TEI BY EXAMPLE

XML RESOURCES

Ron Van den Branden

Edward Vanhoutte

Melissa Terras

TABLE OF CONTENTS

1

1   Although most of the information on this page is still useful, it is subject to revision. Currently, more up-to-date information about tools can be found on <u>Tools section of the TEI Wiki</u>. Either this page will be updated or merged into a more general section on TEI/ XML resources.

1. Editors

1.1 Amaya 9.53 (12 December 2006)

Description

Amaya is a complete web browsing and authoring environment, i.e., a tool used to create and update documents directly on the Web. Using Amaya you can create Web pages and upload them onto a server. Authors can create a document from scratch, they can browse the web and find the information they need, copy and paste it to their pages, and create links to other Web sites. All this is done in a straightforward and simple manner, and actions are performed in a single consistent environment. Editing and browsing functions are integrated seamlessly in a single tool.

Amaya always represents the document internally in a structured way consistent with the Document Type Definition (DTD). A properly structured document enables other tools to further process the data safely. Amaya allows you to display the document structure at the same time as the formatted view, which is portrayed diagrammatically on the screen.

Work on Amaya started at W3C in 1996 to showcase Web technologies in a fully-featured Web client. The main motivation for developing Amaya was to provide a framework that can integrate as many W3C technologies as possible. It is used to demonstrate these technologies in action while taking advantage of their combination in a single, consistent environment.

Amaya started as an HTML + CSS style sheets editor. Since that time it was extended to support XML and an increasing number of XML applications such as the XHTML family, MathML, and SVG. It allows all those vocabularies to be edited simultaneously in compound documents.

Amaya includes a collaborative annotation application based on Resource Description Framework (RDF), XLink, and XPointer. Visit the Annotea project home page.

The current release, Amaya 9.53 supports HTML 4.01, XHTML 1.0, XHTML Basic, XHTML 1.1, HTTP 1.1, MathML 2.0, many CSS 2 features, and includes SVG support (transformation, transparency, and SMIL animation). You can display and partially edit XML documents. It's an internationalized application.

Homepage

http://www.w3.org/Amaya/

1.2 Butterfly XML Editor 1.1 (2 July 2004)

Description

Butterfly XML Editor is an IDE built on top of a new real-time incremental XML parsing algorithm. The editor features syntax and error highlighting, incremental validation, code completion, XSLT pipelines, and side by side DOM and source viewing. It supports XML creation with DTD and W3C XML Schemas. Small-scale beta software, yet very promising.

Homepage

http://www.butterflyxml.com/

1.3 Exchanger XML Lite V3.2 (16 September 2005)

Description

Exchanger XML Lite is a comprehensive multi platform XML Editor bringing you lots of the great features you have come to expect. The XML Editor facilitates easy editing, browsing, managing and conversion of XML Documents.

Exchanger XML Lite is a Java-based product that provides unique functionality for viewing, authoring and editing XML data and documents.

It features W3C XML Schema, Relax NG and DTD based editing, tag prompting and validation, XPath and regular expression searches, schema conversion, XSLT, XQUERY and XSLFO transformations, comprehensive project management, an SVG viewer and conversion, easy SOAP invocations, and more....

The Exchanger XML Lite edition is for use only in non-commercial environments. Starting the XML Editor for the first time you will be prompted with a license agreement that only allows use of the Exchanger XML Lite edition in non-commercial environments. No registration is required. If you would like to be kept informed of new releases please subscribe to our newsletter.

Documentation

A full manual is located at http://www.exchangerxml.com/editor/library.html.

Homepage

http://www.freexmleditor.com/

1.4 First Page 2006, 3rd Edition

Description

First Page 2006 is a professional HTML editing software package which lets you create great websites fast! The visually appealing program comes bundled with over 450+ Javascripts and supports all the latest web languages. First Page 2006 now includes full support for HTML, XHTML, PHP, ASP, Cold Fusion, Javascript, CSS, SSI and Perl.

Homepage

http://www.evrsoft.com

1.5 jEdit 4.2 (28 August 2004)

Description

jEdit is a mature and well-designed programmer's text editor with many attractive features and plugings for a wide range of applications amongst which XML/HTML editing.

Completed with specific XML plugins (see http://plugins.jedit.org/), jEdit supports XML tag completion, on-the fly DTD + W3C XML Schema validation, document structure view etc.

Documentation

A full manual is located at http://surfnet.dl.sourceforge.net/sourceforge/jedit/jedit42manual-a4.pdf and inside the jEdit program.

A nice introduction to setting up jEdit for XML editing can be found at http://www.mith.umd.edu/teaching/tutorials/xml_tei/index.html.

Homepage

http://www.jedit.org

1.6 NoteTab Light 4.95

Description

NoteTab Light is a very complete plain text editor which allows you to create SGML, XML, (X)HTML, XSL, CSS etc. documents. The handy libraries (written in their own library language) enable you to create your own user interface and markup tools.

