

Dr. StrangeBook

or: How I Learned to Stop Worrying and Love XML

Nigel Stanger

Department of Information Science

May 7, 2004

What am I going to talk about?

- ▶ Document publication in web and print formats using XML as the base source.
- ▶ Used for many course documents:
 - ▶ Currently INFO 321 and INFO 405
 - ▶ INFO 212 is next

What am I going to talk about?

The initial stages

The second version of our authoring framework

Problems encountered

What you need to do this yourself

Where to next?

Aren't you talking about DocBook?

Aren't you talking about DocBook?

SH!

Not yet!

Disclaimer

This presentation was not produced
from an XML document!

Beamer is much better at this for now.

But first, a word from our sponsors. . .

Chris Edwards did the entire first implementation and much of the basic infrastructure for the second version.

Richard Pascoe got me back into \LaTeX again, and also introduced it to Chris along with the idea of producing multiple output targets from the same source.

Richard O'Keefe presented a seminar on using SGML for exam papers, which got Chris thinking about using XML for document production.

What we wanted to do

- ▶ All course documents in the same source format.
- ▶ Cross-platform (at least Win32 & Mac OS X).
- ▶ Produce both print and HTML versions.
- ▶ Multiple versions of the same document:
 - ▶ Questions for students
 - ▶ Model answers for students
 - ▶ Notes for teachers
 - ▶ Individual documents vs. combined course handbook

We used to use Word...

(→ ca. 1998)

- ▶ OK, but a typical Microsoft product.
- ▶ Print output typically pretty ugly; HTML even worse :(
- ▶ Messy for managing questions vs. answers vs. notes.

...then we moved to \LaTeX ...

(ca. 1999–2002)

- ▶ No GUI, but so what? It doesn't have that `!$@%^$!` paper clip.
- ▶ Beautiful print output.
- ▶ Web output mostly OK ($\text{\LaTeX}2\text{HTML}$), but still not ideal.

...then Chris began to think about XML

(late 2002)

- ▶ Content-neutral format.
- ▶ Potentially better HTML output using XSLT.
- ▶ We were starting to teach XML + XSLT anyway \Rightarrow good learning exercise!

So why didn't you use DocBook?

So why didn't you use DocBook?

SH!

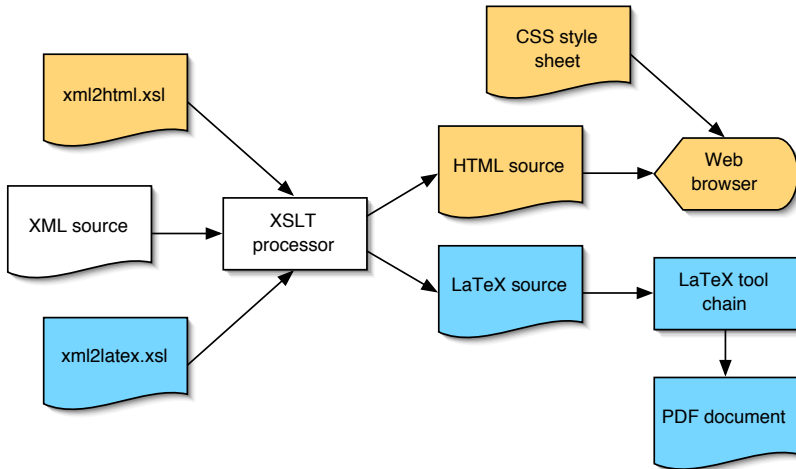
I told you, not yet!!

The first version of the framework

(S1 2003)

- ▶ Lab & tutorial documents, INFO 321 2003.
- ▶ Two monolithic XSL style sheets: XML \rightarrow HTML, XML \rightarrow \LaTeX .
- ▶ Existing \LaTeX tool chain for print output (\LaTeX [\rightarrow DVI \rightarrow PS] \rightarrow PDF).

Workflow for version 1



But not all was well...

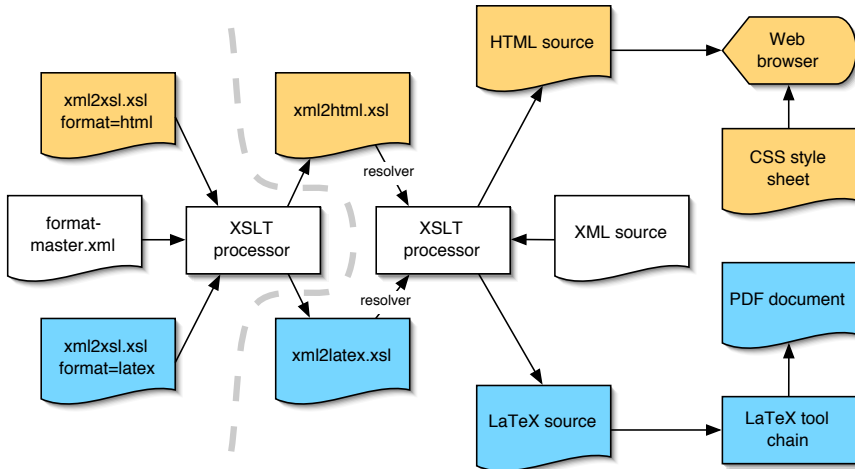
- ▶ Not designed for general documents (i.e., other than labs or tutorials).
- ▶ Two separate style sheets \Rightarrow harder to maintain consistency.

