TEI By Example

TEI by Example. Module 1

Edward Vanhoutte
Ron Van den Branden
Melissa Terras

Centre for Scholarly Editing and Document Studies (CTB), Royal Academy of Dutch Language and Literature, Belgium, Gent, 9 July 2010

Last modified on: 2018-04-28
Licensed under a Creative Commons Attribution ShareAlike 3.0 License
<table>
<thead>
<tr>
<th>section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>TEI Header</td>
<td>1</td>
</tr>
<tr>
<td>2.2</td>
<td>Text</td>
<td>2</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Body</td>
<td>2</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Front</td>
<td>3</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Back</td>
<td>3</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Full Example <code>&lt;text&gt;</code></td>
<td>5</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Unitary or Composite Texts</td>
<td>5</td>
</tr>
<tr>
<td>2.2.6</td>
<td>Summary</td>
<td>5</td>
</tr>
</tbody>
</table>
2. General TEI Document Structure

The TEI makes use of XML as its governing metalanguage. This means that all TEI metadata are expressed as XML elements and thus comply with the World Wide Web Consortium XML Recommendation. Information (plain text) is contained in XML elements, delimited by start tags (e.g.: `<TEI>`) and end tags (e.g.: `</TEI>`). Additional information to these XML elements can be given in attributes, consisting of a name (e.g.: `xml:id`) and a value (e.g.: `text1`). XML comments are delimited by start markers (`<!--`) and end markers (`-->`).

A full TEI document consists of a `<teiHeader>`, documenting all the metadata describing it, and a `<text>`, containing the document proper. This common structure is mandatory for all TEI documents. This basic structural pair is contained by a `<TEI>` element:

```xml
<TEI xmlns="_TBEeg_http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!--...-->  
  </teiHeader>

  <text>
    <!--...-->
  </text>

</TEI>
```

2.1 TEI Header

The TEI header (<teiHeader>) is mandatory and contains descriptive meta-information about the document. The `<teiHeader>` minimally contains a description of the electronic file inside a `<fileDesc>`. The latter element consists of three mandatory components:

- the title statement (<titleStmt>), providing information about the title (<title>), author (<author>) and others responsible for the electronic text
- the publication statement (<publicationStmt>), providing publication details about the electronic text in a structured way or as prose inside a paragraph (<p>)
- a description of the source (<sourceDesc>), documenting bibliographic details about the electronic text's material source (if any) in a structured way or in a prose paragraph (<p>)

......

1 This is an example of a TEI XML text, representing both information and meta-information. This example, as any TEI text, is recognizable as a TEI text by the outermost `<TEI>` element, which is declared in the dedicated TEI namespace (http://www.tei-c.org/ns/1.0).
The actual text (\text{}) contains a single text of any kind. This commonly contains the actual text and other encodings. A text \text{} minimally contains a text body (\textbf{body}). The body contains lower-level text structures like paragraphs (\textbf{p}), or different structures for text genres other than prose: lines for poetry, speeches for drama.

\begin{quote}
For the first time in twenty-five years, Dr Burt Diddledygook decided not to turn up to the annual meeting of the Royal Academy of Whoopledywhaa (RAW). It was a sunny day in late September 1960 bang on noontime and Dr Burt was looking forward to a stroll in the park instead. He hoped his fellow members of the RAW weren't even going to notice his absence.
\end{quote}
2.2.2 Front

Next to the <body>, a text can optionally contain front matter which may be encoded with <front>. Clear examples are title pages, headers, prefaces, or dedications. Prologues in drama or forewords and introductions in prose may also be considered prefatory material. May, because the encoder may choose simply to ignore to encode the front matter of a text as such. With exception of the title page, for which the TEI defines specific elements, front matter should be encoded using the same elements as the rest of a text. This means that there are no dedicated elements to encode prefaces, dedications, abstracts, frontispieces etc. Instead, either numbered or un-numbered divisions <div> with an attribute @type are used to distinguish the different components of a <front>. The following suggested values for the @type attribute may be used for this purpose:

- preface: a foreword or preface addressed to the reader
- ack: a formal declaration of acknowledgement by the author
- dedication: a formal offering or dedication of a text by the author
- abstract: a summary of the content of a text as continuous prose
- contents: a table of contents. A <list> element should be used to mark its structure
- frontispiece: a pictorial frontispiece, possibly including some text

```
<front>
  <div type="dedication">
    <p>In memory of Lisa Wheeman.</p>
  </div>
  <div type="contents">
    <head>Table of Contents</head>
    <list>
      <item>I. The Decision</item>
      <item>II. The Fuss</item>
      <item>III. The Celebration</item>
    </list>
  </div>
</front>
```

2.2.3 Back

All back matter to a text may be grouped within <back>. As is the case with <front>, either numbered or un-numbered divisions <div> with a @type attribute are used to distinguish the different components. The following attribute values may be supplied for the @type in order to distinguish various kinds of division characteristic of back matter:

- appendix: an appended self-contained section of a work, often providing additional information or text
• **glossary**: contains a list `<list>` of terms and their explanations
• **notes**: a section in which textual or other kinds of notes are gathered together
• **bibliogr**: contains a list of bibliographical citations `<listBibl>`
• **index**: any form of index to the work
• **colophon**: a statement appearing at the end of a book describing the conditions of its physical production

```html
<div type="colophon">
  <p>Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South Africa.</p>
</div>
```
2.2.4 Full Example <text>

```xml
<text>
  <front>
    <div type="dedication">In memory of Lisa Wheeman.</div>
  </front>
  <div type="contents">
    <head>Table of Contents</head>
    <list>
      <item>I. The Decision</item>
      <item>II. The Fuss</item>
      <item>III. The Celebration</item>
    </list>
  </div>
  <body>
    For the first time in twenty-five years, Dr Burt Diddledygook decided not to turn up to the annual meeting of the Royal Academy of Whoopledywhaa (RAW). It was a sunny day in late September 1960 bang on noontime and Dr Burt was looking forward to a stroll in the park instead. He hoped his fellow members of the RAW weren't even going to notice his absence.
  </body>
  <back>
    <div type="colophon">Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South Africa.</div>
  </back>
</text>
```

2.2.5 Unitary or Composite Texts

Apart from simple texts, TEI provides means to encode composite texts, either by grouping structurally related texts in a `<group>` element inside `<text>`, or treating them as a corpus of diverse texts, using `<teiCorpus>` as the outermost element.

2.2.6 Summary

The following example shows the empty framework of a basic TEI document structure:
The following example fills this empty framework with the text of the examples:
<TEI xmlns="_TBEeg_http://www.tei-c.org/ns/1.0"/>
<teiHeader>
<fileDesc>
<titleStmt>
<title>The Strange Adventures of Dr. Burt Diddledygook: a machine-readable transcription</title>
<respStmt>
<resp>editor</resp>
<name xml:id="EV">Edward Vanhoutte</name>
</respStmt>
<titleStmt>
<publicationStmt>
<p>Not for distribution.</p>
</publicationStmt>
<sourceDesc>
<p>Transcribed from the diaries of the late Dr. Roy Offire.</p>
</sourceDesc>
</fileDesc>
</teiHeader>
<text>
<front>
<div type="dedication">
<p>In memory of Lisa Wheeman.</p>
</div>
<div type="contents">
<head>Table of Contents</head>
<list>
<item>I. The Decision</item>
<item>II. The Fuss</item>
<item>III. The Celebration</item>
</list>
</div>
</front>
<body>
<p>For the first time in twenty-five years, Dr Burt Diddledygook decided not to turn up to the annual meeting of the Royal Academy of Whoopledywaan (RAW). It was a sunny day in late September 1960 bang on noontime and Dr Burt was looking forward to a stroll in the park instead. He hoped his fellow members of the RAW weren't even going to notice his absence.</p>
</body>
<back>
<div type="colophon">
<p>Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South Africa.</p>
</div>
</back>