Standards for digital encoding
Tomaž Erjavec

Karl-Franzens-Universität Graz

Lecture 2: TEI

What can the TEI do for you?

The TEI provides a framework for the definition of multiple schemas

- it defines and names several hundred useful textual distinctions
- it provides a set of modules that can be used to define schemas making those distinctions
- it provides a customization mechanism for modifying and combining those definitions with new ones using the same conceptual model

---

Where did the TEI come from?

- Originally, a research project within the humanities
  - Sponsored by three professional associations
  - Funded 1990-1994 by US, EU.
- Major influences
  - digital libraries and text collections
  - language corpora
  - scholarly datasets
- International consortium established June 1999 (see http://www.tei-c.org/)

Goals of the TEI

- better interchange and integration of scholarly data
- support for all texts, in all languages, from all periods
- guidance for the perplexed: what to encode — hence, a user-driven codification of existing best practice
- assistance for the specialist: how to encode — hence, a loose framework into which unpredictable extensions can be fitted

These apparently incompatible goals result in a flexible and modular environment

TEI Guidelines

- A set of recommendations for text encoding, covering both generic text structures and some highly specific areas based on (but not limited by) existing practice
- A very large collection of element definitions with associated declarations for various schema languages
- a modular system for creating personalized schemas or DTDs from the foregoing

for the full picture see
    http://www.tei-c.org/Guidelines/

Legacy of the TEI

- a way of looking at what 'text' *really* is
- a codification of current scholarly practice
- (crucially) a set of shared assumptions and priorities about the digital agenda:
  - focus on content and function (rather than presentation)
  - identify generic solutions (rather than application-specific ones)

Users of TEI

- Over 100 projects listed on the TEI project page
- Main areas:
  - digital libraries
  - text-critical editions
  - computer corpora
  - dictionaries

Versions of the Guidelines

- TEI P3 (1994) first public version:
  - SGML + book (1200pp) and soon also on the Web.
- TEI P4 (2002):
  - provides equal *support for XML* and SGML applications using the TEI scheme;
  - error correction, while maintaining backward compatibility: documents conforming to TEI P3 will not become illegal when processed with TEI P4.
- TEI P5 (2007):
  - implements more fundamental changes to the schemas, in line with current practice and identified problems, e.g. uses namespaces
  - no longer backward compatible (but a ~migration P4 to P5 XSLT exists)
  - Relax NG becomes the main schema langauge
  - continuous improvement..

The general structure of TEI documents

- Burnard, Driscoll, Rahtz, TEI Training Course, Sofia 2005:
Slides for TEI overview

TEI Lite

- TEI Lite is a particular parametrisation of TEI that "provides 90% of the elements needed for 90% of users"
- but better to make your own schema..

Lab session

- take personlist data and write XSLT to transform it into teiLite

XSLT

Erjavec: Course at ESSLII 2005 Annotation of Language Resources, Lecture II. XML-Related Recommendations: Formatting and Transforming XML

- ZVON tutorial
- W3schools
- …