

Prototype reimplementation of L^AT_EX 2_ε's role mapping

L^AT_EX Project, initial implementation Ulrike Fischer

v0.8k 2026-04-28

1 Introduction

This file provides support for tag names and symbolic structure names.

At first, it creates tagpdf-ns-latex-lab.def which collects tag names from the L^AT_EX namespace and their role mapping. The names and the roles are not necessarily the final ones!

Later settings of the same tag overwrite previous settings. The syntax of a line is a list of three strings ending with commas:

latex-tag, PDF-tag, name space,

The name space is normally pdf2 with the exception of the following PDF-tags which are only in pdf namespace: Art, BlockQuote, TOC, TOCI, Index, Private, Quote, Note, Reference, BibEntry, Code.

At second it provides an interface to use symbolic structure names which can be used in the `tag` key of `\tagstructbegin`. The symbolic names are strings structured with slashes similar to the names of (tagging) sockets. The list of such names used by L^AT_EX is below in the implementation section.

In PDF 2.0 there are namespaces, so theoretically it is possible to use the same tag name for different purposes and with different role mapping. But practically this is not an option as we build also a global RoleMap dictionary for compatibility with applications that are not PDF 2.0 aware, and this dictionary can only contain unique tag names. So we assume that tag names are used only once across all namespaces. Documents that want to use their own specialized tag names must take care not to clash with existing names, e.g., by using a specific prefix on all their tag names.

<code>\NewStructureName</code>	<code>\NewStructureName{<symbolic name>}</code>
--------------------------------	---

This declares a new symbolic structure name `<symbolic name>` which is initially defined as `NonStruct`. To assign it another role use `\AssignStructureRole` described below.

<code>\UseStructureName</code> ★	<code>\UseStructureName{<symbolic name>}</code>
----------------------------------	---

This uses the new symbolic structure name `<symbolic name>`, e.g.

`tagstructbegin{tag=\UseStructureName{sec/section}}` is the same as `tagstructbegin{tag=Sect}` assuming that `\NewStructureName{sec/section}` and `\AssignStructureRole{sec/section}{Sect}` has been used before.

`\AssignStructureRole` `\AssignStructureRole{<symbolic name>}{<role>}`

This assigns to the symbolic structure name `<symbolic name>` the role `<role>`. `<symbolic name>` must have been declared first with `\NewStructureName`. If `<role>` is a known role it is simply used. If it is unknown, it is declared first and rolemapped to the tag currently stored for `<name>`. So

```

\NewStructureName{sec/section}
% structure with tag 'NonStruct'
\tagstructbegin{tag=\UseStructureName{sec/section}}

\AssignStructureRole{sec/section}{Sect}
% structure with tag 'Sect'
\tagstructbegin{tag=\UseStructureName{sec/section}}

\AssignStructureRole{sec/section}{Abschnitt}
% structure with tag 'Abschnitt' rolemapped to 'Sect'
\tagstructbegin{tag=\UseStructureName{sec/section}}

\AssignStructureRole{sec/section}{Figure}
% structure with tag 'Figure'
\tagstructbegin{tag=\UseStructureName{sec/section}}

```

Assigning a symbolic structure name a role that is not suitable for the structure element can lead to invalid tagging and parent-child warnings!

```

1 <ns-latex-lab>
2 %% \ProvidesExplFile {tagpdf-ns-latex-lab.def} {2023-09-04} {0.8}
3 %% {latex} {https://www.latex-project.org/ns/lab/2022}{ }

```

1.1 Role from tagpdf

The following tags are already set by tagpdf, so deleting here will not remove them.

```

4 title,      Title,    pdf2,
5 part,      Title,    pdf2,

```

Headings differ in article and book class. This here is for article:

```

6 section,    H1,      pdf2,
7 subsection, H2,      pdf2,
8 subsubsection, H3,    pdf2,
9 paragraph,  H4,      pdf2,
10 subparagraph, H5,    pdf2,

```

And this here is used in book:

```

11 %chapter,    H1,      pdf2,
12 %section,    H2,      pdf2,
13 %subsection, H3,      pdf2,
14 %subsubsection, H4,    pdf2,
15 %paragraph,  H5,      pdf2,
16 %subparagraph, H6,    pdf2,

17 list,        L,      pdf2,
18 itemize,     L,      pdf2,

