a mc is used twice.

```

226 \cs_new_protected:Npn \__tag_check_mc_used:n #1 %#1 mcid abscent
227 {
228     \int_compare:nNnT {\l__tag_loglevel_int} > { 2 }
229     {
230         \__tag_check_init_mc_used:
231         \intarray_gset:Nnn \g__tag_check_mc_used_intarray
232             {#1}
233             { \intarray_item:Nn \g__tag_check_mc_used_intarray {#1} + 1 }
234         \int_compare:nNnT
235             {
236                 \intarray_item:Nn \g__tag_check_mc_used_intarray {#1}
237             }
238             >
239             { 1 }
240             {
241                 \msg_warning:nnn { tag } {mc-used-twice} {#1}
242             }
243     }
244 }

```

(End definition for __tag_check_mc_used:n.)

__tag_check_show_MCID_by_page:

This allows to show the mc on a page. Currently unused.

```

245 \cs_new_protected:Npn \__tag_check_show_MCID_by_page:
246 {
247     \tl_set:Nx \l__tag_tmpa_tl
248     {
249         \__tag_ref_value_lastpage:nn
250         {abspage}

```

```

251         {-1}
252     }
253     \int_step_inline:nnnn {1}{1}
254     {
255         \l__tag_tmpa_tl
256     }
257     {
258         \seq_clear:N \l_tmpa_seq
259         \int_step_inline:nnnn
260         {1}
261         {1}
262         {
263             \__tag_ref_value_lastpage:nn
264             {tagmcabs}
265             {-1}
266         }
267         {
268             \int_compare:nT
269             {
270                 \__tag_ref_value:enn
271                 {mcid-###1}
272                 {tagabspage}
273                 {-1}
274                 =
275                 ##1
276             }
277             {
278                 \seq_gput_right:Nx \l_tmpa_seq
279                 {
280                     Page##1-###1-
281                     \__tag_ref_value:enn
282                     {mcid-###1}
283                     {tagmcid}
284                     {-1}
285                 }
286             }
287         }
288         \seq_show:N \l_tmpa_seq
289     }
290 }

```

(End definition for `__tag_check_show_MCID_by_page:.`)

6.5 Checks related to the state of MC on a page or in a split stream

The following checks are currently only usable in generic mode as they rely on the marks defined in the `mc-generic` module. They are used to detect if a `mc-chunk` has been split by a page break or similar and additional end/begin commands are needed.

`__tag_check_mc_in_galley_p:` At first we need a test to decide if `\tag_mc_begin:n` (`tmb`) and `\tag_mc_end:` (`tme`) has been used at all on the current galley. As each command issues two slightly different marks we can do it by comparing firstmarks and botmarks. The test assumes that

`__tag_check_mc_in_galley:` **TF**

the marks have been already mapped into the sequence with `\@@_mc_get_marks:`. As `\seq_if_eq:NNTF` doesn't exist we use the `tl-test`.

```

291 \prg_new_conditional:Npnn \__tag_check_if_mc_in_galley: { T,F,TF }
292 {
293   \tl_if_eq:NNTF \l__tag_mc_firstmarks_seq \l__tag_mc_botmarks_seq
294   { \prg_return_false: }
295   { \prg_return_true: }
296 }

```

(End definition for __tag_check_mc_in_galley:TF.)

`__tag_check_if_mc_tmb_missing_p:` This checks if a extra top mark (“extra-tmb”) is needed. According to the analysis this the case if the firstmarks start with `e-` or `b+`. Like above we assume that the marks content is already in the seq's.

`__tag_check_if_mc_tmb_missing:TF`

```

297 \prg_new_conditional:Npnn \__tag_check_if_mc_tmb_missing: { T,F,TF }
298 {
299   \bool_if:nTF
300   {
301     \str_if_eq_p:ee {\seq_item:Nn \l__tag_mc_firstmarks_seq {1}}{e-}
302     ||
303     \str_if_eq_p:ee {\seq_item:Nn \l__tag_mc_firstmarks_seq {1}}{b+}
304   }
305   { \prg_return_true: }
306   { \prg_return_false: }
307 }

```

(End definition for __tag_check_if_mc_tmb_missing:TF.)

`__tag_check_if_mc_tme_missing_p:` This checks if a extra bottom mark (“extra-tme”) is needed. According to the analysis this the case if the botmarks starts with `b+`. Like above we assume that the marks content is already in the seq's.

`__tag_check_if_mc_tme_missing:TF`

```

308 \prg_new_conditional:Npnn \__tag_check_if_mc_tme_missing: { T,F,TF }
309 {
310   \str_if_eq:eeTF {\seq_item:Nn \l__tag_mc_botmarks_seq {1}}{b+}
311   { \prg_return_true: }
312   { \prg_return_false: }
313 }

```

(End definition for __tag_check_if_mc_tme_missing:TF.)

314 `</package>`

315 `<*debug>`

Code for `tagpdf-debug`. This will probably change over time. At first something for the `mc` commands.

```

316 \msg_new:nnn { tag / debug } {mc-begin} { MC~begin~#1~with~options~\tl_to_str:n{#2}~[\msg_line_no] }
317 \msg_new:nnn { tag / debug } {mc-end} { MC~end~#1~[\msg_line_context:] }
318
319 \cs_new_protected:Npn \__tag_debug_mc_begin_insert:n #1
320 {
321   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
322   {
323     \msg_note:nnnn { tag / debug } {mc-begin} {inserted} { #1 }
324   }

```

```

325 }
326 \cs_new_protected:Npn \__tag_debug_mc_begin_ignore:n #1
327 {
328   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
329   {
330     \msg_note:nnnn { tag / debug } {mc-begin } {ignored} { #1 }
331   }
332 }
333 \cs_new_protected:Npn \__tag_debug_mc_end_insert:
334 {
335   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
336   {
337     \msg_note:nnn { tag / debug } {mc-end} {inserted}
338   }
339 }
340 \cs_new_protected:Npn \__tag_debug_mc_end_ignore:
341 {
342   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
343   {
344     \msg_note:nnn { tag / debug } {mc-end } {ignored}
345   }
346 }

```

And now something for the structures

```

347 \msg_new:nnn { tag / debug } {struct-begin}
348 {
349   Struct~\tag_get:n{struct_num}~begin~#1~with~options::~~\tl_to_str:n{#2}~[\msg_line_context:]
350 }
351 \msg_new:nnn { tag / debug } {struct-end}
352 {
353   Struct~end~#1~[\msg_line_context:]
354 }
355
356 \cs_new_protected:Npn \__tag_debug_struct_begin_insert:n #1
357 {
358   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
359   {
360     \msg_note:nnnn { tag / debug } {struct-begin} {inserted} { #1 }
361     \seq_log:N \g__tag_struct_tag_stack_seq
362   }
363 }
364 \cs_new_protected:Npn \__tag_debug_struct_begin_ignore:n #1
365 {
366   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
367   {
368     \msg_note:nnnn { tag / debug } {struct-begin } {ignored} { #1 }
369   }
370 }
371 \cs_new_protected:Npn \__tag_debug_struct_end_insert:
372 {
373   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
374   {
375     \msg_note:nnn { tag / debug } {struct-end} {inserted}
376     \seq_log:N \g__tag_struct_tag_stack_seq
377   }

```

```
378 }
379 \cs_new_protected:Npn \__tag_debug_struct_end_ignore:
380 {
381   \int_compare:nNnT { \l__tag_loglevel_int } > {0}
382   {
383     \msg_note:nnn { tag / debug } {struct-end } {ignored}
384   }
385 }
386 </debug>
```

Part II

The **tagpdf-user** module

Code related to L^AT_EX2e user commands and document commands

Part of the tagpdf package

1 Setup commands

<code>\tagpdfsetup</code>	<code>\tagpdfsetup{<key val list>}</code>
---------------------------	---

This is the main setup command to adapt the behaviour of tagpdf. It can be used in the preamble and in the document (but not all keys make sense there).

<code>activate_<setup-key></code>

And additional setup key which combine the other activate keys `activate-mc`, `activate-tree`, `activate-struct` and additionally add a document structure.

<code>\tagpdfifluatexTF</code>
<code>\tagpdfifluatexT</code>
<code>\tagpdfifpdfTeX</code>

small wrappers around engine tests. This functions should not be used and will be removed in one of the next versions.

2 Commands related to mc-chunks

<code>\tagmcbegin</code>	<code>\tagmcbegin {<key-val>}</code>
<code>\tagmcend</code>	<code>\tagmcend</code>
<code>\tagmcuse</code>	<code>\tagmcuse{<label>}</code>

These are wrappers around `\tag_mc_begin:n`, `\tag_mc_end:` and `\tag_mc_use:n`. The commands and their argument are documented in the `tagpdf-mc` module. In difference to the `expl3` commands, `\tagmcbegin` issues also an `\ignorespaces`, and `\tagmcend` will issue in horizontal mode an `\unskip`.

<code>\tagmcifinTF</code>	<code>\tagmcifin {<true code>}{<false code>}</code>
---------------------------	---

This is a wrapper around `\tag_mc_if_in:TF`. and tests if an mc is open or not. It is mostly of importance for pdf_lat_ex as lua_lat_ex doesn't mind much if a mc tag is not correctly closed. Unlike the `expl3` command it is not expandable.

The command is probably not of much use and will perhaps disappear in future versions. It normally makes more sense to push/pop an mc-chunk.

3 Commands related to structures

<code>\tagstructbegin</code>	<code>\tagstructbegin {⟨key-val⟩}</code>
<code>\tagstructend</code>	<code>\tagstructend</code>
<code>\tagstructuse</code>	<code>\tagstructuse{⟨label⟩}</code>

These are direct wrappers around `\tag_struct_begin:n`, `\tag_struct_end:` and `\tag_struct_use:n`. The commands and their argument are documented in the `tagpdf-struct` module.

4 Debugging

<code>\ShowTagging</code>	<code>\ShowTagging {⟨key-val⟩}</code>
---------------------------	---------------------------------------

This is a generic function to output various debugging helps. It not necessarily stops the compilation. The keys and their function are described below.

<code>mc-data_⟨⟩(show-key)</code>	<code>mc-data = ⟨number⟩</code>
-----------------------------------	---------------------------------

This key is (currently?) relevant for lua mode only. It shows the data of all mc-chunks created so far. It is accurate only after shipout (and perhaps a second compilation), so typically should be issued after a newpage. The value is a positive integer and sets the first mc-shown. If no value is given, 1 is used and so all mc-chunks created so far are shown.

<code>mc-current_⟨⟩(show-key)</code>	<code>mc-current</code>
--------------------------------------	-------------------------

This key shows the number and the tag of the currently open mc-chunk. If no chunk is open it shows only the state of the abs count. It works in all mode, but the output in luamode looks different.

<code>mc-marks_⟨⟩(show-key)</code>	<code>mc-marks = show use</code>
------------------------------------	----------------------------------

This key helps to debug the page marks. It should only be used at shipout in header or footer.

<code>struct-stack_⟨⟩(show-key)</code>	<code>struct-stack = log show</code>
--	--------------------------------------

This key shows the current structure stack. With `log` the info is only written to the log-file, `show` stops the compilation and shows on the terminal. If no value is used, then the default is `show`.

5 Extension commands

The following commands and code parts are not core command of tagpdf. They either provide work-arounds for missing functionality elsewhere, or do a first step to apply tagpdf commands to document commands.

The commands and keys should be view as experimental!

This part will be regularly revisited to check if the code should go to a better place or can be improved and so can change easily.

5.1 Fake space

`\pdffakespace` (lua-only) This provides a lua-version of the `\pdffakespace` primitive of pdftex.

5.2 Paratagging

This is a first try to make use of the new paragraph hooks in a current LaTeX to automate the tagging of paragraph. It requires sane paragraph nesting, faulty code, e.g. a missing `\par` at the end of a low-level vbox can highly confuse the tagging. The tags should be carefully checked if this is used.

<code>paratagging_␣(setup-key)</code>	<code>paratagging = true false</code>
<code>paratagging-show_␣(setup-key)</code>	<code>paratagging-show = true false</code>

This keys can be used in `\tagpdfsetup` and enable/disable paratagging. `paratagging-show` puts small red numbers at the begin and end of a paragraph. This is meant as a debugging help. The number are boxes and have a (tiny) height, so they can affect typesetting.

<code>\tagpdfparaOn</code>	These commands allow to enable/disable para tagging too and are a bit faster then <code>\tagpdfsetup</code> . But I'm not sure if the names are good.
<code>\tagpdfparaOff</code>	

<code>\tagpdfsuppressmarks</code>	This command allows to suppress the creation of the marks. It takes an argument which should normally be one of the mc-commands, puts a group around it and suppress the marks creation in this group. This command should be used if the begin and end command are at different boxing levels. E.g.
-----------------------------------	--

```
\@hangfrom
{
  \tagstructbegin{tag=H1}%
  \tagmcbegin    {tag=H1}%
  #2
}
{#3\tagpdfsuppressmarks{\tagmcend}\tagstructend}%
```

5.3 Header and footer

Header and footer are automatically excluded from tagging. This can be disabled with the following key. If some real content is in the header and footer, tagging must be restarted there explicitly. The key accepts the values `true` which surrounds the header with an artifact mc-chunk, `false` which disables the automatic tagging, and `pagination` which additionally adds an artifact structure with an pagination attribute.

<code>exclude-header-footer_␣(setup-key)</code>	<code>exclude-header-footer = true false pagination</code>
---	--

5.4 Link tagging

Links need a special structure and cross reference system. This is added through hooks of the l3pdfannot module and will work automatically if tagging is activated.

Links should (probably) have an alternative text in the Contents key. It is unclear which text this should be and how to get it. Currently the code simply adds the fix texts `url` and `ref`. Another text can be added by changing the dictionary value:

```
\pdfannot_dict_put:nnn
{ link/GoTo }
{ Contents }
{ (ref) }
```

6 User commands and extensions of document commands

```
1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-user} {2022-12-22} {0.98}
4   {tagpdf - user commands}
5 </header>
```

7 Setup and preamble commands

```
\tagpdfsetup
6 (base)\NewDocumentCommand \tagpdfsetup { m }{}
7 <*package>
8 \RenewDocumentCommand \tagpdfsetup { m }
9   {
10     \keys_set:nn { __tag / setup } { #1 }
11   }
12 </package>
```

(End definition for \tagpdfsetup. This function is documented on page 32.)

8 Commands for the mc-chunks

```
\tagmcbegin
\tagmcend
\tagmcuse
13 <*base>
14 \NewDocumentCommand \tagmcbegin { m }
15   {
16     \tag_mc_begin:n {#1}
17   }
18
19
20 \NewDocumentCommand \tagmcend { }
21   {
22     \tag_mc_end:
23   }
24
25 \NewDocumentCommand \tagmcuse { m }
```

```

26 {
27   \tag_mc_use:n {#1}
28 }
29 \</base>

```

(End definition for \tagmcbegin, \tagmcend, and \tagmcuse. These functions are documented on page 32.)

\tagmcifinTF This is a wrapper around \tag_mc_if_in: and tests if an mc is open or not. It is mostly of importance for pdf_lat_ex as lua_lat_ex doesn't mind much if a mc tag is not correctly closed. Unlike the expl3 command it is not expandable.

```

30 \<package>
31 \NewDocumentCommand \tagmcifinTF { m m }
32 {
33   \tag_mc_if_in:TF { #1 } { #2 }
34 }
35 \</package>

```

(End definition for \tagmcifinTF. This function is documented on page 32.)

9 Commands for the structure

\tagstructbegin **\tagstructend** **\tagstructuse** These are structure related user commands. There are direct wrapper around the expl3 variants.

```

36 \<base>
37 \NewDocumentCommand \tagstructbegin { m }
38 {
39   \tag_struct_begin:n {#1}
40 }
41
42 \NewDocumentCommand \tagstructend { }
43 {
44   \tag_struct_end:
45 }
46
47 \NewDocumentCommand \tagstructuse { m }
48 {
49   \tag_struct_use:n {#1}
50 }
51 \</base>

```

(End definition for \tagstructbegin, \tagstructend, and \tagstructuse. These functions are documented on page 33.)

10 Debugging

\ShowTagging This is a generic command for various show commands. It takes a keyval list, the various keys are implemented below.

```

52 \<package>
53 \NewDocumentCommand\ShowTagging { m }
54 {
55   \keys_set:nn { __tag / show }{ #1}

```

```

56
57 }

```

(End definition for \ShowTagging. This function is documented on page 33.)

mc-data_␣(show-key) This key is (currently?) relevant for lua mode only. It shows the data of all mc-chunks created so far. It is accurate only after shipout, so typically should be issued after a newpage. With the optional argument the minimal number can be set.

```

58 \keys_define:nn { __tag / show }
59 {
60   mc-data .code:n =
61   {
62     \sys_if_engine luatex:T
63     {
64       \lua_now:e{!tx.__tag.trace.show_all_mc_data(#1,\__tag_get_mc_abs_cnt:,0)}
65     }
66   }
67   ,mc-data .default:n = 1
68 }
69

```

(End definition for mc-data (show-key). This function is documented on page 33.)

mc-current_␣(show-key) This shows some info about the current mc-chunk. It works in generic and lua-mode.

```

70 \keys_define:nn { __tag / show }
71 { mc-current .code:n =
72   {
73     \bool_if:NTF \g__tag_mode_lua_bool
74     {
75       \sys_if_engine luatex:T
76       {
77         \int_compare:nNnTF
78         { -2147483647 }
79         =
80         {
81           \lua_now:e
82           {
83             tex.print
84             (tex.getattribute
85              (luatexbase.attributes.g__tag_mc_cnt_attr))
86           }
87         }
88         {
89           \lua_now:e
90           {
91             ltx.__tag.trace.log
92             (
93               "mc-current:~no~MC~open,~current~absent
94               =\__tag_get_mc_abs_cnt:"
95               ,0
96             )
97             texio.write_nl("")
98           }
99         }

```

```

100     {
101         \lua_now:e
102         {
103             ltx.__tag.trace.log
104             (
105                 "mc-current:~absent=~\__tag_get_mc_abs_cnt:=="
106                 ..
107                 tex.getattribute(luatexbase.attributes.g__tag_mc_cnt_attr)
108                 ..
109                 "~=>tag="
110                 ..
111                 tostring
112                 (ltx.__tag.func.get_tag_from
113                  (tex.getattribute
114                   (luatexbase.attributes.g__tag_mc_type_attr)))
115                 ..
116                 "="
117                 ..
118                 tex.getattribute
119                 (luatexbase.attributes.g__tag_mc_type_attr)
120                 ,0
121             )
122             texio.write_nl("")
123         }
124     }
125 }
126
127 {
128     \msg_note:nn{ tag }{ mc-current }
129 }
130 }
131 }

```

(End definition for `mc-current` (`show-key`). This function is documented on page 33.)

mc-marks_␣(show-key) It maps the mc-marks into the sequences and then shows them. This allows to inspect the first and last mc-Mark on a page. It should only be used in the shipout (header/footer).

```

132 \keys_define:nn { __tag / show }
133 {
134     mc-marks .choice: ,
135     mc-marks / show .code:n =
136     {
137         \__tag_mc_get_marks:
138         \__tag_check_if_mc_in_galley:TF
139         {
140             \iow_term:n {Marks~from~this~page:~}
141         }
142         {
143             \iow_term:n {Marks~from~a~previous~page:~}
144         }
145         \seq_show:N \l__tag_mc_firstmarks_seq
146         \seq_show:N \l__tag_mc_botmarks_seq
147         \__tag_check_if_mc_tmb_missing:T
148         {

```

```

149         \iow_term:n {BDC~missing-on~this~page!}
150     }
151     \__tag_check_if_mc_tme_missing:T
152     {
153         \iow_term:n {EMC~missing-on~this~page!}
154     }
155 },
156 mc-marks / use .code:n =
157 {
158     \__tag_mc_get_marks:
159     \__tag_check_if_mc_in_galley:TF
160     { Marks~from~this~page:~}
161     { Marks~from~a~previous~page:~}
162     \seq_use:Nn \l__tag_mc_firstmarks_seq {,~}\quad
163     \seq_use:Nn \l__tag_mc_botmarks_seq {,~}\quad
164     \__tag_check_if_mc_tmb_missing:T
165     {
166         BDC~missing~
167     }
168     \__tag_check_if_mc_tme_missing:T
169     {
170         EMC~missing
171     }
172 },
173 mc-marks .default:n = show
174 }

```

(End definition for mc-marks (show-key). This function is documented on page 33.)

struct-stack_␣(show-key)

```

175 \keys_define:nn { __tag / show }
176 {
177     struct-stack .choice:
178     ,struct-stack / log .code:n = \seq_log:N \g__tag_struct_tag_stack_seq
179     ,struct-stack / show .code:n = \seq_show:N \g__tag_struct_tag_stack_seq
180     ,struct-stack .default:n = show
181 }

```

(End definition for struct-stack (show-key). This function is documented on page 33.)

11 Commands to extend document commands

The following commands and code parts are not core command of tagpdf. The either provide work arounds for missing functionality elsewhere, or do a first step to apply tagpdf commands to document commands. This part should be regularly revisited to check if the code should go to a better place or can be improved.

11.1 Document structure

```

\__tag_add_document_structure:n
activate␣(setup-key) 182 \cs_new_protected:Npn \__tag_add_document_structure:n #1
183 {
184     \hook_gput_code:nnn{begindocument}{tagpdf}{\tagstructbegin{tag=#1}}

```

```

185   \hook_gput_code:nnn{tagpdf/finish/before}{tagpdf}{\tagstructend}
186 }
187 \keys_define:nn { __tag / setup}
188 {
189   activate .code:n =
190   {
191     \keys_set:nn { __tag / setup }
192     { activate-mc,activate-tree,activate-struct }
193     \__tag_add_document_structure:n {#1}
194   },
195   activate .default:n = Document
196 }

```

(End definition for `__tag_add_document_structure:n` and `activate` (setup-key). This function is documented on page 32.)

11.2 Structure destinations

In TeXlive 2022 pdfTeX and LuaTeX will offer support for structure destinations. The pdfmanagement has already backend support. We activate them if the prerequisites are there: structures should be activated, the code in the pdfmanagement must be there. Structure destinations are actually PDF 2.0 only but they don't harm in older PDF and can improve HTML export.

```

197 \AddToHook{begindocument/before}
198 {
199   \bool_lazy_all:nT
200   {
201     { \g__tag_active_struct_dest_bool }
202     { \g__tag_active_struct_bool }
203     { \cs_if_exist_p:N \pdf_activate_structure_destination: }
204   }
205   {
206     \tl_set:Nn \l_pdf_current_structure_destination_tl { __tag/struct/\g__tag_struct_stack
207     \pdf_activate_structure_destination:
208   }
209 }

```

11.3 Fake space

`\pdffakespace` We need a LuaTeX variant for `\pdffakespace`. This should probably go into the kernel at some time.

```

210 \sys_if_engine_luaTeX:T
211 {
212   \NewDocumentCommand\pdffakespace { }
213   {
214     \__tag_fakespace:
215   }
216 }

```

(End definition for `\pdffakespace`. This function is documented on page 34.)

11.4 Paratagging

The following are some simple commands to enable/disable paratagging. Probably one should add some checks if we are already in a paragraph.

```

\l__tag_para_bool
\l__tag_para_show_bool
\g__tag_para_int
\l__tag_para_tag_default_tl
217 \bool_new:N \l__tag_para_bool
218 \bool_new:N \l__tag_para_show_bool
219 \int_new:N \g__tag_para_begin_int
220 \int_new:N \g__tag_para_end_int
221 \tl_new:N \l__tag_para_tag_default_tl
222 \tl_set:Nn \l__tag_para_tag_default_tl { P }
223 \tl_new:N \l__tag_para_tag_tl
224 \tl_set:Nn \l__tag_para_tag_tl { \l__tag_para_tag_default_tl }

```

(End definition for `\l__tag_para_bool` and others.)

`paratagging_␣(setup-key)` These keys enable/disable locally paratagging, and the debug modus. It can affect the typesetting if `paratagging-show` is used. The small numbers are boxes and they have a (small) height.

```

225 \keys_define:nn { __tag / setup }
226 {
227   paratagging      .bool_set:N = \l__tag_para_bool,
228   paratagging-show .bool_set:N = \l__tag_para_show_bool,
229   paratag          .tl_set:N   = \l__tag_para_tag_tl
230 }
231

```

(End definition for `paratagging (setup-key)` and `paratagging-show (setup-key)`. These functions are documented on page 34.)

This fills the para hooks with the needed code.

```

232 \AddToHook{para/begin}
233 {
234   \bool_if:NT \l__tag_para_bool
235   {
236     \int_gincr:N \g__tag_para_begin_int
237     \tag_struct_begin:n {tag=\l__tag_para_tag_tl}
238     \bool_if:NT \l__tag_para_show_bool
239     { \tag_mc_begin:n{artifact}
240       \llap{\color_select:n{red}\tiny\int_use:N\g__tag_para_begin_int\ }
241       \tag_mc_end:
242     }
243     \tag_mc_begin:n {}
244   }
245 }
246 \AddToHook{para/end}
247 {
248   \bool_if:NT \l__tag_para_bool
249   {
250     \int_gincr:N \g__tag_para_end_int
251     \tag_mc_end:
252     \bool_if:NT \l__tag_para_show_bool
253     { \tag_mc_begin:n{artifact}
254       \rlap{\color_select:n{red}\tiny\int_use:N\g__tag_para_end_int}

```

```

255         \tag_mc_end:
256     }
257     \tag_struct_end:
258 }
259 }
260 \AddToHook{enddocument/info}
261 {
262     \int_compare:nNnF {\g__tag_para_begin_int}={\g__tag_para_end_int}
263     {
264         \msg_error:nnxx
265         {tag}
266         {para-hook-count-wrong}
267         {\int_use:N\g__tag_para_begin_int}
268         {\int_use:N\g__tag_para_end_int}
269     }
270 }

```

In generic mode we need the additional code from the ptagging tests.

```

271 \AddToHook{begindocument/before}
272 {
273     \bool_if:NF \g__tag_mode_lua_bool
274     {
275         \cs_if_exist:NT \@kernel@before@footins
276         {
277             \tl_put_right:Nn \@kernel@before@footins
278             { \__tag_add_missing_mcs_to_stream:Nn \footins {footnote} }
279             \tl_put_right:Nn \@kernel@before@cclv
280             {
281                 \__tag_check_typeout_v:n {====>~In~\token_to_str:N \@makecol\c_space_tl\the\c@
282                 \__tag_add_missing_mcs_to_stream:Nn \@cclv {main}
283             }
284             \tl_put_right:Nn \@mult@ptagging@hook
285             {
286                 \__tag_check_typeout_v:n {====>~In~\string\page@sofar}
287                 \process@cols\mult@firstbox
288                 {
289                     \__tag_add_missing_mcs_to_stream:Nn \count@ {multicol}
290                 }
291                 \__tag_add_missing_mcs_to_stream:Nn \mult@rightbox {multicol}
292             }
293         }
294     }
295 }
296 \end{package}

```

\tagpdfparaOn This two command switch para mode on and off. **\tagpdfsetup** could be used too but **\tagpdfparaOff** is longer.

```

297 \base\newcommand\tagpdfparaOn {}
298 \base\newcommand\tagpdfparaOff{}
299 *package}
300 \renewcommand\tagpdfparaOn {\bool_set_true:N \l__tag_para_bool}
301 \renewcommand\tagpdfparaOff{\bool_set_false:N \l__tag_para_bool}

```

(End definition for **\tagpdfparaOn** and **\tagpdfparaOff**. These functions are documented on page 34.)

`\tagpdfsuppressmarks` This command allows to suppress the creation of the marks. It takes an argument which should normally be one of the mc-commands, puts a group around it and suppress the marks creation in this group. This command should be used if the begin and end command are at different boxing levels. E.g.

```

\@hangfrom
{
  \tagstructbegin{tag=H1}%
  \tagmcbegin {tag=H1}%
  #2
}
{#3\tagpdfsuppressmarks{\tagmcend}\tagstructend}%

302 \NewDocumentCommand\tagpdfsuppressmarks{m}
303   {{\use:c{__tag_mc_disable_marks:} #1}}

```

(End definition for `\tagpdfsuppressmarks`. This function is documented on page 34.)

11.5 Header and footer

Header and footer should normally be tagged as artifacts. The following code requires the new hooks. For now we allow to disable this function, but probably the code should always there at the end. TODO check if Pagination should be changeable.

```

304 \cs_new_protected:Npn\__tag_hook_kernel_before_head:{}
305 \cs_new_protected:Npn\__tag_hook_kernel_after_head:{}
306 \cs_new_protected:Npn\__tag_hook_kernel_before_foot:{}
307 \cs_new_protected:Npn\__tag_hook_kernel_after_foot:{}
308
309 \AddToHook{begindocument}
310 {
311   \cs_if_exist:NT \@kernel@before@head
312   {
313     \tl_put_right:Nn \@kernel@before@head {\__tag_hook_kernel_before_head:}
314     \tl_put_left:Nn \@kernel@after@head {\__tag_hook_kernel_after_head:}
315     \tl_put_right:Nn \@kernel@before@foot {\__tag_hook_kernel_before_foot:}
316     \tl_put_left:Nn \@kernel@after@foot {\__tag_hook_kernel_after_foot:}
317   }
318 }
319
320 \bool_new:N \g__tag_saved_in_mc_bool
321 \cs_new_protected:Npn \__tag_exclude_headfoot_begin:
322 {
323   \bool_set_false:N \l__tag_para_bool
324   \bool_if:NTF \g__tag_mode_lua_bool
325   {
326     \tag_mc_end_push:
327   }
328   {
329     \bool_gset_eq:NN \g__tag_saved_in_mc_bool \g__tag_in_mc_bool
330     \bool_gset_false:N \g__tag_in_mc_bool
331   }
332   \tag_mc_begin:n {artifact}
333 }
334 \cs_new_protected:Npn \__tag_exclude_headfoot_end:

```

```

335 {
336     \tag_mc_end:
337     \bool_if:NTF \g__tag_mode_lua_bool
338     {
339         \tag_mc_begin_pop:n{ }
340     }
341     {
342         \bool_gset_eq:NN \g__tag_in_mc_bool\g__tag_saved_in_mc_bool
343     }
344 }

```

This version allows to use an Artifact structure

```

345 \__tag_attr_new_entry:nn {__tag/attr/pagination}{/0/Artifact/Type/Pagination}
346 \cs_new_protected:Npn \__tag_exclude_struct_headfoot_begin:n #1
347 {
348     \bool_set_false:N \l__tag_para_bool
349     \bool_if:NTF \g__tag_mode_lua_bool
350     {
351         \tag_mc_end_push:
352     }
353     {
354         \bool_gset_eq:NN \g__tag_saved_in_mc_bool \g__tag_in_mc_bool
355         \bool_gset_false:N \g__tag_in_mc_bool
356     }
357     \tag_struct_begin:n{tag=Artifact,attribute-class=__tag/attr/#1}
358     \tag_mc_begin:n {artifact=#1}
359 }
360
361 \cs_new_protected:Npn \__tag_exclude_struct_headfoot_end:
362 {
363     \tag_mc_end:
364     \tag_struct_end:
365     \bool_if:NTF \g__tag_mode_lua_bool
366     {
367         \tag_mc_begin_pop:n{ }
368     }
369     {
370         \bool_gset_eq:NN \g__tag_in_mc_bool\g__tag_saved_in_mc_bool
371     }
372 }

```

And now the keys

`exclude-header-footer␣(setup-key)`

```

373 \keys_define:nn { __tag / setup }
374 {
375     exclude-header-footer .choice:,
376     exclude-header-footer / true .code:n =
377     {
378         \cs_set_eq:NN \__tag_hook_kernel_before_head: \__tag_exclude_headfoot_begin:
379         \cs_set_eq:NN \__tag_hook_kernel_before_foot: \__tag_exclude_headfoot_begin:
380         \cs_set_eq:NN \__tag_hook_kernel_after_head: \__tag_exclude_headfoot_end:
381         \cs_set_eq:NN \__tag_hook_kernel_after_foot: \__tag_exclude_headfoot_end:
382     },
383     exclude-header-footer / pagination .code:n =

```

```

384     {
385         \cs_set:Nn \__tag_hook_kernel_before_head: { \__tag_exclude_struct_headfoot_begin:n {p
386         \cs_set:Nn \__tag_hook_kernel_before_foot: { \__tag_exclude_struct_headfoot_begin:n {p
387         \cs_set_eq:NN \__tag_hook_kernel_after_head: \__tag_exclude_struct_headfoot_end:
388         \cs_set_eq:NN \__tag_hook_kernel_after_foot: \__tag_exclude_struct_headfoot_end:
389     },
390     exclude-header-footer / false .code:n =
391     {
392         \cs_set_eq:NN \__tag_hook_kernel_before_head: \prg_do_nothing:
393         \cs_set_eq:NN \__tag_hook_kernel_before_foot: \prg_do_nothing:
394         \cs_set_eq:NN \__tag_hook_kernel_after_head: \prg_do_nothing:
395         \cs_set_eq:NN \__tag_hook_kernel_after_foot: \prg_do_nothing:
396     },
397     exclude-header-footer .default:n = true,
398     exclude-header-footer .initial:n = true
399 }

```

(End definition for `exclude-header-footer` (setup-key). This function is documented on page 34.)

11.6 Links

We need to close and reopen mc-chunks around links. Currently we handle URI and GoTo (internal) links. Links should have an alternative text in the Contents key. It is unclear which text this should be and how to get it.

```

400 \hook_gput_code:nnn
401 {pdfannot/link/URI/before}
402 {tagpdf}
403 {
404     \tag_mc_end_push:
405     \tag_struct_begin:n { tag=Link }
406     \tag_mc_begin:n { tag=Link }
407     \pdfannot_dict_put:nnx
408     { link/URI }
409     { StructParent }
410     { \tag_struct_parent_int: }
411 }
412
413 \hook_gput_code:nnn
414 {pdfannot/link/URI/after}
415 {tagpdf}
416 {
417     \tag_struct_insert_annot:xx {\pdfannot_link_ref_last:}{\tag_struct_parent_int:}
418     \tag_mc_end:
419     \tag_struct_end:
420     \tag_mc_begin_pop:n{ }
421 }
422
423 \hook_gput_code:nnn
424 {pdfannot/link/GoTo/before}
425 {tagpdf}
426 {
427     \tag_mc_end_push:
428     \tag_struct_begin:n{tag=Link}
429     \tag_mc_begin:n{tag=Link}

```

```

430     \pdfannot_dict_put:nnx
431     { link/GoTo }
432     { StructParent }
433     { \tag_struct_parent_int: }
434 }
435
436 \hook_gput_code:nnn
437 {pdfannot/link/GoTo/after}
438 {tagpdf}
439 {
440     \tag_struct_insert_annot:xx {\pdfannot_link_ref_last:}{\tag_struct_parent_int:}
441     \tag_mc_end:
442     \tag_struct_end:
443     \tag_mc_begin_pop:n{ }
444 }
445 }
446
447 % "alternative descriptions " for PAX3. How to get better text here??
448 \pdfannot_dict_put:nnn
449 { link/URI }
450 { Contents }
451 { (url) }
452
453 \pdfannot_dict_put:nnn
454 { link/GoTo }
455 { Contents }
456 { (ref) }
457
</package>

```

Part III

The tagpdf-tree module

Commands trees and main dictionaries

Part of the tagpdf package

```
1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-tree-code} {2022-12-22} {0.98}
4 {part of tagpdf - code related to writing trees and dictionaries to the pdf}
5 </header>
```

1 Trees, pdfmanagement and finalization code

The code to finish the structure is in a hook. This will perhaps at the end be a kernel hook. TODO check right place for the code The pdfmanagement code is the kernel hook after shipout/lastpage so all code affecting it should be before. Objects can be written later, at least in pdf mode.

```
6 <*package>
7 \hook_gput_code:nnn{begindocument}{tagpdf}
8 {
9   \bool_if:NT \g__tag_active_tree_bool
10   {
11     \sys_if_output_pdf:TF
12     {
13       \AddToHook{enddocument/end} { \__tag_finish_structure: }
14     }
15     {
16       \AddToHook{shipout/lastpage} { \__tag_finish_structure: }
17     }
18   }
19 }
```

1.1 Check structure

__tag_tree_final_checks:

```
20 \cs_new_protected:Npn \__tag_tree_final_checks:
21 {
22   \int_compare:nNnF {\seq_count:N\g__tag_struct_stack_seq}={1}
23   {
24     \msg_warning:nn {tag}{tree-struct-still-open}
25     \int_step_inline:nnn{2}{\seq_count:N\g__tag_struct_stack_seq}
26     {\tag_struct_end:}
27   }
28 }
```

(End definition for __tag_tree_final_checks:.)

1.2 Catalog: MarkInfo and StructTreeRoot

The StructTreeRoot and the MarkInfo entry must be added to the catalog. We do it late so that we can win, but before the pdfmanagement hook.

```
--tag/struct/0 This is the object for the root object, the StructTreeRoot
29 \pdf_object_new:n { __tag/struct/0 }
(End definition for __tag/struct/0.)
30 \hook_gput_code:nnn{shipout/lastpage}{tagpdf}
31 {
32   \bool_if:NT \g__tag_active_tree_bool
33   {
34     \pdfmanagement_add:nnn { Catalog / MarkInfo } { Marked } { true }
35     \pdfmanagement_add:nnx
36       { Catalog }
37       { StructTreeRoot }
38     { \pdf_object_ref:n { __tag/struct/0 } }
39   }
40 }
```

1.3 Writing structure elements

The following commands are needed to write out the structure.

```
\__tag_tree_write_structtreeroot: This writes out the root object.
41 \pdf_version_compare:NnTF < {2.0}
42 {
43   \cs_new_protected:Npn \__tag_tree_write_structtreeroot:
44   {
45     \__tag_prop_gput:cnx
46       { g__tag_struct_0_prop }
47       { ParentTree }
48     { \pdf_object_ref:n { __tag/tree/parenttree } }
49     \__tag_prop_gput:cnx
50       { g__tag_struct_0_prop }
51       { RoleMap }
52     { \pdf_object_ref:n { __tag/tree/rolemap } }
53     \__tag_struct_write_obj:n { 0 }
54   }
55 }
no RoleMap in pdf 2.0
56 {
57   \cs_new_protected:Npn \__tag_tree_write_structtreeroot:
58   {
59     \__tag_prop_gput:cnx
60       { g__tag_struct_0_prop }
61       { ParentTree }
62     { \pdf_object_ref:n { __tag/tree/parenttree } }
63     \__tag_struct_write_obj:n { 0 }
64   }
65 }
(End definition for \__tag_tree_write_structtreeroot:.)
```


`_tag_tree_write_struetelements:` This writes out the other struct elems, the absolute number is in the counter.

```

66 \cs_new_protected:Npn \_tag\_tree\_write\_struetelements:
67 {
68   \int_step_inline:nnnn {1}{1}{\c@g\_tag\_struct\_abs\_int}
69   {
70     \_tag\_struct\_write\_obj:n { ##1 }
71   }
72 }

```

(End definition for `_tag_tree_write_struetelements:.`)

1.4 ParentTree

`__tag/tree/parenttree` The object which will hold the parenttree

```

73 \pdf_object_new:n { __tag/tree/parenttree }

```

(End definition for `__tag/tree/parenttree.`)

The ParentTree maps numbers to objects or (if the number represents a page) to arrays of objects. The numbers refer to two distinct types of entries: page streams and real objects like annotations. The numbers must be distinct and ordered. So we rely on abspage for the pages and put the real objects at the end. We use a counter to have a chance to get the correct number if code is processed twice.

