



CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2013

Week 8

Today's Class

- Text Encoding Initiative (TEI)

Discussion Analysis

[Folger Digital Texts](#)

Text Encoding Initiative

Consortium

- maintains a technical standard
- set of guidelines
- user wiki
- set of tools for development and processing
- currently predominant in social sciences and humanities, in particular
 - textual studies
 - literary studies
 - linguistic studies

Text Encoding Initiative

- membership
- Special Interest Groups (SIG)
 - currently 10 active groups
 - wiki and mailing list
- Activities
 - various boards and membership groups
 - conferences
 - workgroups
 - project information

Text Encoding Initiative

A gentle introduction to the [TEI Guidelines](#)

- guidelines for electronic text encoding and interchange
- define and document a markup language for representing
 - structural
 - rendition
 - logical & semantic
 - analytic
- rules and recommendations rather than standard
- wide variety of possible solutions for encoding material
- freedom of expression of personal textual theory
- current version = P5
- over 513 elements, 214 attributes...
- customisations such as [TEI Lite](#)

Text Encoding Initiative

A gentle introduction to the Guidelines

- structural features
 - organisation of information in text
- TEI defines overall text structure using
 - front
 - body
 - group
 - back
- a consideration of rendition features

Text Encoding Initiative

A gentle introduction to the Guidelines

- rendition features such as
 - distinct fonts
 - colours
 - alignments
 - italics, underline, bold
 - font weight...
- highlighting

[Malory Example](#)

`<hi>italic words...</hi>`

`<hi rend="italic">italic words...</hi>`

Text Encoding Initiative

A gentle introduction to the Guidelines

- logical and semantic features such as
 - emphasis
 - foreign words
 - linguistically distinct words, phrases
- we can also consider
 - quotation marks
 - quotes
 - cited quotation...

Text Encoding Initiative

A gentle introduction to the Guidelines

- analytical features such as
 - notes and comments
 - marking for indexing
 - regularisation
 - editorial statements

Text Encoding Initiative

TEI Modules - Part 1

1. TEI Infrastructure (tei)
2. Common metadata (header)
3. Common Core (core)
4. Default text structure (textstructure)
5. Character and Glyph Documentation (gaiji)

[Header example](#)

Text Encoding Initiative

TEI Modules - Part 2

- 6. Verse (verse)
- 7. Performance texts (drama)
- 8. Transcribed speech (spoken)
- 9. Print dictionaries (dictionaries)
- 10. Manuscript description (msdescription)

Text Encoding Initiative

TEI Modules - Part 3

- 11. Transcription of primary sources (transcr)
- 12. Text criticism (textcrit)
- 13. Names, Dates, People, and Places (namesdates)
- 14. Tables, Formulae, Figures (figures)
- 15. Metadata for language corpora (corpus)

Text Encoding Initiative

TEI Modules - Part 4

- 16. Linking, segmentation, and alignment (linking)
- 17. Analysis and interpretation (analysis)
- 18. Feature structures (iso-fs)
- 19. Graphs, networks, and trees (nets)
- 20. Certainty and Uncertainty (certainty)
- 21. Documentation elements (tagdocs)

[View elements by Module](#)