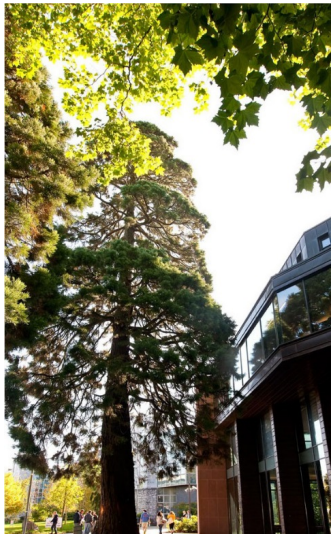


A TRADITION OF INDEPENDENT THINKING

UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Markup to generate markup to generate markup

Peter Flynn

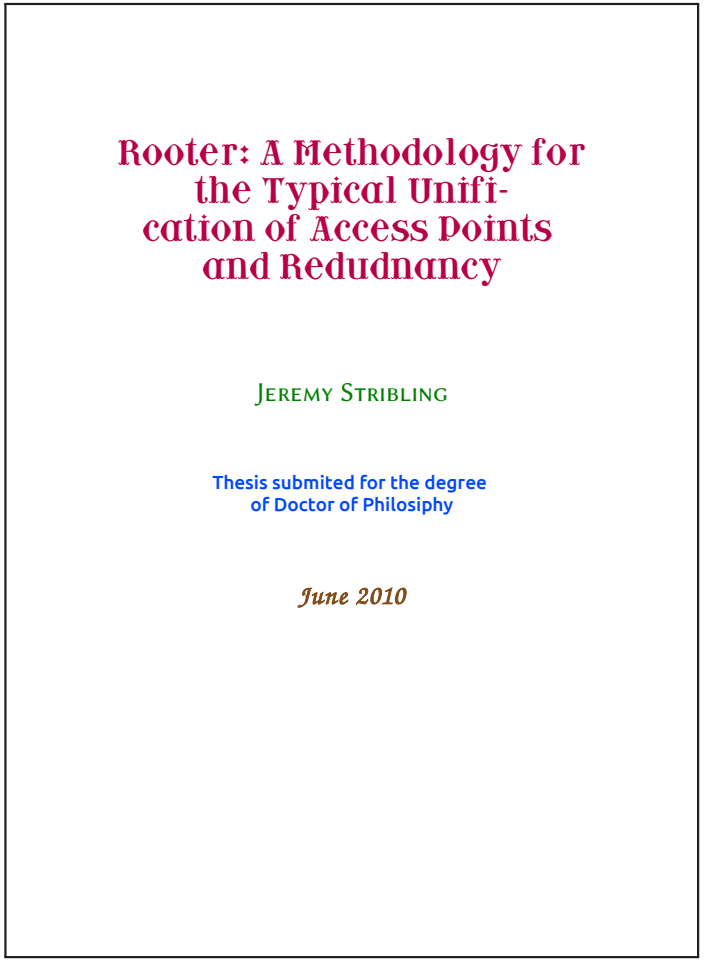
IT Services Electronic Publishing Unit

6 August 2013

Contents

Genesis

Better consistency for theses

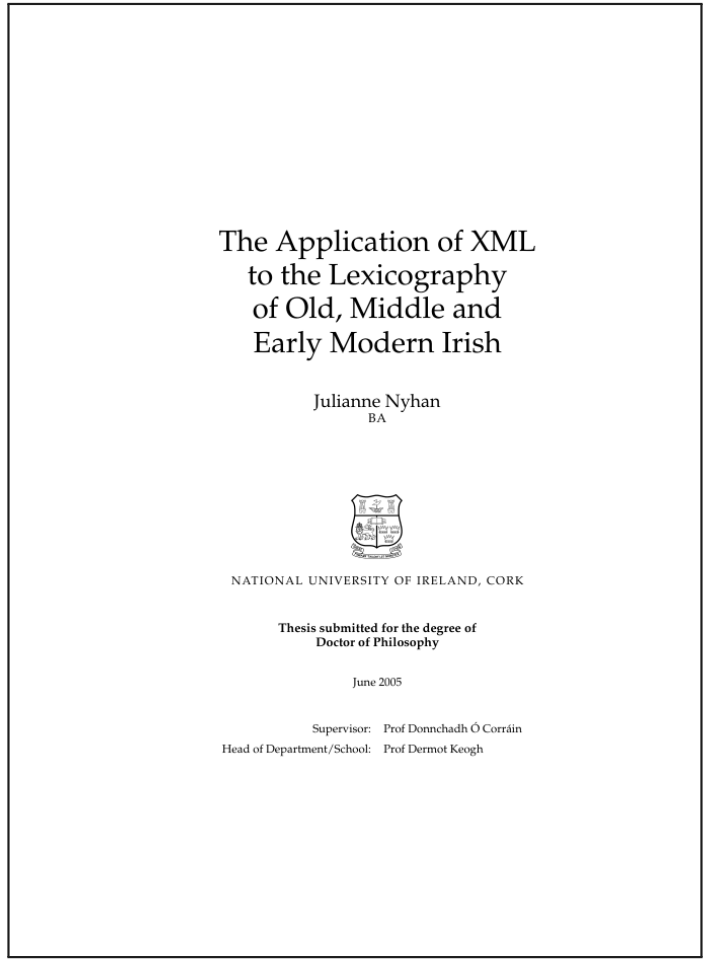


Rooter: A Methodology for the Typical Unification of Access Points and Redudnancy

Jeremy Stribling

Thesis submited for the degree of Doctor of Philosiphy

June 2010



The Application of XML to the Lexicography of Old, Middle and Early Modern Irish

Julianne Nyhan
BA

NATIONAL UNIVERSITY OF IRELAND, CORK

Thesis submitted for the degree of Doctor of Philosophy

June 2005

Supervisor: Prof Donnchadh Ó Corráin
Head of Department/School: Prof Dermot Keogh

LaTeX: the format everyone loves to hate

Consistent, modular, extensible

```
\documentclass{balisage}
\begin{document}
\title{Markup to generate markup to generate markup}
\author{Peter Flynn}
\affiliation{University College Cork}
\date{6 August 2013}
\maketitle
\begin{abstract}
This paper presents an experiment in using DocBook5...
\end{abstract}
\section{Background}
The \LaTeX\ document preparation system provides...
\end{document}
```

A closely-guarded secret known only to a select few million users

Document classes

Consistent layouts (think schema/DTD)

- defaults (reports, books, articles, letters)
- publishers' layouts (books, journals)
- industrial formats (vertical markets)
- business formats (invoices, statements, letters)
- research formats (arts and sciences editions, theses)
- documentation

About 400 classes publicly available

Style packages

Modular design (think plugins)

- typefaces
- formatting variants
- metadata
- languages and writing systems
- recreation
- utilities

