omdoc.sty/cls: Semantic Markup for Open Mathematical Documents in LATEX

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

January 15, 2016

Abstract

The <code>omdoc</code> package is part of the STEX collection, 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).

This package supplies an infrastructure for writing OMDoc documents in IATEX. This includes a simple structure sharing mechanism for STEX that allows to to move from a copy-and-paste document development model to a copy-and-reference model, which conserves space and simplifies document management. The augmented structure can be used by MKM systems for added-value services, either directly from the STEX sources, or after translation.

Contents

1	Inti	Introduction		
2	The	e User Interface	3	
	2.1	Package and Class Options	3	
	2.2	Document Structure	3	
	2.3	Ignoring Inputs	4	
	2.4	Structure Sharing	5	
	2.5	Colors	6	
3	Lim	itations 6		
4	Imp	plementation: The OMDoc Class	7	
	4.1	Class Options	7	
	4.2	Beefing up the document environment	7	
5	Imp	plementation: OMDoc Package	8	
	5.1	Package Options	8	
	5.2	~ ·	8	
	5.3		11	
	5.4		11	
	5.5		12	
	5.6		12	
	5.7		12	

1 Introduction

STEX is a version of TeX/IATeX that allows to markup TeX/IATeX documents semantically without leaving the document format, essentially turning TeX/IATeX into a document format for mathematical knowledge management (MKM). The package supports direct translation to the OMDoc format [Koh06]

The omdoc package supplies macros and environment that allow to label document fragments and to reference them later in the same document or in other documents. In essence, this enhances the document-as-trees model to documents-as-directed-acyclic-graphs (DAG) model. This structure can be used by MKM systems for added-value services, either directly from the STEX sources, or after translation. Currently, trans-document referencing provided by this package can only be used in the STEX collection.

DAG models of documents allow to replace the "Copy and Paste" in the source document with a label-and-reference model where document are shared in the document source and the formatter does the copying during document formatting/presentation.¹

2 The User Interface

The omdoc package generates two files: omdoc.cls, and omdoc.sty. The OM-Doc class is a minimally changed variant of the standard article class that includes the functionality provided by omdoc.sty. The rest of the documentation pertains to the functionality introduced by omdoc.sty.

2.1 Package and Class Options

The omdoc package and class accept the following options:

report	load report.cls instead of article.cls		
book	load book.cls instead of article.cls		
showignores	show the the contents of the ignore environment after all		
showmeta	show the metadata; see metakeys.sty		
showmods	show modules; see modules.sty		
extrefs	allow external references; see sref.sty		
defindex	index definienda; see statements.sty		

2.2 Document Structure

document \documentkeys

EdN:1

The top-level document environment can be given key/value information by the \documentkeys macro in the preamble¹. This can be used to give metadata about the document. For the moment only the id key is used to give an identifier to the

 $^{^{1}\}mathrm{EdNote}$: integrate with latexml's XMRef in the Math mode.

¹We cannot patch the document environment to accept an optional argument, since other packages we load already do; pity.

omdoc element resulting from the LATEXML transformation.

omgroup

The structure of the document is given by the omgroup environment just like in OMDoc. In the LATEX route, the omgroup environment is flexibly mapped to sectioning commands, inducing the proper sectioning level from the nesting of omgroup environments. Correspondingly, the omgroup environment takes an optional key/value argument for metadata followed by a regular argument for the (section) title of the omgroup. The optional metadata argument has the keys id for an identifier, creators and contributors for the Dublin Core metadata [DCM03]; see [Koh15a] for details of the format. The short allows to give a short title for the generated section. If the title contains semantic macros, they need to be protected by \protect, and we need to give the loadmodules key it needs no value. For instance we would have

creators
contributors
short
loadmodules

```
\begin{module}{foo}
\symdef{bar}{B^a_r}
...
\begin{omgroup}[id=barderiv,loadmodules]
   {Introducing $\protect\bar$ Derivations}
```

blindomgroup

STEX automatically computes the sectioning level, from the nesting of omgroup environments. But sometimes, we want to skip levels (e.g. to use a subsection* as an introduction for a chapter). Therefore the omdoc package provides a variant blindomgroup that does not produce markup, but increments the sectioning level and logically groups document parts that belong together, but where traditional document markup relies on convention rather than explicit markup. The blindomgroup environment is useful e.g. for creating frontmatter at the correct level. Example 1 shows a typical setup for the outer document structure of a book with parts and chapters. We use two levels of blindomgroup:

