TEI By Example

TEI by Example. Module 0

Edward Vanhoutte
Ron Van den Branden
Melissa Terras
# TABLE OF CONTENTS

5. TEI: Ground Rules............................................................................................................. 1
  5.1 Guidelines ................................................................................................................... 1
  5.2 TEI Modules ............................................................................................................... 1
  5.3 Using TEI ................................................................................................................... 3
5. TEI: Ground Rules

5.1 Guidelines

The conclusions and the work of the TEI community are formulated as guidelines, rules, and recommendations rather than standards, because it is acknowledged that each scholar must have the freedom of expressing their own theory of text by encoding the features they think important in the text. A wide array of possible solutions to encoding matters is demonstrated in the TEI Guidelines which therefore should be considered a reference manual rather than a tutorial. Mastering the complete TEI encoding scheme implies a steep learning curve, but few projects require a complete knowledge of the TEI. Therefore, a manageable subset of the full TEI encoding scheme was published as TEI Lite, currently describing 145 elements. 1 Originally intended as an introduction and a didactic stepping stone to the full recommendations, TEI Lite has, since its publication in 1995, become one of the most popular TEI customizations and proves to meet the needs of 90% of the TEI community, 90% of the time.

5.2 TEI Modules

A significant part of the rules in the TEI Guidelines apply to the expression of descriptive and structural meta-information about the text. Yet, the TEI defines concepts to represent a much wider array of textual phenomena, amounting to a total of 503 elements and 210 attributes. These are organized into 21 modules, grouping related elements and attributes:

1. Definition of common datatypes and modular class structures used to define the elements and attributes in the other modules.

2. Definition of the elements that make up the header section of TEI documents. Its major parts provide elements to encode detailed metadata about bibliographic aspects of electronic texts, their relationship with the source materials from which they may have been derived, non-bibliographic details, and a complete revision history.

3. Definition of elements and attributes that may occur in any TEI text, of whatever genre. These elements cover textual phenomena like paragraphs, highlighting and quotation, editorial changes (marking of errors, regularisations, additions), data-like structures (names, addresses, dates, numbers, abbreviations), cross-reference mechanisms, lists, notes, graphical elements, bibliographic references, and passages of verse or drama.

4. Definition of elements and attributes that describe the structure of TEI texts, like front matter and title pages, text body, and back matter.

...  

These may contain further divisions, possibly introduced by headings, salutations, opening formulae, and/or concluded by closing formulae, closing salutations, trailing material and postscripts.

5. Definition of specific provisions for representing characters for which no standardised representation (such as defined by the Unicode Consortium http://www.unicode.org/) exists.

6. Definition of specific elements and attributes for dedicated analysis of verse materials, such as caesurae, metrical systems, rhyme schemes, and enjambments.

7. Definition of specific elements and attributes for dedicated analysis of drama materials. These include provisions for encoding specific phenomena in front and back matter, like details about performances, prologues, epilogues, the dramatic setting, and cast lists. Other drama-specific structures include speeches and stage directions. For multimedia performances, elements for the description of screen contents, camera angles, captions, and sound are provided.

8. Definition of elements and attributes for (general purpose) transcription of different kinds of spoken material. These cover phenomena like utterances, pauses, non-lexical sounds, gestures, and shifts in vocal quality. Besides this, specific header elements for describing the vocal source of the transcription are provided.

9. Definition of elements and attributes for representing dictionaries, with provisions for unstructured and structured dictionary entries (possibly grouped). Dictionary entries may be structured with a number of specific elements indicating homonyms, sense, word form, grammatical information, definitions, citations, usage, and etymology.

10. Definition of specific header and structural elements and attributes for the encoding of manuscript sources. Header elements include provisions for detailed documentation of a manuscript's or manuscript part's identification, heading information, contents, physical description, history, and additional information. Dedicated text elements cover phenomena like catchwords, dimensions, heraldry, watermarks, and so on.

11. Definition of elements and attributes for detailed transcription of primary sources. Phenomena covered are facsimiles, more complex additions, deletions, substitutions and restorations, document hands, damage to the source material and illegibility of the text.

12. Definition of elements and attributes for the representation of (different versions texts as) scholarly editions, listing all variation between the versions in a variant apparatus.

13.
Definition of elements and attributes for more detailed analysis of names of persons, organisations, and places, their referents (persons, organisations, and places) and aspects of temporal analyses.

14. Definition of specific elements and attributes for detailed representation of graphical elements in texts, like tables, formulae, and images.

15. Definition of elements and attributes for the encoding of corpora of texts that have been collected according to specific criteria. Most of these elements apply to the documentation of these sampling criteria, and contextual information about the texts, participants, and their communicative setting.

16. Definition of elements and attributes for representing complex systems of cross-references between identified anchor places in TEI texts. Recommendations are given for either in-line or stand-off reference mechanisms.

17. Definition of elements and attributes that allow the association of simple analyses and interpretations with text elements. Mechanisms for the representation of both generic and particularly linguistic analyses are discussed.

18. Definition of elements and attributes for constructing complex analytical frameworks that can be used to represent specific analyses in TEI texts.

19. Definition of elements and attributes for the analytical representation of schematic relationships between nodes in graphs and charts.