```

```

19 enumerate, L, pdf2,
20 description, L, pdf2,
21 quote, BlockQuote, pdf,
22 quotation, BlockQuote, pdf,
23 verbatim, Code, pdf, % overwritten by block
24 item, LI, pdf2,
25 itemlabel, Lbl, pdf2,
26 itembody, LBody, pdf2,
27 footnote, FENote, pdf2,
28 footnotemark, Lbl, pdf2,
29 footnotelabel, Lbl, pdf2,
30 text-unit, Part, pdf2,
31 text, P, pdf2,

```

1.2 From sec code

```

32 section-number, Span, pdf2,

```

1.3 From block code

```

33 theorem-like, Sect, pdf2,
34 verbatim, Code, pdf,
35 codeline, Sub, pdf2,

```

1.4 From float code

```

36 float, Aside, pdf2,
37 figures, Sect, pdf2,
38 tables, Sect, pdf2,

39 </ns-latex-lab>

40 <*package>
41 \ProvidesExplPackage {latex-lab-testphase-names} {\ltxlabnamesdate} {\ltxlabnamesversion}
42 {Code related to the names of tags}

43 <@@=tag>

```

\AssignStructureRole We redefine \AssignStructureRole to set up roles for tagging. The second argument needs expansion so that we do not store things like \UseStructureName{block/verbatim/codeline}} but its final value. Otherwise ...

```

44 \cs_set_protected:Npn \AssignStructureRole#1#2 {
45   \exp_args:Nne \__tag_AssignStructureRole {#1}{#2}
46 }
47 \cs_set_protected:Npn \__tag_AssignStructureRole#1#2
48 {
49   \prop_get:NnNF\g__tag_role_tags_NS_prop{#2}\l__tag_tmp_unused_tl
50   {\exp_args:Ne\tagpdfsetup{role/new-tag=#2/\UseStructureName{#1}}}
51   \tl_set:cn { l__tag_name_#1_tl }{#2}
52 }

```

(End of definition for \AssignStructureRole. This function is documented on page 2.)

```

53 </package>

```

fix implementation and
docu

1.5 Symbolic names

These are the planned symbolic structure names. Every commented name is not decided and not used yet and can change in future.

1.5.1 Generic

In various places we have to use a `Span` structure to add attributes.

```
54 \<package>
55 \NewStructureName{span}
56 \AssignStructureRole{span}{Span}
57 \</package>
```

1.5.2 Paragraphs

```
58 \<package>
59 \NewStructureName{para/textblock}
60 \AssignStructureRole{para/textblock}{text}
61 \NewStructureName{para/semantic}
62 \AssignStructureRole{para/semantic}{text-unit}
63 \</package>
```

1.5.3 Sectioning commands

```
64 \<package>
65 %-1=part, 0=chapter, 1=section etc
66 \NewStructureName{sec/-1}
67 \AssignStructureRole{sec/-1}{Part}
68 \NewStructureName{sec/-1/title}
69 \AssignStructureRole{sec/-1/title}{part}
70 \NewStructureName{sec/-1/titleline}
```

