

smultiling.sty: Multilinguality Support for S_TE_X

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

April 27, 2014

Abstract

The **smultiling** package is part of the S_TE_X 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).

The **smultiling** package adds multilinguality support for S_TE_X, the idea is that multilingual modules in S_TE_X consist of a module signature together with multiple language bindings that inherit symbols from it, which also account for cross-language coordination.

Contents

1	Introduction	2
1.1	S _T E _X Module Signatures	2
2	The User Interface	3
3	Implementation	4
3.1	Class Options	4
3.2	Handling Languages	4
3.3	Signatures	6
3.4	Language Bindings	6

1 Introduction

We have been using \TeX as the encoding for the Semantic Multilingual Glossary of Mathematics (SMGloM; see [Gin+14]). The SMGloM data model has been taxing the representational capabilities of \TeX with respect to multilingual support and verbalization definitions; see [Koh14], which we assume as background reading for this note.

1.1 \TeX Module Signatures

(monolingual) \TeX had the intuition that the symbol definitions ($\text{\textbackslash symdef}$ and $\text{\textbackslash symvariant}$) are interspersed with the text and we generate \TeX module signatures (SMS *.sms files) from the \TeX files. The SMS duplicate “formal” information from the “narrative” \TeX files. In the SMGloM, we extend this idea by making the the SMS primary objects that contain the language-independent part of the formal structure conveyed by the \TeX documents and there may be multiple narrative “language bindings” that are translations of each other – and as we do not want to duplicate the formal parts, those are inherited from the SMS rather than written down in the language binding itself. So instead of

```
\begin{module}[id=foo]
\symdef{bar}{BAR}
\begin{definition}[for=bar]
  A \defiii{big}{array}{raster} ( $\bar{\phantom{x}}$ ) is a\ldots, it is much bigger
  than a \defiii[sar]{small}{array}{raster}.
\end{definition}
\end{module}
```

we now advocate the divided style in the listing below.

```
\usepackage[english,ngerman]{multiling}
\begin{modsig}{foo}
\symdef{bar}{BAR}
\symbol{sar}
\end{modsig}

\begin{modnl}[creators=miko,primary]{foo}{en}
\begin{definition}
  A \defiii[bar]{big}{array}{raster} ( $\bar{\phantom{x}}$ ) is a\ldots, it is much bigger
  than a \defiii[sar]{small}{array}{raster}.
\end{definition}
\end{modnl}

\begin{modnl}[creators=miko]{foo}{de}
\begin{definition}
  Ein \defiii[bar]{gro"ses}{Feld}{Raster} ( $\bar{\phantom{x}}$ ) ist ein\ldots, es
  ist viel gr"o"ser als ein \defiii[sar]{kleines}{Feld}{Raster}.
\end{definition}
\end{modnl}
```

`\end{modnl}`

There the `modsig` environment works exactly like the old `module` environment, only that the `id` attribute has moved into the required argument – anonymous module signatures do not make sense. The `modnl` environment takes two arguments the first is the name of the module signature it provides language bindings for and the second the ISO 639 language specifier of the content language. We add the `primary` key `modnl`, which can specify the primary language binding (the one the others translate from; and which serves as the reference in case of translation conflicts).¹

There is another difference in the multilingual encoding: All symbols are introduced in the module signature, either by a `\symdef` or the new `\symbol` macro.

We retain the old `module` environment as an intermediate stage. It is still useful for monolingual texts. Note that for files with a module, we still have to extract `*.sms` files. It is not completely clear yet, how to adapt the workflows. We clearly need a `lmh` or editor command that transfers an old-style module into a new-style signature/binding combo to prepare it for multilingual treatment.

2 The User Interface

The `smultiling` package accepts all options of the `babel.sty` and just passes them on to it. The options specify which languages can be used in the `gTeX` language bindings.

¹EdNOTE: ©DG: This needs to be implemented in LaTeXML

3 Implementation

Technically, the `smultiling` package is essentially a wrapper around the `babel` package but allows specification of languages by their ISO 639 language codes.

