How To Write A Minimal LaTeXML Binding

Hang Yuan, Jinbo Zhang, Michael Kohlhase
Computer Science, Jacobs University Bremen

March 7, 2015

1 Introduction

LaTeXhas been widely used as a document processor among scholars, especially when one needs to use large quantities of mathematical representations. LaTeX is also a good choice for those who are meticulous about typographical quality of documents.

As a page formatting tool, the primary output format of the LaTeX formatter is PDF; which – with fixed page formats and limited interaction features – is only partially suited for usage in the modern web. The DLMF (Digital Library of Mathematical Functions) developed LaTeXML, a flexible, semantics-preserving LaTeX to XML converter to fix this.

However, for every LaTeX class and package used in a document LaTeXML needs a **LaTeXML binding** – a configuration file that specifies the XML counterpart of the LaTeX command sequences provided by the respective class or package.

Even though the LaTeXML distribution provides bindings for the most commonly used classes and packages, the availability of bindings is still the most severe bottleneck for LaTeXML. The LaTeXML documentation [Milb] is mostly written for developers and quite impenetrable for beginners.

To encourage binding development this how-to tutorial goes through the steps and pitfalls of creating a LaTeXML class binding from scratch. This tutorial does not cover advanced topics related to LaTeXML, for which we refer to the LaTeXML manual [Milb].

We have developed a minimal document class `mockDoc` as an example for this how-to and will go through it step-by-step. All necessary files (and the development version of this tutorial)are available from [Moc], but are also included in the appendix of this document for reference.

This how-to tutorial is structured as follows: Section 2 briefly reviews LaTeXML workflows and the files involved.

2 Using LaTeXML

In this tutorial we we assume a working installation of LaTeXML– see [Mila] for instructions – on a Unix-like system (Linux, Mac OS, etc.).

Given that, we use the command

```
latexmlc mockDoc.tex --format=XML --destination=mockDoc.xml --log=mockDoc.xml.log
```

for converting `mockDoc.tex` into `mockDoc.xml`. BOP:1

Note: Regarding LaTeXML installation, when you think you have finished installing LaTeXML, run a simple command:

```
latexml your_sample.tex
```

to test it. You should be able to see an XML interpretation of `your\_sample.tex` in screen immediately. Under some circumstances LaTeXML doesn't seem to work, maybe you fail to install the prerequisites such as `libxml2` or `libxslt`. EOP:1

[1] Old Part: MK: I do not think that this has any role here; what are you trying to achieve with this?

The conversion from LaTeX to XML is processed by LaTeXML. Basically LaTeXML maps the LaTeX markups to the XML markups, more specifically: macros, primitives and constructors.

2.1 Things We Need

source Here we use `mockDoc.tex` as a minimal example[2] see appendix **??**[3] EdN:2

EdN:3

(LaTeX class) we provide a LaTeX class `doc.cls` for reference; sometimes it is useful to generate PDF for proofreading. Also, the normal situation in developing LaTeXML bindings is that the class/package pre-exists. This file won't be illustrated in this tutorial.

LaTeXML binding the core issue of this tutorial. We use `doc.cls.ltxml` – Section **??** for a step-by-step explanation and appendix A.2 for the end result.

RelaxNG schema LaTeXML needs a RelaxNG schema to infer the output structure. We supply it in compact form (`mockDoc.rnc`; see appendix A.3, which can be converted to the XML form LaTeXML needs (`mockDoc.rng`) via `trang.jar`. The reason for writing `mockDoc.rnc` instead of `mockDoc.rng` is that, `mockDoc.rnc` is much shorter and easier to maintain.

After you have finished writing all the documents above, run the command mentioned before, and then you should be able to see the converted XML file for `mockDoc.tex`. In the following chapters we will explain how to construct `mockDoc.rnc` and `doc.cls.ltxml`[4]. EdN:4

These workflows can be automated via a Unix `makefile` (see appendix B), which re-generates everything when source files have changed. Then only need to issue the command:

```
make
```

BOP:5

you should be able to see the generated `mockDoc.xml` in your current directory. It should be something similar to your expected `mockDoc\_sample.xml`. EOP:5

3 The mockDoc Format

3.1 A minimal Document Format

Actually our `mockDoc`format is probably the smallest one in the world, it is only intended for this tutorial.