`\c@g_tag_parenttree_obj_int` This is a counter for the real objects. It starts at the absolute last page value. It relies on l3ref.

```

74 \newcounter { g\_tag\_parenttree\_obj\_int }
75 \hook_gput_code:nnn{begindocument}{tagpdf}
76 {
77   \int_gset:Nn
78   \c@g\_tag\_parenttree\_obj\_int
79   { \_tag\_ref\_value\_lastpage:nn{abspage}{100} }
80 }

```

(End definition for `\c@g_tag_parenttree_obj_int.`)

We store the number/object references in a tl-var. If more structure is needed one could switch to a seq.

`\g_tag_parenttree_objr_tl`

```

81 \tl_new:N \g\_tag\_parenttree\_objr\_tl

```

(End definition for `\g_tag_parenttree_objr_tl.`)

`_tag_parenttree_add_objr:nn` This command stores a StructParent number and a objref into the tl var. This is only for objects like annotations, pages are handled elsewhere.

```

82 \cs_new_protected:Npn \_tag\_parenttree\_add\_objr:nn #1 #2 %#1 StructParent number, #2 objref
83 {
84   \tl_gput_right:Nx \g\_tag\_parenttree\_objr\_tl
85   {
86     #1 \c\_space\_tl #2 ^^J
87   }
88 }

```

(End definition for `_tag_parenttree_add_objr:nn.`)

`\l__tag_parenttree_content_tl` A tl-var which will get the page related parenttree content.

89 `\tl_new:N \l__tag_parenttree_content_tl`

(End definition for `\l__tag_parenttree_content_tl`.)

`__tag_tree_fill_parenttree:` This is the main command to assemble the page related entries of the parent tree. It wanders through the pages and the mcid numbers and collects all mcid of one page.

90

91 `\cs_new_protected:Npn __tag_tree_fill_parenttree:`

92 {

93 `\int_step_inline:nnnn{1}{1}{__tag_ref_value_lastpage:nn{abspage}{-1}} %not quite clear in`

94 { %page ##1

95 `\prop_clear:N \l__tag_tmpa_prop`

96 `\int_step_inline:nnnn{1}{1}{__tag_ref_value_lastpage:nn{tagmcabs}{-1}}`

97 {

98 `%mcid###1`

99 `\int_compare:nT`

100 `{__tag_ref_value:enn{mcid-###1}{tagabspage}{-1}=##1} %mcid is on current page`

101 `{% yes`

102 `\prop_put:Nxx`

103 `\l__tag_tmpa_prop`

104 `{__tag_ref_value:enn{mcid-###1}{tagmcid}{-1}}`

105 `{\prop_item:Nn \g__tag_mc_parenttree_prop {###1}}`

106 `}`

107 `}`

108 `\tl_put_right:Nx\l__tag_parenttree_content_tl`

109 {

110 `\int_eval:n {##1-1}\c_space_tl`

111 `[\c_space_tl %]`

112 `}`

113 `\int_step_inline:nnnn`

114 `{0}`

115 `{1}`

116 `{ \prop_count:N \l__tag_tmpa_prop -1 }`

117 {

118 `\prop_get:NnNTF \l__tag_tmpa_prop {###1} \l__tag_tmpa_tl`

119 `{% page#1:mcid##1:\l__tag_tmpa_tl :content`

120 `\tl_put_right:Nx \l__tag_parenttree_content_tl`

121 {

122 `\pdf_object_if_exist:eT { __tag/struct/\l__tag_tmpa_tl }`

123 {

124 `\pdf_object_ref:e { __tag/struct/\l__tag_tmpa_tl }`

125 `}`

126 `\c_space_tl`

127 `}`

128 `}`

129 {

130 `\msg_warning:nn { tag } {tree-mcid-index-wrong}`

131 `}`

132 `}`

133 `\tl_put_right:Nn`

134 `\l__tag_parenttree_content_tl`

135 `{%[`

136 `]^^J`

```

137     }
138   }
139 }

```

(End definition for `_tag_tree_fill_parenttree:`)

`_tag_tree_lua_fill_parenttree:` This is a special variant for luatex. lua mode must/can do it differently.

```

140 \cs_new_protected:Npn \_tag_tree_lua_fill_parenttree:
141 {
142   \tl_set:Nn \l__tag_parenttree_content_tl
143   {
144     \lua_now:e
145     {
146       ltx.__tag.func.output_parenttree
147       (
148         \int_use:N\g_shipout_readonly_int
149       )
150     }
151   }
152 }

```

(End definition for `_tag_tree_lua_fill_parenttree:`)

`_tag_tree_write_parenttree:` This combines the two parts and writes out the object. TODO should the check for lua be moved into the backend code?

```

153 \cs_new_protected:Npn \_tag_tree_write_parenttree:
154 {
155   \bool_if:NTF \g__tag_mode_lua_bool
156   {
157     \_tag_tree_lua_fill_parenttree:
158   }
159   {
160     \_tag_tree_fill_parenttree:
161   }
162   \tl_put_right:NV \l__tag_parenttree_content_tl\g__tag_parenttree_objr_tl
163   \pdf_object_write:nmx {__tag/tree/parenttree}{dict}
164   {
165     /Nums\c_space_tl [\l__tag_parenttree_content_tl]
166   }
167 }

```

(End definition for `_tag_tree_write_parenttree:`)

1.5 Rolemap dictionary

The Rolemap dictionary describes relations between new tags and standard types. The main part here is handled in the role module, here we only define the command which writes it to the PDF.

`__tag/tree/rolemap` At first we reserve again an object.

```

168 \pdf_version_compare:NnT < {2.0}
169 {
170   \pdf_object_new:n {__tag/tree/rolemap}
171 }

```

(End definition for __tag/tree/rolemap.)

__tag_tree_write_rolemap: This writes out the rolemap, basically it simply pushes out the dictionary which has been filled in the role module.

```
172 \pdf_version_compare:NnTF < {2.0}
173 {
174   \cs_new_protected:Npn \__tag_tree_write_rolemap:
175   {
176     \prop_map_inline:Nn\g__tag_role_rolemap_prop
177     {
178       \pdfdict_gput:nxx {g__tag_role/RoleMap_dict}
179       {##1}
180       {\pdf_name_from_unicode_e:n{##2}}
181     }
182     \pdf_object_write:nxx {__tag/tree/rolemap }{dict}
183     {
184       \pdfdict_use:n{g__tag_role/RoleMap_dict}
185     }
186   }
187 }
188 {
189   \cs_new_protected:Npn \__tag_tree_write_rolemap:{ }
190 }
191
```

(End definition for __tag_tree_write_rolemap:.)

1.6 Classmap dictionary

Classmap and attributes are setup in the struct module, here is only the code to write it out. It should only done if values have been used.

__tag_tree_write_classmap:

```
192 \cs_new_protected:Npn \__tag_tree_write_classmap:
193 {
194   \tl_clear:N \l__tag_tmpa_tl
195   \seq_gremove_duplicates:N \g__tag_attr_class_used_seq
196   \seq_set_map:NNn \l__tag_tmpa_seq \g__tag_attr_class_used_seq
197   {
198     ##1\c_space_tl
199     <<
200     \prop_item:Nn
201     \g__tag_attr_entries_prop
202     {##1}
203     >>
204   }
205   \tl_set:Nx \l__tag_tmpa_tl
206   {
207     \seq_use:Nn
208     \l__tag_tmpa_seq
209     { \iow_newline: }
210   }
211   \tl_if_empty:NF
212   \l__tag_tmpa_tl

```

```

213     {
214         \pdf_object_new:n { __tag/tree/classmap }
215         \pdf_object_write:nnx
216             { __tag/tree/classmap }
217             {dict}
218             { \l__tag_tmpa_tl }
219         \__tag_prop_gput:cnx
220             { g__tag_struct_0_prop }
221             { ClassMap }
222             { \pdf_object_ref:n { __tag/tree/classmap } }
223     }
224 }

```

(End definition for __tag_tree_write_classmap:.)

1.7 Namespaces

Namespaces are handle in the role module, here is the code to write them out. Namespaces are only relevant for pdf2.0.

__tag/tree/namespaces

```

225 \pdf_object_new:n { __tag/tree/namespaces }

```

(End definition for __tag/tree/namespaces.)

__tag_tree_write_namespaces:

```

226 \cs_new_protected:Npn \__tag_tree_write_namespaces:
227 {
228     \pdf_version_compare:NnF < {2.0}
229     {
230         \prop_map_inline:Nn \g__tag_role_NS_prop
231         {
232             \pdfdict_if_empty:nF {g__tag_role/RoleMapNS_##1_dict}
233             {
234                 \pdf_object_write:nnx {__tag/RoleMapNS/##1}{dict}
235                 {
236                     \pdfdict_use:n {g__tag_role/RoleMapNS_##1_dict}
237                 }
238                 \pdfdict_gput:nnx{g__tag_role/Namespace_##1_dict}
239                 {RoleMapNS}{\pdf_object_ref:n {__tag/RoleMapNS/##1}}
240             }
241             \pdf_object_write:nnx{tag/NS/##1}{dict}
242             {
243                 \pdfdict_use:n {g__tag_role/Namespace_##1_dict}
244             }
245         }
246         \pdf_object_write:nnx {__tag/tree/namespaces}{array}
247         {
248             \prop_map_tokens:Nn \g__tag_role_NS_prop{use_ii:nn}
249         }
250     }
251 }

```

(End definition for __tag_tree_write_namespaces:.)

1.8 Finishing the structure

This assembles the various parts. TODO (when tabular are done or if someone requests it): IDTree

```
\__tag_finish_structure:
252 \hook_new:n {tagpdf/finish/before}
253 \cs_new_protected:Npn \__tag_finish_structure:
254 {
255   \bool_if:NT\g__tag_active_tree_bool
256   {
257     \hook_use:n {tagpdf/finish/before}
258     \__tag_tree_final_checks:
259     \__tag_tree_write_parenttree:
260     \__tag_tree_write_rolemap:
261     \__tag_tree_write_classmap:
262     \__tag_tree_write_namespaces:
263     \__tag_tree_write_structelements: %this is rather slow!!
264     \__tag_tree_write_structtreeroot:
265   }
266 }

(End definition for \__tag_finish_structure:.)
```

1.9 StructParents entry for Page

We need to add to the Page resources the StructParents entry, this is simply the absolute page number.

```
267 \hook_gput_code:nnn{begindocument}{tagpdf}
268 {
269   \bool_if:NT\g__tag_active_tree_bool
270   {
271     \hook_gput_code:nnn{shipout/before} { tagpdf/structparents }
272     {
273       \pdfmanagement_add:nnx
274       { Page }
275       { StructParents }
276       { \int_eval:n { \g_shipout_readonly_int } }
277     }
278   }
279 }
280 </package>
```

Part IV

The **tagpdf-mc-shared** module

Code related to Marked Content (mc-chunks), code shared by all modes

Part of the tagpdf package

1 Public Commands

<code>\tag_mc_begin:n</code>	<code>\tag_mc_begin:n{<key-values>}</code>
<code>\tag_mc_end:</code>	<code>\tag_mc_end:</code>

These commands insert the end code of the marked content. They don't end a group and in generic mode it doesn't matter if they are in another group as the starting commands. In generic mode both commands check if they are correctly nested and issue a warning if not.

<code>\tag_mc_use:n</code>	<code>\tag_mc_use:n{<label>}</code>
----------------------------	---

These command allow to record a marked content that was stashed away before into the current structure. A marked content can be used only once – the command will issue a warning if an mc is use a second time.

<code>\tag_mc_artifact_group_begin:n</code>	<code>\tag_mc_artifact_group_begin:n {<name>}</code>
<code>\tag_mc_artifact_group_end:</code>	<code>\tag_mc_artifact_group_end:</code>

New: 2019-11-20

This command pair creates a group with an artifact marker at the begin and the end. Inside the group the tagging commands are disabled. It allows to mark a complete region as artifact without having to worry about user commands with tagging commands. `<name>` should be a value allowed also for the `artifact` key. It pushes and pops mc-chunks at the begin and end. TODO: document is in tagpdf.tex

<code>\tag_mc_end_push:</code>	<code>\tag_mc_end_push:</code>
<code>\tag_mc_begin_pop:n</code>	<code>\tag_mc_begin_pop:n{<key-values>}</code>

New: 2021-04-22

If there is an open mc chunk, `\tag_mc_end_push:` ends it and pushes its tag of the (global) stack. If there is no open chunk, it puts `-1` on the stack (for debugging) `\tag_mc_begin_pop:n` removes a value from the stack. If it is different from `-1` it opens a tag with it. The reopened mc chunk loses info like the alt text for now.

<code>\tag_mc_if_in_p: *</code>	<code>\tag_mc_if_in:TF {<true code>} {<false code>}</code>
<code>\tag_mc_if_in:TF *</code>	Determines if a mc-chunk is open.

2 Public keys

The following keys can be used with `\tag_mc_begin:n`, `\tagmcbegin`, `\tag_mc_begin_pop:n`,

`tag_␣(mc-key)` This key is required, unless artifact is used. The value is a tag like P or H1 without a slash at the begin, this is added by the code. It is possible to setup new tags. The value of the key is expanded, so it can be a command. The expansion is passed unchanged to the PDF, so it should with a starting slash give a valid PDF name (some ascii with numbers like H4 is fine).

`artifact_␣(mc-key)` This will setup the marked content as an artifact. The key should be used for content that should be ignored. The key can take one of the values **pagination**, **layout**, **page**, **background** and **notype** (this is the default).

`raw_␣(mc-key)` This key allows to add more entries to the properties dictionary. The value must be correct, low-level PDF. E.g. `raw=/Alt (Hello)` will insert an alternative Text.

`alt_␣(mc-key)` This key inserts an `/Alt` value in the property dictionary of the BDC operator. The value is handled as verbatim string, commands are not expanded. The value will be expanded first once.

`actualtext_␣(mc-key)` This key inserts an `/ActualText` value in the property dictionary of the BDC operator. The value is handled as verbatim string, commands are not expanded. The value will be expanded first once.

`label_␣(mc-key)` This key sets a label by which one can call the marked content later in another structure (if it has been stashed with the **stash** key). Internally the label name will start with `tagpdf-`.

`stash_␣(mc-key)` This “stashes” an mc-chunk: it is not inserted into the current structure. It should be normally be used along with a label to be able to use the mc-chunk in another place.

The code is splitted into three parts: code shared by all engines, code specific to luamode and code not used by luamode.

3 Marked content code – shared

```

1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-mc-code-shared} {2022-12-22} {0.98}
4   {part of tagpdf - code related to marking chunks -
5     code shared by generic and luamode }
6 </header>
```


3.1 Variables and counters

MC chunks must be counted. I use a latex counter for the absolute count, so that it is added to `\cl@@ckpt` and restored e.g. in tabulars and align. `\int_new:N \c@g_@@_MCID_int` and `\tl_put_right:Nn\cl@@ckpt{\@elt{g_uf_test_int}}` would work too, but as the name is not expl3 then too, why bother? The absolute counter can be used to label and to check if the page counter needs a reset.

`g__tag_MCID_abs_int`

```
7 (*shared)
8 \newcounter { g__tag_MCID_abs_int }
```

(End definition for g__tag_MCID_abs_int.)

`__tag_get_mc_abs_cnt:` A (expandable) function to get the current value of the cnt.

```
9 \cs_new:Npn \__tag_get_mc_abs_cnt: { \int_use:N \c@g__tag_MCID_abs_int }
```

(End definition for __tag_get_mc_abs_cnt:.)

`\g__tag_MCID_tmp_bypage_int`

The following hold the temporary by page number assigned to a mc. It must be defined in the shared code to avoid problems with labels.

```
10 \int_new:N \g__tag_MCID_tmp_bypage_int
```

(End definition for \g__tag_MCID_tmp_bypage_int.)

`\g__tag_in_mc_bool`

This booleans record if a mc is open, to test nesting.

```
11 \bool_new:N \g__tag_in_mc_bool
```

(End definition for \g__tag_in_mc_bool.)

`\g__tag_mc_parenttree_prop`

For every chunk we need to know the structure it is in, to record this in the parent tree. We store this in a property.

key: absolute number of the mc (tagmcabs)

value: the structure number the mc is in

```
12 \__tag_prop_new:N \g__tag_mc_parenttree_prop
```

(End definition for \g__tag_mc_parenttree_prop.)

`\g__tag_mc_parenttree_prop`

Some commands (e.g. links) want to close a previous mc and reopen it after they did their work. For this we create a stack:

```
13 \seq_new:N \g__tag_mc_stack_seq
```

(End definition for \g__tag_mc_parenttree_prop.)

`\l__tag_mc_artifact_type_tl`

Artifacts can have various types like Pagination or Layout. This stored in this variable.

```
14 \tl_new:N \l__tag_mc_artifact_type_tl
```

(End definition for \l__tag_mc_artifact_type_tl.)

`\l__tag_mc_key_stash_bool`

This booleans store the stash and artifact status of the mc-chunk.

`\l__tag_mc_artifact_bool`

```
15 \bool_new:N \l__tag_mc_key_stash_bool
```

```
16 \bool_new:N \l__tag_mc_artifact_bool
```

(End definition for \l__tag_mc_key_stash_bool and \l__tag_mc_artifact_bool.)

`\l__tag_mc_key_tag_tl` Variables used by the keys. `\l_@@mc_key_properties_tl` will collect a number of values. TODO: should this be a pdfdict now?
`\g__tag_mc_key_tag_tl`
`\l__tag_mc_key_label_tl` 17 `\tl_new:N \l__tag_mc_key_tag_tl`
`\l__tag_mc_key_properties_tl` 18 `\tl_new:N \g__tag_mc_key_tag_tl`
19 `\tl_new:N \l__tag_mc_key_label_tl`
20 `\tl_new:N \l__tag_mc_key_properties_tl`
(End definition for `\l__tag_mc_key_tag_tl` and others.)

3.2 Functions

`__tag_mc_handle_mc_label:n` The commands labels a mc-chunk. It is used if the user explicitly labels the mc-chunk with the `label` key. The argument is the value provided by the user. It stores the attributes
`tagabspace`: the absolute page, `\g_shipout_readonly_int`,
`tagmcabs`: the absolute mc-counter `\c@g_@@MCID_abs_int`,
`tagmcid`: the ID of the chunk on the page `\g_@@MCID_tmp_bypage_int`, this typically settles down after a second compilation. The reference command is defined in `tagpdf.dtx` and is based on `l3ref`.
21 `\cs_new:Nn __tag_mc_handle_mc_label:n`
22 `{`
23 `__tag_ref_label:en{tagpdf-#1}{mc}`
24 `}`
(End definition for `__tag_mc_handle_mc_label:n`.)

`__tag_mc_set_label_used:n` Unlike with structures we can't check if a labeled mc has been used by looking at the P key, so we use a dedicated csname for the test
25 `\cs_new_protected:Npn __tag_mc_set_label_used:n #1 %#1 labelname`
26 `{`
27 `\tl_new:c { g__tag_mc_label_\tl_to_str:n{#1}_used_tl }`
28 `}`
29 `\shared`
(End definition for `__tag_mc_set_label_used:n`.)

`\tag_mc_use:n` These command allow to record a marked content that was stashed away before into the current structure. A marked content can be used only once – the command will issue a warning if an mc is use a second time. The argument is a label name set with the `label` key.

TODO: is testing for struct the right test?

30 `\base\cs_new_protected:Npn \tag_mc_use:n #1 { __tag_whatsits: }`
31 `\shared`
32 `\cs_set_protected:Npn \tag_mc_use:n #1 %#1: label name`
33 `{`
34 `__tag_check_if_active_struct:T`
35 `{`
36 `\tl_set:Nx \l__tag_tmpa_tl { __tag_ref_value:nnn{tagpdf-#1}{tagmcabs}{ } }`
37 `\tl_if_empty:NTF\l__tag_tmpa_tl`
38 `{`
39 `\msg_warning:nnn {tag} {mc-label-unknown} {#1}`
40 `}`
41 `}`

```

42         \cs_if_free:cTF { g__tag_mc_label_\tl_to_str:n{#1}_used_tl }
43         {
44             \__tag_mc_handle_stash:x { \l__tag_tmpa_tl }
45             \__tag_mc_set_label_used:n {#1}
46         }
47         {
48             \msg_warning:nnn {tag}{mc-used-twice}{#1}
49         }
50     }
51 }
52 }
53 </shared>

```

(End definition for \tag_mc_use:n. This function is documented on page 55.)

\tag_mc_artifact_group_begin:n This opens an artifact of the type given in the argument, and then stops all tagging. It
\tag_mc_artifact_group_end: creates a group. It pushes and pops mc-chunks at the begin and end.

```

54 <base>\cs_new_protected:Npn \tag_mc_artifact_group_begin:n #1 {}
55 <base>\cs_new_protected:Npn \tag_mc_artifact_group_end: {}
56 <*shared>
57 \cs_set_protected:Npn \tag_mc_artifact_group_begin:n #1
58 {
59     \tag_mc_end_push:
60     \tag_mc_begin:n {artifact=#1}
61     \tag_stop_group_begin:
62 }
63
64 \cs_set_protected:Npn \tag_mc_artifact_group_end:
65 {
66     \tag_stop_group_end:
67     \tag_mc_end:
68     \tag_mc_begin_pop:n {}
69 }
70 </shared>

```

(End definition for \tag_mc_artifact_group_begin:n and \tag_mc_artifact_group_end:. These functions are documented on page 55.)

\tag_mc_end_push:
\tag_mc_begin_pop:n

```

71 <base>\cs_new_protected:Npn \tag_mc_end_push: {}
72 <base>\cs_new_protected:Npn \tag_mc_begin_pop:n #1 {}
73 <*shared>
74 \cs_set_protected:Npn \tag_mc_end_push:
75 {
76     \__tag_check_if_active_mc:T
77     {
78         \__tag_mc_if_in:TF
79         {
80             \seq_gpush:Nx \g__tag_mc_stack_seq { \tag_get:n {mc_tag} }
81             \__tag_check_mc_pushed_popped:nn
82             { pushed }
83             { \tag_get:n {mc_tag} }
84             \tag_mc_end:
85         }
86     }

```

```

87         \seq_gpush:Nn \g__tag_mc_stack_seq {-1}
88         \__tag_check_mc_pushed_popped:nn { pushed }{-1}
89     }
90 }
91 }
92
93 \cs_set_protected:Npn \tag_mc_begin_pop:n #1
94 {
95     \__tag_check_if_active_mc:T
96     {
97         \seq_gpop:NNTF \g__tag_mc_stack_seq \l__tag_tmpa_tl
98         {
99             \tl_if_eq:NnTF \l__tag_tmpa_tl {-1}
100             {
101                 \__tag_check_mc_pushed_popped:nn {popped}{-1}
102             }
103             {
104                 \__tag_check_mc_pushed_popped:nn {popped}{\l__tag_tmpa_tl}
105                 \tag_mc_begin:n {tag=\l__tag_tmpa_tl,#1}
106             }
107         }
108         {
109             \__tag_check_mc_pushed_popped:nn {popped}{empty~stack,~nothing}
110         }
111     }
112 }

```

(End definition for \tag_mc_end_push: and \tag_mc_begin_pop:n. These functions are documented on page 55.)

3.3 Keys

This are the keys where the code can be shared between the modes.

stash_(mc-key) the two internal artifact keys are use to define the public artifact. For now we add support for the subtypes Header and Footer. Watermark,PageNum, LineNum,Redaction,Bates will be added if some use case emerges. If some use case for /BBox and /Attached emerges, it will be perhaps necessary to adapt the code.

```

113 \keys_define:nn { __tag / mc }
114 {
115     stash .bool_set:N = \l__tag_mc_key_stash_bool,
116     __artifact-bool .bool_set:N = \l__tag_mc_artifact_bool,
117     __artifact-type .choice:,
118     __artifact-type / pagination .code:n =
119     {
120         \tl_set:Nn \l__tag_mc_artifact_type_tl { Pagination }
121     },
122     __artifact-type / pagination/header .code:n =
123     {
124         \tl_set:Nn \l__tag_mc_artifact_type_tl { Pagination/Subtype/Header }
125     },
126     __artifact-type / pagination/footer .code:n =
127     {
128         \tl_set:Nn \l__tag_mc_artifact_type_tl { Pagination/Subtype/Footer }

```

```

129     },
130     __artifact-type / layout      .code:n    =
131     {
132         \tl_set:Nn \l__tag_mc_artifact_type_tl { Layout }
133     },
134     __artifact-type / page        .code:n    =
135     {
136         \tl_set:Nn \l__tag_mc_artifact_type_tl { Page }
137     },
138     __artifact-type / background .code:n    =
139     {
140         \tl_set:Nn \l__tag_mc_artifact_type_tl { Background }
141     },
142     __artifact-type / notype      .code:n    =
143     {
144         \tl_set:Nn \l__tag_mc_artifact_type_tl {}
145     },
146     __artifact-type /            .code:n    =
147     {
148         \tl_set:Nn \l__tag_mc_artifact_type_tl {}
149     },
150 }

```

(End definition for stash (mc-key), __artifact-bool, and __artifact-type. This function is documented on page 56.)

```

151 </shared>

```

Part V

The tagpdf-mc-generic module

Code related to Marked Content (mc-chunks), generic mode

Part of the tagpdf package

1 Marked content code – generic mode

```
1 <@@=tag>
2 <*generic>
3 \ProvidesExplPackage {tagpdf-mc-code-generic} {2022-12-22} {0.98}
4 {part of tagpdf - code related to marking chunks - generic mode}
5 </generic>
6 <*debug>
7 \ProvidesExplPackage {tagpdf-debug-generic} {2022-12-22} {0.98}
8 {part of tagpdf - debugging code related to marking chunks - generic mode}
9 </debug>
```

1.1 Variables

`\g__tag_MCID_byabspage_prop` This property will hold the current maximum on a page it will contain key-value of type $\langle abspagenum \rangle = \langle max\ mcid \rangle$

```
10 <*generic>
11 \__tag_prop_new:N \g__tag_MCID_byabspage_prop
```

(End definition for `\g__tag_MCID_byabspage_prop`.)

`\l__tag_mc_ref_abspage_tl` We need a ref-label system to ensure that the MCID cnt restarts at 0 on a new page This will be used to store the tagabspage attribute retrieved from a label.

```
12 \tl_new:N \l__tag_mc_ref_abspage_tl
```

(End definition for `\l__tag_mc_ref_abspage_tl`.)

`\l__tag_mc_tmpa_tl` temporary variable

```
13 \tl_new:N \l__tag_mc_tmpa_tl
```

(End definition for `\l__tag_mc_tmpa_tl`.)

`\g__tag_mc_marks` a marks register to keep track of the mc's at page breaks and a sequence to keep track of the data for the continuation extra-tmb. We probably will need to track mc-marks in more than one stream, so the seq contains the name of the stream.

```
14 \newmarks \g__tag_mc_marks
```

(End definition for `\g__tag_mc_marks`.)

`\g__tag_mc_main_marks_seq` Each stream has an associated global seq variable holding the bottom marks from the/a previous chunk in the stream. We provide three by default: main, footnote and multicol.
`\g__tag_mc_footnote_marks_seq`
`\g__tag_mc_multicol_marks_seq`
 TODO: perhaps an interface for more streams will be needed.

```
15 \seq_new:N \g__tag_mc_main_marks_seq
16 \seq_new:N \g__tag_mc_footnote_marks_seq
17 \seq_new:N \g__tag_mc_multicol_marks_seq
```

(End definition for `\g__tag_mc_main_marks_seq`, `\g__tag_mc_footnote_marks_seq`, and `\g__tag_mc_multicol_marks_seq`.)

`\l__tag_mc_firstmarks_seq` The marks content contains a number of data which we will have to access and compare, so we will store it locally in two sequences. topmarks is unusable in LaTeX so we ignore it.
`\l__tag_mc_botmarks_seq`

```
18 \seq_new:N \l__tag_mc_firstmarks_seq
19 \seq_new:N \l__tag_mc_botmarks_seq
```

(End definition for `\l__tag_mc_firstmarks_seq` and `\l__tag_mc_botmarks_seq`.)

1.2 Functions

`__tag_mc_begin_marks:nn` Generic mode need to set marks for the page break and split stream handling. We always set two marks to be able to detect the case when no mark is on a page/galley. MC-begin commands will set (b,-,data) and (b+,data), MC-end commands will set (e,-,data) and (e+,data).

```
20 \cs_new_protected:Npn \__tag_mc_begin_marks:nn #1 #2 % #1 tag, #2 label
21 {
22   \tex_marks:D \g__tag_mc_marks
23   {
24     b-, %first of begin pair
25     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
26     \g__tag_struct_stack_current_tl, %structure num
27     #1, %tag
28     \bool_if:NT \l__tag_mc_key_stash_bool{stash}, % stash info
29     #2, %label
30   }
31   \tex_marks:D \g__tag_mc_marks
32   {
33     b+, % second of begin pair
34     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
35     \g__tag_struct_stack_current_tl, %structure num
36     #1, %tag
37     \bool_if:NT \l__tag_mc_key_stash_bool{stash}, % stash info
38     #2, %label
39   }
40 }
41 \cs_generate_variant:Nn \__tag_mc_begin_marks:nn {oo}
42 \cs_new_protected:Npn \__tag_mc_artifact_begin_marks:n #1 % #1 type
43 {
44   \tex_marks:D \g__tag_mc_marks
45   {
46     b-, %first of begin pair
47     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
48     -1, %structure num
```

```

49     #1 %type
50   }
51   \tex_marks:D \g__tag_mc_marks
52   {
53     b+, %first of begin pair
54     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
55     -1, %structure num
56     #1 %Type
57   }
58 }
59
60 \cs_new_protected:Npn \__tag_mc_end_marks:
61 {
62   \tex_marks:D \g__tag_mc_marks
63   {
64     e-, %first of end pair
65     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
66     \g__tag_struct_stack_current_tl, %structure num
67   }
68   \tex_marks:D \g__tag_mc_marks
69   {
70     e+, %second of end pair
71     \int_use:N\c@g__tag_MCID_abs_int, %mc-num
72     \g__tag_struct_stack_current_tl, %structure num
73   }
74 }

```

(End definition for __tag_mc_begin_marks:nn, __tag_mc_artifact_begin_marks:n, and __tag_mc_end_marks:.)

__tag_mc_disable_marks: This disables the marks. They can't be reenabled, so it should only be used in groups.

```

75 \cs_new_protected:Npn \__tag_mc_disable_marks:
76 {
77   \cs_set_eq:NN \__tag_mc_begin_marks:nn \use_none:nn
78   \cs_set_eq:NN \__tag_mc_artifact_begin_marks:n \use_none:n
79   \cs_set_eq:NN \__tag_mc_end_marks: \prg_do_nothing:
80 }

```

(End definition for __tag_mc_disable_marks:.)

__tag_mc_get_marks: This stores the current content of the marks in the sequences. It naturally should only be used in places where it makes sense.

```

81 \cs_new_protected:Npn \__tag_mc_get_marks:
82 {
83   \exp_args:NNx
84   \seq_set_from_clist:Nn \l__tag_mc_firstmarks_seq
85   { \tex_firstmarks:D \g__tag_mc_marks }
86   \exp_args:NNx
87   \seq_set_from_clist:Nn \l__tag_mc_botmarks_seq
88   { \tex_botmarks:D \g__tag_mc_marks }
89 }

```

(End definition for __tag_mc_get_marks:.)

`__tag_mc_store:nnn` This inserts the mc-chunk $\langle mc-num \rangle$ into the structure `struct-num` after the $\langle mc-prev \rangle$. The structure must already exist. The additional mcid dictionary is stored in a property. The item is retrieved when the kid entry is built. We test if there is already an addition and append if needed.

```

90 \cs_new_protected:Npn \__tag_mc_store:nnn #1 #2 #3 % #1 mc-prev, #2 mc-num #3 structure-
    num
91 {
92   %\prop_show:N \g__tag_struct_cont_mc_prop
93   \prop_get:NnTF \g__tag_struct_cont_mc_prop {#1} \l__tag_tmpa_tl
94   {
95     \prop_gput:Nnx \g__tag_struct_cont_mc_prop {#1}{ \l__tag_tmpa_tl \__tag_struct_mcid_d
96   }
97   {
98     \prop_gput:Nnx \g__tag_struct_cont_mc_prop {#1}{ \__tag_struct_mcid_dict:n {#2}}
99   }
100   \prop_gput:Nxx \g__tag_mc_parenttree_prop
101     {#2}
102     {#3}
103 }
104 \cs_generate_variant:Nn \__tag_mc_store:nnn {xxx}

```

(End definition for `__tag_mc_store:nnn`.)

`__tag_mc_insert_extra_tmb:n` `__tag_mc_insert_extra_tme:n` These two functions should be used in the output routine at the place where a mc-literal could be missing due to a page break or some other split. They check (with the help of the marks) if a extra-tmb or extra-tme is needed. The tmb command stores also the mc into the structure, the tme has to store the data for a following extra-tmb. The argument takes a stream name like main or footnote to allow different handling there. The content of the marks must be stored before (with `\@@_mc_get_marks:` or manually) into `\l_@@_mc_firstmarks_seq` and `\l_@@_mc_botmarks_seq` so that the tests can use them.

```

105 \cs_new_protected:Npn \__tag_mc_insert_extra_tmb:n #1 % #1 stream: e.g. main or footnote
106 {
107   \__tag_check_typeout_v:n {>~ first~ \seq_use:Nn \l__tag_mc_firstmarks_seq {,~}}
108   \__tag_check_typeout_v:n {>~ bot~ \seq_use:Nn \l__tag_mc_botmarks_seq {,~}}
109   \__tag_check_if_mc_tmb_missing:TF
110   {
111     \__tag_check_typeout_v:n {>~ TMB~ ~ missing~ --- inserted}
112     %test if artifact
113     \int_compare:nNnTF { \seq_item:cn { g__tag_mc_#1_marks_seq } {3} } = {-
114       1}
115       {
116         \tl_set:Nx \l__tag_tmpa_tl { \seq_item:cn { g__tag_mc_#1_marks_seq } {4} }
117         \__tag_mc_handle_artifact:N \l__tag_tmpa_tl
118       }
119       {
120         \exp_args:Nx
121         \__tag_mc_bdc_mcid:n
122         {
123           \seq_item:cn { g__tag_mc_#1_marks_seq } {4}
124         }
125         \str_if_eq:eeTF

```

```

126         \seq_item:cn { g__tag_mc_#1_marks_seq } {5}
127     }
128     {}
129     {
130         %store
131         \__tag_mc_store:xxx
132         {
133             \seq_item:cn { g__tag_mc_#1_marks_seq } {2}
134         }
135         { \int_eval:n{\c@g__tag_MCID_abs_int} }
136         {
137             \seq_item:cn { g__tag_mc_#1_marks_seq } {3}
138         }
139     }
140     {
141         %stashed -> warning!!
142     }
143 }
144 }
145 {
146     \__tag_check_typeout_v:n {=>~ TMB~ not~ missing}
147 }
148 }
149
150 \cs_new_protected:Npn \__tag_mc_insert_extra_tme:n #1 % #1 stream, eg. main or footnote
151 {
152     \__tag_check_if_mc_tme_missing:TF
153     {
154         \__tag_check_typeout_v:n {=>~ TME~ ~ missing~ --- inserted}
155         \__tag_mc_emc:
156         \seq_gset_eq:cN
157         { g__tag_mc_#1_marks_seq }
158         \l__tag_mc_botmarks_seq
159     }
160     {
161         \__tag_check_typeout_v:n {=>~ TME~ not~ missing}
162     }
163 }

```

(End definition for __tag_mc_insert_extra_tmb:n and __tag_mc_insert_extra_tme:n.)

1.3 Looking at MC marks in boxes

__tag_add_missing_mcs:Nn Assumptions:

- test for tagging active outside;
- mark retrieval also outside.

This takes a box register as its first argument (or the register number in a count register, as used by `multicol`). It adds an extra tmb at the top of the box if necessary and similarly an extra tme at the end. This is done by adding hboxes in a way that the positioning and the baseline of the given box is not altered. The result is written back to the box.

The second argument is the stream this box belongs to and is currently either `main` for the main galley, `footnote` for footnote text, or `multicol` for boxes produced for columns in that environment. Other streams may follow over time.

```

164 \cs_new_protected:Npn \__tag_add_missing_mcs:Nn #1 #2 {
165   \vbadness \@M
166   \vfuzz      \c_max_dim
167   \vbox_set_to_ht:Nnn #1 { \box_ht:N #1 } {
168     \hbox_set:Nn \l__tag_tmpa_box { \__tag_mc_insert_extra_tmb:n {#2} }
169     \hbox_set:Nn \l__tag_tmppb_box { \__tag_mc_insert_extra_tme:n {#2} }
170     \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
171     {
172       \seq_log:c { g__tag_mc_#2_marks_seq}
173     }

```

The box placed on the top gets zero size and thus will not affect the box dimensions of the box we are modifying.

```

174   \box_set_ht:Nn \l__tag_tmpa_box \c_zero_dim
175   \box_set_dp:Nn \l__tag_tmpa_box \c_zero_dim

```

The box added at the bottom will get the depth of the original box. This way we can arrange that from the outside everything looks as before.

```

176   \box_set_ht:Nn \l__tag_tmppb_box \c_zero_dim
177   \box_set_dp:Nn \l__tag_tmppb_box { \box_dp:N #1 }

```

We need to set `\boxmaxdepth` in case the original box has an unusually large depth, otherwise that depth is not preserved when we string things together.

```

178   \boxmaxdepth \@maxdepth
179   \box_use_drop:N      \l__tag_tmpa_box
180   \vbox_unpack_drop:N  #1

```

Back up by the depth of the box as we add that later again.

```

181   \tex_kern:D -\box_dp:N \l__tag_tmppb_box

```

And we don't want any glue added when we add the box.

```

182   \nointerlineskip
183   \box_use_drop:N \l__tag_tmppb_box
184 }
185 }

```

(End definition for `__tag_add_missing_mcs:Nn`.)

`__tag_add_missing_mcs_to_stream:Nn`

This is the main command to add mc to the stream. It is therefor guarded by the mc-boolean.