Sub does not work as part is rolemapped to Title so we use `Span`

```
71 \AssignStructureRole{sec/-1/titleline}{Span}
72 \NewStructureName{sec/-1/number}
73 \AssignStructureRole{sec/-1/number}{section-number}
74 \NewStructureName{sec/0}
75 \AssignStructureRole{sec/0}{Sect}
76 \NewStructureName{sec/0/title}
77 \AssignStructureRole{sec/0/title}{chapter}
78 \NewStructureName{sec/0/titleline}
79 \AssignStructureRole{sec/0/titleline}{Span}
80 \NewStructureName{sec/0/number}
81 \AssignStructureRole{sec/0/number}{section-number}
82 \NewStructureName{sec/1}
83 \AssignStructureRole{sec/1}{Sect}
84 \NewStructureName{sec/1/title}
85 \AssignStructureRole{sec/1/title}{section}
86 \NewStructureName{sec/1/number}
87 \AssignStructureRole{sec/1/number}{section-number}
88 \NewStructureName{sec/2}
89 \AssignStructureRole{sec/2}{Sect}
90 \NewStructureName{sec/2/title}
91 \AssignStructureRole{sec/2/title}{subsection}
```

```

92 \NewStructureName{sec/2/number}
93 \AssignStructureRole{sec/2/number}{section-number}
94 \NewStructureName{sec/3}
95 \AssignStructureRole{sec/3}{Sect}
96 \NewStructureName{sec/3/title}
97 \AssignStructureRole{sec/3/title}{subsubsection}
98 \NewStructureName{sec/3/number}
99 \AssignStructureRole{sec/3/number}{section-number}
100 \NewStructureName{sec/4}
101 \AssignStructureRole{sec/4}{Sect}
102 \NewStructureName{sec/4/title}
103 \AssignStructureRole{sec/4/title}{paragraph}
104 \NewStructureName{sec/4/number}
105 \AssignStructureRole{sec/4/number}{section-number}
106 \NewStructureName{sec/5}
107 \AssignStructureRole{sec/5}{Sect}
108 \NewStructureName{sec/5/title}
109 \AssignStructureRole{sec/5/title}{subparagraph}
110 \NewStructureName{sec/5/number}
111 \AssignStructureRole{sec/5/number}{section-number}

```

1.5.4 Table of contents and similar

```

112 \NewStructureName{toc}
113 \AssignStructureRole{toc}{TOC}
114 \NewStructureName{toc/item}
115 \AssignStructureRole{toc/item}{TOCI}
116 \NewStructureName{toc/item/label}
117 \AssignStructureRole{toc/item/label}{Lb1}
118 \NewStructureName{toc/item/text}
119 \AssignStructureRole{toc/item/text}{Reference}
120 \NewStructureName{toc/item/page}
121 \AssignStructureRole{toc/item/page}{Reference}

```

1.5.5 Title

“title” in this context can be a structure that surrounds the *title block* (or a part from it) with author, title, subtitle, date and perhaps more. Or it can mean only the *title text* (plus perhaps the subtitle). In PDF 2.0 the tag `Title` is described as grouping and block type and it can not be used in a `P` tag, so it is better used for the *title block*. With PDF 1.7 `Title` is currently rolemapped to `P` and the code in `latex-lab-title` uses it for the *title text*. We provide therefore two symbolic names for the two concepts, which both map to `Title`.

```

122 \NewStructureName{title/block}
123 \AssignStructureRole{title/block}{Title}
124 \NewStructureName{title/text}
125 \AssignStructureRole{title/text}{Title}

```

1.5.6 Lists and blocks

`block/inner` is the generic named used for inner blocks in, e.g., the recipe `standard`.

```

126 \NewStructureName{block/inner}
127 \AssignStructureRole{block/inner}{Div}

```

UFi:This should be discussed.

```

128 \NewStructureName{block/itemize}
129 \AssignStructureRole{block/itemize}{itemize}
130 \NewStructureName{block/description}
131 \AssignStructureRole{block/description}{description}
132 \NewStructureName{block/enumerate}
133 \AssignStructureRole{block/enumerate}{enumerate}
134 \NewStructureName{block/list}
135 \AssignStructureRole{block/list}{list}
136 \NewStructureName{block/list/item}
137 \AssignStructureRole{block/list/item}{item}
138 \NewStructureName{block/list/label}
139 \AssignStructureRole{block/list/label}{itemlabel}
140 \NewStructureName{block/list/body}
141 \AssignStructureRole{block/list/body}{itembody}

142 \NewStructureName{block/quote}
143 \AssignStructureRole{block/quote}{quote}

```

decide on tagging for verse

For now we map verse to quote which means it will become BlockQuote in tagging.

```

144 \NewStructureName{block/verse}
145 \AssignStructureRole{block/verse}{quote}

146 \NewStructureName{block/quotation}
147 \AssignStructureRole{block/quotation}{quotation}
148 %
149 \NewStructureName{block/verbatim}
150 \AssignStructureRole{block/verbatim}{verbatim}
151 \NewStructureName{block/verbatim/codeline}
152 \AssignStructureRole{block/verbatim/codeline}{codeline}
153 \NewStructureName{block/verbatim/linenumber}
154 \AssignStructureRole{block/verbatim/linenumber}{Lb1}
155 %
156 \NewStructureName{block/theorem-like}
157 \AssignStructureRole{block/theorem-like}{theorem-like}
158 \NewStructureName{block/theorem-like/caption}
159 \AssignStructureRole{block/theorem-like/caption}{Caption}
160 \NewStructureName{block/theorem-like/label}
161 \AssignStructureRole{block/theorem-like/label}{Lb1}

