

omtext: Semantic Markup for Mathematical Text Fragments in L^AT_EX*

Michael Kohlhase
Jacobs University, Bremen
<http://kwarc.info/kohlhase>

November 16, 2015

Abstract

The **omtext** package is part of the sT_EX collection, a version of T_EX/L^AT_EX that allows to markup T_EX/L^AT_EX documents semantically without leaving the document format, essentially turning T_EX/L^AT_EX into a document format for mathematical knowledge management (MKM).

This package supplies an infrastructure for writing OMDoc text fragments in L^AT_EX.

*Version v1.1 (last revised 2015/11/04)

Contents

1	Introduction	3
2	The User Interface	3
2.1	Package Options	3
2.2	Mathematical Text	3
2.3	Phrase-Level Markup	3
2.4	Block-Level Markup	4
2.5	Index Markup	5
3	Limitations	6
4	Implementation	7
4.1	Package Options	7
4.2	Metadata	8
4.3	Mathematical Text	8
4.4	Phrase-level Markup	10
4.5	Block-Level Markup	11
4.6	Index Markup	13
4.7	L ^A T _E X Commands we interpret differently	15
4.8	Providing IDs for OMDoc Elements	16
4.9	Miscellaneous	18
4.10	Finale	18

1 Introduction

The `omtext` package supplies macros and environment that allow to mark up mathematical texts in \LaTeX , a version of \TeX / \LaTeX that allows to markup \TeX / \LaTeX documents semantically without leaving the document format, essentially turning \TeX / \LaTeX into a document format for mathematical knowledge management (MKM). The package supports direct translation to the OMDoc format [Kohlhase:OMDoc1.2]

2 The User Interface

2.1 Package Options

`showmeta` The `omtext` package takes a single option: `showmeta`. If this is set, then the metadata keys are shown (see [Kohlhase:metakeys:ctan] for details and customization options).

2.2 Mathematical Text

`omtext` The `omtext` environment is used for any text fragment that has a contribution to a text that needs to be marked up. It can have a title, which can be specified via the `title=` `title` key. Often it is also helpful to annotate the `type` key. The standard relations from rhetorical structure theory `abstract`, `introduction`, `conclusion`, `thesis`, `comment`, `antithesis`, `elaboration`, `motivation`, `evidence`, `transition`, `note`, `annote` are recommended as values. Note that some of them are unary relations like `introduction`, which calls for a target. In this case, a target using the `for` key should be specified. The `transition` relation is special in that it is binary (a “transition between two statements”), so additionally, a source should be specified using the `from` key.

Note that the values of the `title` and `type` keys are often displayed in the text.

`display=` This can be turned off by setting the `display` key to the value `flow`. Sometimes we want to specify that a text is a continuation of another, this can be done by giving the identifier of this in the `continues` key.

Finally, there is a set of keys that pertain to the mathematical formulae in the text. The `functions` key allows to specify a list of identifiers that are to be interpreted as functions in the generate content markup. The `theory` specifies a module (see [KohAmb:smmssl:svn]) that is to be pre-loaded in this one¹ Finally, `verbalizes=` `verbalizes` specifies a (more) formal statement (see [Kohlhase:smms:svn]) that this text verbalizes or paraphrases.²

2.3 Phrase-Level Markup

`\phrase` The `phrase` macro allows to mark up phrases with semantic information. It takes an optional `KeyVal` argument with the keys `verbalizes` and `type` as above and

¹EdNOTE: this is not implemented yet.

²EdNOTE: MK:specify the form of the reference.

`style` `style`, `class`, `index` that are disregarded in the L^AT_EX, but copied into the generated content markup.

`index` We use the `\nlex{⟨phrase⟩}` for marking up phrases that serve as natural language examples and `\nlcex{⟨phrase⟩}` for counter-examples (utterances that are not acceptable for some reason). In natural language examples, we sometimes use “co-reference markers” to specify the resolution of anaphora and the like. We

`\nlex`

`\nlcex`

`\coreft` use the `\coreft{⟨phrase⟩}{⟨mark⟩}` to mark up the “target” of a co-reference and

`\corefs` analogously `\corefs` for coreference source – e.g. for an anaphoric reference. The usage is the following:

```
\nlex{If \coreft{a farmer}1 owns \coreft{a donkey}2,
      \corefs{he}2 beats \corefs{it}2.}
```

is formatted to

If a farmer¹ owns a donkey², he₂ beats it₂.

`\sinlinequote` The `\sinlinequote` macro allows to mark up quotes inline and attribute them. The quote itself is given as the argument, possibly preceded by the a specification of the source in a an optional argument. For instance, we would quote Hamlet with

```
\sinlinequote[Hamlet, \cite{Shak:1603:Hamlet}]{To be or not to be}
```

which would appear as “*To be or not to be*” Hamlet, (Shakespeare 1603) in the text. The style in which inline quotations appear in the text can be adapted

`\@sinlinequote` by specializing the macros `\@sinlinequote` — for quotations without source and

`\@@sinlinequote` `\@@sinlinequote` — for quotations with source.