Libraries

In order to work with Edward Vanhoutte's teixlite NoteTab library, Ron Van den Branden's slightly more technical teixliteSA Notetab Library, or the DALF10 and DALF10META Notetab libraries, save the respective libraries (with .clb extension!) in NoteTab Light/Libraries. The Tab "teixlite," "teixliteSA," "DALF10," or "DALF10META" will now appear in the tab-bar at the bottom of the program window. Click to activate the library which will appear in the left margin.

- ° http://www.kantl.be/ctb/project/dalf/tools/DALF10.clb

Homepage

http://www.notetab.com

1.7 OpenOffice 2.1.0 (12 December 2006)

Description

The OpenOffice.org has as its mission statement: "To create, as a community, the leading international office suite that will run on all major platforms and provide access to all functionality and data through open-component based APIs and an XML-based file format."

OpenOffice.org is a multiplatform and multilingual office suite and an open-source project. Compatible with all other major office suites, the product is free to download, use, and distribute. This is a complete free replacement suite for Microsoft Office.

Homepage

http://www.openoffice.org

1.8 Open XML Editor 1.4.3 (21 May 2006)

Description

The Open XML Editor is a freely available tool for XML document editing. It includes a built in XML wellformedness tester and DTD validator. Open XML Editor was written by Dieter Kohler.

Homepage

http://www.philo.de/xmledit/

1.9 StyleAssistant

Description

StyleAssistant is a tool written by Thomas Meinike which helps you create CSS stylesheets by selecting from several drop-down menus.

Homepage

http://www.styleassistant.de/

1.10 TEI-Emacs

Description

A single archive of GNU Emacs, nxml, psgml, and a collection of other useful stuff bundled together for installation on any Windows 32 bit system (30 Mb). Courtesy of the TEI Consortium (last updated Jan 2005). Emacs is a command line driven editor which has been adapted for the Windows platform in this TEI build. Emacs makes intelligent use of selected DTDs and Relax NG schemas and enables you to create and parse valid SGML/XML/(X)HTML/XSL etc. documents.

Homepage

http://www.tei-c.org/Software/tei-emacs/

1.11 TextPad 4.7.3 (19 June 2004)

Description

Textpad is a professional and highly customizable plain text editor. TextPad is designed to provide the power and functionality to satisfy the most demanding text editing requirements. It is Windows hosted, and comes in 16 and 32-bit editions. Huge files can be edited by either - just choose the edition that works best with your PC. The 32-bit edition can edit files up to the limits of virtual memory, and it will work with Windows 9x, ME, NT 4, 2000 and XP.

TextPad has been implemented according to the Windows XP user interface guidelines, so great attention has been paid to making it easy for both beginners and experienced users. In-context help is available for all commands, and in-context menus pop-up with the right mouse button. The Windows multiple document interface allows multiple files to be edited simultaneously, with up to 2 views on each file. Text can be dragged and dropped between files.

In addition to the usual cut and paste capabilities, you can correct the most common typing errors with commands to change case, and transpose words, characters and lines. Other commands let you indent blocks of text, split or join lines, and insert whole files. Any change can be undone or redone, right back to the first one made. Visible bookmarks can be put on lines, and edit commands can be applied to lines with bookmarks.

Frequently used combinations of commands can be saved as keystroke macros, and the spelling checker has dictionaries for 10 languages.

It also has a customizable tools menu, and integral file compare and search commands, with hypertext jumps from the matched text to the corresponding line in the source file (ideal for integrating compilers).

Documentation

Consult the README file and the Help Contents and Index in the Help Menu.

Homepage

http://www.textpad.com

1.12 XML Copy Editor 1.0.0.8 (15 January 2007)

1.13 XRay 2 XML Editor (19 December 2002)

Description

XRay is a free XML editing enviroment. Now in its second major release, XRay provides support for W3C XML Schema (XSD) and an integrated online XML tutorial system. XRay supports real time XML editing, W3C XML Schema validation, and XSLT processing.

Homepage

http://www.architag.com/xray/

1.14 XXE 3.5.1 (27 December 27 2006)

Description

XMLmind XML Editor Standard Edition is a very powerful authoring tool which has all the features needed to edit any XML document, whether conforming to a standard schema (DocBook, DITA, XHTML, etc) or to a proprietary one. As of v3.5.1, Standard Edition has no restrictions in terms of schema. DTD, W3C XML Schema, RELAX NG schema and Schematron are all fully supported. XMLmind XML Editor Standard Edition offers a a word processor-like view for editing XML files, which is configured using W3C's cascading style sheets (CSS).

Documentation

Consult the Help Contents and Index in the Help Menu, and the userguides in the docs\user\ folder of your installation.

Homepage

http://www.xmlmind.com/xmleditor/

2. XML/XSLT Processors

2.1 nsgmls

Description

Nsgmls parses and validates SGML/XML documents, and was written by James Clark as part of the SP SGML toolkit.

Documentation

Once you installed the program, all documentation is in HTML format in the doc directory. Start with doc/index.html. Pay special attention to doc/xml.htm.

Homepage

http://www.jclark.com/sp/

2.2 Runsp2 Run nsgmls

Description

RUNSP2 is designed by Richard Light to let you run the NSGMLS parser in a Windows environment. It provides standard Windows facilities for opening a file to be parsed and running the parser, but goes beyond that by "reading" the error messages, and providing a helpful editing environment in which the user can correct the errors found.