```

For now we map proof to theorem-like:

```

162 \NewStructureName{block/proof}
163 \AssignStructureRole{block/proof}{theorem-like}

```

Bibliographies normally use internally lists and tag names should be adapted by changing the symbolic names of lists locally, so dedicated symbolic names are probably not needed.

```

164 %\NewStructureName{block/bib}
165 %\AssignStructureRole{block/bib}{L}
166 %\NewStructureName{block/bib/label}
167 %\AssignStructureRole{block/bib/label}{Lb1}
168 %\NewStructureName{block/bib/item}
169 %\AssignStructureRole{block/bib/item}{LI}
170 %\NewStructureName{block/bib/body}
171 %\AssignStructureRole{block/bib/body}{LBody}

```

1.5.7 L^AT_EX Code documentation

The class l3doc defines a number of structures for the document of L^AT_EX code. The exact tagging is currently experimental and it is unclear if symbolic names and tag names for this case should be setup here by default. Their setup is therefore done in latex-lab-l3doc-tagging.

1.5.8 References and Links

```
172 \NewStructureName{ref}
173 \AssignStructureRole{ref}{Reference}
174 \NewStructureName{link}
175 \AssignStructureRole{link}{Link}
176 \NewStructureName{cite}
177 \AssignStructureRole{cite}{Reference}
```

1.5.9 Tables

```
178 \NewStructureName{table}
179 \AssignStructureRole{table}{Table}
180 \NewStructureName{table/row}
181 \AssignStructureRole{table/row}{TR}
182 \NewStructureName{table/cell}
183 \AssignStructureRole{table/cell}{TD}
184 \NewStructureName{table/headercell}
185 \AssignStructureRole{table/headercell}{TH}
```

THead and TFoot are currently not used but are perhaps needed later

```
186 %\NewStructureName{tbl/table/head}
187 %\AssignStructureRole{tbl/table/head}{}
188 %\NewStructureName{tbl/table/foot}
189 %\AssignStructureRole{tbl/table/foot}{}%
```

Inside p-cells we have to set the para/semantic structure to Div as Part is not allowed as child of TD and as we can not flatten the paragraphs.

```
190 \NewStructureName{table/para/semantic}
191 \AssignStructureRole{table/para/semantic}{Div}
```

We need also names for layout tables

```
192 \NewStructureName{layouttable}
193 \AssignStructureRole{layouttable}{Div}
194 \NewStructureName{layouttable/row}
195 \AssignStructureRole{layouttable/row}{NonStruct}
196 \NewStructureName{layouttable/cell}
197 \AssignStructureRole{layouttable/cell}{\UseStructureName{para/textblock}}
198 \NewStructureName{layouttable/pcell}
199 \AssignStructureRole{layouttable/pcell}{NonStruct}
```

1.5.10 Floats

```
200 \NewStructureName{float/generic}
201 \AssignStructureRole{float/generic}{float}
202 \NewStructureName{float/figures}
203 \AssignStructureRole{float/figures}{figures}
```

```

204 \NewStructureName{float/tables}
205 \AssignStructureRole{float/tables}{tables}
206 \NewStructureName{float/figure}
207 \AssignStructureRole{float/figure}{float}
208 \NewStructureName{float/table}
209 \AssignStructureRole{float/table}{float}
210 \NewStructureName{float/figure/caption}
211 \AssignStructureRole{float/figure/caption}{Caption}
212 \NewStructureName{float/table/caption}
213 \AssignStructureRole{float/table/caption}{Caption}
214 \NewStructureName{float/figure/label}
215 \AssignStructureRole{float/figure/label}{Lbl}
216 \NewStructureName{float/table/label}
217 \AssignStructureRole{float/table/label}{Lbl}

```

1.5.11 Footnotes

```

218 \NewStructureName{fnote/mark}
219 \AssignStructureRole{fnote/mark}{footnotemark}
220 \NewStructureName{fnote/note}
221 \AssignStructureRole{fnote/note}{footnote}
222 \NewStructureName{fnote/note/label}
223 \AssignStructureRole{fnote/note/label}{footnotelabel}