3.1 Class Options

To initialize the `smultiling` class, we pass on all options to `babel.cls` and record which languages are loaded by defining `\smul@⟨language⟩@loaded` macros.²

The `langfiles` option specifies that for a module `⟨mod⟩`, the module signature file has the name `⟨mod⟩.tex` and the language bindings of language with the ISO 639 language specifier `⟨lang⟩` have the file name `⟨mod⟩.⟨lang⟩.tex`.³

```

1 ⟨*sty⟩
2 \newif\if@langfiles\@langfilesfalse
3 \DeclareOption{langfiles}{\@langfilestrue}
4 \DeclareOption*{\PassOptionsToPackage{\CurrentOption}{babel}}
5 \@namedef{smul@\CurrentOption @loaded}{yes}
6 \ProcessOptions
7 ⟨/sty⟩
8 ⟨*ltxml⟩
9 # -*- CPERL -*-
10 package LaTeXML::Package::Pool;
11 use strict;
12 use LaTeXML::Package;
13 DeclareOption('langfiles',sub {AssignValue('smultiling_langfiles',1,'global')});
14 DeclareOption(undef,sub {PassOptions('babel','sty',ToString(Digest(T_CS('\CurrentOption')))); });
15 ProcessOptions();
16 ⟨/ltxml⟩

    We load babel.sty

17 ⟨*sty⟩
18 \RequirePackage{etoolbox}
19 \RequirePackage{babel}
20 ⟨/sty⟩
21 ⟨*ltxml⟩
22 RequirePackage('babel');
23 ⟨/ltxml⟩

```

3.2 Handling Languages

`\smg@select@language` This macro selects one of the registered languages by its language code by setting the internal `\smg@lang` macro to the argument and then runs the actual selection code in `\smg@select@lang`. This internal code register is only initialized there, the code is generated by the `\smg@register@language` macro below.

```

24 ⟨ltxml⟩RawTeX(

```

²EDNOTE: @DG: We also want to do that in `LaTeXML`

³EDNOTE: implement other schemes, e.g. the onefile scheme.

```

25 <*sty | ltxml>
26 \newcommand\smg@select@lang{}
27 \newcommand\smg@select@language[1]{\def\smg@lang{#1}\smg@select@lang}

\smg@register@language \smg@register@language{<lang>}{<babel>} registers the babel language name
<babel> with its ISO 639 language code <lang> by extending the \smg@select@language
macro.

28 \newcommand\smg@register@language[2]%
29 {\@ifundefined{smul@#1@loaded}{}\appto\smg@select@lang%
30 {\expandafter\ifstrequal\expandafter\smg@lang{#1}{\selectlanguage{#2}}{}}}}

```

Now we register a couple of languages for which we have babel support. Maybe we have to extend this list with others. But then we have to extend the mechanisms.

```

31 \smg@register@language{af}{afrikaans}
32 \smg@register@language{de}{ngerman}
33 \smg@register@language{fr}{french}%
34 \smg@register@language{he}{hebrew}
35 \smg@register@language{hu}{hungarian}
36 \smg@register@language{id}{indonesian}
37 \smg@register@language{ms}{malay}
38 \smg@register@language{nn}{nynorsk}
39 \smg@register@language{pt}{portuguese}
40 \smg@register@language{ru}{russian}
41 \smg@register@language{uk}{ukrainian}
42 \smg@register@language{en}{english}
43 \smg@register@language{es}{spanish}
44 \smg@register@language{sq}{albanian}
45 \smg@register@language{bg}{bulgarian}
46 \smg@register@language{ca}{catalan}
47 \smg@register@language{hr}{croatian}
48 \smg@register@language{cs}{czech}
49 \smg@register@language{da}{danish}
50 \smg@register@language{nl}{dutch}
51 \smg@register@language{eo}{esperanto}
52 \smg@register@language{et}{estonian}
53 \smg@register@language{fi}{finnish}
54 \smg@register@language{ka}{georgian}
55 \smg@register@language{el}{greek}
56 \smg@register@language{is}{icelandic}
57 \smg@register@language{it}{italian}
58 \smg@register@language{la}{latin}
59 \smg@register@language{no}{norsk}
60 \smg@register@language{pl}{polish}
61 \smg@register@language{sr}{serbian}
62 \smg@register@language{sk}{slovak}
63 \smg@register@language{sl}{slovenian}
64 \smg@register@language{sv}{swedish}
65 \smg@register@language{th}{thai}
66 \smg@register@language{tr}{turkish}

```

```

67 \smg@register@language{vi}{vietnamese}
68 \smg@register@language{cy}{welsh}
69 \smg@register@language{hi}{hindi}

