

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

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

February 8, 2016

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 2015/11/22)

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	Declarations (under development)	4
2.5	Block-Level Markup	5
2.6	Index Markup	5
3	Limitations	6
4	Implementation	7
4.1	Package Options	7
4.2	Mathematical Text	7
4.3	Phrase-level Markup	8
4.4	Declarations (under development)	9
4.5	Block-Level Markup	9
4.6	Index Markup	10
4.7	Miscellaneous	11

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 [Koh06]

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 [Koh15a] 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=` 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`, `annotate` 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. 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 [KGA15a]) that is to be pre-loaded in this one¹ Finally, `verbalizes=` specifies a (more) formal statement (see [Koh15b]) 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 `style`, `class`, `index` that are disregarded in the \LaTeX , but copied into the gen-

¹EDNOTE: this is not implemented yet.

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

erated content markup.

`\nlex` We use the `\nlex{<phrase>}` for marking up phrases that serve as natural
`\nlcex` 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
`\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 Declarations (under development)

Declarations are special phrases that carry a lot of meaning in mathematics: they
introduce and further specify the indentifiers available in formulae. They are

`\vdec` marked up via the `\vdec` macro. Inside a declaration we can use the `\vids` macro
`\vids` to mark up the variable names and `\vrest` to mark up the restrictions. In the
`\vrest` simplest case we have a single variable as in “...for all u ”, which we mark up as.

... for all `\vdec{\vids[u]{u}}`.

A more complex example has multiple identifiers embedded in a restriction, as in
“Let $x, y, z \in \mathbb{R}$, such that $x + 2y = z$, then ...”, which we mark up as

```
Let \vdec{\vids[x,y,z]{\minset{x,y,z}\Reals}},
    such that \vrest{\$x+2y=z\$}, then \ldots''
```

3456

³EDNOTE: explain and make better examples

⁴EDNOTE: unify this with the vardef macros from (currently) the modules package.

⁵EDNOTE: talk with Frederic about this see what other examples there are.

⁶EDNOTE: how do we identify the variables in complex restriction patterns. maybe with LMXref,
which we should reinstate for this.

2.5 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 macros `\begin@sblockquote` to `\end@@sblockquote` are used for styling and can 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 actual appearance of the line end comment is determined by the `\@@lec` macro, which can be customized in the document class.

2.6 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 undefined⁷.

`\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`, 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:

The third set of macros does the same for two-word compounds with adjectives, e.g. “wonderful OMDoc document”. `\atwin{wonderful}{OMDoc}{document}`

⁷EdNOTE: implement this and issue the respective error message

```
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

`\atwinalt` will make the necessary index entries under “wonderful” and “document”. Again, 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.

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].

1. none reported yet

4 Implementation

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).⁸

```
1 <*package>
2 \newif\if@omtext@mh@\@omtext@mh@false
3 \DeclareOption{mh}{\@omtext@mh@true
4 \PassOptionsToPackage{\CurrentOption}{modules}}
5 \newif\ifindex\indextrue
6 \DeclareOption{noindex}{\indexfalse}
7 \DeclareOption*{\PassOptionsToPackage{\CurrentOption}{modules}}
8 \ProcessOptions
9 \ifindex\makeindex\fi

10 \if@omtext@mh\RequirePackage{omtext-mh}\fi
11 \RequirePackage{xspace}
12 \RequirePackage{modules}
13 \RequirePackage{comment}
14 \RequirePackage{mdframed}
15 \RequirePackage{latexsym}
```

4.2 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` infrastructure [Koh15a]. 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).

```
16 \srefaddidkey{omtext}
17 \addmetakey[] {omtext}{functions}
18 \addmetakey* {omtext}{display}
19 \addmetakey {omtext}{for}
20 \addmetakey {omtext}{from}
21 \addmetakey {omtext}{type}
22 \addmetakey* {omtext}{title}
23 \addmetakey* {omtext}{start}
24 \addmetakey {omtext}{theory}
25 \addmetakey {omtext}{continues}
26 \addmetakey {omtext}{verbalizes}
27 \addmetakey {omtext}{subject}
```

The next keys handle module loading (see [KGA15b]).

```
28 % \ednote{MK: need to implement these in LaTeXML, I wonder whether there is a general
29 % mechanism like numberit.}\ednote{MK: this needs to be rethought in the light of
30 % |\usemodule|. It is probably obsolete. Is this used? Is this documented?}
```

⁸EDNOTE: need an implementation for \LaTeX ML

```

31 \define@key{omtext}{require}{\requiremodules{#1}{sms}}
32 \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
33 \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.

34 \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.