2.4 Block-Level Markup

`sblockquote` The `sblockquote` environment is the big brother of the `\sinlinequote` macro. It also takes an optional argument to specify the source. Here the four internal

`\begin@sblockquote` macros `\begin@sblockquote` to `\end@@sblockquote` are used for styling and can

`\end@@sblockquote` be adapted by package integrators. Here a quote of Hamlet would marked up as

```
\begin{sblockquote}[Hamlet, \cite{Shak:1603:Hamlet}]\obeylines
  To be, or not to be: that is the question:
  Whether 'tis nobler in the mind to suffer
\end{sblockquote}
```

and would render as

To be, or not to be: that is the question:

Whether 'tis nobler in the mind to suffer

Hamlet, (Shakespeare 1603)

`\lec` The `\lec` macro takes one argument and sets it as a comment at the end of the line, making sure that if the content is too long it is pushed into a new line. We use it internally for placing the of source of the `sblockquote` environment above. The

`\@@lec` actual appearance of the line end comment is determined by the `\@@lec` macro, which can be customized in the document class.

2.5 Index Markup

The `omtext` package provides some extensions for the well-known indexing macros of L^AT_EX. The main reason for introducing these macros is that index markup in OMDoc wraps the indexed terms rather than just marking the spot for cross-referencing. Furthermore the index commands only indexes words unless the `noindex` option is set in the `\usepackage`. The `omtext` package and class make the usual `\index` macro undefined³.

`\indextoo` The `\indextoo` macro renders a word and marks it for the index. Sometimes, we want to index a slightly different form of the word, e.g. for non-standard plurals: while `\indextoo{word}s` works fine, we cannot use this for the word “datum”, which has the plural “data”. For this we have the macro

`\indexalt` `\indexalt`, which takes another argument for the displayed text, allowing us to use `\indexalt{data}{datum}`, which prints “data” but puts “datum” into the index.

The second set of macros adds an infrastructure for two-word compounds. Take for instance the compound “OMDoc document”, which we usually want to add into the index under “OMDoc” and “document”. `\twintoo{OMDoc}{document}` is a variant of `\indextoo` that will do just this. Again, we have a version that prints a variant: This is useful for situations like this the one in Figure 1:

We call `group \twinalt{Abelian}{Abelian}{group}`, iff `\ldots`

will result in the following

We call group Abelian, iff ...

and put “Abelian Group” into the index.

Example 1: Index markup

`\atwintoo` The third set of macros does the same for two-word compounds with adjectives, e.g. “wonderful OMDoc document”. `\atwin{wonderful}{OMdoc}{document}` will make the necessary index entries under “wonderful” and “document”. Again,

`\atwinalt` we have a variant `\atwinalt` whose first argument is the alternative text.

All index macros take an optional first argument that is used for ordering the respective entries in the index.

³EDNOTE: implement this and issue the respective error message

3 Limitations

In this section we document known limitations. If you want to help alleviate them, please feel free to contact the package author. Some of them are currently discussed in the `sTeX` GitHub repository [`sTeX:github:on`].

1. none reported yet

4 Implementation

The `omtext` package generates two files: the \LaTeX package (all the code between `<*package>` and `</package>`) and the \LaTeX XML bindings (between `<*ltxml>` and `</ltxml>`). We keep the corresponding code fragments together, since the documentation applies to both of them and to prevent them from getting out of sync.

The initial setup for \LaTeX XML:

```
1 <*ltxml>
2 package LaTeXXML::Package::Pool;
3 use strict;
4 use LaTeXXML::Package;
5 use LaTeXXML::Util::Pathname;
6 </ltxml>
```

4.1 Package Options

We declare some switches which will modify the behavior according to the package options. Generally, an option `xxx` will just set the appropriate switches to true (otherwise they stay false).⁴

```
7 <*package>
8 \newif\if@omtext@mh@\@omtext@mh@false
9 \DeclareOption{mh}{\@omtext@mh@true}
10 \newif\ifindex\indextrue
11 \DeclareOption{noindex}{\indexfalse}
12 \DeclareOption*{\PassOptionsToPackage{\CurrentOption}{modules}}
13 \ProcessOptions
14 \ifindex\makeindex\fi
15 </package>
16 <*ltxml>
17 DefConditional('\if@omtext');
18 DeclareOption('mh', sub {Digest(T_CS('\@omtexttrue'))}; });
19 DeclareOption('noindex', '');
20 DeclareOption(undef, sub {PassOptions('modules', 'sty', ToString(Digest(T_CS('\CurrentOption'))));});
21 ProcessOptions();
22 </ltxml>

23 <*package>
24 \if@omtext@mh\RequirePackage{omtext-mh}\fi
25 \RequirePackage{xspace}
26 \RequirePackage{modules}
27 \RequirePackage{comment}
28 \RequirePackage{mdframed}
29 \RequirePackage{latexsym}
30 </package>
31 <*ltxml>
32 if(IfCondition(T_CS('\if@omtext')) {RequirePackage('omtext-mh');}
33 RequirePackage('xspace');
```