Homepage

http://www.light.demon.co.uk/runsp/

2.3 Saxon 6.5.3 and Instant Saxon 6.5.3

Description

Saxon is an open source XSLT 1.0 processor developed by Michael Kay. It is a Java application, and can be run directly from the command prompt; no web server or browser is required. The Saxon program will transform the XML document to, say, an HTML document, which can then be placed on a web server.

If you are running Windows (95/98/NT/2000) the simplest way to use it is to download Instant Saxon, which is packaged as a Windows executable. You will need to have Java installed, but that will be there already if you have any recent version Internet Explorer. (On non-Windows platforms you will need to install the full Saxon product and follow the instructions that come with it.) Instant Saxon is a cut-down version of the full Saxon package. It provides an XSLT processor that can be executed directly on Windows 95/98/NT/2000 platforms. It includes the same executable code as full Saxon, but omits source code, API documentation, and sample applications.

Saxon will run with any XML parser that implements the SAX2 interface (in its Java form), but it comes with a copy of fhe Ælfred parser, so you don't need to install one separately.

Homepage

http://saxon.sourceforge.net/

2.4 Saxon 8.8

Description

Saxon 8.8 is an XSLT 2.0 and XQuery 1.0 Processor. Developed by Michael Kay, currently editor of the XSLT 2.0 and XPath 2.0 W3C specifications, it can be considered the reference conformant implementation of these standards.

Since version 8.0, Saxon comes in 2 versions:
° Saxon-B: basic version, freely available
° Saxon-SA: schema-aware, available on a commercial license

Saxon features as (schema-aware) XSLT, XQuery and XPath processor. From version 8.7 onwards, Saxon is available on both the Java and .NET platforms.

Homepage

http://saxon.sourceforge.net/

2.5 XMLStarlet Command Line XML Toolkit, version 1.0.1 (15 March 2005)

Description

XMLStarlet is a set of command line utilities (tools) which can be used to transform, query, validate, and edit XML documents and files using simple set of shell commands in similar way it is done for plain text files using UNIX grep, sed, awk, diff, patch, join, etc commands.

XMLStarlet command line utility is written in C and uses libxml2 and libxslt from http://xmlsoft.org/.

Documentation

Once you installed the program, all documentation is in PDF format in the installation directory. The project website contains the same documentation in HTML form as well as links to user forums. See http://xmlstar.sourceforge.net/docs.php.

Homepage

http://xmlstar.sourceforge.net/

3. Browsers

3.1 Amaya 9.53 (12 December 2006)

Description

Amaya is a complete web browsing and authoring environment, i.e., a tool used to create and update documents directly on the Web. Using Amaya you can create Web pages and upload them onto a server. Authors can create a document from scratch, they can browse the web and find the information they need, copy and paste it to their pages, and create links to other Web sites. All this is done in a straightforward and simple manner, and actions are performed in a single consistent environment. Editing and browsing functions are integrated seamlessly in a single tool.

Amaya always represents the document internally in a structured way consistent with the Document Type Definition (DTD). A properly structured document enables other tools to further process the data safely. Amaya allows you to display the document structure at the same time as the formatted view, which is portrayed diagrammatically on the screen.

Work on Amaya started at W3C in 1996 to showcase Web technologies in a fully-featured Web client. The main motivation for developing Amaya was to provide a framework that can integrate as many W3C technologies as possible. It is used to demonstrate these technologies in action while taking advantage of their combination in a single, consistent environment.

Amaya started as an HTML + CSS style sheets editor. Since that time it was extended to support XML and an increasing number of XML applications such as the XHTML family, MathML, and SVG. It allows all those vocabularies to be edited simultaneously in compound documents.

Amaya includes a collaborative annotation application based on Resource Description Framework (RDF), XLink, and XPointer. Visit the Annotea project home page.

The current release, Amaya 9.53 supports HTML 4.01, XHTML 1.0, XHTML Basic, XHTML 1.1, HTTP 1.1, MathML 2.0, many CSS 2 features, and includes SVG support (transformation, transparency, and SMIL animation). You can display and partially edit XML documents. It's an internationalized application.

Homepage

http://www.w3.org/Amaya/

3.2 Firefox 2.0.0.1 (24 December 2006)

Description

Firefox 2.0 is the popular and fast light weight browser of Mozilla.

Homepage

http://www.mozilla.org/products/firefox/

3.3 SeaMonkey 1.1 (18 January 2007)

Description

SeaMonkey is an open-source web-browsing software suite formerly known as the "Mozilla Application Suite." It offers a complete web-browsing environment, with a browser, email client, HTML editor, IRC chat client and more.

Homepage

http://www.mozilla.org/projects/seamonkey/

3.4 Opera 9.10

Description

Opera is a freeware internet browser which can also visualize XML but without XSL support. XML + CSS is supported, however.

Homepage

http://www.opera.com

3.5 Panorama Pro 2.0

Description

Panorama Pro is/was one of the best SGML browsers, but has been discontinued for some years now. The handy WYSIWYG interface allows you to make quick (proprietary) style sheets. Since XML is SGML, this browser can also display XML.