The LaTeX class only provides one environment: `document` and four macros: `\section`, `\subsection`, `\paragraph`, and `\newline`. A minimal example would be

Listing 1: A Minimal LaTeX Document

```
\documentclass{doc}
\begin{document}
        \section{A brief introduction about Shelley}
                Percy Bysshe Shelley (4 August 1792 -- 8 July 1822) was one of the
                    major English Romantic poets, and is regarded by some critics
                    as amongst the finest lyric poets in the English language.
        \section{Ode to the West Wind (partial)}
                \subsection{I}
                        \paragraph{1.}
                                O wild West Wind, thou breath of Autumn's being, \
                                    newline
                                Thou, from whose unseen presence the leaves dead \
                                    newline
                                Are driven, like ghosts from an enchante fleeing,
                        \paragraph{2.}
```

[2] EdNote: MK: make a minimal one, use that here
[3] EdNote: make other references
[4] EdNote: MK: I think we should rename doc.cls and doc.cls.ltxml to mockDoc.cls, ...
[5] Old Part: put somewhere else

```
                                Yellow, and black, and pale, and hectic red,\
                                    newline
                                Pestilence-stricken multitudes: O thou, \newline
                                Who chariotest to their dark wintry bed
                        \paragraph{3.}
                                The winged seeds, where they lie cold and low,\
                                    newline
                                Each like a corpse within its grave, until \newline
                                Thine azure sister of the Spring shall blow
                        \paragraph{4.}
                                Her clarion o'er the dreaming earth, and fill \
                                    newline (Driving sweet buds like flocks to feed
                                     in air) \newline With living hues and odours
                                    plain and hill:
                        \paragraph{5.}
                                Wild Spirit, which art moving everywhere; \newline
                                    Destroyer and Preserver; hear, O hear!
                \subsection{II}
                        \paragraph{1.}
                                Thou on whose stream, 'mid the steep sky's
                                    commotion, \newline
                                Loose clouds like Earth's decaying leaves are shed,
                                     \newline
                                Shook from the tangled boughs of Heaven and Ocean,
                        \paragraph{2.}
                                Angels of rain and lightning: there are spread \
                                    newline
                                On the blue surface of thine airy surge, \newline
                                Like the bright hair uplifted from the head
                        \paragraph{3.}
                                Of some fierce Maenad, even from the dim verge \
                                    newline
                                Of the horizon to the zenith's height, \newline
                                The locks of the approaching storm. Thou dirge
                        \paragraph{4.}
                                Of the dying year, to which this closing night \
                                    newline
                                Will be the dome of a vast sepulchre \newline
                                Vaulted with all thy congregated might
                        \paragraph{5.}
                                Of vapours, from whose solid atmosphere \newline
                                Black rain, and fire, and hail will burst: O hear!
\end{document}
%%% Local Variables:
%%% mode: latex
%%% TeX-master: t
%%% End:
```

We want to use this document class for generating XML documents, which use the five elements `document`, `\section`, `\subsection`, `\paragraph`, and `\newline`. The XML document corresponding to the LaTeX document from Listing 1 is