About 4,000 packages available

Automation of features which would require repeated manual formatting

Writing your own

Extensible (think programming)

- it's a language
- it's an API
- it can be redefined and customised

But. . .

- no lookahead
- only TEX can parse it reliably
- synchronous typographic editing is rare
- it can be redefined and customised

Authoring *vs* programming

```
In this section we focus our attention on the optical modes of
GaAs\index{GaAs}-Al$_{x}$Ga$_{1-x}$As\index{Al$_{x}$Ga$_{1-x}$As} ($0
\leq x \leq 1$) quantum wells\index{quantum well}.  In studying these
systems we must assign to the ternary alloy the values of the
frequencies for the longitudinal optical and transverse optical modes
given in  section \ref{sec:data} for the GaAs (AlAs) like modes in
this alloy [\ref{bib:5.1}]. The $\beta_{L}$ and $\beta_{T}$
parameters were estimated from the experimental curves of Wang et al.
[\ref{bib:5.2}] (see section \ref{sec:data}).

\unskip\nobreak\hfil\penalty50
  \hskip2em\hbox{}\nobreak\hfil\upshape\footnotesize
  \if\@fullcite\relax\else\@fullcite\fi
  \ifthenelse{\equal{\@quotcite}{}}{%
    \if\@fullcite\relax(Anon.)\fi}{%
       \expandafter\@opttest\@quotcite\sentinel
    \ifoptarg\expandafter\cite\@quotcite
       \else\cite{\@quotcite}\fi}%
    \parfillskip\z@\finalhyphendemerits=0\par}%
  \medskip
  \endlist\@afterheading
```

Writing class and package code

The doc and ltxdoc packages

```
% \begin{macro}{\prelim}
% The \command{prelim} instruction creates a
% minor frontmatter (unnumbered) section.
%    \begin{macrocode}
\newcommand{\prelim}[1]{%
  \ifmadetitle\else\maketitle\fi
  \clearpage
  \section*{#1}\@mkboth{{#1}}{{#1}}%
  \addcontentsline{toc}{section}{#1}}
%    \end{macrocode}
% \end{macro}
```

A form of literate programming; creates a pair of distribution files (`.dtx` and `.ins`)

Enter XML...