- The outer one groups the introductory parts of the book (which we assume to have a sectioning hierarchy topping at the part level). This blindomgroup makes sure that the introductory remarks become a "chapter" instead of a "part".
- Th inner one groups the frontmatter² and makes the preface of the book a section-level construct. Note that here the display=flow on the omgroup environment prevents numbering as is traditional for prefaces.

The \currentsectionlevel macro supplies the name of the current sectioning level, e.g. "chapter", or "subsection". \CurrentSectionLevel is the capitalized variant. They are useful to write something like "In this \currentsectionlevel, we will..." in an omgroup environment, where we do not know which sectioning level we will end up.

\currentsectionlevel \CurrentSectionLevel

2.3 Ignoring Inputs

ignore The ignore environment can be used for hiding text parts from the document

²We shied away from redefining the **frontmatter** to induce a blindomgroup, but this may be the "right" way to go in the future.

```
\begin{document}
\begin{blindomgroup}
\begin{blindomgroup}
\begin{frontmatter}
\maketitle\newpage
\begin{omgroup}[display=flow]{Preface}
... <<pre><<pre>c<<pre>
\end{omgroup}
\clearpage\setcounter{tocdepth}{4}\tableofcontents\clearpage
\end{frontmatter}
\end{blindomgroup}
... <<introductory remarks>> ...
\end{blindomgroup}
\begin{omgroup}{Introduction}
... <<intro>> ...
\end{omgroup}
... <<more chapters>> ...
\bibliographystyle{alpha}\bibliography{kwarc}
\end{document}
```

Example 1: A typical Document Structure of a Book

showignores

structure. The body of the environment is not PDF or DVI output unless the showignores option is given to the omdoc class or package. But in the generated OMDoc result, the body is marked up with a ignore element. This is useful in two situations. For

editing One may want to hide unfinished or obsolete parts of a document

narrative/content markup In STEX we mark up narrative-structured documents. In the generated OMDoc documents we want to be able to cache content objects that are not directly visible. For instance in the statements package [Koh15c] we use the \inlinedef macro to mark up phrase-level definitions, which verbalize more formal definitions. The latter can be hidden by an ignore and referenced by the verbalizes key in \inlinedef.

2.4 Structure Sharing

\STRlabel \STRcopy The \STRlabel macro takes two arguments: a label and the content and stores the the content for later use by \STRcopy[$\langle URL \rangle$] { $\langle label \rangle$ }, which expands to the previously stored content. If the \STRlabel macro was in a different file, then we can give a URL $\langle URL \rangle$ that lets LaTrXML generate the correct reference.

\STRsemantics

The \STRlabel macro has a variant \STRsemantics, where the label argument is optional, and which takes a third argument, which is ignored in LATEX. This allows to specify the meaning of the content (whatever that may mean) in cases, where the source document is not formatted for presentation, but is transformed into some content markup format.²

 $^{^2\}mathrm{EdNote}$: document LMID und LMXREf here if we decide to keep them.

2.5 Colors

\blue \red For convenience, the omdoc package defines a couple of color macros for the color package: For instance \blue abbreviates \textcolor{blue}, so that $\begin{tabular}{l} \begin{tabular}{l} \begin{tabular}$

\black

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 ST_EX GitHub repository [sTeX].

1. when option book which uses \pagestyle{headings} is given and semantic macros are given in the omgroup titles, then they sometimes are not defined by the time the heading is formatted. Need to look into how the headings are made.

4 Implementation: The OMDoc Class

The functionality is spread over the omdoc class and package. The class provides the document environment and the omdoc element corresponds to it, whereas the package provides the concrete functionality.