Listing 2: A Minimal LaTeX Document

```
<?xml version="1.0" encoding="UTF-8"?>
<?latexml searchpaths="/home/la_stravaganza/repos/mockDoc/secondTrial"?>
<?latexml class="doc"?>
<?latexml RelaxNGSchema="mockDoc.rng"?>
<mock:document xmlns:mock="https://kwarc.info/projects/mockDoc">
  <mock:section>
    <mock:title>A brief introduction about Shelley</mock:title>
    <mock:p>Percy Bysshe Shelley (4 August 1792     8 July 1822) was one of the
        major English Romantic poets, and is regarded by some critics as amongst
        the finest lyric poets in the English language.
</mock:p>
  </mock:section>
  <mock:section>
    <mock:title>Ode to the West Wind (partial)</mock:title>
```

```
    <mock:subsection>
      <mock:title>I</mock:title>
      <mock:paragraph>
        <mock:title>1.</mock:title>
        <mock:p>
O wild West Wind, thou breath of Autumn s being, <mock:break/>Thou, from whose
    unseen presence the leaves dead <mock:break/>Are driven, like ghosts from an
    enchante fleeing,
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>2.</mock:title>
        <mock:p>
Yellow, and black, and pale, and hectic red,<mock:break/>Pestilence-stricken
    multitudes: O thou, <mock:break/>Who chariotest to their dark wintry bed
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>3.</mock:title>
        <mock:p>
The winged seeds, where they lie cold and low,<mock:break/>Each like a corpse
    within its grave, until <mock:break/>Thine azure sister of the Spring shall
    blow
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>4.</mock:title>
        <mock:p>
Her clarion o er the dreaming earth, and fill <mock:break/>(Driving sweet buds
    like flocks to feed in air) <mock:break/>With living hues and odours plain and
    hill:
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>5.</mock:title>
        <mock:p>
Wild Spirit, which art moving everywhere; <mock:break/>Destroyer and Preserver;
    hear, O hear!
</mock:p>
      </mock:paragraph>
    </mock:subsection>
    <mock:subsection>
      <mock:title>II</mock:title>
      <mock:paragraph>
        <mock:title>1.</mock:title>
        <mock:p>
<!-- %**** mockDoc.tex Line 25 **** -->Thou on whose stream,  mid the steep
    sky s commotion, <mock:break/>Loose clouds like Earth s decaying leaves are
     shed, <mock:break/>Shook from the tangled boughs of Heaven and Ocean,
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>2.</mock:title>
        <mock:p>
Angels of rain and lightning: there are spread <mock:break/>On the blue surface of
    thine airy surge, <mock:break/>Like the bright hair uplifted from the head
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>3.</mock:title>
        <mock:p>
Of some fierce Maenad, even from the dim verge <mock:break/>Of the horizon to the
    zenith s height, <mock:break/>The locks of the approaching storm. Thou dirge
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>4.</mock:title>
```

```
        <mock:p>
Of the dying year, to which this closing night <mock:break/>Will be the dome of a
    vast sepulchre <mock:break/>Vaulted with all thy congregated might
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>5.</mock:title>
        <mock:p>
Of vapours, from whose solid atmosphere <mock:break/>Black rain, and fire, and hail
     will burst: O hear!
</mock:p>
      </mock:paragraph>
    </mock:subsection>
  </mock:section>
</mock:document>
```

[6] EdN:6

Note the typical format-specific differences between the presentation-oriented LaTeX and more content-oriented XML formats. The sectioning is conveyed by macros in LaTeX – only giving the start cues (here the numbered section headings) – whereas the XML has start and end tags[7] EdN:7

After you link `mockDoc.tex` and `doc.cls.ltxml` by changing your document class in your `mockDoc.tex` into your LaTeXML binding name, in our case,"doc". Put `doc.cls.ltxml` and `mockDoc.tex` in the same folder, LaTeXML will load your binding file automatically, when it tries to do the conversion. BOP:8

EOP:8

3.2 The RelaxNG Schema

Schema is a crucial document that decides how `mockDoc.xml` is constructed. When you are creating your own schema[1], one good approach to test this is to create your expected `mockDoc\_sample.xml` by hand, according to your `mockDoc.tex`, then compare `mockDoc\_sample.xml` with the generated `mockDoc.xml`. You can easily accomplish this by using *emacs nxml mode*[2], in which you have the freedom to write your expected `mockDoc.xml`, while validating your `mockDoc.xml` at the same time. If validation fails, you can see the error message instantly, such that you can debug your `mockDoc.xml` or schema accordingly.[9] EdN:9

In our `mockDoc.rnc`:

```
document = element document {p, section*}
section = element section {title,(p |subsection)*}
```

you can easily see that, under a `document`, there can be either `p` or `section`, and under a `section` there can be a `title` followed by `p` or a `title` followed by a `subsection`. This is because in the first section in `mockDoc.tex`:

```
\section{A brief introduction about Shelley}
    Percy Bysshe Shelley (4 August 1792 -- 8 July 1822)...
```

there is no `subsection` but texts directly. But in the other `section`s, there are `subsection`s. In your schema you need to consider all kinds of possible hierarchy of your elements.

