

HCU Summer School: XML and TEI

Sebastian Rahtz

Lou Burnard

July 2001

Welcome to the HCU Summer Seminars sessions on XML and the TEI! We've tried to crowd far too much into these two days, (and we are still not going to cover all of the topics we'd like to) but we hope you'll find it all worthwhile and enjoyable.

In your pack, you should find

1. Printout of slides for most of the talks
2. Crib sheets for some of the software we'll be using, and for the XML and XSLT languages
3. Instructions for the practical sessions
4. Attendee list and comment sheet
5. A specially produced CD-ROM full of XML resources for you to explore after the Workshop.

1. Overview

This two-day course will combine practical experience with a thorough technical grounding in two key technologies: XML, the W3C's extensible markup language and its associated languages, and TEI, the established markup scheme for preparation of scholarly texts. You will learn how to construct a TEI application (a 'pizza') appropriate to your needs, and will be given an overview of the range of tagging opportunities afforded by the full TEI scheme.

Working with a wide range of textual materials, you will learn how to apply XML markup to facilitate interchange and integration of encoded data. As well as XML itself, we will teach CSS and XSL-T, and you will use these to publish documents over the web or on paper. You will gain experience of a range of XML-based applications and tools, ranging from document management and text retrieval systems to scholarly editing tools and hypertext, with a clear focus on their applicability for digital publishing of TEI-encoded materials.

2. Prerequisites

Some general computing experience and also experience of using a markup language (e.g. HTML) is assumed. This seminar will build on and extend some of the topics covered by the seminar on Digital Resource Creation, but can also be taken on its own.

3. Timetable

3.1. Summary of Day 1 (25th July)

- The XML Refresher.
- TEI Lite: one way to use the TEI DTD.
- Introduction to CSS.
- Introduction to XSLT.

Practical sessions cover editing a TEI-XML text and Web delivery of TEI XML.

3.2. Summary of Day 2 (26th July)

- The TEI Architecture .
- Advanced XSLT.
- XSL-FO.
- XML Strategies and Tools for TEI users.

The practical sessions cover building a DTD, print delivery of TEI XML, and exploiting mark-up in retrieval systems.

3.3. Detailed Timetable

25th July:

Time	Activity	Tutor
9.30-10.15	XML: the refresher	LDB
10.15-11.00	Introduction to TEI Lite	SPQR
11.00-11.30	Coffee break	
11.30-12.00	Exercise: creating and editing XML/TEI texts using Emacs	SPQR
12.00-12.30	Using CSS with XML	SPQR
12.30-13.00	Exercise: using CSS with a TEI text	SPQR
13.00-14.00	Lunch	
14.00-14.45	Metadata Matters: the TEI header	LDB
14.45-15.30	Introduction to XSLT	SPQR
15.30-16.00	Tea	
16.00-16.30	Exercise: simple XSLT	SPQR
16.30-17.00	Escaping to XML	SDR

26th July:

Time	Activity	Tutor
9.30-10.30	The TEI architecture and the pizza chef	LDB
10.30-11.00	Exercise: creating a TEI DTD	LDB
11.00-11.30	Coffee break	
11.30-12.00	More XSLT: using TEI stylesheets	SPQR
12.00-12.30	Exercise: publishing your TEI document to HTML	SPQR
12.30-13.00	Using XML in the Real World	LDB
13.00-14.00	Lunch	
14.00-15.00	An overview of related XML technologies	SDR
15.00-15.30	A look at XSL FO	SPQR
15.30-16.00	Tea	
16.00-16.45	TEI and XML in practice	LDB
16.45-17.30	Hands-on and discussion	LDB

4. Delegates

Dr Steve Ashwell	Health Care Libraries Unit, John Radcliffe, University of Oxford
Emma Barker	History, University of Sheffield
Jane Barton	Papyrology, Ashmolean Museum, Oxford
Dr Anne Bowtell	Literae Humaniores, University of Oxford
Marina Cacioppo	London
Dr Katherine Cocking	English, University of Hull
Amandine Croison	Edinburgh University Library