Homepage

Not Available

4. XML Publication Systems

4.1 Anastasia

Description

Anastasia (Analytic System Tools and SGML/XML integration applications) is a publication system that allows you to script the process of translating XML/SGML documents into presentable output. Though the vendor states the tool can "...create output in any format," specific versions of Anastasia allow for CD-ROM and Web publication.

Developers control the output of the source XML/SGML files by creating "style files." These files are developed using Tcl, and therefore include logical branching and looping capabilities as per the Tcl language. Anastasia can handle documents greater than 2GB in size, and can additionally examine documents not only as a series of hierarchical elements, but also as an informational stream - functionality that allows Anastasia to extract data from documents in "chunks," beginning and ending at any arbitrary point within the document itself.

Specifically optimized for HTML output to Web sites, the Web version of Anastasia is also HTTP protocol aware, allowing for the use of forms using "GET" and "POST," for example, and also allowing for the identification of host or user information that can be used to personalize the resulting output for specific users.

Both Web and CD-ROM output can be generated by Anastasia from a single set of scripts, and both the Web and CD-ROM versions include a built-in XML/SGML search engine.

Anastasia is a SGML/XML publication tool which allows the processing and searching of large documents using tcl scripting.

Homepage

http://sourceforge.net/projects/anastasia

4.2 Apache Cocoon

Description

Apache Cocoon is a web development framework built around the concepts of separation of concerns and component-based web development. Cocoon implements these concepts around the notion of "component pipelines," each component on the pipeline specializing on a particular operation. This makes it possible to use a "building block" approach for web solutions, hooking together components into pipelines without any required programming.

Documentation

Documentation can be found at the homepage, and is offered as a separate download package at http://www.apache.org/dist/cocoon/cocoon-2.1.10-docs.zip.

Homepage

http://cocoon.apache.org/

4.3 eXist Native XML Database

Description

The eXist native XML database features efficient, index-based XQuery processing, extensions for fulltext search, XUpdate support, and tight integration with existing XML development tools like Cocoon. The database is lightweight and may be easily deployed in a number of ways, running either as a stand-alone server process, inside a servlet engine, or directly embedded into an application.

The eXist project team has released two stable release versions: one concluding an older development branch featuring an older (but well-tested) internal indexing scheme, the other being a stable snapshot of the current development version (featuring a new indexing scheme and the locus of further development). Since the project is developed very actively, it is certainly worth checking out the Subversion repository for access to the most recent features: http://exist.svn.sourceforge.net/viewvc/exist/. Above its powerful features, it is characterised by excellent documentation and comes with complete example web applications.

Documentation

This project provides excellent documentation at http://www.exist-db.org/documentation.html and examples at http://demo.exist-db.org/examples.xml.

Homepage

http://www.exist-db.org/

4.4 <teiPublisher>

Description

<teiPublisher>, an extensible, modular and configurable XML-based repository is designed to bridge the gap between having a collection of structured documents which are posted on the Web as static HTML or XML pages, and having a functional digital library. It provides a way to create an online web-deliverable Repository of XML-encoded documents by allowing Repository administrators to configure and then publish a browseable and searchable XML document collection through a Web interface.

The application consists of three components:
° an installer: which is used to install the application on your local machine;
° teiWizard: which prepares your documents for web delivery through a series of steps in which you specify your preferences for browsing and searching;
° teiRepository: a server environment which hosts the documents and displays them via a Web browser.

If you have a collection of XML documents that you want to make available through a Web interface, and you have a computer connected to the Internet on which you can run a simple Web server, then you can use the teiPublisher system.

Homepage

http://teipublisher.sourceforge.net/

4.5 Xaira

Description

Xaira (XML Aware Indexing and Retrieval Architecture) is a new version of SARA, the text searching software originally developed for use with the British National Corpus. This new version has been entirely re-written as a general purpose XML search engine, which will operate on any corpus of well-formed XML documents. It is however best used with TEI-conformant documents.

As of release 1.15, all versions of the full Xaira toolkit are distributed under an open source license. This includes source code for the Xaira indexer, the Xaira daemon, and the Xaira SOAP server, as well as the client software. An installer for Microsoft Windows is also available. The software is under active development; the current release is version 1.22.

Xaira is especially geared to efficient querying of XML-annotated corpora, like the British National Corpus. For this, it uses its own query language.

Documentation

Documentation and tutorials are provided at http://www.xaira.org/.

Homepage

http://xaira.sourceforge.net/

5. Viewers

5.1 XPath Explorer

Description

XPath Explorer (XPE) is a GUI application that lets you interactively experiment with XPath. Given an xpath and URL (to an HTML or XML document), it displays matching nodes and their values. This makes it easy to play with and debug your XPath expression.

Homepage

http://sourceforge.net/projects/xpe

5.2 SVG Viewer

Description

The SVG Viewer allows you to display Scalable Vector Graphics.