Markup of choice for master documents

```
<annotation role="macro" xreflabel="prelim">
  <para>The <command>prelim</command>
    instruction creates a minor frontmatter
    (unnumbered) section.</para>
  <programlisting>
\newcommand{\prelim}[1]{%
  \ifmadetitle\else\maketitle\fi
  \clearpage
  \section*{#1}\@mkboth{{#1}}{{#1}}%
  \addcontentsline{toc}{section}{#1}}
  </programlisting>
</annotation>
```

Data-driven

Organisational rules on classifications

PhD	MA	apsoc	Harvard
LLM	MSW	apsych	APA
DSocSc	MPhil	arthist	Oxford
DOccT	MBA	geog	Harvard
DBA	MBS	german	MLA
DClinDent	MComm	dentsurg	Vancouver
DMus	MEd	chem	ACS
DD	MMus	cs	Default
...	...	...	...

Re-enter XML...

Markup of choice for textual data constraint

```
<constraintdef role="deptoptions">
  <methodsynopsis xml:id="apsoc" arch="acsss">
    <methodname>Harvard</methodname>
    <methodparam>
      <parameter role="school" remap="School of">Applied
        Social Studies</parameter>
      <initializer>kluwer</initializer>
      <modifier>natbib</modifier>
      <modifier>har2nat</modifier>
    </methodparam>
  </methodsynopsis>
  ...
</constraintdef>
```

30 PhD programs and 150 Master's programs among 95 departments

XML to LaTeX

Constructing a class or package file

- DocBook5 for authoring and storage
- Some additional attributes for document control
- Some tag abuse during this experimental phase
- XSLT2 to create the pair of distribution files and maintain a script for the toolchain
- Currently being used to manage 20 classes and packages
- Development version available on CTAN

Modelling the workflow

Contents of a `.dtx` file

1. an initialization block
2. the LaTeX Preamble for the documentation
3. a character checksum table
4. a change history
5. an indexing control block
6. the user documentation
7. **the annotated code**
8. any ancillary files to be distributed with the class or package can be embedded

Initialization

Header and metadata

```
<book xml:id="uccthesis" version="1"
  revision="03" xml:lang="en" xml:base="ucc"
  remap="a4paper,12pt" arch="class"
  audience="lppl" condition="2009/09/24"
  conformance="LaTeX2e" os="all" security="2070"
  userlevel="cls" vendor="UCC" status="beta">
```

Generating the 'tagged' format for extraction

```
% \iffalse
%<*driver>
\ProvidesFile{uccthesis.dtx}
%</driver>
%<class>\NeedsTeXFormat{LaTeX2e}[2009/09/24]
%<class>\ProvidesClass{uccthesis}[2012/12/18 v1.03
  Typesetting a UCC thesis with LaTeX]
...
% \fi
```

Preamble and setup

Using the specifications

```
<methodsynopsis xml:id="apsoc" arch="acsss">
  <methodname>Harvard</methodname>
  <methodparam>
    <parameter role="school" remap="School of">Applied
      Social Studies</parameter>
    <initializer>kluwer</initializer>
    <modifier>natbib</modifier>
    <modifier>har2nat</modifier>
  </methodparam>
</methodsynopsis>
```

Generate user options to select status

```
%    \begin{macrocode}
\DeclareOption{apsoc}{%
  \school{Applied Social Studies}
  \@usebib[natbib,har2nat]{kluwer}{Harvard}{}}
%    \end{macrocode}
```

Invoking packages

Packages are specified in a structure

```
<constraintdef xml:id="clspackages" linkend="options">
  <segmentedlist>
    <segtitle>Packages needed for this class</segtitle>
    <seglistitem>
      <seg>fix-cm</seg>
    </seglistitem>
    <seglistitem>
      <seg role="textwidth=159mm">geometry</seg>
    </seglistitem>
    <seglistitem>
      <seg>graphicx</seg>
    </seglistitem>
    [...]
  </segmentedlist>
</constraintdef>

\RequirePackage{fix-cm}
\RequirePackage[textwidth=159mm]{geometry}
\RequirePackage{graphicx}
...
```

Annotated code

Literate programming, XML-style

```
  <annotation role="macro" xreflabel="so">
    <para>Specify tighter letter-spacing for
      the title page.</para>
    <programlisting>
\sodef\so{}{.1em}{.5em plus.5em}{1em plus1em}
    </programlisting>
  </annotation>
  <annotation role="environment" xreflabel="figure">
    <para>Restyle the Figure float to have
      the caption below.</para>
    <programlisting>
\floatstyle{plain}
\restylefloat{figure}
    </programlisting>
  </annotation>
```

