

Electronic publishing and XML
7th CEEnet workshop on networks technology
Budapest

Sébastien 'sbi' BLONDEEL (<sbi@IDEALX.com>)

August 2001

Abstract

We have studied about XML was becoming a growingly accepted standard, how it could be used in data interchange, and how XSLT could allow one to transform XML documents ...

Contents

1	A few words about L^AT_EX2e	1
2	Example of a L^AT_EX2e document	2
3	DocBook	2
4	DocBook Modular StyleSheets	3
5	Other uses of XML and new applications	3
6	Examples of electronic publishing	3
7	References	4

1 A few words about L^AT_EX2e

Donald Knuth invented T_EX and then Leslie Lamport added a higher level layer automating things and giving stylesheets, a package named L^AT_EX, which then got further improved and named L^AT_EX2e.

It was not meant to be used with XML and XSLT, but it is the best tool I know to produce correct printable documents. Its balancing algorithm, for example, works at a paragraph level with penalties, ensuring a better look to the document. By default, the user has little parameters to act upon and little freedom to change the presentation

and the formatting of the document, which is a good thing since styles were written by people who know what they were doing.

It has a different syntax, not using angle brackets but braces for local changes (the equivalent of inline elements), and `begin` and `end` keywords for environments. The backslash character introduces the commands, and the command name stops at the first non-letter character met (roughly).

But it was many special characters incompatible with those of XML, blank lines are significant, it has many packages with some subtle incompatibilities (for example: the footnote calls inside tables don't work), and I am not sure it can handle UTF-8 encoding, which explains the filter pass I added in the Makefile to handle this presentation (although there are a number of styles and encodings supported; I managed to produce Georgian and Russian documents with it, and you will be able to see on the *Not so short introduction* that it can handle a number of documents.)

It is possible to draw one's fonts with the `metafont` program.

Documentation:

- Comprehensive T_EX archive network.
- Not so short introduction to L^AT_EX2_ε, which exists in 10 languages under CTAN:/info/lshort/.
- tetex documentation

2 Example of a L^AT_EX2_ε document

```
\documentclass{article}
% \usepackage[...] % languages, encodings
\title{The title of my document}
\author{John Doe}
\begin{document}
\maketitle
\begin{abstract}
  This document aims at giving a simple
  example of a \LaTeX2ε{ } document. Understanding
  this language is \emph{interesting}.
\end{abstract}

\section{Introduction}
...
\end{document}
```

3 DocBook

One of the only technical DTDs available.

Pretty complete: hundreds of elements and entities.

Hard to manage and remember; PSGML can help, or else use the reference documentation or start with a document similar to the one you wish to type.

Only one set of stylesheets available, and not working very well, especially the printed version.

References: DocBook.org¹ and Oasis-Open.org². An example of book on line: Linux Device Drivers³.

4 DocBook Modular StyleSheets

Use Jade to make all the transformations and Jadedtex for the printable versions.

Mainly written en DSSSL and aimed at SGML processing (but can manipulate XML documents as well, even if the converse is not true). The XSL version supports less backends.

Can be found on Norman Walsh's website⁴.

May be customized⁵.

5 Other uses of XML and new applications

In case it is too difficult for you to use or install at home the software used here in an easy way, here are some other solutions to get started with XML: there exist quite a number of dynamic web pages or software there to help you.

Microsoft XML parser⁶: Internet Explorer is now one of the only web browsers to offer XML support.

XUL⁷ is an XML vocabulary that was designed to describe the look-and-feel of application-based user interfaces.

On line XML document checker and validator.⁸

On line XML document checker and DOM-Tree representation.⁹

Other XML parsers and validators.¹⁰

6 Examples of electronic publishing

The encoding issue: Cyrillic, Chinese, (Georgian).

Typesetting documents in different languages: the example of lshort¹¹.

¹<http://www.docbook.org/>

²<http://www.oasis-open.org/>

³<http://www.oreilly.com/catalog/linuxdrive2/>

⁴<http://www.nwalsh.com/docbook/dsssl/>

⁵<http://www.nwalsh.com/docbook/dsssl/doc/custom/>

⁶<http://www.microsoft.com/XML/>

⁷<http://www.xulbook.com/>

⁸<http://www.oasis-open.org/cover/xml4j-check00.html>

⁹<http://www.networking.ibm.com/xml/XmlValidatorForm.html>

¹⁰<http://xml.cnec.org/download/parser.php>

¹¹<http://www.ctan.org/tex-archive/info/lshort/>

Typesetting music: the example of LilyPond¹² and Nupedia¹³.

Typesetting DocBook with non-standard tools: the example of DB2LaTeX¹⁴.

Translating documents in a "clever" way: recent examples, and the work on translating a document in several languages¹⁵.

Workflow: the example of internal web forms in my company.

7 References

- DocBook: a complete technical SGML/XML DTD
- T_EX Users Group
- Comprehensive T_EX Archive Network

¹²<http://www.lilypond.org/>

¹³<http://www.nupedia.org/>

¹⁴<http://db2latex.sourceforge.net>

¹⁵<ftp://ftp.mandrakesoft.com/>