

TEI BY EXAMPLE



XML RESOURCES

Ron Van den Branden

Edward Vanhoutte

Melissa Terras

Centre for Scholarly Editing and Document Studies (CTB) , Royal
Academy of Dutch Language and Literature, Belgium, Gent, 9 July 2010

Last updated September 2020

Licensed under a Creative Commons Attribution ShareAlike 3.0 License

TABLE OF CONTENTS

4. XML Publication Systems.....	1
4.1 Anastasia.....	1
4.2 Apache Cocoon.....	2
4.3 eXist Native XML Database.....	2
4.4 <teiPublisher>.....	3
4.5 Xaira.....	4

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/>