⁴EdNOTE: need an implementation for \LaTeX XML

```

34 RequirePackage('modules');
35 RequirePackage('lxRDFa');
36 RequirePackage('latexsym');
37 </ltxml>

```

4.2 Metadata

All the OMDoc elements allow to specify metadata in them, which is modeled by the `omdoc:metadata` element. Since the content of this element is precisely controlled by the Schema, we can afford to auto-open and auto-close it. Thus metadata elements from various sources will just be included into one `omdoc:metadata` element, even if they are supplied by different \TeX bindings. Also we add numbering and location facilities.

```

38 <*ltxml>
39 Tag('omdoc:metadata', afterOpen=>\&numberIt, afterClose=>\&locateIt, autoClose=>1, autoOpen=>1);
40 </ltxml>

```

the `itemize`, `description`, and `enumerate` environments generate `omdoc:li`, `omdoc:di` with `autoclose` inside a CMP. This behavior will be overwritten later, so we remember that we are in a CMP by assigning `_LastSeenCMP`.

```

41 <*ltxml>
42 Tag('omdoc:CMP', afterOpen => sub {AssignValue('_LastSeenCMP', $_[1], 'global');return;});#$
43 </ltxml>

```

the `itemize`, `description`, and `enumerate` environments originally introduced in the `omtext` package do double duty in OMDoc, outside a CMP they are transformed into a `<omgroup layout='itemizedescriptionenumerate'>`, where the text after the macros `\item` come to be the children. If that is only text, then it is enclosed in an `<omtext><CMP>`, otherwise it is left as it is. The optional argument of the `\item` is transformed into the `<metadata><dc:title>` of the generated `\item` element.

```

44 <*ltxml>
45 DefParameterType('IfBeginFollows', sub {
46     my ($gullet) = @_;
47     $gullet->skipSpaces;
48     my $next = $gullet->readToken;
49     $gullet->unread($next);
50     $next = ToString($next);
51     #Hm, falling back to regexp handling, the $gullet->ifNext approach didn't work
52     return 1 unless ($next=~/\^\backslashbegin/);
53     return;
54 },
55 reversion=>'', optional=>1);
56 </ltxml>

```

4.3 Mathematical Text

We define the actions that are undertaken, when the keys are encountered. The first set just records metadata; this is very simple via the `\addmetakey` infrastruc-

ture [**Kohlhase:metakeys:ctan**]. Note that we allow math in the `title` field, so we do not declare it to be `Semiverbatim` (indeed not at all, which allows it by default).

```

57 <*package>
58 \srefaddidkey{omtext}
59 \addmetakey[] {omtext} {functions}
60 \addmetakey* {omtext} {display}
61 \addmetakey {omtext} {for}
62 \addmetakey {omtext} {from}
63 \addmetakey {omtext} {type}
64 \addmetakey* {omtext} {title}
65 \addmetakey* {omtext} {start}
66 \addmetakey {omtext} {theory}
67 \addmetakey {omtext} {continues}
68 \addmetakey {omtext} {verbalizes}
69 \addmetakey {omtext} {subject}
70 </package>
71 <*ltxml>
72 DefKeyVal('omtext','functions','CommaList');
73 DefKeyVal('omtext','display','Semiverbatim');
74 DefKeyVal('omtext','for','Semiverbatim');
75 DefKeyVal('omtext','from','Semiverbatim');
76 DefKeyVal('omtext','type','Semiverbatim');
77 DefKeyVal('omtext','title','Plain'); #Math mode in titles.
78 DefKeyVal('omtext','start','Plain'); #Math mode in start phrases
79 DefKeyVal('omtext','theory','Semiverbatim');
80 DefKeyVal('omtext','continues','Semiverbatim');
81 DefKeyVal('omtext','verbalizes','Semiverbatim');
82 </ltxml>

```

The next keys handle module loading (see [**KohAmb:smmssl:ctan**]).

```

83 % \ednote{MK: need to implement these in LaTeXML, I wonder whether there is a general
84 % mechanism like numberit.}\ednote{MK: this needs to be rethought in the light of
85 % |\usemodule|. It is probably obsolete. Is this used? Is this documented?}
86 <*package>
87 \define@key{omtext}{require}{\requiremodules{#1}{sms}}
88 \define@key{omtext}{module}{\message{module: #1}\importmodule{#1}\def\omtext@theory{#1}}

