Tools for publication and analysis of TEI documents

TEI@Oxford

July 2009
Summary

• Now that it's all in TEI XML, what next?
• Varieties of XML delivery system
• Some real life examples
A reminder: what is XML for?

- exchanging data
  1. between people
  2. between people and machines
  3. between machines

- preserving information
  1. independent of particular applications
  2. independent of medium or hardware

But is XML any good as a means of delivering information?
What tools do we need?

• Appropriately expressive vocabularies (eg TEI XML)
• Syntax-checking document creation tools (ie editors)
• Document transformation tools
• Document delivery tools
• Document storage and management tools
• Programming interfaces
• Specialized applications
What are the options?

• for delivery:
  • transform into HTML on the server and deliver that
  • deliver XML; use clever client-side processing to handle it

• for management:
  • the filesystem
  • a content management system
  • a database management system

• for tools development:
  • generic digital library solutions e.g. xtf
  • Linux-Apache-mySQL-PHP (LAMP) toolkit
  • web applications using xhtml, ajax, etc.
Real life examples 1

The line of least resistance...

• Here are some TEI files. You know what to do. http://www.cnrtl.fr/corpus/estrepublicain/
• Just the data... http://archimedespalimpsest.net
• make the browser do the work: Early America's Digital Archive (http://www.mith2.umd.edu/eada/)
• handling multiple scripts: Samyukta Agama (http://buddhistinformatics.ddbc.edu.tw/BZA/)
• HAL Open Archive (http://hal.archives-ouvertes.fr)
Real life examples 2

• make the client do the work: Artamène, ou Le Grand Cyrus (http://www.artamene.org/)

• combine image and text: Robert Graves' Diary (http://www.tapor.uvic.ca:8080/cocoon/graves/index.xml)

• critical edition of biblical literature: (http://www.sbl-site.org/publications/books_additionalmaterial.aspx)

• Old Bailey proceedings http://www.oldbaileyonline.org

• digital archive: Swinburne Archive (http://swinburnearchive.indiana.edu/swinburne/www/swinburne/)

• Ms cataloguing (http://handrit.is)

• http://www.bvh.univ-tours.fr/Epistemon/index.htm
XTF meets Mr Punch

• 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

http://m.ox.ac.uk:8080/xtf/view?docId=Punch/1914-07-01.xml
Conclusions

- even the simplest of approaches allows you to share your encoding, your analysis
- the more ambitious your markup, the more sophisticated the possibilities
- there is a common set of techniques and tools: you don't need to re-invent the wheel
- TEI XML empowers the data provider: you determine what can be done with your materials