4.1 Class Options

\omdoc@class \ifclass@book To initialize the omdoc class, we declare and process the necessary options. For omdoc.cls this is quite simple. We have options report and book, which set the \omdoc@class macro and pass on the macro to omdoc.sty for further processing. The book option also sets the conditional to true for the frontmatter handling later.

```
1 \( \*\cls \)
2 \def \omdoc@class{article}
3 \DeclareOption{report}{\def \omdoc@class{report}%
4 \PassOptionsToPackage{\CurrentOption}{omdoc}
5 \PassOptionsToPackage{\CurrentOption}{stex}}
6 \newif\ifclass@book\class@bookfalse
7 \DeclareOption{book}{\def \omdoc@class{book}\class@booktrue%
8 \PassOptionsToPackage{\CurrentOption}{omdoc}
9 \PassOptionsToPackage{\CurrentOption}{stex}}
```

the rest of the options are only passed on to omdoc.sty and the class selected by the first options.

We load article.cls, and the desired packages. For the LATEXML bindings, we make sure the right packages are loaded.

```
14 \LoadClass{\omdoc@class}
15 \RequirePackage{etoolbox}
16 \RequirePackage{omdoc}
17 \RequirePackage{stex}
```

4.2 Beefing up the document environment

Now, we will define the environments we need. The top-level one is the document environment, which we redefined so that we can provide keyval arguments.

document

EdN:3

For the moment we do not use them on the LaTeX level, but the document identifier is picked up by LaTeXML.³

```
18 \srefaddidkey{document}
```

19 \newcommand\documentkeys[1]{\metasetkeys{document}{#1}}

³EdNote: faking documentkeys for now. @HANG, please implement

```
20 \let\orig@document=\document  
21 \srefaddidkey{document}  
22 \renewcommand{\document} [1] [] {\metasetkeys{document} {#1} \orig@document}  
23 \langle / cls \rangle
```

5 Implementation: OMDoc Package

5.1 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). The report and book options affect the sectioning behavior of the omgroup environment via the \section@level macro later.

\section@level

- $24 \ \ \text{package}$ 25 \newif\ifshow@ignores\show@ignoresfalse
- 26 \DeclareOption{showignores}{\show@ignorestrue}
- 27 \newcount\section@level\section@level=2
- 28 \DeclareOption{report}{\section@level=0}
- 29 \DeclareOption{book}{\section@level=0}
- 30 \DeclareOption*{\PassOptionsToPackage{\CurrentOption}{sref}}
- 31 \ProcessOptions

Then we need to set up the packages by requiring the sref package to be loaded.

- 32 \RequirePackage{sref}
- 33 \RequirePackage{xspace}
- 34 \RequirePackage{comment}
- 35 \RequirePackage{etoolbox}

5.2 Document Structure

The structure of the document is given by the omgroup environment just like in OMDoc. The hierarchy is adjusted automatically according to the LATEX class in effect.

\currentsectionlevel

For the \currentsectionlevel and \Currentsectionlevel macros we use an internal macro \current@section@level that only contains the keyword (no markup). We initialize it with "document" as a default. In the generated OMDoc, we only generate a text element of class omdoc_currentsectionlevel, wich will be instantiated by CSS later.⁴

- 36 \def\current@section@level{document}%
- 37 \newcommand\currentsectionlevel{\lowercase\expandafter{\current@section@level}\xspace}%
- $38 \verb| newcommand| Currentsectionlevel{expandafter| MakeUppercase| current@section@level| xspace} % and the contraction of the$