```

```

\st@flow We define this macro, so that we can test whether the display key has the value
flow
89 \def\st@flow{flow}

```

We define a switch that allows us to see whether we are inside an `omtext` environment or a statement. It will be used to give better error messages for inline statements.

```

90 \newif\if@in@omtext\@in@omtextfalse

```

```

omtext The omtext environment is different, it does not have a keyword that marks it.
Instead, it can have a title, which is used in a similar way. We redefine the \lec

```

macro so the trailing `\par` does not get into the way.

```

91 \def\omtext@pre@skip{\smallskip}
92 \def\omtext@post@skip{}
93 \providecommand{\stDMemph}[1]{\textbf{#1}}
94 \newenvironment{omtext}[1][\in@omtexttrue%
95 \bgroup\metasetkeys{omtext}{#1}\sref@label{id}{this paragraph}%
96 \def\lec##1{\@lec{##1}}%
97 \ifx\omtext@display\st@flow\else\omtext@pre@skip\par\noindent%
98 \ifx\omtext@title\@empty%
99 \ifx\omtext@start\@empty\else\stDMemph{\omtext@start}\xspace\fi%
100 \else\stDMemph{\omtext@title}:\xspace%
101 \ifx\omtext@start\@empty\else\omtext@start\xspace\fi%
102 \fi% omtext@title empty
103 \fi% omtext@display=flow
104 \ignorespaces}
105 {\egroup\omtext@post@skip\in@omtextfalse}
106 \end{package}
107 \let\omtext\@empty
108 DefEnvironment('omtext' OptionalKeyVals:omtext',
109   "<omdoc:omtext "
110     . "&GetKeyVal(#1,'id')(xml:id='&GetKeyVal(#1,'id'))() "
111     . "&GetKeyVal(#1,'type')(type='&GetKeyVal(#1,'type'))() "
112     . "&GetKeyVal(#1,'for')(for='&GetKeyVal(#1,'for'))() "
113     . "&GetKeyVal(#1,'from')(from='&GetKeyVal(#1,'from'))()>"
114     . "&GetKeyVal(#1,'title')(<dc:title>&GetKeyVal(#1,'title')</dc:title>())"
115     . "&GetKeyVal(#1,'start')(<ltx:text class='startemph'>&GetKeyVal(#1,'start')</ltx:text>)"
116     . "#body"
117     . "</omdoc:omtext>");
118 \end{ltxml}

```

4.4 Phrase-level Markup

`\phrase` For the moment, we do disregard the most of the keys

```

119 \begin{package}
120 \srefaddidkey{phrase}
121 \addmetakey{phrase}{style}
122 \addmetakey{phrase}{class}
123 \addmetakey{phrase}{index}
124 \addmetakey{phrase}{verbalizes}
125 \addmetakey{phrase}{type}
126 \addmetakey{phrase}{only}
127 \newcommand\phrase[2][\metasetkeys{phrase}{#1}%
128 \ifx\prhase@only\@empty\only<\phrase@only>{#2}\else #2\fi}
129 \end{package}
130 \let\phrase\@empty
131 DefKeyVal('phrase','id','Semiverbatim');
132 DefKeyVal('phrase','style','Semiverbatim');
133 DefKeyVal('phrase','class','Semiverbatim');
134 DefKeyVal('phrase','index','Semiverbatim');

```

```

135 DefKeyVal('phrase','verbalizes','Semiverbatim');
136 DefKeyVal('phrase','type','Semiverbatim');
137 DefKeyVal('phrase','only','Semiverbatim');
138 DefConstructor('\phrase OptionalKeyVals:phrase {}',
139     "<ltx:text %&GetKeyVals(#1) ?&GetKeyVal(#1,'only')(rel='beamer:only' content='&GetKeyVal
140 </ltxml>

\coref*
141 <*package>
142 \providecommand\textsubscript[1]{\ensuremath_{\{#1\}}}
143 \newcommand\corefs[2]{#1\textsubscript{#2}}
144 \newcommand\coreft[2]{#1\textsuperscript{#2}}
145 </package>
146 <*ltxml>
147 DefConstructor('\corefs{}{}',
148     "<ltx:text class='coref-source' stex:index='#2'>#1</ltx:text>");
149 DefConstructor('\coreft{}{}',
150     "<ltx:text class='coref-target' stex:index='#2'>#1</ltx:text>");
151 </ltxml>

\n*lex
152 <*package>
153 \newcommand\nlex[1]{\green{\sl{#1}}}
154 \newcommand\nlcex[1]{*\green{\sl{#1}}}
155 </package>
156 <*ltxml>
157 DefConstructor('\nlex{}',"<ltx:text class='nlex'>#1</ltx:text>");
158 DefConstructor('\nlcex{}',"<ltx:text class='nlcex'>#1</ltx:text>");
159 </ltxml>

sinlinequote
160 <*package>
161 \def\@sinlinequote#1{'\sl{#1}}'
162 \def\@@sinlinequote#1#2{\@sinlinequote{#2}~#1}
163 \newcommand\sinlinequote[2]{}
164 {\def\@opt{#1}\ifx\@opt\@empty\@sinlinequote{#2}\else\@@sinlinequote\@opt{#2}\fi}
165 </package>
166 <*ltxml>
167 DefConstructor('\sinlinequote [] {}',
168     "<ltx:quote type='inlinequote'>"
169     . "?#1(<dc:source>#1</dc:source>\n)()"
170     . "#2"
171     . "</ltx:quote>");
172 </ltxml>

