

# How To Write A Minimal L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> Binding

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

L<sup>A</sup>T<sub>E</sub>X has been widely used as a word processing tool among scholars, especially when one needs to use large quantities of mathematical representations. L<sup>A</sup>T<sub>E</sub>X is also a good choice for those who are meticulous about typographical quality of documents. However, L<sup>A</sup>T<sub>E</sub>X lacks a converter tool to XML. The DLMF (Digital Library of Mathematical Functions) developed L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub>, trying to make a new typesetting system that allows users to be able to focus more on the contents, not the style, by providing extensive ways of customizations. In order to achieve this goal, building up the bindings is crucial, and yet L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> seems fairly unfathomable for beginners. We want to make it easier for those who want to pick up using L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> in the future, by going through how to construct a minimal L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> binding step by step. We will use *mockDoc*<sup>1</sup> as a sample in this tutorial. This tutorial does not cover advanced topics related to L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub>, and thus if you are interested in the general theories, please explore the L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> Manual [Mil] to better comprehend how the theories are implemented.

## 1 Using LaTeXML

We are going to talk about various aspects of L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub>, and then we will move onto the workflow of creating your first L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> binding. In this tutorial, we use the command:

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

for converting *mockDoc.tex* into *mockDoc.xml*.

**Note:** Regarding L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> installation, when you think you have finished installing L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub>, run a simple command:

```
1 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 L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> doesn't seem to work, maybe you fail to install the prerequisites such as *libxml2* or *libxslt*<sup>2</sup>.

## 2 How to Create a LaTeXML Binding

The conversion from L<sup>A</sup>T<sub>E</sub>X to XML is processed by L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub>. Basically L<sup>A</sup>T<sub>E</sub>X<sub>ML</sub> maps the L<sup>A</sup>T<sub>E</sub>X markups to the XML markups, more specifically: macros, primitives and constructors.

### 2.1 Things We Need

As your source file. You can write down whatever you want. A minimal example<sup>1</sup> can be found in appendix A<sup>2</sup>

EdN:1  
EdN:2

For XeL<sup>A</sup>T<sub>E</sub>X, which essentially helps you to see what *mockDoc.tex* file looks like in a pdf format. This file won't be illustrated in this tutorial.

---

<sup>1</sup>mockDoc project in Github: <https://github.com/angerhang/mockDoc>

<sup>2</sup>Please visit <http://dlmf.nist.gov/LaTeXML/get.html> for more information.

<sup>1</sup>EdNOTE: MK: make a minimal one, use that here

<sup>2</sup>EdNOTE: make other references

L<sup>A</sup>T<sub>E</sub>X<sub>XML</sub> binding, the core file of this tutorial. `doc.cls.ltxml` is similar to `doc.cls`, but used for the conversion to other formats.

The schema in compact form, which defines the structure of `mockDoc.tex`, crucial for executing tasks like placing the tags correctly and auto closing the tags when needed.

L<sup>A</sup>T<sub>E</sub>X<sub>XML</sub> cannot process the compact form schema, therefore you need `trang.jar` to convert `mockDoc.rnc` into `mockDoc.rng`. 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`.

## 2.2 RelaxNG Schema

Schema is a crucial document that decides how `mockDoc.xml` is constructed. When you are creating your own schema<sup>3</sup>, 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*<sup>4</sup>, 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.

In our `mockDoc.rnc`:

```
1 document = element document {p, section*}
2 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`:

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

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

## 2.3 Minimal L<sup>A</sup>T<sub>E</sub>X<sub>XML</sub>

Actually this binding is not the smallest one in the world, in `doc.cls.ltxml` we covered:

**1 environment:** `document`

**4 control sequences:** `\textbackslash section`, `\textbackslash subsection`, `\textbackslash paragraph`, `\textbackslash newline`

After you link `mockDoc.tex` and `doc.cls.ltxml` by changing your document class in your `mockDoc.tex` into your L<sup>A</sup>T<sub>E</sub>X<sub>XML</sub> binding name, in our case, “`doc`”. Put `doc.cls.ltxml` and `mockDoc.tex` in the same folder, L<sup>A</sup>T<sub>E</sub>X<sub>XML</sub> will load your binding file automatically, when it tries to do the conversion.

### 2.3.1 Basic structure

Since L<sup>A</sup>T<sub>E</sub>X binding is a perl module, we need to initialize a binding file by adding the followings in the beginning of `doc.cls.ltxml`:

```
1 package LaTeXXML::Package::Pool;
2 use strict;
3 use LaTeXXML::Package;
4 use warnings;
```

---

<sup>3</sup>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/>.

<sup>4</sup>Here is a tutorial about Emacs nxml mode: <http://www.emacswiki.org/emacs/NxmlMode>

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

```
1 1;
```

to make sure that perl works properly.

### 2.3.2 Configure namespace

With:

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

We declared the namespace associated the prefix `mock` with the namespace.

### 2.3.3 Define `\newline`

The next task is to teach  $\text{\LaTeX}$ XML new commands used in `mockDoc.tex`. Here is an example:

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

This line defines how  $\text{\LaTeX}$ XML interprets `\textbackslash newline`, as you see,  $\text{\LaTeX}$ XML will translate `\textbackslash newline` to `<mock:break/>` in `mockDoc.xml`.