4 How to Create a LaTeXML Binding

We now come to the central part of our tutorial: writing the LaTeXML binding itself. Generally, a LaTeXML binding file is a Perl module – and therefore underlies Perl syntax, but special high-level

[6] EdNote: MK: actually, we should make an idealized minimal XML example by remove all the XML-isms from the generated one and show it here.

[7] EdNote: MK: are there more? Here would be the place to discuss them.

[8] Old Part: MK: I do not understand this, what do you want to say here?

[1] Before you write your expected xml and schema, having a look at the links below can be beneficial: `http://relaxng.org/compact-tutorial-20030326.html`; `http://www.w3schools.com/xml/`.

[2] Here is a tutorial about Emacs nxml mode: `http://www.emacswiki.org/emacs/NxmlMode`

[9] EdNote: MK: convert all footnotes to citations!

commands simplify expressing the LaTeX-to-XML relation.

4.1 Basic structure

Since LaTeX binding is a perl module, we need to initialize a binding file by adding the followings in the beginning of `doc.cls.ltxml`:

```
package LaTeXML::Package::Pool;
use strict;
use LaTeXML::Package;
use warnings;
```

At the end of `doc.cls.ltxml`, don't forget to include

```
1;
```

to make sure that perl works properly.

4.2 Configure namespace

With:

```
RegisterNamespace('mock'=>"https://kwarc.info/projects/mockDoc");
RelaxNGSchema("mockDoc.rng",'mock'=>"https://kwarc.info/projects/mockDoc");
```

We declared the namespace associated the prefix `mock` with the namespace.[10] EdN:10

4.3 Linebreaks

The next task is to teach LaTeXML new commands used in `mockDoc.tex`. Here is an example:

```
DefConstructor('\newline',"<mock:break/>");
```

This line defines how LaTeXML interprets `\newline`, as you see, LaTeXML will translate `\newline` to `<mock:break/>` in `mockDoc.xml`.

4.4 Sectioning

When dealing with `section`, things get a little tricky, with:

```
DefConstructor('\section{}', "<mock:section><mock:title>#1</mock:title>");
```

we defined `\section`. But, think about the closing tags. In `mockDoc.tex`, we declared where the `\section` starts and where the next `\section` starts, nevertheless, we never wrote something like "Now close this section". Here is why we need `mockDoc.rnc`. This schema file tells LaTeXML what the structure of our document, and with:

```
Tag('mock:section', autoClose=>1);
```

LaTeXML will close the section tags (i.e, adding `</mock:section>`) whenever needed.

4.5 The Document Environment

You may think something like:

```
DefEnvironment('{document}', "<mock:document>#body</mock:document>");
```

is enough for defining `document` environment. You can try it, you will find that all spaces disappear. What we actually wrote in `doc.cls.ltxml` is:

```
DefEnvironment('{document}', "<mock:document>#body</mock:document>", beforeDigest
    => sub { AssignValue(inPreamble => 0); });
```

[10] EdNote: MK: we have to explain the why more, not just what to do; that is generally the case.

This code can prevent the error mentioned before, however, the mechanism of the `beforeDigest` part is out of our discussion in this tutorial.

For an environment, we don't need care about auto-closing, since an environment is always like

```
\begin{*environment-name*}
content...
\end{*environment-name*}
```

where `\end\{*environment-name*\}` will indicate where to close the tags.

4.6 Auto-opening for Paragraphs

Since we also want to write some texts directly under `document`, without any `section`. At this circumstance, we need auto-open for `p`:

```
Tag('mock:p', autoOpen=>1);
```

which will surround such texts.