```

4.5 Block-Level Markup

sbloquote

```

173 <*package>

```

```

174 \def\begin@sblockquote{\begin{quote}\sl}
175 \def\end@sblockquote{\end{quote}}
176 \def\begin@@sblockquote#1{\begin@sblockquote}
177 \def\end@@sblockquote#1{\def\@@lec##1{{\rm ##1}}\@lec{#1}\end@sblockquote}
178 \newenvironment{sblockquote}[1]{}
179 {\def\@opt{#1}\ifx\@opt\@empty\begin@sblockquote\else\begin@@sblockquote\@opt\fi}
180 {\ifx\@opt\@empty\end@sblockquote\else\end@@sblockquote\@opt\fi}
181 \</package>
182 \<*txml>
183 DefEnvironment('sblockquote' [],',
184 "<ltx:quote>?#1(<ltx:note role='source'>#1</ltx:note>())#body</ltx:quote>");
185 \</txml>

```

sboxquote

```

186 \<*package>
187 \newenvironment{sboxquote}[1]{}
188 {\def\@@src{#1}\begin{mdframed}[leftmargin=.5cm,rightmargin=.5cm]}
189 {\@lec{\rm\@@src}\end{mdframed}}
190 \</package>
191 \<*txml>
192 DefEnvironment('sboxquote' [],',
193 "<ltx:quote class='boxed'>?#1(<ltx:note role='source'>#1</ltx:note>())#body</ltx:quote>");
194 \</txml>

```

The line end comment macro makes sure that it will not be forced on the next line unless necessary.

\lec The actual appearance of the line end comment is determined by the \@@lec macro, which can be customized in the document class. The basic one here is provided so that it is not missing.

```

195 \<*package>
196 \providecommand{\@@lec}[1]{(#{1})}
197 \def\@lec#1{\strut\hfil\strut\null\nobreak\hfill\@@lec{#1}}
198 \def\lec#1{\@lec{#1}\par}
199 \</package>
200 \<*txml>
201 DefConstructor('\lec{}',
202 "\n<omdoc:note type='line-end-comment'>#1</omdoc:note>");
203 \</txml>

```

\my*graphics We set up a special treatment for including graphics to respect the intended OM- Doc document structure. The main work is done in the transformation stylesheet though.

```

204 \<txml>RawTeX('
205 \<*txml | package>
206 \newcommand\mygraphics[2][]{\includegraphics[#1]{#2}}
207 \newcommand\mycgraphics[2][]{\begin{center}\mygraphics[#1]{#2}\end{center}}
208 \newcommand\mybgraphics[2][]{\fbox{\mygraphics[#1]{#2}}}
209 \newcommand\mycbgraphics[2][]{\begin{center}\fbox{\mygraphics[#1]{#2}\end{center}}}

```

```

210 </ltxml | package>
211 <ltxml>' );

```

4.6 Index Markup

`\omdoc@index` this is the main internal indexing command. It makes sure that the modules necessary for interpreting the math in the index entries are loaded. If the `loadmodules` key is given, we import the module we are in otherwise all the currently imported modules. We do not have to require the module files, since the index is at the end of the document. If the `at` key is given, then we use that for sorting in the index.

```

212 <*package>
213 \addmetakey{omdoc@index}{at}
214 \addmetakey[false]{omdoc@index}{loadmodules}[true]
215 \newcommand\omdoc@index[2] [] {\ifindex%
216 \metasetkeys{omdoc@index}{#1}%
217 \@bsphack\begingroup\@sanitize%
218 \ifx\omdoc@index@loadmodules\@true%
219 \protected@write\@indexfile{}\string\indexentry%
220 {\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi%
221 {\string\importmodules{\@ifundefined{mod@id}\imported@modules\mod@id}%
222 #2}}{\thepage}}%
223 \else%
224 \protected@write\@indexfile{}\string\indexentry%
225 {\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi#2}{\thepage}}%
226 \fi% loadmodules
227 \endgroup\@esphack\fi}%ifindex