If we aren't in the main stream then processing is a bit more complicated because to get at the marks in the box we need to artificially split it and then look at the split marks.

First argument is the box to update and the second is the "stream". In lua mode the command is a no-op.

```

186 \cs_new_protected:Npn \__tag_add_missing_mcs_to_stream:Nn #1#2
187 {
188   \__tag_check_if_active_mc:T {

```

First set up a temp box for trial splitting.

```

189   \vbadness\maxdimen
190   \box_set_eq:NN \l__tag_tmpa_box #1

```

Split the box to the largest size available. This should give us all content (but to be sure that there is no issue we could test out test box is empty now (not done).

```
191 \vbox_set_split_to_ht:NNn \l__tag_tmpa_box \l__tag_tmpa_box \c_max_dim
```

As a side effect of this split we should now have the first and bottom split marks set up. We use this to set up `\l__tag_mc_firstmarks_seq`

```
192 \exp_args:NNx
193 \seq_set_from_clist:Nn \l__tag_mc_firstmarks_seq
194 { \tex_splitfirstmarks:D \g__tag_mc_marks }
```

Some debugging info:

```
195 % \iow_term:n { First~ mark~ from~ this~ box: }
196 % \seq_log:N \l__tag_mc_firstmarks_seq
```

If this mark was empty then clearly the bottom mark will too be empty. Thus in this case we make use of the saved bot mark from the previous chunk. Note that if this is the first chunk in the stream the global seq would contain a random value, but then we can't end in this branch because the basis assumption is that streams are properly marked up so the first chunk would always have a mark at the beginning!

```
197 \seq_if_empty:NTF \l__tag_mc_firstmarks_seq
198 {
199   \__tag_check_typeout_v:n
200   {
201     No~ marks~ so~ use~ saved~ bot~ mark:~
202     \seq_use:cn {g__tag_mc_#2_marks_seq} {,~} \iow_newline:
203   }
204   \seq_set_eq:Nc \l__tag_mc_firstmarks_seq {g__tag_mc_#2_marks_seq}
```

We also update the bot mark to the same value so that we can later apply `__tag_add_missing_mcs:Nn` with the data structures in place (see assumptions made there).

```
205 \seq_set_eq:NN \l__tag_mc_botmarks_seq \l__tag_mc_firstmarks_seq
206 }
```

If there was a first mark then there is also a bot mark (and it can't be the same as our marks always come in pairs). So if that branch is chosen we update `\l__tag_mc_botmarks_seq` from the bot mark.

```
207 {
208   \__tag_check_typeout_v:n
209   {
210     Pick~ up~ new~ bot~ mark!
211   }
212   \exp_args:NNx
213   \seq_set_from_clist:Nn \l__tag_mc_botmarks_seq
214   { \tex_splitbotmarks:D \g__tag_mc_marks }
215 }
```

Finally we call `__tag_add_missing_mcs:Nn` to add any missing tmb/tme as needed,

```
216 \__tag_add_missing_mcs:Nn #1 {#2}
217 %%
218 \seq_gset_eq:cN {g__tag_mc_#2_marks_seq} \l__tag_mc_botmarks_seq
219 %%
220 }
221 }
```

(End definition for `__tag_add_missing_mcs_to_stream:Nn`.)

`_tag_mc_if_in_p:` This is a test if a mc is open or not. It depends simply on a global boolean: mc-chunks are added linearly so nesting should not be relevant.

`_tag_mc_if_in:TF`
`\tag_mc_if_in_p:` One exception are header and footer (perhaps they are more, but for now it doesn't seem so, so there are no dedicated code to handle this situation): When they are built and added to the page we could be both inside or outside a mc-chunk. But header and footer should ignore this and not push/pop or warn about nested mc. It is therefore important there to set and reset the boolean manually. See the tagpddocu-patches.sty for an example.
`\tag_mc_if_in:TF`

```

222 \prg_new_conditional:Nnn \_tag_mc_if_in: {p,T,F,TF}
223 {
224   \bool_if:NTF \g__tag_in_mc_bool
225   { \prg_return_true: }
226   { \prg_return_false: }
227 }

```

```

228
229 \prg_new_eq_conditional:Nnn \tag_mc_if_in: \_tag_mc_if_in: {p,T,F,TF}

```

(End definition for `_tag_mc_if_in:TF` and `\tag_mc_if_in:TF`. This function is documented on page 55.)

`_tag_mc_bmc:n` These are the low-level commands. There are now equal to the pdfmanagement commands generic mode, but we use an indirection in case luamode need something else.
`_tag_mc_emc:` change 04.08.2018: the commands do not check the validity of the arguments or try to
`_tag_mc_bdc:nn` escape them, this should be done before using them.
`_tag_mc_bdc:nx`

```

230 % #1 tag, #2 properties
231 \cs_set_eq:NN \_tag_mc_bmc:n \pdf_bmc:n
232 \cs_set_eq:NN \_tag_mc_emc: \pdf_emc:
233 \cs_set_eq:NN \_tag_mc_bdc:nn \pdf_bdc:nn
234 \cs_generate_variant:Nn \_tag_mc_bdc:nn {nx}

```

(End definition for `_tag_mc_bmc:n`, `_tag_mc_emc:`, and `_tag_mc_bdc:nn`.)

`_tag_mc_bdc_mcid:nn` This create a BDC mark with an /MCID key. Most of the work here is to get the current
`_tag_mc_bdc_mcid:n` number value for the MCID: they must be numbered by page starting with 0 and then
`_tag_mc_handle_mcid:nn` successively. The first argument is the tag, e.g. P or Span, the second is used to pass
`_tag_mc_handle_mcid:VV` more properties. We also define a wrapper around the low-level command as luamode
will need something different.

```

235 \cs_new_protected:Npn \_tag_mc_bdc_mcid:nn #1 #2
236 {
237   \int_gincr:N \c@g__tag_MCID_abs_int
238   \tl_set:Nx \l__tag_mc_ref_abbrev_t1
239   {
240     \_tag_ref_value:enn %3 args
241     {
242       mcid-\int_use:N \c@g__tag_MCID_abs_int
243     }
244     { tagabbrev }
245     {-1}
246   }
247   \prop_get:NoNTF
248   \g__tag_MCID_byabbrev_prop
249   {
250     \l__tag_mc_ref_abbrev_t1

```

```

251     }
252     \l__tag_mc_tmpa_tl
253     {
254         %key already present, use value for MCID and add 1 for the next
255         \int_gset:Nn \g__tag_MCID_tmp_bypage_int { \l__tag_mc_tmpa_tl }
256         \__tag_prop_gput:Nxx
257             \g__tag_MCID_byabspage_prop
258             { \l__tag_mc_ref_abspage_tl }
259             { \int_eval:n { \l__tag_mc_tmpa_tl +1 } }
260     }
261     {
262         %key not present, set MCID to 0 and insert 1
263         \int_gzero:N \g__tag_MCID_tmp_bypage_int
264         \__tag_prop_gput:Nxx
265             \g__tag_MCID_byabspage_prop
266             { \l__tag_mc_ref_abspage_tl }
267             { 1 }
268     }
269     \__tag_ref_label:en
270     {
271         mcid-\int_use:N \c@g__tag_MCID_abs_int
272     }
273     { mc }
274     \__tag_mc_bdc:nx
275     { #1 }
276     { /MCID~\int_eval:n { \g__tag_MCID_tmp_bypage_int } ~ \exp_not:n { #2 } }
277 }
278 \cs_new_protected:Npn \__tag_mc_bdc_mcid:n #1
279 {
280     \__tag_mc_bdc_mcid:nn { #1 } {}
281 }
282
283 \cs_new_protected:Npn \__tag_mc_handle_mcid:nn #1 #2 % #1 tag, #2 properties
284 {
285     \__tag_mc_bdc_mcid:nn { #1 } { #2 }
286 }
287
288 \cs_generate_variant:Nn \__tag_mc_handle_mcid:nn { VV }

```

(End definition for __tag_mc_bdc_mcid:nn, __tag_mc_bdc_mcid:n, and __tag_mc_handle_mcid:nn.)

__tag_mc_handle_stash:n This is the handler which puts a mc into the the current structure. The argument is the
 __tag_mc_handle_stash:x number of the mc. Beside storing the mc into the structure, it also has to record the
 structure for the parent tree. The name is a bit confusing, it does *not* handle mc with
 the stash key TODO: why does luamode use it for begin + use, but generic mode
 only for begin?

```

289 \cs_new_protected:Npn \__tag_mc_handle_stash:n #1 %1 mcidnum
290 {
291     \__tag_check_mc_used:n { #1 }
292     \__tag_struct_kid_mc_gput_right:nn
293     { \g__tag_struct_stack_current_tl }
294     { #1 }
295     \prop_gput:Nxx \g__tag_mc_parenttree_prop
296     { #1 }

```

```

297     { \g__tag_struct_stack_current_tl }
298   }
299   \cs_generate_variant:Nn \__tag_mc_handle_stash:n { x }

```

(End definition for __tag_mc_handle_stash:n.)

__tag_mc_bmc_artifact: Two commands to create artifacts, one without type, and one with. We define also a wrapper handler as luamode will need a different definition. TODO: perhaps later: more properties for artifacts

```

300 \cs_new_protected:Npn \__tag_mc_bmc_artifact:
301   {
302     \__tag_mc_bmc:n {Artifact}
303   }
304 \cs_new_protected:Npn \__tag_mc_bmc_artifact:n #1
305   {
306     \__tag_mc_bdc:nn {Artifact}{/Type/#1}
307   }
308 \cs_new_protected:Npn \__tag_mc_handle_artifact:N #1
309   % #1 is a var containing the artifact type
310   {
311     \int_gincr:N \c@g__tag_MCID_abs_int
312     \tl_if_empty:NTF #1
313       { \__tag_mc_bmc_artifact: }
314       { \exp_args:NV\__tag_mc_bmc_artifact:n #1 }
315   }

```

(End definition for __tag_mc_bmc_artifact:, __tag_mc_bmc_artifact:n, and __tag_mc_handle_artifact:N.)

__tag_get_data_mc_tag: This allows to retrieve the active mc-tag. It is use by the get command.

```

316 \cs_new:Nn \__tag_get_data_mc_tag: { \g__tag_mc_key_tag_tl }
317 </generic>

```

(End definition for __tag_get_data_mc_tag:.)

\tag_mc_begin:n These are the core public commands to open and close an mc. They don't need to be in the same group or grouping level, but the code expect that they are issued linearly. The tag and the state is passed to the end command through a global var and a global boolean.

\tag_mc_end:

```

318 <base>\cs_new_protected:Npn \tag_mc_begin:n #1 { \__tag_whatsits: }
319 <base>\cs_new_protected:Nn \tag_mc_end:{ \__tag_whatsits: }
320 <*generic | debug>
321 <*generic>
322 \cs_set_protected:Npn \tag_mc_begin:n #1 %#1 keyval
323   {
324     \__tag_check_if_active_mc:T
325     {
326       </generic>
327       <*debug>
328     }
329   }
330 \cs_set_protected:Npn \tag_mc_begin:n #1 %#1 keyval
331   {
332     \__tag_check_if_active_mc:TF
333     {
334       \__tag_debug_mc_begin_insert:n { #1 }

```

```

333 </debug>
334     \group_begin: %hm
335     \__tag_check_mc_if_nested:
336     \bool_gset_true:N \g__tag_in_mc_bool
337 set default MC tags to structure:
338     \tl_set_eq:NN \l__tag_mc_key_tag_tl \g__tag_struct_tag_tl
339     \tl_gset_eq:NN \g__tag_mc_key_tag_tl \g__tag_struct_tag_tl
340     \keys_set:nn { __tag / mc } {#1}
341     \bool_if:NTF \l__tag_mc_artifact_bool
342     { %handle artifact
343       \__tag_mc_handle_artifact:N \l__tag_mc_artifact_type_tl
344       \exp_args:NV
345       \__tag_mc_artifact_begin_marks:n \l__tag_mc_artifact_type_tl
346     }
347     { %handle mcid type
348       \__tag_check_mc_tag:N \l__tag_mc_key_tag_tl
349       \__tag_mc_handle_mcid:VV
350         \l__tag_mc_key_tag_tl
351         \l__tag_mc_key_properties_tl
352       \__tag_mc_begin_marks:oof{\l__tag_mc_key_tag_tl}{\l__tag_mc_key_label_tl}
353       \tl_if_empty:NF {\l__tag_mc_key_label_tl}
354       {
355         \exp_args:NV
356         \__tag_mc_handle_mc_label:n \l__tag_mc_key_label_tl
357       }
358       \bool_if:NF \l__tag_mc_key_stash_bool
359       {
360         \exp_args:NV \__tag_struct_get_tag_info:nNN
361         \g__tag_struct_stack_current_tl
362         \l__tag_tmpa_tl
363         \l__tag_tmpb_tl
364         \__tag_check_parent_child:VVnnN
365         \l__tag_tmpa_tl \l__tag_tmpb_tl
366         {MC}{ }
367         \l__tag_parent_child_check_tl
368         \int_compare:nNnT {\l__tag_parent_child_check_tl}<{0}
369         {
370           \msg_warning:nnxxx
371           { tag }
372           {role-parent-child}
373           { \g__tag_struct_tag_tl/\g__tag_struct_tag_NS_tl }
374           { MC~(=real content) }
375           { 'not~allowed'. }
376         }
377         \__tag_mc_handle_stash:x { \int_use:N \c@g__tag_MCID_abs_int }
378       }
379     }
380   }
381 <*debug>
382   {
383     \__tag_debug_mc_begin_ignore:n { #1 }
384   }
385 </debug>

```



```

386     }
387     <*generic>
388     \cs_set_protected:Nn \tag_mc_end:
389     {
390         \__tag_check_if_active_mc:T
391         {
392             </generic>
393             <*debug>
394             \cs_set_protected:Nn \tag_mc_end:
395             {
396                 \__tag_check_if_active_mc:TF
397                 {
398                     \__tag_debug_mc_end_insert:
399                     </debug>
400                     \__tag_check_mc_if_open:
401                     \bool_gset_false:N \g__tag_in_mc_bool
402                     \tl_gset:Nn \g__tag_mc_key_tag_tl { }
403                     \__tag_mc_emc:
404                     \__tag_mc_end_marks:
405                 }
406             <*debug>
407             {
408                 \__tag_debug_mc_end_ignore:
409             }
410         </debug>
411     }
412 </generic | debug>

```

(End definition for \tag_mc_begin:n and \tag_mc_end:. These functions are documented on page 55.)

1.4 Keys

Definitions are different in luamode. `tag` and `raw` are expanded as `\lua_now:e` in lua does it too and we assume that their values are safe.

```

tag_(mc-key)
raw_(mc-key) 413 <*generic>
alt_(mc-key) 414 \keys_define:nn { __tag / mc }
actualtext_(mc-key) 415 {
label_(mc-key) 416     tag .code:n = % the name (H,P,Span) etc
artifact_(mc-key) 417     {
418         \tl_set:Nx \l__tag_mc_key_tag_tl { #1 }
419         \tl_gset:Nx \g__tag_mc_key_tag_tl { #1 }
420     },
421     raw .code:n =
422     {
423         \tl_put_right:Nx \l__tag_mc_key_properties_tl { #1 }
424     },
425     alt .code:n = % Alt property
426     {
427         \str_set_convert:Noon
428         \l__tag_tmpa_str
429         { #1 }
430         { default }

```

```

431         { utf16/hex }
432         \tl_put_right:Nn \l__tag_mc_key_properties_tl { /Alt~< }
433         \tl_put_right:No \l__tag_mc_key_properties_tl { \l__tag_tmpa_str>~ }
434     },
435     alttext .meta:n = {alt=#1},
436     actualtext .code:n      = % ActualText property
437     {
438         \str_set_convert:Noon
439         \l__tag_tmpa_str
440         { #1 }
441         { default }
442         { utf16/hex }
443         \tl_put_right:Nn \l__tag_mc_key_properties_tl { /ActualText~< }
444         \tl_put_right:No \l__tag_mc_key_properties_tl { \l__tag_tmpa_str>~ }
445     },
446     label .tl_set:N      = \l__tag_mc_key_label_tl,
447     artifact .code:n     =
448     {
449         \exp_args:Nnx
450         \keys_set:nn
451         { __tag / mc }
452         { __artifact-bool, __artifact-type=#1 }
453     },
454     artifact .default:n  = {notype}
455 }
456 </generic>

```

(End definition for tag (mc-key) and others. These functions are documented on page 56.)

Part VI

The tagpdf-mc-luacode module

Code related to Marked Content (mc-chunks), luamode-specific

Part of the tagpdf package

The code is splitted into three parts: code shared by all engines, code specific to luamode and code not used by luamode.

1 Marked content code – luamode code

luamode uses attributes to mark mc-chunks. The two attributes used are defined in the backend file. The backend also load the lua file, as it can contain functions needed elsewhere. The attributes for mc are global (between 0.6 and 0.81 they were local but this was reverted). The attributes are setup only in lua, and one should use the lua functions to set and get them.

`g_@@_mc_type_attr`: the value represent the type

`g_@@_mc_cnt_attr`: will hold the `\c@g_@@_MCID_abs_int` value

Handling attribute needs a different system to number the page wise mcid's: a `\tagmcbegin ... \tagmcend` pair no longer surrounds exactly one mc chunk: it can be split at page breaks. We know the included mcid(s) only after the shipout. So for the `struct -> mcid` mapping we need to record `struct -> mc-cnt` (in `\g_@@_mc_parenttree_prop` and/or a lua table and at shipout `mc-cnt-> {mcid, mcid, ...}` and when building the trees connect both.

Key definitions are overwritten for luatex to store that data in lua-tables. The data for the mc are in `ltx.@@.mc[absnum]`. The fields of the table are:

`tag` : the type (a string)

`raw` : more properties (string)

`label`: a string.

`artifact`: the presence indicates an artifact, the value (string) is the type.

`kids`: a array of tables

`{1={kid=num2,page=pagenum1}, 2={kid=num2,page=pagenum2},...}`,

this describes the chunks the mc has been split to by the traversing code

`parent`: the number of the structure it is in. Needed to build the parent tree.

```
1 <@@=tag>
2 <*luamode>
3 \ProvidesExplPackage {tagpdf-mc-code-lua} {2022-12-22} {0.98}
4   {tagpdf - mc code only for the luamode }
5 </luamode>
```

The main function which wanders through the shipout box to inject the literals. if the new callback is there, it is used.

```
6 <*luamode>
7 \hook_gput_code:nnn{begindocument}{tagpdf/mc}
8 {
```

```

9      \bool_if:NT\g__tag_active_space_bool
10      {
11          \lua_now:e
12          {
13              if~luatexbase.callbacktypes.pre_shipout_filter~then~
14                  luatexbase.add_to_callback("pre_shipout_filter", function(TAGBOX)~
15                      ltx.__tag.func.space_chars_shipout(TAGBOX)~return~true~
16                      end, "tagpdf")~
17              end
18          }
19      \lua_now:e
20      {
21          if~luatexbase.callbacktypes.pre_shipout_filter~then~
22              token.get_next()~
23              end
24          }~\@secondoftwo~\@gobble
25          {
26              \hook_gput_code:nnn{shipout/before}{tagpdf/lua}
27              {
28                  \lua_now:e
29                  { ltx.__tag.func.space_chars_shipout (tex.box["ShipoutBox"]) }
30              }
31          }
32      }
33      \bool_if:NT\g__tag_active_mc_bool
34      {
35          \lua_now:e
36          {
37              if~luatexbase.callbacktypes.pre_shipout_filter~then~
38                  luatexbase.add_to_callback("pre_shipout_filter", function(TAGBOX)~
39                      ltx.__tag.func.mark_shipout(TAGBOX)~return~true~
40                      end, "tagpdf")~
41              end
42          }
43      \lua_now:e
44      {
45          if~luatexbase.callbacktypes.pre_shipout_filter~then~
46              token.get_next()~
47              end
48          }~\@secondoftwo~\@gobble
49          {
50              \hook_gput_code:nnn{shipout/before}{tagpdf/lua}
51              {
52                  \lua_now:e
53                  { ltx.__tag.func.mark_shipout (tex.box["ShipoutBox"]) }
54              }
55          }
56      }
57  }

```

1.1 Commands

`_tag_add_missing_mcs_to_stream:Nn`

This command is used in the output routine by the ptagging code. It should do nothing in luamode.

```
58 \cs_new_protected:Npn \__tag_add_missing_mcs_to_stream:Nn #1#2 {}
```

(End definition for __tag_add_missing_mcs_to_stream:Nn.)

__tag_mc_if_in_p: This tests, if we are in an mc, for attributes this means to check against a number.

```
\__tag_mc_if_in:TF 59 \prg_new_conditional:Nnn \__tag_mc_if_in: {p,T,F,TF}
\tag_mc_if_in_p: 60 {
\tag_mc_if_in:TF 61 \int_compare:nNnTF
62 { -2147483647 }
63 =
64 {\lua_now:e
65 {
66 tex.print(tex.getattribute(luatexbase.attributes.g__tag_mc_type_attr))
67 }
68 }
69 { \prg_return_false: }
70 { \prg_return_true: }
71 }
72
73 \prg_new_eq_conditional:NNn \tag_mc_if_in: \__tag_mc_if_in: {p,T,F,TF}
```

(End definition for __tag_mc_if_in:TF and \tag_mc_if_in:TF. This function is documented on page 55.)

__tag_mc_lua_set_mc_type_attr:n This takes a tag name, and sets the attributes globally to the related number.

```
\__tag_mc_lua_set_mc_type_attr:o 74 \cs_new:Nn \__tag_mc_lua_set_mc_type_attr:n % #1 is a tag name
\__tag_mc_lua_unset_mc_type_attr: 75 {
76 %TODO ltx.__tag.func.get_num_from("#1") seems not to return a suitable number??
77 \tl_set:Nx\l__tag_tmpa_tl{\lua_now:e{ltx.__tag.func.output_num_from ("#1")}} }
78 \lua_now:e
79 {
80 tex.setattribute
81 (
82 "global",
83 luatexbase.attributes.g__tag_mc_type_attr,
84 \l__tag_tmpa_tl
85 )
86 }
87 \lua_now:e
88 {
89 tex.setattribute
90 (
91 "global",
92 luatexbase.attributes.g__tag_mc_cnt_attr,
93 \__tag_get_mc_abs_cnt:
94 )
95 }
96 }
97
98 \cs_generate_variant:Nn\__tag_mc_lua_set_mc_type_attr:n { o }
99
100 \cs_new:Nn \__tag_mc_lua_unset_mc_type_attr:
101 {
102 \lua_now:e
103 {
```

```

104         tex.setattribute
105         (
106             "global",
107             luatexbase.attributes.g__tag_mc_type_attr,
108             -2147483647
109         )
110     }
111     \lua_now:e
112     {
113         tex.setattribute
114         (
115             "global",
116             luatexbase.attributes.g__tag_mc_cnt_attr,
117             -2147483647
118         )
119     }
120 }
121

```

(End definition for _tag_mc_lua_set_mc_type_attr:n and _tag_mc_lua_unset_mc_type_attr:.)

_tag_mc_insert_mcid_kids:n These commands will in the finish code replace the dummy for a mc by the real mcid
_tag_mc_insert_mcid_single_kids:n kids we need a variant for the case that it is the only kid, to get the array right

```

122 \cs_new:Nn \_tag_mc_insert_mcid_kids:n
123 {
124     \lua_now:e { ltx.__tag.func.mc_insert_kids (#1,0) }
125 }
126
127 \cs_new:Nn \_tag_mc_insert_mcid_single_kids:n
128 {
129     \lua_now:e { ltx.__tag.func.mc_insert_kids (#1,1) }
130 }

```

(End definition for _tag_mc_insert_mcid_kids:n and _tag_mc_insert_mcid_single_kids:n.)

_tag_mc_handle_stash:n This is the lua variant for the command to put an mcid absolute number in the current
_tag_mc_handle_stash:x structure.

```

131 \cs_new:Nn \_tag_mc_handle_stash:n %1 mcidnum
132 {
133     \_tag_check_mc_used:n { #1 }
134     \seq_gput_right:cn % Don't fill a lua table due to the command in the item,
135                       % so use the kernel command
136     { g__tag_struct_kids \g__tag_struct_stack_current_tl _seq }
137     {
138         \_tag_mc_insert_mcid_kids:n {#1}%
139     }
140     \lua_now:e
141     {
142         ltx.__tag.func.store_struct_mcabs
143         (
144             \g__tag_struct_stack_current_tl,#1
145         )
146     }
147     \prop_gput:Nxx

```

```

148     \g__tag_mc_parenttree_prop
149     { #1 }
150     { \g__tag_struct_stack_current_tl }
151 }
152
153 \cs_generate_variant:Nn \__tag_mc_handle_stash:n { x }

(End definition for \__tag_mc_handle_stash:n.)

```

\tag_mc_begin:n This is the lua version of the user command. We currently don't check if there is nesting as it doesn't matter so much in lua.

```

154 \cs_set_protected:Nn \tag_mc_begin:n
155 {
156     \__tag_check_if_active_mc:T
157     {
158         \group_begin:
159         %\__tag_check_mc_if_nested:
160         \bool_gset_true:N \g__tag_in_mc_bool
161         \bool_set_false:N \l__tag_mc_artifact_bool
162         \tl_clear:N \l__tag_mc_key_properties_tl
163         \int_gincr:N \c@g__tag_MCID_abs_int

```

set the default tag to the structure:

```

164         \tl_set_eq:NN \l__tag_mc_key_tag_tl \g__tag_struct_tag_tl
165         \tl_gset_eq:NN \g__tag_mc_key_tag_tl \g__tag_struct_tag_tl
166         \lua_now:e
167         {
168             ltx.__tag.func.store_mc_data(\__tag_get_mc_abs_cnt:,"tag"," \g__tag_struct_tag_tl",
169             }
170         \keys_set:nn { __tag / mc }{ label={}, #1 }
171         %check that a tag or artifact has been used
172         \__tag_check_mc_tag:N \l__tag_mc_key_tag_tl
173         %set the attributes:
174         \__tag_mc_lua_set_mc_type_attr:o { \l__tag_mc_key_tag_tl }
175         \bool_if:NF \l__tag_mc_artifact_bool
176         { % store the absolute num name in a label:
177             \tl_if_empty:NF {\l__tag_mc_key_label_tl}
178             {
179                 \exp_args:NV
180                 \__tag_mc_handle_mc_label:n \l__tag_mc_key_label_tl
181             }
182         % if not stashed record the absolute number
183         \bool_if:NF \l__tag_mc_key_stash_bool
184         {
185             \exp_args:NV \__tag_struct_get_tag_info:nNN
186             \g__tag_struct_stack_current_tl
187             \l__tag_tmpa_tl
188             \l__tag_tmpb_tl
189             \__tag_check_parent_child:VVnnN
190             \l__tag_tmpa_tl \l__tag_tmpb_tl
191             {MC}{ }
192             \l__tag_parent_child_check_tl
193             \int_compare:nNnT {\l__tag_parent_child_check_tl}<{0}
194             {
195                 \msg_warning:nnxxx

```

```

196         { tag }
197         {role-parent-child}
198         { \g__tag_struct_tag_t1/\g__tag_struct_tag_NS_t1 }
199         { MC~(=~real content) }
200         { 'not~allowed'. }
201     }
202     \__tag_mc_handle_stash:x { \__tag_get_mc_abs_cnt: }
203 }
204 }
205 \group_end:
206 }
207 }

```

(End definition for \tag_mc_begin:n. This function is documented on page 55.)

\tag_mc_end: TODO: check how the use command must be guarded.

```

208 \cs_set_protected:Nn \tag_mc_end:
209 {
210     \__tag_check_if_active_mc:T
211     {
212         %\__tag_check_mc_if_open:
213         \bool_gset_false:N \g__tag_in_mc_bool
214         \bool_set_false:N\l__tag_mc_artifact_bool
215         \__tag_mc_lua_unset_mc_type_attr:
216         \tl_set:Nn \l__tag_mc_key_tag_t1 { }
217         \tl_gset:Nn \g__tag_mc_key_tag_t1 { }
218     }
219 }

```

(End definition for \tag_mc_end:. This function is documented on page 55.)

__tag_get_data_mc_tag: The command to retrieve the current mc tag. TODO: Perhaps this should use the attribute instead.

```

220 \cs_new:Npn \__tag_get_data_mc_tag: { \g__tag_mc_key_tag_t1 }

```

(End definition for __tag_get_data_mc_tag:.)

1.2 Key definitions

```

tag_␣(mc-key)  TODO: check conversion, check if local/global setting is right.
raw_␣(mc-key)  221 \keys_define:nn { __tag / mc }
alt_␣(mc-key)  222 {
actualtext_␣(mc-key) 223     tag .code:n = %
label_␣(mc-key)    224     {
artifact_␣(mc-key) 225         \tl_set:Nx \l__tag_mc_key_tag_t1 { #1 }
226         \tl_gset:Nx \g__tag_mc_key_tag_t1 { #1 }
227         \lua_now:e
228         {
229             ltx.__tag.func.store_mc_data(\__tag_get_mc_abs_cnt:,"tag","#1")
230         }
231     },
232     raw .code:n =
233     {
234         \tl_put_right:Nx \l__tag_mc_key_properties_t1 { #1 }

```



```

235     \lua_now:e
236     {
237         ltx.__tag.func.store_mc_data(\__tag_get_mc_abs_cnt:,"raw","#1")
238     }
239 },
240 alt .code:n      = % Alt property
241 {
242     \str_set_convert:Noon
243     \l__tag_tmpa_str
244     { #1 }
245     { default }
246     { utf16/hex }
247     \tl_put_right:Nn \l__tag_mc_key_properties_tl { /Alt~< }
248     \tl_put_right:No \l__tag_mc_key_properties_tl { \l__tag_tmpa_str>~ }
249     \lua_now:e
250     {
251         ltx.__tag.func.store_mc_data
252         (
253             \__tag_get_mc_abs_cnt:,"alt","/Alt~<\str_use:N \l__tag_tmpa_str>"
254         )
255     }
256 },
257 alttext .meta:n = {alt=#1},
258 actualtext .code:n      = % Alt property
259 {
260     \str_set_convert:Noon
261     \l__tag_tmpa_str
262     { #1 }
263     { default }
264     { utf16/hex }
265     \tl_put_right:Nn \l__tag_mc_key_properties_tl { /Alt~< }
266     \tl_put_right:No \l__tag_mc_key_properties_tl { \l__tag_tmpa_str>~ }
267     \lua_now:e
268     {
269         ltx.__tag.func.store_mc_data
270         (
271             \__tag_get_mc_abs_cnt:,
272             "actualtext",
273             "/ActualText~<\str_use:N \l__tag_tmpa_str>"
274         )
275     }
276 },
277 label .code:n =
278 {
279     \tl_set:Nn\l__tag_mc_key_label_tl { #1 }
280     \lua_now:e
281     {
282         ltx.__tag.func.store_mc_data
283         (
284             \__tag_get_mc_abs_cnt:,"label","#1"
285         )
286     }
287 },
288 __artifact-store .code:n =

```

```

289     {
290         \lua_now:e
291         {
292             ltx.__tag.func.store_mc_data
293             (
294                 \__tag_get_mc_abs_cnt:,"artifact","#1"
295             )
296         }
297     },
298     artifact .code:n      =
299     {
300         \exp_args:Nnx
301         \keys_set:nn
302         { __tag / mc }
303         { __artifact-bool, __artifact-type=#1, tag=Artifact }
304         \exp_args:Nnx
305         \keys_set:nn
306         { __tag / mc }
307         { __artifact-store=\l__tag_mc_artifact_type_tl }
308     },
309     artifact .default:n   = { notype }
310 }
311
312 </luamode>

```

(End definition for tag (mc-key) and others. These functions are documented on page 56.)

Part VII

The tagpdf-struct module

Commands to create the structure

Part of the tagpdf package

1 Public Commands

<code>\tag_struct_begin:n</code>	<code>\tag_struct_begin:n{<key-values>}</code>
<code>\tag_struct_end:</code>	<code>\tag_struct_end:</code>

These commands start and end a new structure. They don't start a group. They set all their values globally.

<code>\tag_struct_use:n</code>	<code>\tag_struct_use:n{<label>}</code>
--------------------------------	---

These commands insert a structure previously stashed away as kid into the currently active structure. A structure should be used only once, if the structure already has a parent a warning is issued.

<code>\tag_struct_object_ref:n</code>	<code>\tag_struct_object_ref:n{<struct number>}</code>
<code>\tag_struct_object_ref:e</code>	

This is a small wrapper around `\pdf_object_ref:n` to retrieve the object reference of the structure with the number `<struct number>`. This number can be retrieved and stored for the current structure for example with `\tag_get:n{<structnum>}`. Be aware that it can only be used if the structure has already been created and that it doesn't check if the object actually exists!

The following two functions are used to add annotations. They must be used together and with care to get the same numbers. Perhaps some improvements are needed here.

<code>\tag_struct_insert_annot:nn</code>	<code>\tag_struct_insert_annot:nn{<object reference>}{<struct parent number>}</code>
--	--

This inserts an annotation in the structure. `<object reference>` is there reference to the annotation. `<struct parent number>` should be the same number as had been inserted with `\tag_struct_parent_int:` as `StructParent` value to the dictionary of the annotation. The command will increase the value of the counter used by `\tag_struct_parent_int:`.

<code>\tag_struct_parent_int:</code>	<code>\tag_struct_parent_int:</code>
--------------------------------------	--------------------------------------

This gives back the next free `/StructParent` number (assuming that it is together with `\tag_struct_insert_annot:nn` which will increase the number.

2 Public keys

2.1 Keys for the structure commands

<u>tag_□(struct-key)</u>	This is required. The value of the key is normally one of the standard types listed in the main tagpdf documentation. It is possible to setup new tags/types. The value can also be of the form type/NS , where NS is the shorthand of a declared name space. Currently the names spaces pdf , pdf2 , mathml and user are defined. This allows to use a different name space than the one connected by default to the tag. But normally this should not be needed.
<u>stash_□(struct-key)</u>	Normally a new structure inserts itself as a kid into the currently active structure. This key prohibits this. The structure is nevertheless from now on “the current active structure” and parent for following marked content and structures.
<u>label_□(struct-key)</u>	This key sets a label by which one can refer to the structure. It is e.g. used by <code>\tag_struct_use:n</code> (where a real label is actually not needed as you can only use structures already defined), and by the ref key (which can refer to future structures). Internally the label name will start with tagpdfstruct- and it stores the two attributes tagstruct (the structure number) and tagstructobj (the object reference).
<u>parent_□(struct-key)</u>	By default a structure is added as kid to the currently active structure. With the parent key one can choose another parent. The value is a structure number which must refer to an already existing, previously created structure. Such a structure number can for example be have been stored with <code>\tag_get:n</code> , but one can also use a label on the parent structure and then use <code>\ref_value:nn{tagpdfstruct-label}{tagstruct}</code> to retrieve it.
<u>title_□(struct-key)</u> <u>title-o_□(struct-key)</u>	This keys allows to set the dictionary entry /Title in the structure object. The value is handled as verbatim string and hex encoded. Commands are not expanded. title-o will expand the value once.
<u>alt_□(struct-key)</u>	This key inserts an /Alt value in the dictionary of structure object. The value is handled as verbatim string and hex encoded. The value will be expanded first once.
<u>actualtext_□(struct-key)</u>	This key inserts an /ActualText value in the dictionary of structure object. The value is handled as verbatim string and hex encoded. The value will be expanded first once.
<u>lang_□(struct-key)</u>	This key allows to set the language for a structure element. The value should be a bcp-identifier, e.g. de-De .

ref_□(struct-key) This key allows to add references to other structure elements, it adds the */Ref* array to the structure. The value should be a comma separated list of structure labels set with the *label* key. e.g. `ref={label1,label2}`.

E_□(struct-key) This key sets the */E* key, the expanded form of an abbreviation or an acronym (I couldn't think of a better name, so I stucked to E).

AF_□(struct-key) `AF = <object name>`

AFinline_□(struct-key) `AF-inline = <text content>`

AFinline-o_□(struct-key) These keys allows to reference an associated file in the structure element. The value *<object name>* should be the name of an object pointing to the */Filespec* dictionary as expected by `\pdf_object_ref:n` from a current `l3kernel`.

The value **AF-inline** is some text, which is embedded in the PDF as a text file with mime type `text/plain`. **AF-inline-o** is like **AF-inline** but expands the value once.

Future versions will perhaps extend this to more mime types, but it is still a research task to find out what is really needed.

AF can be used more than once, to associate more than one file. The inline keys can be used only once per structure. Additional calls are ignored.

attribute_□(struct-key) This key takes as argument a comma list of attribute names (use braces to protect the commas from the external key-val parser) and allows to add one or more attribute dictionary entries in the structure object. As an example

`\tagstructbegin{tag=TH,attribute= TH-row}`

Attribute names and their content must be declared first in `\tagpdfsetup`.

attribute-class_□(struct-key)

This key takes as argument a comma list of attribute class names (use braces to protect the commas from the external key-val parser) and allows to add one or more attribute classes to the structure object.

Attribute class names and their content must be declared first in `\tagpdfsetup`.

2.2 Setup keys

`newattribute_␣(setup-key)` `newattribute = {⟨name⟩}{⟨Content⟩}`

This key can be used in the setup command `\tagpdfsetup` and allow to declare a new attribute, which can be used as attribute or attribute class. The value are two brace groups, the first contains the name, the second the content.

```
\tagpdfsetup
{
  newattribute =
    {TH-col}{/O /Table /Scope /Column},
  newattribute =
    {TH-row}{/O /Table /Scope /Row},
}
```

`root-AF_␣(setup-key)` `root-AF = ⟨object name⟩`

This key can be used in the setup command `\tagpdfsetup` and allows to add associated files to the root structure. Like `AF` it can be used more than once to add more than one file.

```
1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-struct-code} {2022-12-22} {0.98}
4 {part of tagpdf - code related to storing structure}
5 </header>
```

3 Variables

`\c@g__tag_struct_abs_int` Every structure will have a unique, absolute number. I will use a latex counter for the structure count to have a chance to avoid double structures in align etc.

```
6 <base>\newcounter { g__tag_struct_abs_int }
7 <base>\int_gzero:N \c@g__tag_struct_abs_int
```

(End definition for \c@g__tag_struct_abs_int.)

`\g__tag_struct_objR_seq` a sequence to store mapping between the structure number and the object number. We assume that structure numbers are assign consecutively and so the index of the seq can be used. A seq allows easy mapping over the structures.

```
8 <*package>
9 \__tag_seq_new:N \g__tag_struct_objR_seq
```

(End definition for \g__tag_struct_objR_seq.)

`\g__tag_struct_cont_mc_prop` in generic mode it can happen after a page break that we have to inject into a structure sequence an additional mc after. We will store this additional info in a property. The key is the absolut mc num, the value the pdf directory.

```
10 \__tag_prop_new:N \g__tag_struct_cont_mc_prop
```

(End definition for \g__tag_struct_cont_mc_prop.)