35 \def\omtext@pre@skip{\smallskip}
36 \def\omtext@post@skip{}
37 \providecommand{\stDMemph}[1]{\textbf{#1}}
38 \newenvironment{omtext}[1][]{\@in@omtexttrue%
39 \bgroup\metasetkeys{omtext}{#1}\sref@label{id{this paragraph}}%
40 \def\lec##1{\@lec{##1}}%
41 \ifx\omtext@display\st@flow\else\omtext@pre@skip\par\noindent%
42 \ifx\omtext@title\@empty%
43 \ifx\omtext@start\@empty\else\stDMemph{\omtext@start}\xspace\fi%
44 \else\stDMemph{\omtext@title}:\xspace%
45 \ifx\omtext@start\@empty\else\omtext@start\xspace\fi%
46 \fi% \omtext@title empty
47 \fi% \omtext@display=flow
48 \ignorespaces}
49 {\egroup\omtext@post@skip\@in@omtextfalse}

```

4.3 Phrase-level Markup

```

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

50 \srefaddidkey{phrase}
51 \addmetakey{phrase}{style}
52 \addmetakey{phrase}{class}
53 \addmetakey{phrase}{index}
54 \addmetakey{phrase}{verbalizes}
55 \addmetakey{phrase}{type}
56 \addmetakey{phrase}{only}
57 \newcommand\phrase[2][]{\metasetkeys{phrase}{#1}%
58 \ifx\phrase@only\@empty\only<\phrase@only>{#2}\else #2\fi}

\coref*

59 \providecommand\textsubscript[1]{\ensuremath{_{#1}}}
60 \newcommand\corefs[2]{#1\textsubscript{#2}}
61 \newcommand\coreft[2]{#1\textsuperscript{#2}}

```


`\n*lex`

```
62 \newcommand\nlex[1]{\green{\sl{#1}}}
63 \newcommand\nlcex[1]{*\green{\sl{#1}}}
```

`sinlinequote`

```
64 \def\@sinlinequote#1{'\sl{#1}'}
65 \def\@@sinlinequote#1#2{\@sinlinequote{#2}~#1}
66 \newcommand\sinlinequote[2]{}
67 {\def\@opt{#1}\ifx\@opt\@empty\@sinlinequote{#2}\else\@@sinlinequote\@opt{#2}\fi}
```

4.4 Declarations (under development)

The declaration macros are still under development (i.e. the macros) are still under development and may change at any time. Currently they are completely empty.

```
68 \newcommand\vdec[1]{#1}
69 \newcommand\vids[2][]{#2}
70 \newcommand\vrest[2][]{#2}
```

4.5 Block-Level Markup

`sblockquote`

```
71 \def\begin@sblockquote{\begin{quote}\sl}
72 \def\end@sblockquote{\end{quote}}
73 \def\begin@@sblockquote#1{\begin@sblockquote}
74 \def\end@@sblockquote#1{\def\@lec##1{{\rm #1}}\@lec{#1}\end@sblockquote}
75 \newenvironment{sblockquote}[1]{}
76 {\def\@opt{#1}\ifx\@opt\@empty\begin@sblockquote\else\begin@@sblockquote\@opt\fi}
77 {\ifx\@opt\@empty\end@sblockquote\else\end@@sblockquote\@opt\fi}
```

`sboxquote`

```
78 \newenvironment{sboxquote}[1]{}
79 {\def\@src{#1}\begin{mdframed}[leftmargin=.5cm,rightmargin=.5cm]}
80 {\@lec{\rm\@src}\end{mdframed}}
```

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.

```
81 \providecommand{\@@lec}[1]{( #1 )}
82 \def\@lec#1{\strut\hfil\strut\null\nobreak\hfill\@@lec{#1}}
83 \def\lec#1{\@lec{#1}\par}
```

`\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.

```
84 \newcommand\mygraphics[2] [] {\includegraphics[#1]{#2}}
85 \newcommand\mycgraphics[2] [] {\begin{center}\mygraphics[#1]{#2}\end{center}}
86 \newcommand\mybgraphics[2] [] {\fbox{\mygraphics[#1]{#2}}}
87 \newcommand\mycbgraphics[2] [] {\begin{center}\fbox{\mygraphics[#1]{#2}}\end{center}}
```

4.6 Index Markup

`\omdoc@index` this is the main internal indexing command. It makes sure that the modules neces-
sary 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.