```

Now, we make two interface macros that make use of this:

```

\indexalt
228 \newcommand\indexalt[3] [] [{#2}\omdoc@index[#1]{#3}} % word in text and index
229 </package>
230 <*ltxml>
231 DefConstructor(' \indexalt [] {} {} ',
232     "<omdoc:idx>"
233     . " <omdoc:idt>#2</omdoc:idt>"
234     . " <omdoc:ide ?#1(sort-by=' #1') ()>"
235     . " <omdoc:idp>#3</omdoc:idp>"
236     . " </omdoc:ide>"
237     . "</omdoc:idx>");
238 </ltxml>

\indextoo
239 <*package>
240 \newcommand\indextoo[2] [] [{#2}\omdoc@index[#1]{#2}} % word in text and index
241 </package>
242 <*ltxml>
243 DefConstructor(' \indextoo [] {} ',
244     "<omdoc:idx>"

```

```

245 . " <omdoc:idt>#2</omdoc:idt>"
246 . " <omdoc:ide ?#1(sort-by='#1')()>"
247 . " <omdoc:idp>#2</omdoc:idp>"
248 . " </omdoc:ide>"
249 . " </omdoc:idx>");
250 </ltxml>

```

`\@twin` this puts two-compound words into the index in various permutations

```

251 <*package>
252 \newcommand\@twin[3] [] {\omdoc@index[#1]{#2!#3}\omdoc@index[#1]{#3!#2}}

```

And again we have two interface macros building on this

`\twinalt`

```

253 \newcommand\twinalt[4] [] {#2\@twin[#1]{#3}{#4}}
254 </package>
255 <*ltxml>
256 DefConstructor('\twinalt[]{}{}',
257     "<omdoc:idx>"
258     . " <omdoc:idt>#2</omdoc:idt>"
259     . " <omdoc:ide ?#1(sort-by='#1')()>"
260     . " <omdoc:idp>#2</omdoc:idp>"
261     . " <omdoc:idp>#3</omdoc:idp>"
262     . " </omdoc:ide>"
263     . " </omdoc:idx>");
264 </ltxml>

```

`\twintoo`

```

265 <*package>
266 \newcommand\twintoo[3] [] {{#2 #3}\@twin[#1]{#2}{#3}} % and use the word compound too
267 </package>
268 <*ltxml>
269 DefConstructor('\twintoo[]{}{}',
270     "<omdoc:idx>"
271     . " <omdoc:idt>#2 #3</omdoc:idt>"
272     . " <omdoc:ide ?#1(sort-by='#1')()>"
273     . " <omdoc:idp>#2</omdoc:idp>"
274     . " <omdoc:idp>#3</omdoc:idp>"
275     . " </omdoc:ide>"
276     . " </omdoc:idx>");
277 </ltxml>

```

EdN:5

`\@atwin` this puts adjectivized two-compound words into the index in various permutations⁵

```

278 <*package>
279 \newcommand\@atwin[4] [] {\omdoc@index[#1]{#2!#3!#4}\omdoc@index[#1]{#3!#2 (#4)}}

```

and the two interface macros for this case:

⁵EDNOTE: what to do with the optional argument here and below?

```

\@atwinalt
280 \newcommand\atwinalt[5][{}]{#2\@atwin[#1]{#3}{#4}{#4}}
281 \</package>
282 \<*ltxml>
283 DefConstructor('atwinalt[]{}{}{}{}',
284     "<omdoc:idx>"
285     . "<omdoc:idt>#2</omdoc:idt>"
286     . "<omdoc:ide ?#1(sort-by='#1')()>"
287     . "<omdoc:idp>#2</omdoc:idp>"
288     . "<omdoc:idp>#3</omdoc:idp>"
289     . "<omdoc:idp>#4</omdoc:idp>"
290     . "</omdoc:ide>"
291     . "</omdoc:idx>");
292 \</ltxml>

\atwintoo
293 \<*package>
294 \newcommand\atwintoo[4][{}]{#2 #3 #4\@atwin[#1]{#2}{#3}{#4}} % and use it too
295 \</package>
296 \<*ltxml>
297 DefConstructor('atwintoo[]{}{}{}{}',
298     "<omdoc:idx>"
299     . "<omdoc:idt>#2 #3</omdoc:idt>"
300     . "<omdoc:ide ?#1(sort-by='#1')()>"
301     . "<omdoc:idp>#2</omdoc:idp>"
302     . "<omdoc:idp>#3</omdoc:idp>"
303     . "<omdoc:idp>#4</omdoc:idp>"
304     . "</omdoc:ide>"
305     . "</omdoc:idx>");
306 \</ltxml>