`\g__tag_struct_stack_seq` A stack sequence for the structure stack. When a sequence is opened it's number is put on the stack.

```

11 \seq_new:N      \g__tag_struct_stack_seq
12 \seq_gpush:Nn  \g__tag_struct_stack_seq {0}

(End definition for \g__tag_struct_stack_seq.)

```

`\g__tag_struct_tag_stack_seq` We will perhaps also need the tags. While it is possible to get them from the numbered stack, lets build a tag stack too.

```

13 \seq_new:N      \g__tag_struct_tag_stack_seq
14 \seq_gpush:Nn  \g__tag_struct_tag_stack_seq {Root}

(End definition for \g__tag_struct_tag_stack_seq.)

```

`\g__tag_struct_tag_NS_prop` For the parent-child check, we need the tag and NS of every structure

```

15 \prop_new:N\g__tag_struct_tag_NS_prop

(End definition for \g__tag_struct_tag_NS_prop.)

```

`\g_tag_struct_stack_current_tl` The global variable will hold the current structure number. It is already defined in `tagpdf-base`.
`\l__tag_struct_stack_parent_tmpa_tl` The local temporary variable will hold the parent when we fetch it from the stack.

```

16 </package>
17 <base>\tl_new:N \g__tag_struct_stack_current_tl
18 <base>\tl_gset:Nn \g__tag_struct_stack_current_tl {\int_use:N\c@g__tag_struct_abs_int}
19 <*package>
20 \tl_new:N      \l__tag_struct_stack_parent_tmpa_tl

(End definition for \g__tag_struct_stack_current_tl and \l__tag_struct_stack_parent_tmpa_tl.)

```

I will need at least one structure: the StructTreeRoot normally it should have only one kid, e.g. the document element.

The data of the StructTreeRoot and the StructElem are in properties: `\g_@@_struct_0_prop` for the root and `\g_@@_struct_N_prop`, $N \geq 1$ for the other.

This creates quite a number of properties, so perhaps we will have to do this more efficiently in the future.

All properties have at least the keys

Type StructTreeRoot or StructElem

and the keys from the two following lists (the root has a special set of properties). the values of the prop should be already escaped properly when the entries are created (title,lang,alt,E,actualtext)

`\c__tag_struct_StructTreeRoot_entries_seq` These seq contain the keys we support in the two object types. They are currently no longer used, but are provided as documentation and for potential future checks. They should be adapted if there are changes in the PDF format.

```

21 \seq_const_from_clist:Nn \c__tag_struct_StructTreeRoot_entries_seq
22   {%p. 857/858
23     Type,                % always /StructTreeRoot
24     K,                   % kid, dictionary or array of dictionaries
25     IDTree,              % currently unused
26     ParentTree,          % required,obj ref to the parent tree
27     ParentTreeNextKey,   % optional
28     RoleMap,

```

```

29     ClassMap,
30     Namespaces,
31     AF                                %pdf 2.0
32 }
33
34 \seq_const_from_clist:Nn \c__tag_struct_StructElem_entries_seq
35 {%p 858 f
36     Type,                            %always /StructElem
37     S,                              %tag/type
38     P,                              %parent
39     ID,                             %optional
40     Ref,                            %optional, pdf 2.0 Use?
41     Pg,                             %obj num of starting page, optional
42     K,                              %kids
43     A,                              %attributes, probably unused
44     C,                              %class ""
45     %R,                             %attribute revision number, irrelevant for us as we
46                                     % don't update/change existing PDF and (probably)
47                                     % deprecated in PDF 2.0
48     T,                              %title, value in () or <>
49     Lang,                           %language
50     Alt,                            % value in () or <>
51     E,                              % abbreviation
52     ActualText,
53     AF,                             %pdf 2.0, array of dict, associated files
54     NS,                             %pdf 2.0, dict, namespace
55     PhoneticAlphabet,               %pdf 2.0
56     Phoneme                         %pdf 2.0
57 }

```

(End definition for \c__tag_struct_StructTreeRoot_entries_seq and \c__tag_struct_StructElem_entries_seq.)

3.1 Variables used by the keys

\g__tag_struct_tag_tl Use by the tag key to store the tag and the namespace.

```

58 \tl_new:N \g__tag_struct_tag_tl
59 \tl_new:N \g__tag_struct_tag_NS_tl

```

(End definition for \g__tag_struct_tag_tl and \g__tag_struct_tag_NS_tl.)

\l__tag_struct_key_label_tl This will hold the label value.

```

60 \tl_new:N \l__tag_struct_key_label_tl

```

(End definition for \l__tag_struct_key_label_tl.)

\l__tag_struct_elem_stash_bool This will keep track of the stash status

```

61 \bool_new:N \l__tag_struct_elem_stash_bool

```

(End definition for \l__tag_struct_elem_stash_bool.)

3.2 Variables used by tagging code of basic elements

`\g__tag_struct_dest_num_prop` This variable records for (some or all, not clear yet) destination names the related structure number to allow to reference them in a Ref. The key is the destination. It is currently used by the toc-tagging and sec-tagging code.

```
62 \prop_new:N \g__tag_struct_dest_num_prop
```

(End definition for `\g__tag_struct_dest_num_prop`.)

`\g__tag_struct_ref_by_dest_prop` This variable contains structures whose Ref key should be updated at the end to point to structured related with this destination. As this is probably need in other places too, it is not only a toc-variable.

```
63 \prop_new:N \g__tag_struct_ref_by_dest_prop
```

(End definition for `\g__tag_struct_ref_by_dest_prop`.)

4 Commands

The properties must be in some places handled expandably. So I need an output handler for each prop, to get expandable output see <https://tex.stackexchange.com/questions/424208>. There is probably room here for a more efficient implementation. TODO check if this can now be implemented with the pdfdict commands. The property contains currently non pdf keys, but e.g. object numbers are perhaps no longer needed as we have named object anyway.

```
\__tag_struct_output_prop_aux:nn
\__tag_new_output_prop_handler:n
64 \cs_new:Npn \__tag_struct_output_prop_aux:nn #1 #2 %#1 num, #2 key
65 {
66   \prop_if_in:cnT
67     { g__tag_struct_#1_prop }
68     { #2 }
69   {
70     \c_space_tl/#2~ \prop_item:cn{ g__tag_struct_#1_prop } { #2 }
71   }
72 }
73
74 \cs_new_protected:Npn \__tag_new_output_prop_handler:n #1
75 {
76   \cs_new:cn { __tag_struct_output_prop_#1:n }
77   {
78     \__tag_struct_output_prop_aux:nn {#1}{##1}
79   }
80 }
```

(End definition for `__tag_struct_output_prop_aux:nn` and `__tag_new_output_prop_handler:n`.)

4.1 Initialization of the StructTreeRoot

The first structure element, the StructTreeRoot is special, so created manually. The underlying object is @@/struct/0 which is currently created in the tree code (TODO move it here). The ParentTree and RoleMap entries are added at begin document in the tree code as they refer to object which are setup in other parts of the code. This avoid timing issues.

```

81 \tl_gset:Nn \g__tag_struct_stack_current_tl {0}

__tag_pdf_name_e:n

82 \cs_new:Npn \__tag_pdf_name_e:n #1{\pdf_name_from_unicode_e:n{#1}}
(End definition for \__tag_pdf_name_e:n.)

g__tag_struct_0_prop
g__tag_struct_kids_0_seq
83 \__tag_prop_new:c { g__tag_struct_0_prop }
84 \__tag_new_output_prop_handler:n {0}
85 \__tag_seq_new:c { g__tag_struct_kids_0_seq }
86
87 \__tag_prop_gput:cnx
88 { g__tag_struct_0_prop }
89 { Type }
90 { \pdf_name_from_unicode_e:n {StructTreeRoot} }
91
92 \__tag_prop_gput:cnx
93 { g__tag_struct_0_prop }
94 { S }
95 { \pdf_name_from_unicode_e:n {StructTreeRoot} }
96
97 \prop_gput:Nnn \g__tag_struct_tag_NS_prop {0}{{StructTreeRoot}{pdf}}

Namespaces are pdf 2.0 but it doesn't harm to have an empty entry. We could add a
test, but if the code moves into the kernel, timing could get tricky.

98 \__tag_prop_gput:cnx
99 { g__tag_struct_0_prop }
100 { Namespaces }
101 { \pdf_object_ref:n { __tag/tree/namespaces } }

(End definition for g__tag_struct_0_prop and g__tag_struct_kids_0_seq.)

```

4.2 Filling in the tag info

__tag_struct_set_tag_info:nnm This adds or updates the tag info to a structure given by a number. We need also the original data, so we store both.

```

102 \pdf_version_compare:NnTF < {2.0}
103 {
104   \cs_new_protected:Npn \__tag_struct_set_tag_info:nnm #1 #2 #3
105     %#1 structure number, #2 tag, #3 NS
106     {
107       \__tag_prop_gput:cnx
108       { g__tag_struct_#1_prop }
109       { S }
110       { \pdf_name_from_unicode_e:n {#2} } %
111       \prop_gput:Nnn \g__tag_struct_tag_NS_prop {#1}{{#2}{#3}}

```

```

112     }
113   }
114   {
115     \cs_new_protected:Npn \__tag_struct_set_tag_info:nnn #1 #2 #3
116     {
117       \__tag_prop_gput:cnx
118       { g__tag_struct_#1_prop }
119       { S }
120       { \pdf_name_from_unicode_e:n {#2} } %
121       \prop_get:NnNT \g__tag_role_NS_prop {#3} \l__tag_get_tmpc_tl
122       {
123         \__tag_prop_gput:cnx
124         { g__tag_struct_#1_prop }
125         { NS }
126         { \l__tag_get_tmpc_tl } %
127       }
128       \prop_gput:Nnn \g__tag_struct_tag_NS_prop {#1}{#2}{#3}
129     }
130   }
131   \cs_generate_variant:Nn \__tag_struct_set_tag_info:nnn {eVV}

```

(End definition for __tag_struct_set_tag_info:nnn.)

\cc We also need a way to get the tag info back from parent structures.

```

132 \cs_new_protected:Npn \__tag_struct_get_tag_info:nNN #1 #2 #3
133   {%#1 struct num, #2 tlvar for tag , #3 tlvar for NS
134   {
135     \prop_get:NnNTF \g__tag_struct_tag_NS_prop {#1}\l__tag_get_tmpc_tl
136     {
137       \tl_set:Nx #2{\exp_last_unbraced:NV\use_i:nn \l__tag_get_tmpc_tl}
138       \tl_set:Nx #3{\exp_last_unbraced:NV\use_ii:nn \l__tag_get_tmpc_tl}
139     }
140     {
141       \tl_clear:N#2
142       \tl_clear:N#3
143     }
144   }
145   \cs_generate_variant:Nn \__tag_struct_get_tag_info:nNN {eNN}

```

(End definition for \cc. This function is documented on page ??.)

4.3 Handlings kids

Commands to store the kids. Kids in a structure can be a reference to a mc-chunk, an object reference to another structure element, or a object reference to an annotation (through an OBJR object).

__tag_struct_kid_mc_gput_right:nn
__tag_struct_kid_mc_gput_right:nx

The command to store an mc-chunk, this is a dictionary of type MCR. It would be possible to write out the content directly as unnamed object and to store only the object reference, but probably this would be slower, and the PDF is more readable like this. The code doesn't try to avoid the use of the /Pg key by checking page numbers. That imho only slows down without much gain. In generic mode the page break code will perhaps have to insert an additional mcid after an existing one. For this we use a property list At first an auxiliary to write the MCID dict. This should normally be expanded!

```

146 \cs_new:Npn \__tag_struct_mcid_dict:n #1 %#1 MCID absnum
147 {
148   <<
149   /Type \c_space_tl /MCR \c_space_tl
150   /Pg
151   \c_space_tl
152   \pdf_pageobject_ref:n { \__tag_ref_value:enn{mcid-#1}{tagabspage}{1} }
153   /MCID \c_space_tl \__tag_ref_value:enn{mcid-#1}{tagmcid}{1}
154   >>
155 }
156 \cs_new_protected:Npn \__tag_struct_kid_mc_gput_right:nn #1 #2 %#1 structure num, #2 MCID absnum
157 {
158   \__tag_seq_gput_right:cx
159   { g__tag_struct_kids_#1_seq }
160   {
161     \__tag_struct_mcid_dict:n {#2}
162   }
163   \__tag_seq_gput_right:cn
164   { g__tag_struct_kids_#1_seq }
165   {
166     \prop_item:Nn \g__tag_struct_cont_mc_prop {#2}
167   }
168 }
169 \cs_generate_variant:Nn \__tag_struct_kid_mc_gput_right:nn {nx}
170

```

(End definition for __tag_struct_kid_mc_gput_right:nn.)

__tag_struct_kid_struct_gput_right:nn This commands adds a structure as kid. We only need to record the object reference in the sequence.

__tag_struct_kid_struct_gput_right:xx

```

171 \cs_new_protected:Npn \__tag_struct_kid_struct_gput_right:nn #1 #2 %#1 num of parent struct, #2 num of kid
172 {
173   \__tag_seq_gput_right:cx
174   { g__tag_struct_kids_#1_seq }
175   {
176     \pdf_object_ref:n { __tag/struct/#2 }
177   }
178 }
179
180 \cs_generate_variant:Nn \__tag_struct_kid_struct_gput_right:nn {xx}

```

(End definition for __tag_struct_kid_struct_gput_right:nn.)

__tag_struct_kid_OBJR_gput_right:nn At last the command to add an OBJR object. This has to write an object first. The first argument is the number of the parent structure, the second the (expanded) object reference of the annotation. The last argument is the page object reference

__tag_struct_kid_OBJR_gput_right:xxx

```

181 \cs_new_protected:Npn \__tag_struct_kid_OBJR_gput_right:nnn #1 #2 #3 %#1 num of parent struct,
182                                     %#2 obj reference
183                                     %#3 page object reference
184 {
185   \pdf_object_unnamed_write:nn
186   { dict }
187   {
188     /Type/OBJR/Obj~#2/Pg~#3

```

```

189     }
190     \__tag_seq_gput_right:cx
191     { g__tag_struct_kids_#1_seq }
192     {
193         \pdf_object_ref_last:
194     }
195 }
196
197 \cs_generate_variant:Nn\__tag_struct_kid_OBJR_gput_right:nnn { xxx }
198

```

(End definition for __tag_struct_kid_OBJR_gput_right:nnn.)

__tag_struct_exchange_kid_command:N
__tag_struct_exchange_kid_command:c

In luamode it can happen that a single kid in a structure is split at a page break into two or more mcid. In this case the lua code has to convert put the dictionary of the kid into an array. See issue 13 at tagpdf repo. We exchange the dummy command for the kids to mark this case.

```

199 \cs_new_protected:Npn\__tag_struct_exchange_kid_command:N #1 %#1 = seq var
200 {
201     \seq_gpop_left:NN #1 \l__tag_tmpa_tl
202     \regex_replace_once:nnN
203     { \c{\__tag_mc_insert_mcid_kids:n} }
204     { \c{\__tag_mc_insert_mcid_single_kids:n} }
205     \l__tag_tmpa_tl
206     \seq_gput_left:NV #1 \l__tag_tmpa_tl
207 }
208
209 \cs_generate_variant:Nn\__tag_struct_exchange_kid_command:N { c }

```

(End definition for __tag_struct_exchange_kid_command:N.)

__tag_struct_fill_kid_key:n

This command adds the kid info to the K entry. In lua mode the content contains commands which are expanded later. The argument is the structure number.

```

210 \cs_new_protected:Npn \__tag_struct_fill_kid_key:n #1 %#1 is the struct num
211 {
212     \bool_if:NF\g__tag_mode_lua_bool
213     {
214         \seq_clear:N \l__tag_tmpa_seq
215         \seq_map_inline:cn { g__tag_struct_kids_#1_seq }
216         { \seq_put_right:Nx \l__tag_tmpa_seq { ##1 } }
217         %\seq_show:c { g__tag_struct_kids_#1_seq }
218         %\seq_show:N \l__tag_tmpa_seq
219         \seq_remove_all:Nn \l__tag_tmpa_seq {}
220         %\seq_show:N \l__tag_tmpa_seq
221         \seq_gset_eq:cN { g__tag_struct_kids_#1_seq } \l__tag_tmpa_seq
222     }
223
224     \int_case:nnF
225     {
226         \seq_count:c
227         {
228             g__tag_struct_kids_#1_seq
229         }
230     }

```

```

231 {
232   { 0 }
233   { } %no kids, do nothing
234   { 1 } % 1 kid, insert
235   {
236     % in this case we need a special command in
237     % luamode to get the array right. See issue #13
238     \bool_if:NT\g__tag_mode_lua_bool
239     {
240       \__tag_struct_exchange_kid_command:c
241       {g__tag_struct_kids_#1_seq}
242     }
243     \__tag_prop_gput:cnx { g__tag_struct_#1_prop } {K}
244     {
245       \seq_item:cn
246       {
247         g__tag_struct_kids_#1_seq
248       }
249       {1}
250     }
251   } %
252 }
253 { %many kids, use an array
254   \__tag_prop_gput:cnx { g__tag_struct_#1_prop } {K}
255   {
256     [
257       \seq_use:cn
258       {
259         g__tag_struct_kids_#1_seq
260       }
261       {
262         \c_space_tl
263       }
264     ]
265   }
266 }
267 }
268

```

(End definition for __tag_struct_fill_kid_key:n.)

4.4 Output of the object

__tag_struct_get_dict_content:nN

This maps the dictionary content of a structure into a tl-var. Basically it does what \pdfdict_use:n does. TODO!! this looks over-complicated. Check if it can be done with pdfdict now.

```

269 \cs_new_protected:Npn \__tag_struct_get_dict_content:nN #1 #2 %#1: structure num
270 {
271   \tl_clear:N #2
272   \seq_map_inline:cn
273   {
274     c__tag_struct_
275     \int_compare:nNnTF{#1}={0}{StructTreeRoot}{StructElem}
276     _entries_seq

```

```

277     }
278     {
279         \tl_put_right:Nx
280         #2
281         {
282             \prop_if_in:cnT
283             { g__tag_struct_#1_prop }
284             { ##1 }
285             {
286                 \c_space_tl/##1~

```

Some keys needs the option to format the key, e.g. add brackets for an array

```

287             \cs_if_exist_use:cTF {__tag_struct_format_##1:e}
288             {
289                 { \prop_item:cn{ g__tag_struct_#1_prop } { ##1 } }
290             }
291             {
292                 \prop_item:cn{ g__tag_struct_#1_prop } { ##1 }
293             }
294         }
295     }
296 }
297 }

```

(End definition for __tag_struct_get_dict_content:nN.)

__tag_struct_format_Ref:n Ref is an array, we store only the content to be able to extend it so the formatting command adds the brackets:

```

298 \cs_new:Nn\__tag_struct_format_Ref:n{[#1]}
299 \cs_generate_variant:Nn\__tag_struct_format_Ref:n{e}

```

(End definition for __tag_struct_format_Ref:n.)

__tag_struct_write_obj:n This writes out the structure object. This is done in the finish code, in the tree module and guarded by the tree boolean.

```

300 \cs_new_protected:Npn \__tag_struct_write_obj:n #1 % #1 is the struct num
301 {
302     \pdf_object_if_exist:nTF { __tag/struct/#1 }
303     {
304         \__tag_struct_fill_kid_key:n { #1 }
305         \__tag_struct_get_dict_content:nN { #1 } \l__tag_tmpa_tl
306         \exp_args:Nx
307         \pdf_object_write:nnx
308         { __tag/struct/#1 }
309         {dict}
310         {
311             \l__tag_tmpa_tl
312         }
313     }
314     {
315         \msg_error:nnn { tag } { struct-no-objnum } { #1}
316     }
317 }

```

(End definition for __tag_struct_write_obj:n.)

`_tag_struct_insert_annot:nn` This is the command to insert an annotation into the structure. It can probably be used for xform too.

Annotations used as structure content must

1. add a StructParent integer to their dictionary
2. push the object reference as OBJR object in the structure
3. Add a Structparent/obj-nr reference to the parent tree.

For a link this looks like this

```
(1) \tag_struct_begin:n { tag=Link }
    \tag_mc_begin:n { tag=Link }
    \pdfannot_dict_put:nnx
      { link/URI }
      { StructParent }
      { \int_use:N\c@g_@@_parenttree_obj_int }
    <start link> link text <stop link>
(2+3) \@@_struct_insert_annot:nn {obj ref}{parent num}
      \tag_mc_end:
      \tag_struct_end:
```

```
318 \cs_new_protected:Npn \_tag_struct_insert_annot:nn #1 #2 %1 object reference to the annotat
319                                     %2 structparent number
320 {
321   \bool_if:NT \g__tag_active_struct_bool
322   {
323     %get the number of the parent structure:
324     \seq_get:NNF
325       \g__tag_struct_stack_seq
326       \l__tag_struct_stack_parent_tmpa_tl
327     {
328       \msg_error:nn { tag } { struct-faulty-nesting }
329     }
330     %put the obj number of the annot in the kid entry, this also creates
331     %the OBJR object
332     \ref_label:nn {__tag_objr_page_#2 }{ tagabspage }
333     \_tag_struct_kid_OBJR_gput_right:xxx
334     {
335       \l__tag_struct_stack_parent_tmpa_tl
336     }
337     {
338       #1 %
339     }
340     {
341       \pdf_pageobject_ref:n { \_tag_ref_value:nnn {__tag_objr_page_#2 }{ tagabspage } }
342     }
343     % add the parent obj number to the parent tree:
344     \exp_args:Nnx
345     \_tag_parenttree_add_objr:nn
346     {
347       #2
348     }
349     {
```



```

350         \pdf_object_ref:e { __tag/struct/\l__tag_struct_stack_parent_tmpa_tl }
351     }
352     % increase the int:
353     \stepcounter{ g__tag_parenttree_obj_int }
354 }
355 }

```

(End definition for __tag_struct_insert_annot:nn.)

__tag_get_data_struct_tag: this command allows \tag_get:n to get the current structure tag with the keyword **struct_tag**. We will need to handle nesting

```

356 \cs_new:Npn \__tag_get_data_struct_tag:
357 {
358     \exp_args:Ne
359     \tl_tail:n
360     {
361         \prop_item:cn {g__tag_struct_\g__tag_struct_stack_current_tl _prop}{S}
362     }
363 }
364 \</package>

```

(End definition for __tag_get_data_struct_tag:.)

__tag_get_data_struct_num: this command allows \tag_get:n to get the current structure number with the keyword **struct_num**. We will need to handle nesting

```

365 \< *base>
366 \cs_new:Npn \__tag_get_data_struct_num:
367 {
368     \g__tag_struct_stack_current_tl
369 }
370 \</base>

```

(End definition for __tag_get_data_struct_num:.)

5 Keys

This are the keys for the user commands. we store the tag in a variable. But we should be careful, it is only reliable at the begin.

```

label_(struct-key)
stash_(struct-key)
parent_(struct-key)
tag_(struct-key)
title_(struct-key)
title-o_(struct-key)
alt_(struct-key)
actualtext_(struct-key)
lang_(struct-key)
ref_(struct-key)
E_(struct-key)
371 \< *package>
372 \keys_define:nn { __tag / struct }
373 {
374     label .tl_set:N      = \l__tag_struct_key_label_tl,
375     stash .bool_set:N    = \l__tag_struct_elem_stash_bool,
376     parent .code:n       =
377     {
378         \bool_lazy_and:nnTF
379         {
380             \prop_if_exist_p:c { g__tag_struct_\int_eval:n {#1}_prop }
381         }
382         {
383             \int_compare_p:nNn {#1}<{\c@g__tag_struct_abs_int}
384         }

```

```

385         { \tl_set:Nx \l__tag_struct_stack_parent_tmpa_tl { \int_eval:n {#1} } }
386         {
387             \msg_warning:nxxx { tag } { struct-unknown }
388             { \int_eval:n {#1} }
389             { parent~key~ignored }
390         }
391     },
392     parent .default:n      = {-1},
393     tag .code:n            = % S property
394     {
395         \seq_set_split:Nne \l__tag_tmpa_seq { / } {#1/\prop_item:Ne\g__tag_role_tags_NS_prop{
396         \tl_gset:Nx \g__tag_struct_tag_tl { \seq_item:Nn\l__tag_tmpa_seq {1} }
397         \tl_gset:Nx \g__tag_struct_tag_NS_tl { \seq_item:Nn\l__tag_tmpa_seq {2} }
398         \__tag_check_structure_tag:N \g__tag_struct_tag_tl
399     },
400     title .code:n          = % T property
401     {
402         \str_set_convert:Nnnn
403         \l__tag_tmpa_str
404         { #1 }
405         { default }
406         { utf16/hex }
407         \__tag_prop_gput:cnx
408         { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
409         { T }
410         { <\l__tag_tmpa_str> }
411     },
412     title-o .code:n        = % T property
413     {
414         \str_set_convert:Nonn
415         \l__tag_tmpa_str
416         { #1 }
417         { default }
418         { utf16/hex }
419         \__tag_prop_gput:cnx
420         { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
421         { T }
422         { <\l__tag_tmpa_str> }
423     },
424     alt .code:n            = % Alt property
425     {
426         \str_set_convert:Noon
427         \l__tag_tmpa_str
428         { #1 }
429         { default }
430         { utf16/hex }
431         \__tag_prop_gput:cnx
432         { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
433         { Alt }
434         { <\l__tag_tmpa_str> }
435     },
436     alttext .meta:n        = {alt=#1},
437     actualtext .code:n     = % ActualText property
438     {

```

```

439     \str_set_convert:Noon
440     \l__tag_tmpa_str
441     { #1 }
442     { default }
443     { utf16/hex }
444     \__tag_prop_gput:cnx
445     { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
446     { ActualText }
447     { <\l__tag_tmpa_str>}
448   },
449   lang .code:n          = % Lang property
450   {
451     \__tag_prop_gput:cnx
452     { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
453     { Lang }
454     { (#1) }
455   },

```

Ref is an array, the brackets are added through the formatting command.

```

456   ref .code:n          = % ref property
457   {
458     \tl_clear:N\l__tag_tmpa_tl
459     \clist_map_inline:on {#1}
460     {
461       \tl_put_right:Nx \l__tag_tmpa_tl
462       {~\ref_value:nn{tagpdfstruct-##1}{tagstructobj} }
463     }
464     \__tag_struct_gput_data_ref:ee { \int_eval:n {\c@g__tag_struct_abs_int} } {\l__tag_tmpa_tl}
465   },
466   E .code:n           = % E property
467   {
468     \str_set_convert:Nnon
469     \l__tag_tmpa_str
470     { #1 }
471     { default }
472     { utf16/hex }
473     \__tag_prop_gput:cnx
474     { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
475     { E }
476     { <\l__tag_tmpa_str> }
477   },
478 }

```

(End definition for label (struct-key) and others. These functions are documented on page 84.)

AF_□(struct-key) keys for the AF keys (associated files). They use commands from l3pdffile! The stream variants use txt as extension to get the mimetype. TODO: check if this should be configurable. For math we will perhaps need another extension. AF is an array and can be used more than once, so we store it in a tl. which is expanded. AFinline can be use only once (more quite probably doesn't make sense).

```

479 \cs_new_protected:Npn \__tag_struct_add_AF:nn #1 #2 % #1 struct num #2 object name
480 {
481   \tl_if_exist:cTF
482   {
483     g__tag_struct_#1_AF_tl

```

```

484     }
485     {
486         \tl_gput_right:cx
487         { g__tag_struct_#1_AF_tl }
488         { ~ \pdf_object_ref:n {#2} }
489     }
490     {
491         \tl_new:c
492         { g__tag_struct_#1_AF_tl }
493         \tl_gset:cx
494         { g__tag_struct_#1_AF_tl }
495         { \pdf_object_ref:n {#2} }
496     }
497 }
498 \cs_generate_variant:Nn \__tag_struct_add_AF:nn {en,ee}
499 \keys_define:nn { __tag / struct }
500 {
501     AF .code:n          = % AF property
502     {
503         \pdf_object_if_exist:nTF {#1}
504         {
505             \__tag_struct_add_AF:en { \int_eval:n {\c@g__tag_struct_abs_int} }{#1}
506             \__tag_prop_gput:cnx
507             { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
508             { AF }
509             {
510                 [
511                     \tl_use:c
512                     { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_AF_tl }
513                 ]
514             }
515         }
516     }
517 }
518 },
519 ,AFinline .code:n =
520 {
521     \group_begin:
522     \pdf_object_if_exist:eF {__tag/fileobj\int_use:N\c@g__tag_struct_abs_int}
523     {
524         \pdffile_embed_stream:nxx
525         {#1}
526         {tag-AFfile\int_use:N\c@g__tag_struct_abs_int.txt}
527         {__tag/fileobj\int_use:N\c@g__tag_struct_abs_int}
528         \__tag_struct_add_AF:ee
529         { \int_eval:n {\c@g__tag_struct_abs_int} }
530         { __tag/fileobj\int_use:N\c@g__tag_struct_abs_int }
531         \__tag_prop_gput:cnx
532         { g__tag_struct_\int_use:N\c@g__tag_struct_abs_int_prop }
533         { AF }
534         {
535             [
536                 \tl_use:c

```

```

538         { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_AF_t1 }
539     ]
540 }
541 }
542 \group_end:
543 }
544 ,AFinline-o .code:n =
545 {
546     \group_begin:
547     \pdf_object_if_exist:eF {__tag/fileobj\int_use:N\c@g__tag_struct_abs_int}
548     {
549         \pdffile_embed_stream:oxx
550         {#1}
551         {tag-AFfile\int_use:N\c@g__tag_struct_abs_int.txt}
552         {__tag/fileobj\int_use:N\c@g__tag_struct_abs_int}
553         \__tag_struct_add_AF:ee
554         { \int_eval:n {\c@g__tag_struct_abs_int} }
555         { __tag/fileobj\int_use:N\c@g__tag_struct_abs_int }
556         \__tag_prop_gput:cnx
557         { g__tag_struct_\int_use:N\c@g__tag_struct_abs_int_prop }
558         { AF }
559         {
560             [
561                 \tl_use:c
562                 { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_AF_t1 }
563             ]
564         }
565     }
566     \group_end:
567 }
568 }

```

(End definition for AF (struct-key), AFinline (struct-key), and AFinline-o (struct-key). These functions are documented on page 85.)

root-AF_␣(setup-key) The root structure can take AF keys too, so we provide a key for it. This key is used with `\tagpdfsetup`, not in a structure!

```

569 \keys_define:nn { __tag / setup }
570 {
571     root-AF .code:n =
572     {
573         \pdf_object_if_exist:nTF {#1}
574         {
575             \__tag_struct_add_AF:en { 0 }{#1}
576             \__tag_prop_gput:cnx
577             { g__tag_struct_0_prop }
578             { AF }
579             {
580                 [
581                     \tl_use:c
582                     { g__tag_struct_0_AF_t1 }
583                 ]
584             }
585         }
586     }
587 }

```

```

586         {
587
588         }
589     },
590 }
591 </package>

```

(End definition for root-AF (setup-key). This function is documented on page 86.)

6 User commands

```

\tag_struct_begin:n
\tag_struct_end: 592 <base>\cs_new_protected:Npn \tag_struct_begin:n #1 {\int_gincr:N \c@g__tag_struct_abs_int}
593 <base>\cs_new_protected:Npn \tag_struct_end: {}
594 *package| debug)
595 <package>\cs_set_protected:Npn \tag_struct_begin:n #1 %#1 key-val
596 <debug>\cs_set_protected:Npn \tag_struct_begin:n #1 %#1 key-val
597 {
598 <package>\__tag_check_if_active_struct:T
599 <debug>\__tag_check_if_active_struct:TF
600 {
601     \group_begin:
602     \int_gincr:N \c@g__tag_struct_abs_int
603     \__tag_prop_new:c { g__tag_struct\_int_eval:n { \c@g__tag_struct_abs_int }_prop }
604     \__tag_new_output_prop_handler:n {\int_eval:n { \c@g__tag_struct_abs_int }}
605     \__tag_seq_new:c { g__tag_struct_kids\_int_eval:n { \c@g__tag_struct_abs_int }_seq}
606     \exp_args:Ne
607     \pdf_object_new:n
608     { \__tag/struct/\int_eval:n { \c@g__tag_struct_abs_int } }
609     \__tag_prop_gput:cno
610     { g__tag_struct\_int_eval:n { \c@g__tag_struct_abs_int }_prop }
611     { Type }
612     { /StructElem }
613     \tl_set:Nn \l__tag_struct_stack_parent_tmpa_tl {-1}
614     \keys_set:nn { __tag / struct} { #1 }
615
616     \__tag_struct_set_tag_info:eVV
617     { \int_eval:n {\c@g__tag_struct_abs_int} }
618     \g__tag_struct_tag_tl
619     \g__tag_struct_tag_NS_tl
620     \__tag_check_structure_has_tag:n { \int_eval:n {\c@g__tag_struct_abs_int} }
621     \tl_if_empty:NF
622     \l__tag_struct_key_label_tl
623     {
624         \__tag_ref_label:en{tagpdfstruct-\l__tag_struct_key_label_tl}{struct}
625     }

```

The structure number of the parent is either taken from the stack or has been set with the parent key.

```

625     \int_compare:nNnT { \l__tag_struct_stack_parent_tmpa_tl } = { -1 }
626     {
627         \seq_get:NNF
628         \g__tag_struct_stack_seq
629         \l__tag_struct_stack_parent_tmpa_tl

```

```

630         {
631             \msg_error:nn { tag } { struct-faulty-nesting }
632         }
633     }
634     \seq_gpush:NV \g__tag_struct_stack_seq          \c@g__tag_struct_abs_int
635     \seq_gpush:NV \g__tag_struct_tag_stack_seq      \g__tag_struct_tag_tl
636     \tl_gset:NV   \g__tag_struct_stack_current_tl \c@g__tag_struct_abs_int
637     %\seq_show:N   \g__tag_struct_stack_seq
638     \bool_if:NF
639         \l__tag_struct_elem_stash_bool
640     {

```

check if the tag can be used inside the parent. It only makes sense, if the structure is actually used here, so it is guarded by the stash boolean. For now we ignore the namespace!

```

641         \__tag_struct_get_tag_info:eNN
642         { \l__tag_struct_stack_parent_tmpa_tl }
643         \l__tag_get_parent_tmpa_tl
644         \l__tag_get_parent_tmpb_tl
645         \__tag_check_parent_child:VVVVN
646         \l__tag_get_parent_tmpa_tl
647         \l__tag_get_parent_tmpb_tl
648         \g__tag_struct_tag_tl
649         \g__tag_struct_tag_NS_tl
650         \l__tag_parent_child_check_tl
651         \int_compare:nNnT { \l__tag_parent_child_check_tl } < 0
652         {
653             \msg_warning:nnxxx
654             { tag }
655             { role-parent-child }
656             { \l__tag_get_parent_tmpa_tl / \l__tag_get_parent_tmpb_tl }
657             { \g__tag_struct_tag_tl / \g__tag_struct_tag_NS_tl }
658             { not-allowed~(struct~\g__tag_struct_stack_current_tl) }
659             \cs_set_eq:NN \l__tag_role_remap_tag_tl \g__tag_struct_tag_tl
660             \cs_set_eq:NN \l__tag_role_remap_NS_tl \g__tag_struct_tag_NS_tl
661             \__tag_role_remap:
662             \cs_gset_eq:NN \g__tag_struct_tag_tl \l__tag_role_remap_tag_tl
663             \cs_gset_eq:NN \g__tag_struct_tag_NS_tl \l__tag_role_remap_NS_tl
664             \__tag_struct_set_tag_info:eVV
665             { \int_eval:n { \c@g__tag_struct_abs_int } }
666             \g__tag_struct_tag_tl
667             \g__tag_struct_tag_NS_tl
668         }

```

Set the Parent.

```

669         \__tag_prop_gput:cnx
670         { \g__tag_struct_\int_eval:n { \c@g__tag_struct_abs_int }_prop }
671         { P }
672         {
673             \pdf_object_ref:e { __tag/struct/\l__tag_struct_stack_parent_tmpa_tl }
674         }
675         %record this structure as kid:
676         %\tl_show:N \g__tag_struct_stack_current_tl
677         %\tl_show:N \l__tag_struct_stack_parent_tmpa_tl
678         \__tag_struct_kid_struct_gput_right:xx

```

```

679         { \l__tag_struct_stack_parent_tmpa_tl }
680         { \g__tag_struct_stack_current_tl }
681         %\prop_show:c { g__tag_struct_\g__tag_struct_stack_current_tl _prop }
682         %\seq_show:c {g__tag_struct_kids_\l__tag_struct_stack_parent_tmpa_tl _seq}
683     }
684     %\prop_show:c { g__tag_struct_\g__tag_struct_stack_current_tl _prop }
685     %\seq_show:c {g__tag_struct_kids_\l__tag_struct_stack_parent_tmpa_tl _seq}
686 <debug> \__tag_debug_struct_begin_insert:n { #1 }
687 \group_end:
688 }
689 <debug>{ \__tag_debug_struct_begin_ignore:n { #1 }}
690 }
691 <package>\cs_set_protected:Nn \tag_struct_end:
692 <debug>\cs_set_protected:Nn \tag_struct_end:
693 { %take the current structure num from the stack:
694   %the objects are written later, lua mode hasn't all needed info yet
695   %\seq_show:N \g__tag_struct_stack_seq
696 <package>\__tag_check_if_active_struct:T
697 <debug>\__tag_check_if_active_struct:TF
698 {
699   \seq_gpop:NN \g__tag_struct_tag_stack_seq \l__tag_tmpa_tl
700   \seq_gpop:NNTF \g__tag_struct_stack_seq \l__tag_tmpa_tl
701   {
702     \__tag_check_info_closing_struct:o { \g__tag_struct_stack_current_tl }
703   }
704   { \__tag_check_no_open_struct: }
705   % get the previous one, shouldn't be empty as the root should be there
706   \seq_get:NNTF \g__tag_struct_stack_seq \l__tag_tmpa_tl
707   {
708     \tl_gset:NV \g__tag_struct_stack_current_tl \l__tag_tmpa_tl
709   }
710   {
711     \__tag_check_no_open_struct:
712   }
713   \seq_get:NNT \g__tag_struct_tag_stack_seq \l__tag_tmpa_tl
714   {
715     \tl_gset:NV \g__tag_struct_tag_tl \l__tag_tmpa_tl
716   }
717 <debug>\__tag_debug_struct_end_insert:
718 }
719 <debug>{\__tag_debug_struct_end_ignore:}
720 }
721 </package|debug>

```

(End definition for \tag_struct_begin:n and \tag_struct_end:. These functions are documented on page 83.)