8

EdN:4

 $^{^4{\}rm EDNote}$: MK: we may have to experiment with the more powerful uppercasing macro from mfirstuc.sty once we internationalize.

```
blindomgroup
                                          39 \newcommand\at@begin@blindomgroup[1]{}
                                          40 \newenvironment{blindomgroup}
                                          41 {\advance\section@level by 1\at@begin@blindomgroup\setion@level}
                                          42 {\advance\section@level by -1}
                                         convenience macro: \operatorname{\mathsf{Nomgroup@nonum}}(\operatorname{\mathsf{level}}) + (\operatorname{\mathsf{title}}) makes an unnumbered sec-
       \omgroup@nonum
                                          tioning with title \langle title \rangle at level \langle level \rangle.
                                          43 \newcommand\omgroup@nonum[2]{%
                                          44 \ifx\hyper@anchor\@undefined\else\phantomsection\fi%
                                          45 \addcontentsline{toc}{#1}{#2}\Qnameuse{#1}*{#2}}
           \omgroup@num
                                          convenience macro: \operatorname{\mathsf{Nomgroup@nonum}}\{\langle level \rangle\}\{\langle title \rangle\} makes numbered sectioning
                                          with title \langle title \rangle at level \langle level \rangle. We have to check the short key was given in the
                                          omgroup environment and – if it is use it. But how to do that depends on whether
                                          the rdfmeta package has been loaded.
                                          46 \newcommand\omgroup@num[2]{%
                                          47 \end{c@sect@name} \end{condoc@sect@name} \end{condoc@sect@name}
                                          48 \ifx\omgroup@short\@empty\@nameuse{#1}{#2}%
                                          49 \else\@ifundefined{rdfmeta@sectioning}{\@nameuse{\#1}[\nomeroup@short]{\#2}}{\@mathchar`{42}}
                                          50 {\@nameuse{rdfmeta@#1@old}[\omgroup@short]{#2}}%
                                          51 \fi}
                      omgroup
                                          52 \def\@true{true}
                                          53 \def\@false{false}
                                          54 \srefaddidkey{omgroup}
                                          55 \addmetakey{omgroup}{date}
                                          56 \addmetakey{omgroup}{creators}
                                          57 \addmetakey{omgroup}{contributors}
                                          58 \addmetakey{omgroup}{srccite}
                                          59 \addmetakey{omgroup}{type}
                                          60 \addmetakey*{omgroup}{short}
                                          61 \addmetakey*{omgroup}{display}
                                          62 \addmetakey[false]{omgroup}{loadmodules}[true]
                                          we define a switch for numbering lines and a hook for the beginning of groups:
                                          The \at@begin@omgroup macro allows customization. It is run at the beginning
\at@begin@omgroup
                                          of the omgroup, i.e. after the section heading.
                                          63 \newif\if@@num\@@numtrue
                                          64 \newif\if@frontmatter\@frontmatterfalse
                                          65 \newif\if@backmatter\@backmatterfalse
                                          66 \newcommand\at@begin@omgroup[3][]{}
                                                 Then we define a helper macro that takes care of the sectioning magic. It
                                          comes with its own key/value interface for customization.
                                          67 \addmetakey{omdoc@sect}{name}
                                          68 \addmetakey[false] {omdoc@sect} {clear} [true]
```