```

3.3 Signatures

modsig The `modsig` environment is just a layer over the `module` environment. We also redefine macros that may occur in module signatures so that they do not create markup.

```

70 \newenvironment{modsig}[2][]{%
71 \def\@test{#1}\ifx\@test\@empty\begin{module}[id=#2]\else\begin{module}[id=#2,#1]\fi}
72 {\end{module}}

```

viewsig The `viewsig` environment is just a layer over the `mhview` environment with the keys suitably adapted.

```

73 \newenvironment{viewsig}[4][]{\def\@test{#1}\ifx\@test\@empty%
74 \begin{mhview}[id=#2,ext=tex]{#3}{#4}\else\begin{mhview}[id=#2,#1,ext=tex]{#3}{#4}\fi}
75 {\end{mhview}}
76 \<sty|ltxml>
77 \<ltxml>');

```

3.4 Language Bindings

modnl:

```

78 \<sty>
79 \addmetakey{modnl}{load}
80 \addmetakey*{modnl}{title}
81 \addmetakey*{modnl}{creators}
82 \addmetakey*{modnl}{contributors}
83 \addmetakey{primary}{contributors}[yes]
84 \</sty>
85 \<ltxml>
86 DefKeyVal('modnl','title','Semiverbatim');
87 DefKeyVal('modnl','load','Semiverbatim');
88 DefKeyVal('modnl','creators','Semiverbatim');
89 DefKeyVal('modnl','contributors','Semiverbatim');
90 DefKeyVal('modnl','primary','Semiverbatim');
91 \</ltxml>

```

modnl The `modnl` environment is just a layer over the `module` environment with the keys and language suitably adapted.

```

92 \<sty>
93 \newenvironment{modnl}[3][]{\metasetkeys{modnl}{#1}%
94 \smg@select@language{#3}%
95 \def\@test{#1}\ifx\@test\@empty\begin{module}[id=#2.#3]\else\begin{module}[id=#2.#3,#1]\fi%
96 \if@langfiles\importmodule[load=#2,ext=tex]{#2}\else
97 \ifx\modnl@load\@empty\importmodule{#2}\else\importmodule[ext=tex,load=\modnl@load]{#2}\fi%

```

```

98 \fi}
99 {\end{module}}
100 \</sty>
101 \<!--*ltxml-->
102 DefEnvironment('{modnl} OptionalKeyVals:modnl {}{}',
103     "<omdoc:theory "
104     . 'xml:id="#2.#3">'
105     . "&defined(&GetKeyVal(#1,'creators'))(<dc:creator>&GetKeyVal(#1,'creators')</dc:cr"
106     . "&defined(&GetKeyVal(#1,'title'))(<dc:title>&GetKeyVal(#1,'title')</dc:title>())"
107     . "&defined(&GetKeyVal(#1,'contributors'))(<dc:contributor>&GetKeyVal(#1,'contribut"
108     . "#body"
109     . "</omdoc:theory>",
110     afterDigestBegin=>sub {
111         my ($stomach, $whatsit) = @_;
112         my $keyval = $whatsit->getArg(1);
113         my $signature = ToString($whatsit->getArg(2));
114         if ($keyval) {
115             # If we're not given load, AND the langfiles option is in effect,
116             # default to #2
117             if ((! $keyval->getValue('load')) && (LookupValue('smultiling_langfiles')) {
118                 $keyval->setValue('load',$signature); }
119             # Always load a TeX file
120             $keyval->setValue('ext','tex'); }
121         importmoduleI($stomach,$whatsit)});
122 \</ltxml>%$

```

References

- [Gin+14] Deyan Ginev et al. “The SMGLoM Project and System”. 2014. URL: <http://kwarc.info/kohlhase/submit/cicm14-smglom-system.pdf>.
- [Koh14] Michael Kohlhase. “A Data Model and Encoding for a Semantic, Multilingual Glossary of Mathematics”. In: *Intelligent Computer Mathematics*. Conferences on Intelligent Computer Mathematics. (Coimbra, Portugal, July 7–11, 2014). Ed. by Stephan Watt et al. Lecture Notes in Computer Science. accepted. Springer, 2014. URL: <http://kwarc.info/kohlhase/submit/cicm14-smglom-datamd1.pdf>. Forthcoming.

␣ltxml␣1;