Nigel Curson	CSED, Registry, University of East Anglia
Leon Doughty	Virtual Norfolk (HIS), University of East Anglia
David Elsmore	Social Sciences and Law, Oxford Brookes University
Dr Daniel Gilfillan	Centre for the Study of Women in Society, University of Oregon
Vanessa Gregory	Oxford
Anne Mette Hansen	The Arnamagnaeian Institute, University of Copenhagen
Anthony Hutton	Art and Design, University of Ulster
Martin Jones	University of Sheffield
Max Kaiser	Austrian Literary Archives, Austrian National Library
Janet Kennington	Institute of Historical Research
Katherine Lindsay	University of Sheffield
Ingo Mittendorf	Linguistics, University of Cambridge
David Moffat	Archives, University of Glasgow
Olwen Myhill	Centre for Metropolitan History, Institute of Historical Research
Fiona OrdShrimpton	Archives, University of Glasgow
Harry Platanakis	Philosophy, University of Edinburgh
Chiba Shoju	College of Foreign Languages, Reitaku University, Japan
Olunmi Sogbesan	Bureau of Public Enterprises, Nigeria
Michael Stevens	Cairns Library, John Radcliffe Hospital, Oxford University
Dr Sophia Stos	Research Laboratory for Archaeology, University of Oxford
Patrick Watrin	Linguistique Germanique, Université Catholique de Louvain
Dr Jo Wheeler	External Relations, King's College London

5. Teaching staff

The seminar is taught by Lou Burnard (<Lou.Burnard@oucs.ox.ac.uk>), manager of the HCU; Sebastian Rahtz, Information Manager, Oxford University Computing Services (<Sebastian.Rahtz@oucs.ox.ac.uk>); and Steve DeRose, Scholarly Technology Group, Brown University (<Steven_DeRose@brown.edu>).

6. The TEI / XML CD-ROM

Text Encoding, the business of scholarly markup, was and remains a key activity in applications of information technology across all kinds of academic discipline, from philology and digital libraries to natural language processing and language engineering. The Text Encoding Initiative (TEI), established as a membership consortium in January 2001, has been at the forefront of developments in this area since its *Guidelines* were first published in 1990. On your CD we are proud to present the first glimpse of the next generation TEI Guidelines, re-expressed in XML, the new language of the internet and therefore of us all.

Since its incorporation, the TEI has been developing

- a new comprehensive web site, <http://www.tei-c.org>;
- a complete revision of the published TEI Guidelines;
- a work plan for continued maintenance and development.

This CD presents some of the work in progress: in particular, it includes the following versions of the Guidelines:

- The Preliminary Edition of P4 (June 2001) (includes draft text describing the new XML/SGML version of the Guidelines); also available as PDF

- The P4 (June 2001) DTDs for the new XML/SGML version of the Guidelines
- The Revised Reprint of May 1999 (includes HTML and SGML versions, and DTDs in a variety of packagings)

Work on correcting and enhancing the text of the Preliminary Edition will continue through the summer of 2001; the DTDs are being made available now as a part of our testing procedures. Comments, and particularly reports of any compatibility problems are very welcome, and should be sent to the TEI editors.

The CD also contains a hand-picked collection of the very finest of currently available software and other resources for the discerning TEI and XML user. Highlights include:

- a complete XML authoring system for Windows 9X/2000/NT (combining Emacs, PSGML and the latest version of the TEI dtds)
- a range of currently available free XML processing tools, including xt, saxon, and rpx;
- complete texts of current W3C specifications for XML and related standards;
- a modular suite of XSL stylesheets developed by SebastianRahtz which can be used to format TEI conformant XML documents as print or HTML;
- a select bibliography of essential reading;
- a range of demonstration XML resources, including literary texts;
- loads of other cool stuff.

If you are a Macintosh user, please note that this CD has long (and mixed-case) filenames according to the Joliet specifications. If you are not yet set up to read Joliet CDs, see `README.MAC`.

We've put this together to demonstrate the range and variety of tools and resources now available, and to provide you with currently authoritative documentation. Please note however that we make no guarantees about any of the material included here, all of which is provided on a purely best-endeavours basis. In particular, we don't claim that anything on this CD-ROM will be fit for any particular purpose, nor do we claim any ownership or property rights in the material included here. Almost all of the material here was obtained from publicly accessible web sites, but we have not exhaustively checked its ownership. You are *not* authorized to redistribute or copy (except for your personal use) any of the material on this CD-ROM without further reference to the originating sites, whose addresses are all listed below. And don't even *think* about trying to sell it.