Homepage

http://www.adobe.com/svg/viewer/install/main.html

6. Validating services

6.1 On-line validation

HTML validation

W3C (X)HTML Validation Service http://validator.w3.org/file-upload.html

CSS Validation

W3C CSS Validation Service http://jigsaw.w3.org/css-validator/

7. Miscellaneous Utilities

7.1 HTML Tidy

Description

Tidy is a utility developed by Dave Raggett which can be used to clean up HTML files, and also to convert them to XML.

Homepage

http://www.w3.org/People/Raggett/tidy/

7.2 Near & Far Light 1.30

Description

Near & Far Light presents you with a tree structure visualization of an imported DTD.

Homepage

Unavailable

7.3 Winzip 11.0

Description

Winzip unzips compressed archives and zips uncompressed ones.

Homepage

http://www.winzip.com/

8. DTDs

8.1 DALF

Description

DALF is an acronym for "Digital Archive of Letters in Flanders." It is envisioned as a growing textbase of correspondence material which can generate different products for both academia and a wider audience, and thus provide a tool for diverse research disciplines ranging from literary criticism to historical, diachronic, synchronic, and sociolinguistic research. The input of this textbase will consist of the materials produced in separate electronic edition projects. The DALF project can be expected to stimulate new electronic edition projects, as well as the international debate on electronic editions of manuscripts.

The DALF DTD is defined as a customization of the TEI.

Homepage

http://ctb.kantl.be/project/dalf/.

8.2 HTML 4.01

Description

HTML 4.01 is a subversion of HTML 4. In addition to the text, multimedia, and hyperlink features of the previous versions of HTML (HTML 3.2 [HTML32] and HTML 2.0 [RFC1866]), HTML 4 supports more multimedia options, scripting languages, style sheets, better printing facilities, and documents that are more accessible to users with disabilities. HTML 4 also takes great strides towards the internationalization of documents, with the goal of making the Web truly World Wide.

Homepage

http://www.w3.org/TR/html401/

8.3 Teixlite

Description

Teixlite describes a manageable subset of the full TEI encoding scheme. The scheme documented here can be used to encode a wide variety of commonly encountered textual features, in such a way as to maximize the usability of electronic transcriptions and to facilitate their interchange among scholars using different computer systems. It is also fully compatible with the full TEI scheme, as defined by TEI document P4, *Guidelines for Electronic Text Encoding and Interchange*, published by the TEI Consortium in 2002.

Homepage

http://www.tei-c.org

8.4 XHTML 1.0

Description

The Extensible HyperText Markup Language (XHTML) is a family of current and future document types and modules that reproduce, subset, and extend HTML, reformulated in XML. XHTML Family document types are all XML-based, and ultimately are designed to work in conjunction with XML-based user agents. XHTML is the successor of HTML, and a series of specifications has been developed for XHTML.

XHTML 1.0 is the W3C's first Recommendation for XHTML, following on from earlier work on HTML 4.01, HTML 4.0, HTML 3.2 and HTML 2.0. With a wealth of features, XHTML 1.0 is a reformulation of HTML 4.01 in XML, and combines the strength of HTML 4 with the power of XML.

XHTML 1.0 is the first major change to HTML since HTML 4.0 was released in 1997. It brings the rigor of XML to Web pages and is the keystone in W3C's work to create standards that provide richer Web pages on an ever increasing range of browser platforms including cell phones, televisions, cars, wallet sized wireless communicators, kiosks, and desktops.

XHTML 1.0 is the first step and the HTML Working Group is busy on the next. XHTML 1.0 reformulates HTML as an XML application. This makes it easier to process and easier to maintain. XHTML 1.0 borrows elements and attributes from W3C's earlier work on HTML 4, and can be interpreted by existing browsers, by following a few simple guidelines. This allows you to start using XHTML now!

You can roll over your old HTML documents into XHTML using an Open Source HTML Tidy utility. This tool also cleans up markup errors, removes clutter and prettifies the markup making it easier to maintain.

XHTML 1.0 is specified in three "flavors." You specify which of these variants you are using by inserting a line at the beginning of the document. For example, the HTML for this document starts with a line which says that it is using XHTML 1.0 Strict. Thus, if you want to validate the document, the tool used knows which variant you are using. Each variant has its own DTD - Document Type Definition - which sets out the rules and regulations for using HTML in a succinct and definitive manner.

XHTML 1.0 Strict

Use this when you want really clean structural mark-up, free of any markup associated with layout. Use this together with W3C's Cascading Style Sheet language (CSS) to get the font, color, and layout effects you want.

XHTML 1.0 Transitional

Many people writing Web pages for the general public to access might want to use this flavor of XHTML 1.0. The idea is to take advantage of XHTML features including style sheets but nonetheless to make small adjustments to your markup for the benefit of those viewing your pages with older browsers which can't understand style sheets. These include using the body element with bgcolor, text and link attributes.

XHTML 1.0 Frameset

Use this when you want to use Frames to partition the browser window into two or more frames.

Homepage

http://www.w3.org/TR/xhtml1/

9. Text Encoding Initiative (TEI)

9.1 An introduction to the TEI and the TEI Consortium

Article

Vanhoutte, Edward. 2004. "An Introduction to the TEI and the TEI Consortium." In: Mats Dahlström, Espen S. Ore, & Edward Vanhoutte (eds.), *Electronic Scholarly Editing-Some Northern European Approaches.* A Special Issue of *Literary and Linguistic Computing*, 19/1. 9–16.