\tag_struct_use:n This command allows to use a stashed structure in another place. TODO: decide how it should be guarded. Probably by the struct-check.

```

722 <base>\cs_new_protected:Npn \tag_struct_use:n #1 {}
723 <*package>
724 \cs_set_protected:Npn \tag_struct_use:n #1 %#1 is the label
725 {
726   \__tag_check_if_active_struct:T

```



```

727 {
728   \prop_if_exist:cTF
729   { g__tag_struct\___tag_ref_value:enn{tagpdfstruct-#1}{tagstruct}{unknown}_prop } %
730   {
731     \__tag_check_struct_used:n {#1}
732     %add the label structure as kid to the current structure (can be the root)
733     \__tag_struct_kid_struct_gput_right:xx
734     { \g__tag_struct_stack_current_tl }
735     { \__tag_ref_value:enn{tagpdfstruct-#1}{tagstruct}{0} }
736     %add the current structure to the labeled one as parents
737     \__tag_prop_gput:cnx
738     { g__tag_struct\___tag_ref_value:enn{tagpdfstruct-#1}{tagstruct}{0}_prop }
739     { P }
740     {
741       \pdf_object_ref:e { __tag/struct/\g__tag_struct_stack_current_tl }
742     }
743   }
744   \__tag_struct_get_tag_info:eNN
745   { \__tag_ref_value:enn{tagpdfstruct-#1}{tagstruct}{0} }
746   \l__tag_tmpa_tl
747   \l__tag_tmpb_tl
748   \__tag_check_parent_child:VVVVN
749   \g__tag_struct_tag_tl
750   \g__tag_struct_tag_NS_tl
751   \l__tag_tmpa_tl
752   \l__tag_tmpb_tl
753   \l__tag_parent_child_check_tl
754   \int_compare:nNnT { \l__tag_parent_child_check_tl } < 0
755   {
756     \cs_set_eq:NN \l__tag_role_remap_tag_tl \g__tag_struct_tag_tl
757     \cs_set_eq:NN \l__tag_role_remap_NS_tl \g__tag_struct_tag_NS_tl
758     \__tag_role_remap:
759     \cs_gset_eq:NN \g__tag_struct_tag_tl \l__tag_role_remap_tag_tl
760     \cs_gset_eq:NN \g__tag_struct_tag_NS_tl \l__tag_role_remap_NS_tl
761     \__tag_struct_set_tag_info:eVV
762     { \int_eval:n { \c@g__tag_struct_abs_int } }
763     \g__tag_struct_tag_tl
764     \g__tag_struct_tag_NS_tl
765   }
766   {
767     \msg_warning:nnn{ tag }{struct-label-unknown}{#1}
768   }
769 }
770 }
771 \end{package}

```

(End definition for `\tag_struct_use:n`. This function is documented on page 83.)

`\tag_struct_object_ref:n` This is a command that allows to reference a structure. The argument is the number which can be get for the current structure with `\tag_get:n{struct_num}` TODO check if it should be in base too.

```

772 (*package)
773 \cs_new:Npn \tag_struct_object_ref:n #1
774 {
775   \pdf_object_ref:n {__tag/struct/#1}
776 }
777 \cs_generate_variant:Nn \tag_struct_object_ref:n {e}

(End definition for \tag_struct_object_ref:n. This function is documented on page 83.)

```

\tag_struct_gput:nnn This is a command that allows to update the data of a structure. The first argument is the number of the structure, the second a keyword referring to a function, the third the value. Currently the only keyword is **ref**

```

778 \cs_new_protected:Npn \tag_struct_gput:nnn #1 #2 #3
779 {
780   \cs_if_exist_use:cF {__tag_struct_gput_data_#2:nn}
781   { %warning??
782     \use_none:nn
783   }
784   {#1}{#3}
785 }
786 \cs_generate_variant:Nn \tag_struct_gput:nnn {ene,nne}

(End definition for \tag_struct_gput:nnn. This function is documented on page ??.)

```

__tag_struct_gput_data_ref:nn

```

787 \cs_new_protected:Npn \__tag_struct_gput_data_ref:nn #1 #2
788   % #1 receiving struct num, #2 list of object ref
789   {
790     \prop_get:cnN
791     { g__tag_struct_#1_prop }
792     {Ref}
793     \l__tag_get_tmpc_tl
794     \__tag_prop_gput:cnx
795     { g__tag_struct_#1_prop }
796     { Ref }
797     { \quark_if_no_value:NF\l__tag_get_tmpc_tl { \l__tag_get_tmpc_tl\c_space_tl }#2 }
798   }
799 \cs_generate_variant:Nn \__tag_struct_gput_data_ref:nn {ee}

(End definition for \__tag_struct_gput_data_ref:nn.)

```

\tag_struct_insert_annot:nn This are the user command to insert annotations. They must be used together to get the numbers right. They use a counter to the **StructParent** and **\tag_struct_insert_annot:nn** increases the counter given back by **\tag_struct_parent_int:.**

\tag_struct_insert_annot:xx It must be used together with **\tag_struct_parent_int:** to insert an annotation.

\tag_struct_parent_int: TODO: decide how it should be guarded if tagging is deactivated.

```

800 \cs_new_protected:Npn \tag_struct_insert_annot:nn #1 #2 %#1 should be an object reference
801                                     %#2 struct parent num
802 {
803   \__tag_check_if_active_struct:T
804   {
805     \__tag_struct_insert_annot:nn {#1}{#2}
806   }
807 }

```

```

808
809 \cs_generate_variant:Nn \tag_struct_insert_annot:nn {xx}
810 \cs_new:Npn \tag_struct_parent_int: {\int_use:c { c@g__tag_parenttree_obj_int }}
811
812 \</package>
813

```

(End definition for `\tag_struct_insert_annot:nn` and `\tag_struct_parent_int:`. These functions are documented on page 83.)

7 Attributes and attribute classes

```

814 \*header>
815 \ProvidesExplPackage {tagpdf-attr-code} {2022-12-22} {0.98}
816 {part of tagpdf - code related to attributes and attribute classes}
817 \</header>

```

7.1 Variables

<pre> \g__tag_attr_entries_prop \g__tag_attr_class_used_seq \g__tag_attr_objref_prop \l__tag_attr_value_tl </pre>	<pre> \g__@@_attr_entries_prop will store attribute names and their dictionary content. \g__@@_attr_class_used_seq will hold the attributes which have been used as class name. \l__@@_attr_value_tl is used to build the attribute array or key. Everytime an attribute is used for the first time, and object is created with its content, the name-object reference relation is stored in \g__@@_attr_objref_prop </pre>
---	---

```

818 \*package>
819 \prop_new:N \g__tag_attr_entries_prop
820 \seq_new:N \g__tag_attr_class_used_seq
821 \tl_new:N \l__tag_attr_value_tl
822 \prop_new:N \g__tag_attr_objref_prop %will contain obj num of used attributes

```

(End definition for `\g__tag_attr_entries_prop` and others.)

7.2 Commands and keys

<pre> __tag_attr_new_entry:nn newattribute_␣(setup-key) </pre>	<p>This allows to define attributes. Defined attributes are stored in a global property. <code>newattribute</code> expects two brace group, the name and the content. The content typically needs an <code>/O</code> key for the owner. An example look like this.</p>
---	--

```

\tagpdfsetup
{
  newattribute =
    {TH-col}{/O /Table /Scope /Column},
  newattribute =
    {TH-row}{/O /Table /Scope /Row},
}

823 \cs_new_protected:Npn \__tag_attr_new_entry:nn #1 #2 %#1:name, #2: content
824 {
825   \prop_gput:Nen \g__tag_attr_entries_prop
826     {\pdf_name_from_unicode_e:n{#1}}{#2}
827 }
828
829 \keys_define:nn { __tag / setup }
830 {

```

```

831     newattribute .code:n =
832     {
833       \__tag_attr_new_entry:nn #1
834     }
835   }

```

(End definition for __tag_attr_new_entry:nn and newattribute (setup-key). This function is documented on page 86.)

attribute-class_␣(struct-key) attribute-class has to store the used attribute names so that they can be added to the ClassMap later.

```

836 \keys_define:nn { __tag / struct }
837 {
838   attribute-class .code:n =
839   {
840     \clist_set:No \l__tag_tmpa_clist { #1 }
841     \seq_set_from_clist:NN \l__tag_tmpb_seq \l__tag_tmpa_clist

```

we convert the names into pdf names with slash

```

842     \seq_set_map_x:NNn \l__tag_tmpa_seq \l__tag_tmpb_seq
843     {
844       \pdf_name_from_unicode_e:n {##1}
845     }
846     \seq_map_inline:Nn \l__tag_tmpa_seq
847     {
848       \prop_if_in:NnF \g__tag_attr_entries_prop {##1}
849       {
850         \msg_error:nnn { tag } { attr-unknown } { ##1 }
851       }
852       \seq_gput_left:Nn\g__tag_attr_class_used_seq { ##1}
853     }
854     \tl_set:Nx \l__tag_tmpa_tl
855     {
856       \int_compare:nT { \seq_count:N \l__tag_tmpa_seq > 1 }{[]}
857       \seq_use:Nn \l__tag_tmpa_seq { \c_space_tl }
858       \int_compare:nT { \seq_count:N \l__tag_tmpa_seq > 1 }{[]}
859     }
860     \int_compare:nT { \seq_count:N \l__tag_tmpa_seq > 0 }
861     {
862       \__tag_prop_gput:cnx
863       { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
864       { C }
865       { \l__tag_tmpa_tl }
866       %\prop_show:c { g__tag_struct_\int_eval:n {\c@g__tag_struct_abs_int}_prop }
867     }
868   }
869 }

```

(End definition for attribute-class (struct-key). This function is documented on page 85.)

attribute_␣(struct-key)

```

870 \keys_define:nn { __tag / struct }
871 {
872   attribute .code:n = % A property (attribute, value currently a dictionary)
873   {

```

```

874 \clist_set:No \l__tag_tmpa_clist { #1 }
875 \seq_set_from_clist:NN \l__tag_tmpb_seq \l__tag_tmpa_clist
we convert the names into pdf names with slash
876 \seq_set_map_x:NNn \l__tag_tmpa_seq \l__tag_tmpb_seq
877 {
878   \pdf_name_from_unicode_e:n {##1}
879 }
880 \tl_set:Nx \l__tag_attr_value_tl
881 {
882   \int_compare:nT { \seq_count:N \l__tag_tmpa_seq > 1 }{[}%]
883 }
884 \seq_map_inline:Nn \l__tag_tmpa_seq
885 {
886   \prop_if_in:NnF \g__tag_attr_entries_prop {##1}
887   {
888     \msg_error:nnn { tag } { attr-unknown } { ##1 }
889   }
890   \prop_if_in:NnF \g__tag_attr_objref_prop {##1}
891   {%\prop_show:N \g__tag_attr_entries_prop
892     \pdf_object_unnamed_write:nx
893     { dict }
894     {
895       \prop_item:Nn \g__tag_attr_entries_prop {##1}
896     }
897     \prop_gput:Nnx \g__tag_attr_objref_prop {##1} {\pdf_object_ref_last:}
898   }
899   \tl_put_right:Nx \l__tag_attr_value_tl
900   {
901     \c_space_tl
902     \prop_item:Nn \g__tag_attr_objref_prop {##1}
903   }
904 % \tl_show:N \l__tag_attr_value_tl
905 }
906 \tl_put_right:Nx \l__tag_attr_value_tl
907 { %[
908   \int_compare:nT { \seq_count:N \l__tag_tmpa_seq > 1 }{[}%]
909 }
910 % \tl_show:N \l__tag_attr_value_tl
911 \__tag_prop_gput:cnx
912 { g__tag_struct \int_eval:n {\c@g__tag_struct_abs_int}_prop }
913 { A }
914 { \l__tag_attr_value_tl }
915 },
916 }
917 \end{package}

```

(End definition for attribute (struct-key). This function is documented on page 85.)

Part VIII

The tagpdf-luatex.def Driver for luatex Part of the tagpdf package

```
1 <@@=tag>
2 <*luatex>
3 \ProvidesExplFile {tagpdf-luatex.def} {2022-12-22} {0.98}
4 {tagpdf~driver~for~luatex}
```

1 Loading the lua

The space code requires that the fall back font has been loaded and initialized, so we force that first. But perhaps this could be done in the kernel.

```
5 {
6   \fontencoding{TU}\fontfamily{lmr}\fontseries{m}\fontshape{n}\fontsize{10pt}{10pt}\selectfont
7 }
8 \lua_now:e { tagpdf=require('tagpdf.lua') }
```

The following defines wrappers around prop and seq commands to store the data also in lua tables. I probably want also lua tables I put them in the ltx.@@.tables namespaces. The tables will be named like the variables but without backslash. To access such a table with a dynamical name create a string and then use ltx.@@.tables[string]. Old code, I'm not quite sure if this was a good idea. Now I have mix of table in ltx.@@.tables and ltx.@@.mc/struct. And a lot is probably not needed. TODO: this should be cleaned up, but at least roles are currently using the table!

```

    \__tag_prop_new:N
    \__tag_seq_new:N
    \__tag_prop_gput:Nnn
    \__tag_seq_gput_right:Nn
    \__tag_seq_item:cn
    \__tag_prop_item:cn
    \__tag_seq_show:N
    \__tag_prop_show:N
9 \cs_set_protected:Npn \__tag_prop_new:N #1
10 {
11   \prop_new:N #1
12   \lua_now:e { ltx.@@.tables.\cs_to_str:N#1 = {} }
13 }
14
15
16 \cs_set_protected:Npn \__tag_seq_new:N #1
17 {
18   \seq_new:N #1
19   \lua_now:e { ltx.@@.tables.\cs_to_str:N#1 = {} }
20 }
21
22
23 \cs_set_protected:Npn \__tag_prop_gput:Nnn #1 #2 #3
24 {
25   \prop_gput:Nnn #1 { #2 } { #3 }
26   \lua_now:e { ltx.@@.tables.\cs_to_str:N#1 ["#2"] = "#3" }
27 }
28
29
```

```

30 \cs_set_protected:Npn \__tag_seq_gput_right:Nn #1 #2
31 {
32   \seq_gput_right:Nn #1 { #2 }
33   \lua_now:e { table.insert(ltx.__tag.tables.\cs_to_str:N#1, "#2") }
34 }
35
36 %Hm not quite sure about the naming
37
38 \cs_set:Npn \__tag_seq_item:cn #1 #2
39 {
40   \lua_now:e { tex.print(ltx.__tag.tables.#1[#2]) }
41 }
42
43 \cs_set:Npn \__tag_prop_item:cn #1 #2
44 {
45   \lua_now:e { tex.print(ltx.__tag.tables.#1["#2"]) }
46 }
47
48 %for debugging commands that show both the seq/prop and the lua tables
49 \cs_set_protected:Npn \__tag_seq_show:N #1
50 {
51   \seq_show:N #1
52   \lua_now:e { ltx.__tag.trace.log ("lua-sequence-array~\cs_to_str:N#1",1) }
53   \lua_now:e { ltx.__tag.trace.show_seq (ltx.__tag.tables.\cs_to_str:N#1) }
54 }
55
56 \cs_set_protected:Npn \__tag_prop_show:N #1
57 {
58   \prop_show:N #1
59   \lua_now:e {ltx.__tag.trace.log ("lua-property-table~\cs_to_str:N#1",1) }
60   \lua_now:e {ltx.__tag.trace.show_prop (ltx.__tag.tables.\cs_to_str:N#1) }
61 }

```

(End definition for __tag_prop_new:N and others.)

62 \langle /luatex \rangle

The module declaration

```

63  $\langle$ *lua $\rangle$ 
64 -- tagpdf.lua
65 -- Ulrike Fischer
66
67 local ProvidesLuaModule = {
68   name      = "tagpdf",
69   version   = "0.98",      --TAGVERSION
70   date      = "2022-12-22", --TAGDATE
71   description = "tagpdf lua code",
72   license    = "The LATEX Project Public License 1.3c"
73 }
74
75 if luatexbase and luatexbase.provides_module then
76   luatexbase.provides_module (ProvidesLuaModule)
77 end
78
79 --[[

```

```

80 The code has quite probably a number of problems
81 - more variables should be local instead of global
82 - the naming is not always consistent due to the development of the code
83 - the traversing of the shipout box must be tested with more complicated setups
84 - it should probably handle more node types
85 -
86 --]]
87

```

Some comments about the lua structure.

```

88 --[[
89 the main table is named ltx.__tag. It contains the functions and also the data
90 collected during the compilation.
91
92 ltx.__tag.mc      will contain mc connected data.
93 ltx.__tag.struct  will contain structure related data.
94 ltx.__tag.page    will contain page data
95 ltx.__tag.tables  contains also data from mc and struct (from older code). This needs cleaning
96                 There are certainly dublettes, but I don't dare yet ...
97 ltx.__tag.func    will contain (public) functions.
98 ltx.__tag.trace   will contain tracing/logging functions.
99 local funktions  starts with __
100 functions meant for users will be in ltx.tag
101
102 functions
103 ltx.__tag.func.get_num_from (tag):   takes a tag (string) and returns the id number
104 ltx.__tag.func.output_num_from (tag): takes a tag (string) and prints (to tex) the id number
105 ltx.__tag.func.get_tag_from (num):   takes a num and returns the tag
106 ltx.__tag.func.output_tag_from (num): takes a num and prints (to tex) the tag
107 ltx.__tag.func.store_mc_data (num,key,data): stores key=data in ltx.__tag.mc[num]
108 ltx.__tag.func.store_mc_label (label,num): stores label=num in ltx.__tag.mc.labels
109 ltx.__tag.func.store_mc_kid (mcnum,kid,page): stores the mc-kids of mcnum on page page
110 ltx.__tag.func.store_mc_in_page(mcnum,mcpagecnt,page): stores in the page table the number of
111 ltx.__tag.func.store_struct_mcabs (structnum,mcnum): stores relations structnum<->mcnum (abs.)
112 ltx.__tag.func.mc_insert_kids (mcnum): inserts the /K entries for mcnum by wandering through
113 ltx.__tag.func.mark_page_elements(box,mcpagecnt,mccntprev,mcopen,name,mctypeprev) : the main
114 ltx.__tag.func.mark_shipout (): a wrapper around the core function which inserts the last EM
115 ltx.__tag.func.fill_parent_tree_line (page): outputs the entries of the parenttree for this p
116 ltx.__tag.func.output_parenttree(): outputs the content of the parenttree
117 ltx.__tag.func.pdf_object_ref(name): outputs the object reference for the object name
118 ltx.__tag.func.markspaceon(), ltx.__tag.func.markspaceoff(): (de)activates the marking of pos
119 ltx.__tag.trace.show_mc_data (num,loglevel): shows ltx.__tag.mc[num] is the current log leve
120 ltx.__tag.trace.show_all_mc_data (max,loglevel): shows a maximum about mc's if the current l
121 ltx.__tag.trace.show_seq: shows a sequence (array)
122 ltx.__tag.trace.show_struct_data (num): shows data of structure num
123 ltx.__tag.trace.show_prop: shows a prop
124 ltx.__tag.trace.log
125 ltx.__tag.trace.showspace : boolean
126 --]]
127

```

This set-ups the main attribute registers. The mc_type attribute stores the type (P, Span etc) encoded as a num, The mc_cnt attribute stores the absolute number and allows so to see if a node belongs to the same mc-chunk.

The interwordspace attr is set by the function `@@_mark_spaces`, and marks the place where spaces should be inserted. The interwordfont attr is set by the function `@@_mark_spaces` too and stores the font, so that we can decide which font to use for the real space char.

```
128 local mctypeattributeid = luatexbase.new_attribute ("g__tag_mc_type_attr")
129 local mcntattributeid   = luatexbase.new_attribute ("g__tag_mc_cnt_attr")
130 local iwspaceattributeid = luatexbase.new_attribute ("g__tag_interwordspace_attr")
131 local iwfontattributeid  = luatexbase.new_attribute ("g__tag_interwordfont_attr")
```

with this token we can query the state of the boolean and so detect if unmarked nodes should be marked as attributes

```
132 local tagunmarkedbool= token.create("g__tag_tagunmarked_bool")
133 local truebool       = token.create("c_true_bool")
```

Now a number of local versions from global tables. Not all is perhaps needed, most node variants were copied from lua-debug.

```
134 local catlatex      = luatexbase.registernumber("catcodetable@latex")
135 local tableinsert    = table.insert
136 local nodeid         = node.id
137 local nodecopy       = node.copy
138 local nodegetattribute = node.get_attribute
139 local nodesetattribute = node.set_attribute
140 local nodehasattribute = node.has_attribute
141 local nodenew        = node.new
142 local nodetail       = node.tail
143 local nodeslide      = node.slide
144 local noderemove     = node.remove
145 local nodetraverseid = node.traverse_id
146 local nodetraverse   = node.traverse
147 local nodeinsertafter = node.insert_after
148 local nodeinsertbefore = node.insert_before
149 local pdfpageref     = pdf.pageref
150
151 local HLIST          = node.id("hlist")
152 local VLIST          = node.id("vlist")
153 local RULE            = node.id("rule")
154 local DISC            = node.id("disc")
155 local GLUE            = node.id("glue")
156 local GLYPH          = node.id("glyph")
157 local KERN            = node.id("kern")
158 local PENALTY         = node.id("penalty")
159 local LOCAL_PAR       = node.id("local_par")
160 local MATH            = node.id("math")
```

Now we setup the main table structure. ltx is used by other latex code too!

```
161 ltx          = ltx          or { }
162 ltx.__tag     = ltx.__tag     or { }
163 ltx.__tag.mc   = ltx.__tag.mc   or { } -- mc data
164 ltx.__tag.struct = ltx.__tag.struct or { } -- struct data
165 ltx.__tag.tables = ltx.__tag.tables or { } -- tables created with new prop and new seq.
166                                     -- wasn't a so great idea ...
167                                     -- g__tag_role_tags_seq used by tag<-> is in this table
168                                     -- used for pure lua tables too now!
169 ltx.__tag.page  = ltx.__tag.page  or { } -- page data, currently only i->{0->mcnum,1->mcnum}
170 ltx.__tag.trace = ltx.__tag.trace or { } -- show commands
```

```

171 ltx.__tag.func      = ltx.__tag.func    or { } -- functions
172 ltx.__tag.conf      = ltx.__tag.conf    or { } -- configuration variables

```

2 Logging functions

`__tag_log` This rather simple log function takes as argument a message (string) and a number and will output the message to the log/terminal if the current loglevel is greater or equal than num.

```

173 local __tag_log =
174   function (message,loglevel)
175     if (loglevel or 3) <= tex.count["l__tag_loglevel_int"] then
176       texio.write_nl("tagpdf: ".. message)
177     end
178   end
179
180 ltx.__tag.trace.log = __tag_log

```

(End definition for `__tag_log` and `ltx.__tag.trace.log`.)

`ltx.__tag.trace.show_seq` This shows the content of a seq as stored in the tables table. It is used by the `\@@_seq_show:N` function. It is not used in user commands, only for debugging, and so requires log level >0.

```

181 function ltx.__tag.trace.show_seq (seq)
182   if (type(seq) == "table") then
183     for i,v in ipairs(seq) do
184       __tag_log ("[" .. i .. "]" => " .. tostring(v),1)
185     end
186   else
187     __tag_log ("sequence " .. tostring(seq) .. " not found",1)
188   end
189 end

```

(End definition for `ltx.__tag.trace.show_seq`.)

`__tag_pairs_prop` This shows the content of a prop as stored in the tables table. It is used by the `\@@_prop_show:N` function.

```

190 local __tag_pairs_prop =
191   function (prop)
192     local a = {}
193     for n in pairs(prop) do tableinsert(a, n) end
194     table.sort(a)
195     local i = 0           -- iterator variable
196     local iter = function () -- iterator function
197       i = i + 1
198       if a[i] == nil then return nil
199       else return a[i], prop[a[i]]
200     end
201   end
202   return iter
203 end
204
205
206 function ltx.__tag.trace.show_prop (prop)

```

```

207 if (type(prop) == "table") then
208   for i,v in __tag_pairs_prop (prop) do
209     __tag_log ("[" .. i .. "] => " .. tostring(v),1)
210   end
211 else
212   __tag_log ("prop " .. tostring(prop) .. " not found or not a table",1)
213 end
214 end

```

(End definition for __tag_pairs_prop and ltx.__tag.trace.show_prop.)

ltx.__tag.trace.show_mc_data This shows some data for a mc given by num. If something is shown depends on the log level. The function is used by the following function and then in \ShowTagging

```

215 function ltx.__tag.trace.show_mc_data (num,loglevel)
216 if ltx.__tag and ltx.__tag.mc and ltx.__tag.mc[num] then
217   for k,v in pairs(ltx.__tag.mc[num]) do
218     __tag_log ("mc"..num..": "..tostring(k)..=>"..tostring(v),loglevel)
219   end
220   if ltx.__tag.mc[num]["kids"] then
221     __tag_log ("mc" .. num .. " has " .. #ltx.__tag.mc[num]["kids"] .. " kids",loglevel)
222     for k,v in ipairs(ltx.__tag.mc[num]["kids"]) do
223       __tag_log ("mc " .. num .. " kid "..k.." =>" .. v.kid.." on page " ..v.page,loglevel)
224     end
225   end
226 else
227   __tag_log ("mc"..num.." not found",loglevel)
228 end
229 end

```

(End definition for ltx.__tag.trace.show_mc_data.)

ltx.__tag.trace.show_all_mc_data This shows data for the mc's between min and max (numbers). It is used by the \ShowTagging function.

```

230 function ltx.__tag.trace.show_all_mc_data (min,max,loglevel)
231 for i = min, max do
232   ltx.__tag.trace.show_mc_data (i,loglevel)
233 end
234 texio.write_nl("")
235 end

```

(End definition for ltx.__tag.trace.show_all_mc_data.)

ltx.__tag.trace.show_struct_data This function shows some struct data. Unused but kept for debugging.

```

236 function ltx.__tag.trace.show_struct_data (num)
237 if ltx.__tag and ltx.__tag.struct and ltx.__tag.struct[num] then
238   for k,v in ipairs(ltx.__tag.struct[num]) do
239     __tag_log ("struct "..num..": "..tostring(k)..=>"..tostring(v),1)
240   end
241 else
242   __tag_log ("struct "..num.." not found ",1)
243 end
244 end

```

(End definition for ltx.__tag.trace.show_struct_data.)

3 Helper functions

3.1 Retrieve data functions

`__tag_get_mc_cnt_type_tag` This takes a node as argument and returns the mc-cnt, the mc-type and and the tag (calculated from the mc-cnt.

```

245 local __tag_get_mc_cnt_type_tag = function (n)
246   local mccnt      = nodegetattribute(n,mccntattributeid) or -1
247   local mctype     = nodegetattribute(n,mctypeattributeid) or -1
248   local tag        = ltx.__tag.func.get_tag_from(mctype)
249   return mccnt,mctype,tag
250 end

```

(End definition for `__tag_get_mc_cnt_type_tag`.)

`__tag_get_mathsubtype` This function allows to detect if we are at the begin or the end of math. It takes as argument a mathnode.

```

251 local function __tag_get_mathsubtype (mathnode)
252   if mathnode.subtype == 0 then
253     subtype = "beginmath"
254   else
255     subtype = "endmath"
256   end
257   return subtype
258 end

```

(End definition for `__tag_get_mathsubtype`.)

`ltx.__tag.tables.role_tag_attribute` The first is a table with key a tag and value a number (the attribute) The second is an array with the attribute value as key.

```

259 ltx.__tag.tables.role_tag_attribute = {}
260 ltx.__tag.tables.role_attribute_tag = {}

```

(End definition for `ltx.__tag.tables.role_tag_attribute`.)

`ltx.__tag.func.alloctag`

```

261 local __tag_alloctag =
262   function (tag)
263     if not ltx.__tag.tables.role_tag_attribute[tag] then
264       table.insert(ltx.__tag.tables.role_attribute_tag,tag)
265       ltx.__tag.tables.role_tag_attribute[tag]=#ltx.__tag.tables.role_attribute_tag
266       __tag_log ("Add " ..tag.." " ..ltx.__tag.tables.role_tag_attribute[tag],3)
267     end
268   end
269 ltx.__tag.func.alloctag = __tag_alloctag

```

(End definition for `ltx.__tag.func.alloctag`.)

`__tag_get_num_from` These functions take as argument a string `tag`, and return the number under which is it recorded (and so the attribute value). The first function outputs the number for lua, while the `output` function outputs to tex.

`ltx.__tag.func.get_num_from`
`ltx.__tag.func.output_num_from`

```

270 local __tag_get_num_from =
271   function (tag)
272     if ltx.__tag.tables.role_tag_attribute[tag] then

```

```

273     a = ltx.__tag.tables.role_tag_attribute[tag]
274     else
275         a = -1
276     end
277     return a
278 end
279
280 ltx.__tag.func.get_num_from = __tag_get_num_from
281
282 function ltx.__tag.func.output_num_from (tag)
283     local num = __tag_get_num_from (tag)
284     tex.sprint(catlatex,num)
285     if num == -1 then
286         __tag_log ("Unknown tag "..tag.." used")
287     end
288 end

```

(End definition for __tag_get_num_from, ltx.__tag.func.get_num_from, and ltx.__tag.func.output_num_from.)

__tag_get_tag_from These functions are the opposites to the previous function: they take as argument a number (the attribute value) and return the string tag. The first function outputs the string for lua, while the output function outputs to tex.

```

289 local __tag_get_tag_from =
290 function (num)
291     if ltx.__tag.tables.role_attribute_tag[num] then
292         a = ltx.__tag.tables.role_attribute_tag[num]
293     else
294         a = "UNKNOWN"
295     end
296     return a
297 end
298
299 ltx.__tag.func.get_tag_from = __tag_get_tag_from
300
301 function ltx.__tag.func.output_tag_from (num)
302     tex.sprint(catlatex,__tag_get_tag_from (num))
303 end

```

(End definition for __tag_get_tag_from, ltx.__tag.func.get_tag_from, and ltx.__tag.func.output_tag_from.)

ltx.__tag.func.store_mc_data This function stores for key=data for mc-chunk num. It is used in the tagpdf-mc code, to store for example the tag string, and the raw options.

```

304 function ltx.__tag.func.store_mc_data (num,key,data)
305     ltx.__tag.mc[num] = ltx.__tag.mc[num] or { }
306     ltx.__tag.mc[num][key] = data
307     __tag_log ("INFO TEX-STORE-MC-DATA: "..num.." => "..tostring(key).. " => "..tostring(data),3)
308 end

```

(End definition for ltx.__tag.func.store_mc_data.)

ltx.__tag.func.store_mc_label This function stores the label=num relationship in the labels subtable. TODO: this is probably unused and can go.

```

309 function ltx.__tag.func.store_mc_label (label,num)

```

```

310 ltx.__tag.mc["labels"] = ltx.__tag.mc["labels"] or { }
311 ltx.__tag.mc.labels[label] = num
312 end

```

(End definition for ltx.__tag.func.store_mc_label.)

ltx.__tag.func.store_mc_kid This function is used in the traversing code. It stores a sub-chunk of a mc mcnum into the kids table.

```

313 function ltx.__tag.func.store_mc_kid (mcnum,kid,page)
314 ltx.__tag.trace.log("INFO TAG-STORE-MC-KID: "..mcnum.." => " .. kid.." on page " .. page,3)
315 ltx.__tag.mc[mcnum]["kids"] = ltx.__tag.mc[mcnum]["kids"] or { }
316 local kidtable = {kid=kid,page=page}
317 tableinsert(ltx.__tag.mc[mcnum]["kids"], kidtable )
318 end

```

(End definition for ltx.__tag.func.store_mc_kid.)

ltx.__tag.func.mc_num_of_kids This function returns the number of kids a mc mcnum has. We need to account for the case that a mc can have no kids.

```

319 function ltx.__tag.func.mc_num_of_kids (mcnum)
320 local num = 0
321 if ltx.__tag.mc[mcnum] and ltx.__tag.mc[mcnum]["kids"] then
322     num = #ltx.__tag.mc[mcnum]["kids"]
323 end
324 ltx.__tag.trace.log ("INFO MC-KID-NUMBERS: " .. mcnum .. "has " .. num .. "KIDS",4)
325 return num
326 end

```

(End definition for ltx.__tag.func.mc_num_of_kids.)

3.2 Functions to insert the pdf literals

__tag_insert_emc_node This insert the emc node.

```

327 local function __tag_insert_emc_node (head,current)
328 local emcnode = nodenew("whatsit","pdf_literal")
329     emcnode.data = "EMC"
330     emcnode.mode=1
331     head = node.insert_before(head,current,emcnode)
332 return head
333 end

```

(End definition for __tag_insert_emc_node.)

__tag_insert_bmc_node This inserts a simple bmc node

```

334 local function __tag_insert_bmc_node (head,current,tag)
335 local bmcnode = nodenew("whatsit","pdf_literal")
336     bmcnode.data = "/"..tag.." BMC"
337     bmcnode.mode=1
338     head = node.insert_before(head,current,bmcnode)
339 return head
340 end

```

(End definition for __tag_insert_bmc_node.)

`__tag_insert_bdc_node` This inserts a bcd node with a fix dict. TODO: check if this is still used, now that we create properties.

```

341 local function __tag_insert_bdc_node (head,current,tag,dict)
342   local bdcnode = nodenew("whatsit","pdf_literal")
343   bdcnode.data = "/"..tag.."<<"..dict..">> BDC"
344   bdcnode.mode=1
345   head = node.insert_before(head,current,bdcnode)
346   return head
347 end

```

(End definition for __tag_insert_bdc_node.)

`__tag_pdf_object_ref` This allows to reference a pdf object reserved with the l3pdf command by name. The return value is `n 0 R`, if the object doesn't exist, `n` is 0. TODO: is uses internal l3pdf commands, this should be properly supported by l3pdf

`ltx.__tag.func.pdf_object_ref`

```

348 local function __tag_pdf_object_ref (name)
349   local tokenname = 'c__pdf_backend_object_'..name..'__int'
350   local object = token.create(tokenname).index..' 0 R'
351   return object
352 end
353 ltx.__tag.func.pdf_object_ref=__tag_pdf_object_ref

```

(End definition for __tag_pdf_object_ref and ltx.__tag.func.pdf_object_ref.)

4 Function for the real space chars

`__tag_show_spacemark` A debugging function, it is used to inserts red color markers in the places where space chars can go, it can have side effects so not always reliable, but ok.

```

354 local function __tag_show_spacemark (head,current,color,height)
355   local markcolor = color or "1 0 0"
356   local markheight = height or 10
357   local pdfstring = node.new("whatsit","pdf_literal")
358   pdfstring.data =
359     string.format("q "..markcolor.." RG "..markcolor.." rg 0.4 w 0 %g m 0 %g l S Q",-
360       3,markheight)
361   head = node.insert_after(head,current,pdfstring)
362   return head
363 end

```

(End definition for __tag_show_spacemark.)

`__tag_fakespace` This is used to define a lua version of `\pdffakespace`

`ltx.__tag.func.fakespace`

```

363 local function __tag_fakespace()
364   tex.setattribute(iwspaceattributeid,1)
365   tex.setattribute(iwfontattributeid,font.current())
366 end
367 ltx.__tag.func.fakespace = __tag_fakespace

```

(End definition for __tag_fakespace and ltx.__tag.func.fakespace.)

`__tag_mark_spaces` a function to mark up places where real space chars should be inserted. It only sets attributes, these are then be used in a later traversing which inserts the actual spaces. When space handling is activated this function is inserted in some callbacks.

```

368 --[[ a function to mark up places where real space chars should be inserted
369       it only sets an attribute.
370 --]]
371
372 local function __tag_mark_spaces (head)
373     local inside_math = false
374     for n in nodetraverse(head) do
375         local id = n.id
376         if id == GLYPH then
377             local glyph = n
378             if glyph.next and (glyph.next.id == GLUE)
379                 and not inside_math and (glyph.next.width > 0)
380             then
381                 nodesetattribute(glyph.next,iwspaceattributeid,1)
382                 nodesetattribute(glyph.next,iwfontattributeid,glyph.font)
383                 -- for debugging
384                 if ltx.__tag.trace.showspace then
385                     __tag_show_spacemark (head,glyph)
386                 end
387             elseif glyph.next and (glyph.next.id==KERN) and not inside_math then
388                 local kern = glyph.next
389                 if kern.next and (kern.next.id== GLUE) and (kern.next.width > 0)
390                 then
391                     nodesetattribute(kern.next,iwspaceattributeid,1)
392                     nodesetattribute(kern.next,iwfontattributeid,glyph.font)
393                 end
394             end
395             -- look also back
396             if glyph.prev and (glyph.prev.id == GLUE)
397                 and not inside_math
398                 and (glyph.prev.width > 0)
399                 and not nodehasattribute(glyph.prev,iwspaceattributeid)
400             then
401                 nodesetattribute(glyph.prev,iwspaceattributeid,1)
402                 nodesetattribute(glyph.prev,iwfontattributeid,glyph.font)
403                 -- for debugging
404                 if ltx.__tag.trace.showspace then
405                     __tag_show_spacemark (head,glyph)
406                 end
407             end
408             elseif id == PENALTY then
409                 local glyph = n
410                 -- ltx.__tag.trace.log ("PENALTY " .. n.subtype .. "VALUE" .. n.penalty,3)
411                 if glyph.next and (glyph.next.id == GLUE)
412                     and not inside_math and (glyph.next.width > 0) and n.subtype==0
413                 then
414                     nodesetattribute(glyph.next,iwspaceattributeid,1)
415                     -- nodesetattribute(glyph.next,iwfontattributeid,glyph.font)
416                     -- for debugging
417                     if ltx.__tag.trace.showspace then
418                         __tag_show_spacemark (head,glyph)

```



```

419         end
420     end
421     elseif id == MATH then
422         inside_math = (n.subtype == 0)
423     end
424 end
425 return head
426 end

```

(End definition for `__tag_mark_spaces`.)

```

__tag_activate_mark_space  These functions add/remove the function which marks the spaces to the callbacks
ltx.__tag.func.markspaceon pre_linebreak_filter and hpack_filter
ltx.__tag.func.markspaceoff

427 local function __tag_activate_mark_space ()
428     if not luatexbase.in_callback ("pre_linebreak_filter","markspaces") then
429         luatexbase.add_to_callback("pre_linebreak_filter",__tag_mark_spaces,"markspaces")
430         luatexbase.add_to_callback("hpack_filter",__tag_mark_spaces,"markspaces")
431     end
432 end
433
434 ltx.__tag.func.markspaceon=__tag_activate_mark_space
435
436 local function __tag_deactivate_mark_space ()
437     if luatexbase.in_callback ("pre_linebreak_filter","markspaces") then
438         luatexbase.remove_from_callback("pre_linebreak_filter","markspaces")
439         luatexbase.remove_from_callback("hpack_filter","markspaces")
440     end
441 end
442
443 ltx.__tag.func.markspaceoff=__tag_deactivate_mark_space

(End definition for __tag_activate_mark_space, ltx.__tag.func.markspaceon, and ltx.__tag.func.markspaceoff.)

We need two local variable to setup a default space char.

444 local default_space_char = node.new(GLYPH)
445 local default_fontid      = font.id("TU/lmr/m/n/10")
446 default_space_char.char    = 32
447 default_space_char.font    = default_fontid

```

`__tag_space_chars_shipout` These is the main function to insert real space chars. It inserts a glyph before every glue which has been marked previously. The attributes are copied from the glue, so if the tagging is done later, it will be tagged like it.

```

448 local function __tag_space_chars_shipout (box)
449     local head = box.head
450     if head then
451         for n in node.traverse(head) do
452             local spaceattr = node.getattribute(n,iwspaceattributeid) or -1
453             if n.id == HLIST then -- enter the hlist
454                 __tag_space_chars_shipout (n)
455             elseif n.id == VLIST then -- enter the vlist
456                 __tag_space_chars_shipout (n)
457             elseif n.id == GLUE then
458                 if ltx.__tag.trace.showspaces and spaceattr==1 then
459                     __tag_show_spacemark (head,n,"0 1 0")
460                 end

```

```

461     if spaceattr==1 then
462         local space
463         local space_char = node.copy(default_space_char)
464         local curfont = node.get_attribute(n, i, font_attribute_id)
465         ltx.__tag.trace.log("INFO SPACE-FUNCTION-FONT: ".. tostring(curfont), 3)
466         if curfont and luaotfload.aux.slot_of_name(curfont, "space") then
467             space_char.font = curfont
468         end
469         head, space = node.insert_before(head, n, space_char) --
470         n.width = n.width - space.width
471         space.attr = n.attr
472     end
473 end
474 end
475 end
476 end
477
478 function ltx.__tag.func.space_chars_shipout (box)
479     __tag_space_chars_shipout (box)
480 end

```

(End definition for `__tag_space_chars_shipout` and `ltx.__tag.func.space_chars_shipout`.)