### 2.3.4 Define `\section`

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

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

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

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

$\text{\LaTeX}$ XML will close the section tags (i.e, adding `</mock:section>`) whenever needed.

### 2.3.5 Define document

You may think something like:

```
1 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:

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

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 autoclosing, since an environment is always like

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

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

### 2.3.6 Autoopen for `p`

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

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

which will surround such texts.

### 3 Conclusion

Thank you for following this tutorial to the end. After processing the `makefile` (see F), with command:

```
1 make
```

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`.

### References

`mockDoc.tex` `doc.cls` `ls-ltx.xml` [Mil] — Bruce R. Miller. *LaTeXML The Manual*. <http://dlmf.nist.gov/LaTeXML/manual.pdf>.

## A mockDoc Example

```
1 \documentclass{doc}
2 \begin{document}
3   \section{A brief introduction about Shelley}
4     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.
5   \section{Ode to the West Wind (partial)}
6     \subsection{I}
7       \paragraph{1.}
8         O wild West Wind, thou breath of Autumn's being, \
9           \newline
10          Thou, from whose unseen presence the leaves dead \
11            \newline
12          Are driven, like ghosts from an enchanter fleeing,
13        \paragraph{2.}
14          Yellow, and black, and pale, and hectic red, \
15            \newline
16          Pestilence-stricken multitudes: O thou, \newline
17          Who chariotest to their dark wintry bed
18        \paragraph{3.}
19          The winged seeds, where they lie cold and low, \
20            \newline
21          Each like a corpse within its grave, until \newline
22          Thine azure sister of the Spring shall blow
23        \paragraph{4.}
24          Her clarion o'er the dreaming earth, and fill \
25            \newline (Driving sweet buds like flocks to feed
26            in air) \newline With living hues and odours
27            plain and hill:
28        \paragraph{5.}
29          Wild Spirit, which art moving everywhere; \newline
30          Destroyer and Preserver; hear, O hear!
31      \subsection{II}
32        \paragraph{1.}
33          Thou on whose stream, 'mid the steep sky's
34            commotion, \newline
35          Loose clouds like Earth's decaying leaves are shed,
36            \newline
37          Shook from the tangled boughs of Heaven and Ocean,
38        \paragraph{2.}
39          Angels of rain and lightning: there are spread \
40            \newline
41          On the blue surface of thine airy surge, \newline
42          Like the bright hair uplifted from the head
43        \paragraph{3.}
44          Of some fierce Maenad, even from the dim verge \
45            \newline
46          Of the horizon to the zenith's height, \newline
47          The locks of the approaching storm. Thou dirge
48        \paragraph{4.}
49          Of the dying year, to which this closing night \
50            \newline
51          Will be the dome of a vast sepulchre \newline
52          Vaulted with all thy congregated might
53        \paragraph{5.}
54          Of vapours, from whose solid atmosphere \newline
55          Black rain, and fire, and hail will burst: O hear!
56 \end{document}
57 %%% Local Variables:
58 %%% mode: latex
59 %%% TeX-master: t
60 %%% End:
```

## B The mockDoc Class

```

1  % File: doc.cls
2  % Author: Jinbo Zhang
3  % Date: 3 Feb, 2015
4
5  \NeedsTeXFormat{LaTeX2e}
6  \ProvidesClass{doc}
7  \RequirePackage{ifthen}
8
9  \renewcommand\normalsize{\fontsize{10pt}{12pt}\selectfont}
10 \setlength{\textwidth}{6.5in}
11 \setlength{\textheight}{8in}
12
13 \newcommand\large{\@setfontsize\large\@xipt{14}}
14 \newcommand\Large{\@setfontsize\Large\@xivpt{18}}
15
16
17 % define \paragraph
18 \newcommand{\paragraph}[1]{
19     \newline\newline
20     \bfseries #1
21     \normalfont
22 }
23
24 % define \section
25 \newcounter{SectionCount}
26 \newcommand{\section}[1]{
27     \ifthenelse{\value{SectionCount}=0}{\{\}\newline\newline\newline}
28     \Large
29     \stepcounter{SectionCount}
30     \noindent\bfseries\arabic{SectionCount}\hspace{4mm} #1
31     \normalfont
32     \newline\newline
33 }
34
35 % define \subsection
36 \newcounter{SubCount}[SectionCount]
37 \newcommand{\subsection}[1]{
38     \ifthenelse{\value{SubCount}=0}{\{\}\newline\newline}
39     \large
40     \stepcounter{SubCount}
41     \bfseries\arabic{SectionCount}.\arabic{SubCount}\hspace{3mm} #1
42     \normalfont
43 }
44
45
46 \endinput

```

## C The mockDoc Class Binding