5 Conclusion

[11] EdN:11

References

[Mila] Bruce R. Miller. *Get LaTeXML*. URL: `http://dlmf.nist.gov/LaTeXML/get.html` (visited on 02/22/2015).

[Milb] Bruce R. Miller. *LaTeXML The Manual*. URL: `http://dlmf.nist.gov/LaTeXML/manual.pdf` (visited on 02/22/2015).

[Moc] *mockDoc, a minimal LaTeXML class binding and HowTo*. URL: `https://github.com/angerhang/mockDoc` (visited on 03/07/2015).

[11] EdNote: MK: say something here

A Appendix

A.1 The mockDoc Class

```
% File: doc.cls
% Author: Jinbo Zhang
% Date: 3 Feb, 2015

\NeedsTeXFormat{LaTeX2e}
\ProvidesClass{doc}
\RequirePackage{ifthen}

\renewcommand\normalsize{\fontsize{10pt}{12pt}\selectfont}
\setlength{\textwidth}{6.5in}
\setlength{\textheight}{8in}

\newcommand\large{\@setfontsize\large\@xiipt{14}}
\newcommand\Large{\@setfontsize\Large\@xivpt{18}}


% define \paragraph
\newcommand{\paragraph}[1]{
        \newline\newline
        \bfseries #1
        \normalfont
}

% define \section
\newcounter{SectionCount}
\newcommand{\section}[1]{
        \ifthenelse{\value{SectionCount}=0}{}{\newline\newline\newline}
        \Large
        \stepcounter{SectionCount}
        \noindent\bfseries\arabic{SectionCount}\hspace{4mm} #1
        \normalfont
        \newline\newline
}

% define \subsection
\newcounter{SubCount}[SectionCount]
\newcommand{\subsection}[1]{
        \ifthenelse{\value{SubCount}=0}{}{\newline\newline}
        \large
        \stepcounter{SubCount}
        \bfseries\arabic{SectionCount}.\arabic{SubCount}\hspace{3mm} #1
        \normalfont
}


\endinput
```

A.2 The mockDoc Class Binding

```
package LaTeXML::Package::Pool;
use strict;
use LaTeXML::Package;
use warnings;

#Document Structure
RegisterNamespace('mock'=>"https://kwarc.info/projects/mockDoc");
RelaxNGSchema("mockDoc.rng",'mock'=>"https://kwarc.info/projects/mockDoc");

#---------------------------------------------------------------------
DefEnvironment('{document}', "<mock:document>#body</mock:document>", beforeDigest
    => sub { AssignValue(inPreamble => 0); });
DefConstructor('\section{}', "<mock:section><mock:title>#1</mock:title>");
DefConstructor('\subsection{}', "<mock:subsection><mock:title>#1</mock:title>");
```

```
DefConstructor('\paragraph{}', "<mock:paragraph><mock:title>#1</mock:title><mock:p>
    ");
DefConstructor('\newline', "<mock:break/>");

#autoClose
Tag('mock:paragraph', autoClose=>1);
Tag('mock:section', autoClose=>1);
Tag('mock:subsection', autoClose=>1);
Tag('mock:p', autoClose=>1);
Tag('mock:p', autoOpen=>1);

#make sure Perl work
1;
```

A.3 mockDoc RelaxNG schema

```
default namespace md = "https://kwarc.info/projects/mockDoc"

start = document
document = element document {p, section*}
section = element section {title,(p |subsection)*}
subsection = element subsection {title,paragraph*}
paragraph = element paragraph { title, p }
title = element title { text }
p = element p { (text|element break { empty })*}
```