20. Definition of elements for detailed attribution of certainty for the encoding in a TEI text, as well as the identification of the responsibility for these encodings.

21. Definition of elements and attributes for the documentation of the encoding scheme used in TEI texts. This module provides means to define elements, attributes, element and attribute classes, either by changing existing definitions or by creating new ones.

Each of these modules and the use of the elements they define are discussed extensively in a dedicated chapter of the TEI Guidelines.

5.3 Using TEI

Among more technical ones, Steven DeRose pointed out substantial advantages of XML to the TEI community: by allowing for more flexible automatic parsing strategies and easy delivery of electronic documents with cheap ubiquitous tools such as web browsers, XML could spread the notion of descriptive markup to a wide audience that will thus be acquainted with the concepts articulated in the TEI Guidelines.²

......

In order to use TEI for the encoding of texts, users must make sure that their texts belong to the TEI namespace (http://www.tei-c.org/ns/1.0) and adhere to the requirements of the text model proposed by the TEI. In order to facilitate this conformance, it is possible (and strongly suggested) to associate TEI texts with formal representations of this text model. These formal structural grammars of a TEI compatible model of the text can be expressed in a number of ways, commonly referred to as a TEI schema. Technically, a TEI schema can be expressed in a variety of formal languages such as Document Type Definition (http://www.w3.org/TR/REC-xml/#dt-doctype), W3C XML Schema (http://www.w3.org/XML/Schema), or the RELAX NG schema language (http://www.relaxng.org/). It is important to notice that no such thing as ‘the TEI schema’ exists. Rather, users are expected to select their desired TEI elements and attributes from the TEI modules, possibly with alterations or extensions where required. In this way, TEI offers a stable base with unambiguous means for the representation of basic textual phenomena, while providing standardized mechanisms for user customization for uncovered features. It is a particular feature of TEI that these abstract text models themselves can be expressed as TEI texts, using the documentation elements defined in the dedicated module Documentation Elements. A minimal TEI customization file looks as follows:
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>A TBE customisation</title>
        <author>The TBE Crew</author>
      </titleStmt>
      <publicationStmt>
        <p>for use by whoever wants it</p>
      </publicationStmt>
      <sourceDesc>
        <p>created on Thursday 24th July 2008 10:20:17 AM by the form at https://roma.tei-c.org/</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <front>
      <divGen type="toc"/>
    </front>
    <body>
      <p>My TEI Customization starts with modules tei, core, header, and textstructure</p>
      <schemaSpec ident="TBEcustom" docLang="en" xml:lang="en" prefix=""
        <moduleRef key="tei"/>
        <moduleRef key="header"/>
        <moduleRef key="core"/>
        <moduleRef key="textstructure"/>
      </schemaSpec>
    </body>
  </text>
</TEI>

Besides the common minimal TEI structure (<teiHeader> and <text>), a TEI customization file has one specific element which defines the TEI schema (<schemaSpec>). A TEI schema must minimally include the modules which define the minimal TEI text structure: the TEI infrastructure module, the core module with all common TEI elements, the header module defining all teiHeader elements, and the textstructure module defining the elements representing the minimal structure of TEI texts.

In the vein of Literary Programming  http://www.literateprogramming.com/, a TEI customisation file not only contains the formal declaration of TEI elements inside <schemaSpec>, but may also contain prose documentation of the TEI encoding scheme it
defines. Consequently, TEI customisation files are commonly called *ODD files* (One Document Does it all), because they serve as a source for the derivation of

- a formal TEI schema
- human-friendly documentation of the TEI encoding scheme

In order to accommodate the process of creating customised TEI schemas and prose documentation, the TEI has developed a dedicated piece of software called *Roma* [https://roma.tei-c.org/](https://roma.tei-c.org/). This is a dedicated ODD processor, offering an intuitive web-based interface for the creation and basic editing of ODD files, generation of according TEI schemas and prose documentation in a number of presentation formats.

A TEI schema, stating all structural conditions and restraints for the elements and attributes in TEI texts can then be used to automatically validate actual TEI documents with an XML parser. Consider, for example, following fragments:

[A]  
```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>A sample TEI document</title>
      </titleStmt>
      <publicationStmt>
        <publisher>KANTL</publisher>
        <pubPlace>Ghent</pubPlace>
        <date when="2009"/>
      </publicationStmt>
      <sourceDesc>
        <p>No source, born digital</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>This is a sample paragraph, illustrating a <name type="organisation">TEI</name> document.</p>
    </body>
  </text>
</TEI>
```

[B]  
```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <text>
    <body>
      <p>This is a sample paragraph, illustrating a <gi>orgName</gi> TEI <gi>orgName</gi> document.</p>
    </body>
  </text>
</TEI>
```
When validated against a TEI schema derived from the previous ODD file, file [A] will be recognised as a valid TEI document, while file [B] won't:

- The TEI prescribes that the `<teiHeader>` must be present in each document, and that it precedes the `<text>` part.
- The minimal set of TEI modules does not include the specialised `<persName>` element. Although it is a TEI element, using it requires selection of the appropriate TEI module in the ODD file (in this case, the module for *Names, Dates, People, and Places*).