5 Function for the tagging

`ltx.__tag.func.mc_insert_kids`

This is the main function to insert the K entry into a StructElem object. It is used in `tagpdf-mc-luacode` module. The `single` attribute allows to handle the case that a single mc on the tex side can have more than one kid after the processing here, and so we get the correct array/non array setup.

```

481 function ltx.__tag.func.mc_insert_kids (mcnum, single)
482     if ltx.__tag.mc[mcnum] then
483         ltx.__tag.trace.log("INFO TEX-MC-INSERT-KID-TEST: " .. mcnum, 4)
484         if ltx.__tag.mc[mcnum]["kids"] then
485             if #ltx.__tag.mc[mcnum]["kids"] > 1 and single==1 then
486                 tex.sprint("(")
487             end
488             for i, kidstable in ipairs( ltx.__tag.mc[mcnum]["kids"] ) do
489                 local kidnum = kidstable["kid"]
490                 local kidpage = kidstable["page"]
491                 local kidpageobjnum = pdfpageref(kidpage)
492                 ltx.__tag.trace.log("INFO TEX-MC-INSERT-KID: " .. mcnum ..
493                     " insert KID " .. i ..
494                     " with num " .. kidnum ..
495                     " on page " .. kidpage .. "/" .. kidpageobjnum, 3)
496                 tex.sprint(catlatex, "<</Type /MCR /Pg ".. kidpageobjnum .. " 0 R /MCID ".. kidnum .. ">> " )
497             end
498             if #ltx.__tag.mc[mcnum]["kids"] > 1 and single==1 then
499                 tex.sprint("]")
500             end
501         else
502             -- this is typically not a problem, e.g. empty hbox in footer/header can
503             -- trigger this warning.
504             ltx.__tag.trace.log("WARN TEX-MC-INSERT-NO-KIDS: ".. mcnum .. " has no kids", 2)

```

```

505     if single==1 then
506         tex.sprint("null")
507     end
508 end
509 else
510     ltx.__tag.trace.log("WARN TEX-MC-INSERT-MISSING: "..mcnum.." doesn't exist",0)
511 end
512 end

```

(End definition for ltx.__tag.func.mc_insert_kids.)

ltx.__tag.func.store_struct_mcabs This function is used in the tagpdf-mc-luacode. It store the absolute count of the mc into the current structure. This must be done ordered.

```

513 function ltx.__tag.func.store_struct_mcabs (structnum,mcnum)
514     ltx.__tag.struct[structnum]=ltx.__tag.struct[structnum] or { }
515     ltx.__tag.struct[structnum]["mc"]=ltx.__tag.struct[structnum]["mc"] or { }
516     -- a structure can contain more than on mc chunk, the content should be ordered
517     tableinsert(ltx.__tag.struct[structnum]["mc"],mcnum)
518     ltx.__tag.trace.log("INFO TEX-MC-INTO-STRUCT: "..
519         mcnum.." inserted in struct "..structnum,3)
520     -- but every mc can only be in one structure
521     ltx.__tag.mc[mcnum]= ltx.__tag.mc[mcnum] or { }
522     ltx.__tag.mc[mcnum]["parent"] = structnum
523 end
524

```

(End definition for ltx.__tag.func.store_struct_mcabs.)

ltx.__tag.func.store_mc_in_page This is used in the traversing code and stores the relation between abs count and page count.

```

525 -- pay attention: lua counts arrays from 1, tex pages from one
526 -- mcid and arrays in pdf count from 0.
527 function ltx.__tag.func.store_mc_in_page (mcnum,mcpagencnt,page)
528     ltx.__tag.page[page] = ltx.__tag.page[page] or {}
529     ltx.__tag.page[page][mcpagencnt] = mcnum
530     ltx.__tag.trace.log("INFO TAG-MC-INTO-PAGE: page " .. page ..
531         ": inserting MCID " .. mcpagencnt .. " => " .. mcnum,3)
532 end

```

(End definition for ltx.__tag.func.store_mc_in_page.)

ltx.__tag.func.mark_page_elements This is the main traversing function. See the lua comment for more details.

```

533 --[[
534     Now follows the core function
535     It wades through the shipout box and checks the attributes
536     ARGUMENTS
537     box: is a box,
538     mcpagencnt: num, the current page cnt of mc (should start at -1 in shipout box), needed for
539     mccntprev: num, the attribute cnt of the previous node/whatever - if different we have a c
540     mcopen: num, records if some bdc/emc is open
541     These arguments are only needed for log messages, if not present are replaces by fix stri
542     name: string to describe the box
543     mctypeprev: num, the type attribute of the previous node/whatever
544
545     there are lots of logging messages currently. Should be cleaned up in due course.

```

```

546     One should also find ways to make the function shorter.
547 --]]
548
549 function ltx.__tag.func.mark_page_elements (box,mcpagecnt,mccntprev,mcopen,name,mctypeprev)
550     local name = name or ("SOMEBOX")
551     local mctypeprev = mctypeprev or -1
552     local abspage = status.total_pages + 1 -- the real counter is increased
553                                           -- inside the box so one off
554                                           -- if the callback is not used. (???)
555     ltx.__tag.trace.log ("INFO TAG-ABSPAGE: " .. abspage,3)
556     ltx.__tag.trace.log ("INFO TAG-ARGS: pagecnt".. mcpagecnt..
557                         " prev "..mccntprev ..
558                         " type prev "..mctypeprev,4)
559     ltx.__tag.trace.log ("INFO TAG-TRAVERSING-BOX: ".. tostring(name)..
560                         " TYPE ".. node.type(node.getid(box)),3)
561     local head = box.head -- ShipoutBox is a vlist?
562     if head then
563         mccnthead, mctypehead,taghead = __tag_get_mc_cnt_type_tag (head)
564         ltx.__tag.trace.log ("INFO TAG-HEAD: " ..
565                             node.type(node.getid(head))..
566                             " MC"..tostring(mccnthead)..
567                             " => TAG " .. tostring(mctypehead)..
568                             " => ".. tostring(taghead),3)
569     else
570         ltx.__tag.trace.log ("INFO TAG-NO-HEAD: head is "..
571                             tostring(head),3)
572     end
573     for n in node.traverse(head) do
574         local mccnt, mctype, tag = __tag_get_mc_cnt_type_tag (n)
575         local spaceattr = node.getattribute(n,iwspaceattributeid) or -1
576         ltx.__tag.trace.log ("INFO TAG-NODE: "..
577                             node.type(node.getid(n))..
578                             " MC".. tostring(mccnt)..
579                             " => TAG ".. tostring(mctype)..
580                             " => " .. tostring(tag),3)
581         if n.id == HLIST
582         then -- enter the hlist
583             mcopen,mcpagecnt,mccntprev,mctypeprev=
584                 ltx.__tag.func.mark_page_elements (n,mcpagecnt,mccntprev,mcopen,"INTERNAL HLIST",mctypeprev)
585         elseif n.id == VLIST then -- enter the vlist
586             mcopen,mcpagecnt,mccntprev,mctypeprev=
587                 ltx.__tag.func.mark_page_elements (n,mcpagecnt,mccntprev,mcopen,"INTERNAL VLIST",mctypeprev)
588         elseif n.id == GLUE and not n.leader then -- at glue real space chars are inserted, but they
589                                                     -- been done if the previous shipout wandering, so here it
590         elseif n.id == LOCAL_PAR then -- local_par is ignored
591         elseif n.id == PENALTY then -- penalty is ignored
592         elseif n.id == KERN then -- kern is ignored
593             ltx.__tag.trace.log ("INFO TAG-KERN-SUBTYPE: "..
594                                 node.type(node.getid(n)).." "..n.subtype,4)
595         else
596             -- math is currently only logged.
597             -- we could mark the whole as math
598             -- for inner processing the mlist_to_hlist callback is probably needed.
599             if n.id == MATH then

```

```

600     ltx.__tag.trace.log("INFO TAG-MATH-SUBTYPE: "..
601         node.type(node.getid(n)).." " ..__tag_get_mathsubtype(n),4)
602 end
603 -- endmath
604 ltx.__tag.trace.log("INFO TAG-MC-COMPARE: current "..
605     mccnt.." prev "..mccntprev,4)
606 if mccnt~=mccntprev then -- a new mc chunk
607     ltx.__tag.trace.log ("INFO TAG-NEW-MC-NODE: "..
608         node.type(node.getid(n))..
609         " MC"..tostring(mccnt)..
610         " <=> PREVIOUS "..tostring(mccntprev),4)
611 if mcpopen~=0 then -- there is a chunk open, close it (hope there is only one ...
612     box.list=__tag_insert_emc_node (box.list,n)
613     mcpopen = mcpopen - 1
614     ltx.__tag.trace.log ("INFO TAG-INSERT-EMC: " ..
615         mcpagecnt .. " MCOPEN = " .. mcpopen,3)
616 if mcpopen ~=0 then
617     ltx.__tag.trace.log ("WARN TAG-OPEN-MC: " .. mcpopen,1)
618 end
619 end
620 if ltx.__tag.mc[mccnt] then
621     if ltx.__tag.mc[mccnt]["artifact"] then
622         ltx.__tag.trace.log("INFO TAG-INSERT-ARTIFACT: "..
623             tostring(ltx.__tag.mc[mccnt]["artifact"]),3)
624     if ltx.__tag.mc[mccnt]["artifact"] == "" then
625         box.list = __tag_insert_bmc_node (box.list,n,"Artifact")
626     else
627         box.list = __tag_insert_bdc_node (box.list,n,"Artifact", "/Type /"..ltx.__tag.mc[mccnt]
628     end
629 else
630     ltx.__tag.trace.log("INFO TAG-INSERT-TAG: "..
631         tostring(tag),3)
632     mcpagecnt = mcpagecnt +1
633     ltx.__tag.trace.log ("INFO TAG-INSERT-BDC: "..mcpagecnt,3)
634     local dict= "/MCID "..mcpagecnt
635     if ltx.__tag.mc[mccnt]["raw"] then
636         ltx.__tag.trace.log("INFO TAG-USE-RAW: "..
637             tostring(ltx.__tag.mc[mccnt]["raw"]),3)
638         dict= dict .. " " .. ltx.__tag.mc[mccnt]["raw"]
639     end
640     if ltx.__tag.mc[mccnt]["alt"] then
641         ltx.__tag.trace.log("INFO TAG-USE-ALT: "..
642             tostring(ltx.__tag.mc[mccnt]["alt"]),3)
643         dict= dict .. " " .. ltx.__tag.mc[mccnt]["alt"]
644     end
645     if ltx.__tag.mc[mccnt]["actualtext"] then
646         ltx.__tag.trace.log("INFO TAG-USE-ACTUALTEXT: "..
647             tostring(ltx.__tag.mc[mccnt]["actualtext"]),3)
648         dict= dict .. " " .. ltx.__tag.mc[mccnt]["actualtext"]
649     end
650     box.list = __tag_insert_bdc_node (box.list,n,tag, dict)
651     ltx.__tag.func.store_mc_kid (mccnt,mcpagecnt,abspage)
652     ltx.__tag.func.store_mc_in_page(mccnt,mcpagecnt,abspage)
653     ltx.__tag.trace.show_mc_data (mccnt,3)

```

```

654         end
655         mcopen = mcopen + 1
656     else
657         if tagunmarkedbool.mode == truebool.mode then
658             ltx.__tag.trace.log("INFO TAG-NOT-TAGGED: this has not been tagged, using artifact",2)
659             box.list = __tag_insert_bmc_node (box.list,n,"Artifact")
660             mcopen = mcopen + 1
661         else
662             ltx.__tag.trace.log("WARN TAG-NOT-TAGGED: this has not been tagged",1)
663         end
664     end
665     mccntprev = mccnt
666 end
667 end -- end if
668 end -- end for
669 if head then
670     mccnthead, mctypehead, taghead = __tag_get_mc_cnt_type_tag (head)
671     ltx.__tag.trace.log ("INFO TAG-ENDHEAD: " ..
672                         node.type(node.getid(head))..
673                         " MC"..tostring(mccnthead)..
674                         " => TAG "..tostring(mctypehead)..
675                         " => "..tostring(taghead),4)
676 else
677     ltx.__tag.trace.log ("INFO TAG-ENDHEAD: ".. tostring(head),4)
678 end
679 ltx.__tag.trace.log ("INFO TAG-QUITTING-BOX "..
680                     tostring(name)..
681                     " TYPE ".. node.type(node.getid(box)),4)
682 return mcopen,mcpagecnt,mccntprev,mctypeprev
683 end
684

```

(End definition for ltx.__tag.func.mark_page_elements.)

ltx.__tag.func.mark_shipout This is the function used in the callback. Beside calling the traversing function it also checks if there is an open MC-chunk from a page break and insert the needed EMC literal.

```

685 function ltx.__tag.func.mark_shipout (box)
686     mcopen = ltx.__tag.func.mark_page_elements (box,-1,-100,0,"Shipout",-1)
687     if mcopen~=0 then -- there is a chunk open, close it (hope there is only one ...
688         local emcnode = nodenew("whatsit","pdf_literal")
689         local list = box.list
690         emcnode.data = "EMC"
691         emcnode.mode=1
692         if list then
693             list = node.insert_after (list,node.tail(list),emcnode)
694             mcopen = mcopen - 1
695             ltx.__tag.trace.log ("INFO SHIPOUT-INSERT-LAST-EMC: MCOPEN " .. mcopen,3)
696         else
697             ltx.__tag.trace.log ("WARN SHIPOUT-UPS: this shouldn't happen",0)
698         end
699         if mcopen ~=0 then
700             ltx.__tag.trace.log ("WARN SHIPOUT-MC-OPEN: " .. mcopen,1)
701         end
702     end

```

```

702 end
703 end

```

(End definition for ltx.__tag.func.mark_shipout.)

6 Parenttree

```

ltx.__tag.func.fill_parent_tree_line
ltx.__tag.func.output_parenttree

```

These functions create the parent tree. The second, main function is used in the tagpdf-tree code. TODO check if the tree code can move into the backend code.

```

704 function ltx.__tag.func.fill_parent_tree_line (page)
705     -- we need to get page-> i=kid -> mcnum -> structnum
706     -- pay attention: the kid numbers and the page number in the parent tree start with 0!
707     local numsentry = ""
708     local pdfpage = page-1
709     if ltx.__tag.page[page] and ltx.__tag.page[page][0] then
710         mcchunks=#ltx.__tag.page[page]
711         ltx.__tag.trace.log("INFO PARENTTREE-NUM: page "..
712             page.." has "..mcchunks.." +1 Elements ",4)
713         for i=0,mcchunks do
714             -- what does this log??
715             ltx.__tag.trace.log("INFO PARENTTREE-CHUNKS:  "..
716                 ltx.__tag.page[page][i],4)
717         end
718         if mcchunks == 0 then
719             -- only one chunk so no need for an array
720             local mcnum = ltx.__tag.page[page][0]
721             local structnum = ltx.__tag.mc[mcnum]["parent"]
722             local propname = "g__tag_struct"..structnum.."_prop"
723             --local objref = ltx.__tag.tables[propname]["objref"] or "XXXX"
724             local objref = __tag_pdf_object_ref('__tag/struct/'..structnum)
725             ltx.__tag.trace.log("INFO PARENTTREE-STRUCT-OBJREF: =====>.."
726                 tostring(objref),5)
727             numsentry = pdfpage .. " [".. objref .. "]"
728             ltx.__tag.trace.log("INFO PARENTTREE-NUMENTRY: page " ..
729                 page.. " num entry = ".. numsentry,3)
730         else
731             numsentry = pdfpage .. " ["
732             for i=0,mcchunks do
733                 local mcnum = ltx.__tag.page[page][i]
734                 local structnum = ltx.__tag.mc[mcnum]["parent"] or 0
735                 local propname = "g__tag_struct"..structnum.."_prop"
736                 --local objref = ltx.__tag.tables[propname]["objref"] or "XXXX"
737                 local objref = __tag_pdf_object_ref('__tag/struct/'..structnum)
738                 numsentry = numsentry .. " [".. objref
739             end
740             numsentry = numsentry .. "]"
741             ltx.__tag.trace.log("INFO PARENTTREE-NUMENTRY: page " ..
742                 page.. " num entry = ".. numsentry,3)
743         end
744     else
745         ltx.__tag.trace.log ("INFO PARENTTREE-NO-DATA: page "..page,3)
746     end
747     return numsentry

```

```

748 end
749
750 function ltx.__tag.func.output_parenttree (abspage)
751   for i=1,abspage do
752     line = ltx.__tag.func.fill_parent_tree_line (i) .. "^^J"
753     tex.sprint(catlatex,line)
754   end
755 end

(End definition for ltx.__tag.func.fill_parent_tree_line and ltx.__tag.func.output_parenttree.)

756  $\langle$ /lua $\rangle$ 

```


Part IX

The tagpdf-roles module

Tags, roles and namespace code

Part of the tagpdf package

```
add-new-tag_(setup-key)
tag_(rolemap-key)
namespace_(rolemap-key)
role_(rolemap-key)
role-namespace_(rolemap-key)
```

The `add-new-tag` key can be used in `\tagpdfsetup` to declare and rolemap new tags. It takes as value a key-value list or a simple `new-tag/old-tag`.

The key-value list knows the following keys:

tag This is the name of the new tag as it should then be used in `\tagstructbegin`.

namespace This is the namespace of the new tag. The value should be a shorthand of a namespace. The allowed values are currently `pdf`, `pdf2`, `mathml`, `latex`, `latex-book` and `user`. The default value (and recommended value for a new tag) is `user`. The public name of the user namespace is `tag/NS/user`. This can be used to reference the namespace e.g. in attributes.

role This is the tag the tag should be mapped too. In a PDF 1.7 or earlier this is normally a tag from the `pdf` set, in PDF 2.0 from the `pdf`, `pdf2` and `mathml` set. It can also be a user tag. The tag must be declared before, as the code retrieves the class of the new tag from it. The PDF format allows mapping to be done transitively. But tagpdf can't/won't check such unusual role mapping.

role-namespace If the role is a known tag the default value is the default namespace of this tag. With this key a specific namespace can be forced.

Namespaces are mostly a PDF 2.0 property, but it doesn't harm to set them also in a PDF 1.7 or earlier.

```
1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-roles-code} {2022-12-22} {0.98}
4 {part of tagpdf - code related to roles and structure names}
5 </header>
```

1 Code related to roles and structure names

```
6 <*package>
```

1.1 Variables

Tags are used in structures (`\tagstructbegin`) and mc-chunks (`\tagmcbegin`).

They have a name (a string), in lua a number (for the lua attribute), and in PDF 2.0 belong to one or more name spaces, with one being the default name space.

Tags of structures are classified, e.g. as grouping, inline or block level structure (and a few special classes like lists and tables), and must follow containments rules depending on their classification (for example a inline structure can not contain a block level structure). New tags inherit their classification from their rolemapping to the standard namespaces (`pdf` and/or `pdf2`). We store this classification as it will probably be needed for tests but currently the data is not much used. The classification for math (and the containment rules) is unclear currently and so not set.

The attribute number is only relevant in lua and only for the MC chunks (so tags with the same name from different names spaces can have the same number), and so only stored if luatex is detected.

Due to the namespaces the storing and processing of tags and there data are different in various places for PDF 2.0 and PDF <2.0, which makes things a bit difficult and leads to some duplications. Perhaps at some time there should be a clear split.

This are the main variables used by the code:

`\g__tag_role_tags_NS_prop` This is the core list of tag names. It uses tags as keys and the shorthand (e.g. `pdf2`, or `mathml`) of the default name space as value.

In pdf 2.0 the value is needed in the structure dictionaries.

`\g__tag_role_tags_class_prop` This contains for each tag a classification type. It is used in pdf <2.0.

`\g__tag_role_NS_prop` This contains the names spaces. The values are the object references. They are used in pdf 2.0.

`\g__tag_role_rolemap_prop` This contains for each tag the role to a standard tag. It is used in pdf<2.0 for tag checking and to fill at the end the RoleMap dictionary.

`g_@@_role/RoleMap_dict` This dictionary contains the standard rolemaps. It is relevant only for pdf <2.0.

`\g__tag_role_NS_<ns>_prop` This prop contains the tags of a name space and their role. The props are also use for remapping. As value they contain two brace groups: tag and namespace. In pdf <2.0 the namespace is empty.

`\g__tag_role_NS_<ns>_class_prop` This prop contains the tags of a name space and their type. The value is only needed for pdf 2.0.

`\g__tag_role_index_prop` This prop contains the standard tags (pdf in pdf<2.0, pdf, pdf2 + mathml in pdf 2.0) as keys, the values are a two-digit number. These numbers are used to get the containment rule of two tags from the intarray.

`\l__tag_role_debug_prop` This property is used to pass some info around for info messages or debugging.

`\g__tag_role_tags_NS_prop` This is the core list of tag names. It uses tags as keys and the shorthand (e.g. `pdf2`, or `mathml`) of the default name space as value. We store the default name space also in pdf <2.0, even if not needed: it doesn't harm and simplifies the code. There is no need to access this from lua, so we use the standard prop commands.

`\prop_new:N \g__tag_role_tags_NS_prop`

(End definition for `\g__tag_role_tags_NS_prop`.)

`\g__tag_role_tags_class_prop` With pdf 2.0 we store the class in the NS dependant props. With pdf <2.0 we store for now the type(s) of a tag in a common prop. Tags that are rolemapped should get the type from the target.

`\prop_new:N \g__tag_role_tags_class_prop`

(End definition for `\g__tag_role_tags_class_prop`.)

`\g__tag_role_NS_prop` This holds the list of supported name spaces. The keys are the name tagpdf will use, the values the object reference. The urls identifier are stored in related dict object.

mathml <http://www.w3.org/1998/Math/MathML>

pdf2 <http://iso.org/pdf/ssn>

pdf <http://iso.org/pdf/ssn> (default)

user `\c__tag_role_userNS_id_str` (random id, for user tags)

latex <https://www.latex-project.org/ns/dft/2022>

latex-book <https://www.latex-project.org/ns/book/2022>

latex-inline <https://www.latex-project.org/ns/inline/2022>

More namespaces are possible and their objects references and their rolemaps must be collected so that an array can be written to the StructTreeRoot at the end (see tagpdf-tree). We use a prop to store the object reference as it will be needed rather often.

`\prop_new:N \g__tag_role_NS_prop`

(End definition for `\g__tag_role_NS_prop`.)

`\g__tag_role_index_prop` This prop contains the standard tags (pdf in pdf<2.0, pdf, pdf2 + mathml in pdf 2.0) as keys, the values are a two-digit number. These numbers are used to get the containment rule of two tags from the intarray.

`\prop_new:N \g__tag_role_index_prop`

(End definition for `\g__tag_role_index_prop`.)

`\l__tag_role_debug_prop` This variable is used to pass more infos to debug messages.

`\prop_new:N \l__tag_role_debug_prop`

(End definition for `\l__tag_role_debug_prop`.)

We need also a bunch of temporary variables.

`\l__tag_role_tag_tmpa_tl`

`\l__tag_role_tag_namespace_tmpa_tl`

`\tl_new:N \l__tag_role_tag_tmpa_tl`

`\l__tag_role_role_tmpa_tl`

`\tl_new:N \l__tag_role_tag_namespace_tmpa_tl`

`\l__tag_role_role_namespace_tmpa_tl`

`\tl_new:N \l__tag_role_role_tmpa_tl`

`\l__tag_role_tmpa_seq`

`\tl_new:N \l__tag_role_role_namespace_tmpa_tl`

`\seq_new:N \l__tag_role_tmpa_seq`

(End definition for `\l__tag_role_tag_tmpa_tl` and others.)

1.2 Namespaces

The following commands setups a name space. With pdf version <2.0 this is only a prop with the rolemap. With pdf 2.0 a dictionary must be set up. Such a name space dictionaries can contain an optional /Schema and /RoleMapNS entry. We only reserve the objects but delay the writing to the finish code, where we can test if the keys and the name spaces are actually needed. This commands setups objects for the name space and its rolemap. It also initialize a dict to collect the rolemaps if needed, and a property with the tags of the name space and their rolemapping for loops. It is unclear if a reference to a schema file will be ever needed, but it doesn't harm

`g__tag_role/RoleMap_dict` This is the object which contains the normal RoleMap. It is probably not needed in pdf
`\g__tag_role_rolemap_prop` 2.0 but currently kept.

```
17 \pdfdict_new:n {g__tag_role/RoleMap_dict}
18 \prop_new:N \g__tag_role_rolemap_prop
```

(End definition for `g__tag_role/RoleMap_dict` and `\g__tag_role_rolemap_prop`.)

```
\__tag_role_NS_new:nnn \__tag_role_NS_new:nnn{<shorthand>}{<URI-ID>}Schema
```

```
\__tag_role_NS_new:nnn
```

```
19 \pdf_version_compare:NnTF < {2.0}
20 {
21   \cs_new_protected:Npn \__tag_role_NS_new:nnn #1 #2 #3
22   {
23     \prop_new:c { g__tag_role_NS_#1_prop }
24     \prop_new:c { g__tag_role_NS_#1_class_prop }
25     \prop_gput:Nnx \g__tag_role_NS_prop {#1}{ }
26   }
27 }
28 {
29   \cs_new_protected:Npn \__tag_role_NS_new:nnn #1 #2 #3
30   {
31     \prop_new:c { g__tag_role_NS_#1_prop }
32     \prop_new:c { g__tag_role_NS_#1_class_prop }
33     \pdf_object_new:n {tag/NS/#1}
34     \pdfdict_new:n {g__tag_role/Namespace_#1_dict}
35     \pdf_object_new:n {__tag/RoleMapNS/#1}
36     \pdfdict_new:n {g__tag_role/RoleMapNS_#1_dict}
37     \pdfdict_gput:nnn
38       {g__tag_role/Namespace_#1_dict}
39       {Type}
40       {/Namespace}
41     \pdf_string_from_unicode:nnN{utf8/string}{#2}\l__tag_tmpa_str
42     \tl_if_empty:NF \l__tag_tmpa_str
43     {
44       \pdfdict_gput:nnx
45         {g__tag_role/Namespace_#1_dict}
46         {NS}
47         {\l__tag_tmpa_str}
48     }
49     %RoleMapNS is added in tree
50     \tl_if_empty:nF {#3}
```

```

51     {
52       \pdfdict_gput:nnx{g__tag_role/Namespace_#1_dict}
53       {Schema}{#3}
54     }
55     \prop_gput:Nnx \g__tag_role_NS_prop {#1}{\pdf_object_ref:n{tag/NS/#1}~}
56   }
57 }

```

(End definition for _tag_role_NS_new:nnn.)

We need an id for the user space. For the tests it should be possible to set it to a fix value. So we use random numbers which can be fixed by setting a seed. We fake a sort of GUID but do not try to be really exact as it doesn't matter ...

\c__tag_role_userNS_id_str

```

58 \str_const:Nx \c__tag_role_userNS_id_str
59 { data:,
60   \int_to_Hex:n{\int_rand:n {65535}}
61   \int_to_Hex:n{\int_rand:n {65535}}
62   -
63   \int_to_Hex:n{\int_rand:n {65535}}
64   -
65   \int_to_Hex:n{\int_rand:n {65535}}
66   -
67   \int_to_Hex:n{\int_rand:n {65535}}
68   -
69   \int_to_Hex:n{\int_rand:n {16777215}}
70   \int_to_Hex:n{\int_rand:n {16777215}}
71 }

```

(End definition for \c__tag_role_userNS_id_str.)

Now we setup the standard names spaces. The mathml space is currently only loaded for pdf 2.0.

```

72 \_tag_role_NS_new:nnn {pdf} {http://iso.org/pdf/ssn}{ }
73 \_tag_role_NS_new:nnn {pdf2} {http://iso.org/pdf2/ssn}{ }
74 \pdf_version_compare:NnF < {2.0}
75 {
76   \_tag_role_NS_new:nnn {mathml}{http://www.w3.org/1998/Math/MathML}{ }
77 }
78 \_tag_role_NS_new:nnn {latex} {https://www.latex-project.org/ns/dfl/2022}{ }
79 \_tag_role_NS_new:nnn {latex-book} {https://www.latex-project.org/ns/book/2022}{ }
80 \_tag_role_NS_new:nnn {latex-inline} {https://www.latex-project.org/ns/inline/2022}{ }
81 \exp_args:Nnx
82   \_tag_role_NS_new:nnn {user}{\c__tag_role_userNS_id_str}{ }

```

1.3 Adding a new tag

Both when reading the files and when setting up a tag manually we have to store data in various places.

_tag_role_alloctag:nnn

This command allocates a new tag without role mapping. In the lua backend it will also record the attribute value.

```

83 \pdf_version_compare:NnTF < {2.0}
84 {

```

```

85 \sys_if_engine luatex:TF
86 {
87   \cs_new_protected:Npn \__tag_role_alloctag:nnn #1 #2 #3 % #1 tagname, ns, type
88   {
89     \lua_now:e { ltx.__tag.func.alloctag ('#1') }
90     \prop_gput:Nnn \g__tag_role_tags_NS_prop {#1}{#2}
91     \prop_gput:cnn {g__tag_role_NS_#2_prop} {#1}{{}}{}
92     \prop_gput:Nnn \g__tag_role_tags_class_prop {#1}{#3}
93     \prop_gput:cnn {g__tag_role_NS_#2_class_prop} {#1}{--UNUSED--}
94   }
95 }
96 {
97   \cs_new_protected:Npn \__tag_role_alloctag:nnn #1 #2 #3
98   {
99     \prop_gput:Nnn \g__tag_role_tags_NS_prop {#1}{#2}
100    \prop_gput:cnn {g__tag_role_NS_#2_prop} {#1}{{}}{}
101    \prop_gput:Nnn \g__tag_role_tags_class_prop {#1}{#3}
102    \prop_gput:cnn {g__tag_role_NS_#2_class_prop} {#1}{--UNUSED--}
103  }
104 }
105 }
106 {
107   \sys_if_engine luatex:TF
108   {
109     \cs_new_protected:Npn \__tag_role_alloctag:nnn #1 #2 #3 % #1 tagname, ns, type
110     {
111       \lua_now:e { ltx.__tag.func.alloctag ('#1') }
112       \prop_gput:Nnn \g__tag_role_tags_NS_prop {#1}{#2}
113       \prop_gput:cnn {g__tag_role_NS_#2_prop} {#1}{{}}{}
114       \prop_gput:Nnn \g__tag_role_tags_class_prop {#1}{--UNUSED--}
115       \prop_gput:cnn {g__tag_role_NS_#2_class_prop} {#1}{#3}
116     }
117   }
118   {
119     \cs_new_protected:Npn \__tag_role_alloctag:nnn #1 #2 #3
120     {
121       \prop_gput:Nnn \g__tag_role_tags_NS_prop {#1}{#2}
122       \prop_gput:cnn {g__tag_role_NS_#2_prop} {#1}{{}}{}
123       \prop_gput:Nnn \g__tag_role_tags_class_prop {#1}{--UNUSED--}
124       \prop_gput:cnn {g__tag_role_NS_#2_class_prop} {#1}{#3}
125     }
126   }
127 }
128 \cs_generate_variant:Nn \__tag_role_alloctag:nnn {nnV}

```

(End definition for __tag_role_alloctag:nnn.)

1.3.1 pdf 1.7 and earlier

`__tag_role_add_tag:nn` The pdf 1.7 version has only two arguments: new and rolemap name. The role must be an existing tag and should not be empty. We allow to change the role of an existing tag: as the rolemap is written at the end not confusion can happen.

```

129 \cs_new_protected:Nn \__tag_role_add_tag:nn % (new) name, reference to old
130 {

```

checks and messages

```

131     \__tag_check_add_tag_role:nn {#1}{#2}
132     \prop_if_in:NnF \g__tag_role_tags_NS_prop {#1}
133     {
134         \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
135         {
136             \msg_info:nnn { tag }{new-tag}{#1}
137         }
138     }

```

now the addition

```

139     \prop_get:NnN \g__tag_role_tags_class_prop {#2}\l__tag_tmpa_tl
140     \quark_if_no_value:NT \l__tag_tmpa_tl
141     {
142         \tl_set:Nn\l__tag_tmpa_tl{--UNKNOWN--}
143     }
144     \__tag_role_alloctag:nnV {#1}{user}\l__tag_tmpa_tl

```

We resolve rolemapping recursively so that all targets are stored as standard tags.

```

145     \tl_if_empty:nF { #2 }
146     {
147         \prop_get:NnN \g__tag_role_rolemapping_prop {#2}\l__tag_tmpa_tl
148         \quark_if_no_value:NTF \l__tag_tmpa_tl
149         {
150             \prop_gput:Nnx \g__tag_role_rolemapping_prop {#1}{\tl_to_str:n{#2}}
151         }
152         {
153             \prop_gput:NnV \g__tag_role_rolemapping_prop {#1}\l__tag_tmpa_tl
154         }
155     }
156 }
157 \cs_generate_variant:Nn \__tag_role_add_tag:nn {VV,ne}

```

(End definition for __tag_role_add_tag:nn.)

1.3.2 The pdf 2.0 version

__tag_role_add_tag:nnnn The pdf 2.0 version takes four arguments: tag/namespace/role/namespace

```

158 \cs_new_protected:Nn \__tag_role_add_tag:nnnn %tag/namespace/role/namespace
159 {
160     \__tag_check_add_tag_role:nnn {#1/#2}{#3}{#4}
161     \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
162     {
163         \msg_info:nnn { tag }{new-tag}{#1}
164     }
165     \prop_get:cnN { g__tag_role_NS_#4_class_prop } {#3}\l__tag_tmpa_tl
166     \quark_if_no_value:NT \l__tag_tmpa_tl
167     {
168         \tl_set:Nn\l__tag_tmpa_tl{--UNKNOWN--}
169     }
170     \__tag_role_alloctag:nnV {#1}{#2}\l__tag_tmpa_tl
171     \pdfdict_gput:nnx {g__tag_role/RoleMapNS_#2_dict}{#1}
172     {
173         [
174             \pdf_name_from_unicode_e:n{#3}

```

```

175         \c_space_tl
176         \pdf_object_ref:n {tag/NS/#4}
177     ]
178 }

```

We resolve rolemapping recursively so that all targets are stored as standard tags for the tests.

```

179     \tl_if_empty:nF { #2 }
180     {
181         \prop_get:cnN { g__tag_role_NS_#4_prop } {#3}\l__tag_tmpa_tl
182         \quark_if_no_value:NTF \l__tag_tmpa_tl
183         {
184             \prop_gput:cnx { g__tag_role_NS_#2_prop } {#1}
185             {{\tl_to_str:n{#3}}{\tl_to_str:n{#4}}}
186         }
187         {
188             \prop_gput:cno { g__tag_role_NS_#2_prop } {#1}{\l__tag_tmpa_tl}
189         }
190     }
191 }
192 \cs_generate_variant:Nn \__tag_role_add_tag:nnnn {VVVV}

```

(End definition for __tag_role_add_tag:nnnn.)

1.4 Helper command to read the data from files

In this section we setup the helper command to read namespace files.

__tag_role_read_namespace_line:nw

This command will process a line in the name space file. The first argument is the name of the name space. The definition differ for pdf 2.0. as we have proper name spaces there. With pdf<2.0 not special name spaces shouldn't update the default role or add to the rolemap again. We use a boolean here.

```

193 \bool_new:N\l__tag_role_update_bool
194 \bool_set_true:N \l__tag_role_update_bool
195 \pdf_version_compare:NnTF < {2.0}
196 {
197     \cs_new_protected:Npn \__tag_role_read_namespace_line:nw #1#2,#3,#4,#5,#6\q_stop %
198     % #1 NS, #2 tag, #3 rolemap, #4 NS rolemap #5 type
199     {
200         \tl_if_empty:nF { #2 }
201         {
202             \bool_if:NTF \l__tag_role_update_bool
203             {
204                 \tl_if_empty:nTF {#5}
205                 {
206                     \prop_get:NnN \g__tag_role_tags_class_prop {#3}\l__tag_tmpa_tl
207                     \quark_if_no_value:NT \l__tag_tmpa_tl
208                     {
209                         \tl_set:Nn\l__tag_tmpa_tl{--UNKNOWN--}
210                     }
211                 }
212                 {
213                     \tl_set:Nn \l__tag_tmpa_tl {#5}
214                 }

```



```

215     \__tag_role_alloctag:nnV {#2}{#1}\l__tag_tmpa_tl
216     \tl_if_empty:nF {#3}
217     {
218         \__tag_role_add_tag:nn {#2}{#3}
219     }
220     \prop_gput:cnn {g__tag_role_NS_#1_prop} {#2}{#{#3}{}}
221 }
222 {
223     \prop_gput:cnn {g__tag_role_NS_#1_prop} {#2}{#{#3}{}}
224     \prop_gput:cnn {g__tag_role_NS_#1_class_prop} {#2}{--UNUSED--}
225 }
226 }
227 }
228 }
229 {
230 \cs_new_protected:Npn \__tag_role_read_namespace_line:nw #1#2,#3,#4,#5,#6\q_stop %
231 % #1 NS, #2 tag, #3 rolemap, #4 NS rolemap #5 type
232 {
233     \tl_if_empty:nF {#2}
234     {
235         \tl_if_empty:nTF {#5}
236         {
237             \prop_get:cnN { g__tag_role_NS_#4_class_prop } {#3}\l__tag_tmpa_tl
238             \quark_if_no_value:NT \l__tag_tmpa_tl
239             {
240                 \tl_set:Nn\l__tag_tmpa_tl{--UNKNOWN--}
241             }
242         }
243         {
244             \tl_set:Nn \l__tag_tmpa_tl {#5}
245         }
246         \__tag_role_alloctag:nnV {#2}{#1}\l__tag_tmpa_tl
247         \bool_lazy_and:nnT
248         { ! \tl_if_empty:n {#3} } { ! \str_if_eq_p:nn {#1}{pdf2} }
249         {
250             \__tag_role_add_tag:nnnn {#2}{#1}{#3}{#4}
251             \prop_gput:cnn {g__tag_role_NS_#1_prop} {#2}{#{#3}{#4}}
252         }
253     }
254 }
255 }

```

(End definition for __tag_role_read_namespace_line:nw.)