```
88 \addmetakey{omdoc@index}{at}
89 \addmetakey[false]{omdoc@index}{loadmodules}[true]
90 \newcommand\omdoc@index[2] [] {\ifindex%
91 \metasetkeys{omdoc@index}{#1}%
92 \@bsphack\begingroup\@sanitize%
93 \ifx\omdoc@index@loadmodules\@true%
94 \protected@write\@indexfile{}\string\indexentry%
95 {\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi%
96 {\string\usemodules{\@ifundefined{mod@id}\used@modules\mod@id}%
97 #2}}{\thepage}}%
98 \else%
99 \protected@write\@indexfile{}\string\indexentry%
100 {\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi#2}{\thepage}}%
101 \fi% loadmodules
102 \endgroup\@esphack\fi}%ifindex
```

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

```
\indexalt
103 \newcommand\indexalt[3] [] {{#2}\omdoc@index[#1]{#3}} % word in text and index
```

```
\indextoo
104 \newcommand\indextoo[2] [] {{#2}\omdoc@index[#1]{#2}} % word in text and index
```

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

```
105 \newcommand\@twin[3] [] {\omdoc@index[#1]{#2!#3}\omdoc@index[#1]{#3!#2}}
```

And again we have two interface macros building on this

```
\twinalt
106 \newcommand\twinalt[4] [] {{#2}\@twin[#1]{#3}{#4}}
```

EdN:9

```

\twinalt
107 \newcommand\twintoo[3] []{{#2 #3}\@twin[#1]{#2}{#3}} % and use the word compound too

\@twin this puts adjectivized two-compound words into the index in various permutations9
108 \newcommand\@twin[4] []{\omdoc@index[#1]{#2!#3!#4}\omdoc@index[#1]{#3!#2 (#4)}}

and the two interface macros for this case:

\@twinalt
109 \newcommand\@twinalt[5] []{{#2\@twin[#1]{#3}{#4}{#4}}}

\atwintoo
110 \newcommand\atwintoo[4] []{{#2 #3 #4}\@twin[#1]{#2}{#3}{#4}} % and use it too

```

4.7 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.

```

111 \newcommand\hateq{\ensuremath{\hat{=}}\xspace}
112 \newcommand\hatequiv{\ensuremath{\hat{\equiv}}\xspace}
113 \@ifundefined{ergo}%
114 {\newcommand\ergo{\ensuremath{\leadsto}\xspace}}%
115 {\renewcommand\ergo{\ensuremath{\leadsto}\xspace}}%
116 \newcommand\reflect@squig[2]{\reflectbox{$\m@th#1\rightsquigarrow$}}%
117 \newcommand\ogre{\ensuremath{\mathrel{\mathpalette\reflect@squig\relax}}\xspace}%
118 </package>

```

⁹EdNOTE: what to do with the optional argument here and below?

Index

Numbers written in *italic* refer to the page where the corresponding entry is described; numbers underlined refer to the code line of the definition; numbers in *roman* refer to the code lines where the entry is used.

Abelian		group	
group,	<u>6</u>	Abelian,	6

References

- [KGA15a] Michael Kohlhase, Deyan Ginev, and Rares Ambrus. *modules.sty: Semantic Macros and Module Scoping in sTeX*. Tech. rep. 2015. URL: <https://github.com/KWARC/sTeX/raw/master/sty/modules/modules.pdf>.
- [KGA15b] Michael Kohlhase, Deyan Ginev, and Rares Ambrus. *modules.sty: Semantic Macros and Module Scoping in sTeX*. Tech. rep. Comprehensive T_EX Archive Network (CTAN), 2015. URL: <http://www.ctan.org/get/macros/latex/contrib/stex/modules/modules.pdf>.
- [KGA15c] Michael Kohlhase, Deyan Ginev, and Rares Ambrus. *modules.sty: Semantic Macros and Module Scoping in sTeX*. Self-documenting L^AT_EX package. 2015.
- [Koh06] Michael Kohlhase. OMDoc – *An open markup format for mathematical documents [Version 1.2]*. LNAI 4180. Springer Verlag, Aug. 2006. URL: <http://omdoc.org/pubs/omdoc1.2.pdf>.
- [Koh15a] Michael Kohlhase. *metakeys.sty: A generic framework for extensible Metadata in L^AT_EX*. Tech. rep. Comprehensive T_EX Archive Network (CTAN), 2015. URL: <http://www.ctan.org/tex-archive/macros/latex/contrib/stex/metakeys/metakeys.pdf>.
- [Koh15b] Michael Kohlhase. *statements.sty: Structural Markup for Mathematical Statements*. Tech. rep. 2015. URL: <https://github.com/KWARC/sTeX/raw/master/sty/statements/statements.pdf>.
- [sTeX] *KWARC/sTeX*. URL: <https://svn.kwarc.info/repos/stex> (visited on 05/15/2015).