```

4.7 L^AT_EX Commands we interpret differently

The first think we have to take care of are the paragraphs, we want to generate OMDoc that uses the `ltx:p` element for paragraphs inside CMPs. For that we have modified the DTD only to allowed `ltx:p` elements in `omdoc:CMP` (in particular no text). Then we instruct the `\par` macro to close a `ltx:p` element if possible. The next `ltx:p` element is then opened automatically, since we make `ltx:p` and `omdoc:CMP` autoclose and autoopen.

```

307 \<*ltxml>
308 Tag('omdoc:CMP', autoClose=>1, autoOpen=>1);
309 Tag('omdoc:omtext', autoClose=>1, autoOpen=>1);
310 Tag('ltx:p', autoClose=>1, autoOpen=>1);
311 \</ltxml>

```

the rest of the reinterpretations is quite simple, we either disregard presentational markup or we re-interpret it in terms of OMDoc.⁶

⁶EDNOTE: MK: we should probably let LaTeXML deal with these and allow more text in the `omdoc+ltxml.xsl`

```

312 <*package>
313 \def\omspace#1{\hspace*{#1}}
314 </package>
315 <*ltxml>
316 DefConstructor('\footnote[]{}',
317     "<omdoc:note type='foot' ?#1(mark='#1')>#2</omdoc:note>");
318 DefConstructor('\footnotemark[]', "");
319 DefConstructor('\footnotetext[]{}',
320     "<omdoc:note class='foot' ?#1(mark='#1')>#2</omdoc:note>");
321 </ltxml>

```

4.8 Providing IDs for OMDoc Elements

To provide default identifiers, we tag all OMDoc elements that allow `xml:id` attributes by executing the `numberIt` procedure below. Furthermore, we use the `locateIt` procedure to give source links.

```

322 <*ltxml>
323 Tag('omdoc:omtext', afterOpen=>\&numberIt, afterClose=>\&locateIt);
324 Tag('omdoc:omgroup', afterOpen=>\&numberIt, afterClose=>\&locateIt);
325 Tag('omdoc:CMP', afterOpen=>\&numberIt, afterClose=>\&locateIt);
326 Tag('omdoc:idx', afterOpen=>\&numberIt, afterClose=>\&locateIt);
327 Tag('omdoc:ide', afterOpen=>\&numberIt, afterClose=>\&locateIt);
328 Tag('omdoc:idt', afterOpen=>\&numberIt, afterClose=>\&locateIt);
329 Tag('omdoc:note', afterOpen=>\&numberIt, afterClose=>\&locateIt);
330 Tag('omdoc:metadata', afterOpen=>\&numberIt, afterClose=>\&locateIt);
331 Tag('omdoc:meta', afterOpen=>\&numberIt, afterClose=>\&locateIt);
332 Tag('omdoc:resource', afterOpen=>\&numberIt, afterClose=>\&locateIt);
333 Tag('omdoc:recurse', afterOpen=>\&numberIt, afterClose=>\&locateIt);
334 Tag('omdoc:imports', afterOpen=>\&numberIt, afterClose=>\&locateIt);
335 Tag('omdoc:theory', afterOpen=>\&numberIt, afterClose=>\&locateIt);
336 Tag('omdoc:ignore', afterOpen=>\&numberIt, afterClose=>\&locateIt);
337 Tag('omdoc:ref', afterOpen=>\&numberIt, afterClose=>\&locateIt);
338 </ltxml>

```

We also have to number some L^AT_EX XML tags, so that we do not get into trouble with the OMDoc tags inside them.

```

339 <*ltxml>
340 Tag('ltx:p', afterOpen=>\&numberIt, afterClose=>\&locateIt);
341 Tag('ltx:tabular', afterOpen=>\&numberIt, afterClose=>\&locateIt);
342 Tag('ltx:thead', afterOpen=>\&numberIt, afterClose=>\&locateIt);
343 Tag('ltx:td', afterOpen=>\&numberIt, afterClose=>\&locateIt);
344 Tag('ltx:tr', afterOpen=>\&numberIt, afterClose=>\&locateIt);
345 Tag('ltx:caption', afterOpen=>\&numberIt, afterClose=>\&locateIt);
346 Tag('ltx:Math', afterOpen=>\&numberIt, afterClose=>\&locateIt);
347 </ltxml>

```

The `numberIt` procedure gets the prefix from first parent with an `xml:id` attribute and then extends it with a label that reflects the number of preceding siblings, provided that there is not already an identifier. Additionally, it estimates an XPointer