__tag_role_read_namespace:n This command reads the namespace file.

```

256 \cs_new_protected:Npn \__tag_role_read_namespace:n #1 %name of namespace
257 {
258     \prop_if_exist:cF {g__tag_role_NS_#1_prop}
259     { \msg_warning:nnn {tag}{namespace-unknown}{#1} }
260     \file_if_exist:nTF { tagpdf-ns-#1.def}
261     {
262         \ior_open:Nn \g_tmpa_ior {tagpdf-ns-#1.def}
263         \msg_info:nnn {tag}{read-namespace}{#1}
264         \ior_map_inline:Nn \g_tmpa_ior

```

```

265         {
266             \_tag_role_read_namespace_line:nw {#1} ##1,,,\q_stop
267         }
268         \ior_close:N\g_tmpa_ior
269     }
270     {
271         \msg_warning:nnn{tag}{namespace-missing}{#1}
272     }
273 }
274

```

(End definition for _tag_role_read_namespace:n.)

1.5 Reading the default data

The order is important as we want pdf2 and latex as default.

```

275 \_tag_role_read_namespace:n {pdf}
276 \_tag_role_read_namespace:n {pdf2}
277 \pdf_version_compare:NnF < {2.0}
278   {\_tag_role_read_namespace:n {mathml}}
279 \bool_set_false:N\l__tag_role_update_bool
280 \_tag_role_read_namespace:n {latex-inline}
281 \_tag_role_read_namespace:n {latex-book}
282 \bool_set_true:N\l__tag_role_update_bool
283 \_tag_role_read_namespace:n {latex}
284 \_tag_role_read_namespace:n {pdf}
285 \_tag_role_read_namespace:n {pdf2}

    But is the class provides a \chapter command then we switch
286 \pdf_version_compare:NnTF < {2.0}
287   {
288       \hook_gput_code:nnn {begindocument}{tagpdf}
289       {
290           \cs_if_exist:NT \chapter
291           {
292               \prop_map_inline:cn{g__tag_role_NS_latex-book_prop}
293               {
294                   \_tag_role_add_tag:ne {#1}{\use_i:nn #2\c_empty_tl\c_empty_tl}
295               }
296           }
297       }
298   }
299   {
300       \hook_gput_code:nnn {begindocument}{tagpdf}
301       {
302           \cs_if_exist:NT \chapter
303           {
304               \prop_map_inline:cn{g__tag_role_NS_latex-book_prop}
305               {
306                   \prop_gput:Nnn \g__tag_role_tags_NS_prop { #1 }{ latex-book }
307               }
308           }
309       }
310   }

```

1.6 Parent-child rules

PDF define various rules about which tag can be a child of another tag. The following code implements the matrix to allow to use it in tests.

`\g__tag_role_parent_child_intarray` This intarray will store the rule as a number. For parent nm and child ij (n,m,i,j digits) the rule is at position nmij. As we have around 56 tags, we need roughly a size 6000.

```
311 \intarray_new:Nn \g__tag_role_parent_child_intarray {6000}
```

(End definition for `\g__tag_role_parent_child_intarray`.)

`\c__tag_role_rules_prop` These two properties map the rule strings to numbers and back. There are in tagpdf-data.dtx near the csv files for easier maintenance.

(End definition for `\c__tag_role_rules_prop` and `\c__tag_role_rules_num_prop`.)

`_tag_store_parent_child_rule:nnm` The helper command is used to store the rule. It assumes that parent and child are given as 2-digit number!

```
312 \cs_new_protected:Npn \_tag_store_parent_child_rule:nnn #1 #2 #3 % num parent, num child, #3
313 {
314   \intarray_gset:Nnn \g__tag_role_parent_child_intarray
315     { #1#2 }{0\prop_item:Nn\c__tag_role_rules_prop{#3}}
316 }
```

(End definition for `_tag_store_parent_child_rule:nnn`.)

1.6.1 Reading in the csv-files

This counter will be used to identify the first (non-comment) line

```
317 \int_zero:N \l__tag_tmpa_int
```

Open the file depending on the PDF version

```
318 \pdf_version_compare:NnTF < {2.0}
319 {
320   \ior_open:Nn \g_tmpa_ior {tagpdf-parent-child.csv}
321 }
322 {
323   \ior_open:Nn \g_tmpa_ior {tagpdf-parent-child-2.csv}
324 }
```

Now the main loop over the file

```
325 \ior_map_inline:Nn \g_tmpa_ior
326 {
```

ignore lines containing only comments

```
327   \tl_if_empty:nF{#1}
328   {
```

count the lines ...

```
329     \int_incr:N\l__tag_tmpa_int
```

put the line into a seq. Attention! empty cells are dropped.

```
330     \seq_set_from_clist:Nn\l__tag_tmpa_seq { #1 }
331     \int_compare:nNnTF {\l__tag_tmpa_int}=1
```

This handles the header line. It gives the tags 2-digit numbers

```

332     {
333         \seq_map_indexed_inline:Nn \l__tag_tmpa_seq
334         {
335             \prop_gput:Nnx\g__tag_role_index_prop
336             {##2}
337             {\int_compare:nNnT{##1}<{10}{0}##1}
338         }
339     }

```

now the data lines.

```

340     {
341         \seq_set_from_clist:Nn\l__tag_tmpa_seq { #1 }
get the name of the child tag from the first column
342         \seq_pop_left:NN\l__tag_tmpa_seq\l__tag_tmpa_tl
get the number of the child, and store it in \l__tag_tmpb_tl
343         \prop_get:NVN \g__tag_role_index_prop \l__tag_tmpa_tl \l__tag_tmpb_tl
remove column 2+3
344         \seq_pop_left:NN\l__tag_tmpa_seq\l__tag_tmpa_tl
345         \seq_pop_left:NN\l__tag_tmpa_seq\l__tag_tmpa_tl

```

Now map over the rest. The index ##1 gives us the number of the parent, ##2 is the data.

```

346         \seq_map_indexed_inline:Nn \l__tag_tmpa_seq
347         {
348             \exp_args:Nnx
349             \__tag_store_parent_child_rule:nnn {##1}{\l__tag_tmpb_tl}{ ##2 }
350         }
351     }
352 }
353 }

```

close the read handle.

```

354 \ior_close:N\g_tmpa_ior

```

The Root, Hn and mathml tags are special and need to be added explicitly

```

355 \prop_get:NnN\g__tag_role_index_prop{StructTreeRoot}\l__tag_tmpa_tl
356 \prop_gput:Nnx\g__tag_role_index_prop{Root}{\l__tag_tmpa_tl}
357 \prop_get:NnN\g__tag_role_index_prop{Hn}\l__tag_tmpa_tl
358 \pdf_version_compare:NnTF < {2.0}
359 {
360     \int_step_inline:nn{6}
361     {
362         \prop_gput:Nnx\g__tag_role_index_prop{H#1}{\l__tag_tmpa_tl}
363     }
364 }
365 {
366     \int_step_inline:nn{10}
367     {
368         \prop_gput:Nnx\g__tag_role_index_prop{H#1}{\l__tag_tmpa_tl}
369     }

```

all mathml tags are currently handled identically

```

370 \prop_get:NnN\g__tag_role_index_prop {mathml}\l__tag_tmpa_tl
371 \prop_get:NnN\g__tag_role_index_prop {math}\l__tag_tmpb_tl
372 \prop_map_inline:Nn \g__tag_role_NS_mathml_prop
373 {
374   \prop_gput:NnV\g__tag_role_index_prop{#1}\l__tag_tmpa_tl
375 }
376 \prop_gput:NnV\g__tag_role_index_prop{math}\l__tag_tmpb_tl
377 }

```

1.6.2 Retrieving the parent-child rule

`\l__tag_role_real_parent_tl` Part, Div and NonStruct have no own rules, instead the parent(s) have to be inspected. To store this real parent we use this tlvar

```

378 \tl_new:N \l__tag_role_real_parent_tl

```

(End definition for `\l__tag_role_real_parent_tl`.)

`_tag_role_get_parent_child_rule:nnN` This command retrieves the rule (as a number) and stores it in the tl-var. TODO check temporary variables. Check if the tl-var should be fix. The arguments should be standard tags for which a rule exist and role mapping should have already be done.

```

379 \tl_new:N \l__tag_parent_child_check_tl
380 \cs_new_protected:Npn \_tag_role_get_parent_child_rule:nnN #1 #2 #3
381 % #1 parent (string) #2 child (string) #3 tl for state
382 {
383   \tl_set:Nn \l__tag_role_real_parent_tl {#1}
384   \prop_get:NnN \g__tag_role_index_prop{#1}\l__tag_tmpa_tl
385   \prop_get:NnN \g__tag_role_index_prop{#2}\l__tag_tmpb_tl
386   \bool_lazy_and:nnTF
387     { ! \quark_if_no_value_p:N \l__tag_tmpa_tl }
388     { ! \quark_if_no_value_p:N \l__tag_tmpb_tl }
389   {

```

Get the rule from the intarray

```

390   \tl_set:Nx#3
391   {
392     \intarray_item:Nn
393     \g__tag_role_parent_child_intarray
394     {\l__tag_tmpa_tl\l__tag_tmpb_tl}
395   }

```

If the state is we have to check the parents from the stack and use the first which is not Part, Div or NonStruct

```

396   \int_compare:nNnT
397     {#3} = {\prop_item:Nn\c__tag_role_rules_prop{}}
398   {
399     \seq_set_eq:NN \l__tag_role_tmpa_seq \g__tag_struct_tag_stack_seq

```

we must take the current child from the stack if is already there, depending on location the check is called, this could also remove the parent, but that is ok too.

```

400     \seq_pop_left:NN \l__tag_role_tmpa_seq\l__tag_get_tmpc_tl
401     \seq_map_inline:Nn\l__tag_role_tmpa_seq
402     {
403       \tl_if_in:nnF {-Part-Div-NonStruct-}{-#1-}

```

```

404         {
405             \tl_set:Nn\l__tag_role_real_parent_tl {##1}
406             \int_zero:N\l__tag_tpa_int
407             \__tag_role_get_parent_child_rule:nnN {##1}{#2}#3
408             \int_set:Nn\l__tag_tpa_int{1}
409             \seq_map_break:
410         }
411     }
412 }

```

This is the message, this can perhaps go into debug mode.

```

413 \group_begin:
414 \int_compare:nNnT {\l__tag_tpa_int*\l__tag_loglevel_int} > { 0 }
415 {
416     \prop_get:NVNF\c__tag_role_rules_num_prop #3 \l__tag_tpa_tl
417     {
418         \tl_set:Nn \l__tag_tpa_tl {unknown}
419     }
420     \tl_set:Nn \l__tag_tpb_tl {#1}
421     \msg_note:nnxxx
422     { tag }
423     { role-parent-child }
424     { #1
425         \tl_if_eq:NNTF\l__tag_tpb_tl\l__tag_role_real_parent_tl
426         {
427             \bool_lazy_and:nnT
428             {
429                 \prop_if_in_p:Nn \l__tag_role_debug_prop {parent}
430             }
431             {
432                 !\str_if_eq_p:ee {#1}{\prop_item:Nn\l__tag_role_debug_prop {parent}}
433             }
434             {
435                 \c_space_tl (from~\prop_item:Nn\l__tag_role_debug_prop {parent})
436             }
437         }
438         {
439             \c_space_tl(inherited~from~\l__tag_role_real_parent_tl)
440         }
441     }
442     {
443         #2
444         \bool_lazy_and:nnT
445         {
446             \prop_if_in_p:Nn \l__tag_role_debug_prop {child}
447         }
448         {
449             !\str_if_eq_p:ee {#2}{\prop_item:Nn\l__tag_role_debug_prop {child}}
450         }
451         {
452             \c_space_tl (from~\prop_item:Nn\l__tag_role_debug_prop {child})
453         }
454     }
455     { '#3~(\l__tag_tpa_tl)' }
456 }

```

```

457         \group_end:
458     }
459     {
460         \tl_set:Nn#3 {0}
461         \msg_warning:nnxxx
462             { tag }
463             {role-parent-child}
464             { #1 }
465             { #2 }
466             { unknown! }
467     }
468 }
469 \cs_generate_variant:Nn\__tag_role_get_parent_child_rule:nnN {VVN}

(End definition for \__tag_role_get_parent_child_rule:nnN.)

```

__tag_check_parent_child:nnnnN

This is the main command. It has to retrieve the standard tags for a comparison. In pdf 2.0 the name spaces of the tags are relevant, so we have arguments for them, but in pdf <2.0 they are ignored and can be left empty.

```

470 \pdf_version_compare:NnTF < {2.0}
471 {
472     \cs_new_protected:Npn \__tag_check_parent_child:nnnnN #1 #2 #3 #4 #5
473     {

```

for debugging messages we store the arguments.

```

474         \prop_put:Nnn \l__tag_role_debug_prop {parent} {#1}
475         \prop_put:Nnn \l__tag_role_debug_prop {child} {#3}

```

get the standard tags through rolemapping if needed at first the parent

```

476         \prop_get:NnNTF \g__tag_role_index_prop {#1}\l__tag_tmpa_tl
477         {
478             \tl_set:Nn \l__tag_tmpa_tl {#1}
479         }
480         {
481             \prop_get:NnNF \g__tag_role_rolemap_prop {#1}\l__tag_tmpa_tl
482             {
483                 \tl_set:Nn \l__tag_tmpa_tl {\q_no_value}
484             }
485         }

```

now the child

```

486         \prop_get:NnNTF \g__tag_role_index_prop {#3}\l__tag_tmppb_tl
487         {
488             \tl_set:Nn \l__tag_tmppb_tl {#3}
489         }
490         {
491             \prop_get:NnNF \g__tag_role_rolemap_prop {#3}\l__tag_tmppb_tl
492             {
493                 \tl_set:Nn \l__tag_tmppb_tl {\q_no_value}
494             }
495         }

```

if we got tags for parent and child we call the checking command

```

496         \bool_lazy_and:nnTF
497             { ! \quark_if_no_value_p:N \l__tag_tmpa_tl }
498             { ! \quark_if_no_value_p:N \l__tag_tmppb_tl }

```

```

499     {
500         \__tag_role_get_parent_child_rule:VvN \l__tag_tmpa_tl \l__tag_tmppb_tl #5
501     }
502     {
503         \tl_set:Nn #5 {0}
504         \msg_warning:nnxxx
505         { tag }
506         {role-parent-child}
507         { #1 }
508         { #3 }
509         { unknown! }
510     }
511 }
512 \cs_new_protected:Npn \__tag_check_parent_child:nnN #1#2#3
513 {
514     \__tag_check_parent_child:nnnnN {#1}{#2}{#3}
515 }
516 }

```

and now the pdf 2.0 version The version with three arguments retrieves the default namespace and then calls the full command. Not sure if this will ever be needed but we leave it for now.

```

517 {
518     \cs_new_protected:Npn \__tag_check_parent_child:nnN #1 #2 #3
519     {
520         \prop_get:NnN\g__tag_role_tags_NS_prop {#1}\l__tag_role_tag_namespace_tmpa_tl
521         \prop_get:NnN\g__tag_role_tags_NS_prop {#2}\l__tag_role_tag_namespace_tmppb_tl
522         \str_if_eq:nnT{#2}{MC}{\tl_clear:N \l__tag_role_tag_namespace_tmppb_tl}
523         \bool_lazy_and:nnTF
524         { ! \quark_if_no_value_p:N \l__tag_role_tag_namespace_tmpa_tl }
525         { ! \quark_if_no_value_p:N \l__tag_role_tag_namespace_tmppb_tl }
526         {
527             \__tag_check_parent_child:nVnVN
528             {#1}\l__tag_role_tag_namespace_tmpa_tl
529             {#2}\l__tag_role_tag_namespace_tmppb_tl
530             #3
531         }
532         {
533             \tl_set:Nn #3 {0}
534             \msg_warning:nnxxx
535             { tag }
536             {role-parent-child}
537             { #1 }
538             { #2 }
539             { unknown! }
540         }
541     }
542 }

```

and now the real command.

```

542     \cs_new_protected:Npn \__tag_check_parent_child:nnnnN #1 #2 #3 #4 #5 %tag,NS,tag,NS, tl var
543     {
544         \prop_put:Nnn \l__tag_role_debug_prop {parent} {#1/#2}
545         \prop_put:Nnn \l__tag_role_debug_prop {child} {#3/#4}

```


If the namespace is empty, we assume a standard tag, otherwise we retrieve the rolemapping from the namespace

```

546     \tl_if_empty:nTF {#2}
547     {
548         \tl_set:Nn \l__tag_tmpa_tl {#1}
549     }
550     {
551         \prop_get:cnNTF
552         { g__tag_role_NS_#2_prop }
553         {#1}
554         \l__tag_tmpa_tl
555         {
556             \tl_set:Nx \l__tag_tmpa_tl {\tl_head:N\l__tag_tmpa_tl}
557             \tl_if_empty:NT\l__tag_tmpa_tl
558             {
559                 \tl_set:Nn \l__tag_tmpa_tl {#1}
560             }
561         }
562         {
563             \tl_set:Nn \l__tag_tmpa_tl {\q_no_value}
564         }
565     }

```

and the same for the child If the namespace is empty, we assume a standard tag, otherwise we retrieve the rolemapping from the namespace

```

566     \tl_if_empty:nTF {#4}
567     {
568         \tl_set:Nn \l__tag_tmpb_tl {#3}
569     }
570     {
571         \prop_get:cnNTF
572         { g__tag_role_NS_#4_prop }
573         {#3}
574         \l__tag_tmpb_tl
575         {
576             \tl_set:Nx \l__tag_tmpb_tl { \tl_head:N\l__tag_tmpb_tl }
577             \tl_if_empty:NT\l__tag_tmpb_tl
578             {
579                 \tl_set:Nn \l__tag_tmpb_tl {#3}
580             }
581         }
582         {
583             \tl_set:Nn \l__tag_tmpb_tl {\q_no_value}
584         }
585     }

```

and now get the relation

```

586     \bool_lazy_and:nnTF
587     { ! \quark_if_no_value_p:N \l__tag_tmpa_tl }
588     { ! \quark_if_no_value_p:N \l__tag_tmpb_tl }
589     {
590         \__tag_role_get_parent_child_rule:VnN \l__tag_tmpa_tl \l__tag_tmpb_tl #5
591     }
592     {

```

```

593         \tl_set:Nn #5 {0}
594         \msg_warning:nnxxx
595         { tag }
596         {role-parent-child}
597         { #1 }
598         { #3 }
599         { unknown! }
600     }
601 }
602 }
603 \cs_generate_variant:Nn\__tag_check_parent_child:nnN {VVM}
604 \cs_generate_variant:Nn\__tag_check_parent_child:nnnnN {VVVVN,nVnVN,VVnnN}

(End definition for __tag_check_parent_child:nnnnN.)

```

1.7 Remapping of tags

In some context it can be necessary to remap or replace the tags. That means instead of tag=H1 or tag=section one wants the effect of tag=Span. Or instead of tag=P one wants tag=Code.

The following command provide some general interface for this. The core idea is that before a tag is set it is fed through a function that can change it. We want to be able to chain such functions, so all of them manipulate the same variables.

```

\l__tag_role_remap_tag_tl
\l__tag_role_remap_NS_tl
605 \tl_new:N \l__tag_role_remap_tag_tl
606 \tl_new:N \l__tag_role_remap_NS_tl

(End definition for \l__tag_role_remap_tag_tl and \l__tag_role_remap_NS_tl.)

\__tag_role_remap: This function is used in the structure and the mc code before using a tag. By default it
does nothing with the tl vars. Perhaps this should be a hook?
607 \cs_new_protected:Npn \__tag_role_remap: { }

(End definition for \__tag_role_remap:.)

\__tag_role_remap_id: This is copy in case we have to restore the main command.
608 \cs_set_eq:NN \__tag_role_remap_id: \__tag_role_remap:

(End definition for \__tag_role_remap_id:.)

\__tag_role_remap_inline: The mapping is meant to “degrade” tags, e.g. if used inside some complex object. The
pdf<2.0 code maps the tag to the new role, the pdf 2.0 code only switch the NS.
609 \pdf_version_compare:NnTF < {2.0}
610 {
611     \cs_new_protected:Npn \__tag_role_remap_inline:
612     {
613         \prop_get:cVNT { g__tag_role_NS_latex-inline_prop } \l__tag_role_remap_tag_tl \l__tag_role_remap_NS_tl
614         {
615             \tl_set:Nx \l__tag_role_remap_tag_tl
616             {
617                 \exp_last_unbraced:NV \use_i:nn \l__tag_tmpa_tl
618             }
619             \tl_set:Nx \l__tag_role_remap_NS_tl

```

```

620         {
621             \exp_last_unbraced:N\use_ii:nn \l__tag_tmpa_tl
622         }
623     }
624     \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
625     {
626         \msg_note:nnx { tag } { role-remapping }{ \l__tag_role_remap_tag_tl }
627     }
628 }
629 }
630 {
631     \cs_new_protected:Npn \__tag_role_remap_inline:
632     {
633         \prop_get:cVNT { g__tag_role_NS_latex-inline_prop }{\l__tag_role_remap_tag_tl\l__tag_tr
634         {
635             \tl_set:Nn\l__tag_role_remap_NS_tl {latex-inline}
636         }
637         \int_compare:nNnT {\l__tag_loglevel_int} > { 0 }
638         {
639             \msg_note:nnx { tag } { role-remapping }{ \l__tag_role_remap_tag_tl/latex-
inline }
640         }
641     }
642 }

```

(End definition for __tag_role_remap_inline:.)

1.8 Key-val user interface

The user interface uses the key `add-new-tag`, which takes either a keyval list as argument, or a tag/role.

```

tag_(rolemap-key)
tag-namespace_(rolemap-key) 643 \keys_define:nn { __tag / tag-role }
role_(rolemap-key)          644 {
role-namespace_(rolemap-key) 645     ,tag .tl_set:N = \l__tag_role_tag_tmpa_tl
add-new-tag_(setup-key)      646     ,tag-namespace .tl_set:N = \l__tag_role_tag_namespace_tmpa_tl
                             647     ,role .tl_set:N = \l__tag_role_role_tmpa_tl
                             648     ,role-namespace .tl_set:N = \l__tag_role_role_namespace_tmpa_tl
                             649 }
650
651 \keys_define:nn { __tag / setup }
652 {
653     add-new-tag .code:n =
654     {
655         \keys_set_known:nnnN
656         {__tag/tag-role}
657         {
658             tag-namespace=user,
659             role-namespace=, %so that we can test for it.
660             #1
661             }{__tag/tag-role}\l_tmpa_tl
662         \tl_if_empty:NF \l_tmpa_tl
663         {

```

```

664     \exp_args:NNno \seq_set_split:Nnn \l_tmpa_seq { / } {\l_tmpa_tl/}
665     \tl_set:Nx \l__tag_role_tag_tmpa_tl { \seq_item:Nn \l_tmpa_seq {1} }
666     \tl_set:Nx \l__tag_role_role_tmpa_tl { \seq_item:Nn \l_tmpa_seq {2} }
667   }
668   \tl_if_empty:NT \l__tag_role_role_namespace_tmpa_tl
669   {
670     \prop_get:NVNTF
671       \g__tag_role_tags_NS_prop
672       \l__tag_role_role_tmpa_tl
673       \l__tag_role_role_namespace_tmpa_tl
674     {
675       \prop_if_in:NVF\g__tag_role_NS_prop \l__tag_role_role_namespace_tmpa_tl
676       {
677         \tl_set:Nn \l__tag_role_role_namespace_tmpa_tl {user}
678       }
679     }
680     {
681       \tl_set:Nn \l__tag_role_role_namespace_tmpa_tl {user}
682     }
683   }
684   \pdf_version_compare:NnTF < {2.0}
685   {
686     %TODO add check for emptyness?
687     \__tag_role_add_tag:VV
688       \l__tag_role_tag_tmpa_tl
689       \l__tag_role_role_tmpa_tl
690   }
691   {
692     \__tag_role_add_tag:VVVV
693       \l__tag_role_tag_tmpa_tl
694       \l__tag_role_tag_namespace_tmpa_tl
695       \l__tag_role_role_tmpa_tl
696       \l__tag_role_role_namespace_tmpa_tl
697   }
698 }
699 }
700 \end{package}

```

(End definition for tag (rolemap-key) and others. These functions are documented on page 129.)

Part X

The tagpdf-space module

Code related to real space chars

Part of the tagpdf package

`interwordspace_␣(setup-key)` This key allows to activate/deactivate the real space chars if the engine supports it. The allowed values are `true`, `on`, `false`, `off`.

`show-spaces_␣(setup-key)` This key works only with luatex and shows with small red bars where spaces have been inserted. This is only for debugging and is not completely reliable (and change affect other literals and tagging), so it should be used with care.

```
1 <@@=tag>
2 <*header>
3 \ProvidesExplPackage {tagpdf-space-code} {2022-12-22} {0.98}
4 {part of tagpdf - code related to real space chars}
5 </header>
```

1 Code for interword spaces

The code is engine/backend dependant. Basically only pdftex and luatex support real space chars. Most of the code for luatex which uses attributes is in the lua code, here are only the keys.

```
interwordspace_␣(setup-key)
show-spaces_␣(setup-key)
6 (*package)
7 \keys_define:nn { __tag / setup }
8 {
9     interwordspace .choices:nn = { true, on }
10     { \msg_warning:nnx {tag}{sys-no-interwordspace}{\c_sys_engine_str} },
11     interwordspace .choices:nn = { false, off }
12     { \msg_warning:nnx {tag}{sys-no-interwordspace}{\c_sys_engine_str} },
13     interwordspace .default:n = true,
14     show-spaces .bool_set:N = \l__tag_showspaces_bool
15 }
16 \sys_if_engine_pdftex:T
17 {
18     \sys_if_output_pdf:TF
19     {
20         \pdfglyptounicode{space}{0020}
21         \keys_define:nn { __tag / setup }
22         {
23             interwordspace .choices:nn = { true, on } { \pdfinterwordspaceon },
24             interwordspace .choices:nn = { false, off } { \pdfinterwordspaceon },
25             interwordspace .default:n = true,
```

```

26     show-spaces .bool_set:N = \l__tag_showspaces_bool
27   }
28 }
29 {
30   \keys_define:nn { __tag / setup }
31   {
32     interwordspace .choices:nn = { true, on, false, off }
33     { \msg_warning:nnn {tag}{sys-no-interwordspace}{dvi} },
34     interwordspace .default:n = true,
35     show-spaces .bool_set:N = \l__tag_showspaces_bool
36   }
37 }
38 }
39
40
41 \sys_if_engine luatex:T
42 {
43   \keys_define:nn { __tag / setup }
44   {
45     interwordspace .choices:nn =
46       { true, on }
47     {
48       \bool_gset_true:N \g__tag_active_space_bool
49       \lua_now:e{\ltx.__tag.func.markspaceon()}
50     },
51     interwordspace .choices:nn =
52       { false, off }
53     {
54       \bool_gset_false:N \g__tag_active_space_bool
55       \lua_now:e{\ltx.__tag.func.markspaceoff()}
56     },
57     interwordspace .default:n = true,
58     show-spaces .choice:,
59     show-spaces / true .code:n =
60       {\lua_now:e{\ltx.__tag.trace.showspaces=true}},
61     show-spaces / false .code:n =
62       {\lua_now:e{\ltx.__tag.trace.showspaces=nil}},
63     show-spaces .default:n = true
64   }
65 }

```

(End definition for `interwordspace` (setup-key) and `show-spaces` (setup-key). These functions are documented on page 149.)

`__tag_fakespace:` For luatex we need a command for the fake space as equivalent of the pdftex primitive.

```

66 \sys_if_engine luatex:T
67 {
68   \cs_new_protected:Nn \__tag_fakespace:
69   {
70     \group_begin:
71     \lua_now:e{\ltx.__tag.func.fakespace()}
72     \skip_horizontal:n{\c_zero_skip}
73     \group_end:
74   }

```

```
75 }  
76 \end{package}  
  
(End definition for \_tag_fakespace:.)
```

Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

Symbols

`\` 10, 23, 37, 38, 39, 47
`_` 240, 254

A

`activate_`(setup-key) 32, 182
`activate-all_`(setup-key) 6, 228
`activate-mc_`(setup-key) 6, 228
`activate-space_`(setup-key) 6, 228
`activate-struct_`(setup-key) 6, 228
`activate-tree_`(setup-key) 6, 228
`actualtext_`(mc-key) 56, 221, 413
`actualtext_`(struct-key) 84, 371
`add-new-tag_`(setup-key) 129, 643
`\AddToHook` 13,
16, 50, 79, 197, 232, 246, 260, 271, 309
`AF_`(struct-key) 85, 479
`AFinline_`(struct-key) 85, 479
`AFinline-o_`(struct-key) 85, 479
`alt_`(mc-key) 56, 221, 413
`alt_`(struct-key) 84, 371
`artifact_`(mc-key) 56, 221, 413
artifact-bool internal commands:
 `__artifact-bool` 113
artifact-type internal commands:
 `__artifact-type` 113
`attr-unknown` 18, 42
`attribute_`(struct-key) 85, 870
`attribute-class_`(struct-key) ... 85, 836

B

bool commands:
 `\bool_gset_eq:NN` ... 329, 342, 354, 370
 `\bool_gset_false:N`
 39, 54, 213, 330, 355, 401
 `\bool_gset_true:N` 38, 48, 119, 160, 336
 `\bool_if:NTF` 9,
 9, 18, 28, 32, 33, 37, 73, 155, 175,
 183, 202, 212, 224, 234, 238, 238,
 248, 252, 255, 269, 270, 273, 279,
 321, 324, 337, 340, 349, 357, 365, 638
 `\bool_if:nTF` 6, 299
 `\bool_lazy_all:nTF` 72, 199
 `\bool_lazy_and:nnTF` 89, 99,
 247, 378, 386, 427, 444, 496, 523, 586
 `\bool_lazy_and_p:nn` 8
 `\bool_new:N` 11,
 15, 16, 37, 61, 114, 115, 116, 117,
 118, 120, 122, 124, 193, 217, 218, 320

`\bool_set_false:N` 161, 185, 186, 191,
192, 204, 205, 214, 279, 301, 323, 348
`\bool_set_true:N` 121,
123, 194, 196, 197, 217, 218, 282, 300

box commands:

`\box_dp:N` 177, 181
`\box_ht:N` 167
`\box_new:N` 109, 110
`\box_set_dp:Nn` 175, 177
`\box_set_eq:NN` 190
`\box_set_ht:Nn` 174, 176
`\box_use_drop:N` 179, 183
`\boxmaxdepth` 67, 178

C

`\c` 203, 204
c@g internal commands:
 `\c@g__tag_MCID_abs_int`
 9, 25, 34, 47, 54, 65, 71, 73,
 135, 148, 163, 237, 242, 271, 311, 376
 `\c@g__tag_parenttree_obj_int` ... 74
 `\c@g__tag_struct_abs_int`
 6, 18, 68, 137, 140,
 142, 383, 408, 420, 432, 445, 452,
 464, 474, 505, 507, 512, 523, 527,
 528, 530, 531, 533, 538, 547, 551,
 552, 554, 555, 557, 562, 592, 602,
 603, 604, 605, 608, 610, 616, 619,
 634, 636, 665, 670, 761, 863, 866, 912
`\cc` 132
`\chapter` 138, 290, 302
clist commands:
 `\clist_const:Nn` 111, 112
 `\clist_map_inline:nn` 135, 459
 `\clist_new:N` 107
 `\clist_set:Nn` 840, 874
color commands:
 `\color_select:n` 240, 254
cs commands:
 `\cs_generate_variant:Nn` 41,
 98, 104, 126, 127, 128, 128, 129,
 130, 130, 131, 131, 132, 133, 134,
 135, 145, 153, 157, 157, 162, 169,
 176, 177, 178, 179, 180, 180, 181,
 192, 197, 209, 234, 288, 299, 299,
 469, 498, 603, 604, 777, 786, 799, 809
 `\cs_gset_eq:NN` 224, 662, 663, 758, 759
 `\cs_if_exist:NTF` 81, 275, 290, 302, 311

mc-pushed 18, 14
mc-tag-missing 18, 8
mc-used-twice 18, 12
\MessageBreak 15, 19, 20, 21
msg commands:
 \msg_error:nn 112, 133, 328, 631
 \msg_error:nnn
 149, 160, 168, 179, 214, 315, 850, 888
 \msg_error:nnnn 264
 \msg_info:nn
 126, 136, 163, 202, 206, 263
 \msg_info:nnnn 156, 175
 \msg_line_context: . 316, 317, 349, 353
 \g_msg_module_name_prop 30, 34
 \g_msg_module_type_prop 33
 \msg_new:nn 7, 8, 9,
 12, 13, 14, 15, 16, 22, 24, 25, 28, 29,
 31, 33, 35, 42, 43, 44, 45, 46, 48, 50,
 51, 52, 53, 54, 55, 57, 316, 317, 347, 351
 \msg_new:nnnn 60
 \msg_note:nn 128
 \msg_note:nnn
 337, 344, 375, 383, 626, 639
 \msg_note:nnnn 323, 330, 360, 368
 \msg_note:nnnnn 421
 \msg_warning:nn 24, 130
 \msg_warning:nnn
 10, 12, 33, 39, 48, 119,
 142, 187, 195, 218, 241, 259, 271, 767
 \msg_warning:nnnn 387
 \msg_warning:nnnnn
 195, 369, 461, 504, 534, 594, 653

N

namespace_(rolemap-key) 129
new-tag 18, 50
newattribute_(setup-key) 86, 823
\newcommand 297, 298
\newcounter 6, 8, 74
\NewDocumentCommand 6,
14, 20, 25, 31, 37, 42, 47, 53, 212, 302
\newlabeldata 69
\newmarks 14
no-struct-dest_(setup-key) 6, 228
\nointerlineskip 182

P

\PackageError 13
\PackageWarning 28
para-hook-count-wrong 19, 60
paratagging_(setup-key) 34, 225
paratagging-show_(setup-key) ... 34, 225
parent_(struct-key) 84, 371

pdf commands:

\pdf_activate_structure_destination:
..... 203, 207
\pdf_bdc:nn 233
\pdf_bmc:n 231
\l_pdf_current_structure_
 destination_tl 206
\pdf_emc: 232
\pdf_name_from_unicode_e:n .. 82,
90, 95, 110, 120, 174, 180, 826, 844, 878
\pdf_object_if_exist:n 125
\pdf_object_if_exist:nTF
..... 122, 140, 302, 503, 523, 547, 573
\pdf_object_new:n
..... 29, 33, 35, 73, 170, 214, 225, 607
\pdf_object_ref:n 38, 48, 52,
55, 62, 101, 124, 126, 142, 176, 176,
222, 239, 350, 488, 495, 673, 741, 775
\pdf_object_ref_last: 193, 897
\pdf_object_unnamed_write:nn 185, 892
\pdf_object_write:nnn
..... 163, 182, 215, 234, 241, 246, 307
\pdf_pageobject_ref:n 152, 341
\pdf_string_from_unicode:nnN ... 41
\pdf_uncompress: 262
\pdf_version_compare:NnTF
19, 41, 74, 83, 102, 168, 172, 195,
228, 277, 286, 318, 358, 470, 609, 684

pdfannot commands:

\pdfannot_dict_put:nnn
..... 127, 407, 430, 448, 453
\pdfannot_link_ref_last: ... 417, 440

pdfdict commands:

\pdfdict_gput:nnn
..... 37, 44, 52, 171, 178, 238
\pdfdict_if_empty:nTF 232
\pdfdict_new:n 17, 34, 36
\pdfdict_use:n 184, 236, 243
\pdffakespace 34, 210

pdffile commands:

\pdffile_embed_stream:nnn
..... 128, 525, 549
\pdfglyphtounicode 20
\pdfinterwordspaceon 23, 24

pdfmanagement commands:

\pdfmanagement_add:nnn
..... 34, 35, 254, 256, 258, 273
\pdfmanagement_if_active_p: .. 9, 10
\pdfmanagement_remove:nn 260

prg commands:

\prg_do_nothing:
..... 79, 224, 392, 393, 394, 395
\prg_generate_conditional_
 variant:Nnn 125