69 \addmetakey{omdoc@sect}{ref}

```
70 \addmetakey[false]{omdoc@sect}{num}[true]
71 \newcommand\omdoc@sectioning[3][]{\metasetkeys{omdoc@sect}{#1}%
72 \ifx\omdoc@sect@clear\@true\cleardoublepage\fi%
73 \if@@num% numbering not overridden by frontmatter, etc.
74 \ifx\omdoc@sect@num\@true\omgroup@num{#2}{#3}\else\omgroup@nonum{#2}{#3}\fi%
75 \def\current@section@level{\omdoc@sect@name}%
76 \else\omgroup@nonum{#2}{#3}\fi}
and another one, if redefines the \addtocontentsline macro of LATEX to import
the respective macros. It takes as an argument a list of module names.<sup>5</sup>
77 \newcommand\omgroup@redefine@addtocontents[1]{\edef\@@import{#1}%
78 \@for\@I:=\@@import\do{\edef\@path{\csname module@\@I @path\endcsname}%
79 \@ifundefined{tf@toc}\relax{\protected@write\tf@toc{}{\string\@requiremodules{\@path}{sms}}}}
80 \ifx\hyper@anchor\@undefined% hyperref.sty loaded?
81 \def\addcontentsline##1##2##3{%
82 \addtocontents{##1}{\protect\contentsline{##2}{\string\importmodules{#1}##3}{\thepage}}}
83 \else\def\addcontentsline##1##2##3{\%
84 \addtocontents{##1}{\protect\contentsline{##2}{\string\importmodules{#1}##3}{\thepage}{\@curren
85 \fi}% hypreref.sty loaded?
now the omgroup environment itself. This takes care of the table of contents via
the helper macro above and then selects the appropriate sectioning command from
article.cls.
86 \newenvironment{omgroup}[2][]% keys, title
87 {\metasetkeys{omgroup}{#1}\sref@target%
88 \ifx\omgroup@display\st@flow\@@numfalse\fi
89 \if@frontmatter\@@numfalse\fi
If the loadmodules key is set on \begin{omgroup}, we redefine the \addcontetsline
macro that determines how the sectioning commands below construct the entries
for the table of contents.
90 \ifx\omgroup@loadmodules\@true%
91 \omgroup@redefine@addtocontents{\@ifundefined{mod@id}\imported@modules%
92 {\@ifundefined{module@\mod@id @path}{\imported@modules}\mod@id}}\fi%
now we only need to construct the right sectioning depending on the value of
\section@level.
93 \advance\section@level by 1
94 \ifcase\section@level%
95 \or\omdoc@sectioning[name=Part,clear,num]{part}{#2}%
96 \or\omdoc@sectioning[name=Chapter,clear,num]{chapter}{#2}%
97 \or\omdoc@sectioning[name=Section,num]{section}{#2}%
98 \or\omdoc@sectioning[name=Subsection,num] {subsection}{#2}%
99 \or\omdoc@sectioning[name=Subsubsection,num] {subsubsection}{#2}%
100 \or\omdoc@sectioning[name=Paragraph,ref=this paragraph]{paragraph}{#2}%
101 \or\omdoc@sectioning[name=Subparagraph,ref=this subparagraph]{paragraph}{#2}%
102 \fi% \ifcase
103 \at@begin@omgroup[#1]\section@level{#2}}% for customization
104 {\advance\section@level by -1}
```

EdN:5

 $^{^5\}mathrm{EdNote}\colon$ MK: the extension sms is hard-coded here, but should not be. This will not work in multilingual settings.

5.3 Front and Backmatter

Index markup is provided by the omtext package [Koh15b], so in the omdoc package we only need to supply the corresponding \printindex command, if it is not already defined

```
\printindex
            105 \providecommand\printindex{\IfFileExists{\jobname.ind}{\input{\jobname.ind}}{}}
            106 (/package)
frontmatter book.cls already has a \frontmatter macro, so we have to redefine the front
             matter environment in this case.
            107 ⟨*cls⟩
            108 \ifclass@book
            109 \renewenvironment{frontmatter}
            110 {\Cfrontmattertrue\cleardoublepage\Cmainmatterfalse\pagenumbering{roman}}
            111 {\@frontmatterfalse\setcounter{page}{1}\pagenumbering{arabic}}
            112 \else
            113 \newenvironment{frontmatter}
            114 {\@frontmattertrue\pagenumbering{roman}}
            115 {\@frontmatterfalse\setcounter{page}{1}\pagenumbering{arabic}}
            116 \fi
            117 %
                     \End{macrocode}
            118 % \end{environment}
            119 %
            120 % \begin{environment}{backmatter}
            121 %
                    |book.cls| already has a |backmatter| macro, so we have to redefine the back
            122 %
                   matter environment in this case.
            123 %
                    \begin{macrocode}
            124 \ifclass@book
            125 \renewenvironment{backmatter}
            126 {\cleardoublepage\@mainmatterfalse\@backmattertrue}
            127 {\@backmatterfalse}
            128 \else
            129 \newenvironment{backmatter}{\@backmattertrue}{\@backmatterfalse}
            130 \fi
```

