

The TEI Toolbox

Lou Burnard Consulting

mai 2014

Now that's in TEI-XML what can we do with it?

- The TEI does not provide a solution out of the box ..
- Historically, it has always kept its distance from software development
- The **Guidelines** focus on an abstract view of what text objects are, not what you might do with them
- What you do with them, and therefore which tools you use, is up to you.

OK, but ...

The TEI Guidelines themselves are in XML, like any other TEI document

- so in principal they can be processed with any XML software
- you just have to build them

For its own needs, the TEI has developed tools to

- generate and process customized schemas ("ODD")
- display documents in various common formatrs (XHTML, PDF, Word, Open Office, ePUB...)
- a generic architecture for transforming TEI documents

The TEI-XML toolbox

To deal with XML documents you need to be able to

- 1 create, modify, store them ...
- 2 transform or display them ..
- 3 search and analyse them ...
- 4 store, manage, archive, preserve them ...

1. Creating and editing XML documents

- oXygen is not the only XML editor in the world!
- See http://wiki.tei-c.org/index.php/Category:Editing_tools
- You can also create documents in your favourite word processor (Word, Open Office) and do an automatic conversion to TEI XML
- For highly structured documents, you can capture them by means of a form, or some other simple interface
- See (for example) Transcribe Bentham and other crowd sourcing projects

2. Transformation for display

The TEI maintains a modular suite of XSLT TEI Stylesheets developed by Sebastian Rahtz and now available :

- packaged within oXygen, as part of the TEI Framework
- downloadable from gitHub
(<https://github.com/TEIC/Stylesheets>)
- as a component of the EGE web service (used by OxGarage)

The stylesheet library is very flexible and complete, but doesn't do everything! for some other tools see the wiki list
<http://wiki.tei-c.org/index.php/Category:Tools>

Why use the TEI Stylesheets

- they provide solutions for problems you may not have anticipated
- better to participate in an open source project than to re-invent your own wheel
- they are maintained along with the TEI Guidelines
- they are modular and extensible and well 'architected'

Why NOT use the TEI Stylesheets

Although generic in scope they are not omnipotent!

- Not every TEI element is treated!
- Not every attribute (in particular those unspecified by the TEI)

To use them well, you need to know :

- how to hack XSLT a bit
- add your own personalisation layer (a profile)

Transformations already provided

	from TEI	to TEI
Word processor formats		
OOXML (docx), ODF (odt)	*	*
LaTeX, PDF	*	
ePub, XSLFO	*	
Web Formats		
HTML5, JSON, RDF	*	
Wordpress		*
Markdown	*	*
Schema Languages(via ODD)		
XML DTD, RELAXNG, Schematron, W3C Schema	*	
Other XML formats		
TEI P4		*
NLM	*	
xlsx	Excel (OOXML)	x
Verbatim XML, Docbook	*	*
Other non-XML formats		
Cocoa	*	
Plain text	*	

Publishing systems

Putting TEI on the web: lots of choice:

- plug-ins for popular CMS (eg Drupal, Zotero, Omeka ...)
- TEI-friendly CMS (eg. Kiln, TEI Boilerplate, Lodel)
- generic XML solutions (e.g. xtf)

XTF : a digital library creator

- Extensible Text Framework or XTF
(<http://www.cdlib.org/inside/projects/xtf/>)
is a collection of server-side scripts from the California Digital Library
- If you are already running Apache and Tomcat, and have access to a website, you can set up a default xtf application for TEI files in 5 minutes
- (more or less)
- Library model

Some examples...

- Victorian Women Writers <http://webapp1.dlib.indiana.edu/vwwp/projectinfo/technical.doc>
- Names and places in Rabelais
<http://renom.univ-tours.fr/>

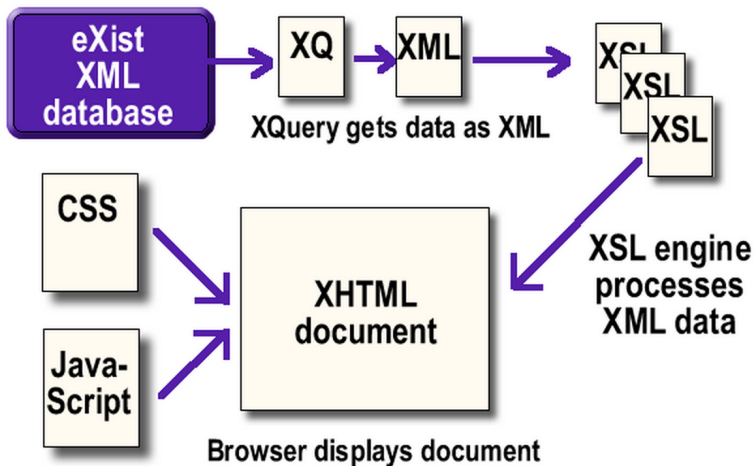
Searching and analysing ..

Generic XML databases

- baseX <http://basex.org>
- eXist <http://exist-db.org>

The solution par excellence for projects with major collections of TEI XML documents

Typical architecture



Some examples...

- Colonial Despatches:
<http://bcgenesis.uvic.ca/docsByDate.htm>
- Carl Maria van Weber
Archive<http://weber-gesamtausgabe.de/en/A002068/Correspondence>

Research tools

Particular scientific communities have developed their own TEI-aware tools:

- 'textometrie' : lexical statistics
- developing critical apparatus
- analysis of transcribed speech
- linguistic analysis

for example ...