The second version of the framework

(2004)

- ▶ Single monolithic master format document combining both HTML and \LaTeX XSLT templates.
- ▶ Master format document processed through separate XSL master style sheets to produce XML \rightarrow HTML & XML \rightarrow \LaTeX style sheets.
- ▶ Generalised to other types of documents.

Workflow for version 2



Features

- ▶ The usual paragraph formatting, etc.
- ▶ Moderately complex tabular structures (including multi-column & multi-row cells).
- ▶ Hyperlinks & cross-references.
- ▶ Floating matter (figures, tables).
- ▶ Images in various formats.
- ▶ *Very* basic maths.
- ▶ Conditional processing based on format (\LaTeX /HTML).
- ▶ Raw code for that *really* crufty stuff.
- ▶ etc. . .

Examples

BUT WHAT ABOUT DOCBOOK?

BUT WHAT ABOUT DOCBOOK?

All right, all right!!

BUT WHAT ABOUT DOCBOOK?

- ▶ DocBook is a set of comprehensive SGML & XSL style sheets for producing technical computing documents from XML source, managed by OASIS. (see <http://www.docbook.org/>)
- ▶ Why didn't we use it?

BUT WHAT ABOUT DOCBOOK?

- ▶ DocBook is a set of comprehensive SGML & XSL style sheets for producing technical computing documents from XML source, managed by OASIS. (see <http://www.docbook.org/>)
- ▶ Why didn't we use it? **We didn't know about it!**

BUT WHAT ABOUT DOCBOOK?

- ▶ DocBook is a set of comprehensive SGML & XSL style sheets for producing technical computing documents from XML source, managed by OASIS. (see <http://www.docbook.org/>)
- ▶ Why didn't we use it? We didn't know about it!
- ▶ Our framework is remarkably similar but not quite as powerful.
- ▶ But we seem to do some things a little better :)
- ▶ Use formatting objects to output direct to PDF.

Problems encountered with version 2

- ▶ Sometimes need to be careful about white space.
- ▶ Sometimes things just don't work \Rightarrow embedded raw code.
- ▶ \LaTeX -only vs. HTML-only features can be a nuisance.
- ▶ Master format document needs to be modularised.

Platform issues

- ▶ Different T_EX distributions (fpT_EX vs. teT_EX).
- ▶ Different XSLT processors (SAXON vs. Xalan-C vs. Xalan-Java) with different command-line conventions.
- ▶ Line breaks!
- ▶ Compatible vector drawing tools.
- ▶ Differing directory path conventions.
- ▶ Where to find the style sheets?

The essential software

- ▶ The XSL style sheets!
- ▶ XSLT processor (we use Apache Xalan or SAXON).
- ▶ A T_EX distribution (teT_EX).
- ▶ Something to edit XML with (GVim, BBEdit).

Nice to have

- ▶ GhostScript.
- ▶ GNU make.
- ▶ Vector drawing tool (Visio, OmniGraffle).
- ▶ Graphics manipulation tools (epstool, ImageMagick, ...).
- ▶ \LaTeX spelling checker (aspell, Excalibur).
- ▶ Version control (CVS).
- ▶ Apache XML-Commons resolver for locating style sheets on the fly. (see <http://xml.apache.org/commons/>)

Did we achieve our goals?

- ▶ Cross-platform (Win32/Cygwin, Mac OS X, should work fine on any Unix derivative).
- ▶ \LaTeX gives high-quality print output.
- ▶ Good (and improving) HTML output.
- ▶ Customisable & extensible.
- ▶ Sufficient geek factor :)

Did we achieve our goals?

- ▶ Cross-platform (Win32/Cygwin, Mac OS X, should work fine on any Unix derivative).
- ▶ \LaTeX gives high-quality print output.
- ▶ Good (and improving) HTML output.
- ▶ Customisable & extensible.
- ▶ Sufficient geek factor :)

YES

Beyond the current version

- ▶ Roll out for INFO 212 (S2 2004).
- ▶ Investigate moving to DocBook (investigations in progress):
 - ▶ Should be relatively easy to write an XSL style sheet to convert our markup to DocBook
 - ▶ Need a customisation layer for “our” stuff
 - ▶ Default PDF output too Word-like; also needs customisation (Apache FOP vs. PassiveTeX vs. ??)
- ▶ Find a good cross-platform vector drawing tool! PGF? Sodipodi? Skencil? Kivio? Others? (but **not** XFig!!!)
- ▶ SVG for graphics?
- ▶ Lecture slides in XML?

Questions?