

From XML to T_EX,
a short overview
of available methods

Simon Pepping

EuroT_EX 2001

27 September 2001

1. Directly from XML to T_EX
with ConT_EXt

by Hans Hagen
or Berend de Boer

2. Directly from XML to T_EX with `xmltex`

Formatting of an XML file:
specify markup
for opening and closing
of each element:

```
\XMLelement{element name}%  
  {attribute commands}%  
  {start element or \xmlgrab}{end element}  
\XMLattribute{attribute name}%  
  {attribute control sequence}
```

See eurotex.xmt

Package and file specification:

`xmltex.cfg` or `program.cfg`:

`\NAMESPACE{URI}{xmt-file}`

`\PUBLIC{FPI}{file}`

`\SYSTEM{URI}{file}`

`\NAME{element-name}{xmt-file}`

`\XMLNS{element-name}{URI}`

Run with `xmltex`:

```
(pdf)xmltex program.xml
```

Run with \LaTeX :

Make tex file `program.tex`:

```
\def\xmlfile{program.xml}  
\input xmltex.tex  
\end{document}
```

Command: `(pdf)latex program.tex`

See Proceedings TUG2000,
TUGboat 21, No. 3, pp. 193-199.

3. The canonical way:
XML + XSL (T + FO)
+ formatter

3.a) XSLT

by Berend de Boer

3.b) XSL FO

Processing model

(see [two-processes.html](#)):

XML file + XSL stylesheet

XSLT
→ FOT

XSL formatter
→ rendered document

A FOT (Formatting Object Tree)
is an abstract description
of the page layout

From XML to FOT via XSLT:

`program-tex.xsl`

Note slightly modified structure
of `program.xml`

/: apply templates

program: page layout setup

day:

- refer to page layout
- start page sequence
- set header and footer
- start flow for body
- set title in a block-container containing a block for each line
- start table
- column setup
- apply templates
- close table

day with mode rh:
used in header and footer

item:

- start table row
- table cell with block with time
- apply templates
- close table row

break:

- as item
- instead of applying templates:
table cell spanning two columns
with block with type of break

opening: **as** break

outing:

- much like break
- the word 'Outing' in inline FO
to get italics
- apply templates for the description
of the outing

presentation (in item):

- determine how many children
in the parent item,
and set the number of
columns spanned accordingly
- start table cell
- start block
- apply templates
for text of presentation
- close block
- close table cell

discussion (in item):

- as presentation
- the word 'Discussion' in inline FO
to get italics

author (in presentation):

- open inline FO to get italics
- apply templates for author name
- close inline FO
- add comma

title (in presentation):

- apply templates for text of title

Create FOT with XSL processor:

- xalan from xml apache,
- xt from James Clark

Command:

```
java -cp path-to-xalan.jar:\  
path-to-xerces.jar \  
org.apache.xalan.xslt.Process \  
-in xml-file -xsl xsl-file -out fot-file
```

3.c) Passive $\text{T}_\text{E}\text{X}$ as formatter

Command:

`(pdf)xmltex program-tex.fot`
or `(pdf)latex program.tex`
with `program-tex.fot` as XML file

`xmltex` as parser of FOT file

input: FOT file

`fotex.xmt` as `xmltex` package

for `fo` name space,

configured in `xmltex.cfg`.

Nonimplemented elements:
spanned columns

See Proceedings TUG2000,
TUGboat 21, No. 3, pp. 222-234.

3.d) FOP as formatter

input: FOT file,
or XML file + XSL stylesheet
`xerces` as parser of FOT file, or
`xalan` as XSL processor,
which uses `xerces`
as parser of XML file
and XSL stylesheet

Command:

```
java -cp path-to-xerces.jar:\
path-to-fop.jar:\
path-to-batik.jar \
org.apache.fop.apps.Fop \
fot-file -pdf pdf-file
java -cp path-to-xalan.jar:\
path-to-xerces.jar:\
path-to-fop.jar:path-to-batik.jar \
org.apache.fop.apps.Fop\\
-xml xml-file -xsl xsl-file -pdf pdf-file
```

Directly from XML/XSL to PDF:

bad error reporting

Errors in XSL stylesheet:

ERROR null

Debugging: explicitly create

FOT file using xalan

Nonimplemented elements:

`block-container, simple-link`

4. Without writing
your own XSL stylesheet,
using DocBook

Write XML file according to

DocBook DTD

Format using DocBook

XSL style sheets

Versions for print (fo) and for html

Command:

```
java -cp path-to-xalan.jar:\  
path-to-xerces.jar:\  
path-to-fop.jar:path-to-batik.jar \  
org.apache.fop.apps.Fop -xml xml-file \  
-xsl path-to-fo-docbook.xsl -pdf pdf-file
```


Nonimplemented elements
in the stylesheet:
ackno

Nonimplemented FOs
and properties
in the formatter:

- FOP: `table-layout`,
`column-width in %`,
`linefeed-treatment`
- Passive $\text{T}_{\text{E}}\text{X}$:
`start-indent="body-start ()"`
($\text{T}_{\text{E}}\text{X}$ macro problem)

Customizing the output:

There are many variables that can be set
Import statement guarantees precedence
of customized values

See `cust-fo-docbook.xsl`

Command:

```
java -cp path-to-xalan.jar:\  
path-to-xerces.jar:\  
path-to-fop.jar:path-to-batik.jar \  
org.apache.fop.apps.Fop -xml xml-file \  
-xsl cust-fo-docbook.xsl -pdf pdf-file
```

URLs: See XSL.html

5. Improving the stylesheet

program-fop-5.xsl

Reusing a single template:

`time:`

- create a template that prints the time in a table cell
- call it in `item`, `break`, `opening`, `outing`

Reordering children:

presentation:

- apply templates in a specific, non-document order: author, title, presenter
- add a comma only if there are an author and a title

Variable number of concurrent presentations/discussions:

`program`:

- determine the maximum number of columns required (maximum number of children of `item`)
- apply templates with this as a parameter

`day`:

- in the column specification: specify the maximum number of columns just found and passed as a parameter
- apply templates with the maximum number of columns as a parameter

`item`:

- determine the number of columns spanned by the children as

`floor(max-cols/num-children)`

- apply templates with this as a parameter

`presentation, discussion`:

- set the number of columns spanned as the parameter passed in by `item`

Variable number of concurrent
presentations/discussions (cont'd):

break, opening, outing:

- set the number of columns spanned as the maximum number of columns

This was:

From XML to T_EX,
a short overview
of available methods