<http://textometrie.ens-lyon.fr/>



Some "TEI-friendly" applications

- Ancient Wisdoms : kiln <http://www.ancientwisdoms.ac.uk/method/software-install/>
- Shelley-Godwin archive : shared canvas viewer <http://shelleygodwinarchive.org/about>
- Letters and mss of 19th c Berlin <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/manuscript?Sandmann+en#5>
- Bibliotheque Virtuelle des Humanistes : philologica <http://www.bvh.univ-tours.fr/Epistemon/philologic.asp>

<http://www.histei.info>

HisTEI

A framework for Oxygen XML Editor allowing researchers to transcribe historical documents in TEI

Pages

[Home](#)

Useful Links

[Emerging Standards Project](#)

[HisTEI on GitHub](#)

[Medieval Irish Bilingualism](#)

[Oxygen XML Editor](#)

[Text Encoding Initiative \(TEI\)](#)

Home

Welcome to the HisTEI project!

HisTEI is an open-source framework for the Oxygen XML Editor that allows historians and historical linguists to transcribe texts in TEI in a word-processor-like view. The source code is hosted at [GitHub](#) and is comprised of Java modules, CSS, XML schemas and special settings that Oxygen uses to put them all together in framework files.

Download

HisTEI offers a base set of these files, that projects can then customize for their own use. Currently, there are three frameworks available:

1. HisTEI: base framework for English
2. EMST: framework for English used by the [Emerging Standards project](#)
3. Amsterdam: framework for Dutch used as part of my dissertation project
4. MIB: framework for the [Medieval Irish Bilingualism project](#) - <http://mib.histei.info/mib.xml>

These can all be accessed in the /dist folder on GitHub. It's much easier to add their URL to Oxygen though, so you can get automatic updates. Documentation is now on the GitHub wiki, so check that out for instructions:

Getting Started HisTEI

Contributors

 [Mike](#)

 [Dutch Corpus](#)

<http://www.juxtasoftware.org>



juxta
Compare • Collate • Discover

Home
Juxta Commons
Screenshots
Download
About
User Manual

About

Juxta is an open-source tool for comparing and collating multiple witnesses to a single textual work. Originally designed to aid scholars and editors examine the history of a text from manuscript to print versions, Juxta offers a number of possibilities for humanities computing and textual scholarship.

As a standalone desktop application, Juxta allows users to complete many of the necessary operations of textual criticism on digital texts (TXT and XML). With this software, you can add or remove witnesses to a comparison set, switch the base text at will. Once you've collated a comparison, Juxta also offers several kinds of analytic visualizations. By default, it displays a heat map of all textual variants and allows the user to locate — at the level of any textual unit — all witness variations from the base text. Users can switch to a side by side collation view, which gives a split frame comparison of a base text with a witness text. A histogram of Juxta collations is particularly useful for long documents; this visualization displays the density of all variation from the base text and serves as a useful finding aid for specific variants.

The desktop version of Juxta also allows users to annotate Juxta-revealed comparisons and save the results, and can output a lemmatized schedule (in HTML format) of the textual variants in any set of comparisons. It can run on any modern Macintosh, Windows, or Unix computer with Java 1.5 installed.

Google Group for developers
Subscribe to RSS feed

<http://v-machine.org/>

VM4.0 VERSIONING
MACHINE v4.0

A TOOL FOR DISPLAYING & COMPARING
DIFFERENT VERSIONS OF LITERARY TEXTS

[Sample Texts](#)

[Download](#)

[Documentation](#)

[Credits](#)

[Report a Bug](#)

The Versioning Machine is a framework and an interface for displaying multiple versions of text encoded according to the Text Encoding Initiative (TEI) Guidelines. VM 4.0 has been updated to be P5 compatible. While the VM provides for features typically found in critical editions, such as annotation and introductory material, it also takes advantage of the opportunities afforded by electronic publication to allow for the comparison diplomatic versions of witnesses, and the ability to easily compare an image of the manuscript with a diplomatic version.

The Versioning Machine is also a tool for textual editors, providing an environment that allows editors to immediately see the consequences of their editorial decisions. The Versioning Machine can be used locally on a Mac or a PC, or it can be mounted on the WWW for public access. The documentation provided with the software not only provides information about the use of the software, but builds upon the Critical Apparatus chapter of the TEI Guidelines to give further guidance to those who wish to use this method of encoding.

[HOME](#) | [SAMPLE TEXTS](#) | [DOWNLOAD](#) | [DOCUMENTATION](#) | [CREDITS](#) | [REPORT A BUG](#)

A DIGITAL COLLECTIONS & RESEARCH PROJECT – LAST MODIFIED: MONDAY, JULY 25, 2011

© SUSAN SCHREIBMAN, 2002, 2010

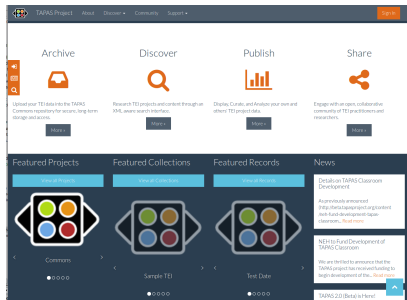


Other examples ...

<http://www.tei-c.org/Activities/Projects>

Preservation and archiving

- Who supports your website?
- Who maintains its infrastructure?
- Where will you keep your data?



<http://www.tapasproject.org/>

Conclusions

- even the simplest of approaches enables you to profit from your markup
- the richer your markup, the more you can do with it
- there's a common set of methods and techniques: no need to reinvent the wheel
- TEI XML is about empowering the data provider : it's up to you to decide how your material is processed