DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2014

Week 7

Today's Class

- Discussion Analysis (Folger Digital Texts)
- SciFi Authors Update
- TEI

Discussion Analysis

Folger Digital Texts

SciFi Authors Update

[SciFi Authors Sheet](#)

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

Text Encoding Initiative

Getting Started: Common Data Structure & Elements

- expressed using XML
- elements for information
- attributes for additional information
- comments use standard delimiters
    <!--this is a comment in TEI-->

Text Encoding Initiative

TEI Document

- <teiHeader> = documents all metadata for the TEI document
- <text> = contains the document proper
- these elements are mandatory for all TEI documents
- this structure is contained within the <TEI> element

```
<TEI>
<teiHeader>...</teiHeader>
<text>...</text>
</TEI>
```

Text Encoding Initiative

TEI Header <teiHeader>

- mandatory for all TEI documents
- <fileDesc> = contains a description of the electronic file
- <fileDesc> consists of at least three mandatory elements
  - <titleStmt>
  - <publicationStmt>
  - <sourceDesc>

Text Encoding Initiative

```xml
<teiHeader>
  <fileDesc>
   <titleStmt>
     <title>Around the World in Eighty Days</title>
     <respStmt>
       <resp>editor</resp>
       <name xml:id="PS">Peter Shillingsburg</name>
     </respStmt>
   </titleStmt>
   <publicationStmt>
     <p>Not for distribution.</p>
   </publicationStmt>
   <sourceDesc>
     <p>Transcribed from 1873 English Edition</p>
   </sourceDesc>
  </fileDesc>
</teiHeader>
```

Text Encoding Initiative

```
<!-- transcriber note -->
<note resp="#NH">Possible mis-interpretation of author correction.
</note>

<!-- editorial note -->
<note resp="#AL">correction meant as addition and not
correction</note>
```

Text Encoding Initiative

```
<text>
   <body>
```

- minimally contains <body>
- <body> contains lower level text structures such as
         - <p>
         - or different structures for genres other than prose

```
<text>
<body>
<p>some text...</p>
</body>
</text>
```

Text Encoding Initiative

```
<text>
   <front>
```

- optional element containing front matter such as
         - title pages, headers, prefaces, dedications...
- also consider prologues for drama, forewords and introductions for prose
- except for title, front matter should be encoded using generic standard elements

```
<front>
<div type="dedication">
<p>a small dedication...</p>
</div>
</front>
```

Text Encoding Initiative

```
<text>
   <back>
```

- optional element and grouping
- may contain all back matter for a text
- either numbered or un-numbered divisions with @type attribute

    - appendix
    - glossary (<list> of terms and their explanations)
    - notes
    - bibliogr (list of bibliographical citations <listBibl>)
    - index
    - colophon

```
<back>
<div type="colophon">
<p>Printed and bound by.....</p>
</div>
</back>
```

Text Encoding Initiative

```
<text>
<front>
 <div type="dedication">
   <p>personal dedication...</p>
 </div>
 <div type="contents">
   <head>Table of Contents</head>
   <list>
     <item>1. No.1...</item>
     <item>2. No.2...</item>
     <item>3. No.3...</item>
   </list>
 </div>
</front>
<body>
 <p>some body text goes here....</p>
</body>
<back>
 <div type="colophon">
   <p>Physical book conditions...</p>
 </div>
</back>
</text>
```

Text Encoding Initiative

Unitary or Composite Texts

<group>

- group structurally related text

<teiCorpus>

- corpus of diverse texts

DIGH 400 - Introduction to Digital Humanities Research

```xml
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          <!--Title-->
        </title>
      </titleStmt>
      <publicationStmt>
        <p>
          <!--Publication Information-->
        </p>
      </publicationStmt>
      <sourceDesc>
        <p>
          <!--Information about the source-->
        </p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <!--Some structural division, paragraph, line group, speech, ...-->
    </body>
  </text>
</TEI>
```