position in the original document of the command sequence which produced the tag. The `locateIt` subroutine is a sibling of `numberIt` as it is required as an `afterClose` handle for tags produced by L^AT_EX environments, as opposed to commands. `locateIt` estimates an XPointer end position of the L^AT_EX environment, allowing to meaningfully locate the entire environment at the source.

```

348 <*!xml>
349 sub numberIt {
350   my($document,$node,$whatsit)=@_;
351   my(@parents)=$document->findnodes('ancestor::*[@xml:id]', $node);
352   my $prefix= (@parents ? $parents[$#parents]->getAttribute('xml:id')." " : '');
353   my(@siblings)=$document->findnodes('preceding-sibling::*[@xml:id]', $node);
354   my $n = scalar(@siblings)+1;
355   my $id = ($node -> getAttribute('xml:id'));
356   my $localname = $node->localname;
357   $node->setAttribute('xml:id'=>$prefix.$localname.$n) unless $id;
358   my $about = $node -> getAttribute('about');
359   $node->setAttribute('about'=>'#'.$node->getAttribute('xml:id')) unless $about;
360   #Also, provide locators:
361   my $locator = $whatsit && $whatsit->getProperty('locator');
362   #Need to inherit locators if missing:
363   $locator = (@parents ? $parents[$#parents]->getAttribute('stex:srcref') : '') unless $locator;
364   if ($locator) {
365     # There is a BUG with namespace declarations (or am I using the API wrongly??) which
366     # does not recognize the stex namespace. Hence, I need to redeclare it...
367     my $parent=$document->getNode;
368     if(! defined $parent->lookupNamespacePrefix("http://kwarc.info/ns/sTeX"))
369       { # namespace not already declared?
370         $document->getDocument->documentElement->setNamespace("http://kwarc.info/ns/sTeX","stex");
371       }
372     $node->setAttribute('stex:srcref'=>$locator);
373   }return;}
374
375 sub locateIt {
376   my($document,$node,$whatsit)=@_;
377   #Estimate trailer and locator:
378   my $locator = $node->getAttribute('stex:srcref');
379   return unless $locator; # Nothing to do here...
380   my $trailer = $whatsit && $whatsit->getProperty('trailer');
381   $trailer = $trailer->getLocator if $trailer;
382   $trailer = $locator unless $trailer; # bootstrap
383   # TODO: Both should be local, or both remote, any mixture or undefinedness will produce garba
384   my $file_path = LookupValue('SOURCEFILE');
385   my $baselocal = LookupValue('BASELOCAL');
386   # Hmm, we only care about relative paths, so let's just do a URL->pathname map
387   $file_path=~s/^\w+:\:\/\/ if $file_path;
388   $baselocal=~s/^\w+:\:\/\/ if $baselocal;
389   if ($file_path && $baselocal && ($locator =~ s/^(\[^\#\]+\)\#\/\#\/)) {
390     my $relative_path = pathname_relative($file_path,$baselocal);
391     $locator = $relative_path.$locator;

```

```

392 }
393 if ($locator =~ /^(.+from=\d+;\d+)/) {
394     my $from = $1;
395     if ($trailer =~ /(,to=\d+;\d+.)$/) {
396         my $to = $1;
397         $locator = $from.$to;
398     } else { Error("stex","locator",undef, "Trailer is garbled, expect nonsense in stex:srcref
399 } else { Error("stex","locator",undef, "Locator \"\$locator\" is garbled, expect nonsense in s
400 my $parent = $document->getNode;
401 if(! defined $parent->lookupNamespacePrefix("http://kwarc.info/ns/sTeX"))
402     { # namespace not already declared?
403         $document->getDocument->documentElement->setNamespace("http://kwarc.info/ns/sTeX","stex",
404     }
405     $node->setAttribute('stex:srcref' => $locator);
406     return;
407 }
408 </ltxml>#&

```

4.9 Miscellaneous

Some shortcuts that use math symbols but are not mathematical at all; in particular, they should not be translated by L^AT_EX_ML.

```

409 <*package>
410 \newcommand\hateq{\ensuremath{\hat{=}}\xspace}
411 \newcommand\hatequiv{\ensuremath{\hat{\equiv}}\xspace}
412 \@ifundefined{ergo}%
413 {\newcommand\ergo{\ensuremath{\leadsto}\xspace}}%
414 {\renewcommand\ergo{\ensuremath{\leadsto}\xspace}}%
415 \newcommand{\reflect@squig}[2]{\reflectbox{$\m@th#1\rightsquigarrow$}}%
416 \newcommand\ogre{\ensuremath{\mathrel{\mathpalette\reflect@squig\relax}}\xspace}%
417 </package>
418 <*ltxml>
419 DefMacro('\'hateq','\'@hateq\xspace');
420 DefConstructor('\'@hateq','\'x{2259}');
421 DefMacro('\'hatequiv','\'@hatequiv\xspace');
422 DefConstructor('\'@hatequiv','\'x{2A6F}');
423 DefMacro('\'ergo','\'@ergo\xspace');
424 DefConstructor('\'@ergo','\'x{219D}');
425 DefMacro('\'ogre','\'@ogre\xspace');
426 DefConstructor('\'@ogre','\'x{2B3F}');
427 </ltxml>

```

4.10 Finale

We need to terminate the file with a success mark for perl.

```

428 <ltxml>1;

```