```

1.5.12 Graphics

```

224 \NewStructureName{graphic}
225 \AssignStructureRole{graphic}{Figure}
226 \NewStructureName{graphic/symbol}
227 \AssignStructureRole{graphic/symbol}{Span}

```

1.5.13 Marginnotes

```

228 \NewStructureName{marginnote}
229 \AssignStructureRole{marginnote}{Aside}

```

1.5.14 Minipages and parboxes

```

230 \NewStructureName{minipage}
231 \AssignStructureRole{minipage}{Div}
232 \NewStructureName{parbox}
233 \AssignStructureRole{parbox}{Div}

```

1.5.15 Text

```

234 \NewStructureName{text/emph}
235 \AssignStructureRole{text/emph}{Em}
236 \NewStructureName{text/verb}
237 \AssignStructureRole{text/verb}{Code}

```


1.5.16 Math

Math uses internally MathML and this should not be remapped so we need only names for the outer Formula tag and perhaps for a label:

```
238 \NewStructureName{math/inline}
239 \AssignStructureRole{math/inline}{Formula}
240 \NewStructureName{math/display}
241 \AssignStructureRole{math/display}{Formula}
242 \NewStructureName{math/label}
243 \AssignStructureRole{math/label}{Lbl}
```

1.5.17 Index and glossaries

Tagging not decided yet.

1.6 Attributes

The latex-lab code sets up a number of kernel attributes. These are collected here:

```
244 \tagpdfsetup
245 {
```

1.7 Text attributes

These attributes are for super- and subscripts and are used in latex-lab-text:

```
246 role/new-attribute = {sub} {/O /Layout /TextPosition/Sub},
247 role/new-attribute = {sup} {/O /Layout /TextPosition/Sup},
```

These are for text alignment and they are used in latex-lab-block: role/new-attribute = justify /O /Layout /TextAlign/Justify, role/new-attribute = center /O /Layout /TextAlign/Center, role/new-attribute = raggedright /O /Layout /TextAlign/Start, role/new-attribute = raggedleft /O /Layout /TextAlign/End,

1.8 Math attributes

```
248 role/new-attribute = {inline} {/O /Layout /Placement/Inline},
249 role/new-attribute = {display} {/O /Layout /Placement/Block},
```

1.9 List attributes

```
250 role/new-attribute = {itemize}
251 {/O /List /ListNumbering/Unordered},
252 role/new-attribute = {enumerate}
253 {/O /List /ListNumbering/Ordered},
254 role/new-attribute = {description}
255 {/O /List /ListNumbering/Description},
```

Initially, we had /None for the basic list environment, but that is not allowed in PDF/UA-2 if the list contains any Lbl tags. So the list attribute uses Unordered too.

```
256 role/new-attribute = {list}{/O /List /ListNumbering/Unordered},
```

1.10 Table attributes

The code predeclares three attribute names for the scope of header, TH-col, TH-row, TH-both. The latex-lab-table code declares on the fly additional attributes colspan-*<number>* and rowspan-*<number>* with the content /O /Table /ColSpan *<number>* and /O /Table /RowSpan *<number>*

```

257   role/new-attribute = {TH-col}   {/O /Table /Scope /Column},
258   role/new-attribute = {TH-row}   {/O /Table /Scope /Row},
259   role/new-attribute = {TH-both}  {/O /Table /Scope /Both},
260   role/new-attribute = {ARIA-role-presentation}
261                                   {/O /ARIA-1.1/role (presentation)}

262 } %end tagpdfsetup

```

The three main table attributes are put directly into the class map TODO: either tagpdf should do that directly all roles or offer an interface:

```

263 \ExplSyntaxOn
264   \seq_gput_left:N\g__tag_attr_class_used_seq
265     {\pdf_name_from_unicode_e:n{TH-col}}
266   \seq_gput_left:N\g__tag_attr_class_used_seq
267     {\pdf_name_from_unicode_e:n{TH-row}}
268   \seq_gput_left:N\g__tag_attr_class_used_seq
269     {\pdf_name_from_unicode_e:n{TH-both}}
270 \ExplSyntaxOff

271 </package>

```