Homepage

http://www.tei-c.org/

9.2 TEI P4: Guidelines for Electronic Text Encoding and Interchange

Description

TEI P4 is the current version of the Guidelines, and should be cited as Sperberg-McQueen, C.M. and Burnard, L. (eds.) 2002. *TEI P4: Guidelines for Electronic Text Encoding and Interchange.* Oxford, Providence, Charlottesville, Bergen: Text Encoding Initiative Consortium.. The chief objective of this revision was to implement proper XML support in the Guidelines, while ensuring that documents produced to earlier TEI specifications remained usable with the new version.

Homepage

http://www.tei-c.org/P4X/

9.3 TEI U5: TEILite

Description

TEILite. *TEI U5: Encoding for Interchange: an introduction to the TEI.* This document provides an introduction to the recommendations of the Text Encoding Initiative (TEI), by describing a manageable subset of the full TEI encoding scheme. The scheme documented here can be used to encode a wide variety of commonly encountered textual features, in such a way as to maximize the usability of electronic transcriptions and to facilitate their interchange among scholars using different computer systems. It is also fully compatible with the full TEI scheme, as defined by TEI document P4, *Guidelines for Electronic Text Encoding and Interchange*, published by the TEI Consortium in 2002.

This document is available in HTML on the TEI Consortium Webpage.

Homepage

http://www.tei-c.org/Lite/

10. DALF

10.1 DALF

Description

DALF is an acronym for "Digital Archive of Letters in Flanders." It is envisioned as a growing textbase of correspondence material which can generate different products for both academia and a wider audience, and thus provide a tool for diverse research disciplines ranging from literary criticism to historical, diachronic, synchronic, and sociolinguistic research. The input of this textbase will consist of the materials produced in separate electronic edition projects. The DALF project can be expected to stimulate new electronic edition projects, as well as the international debate on electronic editions of manuscripts.

The DALF DTD is defined as a customization of the TEI.

Description in Dutch

Ron Van den Branden, DALF: Een Overzicht

DTD

http://ctb.kantl.be/project/dalf/dalfdoc/DTDfiles.html

Guidelines

Edward Vanhoutte & Ron Van den Branden (eds.) 2003. *DALF guidelines for the description and encoding of modern correspondence material.* Version 1.0. Gent: CTB-KANTL. http://www.kantl.be/ctb/project/dalf/dalfdoc/index.html.

11. Specifications

11.1 eXtensible Markup Language 1.0 Specification (Third edition - W3C Recommendation 4 February 2004)

Description

The Extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

Location

Web site http://www.w3.org/TR/2004/REC-xml-20040204/REC-xml-20040204-review.html.

11.2 The Annotated eXtensible Markup Language 1.0 Specification

Description

The official XML 1.0 specification, with detailed explanatory and historical annotations by one of its editors, Tim Bray.

Location

Web site http://www.xml.com/axml/axml.html

11.3 The Extensible Stylesheet Language Family (XSL)

Description

XSL is a family of recommendations for defining XML document transformation and presentation. It consists of 2 major transformation languages:

° XSL Transformations (XSLT): a language for transforming XML to other structured formats (XML, HTML, plain text)

° XSL Formatting Objects (XSL-FO): an XML vocabulary for specifying formatting semantics that can be interpreted to produce common layout formats (PDF, PostScript,...)

Both come in various generations.

XSLT

■ XSL Transformations (XSLT), Version 1.0 (W3C Recommendation, 16 November 1999): now commonly supported by most XSLT processing software. Although rather limited compared to its successor, it guarantees the most stable and widespread option for XML transformation.

■ XSL Transformations (XSLT), Version 2.0 (W3C Recommendation, 23 January 2007): with its many extensions to the previous version and support of the XPath 2 standard, it practically functions as a different (and much more powerful) language for XML transformations. This version of the standard is still marginally implemented in XSLT processors (with Saxon as notable exception).

XSL-FO

■ Extensible Stylesheet Language (XSL), Version 1.0 (W3C Recommendation, 15 October 2001): this specification defines an XML grammar expressing formatting objects, that can be interpreted by XSL-FO processors to different layout formats (PDF, PostScript,...).

■ Extensible Stylesheet Language (XSL), Version 1.1 (W3C Recommendation, 05 December 2006): this version of the standard adds minor new features.

Location

The The Extensible Stylesheet Language Family (XSL) webpage http://www.w3.org/Style/XSL/ contains links to the different (versions of the) eXtensible Stylesheet standards:

º   XSLT 1.0: http://www.w3.org/TR/xslt.

º   XSLT 2.0: http://www.w3.org/TR/xslt20/.

º   XSL-FO 1.0: http://www.w3.org/TR/xsl/.

W3C Candidate Recommendations:

º   XSL-FO 1.1: http://www.w3.org/TR/xsl11/.