\prg_new_conditional:Nnn	59, 222
\prg_new_conditional:Npnn	66, 87, 97, 291, 297, 308
\prg_new_eq_conditional:NNn	73, 229
\prg_return_false:	67, 69, 84, 94, 104, 226, 294, 306, 312
\prg_return_true:	70, 81, 91, 101, 225, 295, 305, 311
\prg_set_conditional:Npnn	70
\ProcessOptions	41
prop commands:	
\prop_clear:N	95
\prop_count:N	116
\prop_get:NnN	139, 147, 165, 181, 206, 237, 343, 355, 357, 370, 371, 384, 385, 520, 521, 790
\prop_get:NnNTF	93, 118, 121, 135, 137, 152, 171, 247, 416, 476, 481, 486, 491, 551, 571, 613, 633, 670
\prop_gpop:NnN	214
\prop_gput:Nnn	25, 25, 30, 33, 34, 55, 90, 91, 92, 93, 95, 97, 98, 99, 100, 100, 101, 102, 111, 112, 113, 114, 115, 121, 122, 123, 124, 128, 129, 147, 150, 153, 169, 184, 188, 206, 220, 223, 224, 251, 295, 306, 335, 356, 362, 368, 374, 376, 825, 897
\prop_gremove:Nn	209
\prop_if_exist:NTF	258, 728
\prop_if_exist_p:N	380
\prop_if_in:NnTF	66, 109, 117, 132, 216, 282, 675, 848, 886, 890
\prop_if_in_p:Nn	429, 446
\prop_item:Nn	34, 70, 105, 131, 166, 172, 200, 289, 292, 315, 361, 395, 397, 432, 435, 449, 452, 895, 902
\prop_map_inline:Nn	176, 230, 292, 304, 372
\prop_map_tokens:Nn	248
\prop_new:N	7, 8, 9, 10, 11, 11, 15, 18, 23, 24, 31, 32, 62, 63, 104, 167, 199, 819, 822
\prop_put:Nnn	102, 130, 474, 475, 544, 545
\prop_show:N	58, 92, 174, 681, 684, 866, 891
\ProvidesExplFile	3
\ProvidesExplPackage	3, 3, 3, 3, 3, 3, 3, 3, 3, 7, 26, 46, 815
Q	
\quad	162, 163
quark commands:	
\q_no_value	483, 493, 563, 583
\quark_if_no_value:NTF	140, 148, 166, 182, 207, 215, 238, 797
\quark_if_no_value_p:N	387, 388, 497, 498, 524, 525, 587, 588
\q_stop	197, 230, 266
R	
raw_(mc-key)	56, 221, 413
ref_(struct-key)	85, 371
ref commands:	
\ref_attribute_gset:nmmn	136, 138, 145, 147, 149
\ref_label:nn	132, 154, 332
\ref_value:nn	84, 462
\ref_value:nmm	6, 81, 81, 83, 160, 165
ref internal commands:	
_ref_value:nmm	86, 89
regex commands:	
\regex_replace_once:nmN	202
\renewcommand	300, 301
\RenewDocumentCommand	8
\RequirePackage	20, 42, 43, 272, 275, 281, 284
\rlap	254
role_(rolemap-key)	129, 643
role-missing	18, 43
role-namespaced_(rolemap-key)	129, 643
role-parent-child	46, 48
role-tag	18, 50
role-unknown	18, 43
role-unknown-tag	18, 43
root-AF_(setup-key)	86, 569
S	
\selectfont	6
seq commands:	
\seq_clear:N	214, 258
\seq_const_from_clist:Nn	21, 34
\seq_count:N	22, 25, 226, 856, 858, 860, 882, 908
\seq_get:NNTF	324, 627, 706, 713
\seq_gpop:NN	699
\seq_gpop:NNTF	97, 700
\seq_gpop_left:NN	201
\seq_gpush:Nn	12, 14, 80, 87, 634, 635
\seq_gput_left:Nn	206, 852
\seq_gput_right:Nn	32, 134, 170, 278
\seq_gremove_duplicates:N	195
\seq_gset_eq:NN	156, 218, 221
\seq_if_empty:NTF	197
\seq_item:Nn	113, 115, 122, 126, 133, 137, 171, 245, 301, 303, 310, 396, 397, 665, 666
\seq_log:N	172, 178, 196, 207, 361, 376
\seq_map_break:	409

\seq_map_indexed_inline:Nn	. 333, 346
\seq_map_inline:Nn 215, 272, 401, 846, 884
\seq_new:N 11, 13, 13, 15, 16, 16, 17, 18, 18, 19, 105, 106, 168, 820
\seq_pop_left:NN	... 342, 344, 345, 400
\seq_put_right:Nn 216
\seq_remove_all:Nn 219
\seq_set_eq:NN 204, 205, 399
\seq_set_from_clist:NN 841, 875
\seq_set_from_clist:Nn 84, 87, 193, 213, 330, 341
\seq_set_map:NNn 196
\seq_set_map_x:NNn 842, 876
\seq_set_split:Nnn 133, 395, 664
\seq_show:N	. 51, 145, 146, 173, 179, 217, 218, 220, 288, 637, 682, 685, 695
\seq_use:Nn 38, 107, 108, 162, 163, 202, 207, 257, 857
\l_tmpa_seq	258, 278, 288, 664, 665, 666
shipout commands:	
\g_shipout_readonly_int 72, 146, 148, 276
show-spaces␣(setup-key) 149, 6
ShowTagging 16, 33, 52
skip commands:	
\skip_horizontal:n 72
\c_zero_skip 72
stash␣(mc-key) 56, 113
stash␣(struct-key) 84, 371
\stepcounter 353
str commands:	
\str_case:nnTF 52
\str_const:Nn 58
\str_if_eq:nnTF 124, 310, 522
\str_if_eq_p:nn	248, 301, 303, 432, 449
\str_new:N 103
\str_set_convert:Nnnn	.. 134, 242, 260, 402, 414, 426, 427, 438, 439, 468
\str_use:N 253, 273
\string 20, 21, 22, 286
struct-faulty-nesting 18, 25
struct-label-unknown 18, 31
struct-missing-tag 18, 28
struct-no-objnum 18, 24
struct-show-closing 18, 33
struct-stack␣(show-key) 33, 175
struct-unknown 22
struct-used-twice 18, 29
sys commands:	
\c_sys_backend_str 52
\c_sys_engine_str 10, 12
\sys_if_engine luatex:TF 38, 41, 62, 66, 75, 85, 107, 210, 264
\sys_if_engine_pdftex:TF 11, 18
sys-no-interwordspace 19, 57
T	
tabsorder␣(setup-key) 6, 252
tag␣(mc-key) 56, 221, 413
tag␣(rolemap-key) 129, 643
tag␣(struct-key) 84, 371
tag commands:	
\tag_get:n 16, 83, 84, 97, 64, 64, 80, 83, 349
\tag_if_active: 66, 70
\tag_if_active:TF 16, 65, 202
\tag_if_active_p: 16, 65
\tag_mc_artifact_group_begin:n 55, 54, 54, 57
\tag_mc_artifact_group_end: 55, 54, 55, 64
\tag_mc_begin:n 10, 55, 16, 60, 105, 154, 154, 239, 243, 253, 318, 318, 322, 328, 332, 358, 406, 429
\tag_mc_begin_pop:n 55, 68, 71, 72, 93, 339, 367, 420, 443
\tag_mc_end: 55, 22, 67, 84, 208, 208, 241, 251, 255, 318, 319, 336, 363, 388, 394, 418, 441
\tag_mc_end_push: 55, 59, 71, 71, 74, 326, 351, 404, 427
\tag_mc_if_in: 73, 229
\tag_mc_if_in:TF 55, 33, 59, 222
\tag_mc_if_in_p: 55, 59, 222
\tag_mc_use:n 55, 27, 30, 30, 32
\tag_start: 6, 182, 194, 224
\tag_start:n 6, 182, 212, 226
\tag_stop: 6, 182, 189, 223
\tag_stop:n 6, 182, 200, 225
\tag_stop_group_begin:	6, 61, 182, 182
\tag_stop_group_end:	. 6, 66, 182, 188
\tag_struct_begin:n 83, 39, 237, 357, 405, 428, 592, 592, 595, 596
\tag_struct_end: 83, 26, 44, 257, 364, 419, 442, 592, 593, 691, 692
\tag_struct_gput:nnn	.. 778, 778, 786
\tag_struct_insert_annot:n 83, 106, 417, 440, 800, 800, 809
\tag_struct_object_ref:n 83, 772, 773, 777
\tag_struct_parent_int: 83, 106, 410, 417, 433, 440, 800, 810
\tag_struct_use:n 83, 84, 49, 722, 722, 724
tag internal commands:	
_tag_activate_mark_space 427

\g__tag_active_mc_bool	33, 75, 89, <u>114</u> , 232	__tag_check_info_closing_-	
\l__tag_active_mc_bool	78, 89, <u>120</u> , 186, 192, 197, 205, 218	struct:n	<u>122</u> , 122, 130, 702
\g__tag_active_space_bool	9, 48, 54, <u>114</u> , 231	__tag_check_init_mc_used:	221, 221, 224, 230
\g__tag_active_struct_bool	74, 99, <u>114</u> , 202, 234, 321	__tag_check_mc_if_nested:	159, <u>183</u> , 183, 335
\l__tag_active_struct_bool	77, 99, <u>120</u> , 185, 191, 196, 204, 217	__tag_check_mc_if_open:	<u>183</u> , 191, 212, 400
\g__tag_active_struct_dest_bool	<u>114</u> , 201, 238	__tag_check_mc_in_galley:TF ..	<u>291</u>
\g__tag_active_tree_bool	9, 32, 76, <u>114</u> , 233, 255, 269	__tag_check_mc_in_galley_p: ..	<u>291</u>
__tag_add_document_structure:n ..	<u>182</u> , 182, 193	__tag_check_mc_pushed_popped:nn	81, 88, 101, 104, 109, <u>198</u> , 198
__tag_add_missing_mcs:Nn	68, <u>164</u> , 164, 216	__tag_check_mc_tag:N	172, <u>210</u> , 210, 347
__tag_add_missing_mcs_to_-		__tag_check_mc_used:n	133, <u>226</u> , 226, 291
stream:Nn	58, 58, <u>186</u> , 186, 278, 282, 289, 291	\g__tag_check_mc_used_intarray ..	<u>221</u> , 231, 233, 236
\g__tag_attr_class_used_seq	195, 196, <u>818</u> , 852	__tag_check_no_open_struct: ..	<u>131</u> , 131, 704, 711
\g__tag_attr_entries_prop	201, <u>818</u> , 825, 848, 886, 891, 895	__tag_check_parent_child:nnN ..	512, 518, 603
__tag_attr_new_entry:nn	345, <u>823</u> , 823, 833	__tag_check_parent_child:nnnnN ..	<u>470</u>
\g__tag_attr_objref_prop	<u>818</u> , 890, 897, 902	__tag_check_parent_child:nnnnN ..	189,
\l__tag_attr_value_tl	<u>818</u> , 880, 899, 904, 906, 910, 914	363, 472, 514, 527, 542, 604, 645, 747	
__tag_check_add_tag_role:nn	131, <u>145</u> , 145	__tag_check_show_MCID_by_page: ..	<u>245</u> , 245
__tag_check_add_tag_role:nnn	160, 164	__tag_check_struct_used:n	<u>135</u> , 135, 731
__tag_check_if_active_mc:	87	__tag_check_structure_has_tag:n	<u>107</u> , 107, 619
__tag_check_if_active_mc:TF ..	76, 87, 95, 156, 188, 210, 324, 330, 390, 396	__tag_check_structure_tag:N	<u>115</u> , 115, 398
__tag_check_if_active_struct: ..	97	__tag_check_typeof_v:n	59, 59, 107, 108, 111,
__tag_check_if_active_struct:TF ..	34, <u>87</u> , 598, 599, 696, 697, 726, 803	146, 154, 161, 199, 208, 245, 281, 286	
__tag_check_if_mc_in_galley: ..	291	__tag_debug_mc_begin_ignore:n ..	326, 383
__tag_check_if_mc_in_galley:TF ..	138, 159	__tag_debug_mc_begin_insert:n ..	319, 332
__tag_check_if_mc_tmb_missing: ..	297	__tag_debug_mc_end_ignore: ..	340, 408
__tag_check_if_mc_tmb_missing:TF ..	109, 147, 164, <u>297</u>	__tag_debug_mc_end_insert: ..	333, 398
__tag_check_if_mc_tmb_missing_-		__tag_debug_struct_begin_-	
p:	297	ignore:n	364, 689
__tag_check_if_mc_tme_missing: ..	308	__tag_debug_struct_begin_-	
__tag_check_if_mc_tme_missing:TF ..	151, 152, 168, <u>308</u>	insert:n	356, 686
__tag_check_if_mc_tme_missing_-		__tag_debug_struct_end_ignore: ..	379, 719
p:	<u>308</u>	__tag_debug_struct_end_insert: ..	371, 717
		__tag_exclude_headfoot_begin: ..	321, 378, 379

__tag_exclude_headfoot_end: ...	\l__tag_mc_artifact_bool
..... 334, 380, 381 15, 116, 161, 175, 214, 340
__tag_exclude_struct_headfoot_-	\l__tag_mc_artifact_type_tl
begin:n 346, 385, 386 14, 120, 124, 128,
__tag_exclude_struct_headfoot_-	132, 136, 140, 144, 148, 307, 342, 344
end: 361, 387, 388	__tag_mc_bdc:nn 230, 233, 234, 274, 306
__tag_fakespace 363	__tag_mc_bdc_mcid:n .. 120, 235, 278
__tag_fakespace: 66, 68, 214	__tag_mc_bdc_mcid:nn
__tag_finish_structure: 235, 235, 280, 285
..... 13, 16, 252, 253	__tag_mc_begin_marks:nn
__tag_get_data_mc_tag: 20, 20, 41, 77, 351
..... 220, 220, 316, 316	__tag_mc_bmc:n 230, 231, 302
__tag_get_data_struct_num: 365, 366	__tag_mc_bmc_artifact: 300, 300, 313
__tag_get_data_struct_tag: 356, 356	__tag_mc_bmc_artifact:n 300, 304, 314
__tag_get_mathsubtype 251	\l__tag_mc_botmarks_seq
__tag_get_mc_abs_cnt: .. 9, 9, 19, 68, 18, 87, 108,
20, 64, 93, 94, 105, 168, 187, 195,	146, 158, 163, 205, 213, 218, 293, 310
202, 214, 229, 237, 253, 271, 284, 294	__tag_mc_disable_marks: 75, 75
__tag_get_mc_cnt_type_tag 245	__tag_mc_emc: 155, 230, 232, 403
__tag_get_num_from 270	__tag_mc_end_marks: . 20, 60, 79, 404
\l__tag_get_parent_tmpa_tl	\l__tag_mc_firstmarks_seq
..... 101, 643, 646, 656 68, 18, 84, 107, 145, 162,
\l__tag_get_parent_tmpa_tl\l__-	193, 196, 197, 204, 205, 293, 301, 303
_tag_get_parent_tmpb_tl\l__-	\g__tag_mc_footnote_marks_seq ... 15
_tag_tmpa_str 98	__tag_mc_get_marks: 81, 81, 137, 158
\l__tag_get_parent_tmpb_tl	__tag_mc_handle_artifact:N
..... 102, 644, 647, 656 116, 300, 308, 342
__tag_get_tag_from 289	__tag_mc_handle_mc_label:n
\l__tag_get_tmpe_tl 98, 21, 21, 180, 355
121, 126, 135, 137, 138, 400, 793, 797	__tag_mc_handle_mcid:nn
__tag_hook_kernel_after_foot: 235, 283, 288, 348
..... 307, 316, 381, 388, 395	__tag_mc_handle_stash:n 44,
__tag_hook_kernel_after_head: ..	131, 131, 153, 202, 289, 289, 299, 376
..... 305, 314, 380, 387, 394	__tag_mc_if_in: 59, 73, 222, 229
__tag_hook_kernel_before_foot: .	__tag_mc_if_in:TF 59, 78, 185, 193, 222
..... 306, 315, 379, 386, 393	__tag_mc_if_in_p: 59, 222
__tag_hook_kernel_before_head: .	__tag_mc_insert_extra_tmb:n ...
..... 304, 313, 378, 385, 392 105, 105, 168
\g__tag_in_mc_bool	__tag_mc_insert_extra_tme:n ...
..... 11, 18, 160, 213, 224, 105, 150, 169
329, 330, 336, 342, 354, 355, 370, 401	__tag_mc_insert_mcid_kids:n ...
__tag_insert_bdc_node 341 122, 122, 138, 203
__tag_insert_bmc_node 334	__tag_mc_insert_mcid_single_-
__tag_insert_emc_node 327	kids:n 122, 127, 204
__tag_lastpagelabel: 62, 63, 80	\l__tag_mc_key_label_tl
__tag_log 173	. 17, 177, 180, 279, 351, 352, 355, 446
\l__tag_loglevel_int	\l__tag_mc_key_properties_tl ...
..... 113, 124, 134, 154, 17, 162, 234, 247, 248,
161, 170, 173, 201, 204, 228, 241,	265, 266, 350, 423, 432, 433, 443, 444
244, 247, 248, 249, 321, 328, 335,	\l__tag_mc_key_stash_bool
342, 358, 366, 373, 381, 414, 624, 637 15, 28, 37, 115, 183, 357
__tag_mark_spaces 368	\g__tag_mc_key_tag_tl ... 17, 19,
__tag_mc_artifact_begin_marks:n	165, 217, 220, 226, 316, 338, 402, 419
..... 20, 42, 78, 344	

\l__tag_mc_key_tag_tl [17](#), [164](#), [172](#),
 [174](#), [216](#), [225](#), [337](#), [347](#), [349](#), [351](#), [418](#)
 __tag_mc_lua_set_mc_type_attr:n
 [74](#), [74](#), [98](#), [174](#)
 __tag_mc_lua_unset_mc_type_-
 attr: [74](#), [100](#), [215](#)
 \g__tag_mc_main_marks_seq [15](#)
 \g__tag_mc_marks [14](#),
 [22](#), [31](#), [44](#), [51](#), [62](#), [68](#), [85](#), [88](#), [194](#), [214](#)
 \g__tag_mc_multicol_marks_seq ... [15](#)
 \g__tag_mc_parenttree_prop
 [12](#), [13](#), [100](#), [105](#), [148](#), [295](#)
 \l__tag_mc_ref_abspage_tl
 [12](#), [238](#), [250](#), [258](#), [266](#)
 __tag_mc_set_label_used:n [25](#), [25](#), [45](#)
 \g__tag_mc_stack_seq [13](#), [80](#), [87](#), [97](#), [207](#)
 __tag_mc_store:nnn . [90](#), [90](#), [104](#), [131](#)
 \l__tag_mc_tmpa_tl .. [13](#), [252](#), [255](#), [259](#)
 g__tag_MCID_abs_int [7](#)
 \g__tag_MCID_byabspage_prop
 [10](#), [248](#), [257](#), [265](#)
 \g__tag_MCID_tmp_bypage_int
 [10](#), [150](#), [255](#), [263](#), [276](#)
 \g__tag_mode_lua_bool
 [37](#), [38](#), [39](#), [73](#), [155](#), [212](#),
 [238](#), [270](#), [273](#), [279](#), [324](#), [337](#), [349](#), [365](#)
 __tag_new_output_prop_handler:n
 [64](#), [74](#), [84](#), [604](#)
 __tag_pairs_prop [190](#)
 \g__tag_para_begin_int
 [219](#), [236](#), [240](#), [262](#), [267](#)
 \l__tag_para_bool
 [217](#), [227](#), [234](#), [248](#), [300](#), [301](#), [323](#), [348](#)
 \g__tag_para_end_int
 [220](#), [250](#), [254](#), [262](#), [268](#)
 \g__tag_para_int [217](#)
 \l__tag_para_show_bool
 [217](#), [228](#), [238](#), [252](#)
 \l__tag_para_tag_default_tl ... [217](#)
 \l__tag_para_tag_tl [223](#), [224](#), [229](#), [237](#)
 \l__tag_parent_child_check_tl ...
 [192](#),
 [193](#), [366](#), [367](#), [379](#), [650](#), [651](#), [752](#), [753](#)
 __tag_parenttree_add_objr:nn ...
 [82](#), [82](#), [345](#)
 \l__tag_parenttree_content_tl ...
 [89](#), [108](#), [120](#), [134](#), [142](#), [162](#), [165](#)
 \g__tag_parenttree_objr_tl [81](#), [84](#), [162](#)
 __tag_pdf_name_e:n [82](#), [82](#)
 __tag_pdf_object_ref [348](#)
 __tag_prop_gput:Nnn
 [9](#), [23](#), [45](#), [49](#), [59](#),
 [87](#), [92](#), [98](#), [107](#), [117](#), [123](#), [167](#), [169](#),
 [176](#), [219](#), [243](#), [254](#), [256](#), [264](#), [407](#),
 [419](#), [431](#), [444](#), [451](#), [473](#), [506](#), [532](#),
 [556](#), [576](#), [609](#), [669](#), [737](#), [794](#), [862](#), [911](#)
 __tag_prop_item:Nn .. [9](#), [43](#), [167](#), [172](#)
 __tag_prop_new:N [9](#),
 [9](#), [10](#), [11](#), [12](#), [83](#), [167](#), [167](#), [178](#), [603](#)
 __tag_prop_show:N [9](#), [56](#), [167](#), [174](#), [181](#)
 __tag_ref_label:nn
 [23](#), [151](#), [151](#), [157](#), [269](#), [623](#)
 __tag_ref_value:nnn ... [36](#), [100](#),
 [104](#), [138](#), [152](#), [153](#), [158](#), [158](#), [162](#),
 [240](#), [270](#), [281](#), [341](#), [729](#), [735](#), [738](#), [744](#)
 __tag_ref_value_lastpage:nn ...
 [79](#), [93](#), [96](#), [163](#), [163](#), [249](#), [263](#)
 \c__tag_refmc_clist [111](#)
 \c__tag_refstruct_clist [111](#)
 g__tag_role/RoleMap_dict [17](#)
 __tag_role_add_tag:nn
 [129](#), [129](#), [157](#), [218](#), [294](#), [687](#)
 __tag_role_add_tag:nnnn
 [158](#), [158](#), [192](#), [250](#), [692](#)
 __tag_role_alloctag:nnn [83](#),
 [87](#), [97](#), [109](#), [119](#), [128](#), [144](#), [170](#), [215](#), [246](#)
 \l__tag_role_debug_prop
 [130](#), [11](#), [429](#), [432](#),
 [435](#), [446](#), [449](#), [452](#), [474](#), [475](#), [544](#), [545](#)
 __tag_role_get_parent_child_-
 rule:nnN [379](#), [380](#), [407](#), [469](#), [500](#), [590](#)
 \g__tag_role_index_prop . [130](#), [10](#),
 [335](#), [343](#), [355](#), [356](#), [357](#), [362](#), [368](#),
 [370](#), [371](#), [374](#), [376](#), [384](#), [385](#), [476](#), [486](#)
 \g__tag_role_NS<ns>_class_prop [130](#)
 \g__tag_role_NS<ns>_prop [130](#)
 \g__tag_role_NS_mathml_prop ... [372](#)
 __tag_role_NS_new:nnn [132](#),
 [19](#), [21](#), [29](#), [72](#), [73](#), [76](#), [78](#), [79](#), [80](#), [82](#)
 \g__tag_role_NS_prop
 ... [130](#), [9](#), [25](#), [55](#), [121](#), [230](#), [248](#), [675](#)
 \g__tag_role_parent_child_-
 intarray [311](#), [314](#), [393](#)
 __tag_role_read_namespace:n ...
 [256](#), [256](#),
 [275](#), [276](#), [278](#), [280](#), [281](#), [283](#), [284](#), [285](#)
 __tag_role_read_namespace_-
 line:nw [193](#), [197](#), [230](#), [266](#)
 \l__tag_role_real_parent_tl ...
 [378](#), [383](#), [405](#), [425](#), [439](#)
 __tag_role_remap:
 [607](#), [607](#), [608](#), [661](#), [757](#)
 __tag_role_remap_id: [608](#), [608](#)
 __tag_role_remap_inline:
 [609](#), [611](#), [631](#)
 \l__tag_role_remap_NS_tl
 ... [605](#), [619](#), [635](#), [660](#), [663](#), [756](#), [759](#)

\l__tag_role_remap_tag_tl [605](#), [613](#),
 [615](#), [626](#), [633](#), [639](#), [659](#), [662](#), [755](#), [758](#)
 \l__tag_role_role_namespace_
 tmpa_tl [12](#),
 [648](#), [668](#), [673](#), [675](#), [677](#), [681](#), [696](#)
 \l__tag_role_role_tmpa_tl
 [12](#), [647](#), [666](#), [672](#), [689](#), [695](#)
 \g__tag_role_rolemap_prop
 . [130](#), [17](#), [147](#), [150](#), [153](#), [176](#), [481](#), [491](#)
 \c__tag_role_rules_num_prop [312](#), [416](#)
 \c__tag_role_rules_prop [312](#), [315](#), [397](#)
 \l__tag_role_tag_namespace_tmpa_
 tl [12](#), [520](#), [524](#), [528](#), [646](#), [694](#)
 \l__tag_role_tag_namespace_tmpb_
 tl [521](#), [522](#), [525](#), [529](#)
 \l__tag_role_tag_tmpa_tl
 [12](#), [645](#), [665](#), [688](#), [693](#)
 \g__tag_role_tags_class_prop ...
 ... [130](#), [8](#), [92](#), [101](#), [114](#), [123](#), [139](#), [206](#)
 \g__tag_role_tags_NS_prop
 ... [130](#), [7](#), [90](#), [99](#), [112](#), [117](#), [121](#),
 [132](#), [152](#), [216](#), [306](#), [395](#), [520](#), [521](#), [671](#)
 \l__tag_role_tmpa_seq [12](#), [399](#), [400](#), [401](#)
 \l__tag_role_update_bool
 [193](#), [194](#), [202](#), [279](#), [282](#)
 \c__tag_role_userNS_id_str
 [131](#), [58](#), [82](#)
 \g__tag_saved_in_mc_bool
 [320](#), [329](#), [342](#), [354](#), [370](#)
 __tag_seq_gput_right:Nn [9](#),
 [30](#), [158](#), [163](#), [167](#), [170](#), [173](#), [177](#), [190](#)
 __tag_seq_item:Nn ... [9](#), [38](#), [167](#), [171](#)
 __tag_seq_new:N
 [9](#), [9](#), [16](#), [85](#), [167](#), [168](#), [179](#), [605](#)
 __tag_seq_show:N . [9](#), [49](#), [167](#), [173](#), [180](#)
 __tag_show_spacemark [354](#)
 \l__tag_showspaces_bool ... [14](#), [26](#), [35](#)
 __tag_space_chars_shipout [448](#)
 \g__tag_state_prop . [199](#), [206](#), [209](#), [214](#)
 __tag_store_parent_child_
 rule:nnn [312](#), [312](#), [349](#)
 g__tag_struct_0_prop [83](#)
 __tag_struct_add_AF:nn
 [479](#), [498](#), [505](#), [529](#), [553](#), [575](#)
 \g__tag_struct_cont_mc_prop
 [10](#), [92](#), [93](#), [95](#), [98](#), [166](#)
 \g__tag_struct_dest_num_prop ... [62](#)
 \l__tag_struct_elem_stash_bool ..
 [61](#), [375](#), [639](#)
 __tag_struct_exchange_kid_
 command:N [199](#), [199](#), [209](#), [240](#)
 __tag_struct_fill_kid_key:n ...
 [210](#), [210](#), [304](#)
 __tag_struct_format_Ref:n
 [298](#), [298](#), [299](#)
 __tag_struct_get_dict_content:nN
 [269](#), [269](#), [305](#)
 __tag_struct_get_tag_info:nnn ..
 [132](#), [145](#), [185](#), [359](#), [641](#), [743](#)
 __tag_struct_gput_data_ref:nn ..
 [464](#), [787](#), [787](#), [799](#)
 __tag_struct_insert_annot:nn ...
 [318](#), [318](#), [805](#)
 \l__tag_struct_key_label_tl
 [60](#), [374](#), [621](#), [623](#)
 __tag_struct_kid_mc_gput_
 right:nn [146](#), [156](#), [169](#), [292](#)
 __tag_struct_kid_OBJR_gput_
 right:nnn [181](#), [181](#), [197](#), [333](#)
 __tag_struct_kid_struct_gput_
 right:nn ... [171](#), [171](#), [180](#), [678](#), [733](#)
 g__tag_struct_kids_0_seq [83](#)
 __tag_struct_mcid_dict:n
 [95](#), [98](#), [146](#), [161](#)
 \g__tag_struct_objR_seq [8](#)
 __tag_struct_output_prop_aux:nn
 [64](#), [64](#), [78](#)
 \g__tag_struct_ref_by_dest_prop . [63](#)
 __tag_struct_set_tag_info:nnn ..
 [102](#), [104](#), [115](#), [131](#), [615](#), [664](#), [760](#)
 \g__tag_struct_stack_current_tl .
 [16](#), [26](#), [35](#),
 [66](#), [72](#), [81](#), [136](#), [144](#), [150](#), [186](#), [206](#),
 [293](#), [297](#), [360](#), [361](#), [368](#), [636](#), [658](#),
 [676](#), [680](#), [681](#), [684](#), [702](#), [708](#), [734](#), [741](#)
 \l__tag_struct_stack_parent_
 tmpa_tl
 [16](#), [326](#), [335](#), [350](#), [385](#), [613](#),
 [625](#), [629](#), [642](#), [673](#), [677](#), [679](#), [682](#), [685](#)
 \g__tag_struct_stack_seq [11](#),
 [22](#), [25](#), [325](#), [628](#), [634](#), [637](#), [695](#), [700](#), [706](#)
 \c__tag_struct_StructElem_
 entries_seq [21](#)
 \c__tag_struct_StructTreeRoot_
 entries_seq [21](#)
 \g__tag_struct_tag_NS_prop
 [15](#), [97](#), [111](#), [128](#), [135](#)
 \g__tag_struct_tag_NS_tl
 [58](#), [198](#), [372](#), [397](#), [618](#), [649](#),
 [657](#), [660](#), [663](#), [667](#), [749](#), [756](#), [759](#), [763](#)
 \g__tag_struct_tag_stack_seq ...
 [13](#), [38](#),
 [178](#), [179](#), [361](#), [376](#), [399](#), [635](#), [699](#), [713](#)
 \g__tag_struct_tag_tl
 . [58](#), [164](#), [165](#), [168](#), [198](#), [337](#), [338](#),
 [372](#), [396](#), [398](#), [617](#), [635](#), [648](#), [657](#),
 [659](#), [662](#), [666](#), [715](#), [748](#), [755](#), [758](#), [762](#)

<code>__tag_struct_write_obj:n</code>	<code>__tag_tree_write_rolemap:</code>
..... 53, 63, 70, 300, 300 172, 174, 189, 260
<code>\g__tag_tagunmarked_bool</code> ... 124, 250	<code>__tag_tree_write_structelements:</code>
<code>\l__tag_tmpa_box</code> 66, 66, 263
..... 98, 168, 174, 175, 179, 190, 191	<code>__tag_tree_write_structtreeroot:</code>
<code>\l__tag_tmpa_clist</code> 41, 43, 57, 264
..... 98, 840, 841, 874, 875	<code>__tag_whatsits:</code> 30, 54, 55, 58, 318, 319
<code>\l__tag_tmpa_int</code>	<code>tag-namespace_ (rolemap-key)</code> 643
..... 98, 317, 329, 331, 406, 408, 414	<code>tag/struct/0</code> internal commands:
<code>\l__tag_tmpa_prop</code> 95, 98, 103, 116, 118	<code>__tag/struct/0</code> 29
<code>\l__tag_tmpa_seq</code>	<code>tag/tree/namespaces</code> internal commands:
..... 98, 196, 208, 214, 216, 218, 219,	<code>__tag/tree/namespaces</code> 225
220, 221, 330, 333, 341, 342, 344,	<code>tag/tree/parenttree</code> internal commands:
345, 346, 395, 396, 397, 842, 846,	<code>__tag/tree/parenttree</code> 73
856, 857, 858, 860, 876, 882, 884, 908	<code>tag/tree/rolemap</code> internal commands:
<code>\l__tag_tmpa_str</code> 41,	<code>__tag/tree/rolemap</code> 168
42, 47, 103, 243, 248, 253, 261, 266,	<code>tagabspage</code> 6, 136
273, 403, 410, 415, 422, 427, 428,	<code>tagmcabs</code> 6, 136
433, 434, 439, 440, 444, 447, 469, 476	<code>\tagmcbegin</code> 32, 129, 13
<code>\l__tag_tmpa_tl</code>	<code>\tagmcend</code> 32, 13
..... 36, 37, 44, 77, 84, 93, 95,	<code>tagmcid</code> 6, 136
97, 98, 99, 104, 105, 115, 116, 118,	<code>\tagmcifin</code> 32
119, 122, 124, 139, 140, 142, 144,	<code>\tagmcifinTF</code> 32, 30
147, 148, 153, 165, 166, 168, 170,	<code>\tagmcuse</code> 32, 13
181, 182, 187, 188, 190, 194, 201,	<code>\tagpdfifluatexT</code> 32
205, 205, 206, 206, 207, 209, 212,	<code>\tagpdfifluatexTF</code> 32
213, 214, 215, 215, 218, 237, 238,	<code>\tagpdfifpdfTeX</code> 32
240, 244, 246, 247, 255, 305, 311,	<code>\tagpdfparaOff</code> 34, 297
342, 343, 344, 345, 355, 356, 357,	<code>\tagpdfparaOn</code> 34, 297
361, 362, 364, 368, 370, 374, 384,	<code>\tagpdfsetup</code> 32, 85, 86, 129, 6
387, 394, 416, 418, 455, 458, 461,	<code>\tagpdfsuppressmarks</code> 34, 302
464, 476, 478, 481, 483, 497, 500,	<code>tagstruct</code> 6, 136
548, 554, 556, 557, 559, 563, 587,	<code>\tagstructbegin</code> 33, 129, 36, 184
590, 613, 617, 621, 633, 699, 700,	<code>\tagstructend</code> 33, 36, 185
706, 708, 713, 715, 745, 750, 854, 865	<code>tagstructobj</code> 6, 136
<code>\l__tag_tmpb_box</code>	<code>\tagstructuse</code> 33, 36
..... 98, 169, 176, 177, 181, 183	<code>tagunmarked_ (setup-key)</code> 6, 250
<code>\l__tag_tmpb_seq</code> 98, 841, 842, 875, 876	<code>TeX</code> and <code>L^AT_EX 2_ε</code> commands:
<code>\l__tag_tmpb_tl</code> 140, 98,	<code>\@M</code> 165
188, 190, 343, 349, 362, 364, 371,	<code>\@auxout</code> 67
376, 385, 388, 394, 420, 425, 486,	<code>\@bsphack</code> 153
488, 491, 493, 498, 500, 568, 574,	<code>\@cclv</code> 282
576, 577, 579, 583, 588, 590, 746, 751	<code>\@esphack</code> 155
<code>__tag_tree_fill_parenttree:</code> ...	<code>\@gobble</code> 24, 48
..... 90, 91, 160	<code>\@ifpackageloaded</code> 28
<code>__tag_tree_final_checks:</code> 20, 20, 258	<code>\@kernel@after@foot</code> 316
<code>__tag_tree_lua_fill_parenttree:</code>	<code>\@kernel@after@head</code> 314
..... 140, 140, 157	<code>\@kernel@before@cclv</code> 279
<code>__tag_tree_write_classmap:</code>	<code>\@kernel@before@foot</code> 315
..... 192, 192, 261	<code>\@kernel@before@footins</code> 275, 277
<code>__tag_tree_write_namespaces:</code> ...	<code>\@kernel@before@head</code> 311, 313
..... 226, 226, 262	<code>\@makecol</code> 281
<code>__tag_tree_write_parenttree:</code> ...	<code>\@maxdepth</code> 178
..... 153, 153, 259	<code>\@mult@ptagging@hook</code> 284

<code>\@secondoftwo</code>	24, 48	59, 60, 81, 89, 98, 99, 100, 101, 102, 221, 223, 378, 379, 491, 605, 606, 821	
<code>\c@page</code>	281	<code>\tl_put_left:Nn</code>	314, 316
<code>\count@</code>	289	<code>\tl_put_right:Nn</code>	108, 120, 133, 162, 234, 247, 248, 265, 266, 277, 279, 279, 284, 313, 315, 423, 432, 433, 443, 444, 461, 899, 906
<code>\mult@firstbox</code>	287	<code>\tl_set:Nn</code> ...	36, 77, 115, 120, 124, 128, 132, 136, 137, 138, 140, 142, 142, 144, 148, 168, 205, 206, 209, 213, 216, 222, 224, 225, 238, 240, 244, 247, 279, 383, 385, 390, 405, 418, 418, 420, 460, 478, 483, 488, 493, 503, 533, 548, 556, 559, 563, 568, 576, 579, 583, 593, 613, 615, 619, 635, 665, 666, 677, 681, 854, 880
<code>\mult@rightbox</code>	291	<code>\tl_set_eq:NN</code>	164, 337
<code>\page@sofar</code>	286	<code>\tl_show:N</code>	676, 677, 904, 910
<code>\process@cols</code>	287	<code>\tl_tail:n</code>	359
tex commands:		<code>\tl_to_str:n</code> 27, 42, 87, 150, 185, 316, 349
<code>\tex_botmarks:D</code>	88	<code>\tl_use:N</code>	92, 511, 537, 561, 581
<code>\tex_firstmarks:D</code>	85	<code>\l_tmpa_tl</code> ..	140, 152, 171, 661, 662, 664
<code>\tex_kern:D</code>	181	token commands:	
<code>\tex_marks:D</code>	22, 31, 44, 51, 62, 68	<code>\token_to_str:N</code>	69, 281
<code>\tex_special:D</code>	58	tree-mcid-index-wrong	19, <u>55</u>
<code>\tex_splitbotmarks:D</code>	214	tree-struct-still-open	<u>35</u>
<code>\tex_splitfirstmarks:D</code>	194		
<code>\the</code>	281		
<code>\tiny</code>	240, 254		
<code>title_␣(struct-key)</code>	84, <u>371</u>		
<code>title-o_␣(struct-key)</code>	84, <u>371</u>		
tl commands:			
<code>\c_empty_tl</code>	294		
<code>\c_space_tl</code> ..	70, 86, 110, 111, 126, 149, 151, 153, 165, 175, 198, 262, 281, 286, 435, 439, 452, 797, 857, 901		
<code>\tl_clear:N</code>			
....	141, 142, 162, 194, 271, 458, 522		
<code>\tl_gput_right:Nn</code>	84, 486		
<code>\tl_gset:Nn</code>	18, 81, 217, 226, 396, 397, 402, 419, 493, 636, 708, 715		
<code>\tl_gset_eq:NN</code>	165, 338		
<code>\tl_head:N</code>	556, 576		
<code>\tl_if_empty:NTF</code> ..	37, 42, 177, 211, 212, 312, 352, 557, 577, 620, 662, 668		
<code>\tl_if_empty:nTF</code>			
.....	50, 145, 147, 166, 179, 200, 204, 216, 233, 235, 327, 546, 566		
<code>\tl_if_empty_p:n</code>	248		
<code>\tl_if_eq:NNTF</code>	293, 425		
<code>\tl_if_eq:NnTF</code>	99		
<code>\tl_if_exist:NTF</code>	91, 481		
<code>\tl_if_in:nNTF</code>	403		
<code>\tl_new:N</code>	12, 12, 13, 13, 14, 14, 15, 17, 17, 18, 19, 20, 20, 27, 58,		

U

<code>\unskip</code>	32
use commands:	
<code>\use:N</code>	64, 303
<code>\use:n</code>	34
<code>\use_i:nn</code>	137, 294, 617
<code>\use_ii:nn</code>	138, 248, 621
<code>\use_none:n</code>	59, 78
<code>\use_none:nn</code>	77, 782

V

<code>\vbadness</code>	165, 189
vbox commands:	
<code>\vbox_set_split_to_ht:NNn</code>	191
<code>\vbox_set_to_ht:Nnn</code>	167
<code>\vbox_unpack_drop:N</code>	180
<code>\vfuzz</code>	166