```

1  package LaTeXML::Package::Pool;
2  use strict;
3  use LaTeXML::Package;
4  use warnings;
5
6  #Document Structure
7  RegisterNamespace('mock'=>"https://kwarc.info/projects/mockDoc");
8  RelaxNGSchema("mockDoc.rng", 'mock'=>"https://kwarc.info/projects/mockDoc");
9
10 #-----
11 DefEnvironment('{document}', "<mock:document>#body</mock:document>", beforeDigest
    => sub { AssignValue(inPreamble => 0); });
12 DefConstructor('\section{', "<mock:section><mock:title>#1</mock:title>");
13 DefConstructor('\subsection{', "<mock:subsection><mock:title>#1</mock:title>");
14 DefConstructor('\paragraph{', "<mock:paragraph><mock:title>#1</mock:title><mock:p>
    ");
15 DefConstructor('\newline', "<mock:break/>");
16

```

```

17 #autoClose
18 Tag('mock:paragraph', autoClose=>1);
19 Tag('mock:section', autoClose=>1);
20 Tag('mock:subsection', autoClose=>1);
21 Tag('mock:p', autoClose=>1);
22 Tag('mock:p', autoOpen=>1);
23
24 #make sure Perl work
25 1;

```

## D mockDoc RelaxNG schema

```

1 default namespace md = "https://kwarc.info/projects/mockDoc"
2
3 start = document
4 document = element document {p, section*}
5 section = element section {title,(p |subsection)*}
6 subsection = element subsection {title,paragraph*}
7 paragraph = element paragraph { title, p }
8 title = element title { text }
9 p = element p { (text|element break { empty })*}

```

## E Generate XML

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?latexml searchpaths="/home/la_stravaganza/repos/mockDoc/secondTrial"?>
3 <?latexml class="doc"?>
4 <?latexml RelaxNGSchema="mockDoc.rng"?>
5 <mock:document xmlns:mock="https://kwarc.info/projects/mockDoc">
6   <mock:section>
7     <mock:title>A brief introduction about Shelley</mock:title>
8     <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.
9   </mock:p>
10  </mock:section>
11  <mock:section>
12    <mock:title>Ode to the West Wind (partial)</mock:title>
13    <mock:subsection>
14      <mock:title>I</mock:title>
15      <mock:paragraph>
16        <mock:title>1.</mock:title>
17        <mock:p>
18          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
          enchanter fleeing,
19        </mock:p>
20        </mock:paragraph>
21        <mock:paragraph>
22          <mock:title>2.</mock:title>
23          <mock:p>
24            Yellow, and black, and pale, and hectic red,<mock:break/>Pestilence-stricken
            multitudes: O thou, <mock:break/>Who chariotest to their dark wintry bed
25          </mock:p>
26          </mock:paragraph>
27          <mock:paragraph>
28            <mock:title>3.</mock:title>
29            <mock:p>
30              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
31            </mock:p>
32            </mock:paragraph>
33            <mock:paragraph>
34              <mock:title>4.</mock:title>

```

```

35     <mock:p>
36 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:
37 </mock:p>
38     </mock:paragraph>
39     <mock:paragraph>
40     <mock:title>5.</mock:title>
41     <mock:p>
42 Wild Spirit, which art moving everywhere; <mock:break/>Destroyer and Preserver;
    hear, O hear!
43 </mock:p>
44     </mock:paragraph>
45 </mock:subsection>
46 <mock:subsection>
47     <mock:title>II</mock:title>
48     <mock:paragraph>
49     <mock:title>1.</mock:title>
50     <mock:p>
51 <!-- ***** 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,
52 </mock:p>
53     </mock:paragraph>
54     <mock:paragraph>
55     <mock:title>2.</mock:title>
56     <mock:p>
57 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
58 </mock:p>
59     </mock:paragraph>
60     <mock:paragraph>
61     <mock:title>3.</mock:title>
62     <mock:p>
63 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
64 </mock:p>
65     </mock:paragraph>
66     <mock:paragraph>
67     <mock:title>4.</mock:title>
68     <mock:p>
69 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
70 </mock:p>
71     </mock:paragraph>
72     <mock:paragraph>
73     <mock:title>5.</mock:title>
74     <mock:p>
75 Of vapours, from whose solid atmosphere <mock:break/>Black rain, and fire, and hail
    will burst: O hear!
76 </mock:p>
77     </mock:paragraph>
78 </mock:subsection>
79 </mock:section>
80 </mock:document>

```

## F A Makefile for Automation

```

1 #makefile for using latexml and pdflatex to generate *.pdf and *.xml
2 #declaration of variables
3 #set .tex as source. In our case only mockDoc is available
4 #name .xml and .pdf based on .tex
5 SRC = $(shell ls *.tex)
6 XML = $(SRC:%.tex=%.xml)
7 PDF = $(SRC:%.tex=%.pdf)
8
9 all: $(XML) $(PDF)

```



```
10
11 mockDoc.rng: mockDoc.rnc
12     java -jar trang.jar -I rnc -O rng mockDoc.rnc mockDoc.rng
13
14 #the codes below follow the usage of variables mentioned above
15 #  $@  : object filename.    $ <  :source file name
16 $(XML): %.xml: %.tex mockDoc.rng doc.cls.ltxml
17     latexmlc $< --format=XML --destination=$@ --log=$@.log
18
19 $(PDF): %.pdf: %.tex doc.cls
20     xelatex $<
```