11.4 XML Path Language (XPath)

Description

XPath is a language for addressing parts of an XML document, designed to be used by various other XML-related standards, such as XSLT, XSL-FO, XQuery, XPointer. Parallel with development of those standards, various generations of the XPath specification exist:

º   XML Path Language (XPath), Version 1.0 (W3C Recommendation, 16 November 1999): in support of its primary purpose (to address parts of an XML document), XPath also provides basic facilities for manipulation of strings, numbers and booleans.

º   XML Path Language (XPath), Version 2.0 (W3C Recommendation, 23 January 2007): XPath 2.0 is a superset of XPath 1.0, with the added capability to support a richer set of data types, and to take advantage of the type information that becomes available when documents are validated using XML Schema. To support richer type sets, XPath 2.0 offers a greatly-expanded set of functions and operators.

Location

The The Extensible Stylesheet Language Family (XSL) webpage http://www.w3.org/Style/XSL/ contains links to the different (versions of the) XPath standards:

º   XPath 1.0 W3C Recommendation: http://www.w3.org/TR/xpath/.

º   XPath W3C Recommendation: http://www.w3.org/TR/xpath20/.

11.5 XML Query Language (XQuery)

Description

XQuery is a query language for XML structures that provides the means to extract and manipulate data from XML documents or any data source that can be viewed as XML, such as relational databases or office documents. As an extension to XPath Version 2.0 in the form of a functional programming language, it enables powerful manipulation capacities to XML document collections. XQuery is defined as XQuery 1.0: An XML Query Language (W3C Recommendation, 23 January 2007)

Location

The W3C XML Query workgroup webpage http://www.w3.org/XML/Query/ contains links to various surrounding documents and specifications. The most important one is the XQuery 1.0 W3C Recommendation: http://www.w3.org/TR/xquery/.

11.6 XHTML 1.0 Specification: The Extensible HyperText Markup Language (W3C Recommendation 26 January 2000).

Description

This specification defines XHTML 1.0, a reformulation of HTML 4 as an XML 1.0 application, and three DTDs corresponding to the ones defined by HTML 4. The semantics of the elements and their attributes are defined in the W3C Recommendation for HTML 4. These semantics provide the foundation for future extensibility of XHTML. Compatibility with existing HTML user agents is possible by following a small set of guidelines.

Location

Web site http://www.w3.org/TR/xhtml1/.

11.7 HTML 4.01 Specification (W3C Recommendation 24 December 1999).

Description

This specification defines the HyperText Markup Language (HTML), the publishing language of the World Wide Web. This specification defines HTML 4.01, which is a subversion of HTML 4. In addition to the text, multimedia, and hyperlink features of the previous versions of HTML (HTML 3.2 [HTML32] and HTML 2.0 [RFC1866]), HTML 4 supports more multimedia options, scripting languages, style sheets, better printing facilities, and documents that are more accessible to users with disabilities. HTML 4 also takes great strides towards the internationalization of documents, with the goal of making the Web truly World Wide.

HTML 4 is an SGML application conforming to International Standard ISO 8879 — Standard Generalized Markup Language [ISO8879].

Location

Web site http://www.w3.org/TR/html4/.

11.8 CSS2 Specification (W3C recommendation 12 May 1998).

Description

This specification defines Cascading Style Sheets, level 2 (CSS2). CSS2 is a style sheet language that allows authors and users to attach style (e.g., fonts, spacing, and aural cues) to structured documents (e.g., HTML documents and XML applications). By separating the presentation style of documents from the content of documents, CSS2 simplifies Web authoring and site maintenance.

CSS2 builds on CSS1 (see and, with very few exceptions, all valid CSS1 style sheets are valid CSS2 style sheets. CSS2 supports media-specific style sheets so that authors may tailor the presentation of their documents to visual browsers, aural devices, printers, braille devices, handheld devices, etc. This specification also supports content positioning, downloadable fonts, table layout, features for internationalization, automatic counters and numbering, and some properties related to user interface.

Location

Web site http://www.w3.org/TR/REC-CSS2/.

12. Bonus Tracks

12.1 EZThumbs 2.6

Description

Easy Thumbnails is a handy new freeware utility for creating accurate thumbnail images and scaled-down copies from a wide range of popular picture formats. An elegant interface makes it simplicity itself to find your images and select them for processing individually, in batches, or in whole folders, using a well-designed file selector and built-in image viewer. You can use slider controls to rotate images and adjust their contrast, brightness, sharpness and quality, and preview the results.

Thumbnails can be created in any existing folder or a new folder, and you can identify them clearly by adding a prefix or suffix to their titles. If you're an image processing enthusiast, you will appreciate the option of choosing from a selection of six size-reduction algorithms for the best possible results.

Easy Thumbnails is a genuine freeware product, without the annoyance of advertising, intrusive spyware components, or nag screens. It comes from Fookes Software, the developers of the award-winning NoteTab text editors.

Homepage

http://www.fookes.com/ezthumbs/index.html

13. Guides to Good Practice

13.1 Text Encoding & Markup

- Alan Morrison, Michael Popham, and Karen Wikander. *Creating and Documenting Electronic Texts: A Guide to Good Practice.* AHDS Guides to Good Practice. http://ota.ahds.ac.uk/documents/creating/.
- *TEI Guides to local Practice.* http://www.tei-c.org/Support/Learn/tutorials.xml.