Annotated code

Literate programming, LaTeX-style

```
% \begin{macro}{\so}
% Specify tighter letter-spacing
%   for the title page.\par
%    \begin{macrocode}
\sodef\so{}{.1em}{.5em plus.5em}{1em plus1em}
%    \end{macrocode}
% \end{macro}
% \begin{environment}{figure}
% Restyle the Figure float to have
%   the caption below.\par
%    \begin{macrocode}
\floatstyle{plain}
\restylefloat{figure}
%    \end{macrocode}
% \end{environment}
```

Annotated code

Literate programming, documentation-style

`\so` Specify tighter letter-spacing for the title page.

```
1434 \sodef\so{}{.1em}{.5em plus.5em}{1em plus1em}
```

`figure` Restyle the Figure float to have the caption below.

```
1439 \floatstyle{plain}
1440 \restylefloat{figure}
```

User documentation

Straightforward DocBook

```
<sect2 xml:id="profsup">
  <title>Professor and Supervisor</title>
  <para>Give the name of your Professor (or Head of College,
    School, Department, or Discipline) with
    <command>professor</command>. Give the name of your
    Supervisor with <command>supervisor</command> (but see
    the note below). These commands are compulsory.</para>
  <programlisting language="LaTeX"
    annotations="\professor,\supervisor,\supervisors">
\professor{Dr F Händel}
\supervisor{Dr PDQ Bach}
\supervisors{Dr WA Mozart\\Mr L van Beethoven}
  </programlisting>
  <para>If you have more than one Supervisor, use
    <command>supervisors</command> (plural) instead, and
    separate their names with a double backslash
    (<systemitem>\\</systemitem>).</para>
</sect2>
```

User documentation

Not-quite-so-straightforward DocTEX

```
% \subsubsection{Professor and Supervisor}\label{profsup}
% Give the name of your Professor (or Head of College,
%     School, Department, or Discipline) with
%     {\ttfamily{}\textbackslash{}professor}. Give the name of your
%     Supervisor with {\ttfamily{}\textbackslash{}supervisor} (but see
%     the note below). These commands are compulsory.\par
% \iffalse
%<*ignore>
% \fi
\begin{lstlisting}[language={[LaTeX]TeX},
   emph={\professor,\supervisor,\supervisors}]
\professor{Dr F Händel}
\supervisor{Dr PDQ Bach}
\supervisors{Dr WA Mozart\\Mr L van Beethoven}
\end{lstlisting}
% \iffalse
%</ignore>
% \fi
% If you have more than one Supervisor, use
%     {\ttfamily{}\textbackslash{}supervisors} (plural) instead, and
%     separate their names with a double backslash
%     (\verb|\\|).\par
```

User documentation

Documentation PDF

2.2.1 Professor and Supervisor

Give the name of your Professor (or Head of College, School, Department or Discipline) with the command `\professor`. Give the name of your Supervisor with the command `\supervisor` (but see the note below). These commands are compulsory.

```
\professor{Dr F Händel}
\supervisor{Dr PDQ Bach}
\supervisors{Dr WA Mozart\\Mr L van Beethoven}
```

If you have more than one Supervisor, use the command `\supervisors` (plural) instead, and separate their names with a double backslash (`\\`).

Conclusions

Markup load overcome with XSLT2

- Many of the ‘features’ of LaTeX syntax, avoiding common errors like using ‘fragile’ commands inside other commands.
- Missing or optional features in LaTeX packages because it can look ahead, which LaTeX cannot.
- Precalculate widths and other values for the same reason.

Use of XML also makes it easy to query settings and classifications using standard query tools.

Conclusions

Tag abuse

`exceptionname`	Keywords of RFC 2119:1997
`methodsynopsis`	Structured university data
`entry`	Attributes for formatting
`annotation`	Container for annotated code
`cover`	Package setups
`constraintdef`	Lists of packages
`procedure`	Default settings
`cmdsynopsis`	Extra command output
`type`	Typographical treatment

RelaxNG modification layer is planned once the system is stable.

Conclusions

YES WE CAN

1. Using XML to define and maintain LaTeX document classes and packages works.
2. Benefits of reusability appear only with multiple documents.
3. Requires significant knowledge of XML and DocBook.
4. Saves time and effort when writing and maintaining the files.

Thank you