A.4 Generated XML

```
<?xml version="1.0" encoding="UTF-8"?>
<?latexml searchpaths="/home/la_stravaganza/repos/mockDoc/secondTrial"?>
<?latexml class="doc"?>
<?latexml RelaxNGSchema="mockDoc.rng"?>
<mock:document xmlns:mock="https://kwarc.info/projects/mockDoc">
  <mock:section>
    <mock:title>A brief introduction about Shelley</mock:title>
    <mock:p>Percy Bysshe Shelley (4 August 1792     8 July 1822) was one of the
        major English Romantic poets, and is regarded by some critics as amongst
        the finest lyric poets in the English language.
</mock:p>
  </mock:section>
  <mock:section>
    <mock:title>Ode to the West Wind (partial)</mock:title>
    <mock:subsection>
      <mock:title>I</mock:title>
      <mock:paragraph>
        <mock:title>1.</mock:title>
        <mock:p>
O wild West Wind, thou breath of Autumn s being, <mock:break/>Thou, from whose
    unseen presence the leaves dead <mock:break/>Are driven, like ghosts from an
    enchante fleeing,
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>2.</mock:title>
        <mock:p>
Yellow, and black, and pale, and hectic red,<mock:break/>Pestilence-stricken
    multitudes: O thou, <mock:break/>Who chariotest to their dark wintry bed
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>3.</mock:title>
        <mock:p>
The winged seeds, where they lie cold and low,<mock:break/>Each like a corpse
    within its grave, until <mock:break/>Thine azure sister of the Spring shall
    blow
</mock:p>
      </mock:paragraph>
```

```
      <mock:paragraph>
        <mock:title>4.</mock:title>
        <mock:p>
Her clarion o  er the dreaming earth, and fill <mock:break/>(Driving sweet buds
    like flocks to feed in air) <mock:break/>With living hues and odours plain and
    hill:
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>5.</mock:title>
        <mock:p>
Wild Spirit, which art moving everywhere; <mock:break/>Destroyer and Preserver;
    hear, O hear!
</mock:p>
      </mock:paragraph>
    </mock:subsection>
    <mock:subsection>
      <mock:title>II</mock:title>
      <mock:paragraph>
        <mock:title>1.</mock:title>
        <mock:p>
<!-- %**** mockDoc.tex Line 25 **** -->Thou on whose stream,  mid the steep
    sky s commotion, <mock:break/>Loose clouds like Earth s decaying leaves are
     shed, <mock:break/>Shook from the tangled boughs of Heaven and Ocean,
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>2.</mock:title>
        <mock:p>
Angels of rain and lightning: there are spread <mock:break/>On the blue surface of
    thine airy surge, <mock:break/>Like the bright hair uplifted from the head
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>3.</mock:title>
        <mock:p>
Of some fierce Maenad, even from the dim verge <mock:break/>Of the horizon to the
    zenith s height, <mock:break/>The locks of the approaching storm. Thou dirge
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>4.</mock:title>
        <mock:p>
Of the dying year, to which this closing night <mock:break/>Will be the dome of a
    vast sepulchre <mock:break/>Vaulted with all thy congregated might
</mock:p>
      </mock:paragraph>
      <mock:paragraph>
        <mock:title>5.</mock:title>
        <mock:p>
Of vapours, from whose solid atmosphere <mock:break/>Black rain, and fire, and hail
     will burst: O hear!
</mock:p>
      </mock:paragraph>
    </mock:subsection>
  </mock:section>
</mock:document>
```

B A Makefile for Automation

```
#makefile for using latexml and pdflatex to generate *.pdf and *.xml
#declaration of variables
#set .tex as source. In our case only mockDoc is available
#name .xml and .pdf based on .tex
SRC = $(shell ls *.tex)
XML = $(SRC:%.tex=%.xml)
PDF = $(SRC:%.tex=%.pdf)
```

```
all: $(XML) $(PDF)

mockDoc.rng: mockDoc.rnc
	java -jar trang.jar -I rnc -O rng mockDoc.rnc mockDoc.rng

#the codes below follow the usage of variables mentioned above
#   $@   : object filename.   $ <   :source file name
$(XML): %.xml: %.tex mockDoc.rng doc.cls.ltxml
	latexmlc $<  --format=XML --destination=$@ --log=$@.log

$(PDF): %.pdf: %.tex doc.cls
	xelatex $<
```