13.2 Digitization

- Catherine Grout, Phill Purdy, and Janine Rymer. *Creating Digital Resources for the Visual Arts: Standards and Good Practice.* AHDS Guides to Good Practice. http://vads.ahds.ac.uk/guides/creating_guide/contents.html.

- Sean Townsend, Cressida Chappell, and Oscar Struijvé. *Digitising History: A Guide to Creating Digital Resources from Historical Documents.* AHDS Guides to Good Practice. http://hds.essex.ac.uk/g2gp/digitising_history/index.asp.

- TASI. *Basic Guidelines for Image Capture and Optimisation.* http://www.tasi.ac.uk/advice/creating/img_capt.html.

- TASI. *An Introduction to Making Digital Image Archives.* http://www.tasi.ac.uk/advice/overview.html.

- TASI. *The Digital Image.* http://www.tasi.ac.uk/advice/creating/image.html.

14. Journals & Series

- *CLC Web Comparative Literature and Culture*: online since 1999, publishes peer reviewed articles on all aspects of its topic. http://clcwebjournal.lib.purdue.edu/.

- *Computers and the Humanities*: http://kapis.www.wkap.nl/kapis/CGI-BIN/WORLD/journalhome.htm?0010-4817.

- *Computers and Texts*: journal/newsletter of the CTI Centre for Textual Studies. http://info.ox.ac.uk/ctitext/publish/comtxt/.

- *Computing in the Humanities Working Papers*: an interdisciplinary series of refereed publications on computer-assisted research http://www.chass.utoronto.ca/epc/chwp/.

- *Cultural & Heritage Informatics Quarterly*: http://kapis.www.wkap.nl/kaphtml.htm/IFA1042-1467.

- *DigiNews*: a journal of the RLG (Research Library Group). http://www.thames.rlg.org/preserv/diginews/.

- *D-Lib forum and magazine*: widely-read electronic periodical and online resource center for digital library research. http://www.dlib.org/dlib/

- *Electronic Book Review (EBR)*: not just reviews; EBR publishes full-length articles, work in experimental media, and roundtable dialogues. http://www.electronicbookreview.com/

- *Human IT*: a quarterly peer reviewed journal published by The Centre for Information Technology Studies as a Human Science (ITH) at the Swedish School for Library and Information Studies (SSLIS), University College of Borås and Göteborg University; many articles of interest to the humanities computing community, both in Swedish and, to an increasing degree, in English. Also published in print. http://www.hb.se/bhs/ith/humanit-eng.htm.

- *Jahrbuch für Computerphilologie*: the online component of a printed journal offering a form for discussion of the numerous possibilities computing offers to literary studies. http://www.computerphilologie.uni-muenchen.de/ejournal.html.

- *Journal of Electronic Publishing*: publishes articles on all aspects of its topic. http://www.press.umich.edu/jep.

- *Literary and Linguistic Computing*: the official peer reviewed journal of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities. http://www3.oup.co.uk/jnls/list/litlin/.

- *Office for Humanities Communication*: based at King's College London, has currently 16 titles in print. http://www.ohc.kcl.ac.uk/.

- *Postmodern Culture*: frequent articles on the cultural impact of computing and new media. Oldest peer-reviewed electronic journal in the humanities (founded 1990). http://www.iath.virginia.edu/pmc/.

- *Revue. Informatique et Statistiques dans les Sciences Humaines*: Belgian journal with articles in French and English concerning the applications of computing techniques and statistics to the humanities. http://www.ulg.ac.be/cipl/rissh/rissh.htm.

- *Text Technology. The Journal of Computer Text Processing*: a journal dedicated to computer technologies and their applications to text. http://cheiron.mcmaster.ca/texttech/.

15. Useful URLs

15.1 XML Tools and Resources

- http://www.w3.org/ The World Wide Web Consortium

- http://www.garshol.priv.no/download/xmltools/ Lars Garshol's database of free XML software

- http://www.xml.com Premier high gloss industry portal hosted by OReilly

- http://xml.startkabel.nl All you need on one page—also available in Dutch

- http://www.oasis-open.org/cover/sgml-xml.html The original all inclusive overview of everything

- http://www.zvon.org The guide to the XML galaxy with downloadable XML/XSL tutorials

15.2 XSL Tools and Resources

- http://www.exslt.org EXSL is a community initiative to provide extensions to XSLT

- http://www.jenitennison.com/xslt The pages of one of the gurus of XSLT

- http://www.oasis-open.org/cover/sgml-xml.html The original all inclusive overview of everything

- http://www.zvon.org The guide to the XML galaxy with downloadable XML/XSL tutorials

- http://www.w3.org/ The World Wide Web Consortium

15.3 Digitization

- http://scantips.com/ A Few Scanning Tips: a very useful site with some good information for all levels, by Wayne Fulton

- http://www.tasi.ac.uk/index.html TASI Technical Advisory Service for Images

- http://vads.ahds.ac.uk Visual Arts Data Service