`smultiling.sty`: Multilingual Support for sTeX

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

April 27, 2014

Abstract

The `smultiling` 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).

The `smultiling` package adds multilinguality support for sTeX, the idea is that multilingual modules in sTeX 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

We have been using sTeX as the encoding for the Semantic Multilingual Glossary of Mathematics (SMGloM; see [**IanJucKoh:sps14**]). The SMGloM data model has been taxing the representational capabilities of sTeX with respect to multilingual support and verbalization definitions; see [**Kohlhase:dmesmgm14**], which we assume as background reading for this note.

1.1 sTeX Module Signatures

(monolingual) sTeX had the intuition that the symbol definitions (`\symdef` and `\symvariant`) are interspersed with the text and we generate sTeX module signatures (SMS `*.sms` files) from the sTeX files. The SMS duplicate "formal" information from the "narrative" sTeX 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 sTeX 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} ($\bar$) is a\ldots
\end{definition}
\end{module}
```

EdN:1 we now advocate the divided style in the listing below[1].

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

\begin{modnl}[creators=miko,primary]{foo}{en}
\begin{definition}
  A \defiii[bar]{big}{array}{raster} ($\bar$) is a\ldots
\end{definition}
\end{modnl}

\begin{modnl}[creators=miko]{foo}{de}
\begin{definition}
  Ein \defiii[bar]{gro"ses}{Feld}{Raster} ($\bar$) is a\ldots
\end{definition}
\end{modnl}
```

[1] EdNote: MK: the names of the environments are still very much in the air. "modsig" I rather like, but "modnl" is terrible

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 to the optional argument of `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).

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 sTeX language bindings.

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

EdN:2 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]

`langfiles` 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] EdN:3

```
⟨*sty⟩
\newif\if@langfiles\@langfilesfalse
\DeclareOption{langfiles}{\@langfilestrue}
\DeclareOption*{\PassOptionsToPackage{\CurrentOption}{babel}
\@namedef{smul@\CurrentOption @loaded}{yes}}
\ProcessOptions
⟨/sty⟩
⟨*ltxml⟩
# -*- CPERL -*-
package LaTeXML::Package::Pool;
use strict;
use LaTeXML::Package;
DeclareOption('langfiles',sub {AssignValue('smultiling_langfiles',1,'global');});
DeclareOption(undef,sub {PassOptions('babel','sty',ToString(Digest(T_CS('\CurrentOption')))); }
ProcessOptions();
⟨/ltxml⟩
```

We load `babel.sty`

```
⟨*sty⟩
\RequirePackage{etoolbox}
\RequirePackage{babel}
⟨/sty⟩
⟨*ltxml⟩
RequirePackage('babel');
⟨/ltxml⟩
```

3.2 Handling Languages

`\smg@select@language` This macro selects one of the registered languages by its langauage 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.

```
⟨ltxml⟩RawTeX('
```

[2]EdNote: @DG: We also want to do that in LaTeXML

[3]EdNote: implement other schemes, e.g. the onefile scheme.

```
⟨*sty | ltxml⟩
\newcommand\smg@select@lang{}
\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 languge code ⟨*lang*⟩ by extending the `\smg@select@language` macro.

```
\newcommand\smg@register@language[2]%
{\@ifundefined{smul@#1@loaded}{}{\appto\smg@select@lang%
{\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.

```
\smg@register@language{af}{afrikaans}
\smg@register@language{de}{ngerman}
\smg@register@language{fr}{french}%
\smg@register@language{he}{hebrew}
\smg@register@language{hu}{hungarian}
\smg@register@language{id}{indonesian}
\smg@register@language{ms}{malay}
\smg@register@language{nn}{nynorsk}
\smg@register@language{pt}{portuguese}
\smg@register@language{ru}{russian}
\smg@register@language{uk}{ukrainian}
\smg@register@language{en}{english}
\smg@register@language{es}{spanish}
\smg@register@language{sq}{albanian}
\smg@register@language{bg}{bulgarian}
\smg@register@language{ca}{catalan}
\smg@register@language{hr}{croatian}
\smg@register@language{cs}{czech}
\smg@register@language{da}{danish}
\smg@register@language{nl}{dutch}
\smg@register@language{eo}{esperanto}
\smg@register@language{et}{estonian}
\smg@register@language{fi}{finnish}
\smg@register@language{ka}{georgian}
\smg@register@language{el}{greek}
\smg@register@language{is}{icelandic}
\smg@register@language{it}{italian}
\smg@register@language{la}{latin}
\smg@register@language{no}{norsk}
\smg@register@language{pl}{polish}
\smg@register@language{sr}{serbian}
\smg@register@language{sk}{slovak}
\smg@register@language{sl}{slovenian}
\smg@register@language{sv}{swedish}
\smg@register@language{th}{thai}
\smg@register@language{tr}{turkish}
```

```
\smg@register@language{vi}{vietnamese}
\smg@register@language{cy}{welsh}
\smg@register@language{hi}{hindi}
```

3.3 Language Bindings

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.

```
\newenvironment{modsig}[2][]{%
\def\@test{#1}\ifx\@test\@empty\begin{module}[id=#2]\else\begin{module}[id=#2,#1]\fi}
{\end{module}}
⟨*sty | ltxml⟩
⟨ltxml⟩ ');
```

modnl:*

```
⟨*sty⟩
\addmetakey{modnl}{load}
\addmetakey*{modnl}{title}
\addmetakey*{modnl}{creators}
\addmetakey*{modnl}{contributors}
⟨/sty⟩
⟨*ltxml⟩
DefKeyVal('modnl','title','Semiverbatim');
DefKeyVal('modnl','load','Semiverbatim');
DefKeyVal('modnl','creators','Semiverbatim');
DefKeyVal('modnl','contributors','Semiverbatim');
⟨/ltxml⟩
```

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

```
⟨*sty⟩
\newenvironment{modnl}[3][]{\metasetkeys{modnl}{#1}%
\smg@select@language{#3}%
\def\@test{#1}\ifx\@test\@empty\begin{module}[id=#2.#3]\else\begin{module}[id=#2.#3,#1]\fi%
\if@langfiles\importmodule[load=#2,ext=tex]{#2}\else
\ifx\modnl@load\@empty\importmodule{#2}\else\importmodule[ext=tex,load=\modnl@load]{#2}\fi%
\fi}
{\end{module}}
⟨/sty⟩
⟨*ltxml⟩
DefEnvironment('{modnl} OptionalKeyVals:modnl {}{}',
        "<omdoc:theory "
        . 'xml:id="#2.#3">'
        .   "?&defined(&GetKeyVal(#1,'creators'))(<dc:creator>&GetKeyVal(#1,'creators')</dc:cr
        .   "?&defined(&GetKeyVal(#1,'title'))(<dc:title>&GetKeyVal(#1,'title')</dc:title>)()"
        .   "?&defined(&GetKeyVal(#1,'contributors'))(<dc:contributor>&GetKeyVal(#1,'contribut
        .   "#body"
        . "</omdoc:theory>",
 afterDigestBegin=>sub {
   my ($stomach, $whatsit) = @_;
   my $keyval = $whatsit->getArg(1);
   my $signature = ToString($whatsit->getArg(2));
   if ($keyval) {
     # If we're not given load, AND the langfiles option is in effect,
     # default to #2
     if ((! $keyval->getValue('load')) && (LookupValue('smultiling_langfiles'))) {
       $keyval->setValue('load',$signature); }
     # Always load a TeX file
     $keyval->setValue('ext','tex'); }
   importmoduleI($stomach,$whatsit)});
⟨/ltxml⟩
⟨ltxml⟩1;
```