5.4 Ignoring Inputs

```
ignore
```

131 (/cls)

```
132 \*package\
133 \ifshow@ignores
134 \addmetakey{ignore}{type}
135 \addmetakey{ignore}{comment}
136 \newenvironment{ignore}[1][]
137 {\metasetkeys{ignore}{#1}\textless\ignore@type\textgreater\bgroup\itshape}
138 {\egroup\textless/\ignore@type\textgreater}
139 \renewenvironment{ignore}{}{}\else\excludecomment{ignore}\fi
```

5.5 Structure Sharing

```
EdN:6
                             140 \providecommand{\lxDocumentID}[1]{}%
                             141 \def\LXMID#1#2{\expandafter\gdef\csname xmarg#1\endcsname{#2}\csname xmarg#1\endcsname}
                             142 \def\LXMRef#1{\csname xmarg#1\endcsname}
                  \STRlabel The main macro, it is used to attach a label to some text expansion. Later on,
                              using the \STRcopy macro, the author can use this label to get the expansion
                              originally assigned.
                             143 \long\def\STRlabel#1#2{\STRlabeldef{#1}{#2}{#2}}
                              The \STRcopy macro is used to call the expansion of a given label. In case the
                   \STRcopy
EdN:7
                              label is not defined it will issue a warning.<sup>7</sup>
                             144 \verb|\newcommand\STRcopy[2][]{\expandafter\ifx\csname STR0#2\endcsname\relax}
                             145 \message{STR warning: reference #2 undefined!}
                             146 \else\csname STR@#2\endcsname\fi}
              \STRsemantics if we have a presentation form and a semantic form, then we can use
                             147 \newcommand\STRsemantics[3][]{#2\def\@test{#1}\ifx\@test\@empty\STRlabeldef{#1}{#2\fi}
               \STRlabeldef This is the macro that does the actual labeling. Is it called inside \STRlabel
                             148 \def\STRlabeldef#1{\expandafter\gdef\csname STR@#1\endcsname}
                                     Colors
                              5.6
 blue, red, green, magenta We will use the following abbreviations for colors from color.sty
                             149 \def\black#1{\textcolor{black}{#1}}
                             150 \def\gray#1{\textcolor{gray}{#1}}
                             151 \def\blue#1{\textcolor{blue}{#1}}
                             152 \def\red#1{\textcolor{red}{#1}}
                             153 \def\green#1{\textcolor{green}{#1}}
                             154 \def\cyan#1{\textcolor{cyan}{#1}}
                             155 \def\magenta#1{\textcolor{magenta}{#1}}
```

Leftovers 5.7

159 \newcommand\baseURI[2][]{} 160 (/package)

156 \def\brown#1{\textcolor{brown}{#1}} 157 \def\yellow#1{\textcolor{yellow}{#1}} 158 \def\orange#1{\textcolor{orange}{#1}}

 $^{^6\}mathrm{EdNote}$: The following is simply copied over from the latexml package, which we eliminated, we should integrate better.

 $^{^7{}m EdNote}$: MK: we need to do something about the ref!

References

- [DCM03] The DCMI Usage Board. DCMI Metadata Terms. DCMI Recommendation. Dublin Core Metadata Initiative, 2003. URL: http://dublincore.org/documents/dcmi-terms/.
- [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. dcm.sty: An Infrastructure for marking up Dublin Core Metadata in LATEX documents. Tech. rep. Comprehensive TEX Archive Network (CTAN), 2015. URL: http://www.ctan.org/tex-archive/macros/latex/contrib/stex/dcm/dcm.pdf.
- [Koh15b] Michael Kohlhase. omtext: Semantic Markup for Mathematical Text Fragments in LATEX. Tech. rep. Comprehensive TEX Archive Network (CTAN), 2015. URL: http://www.ctan.org/tex-archive/macros/latex/contrib/stex/omtext/omtext.pdf.
- [Koh15c] Michael Kohlhase. statements.sty: Structural Markup for Mathematical Statements. Tech. rep. Comprehensive T_EX Archive Network (CTAN), 2015. URL: http://www.ctan.org/tex-archive/macros/latex/contrib/stex/statements/statements.pdf.
- [sTeX] KWARC/sTeX. URL: https://svn.kwarc.info/repos/stex (visited on 05/15/2015).