TEI By Example

TEI by Example. Module 2

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 OF CONTENTS

4. The Header of a Complex Text........................................................................................................... 1
Before we end, let's go back to where we left you: the library, in front of the library catalogue or computer screen. Prepared for the possibility that this exemplar may be in loan, you find another reference to *The Wild Ass's Skin* in following record (look for "505 8 0 |gPhilosophic and analytic studies: v. 41. The|magic skin "; *The Magic Skin* is an alternative title for the English translation):
Now that's a record! If you thought the truckload of possibilities for the description of electronic texts in the TEI header set your head spinning already, imagine what an electronic edition of *La Comédie Humaine* might look like! Code 300 tells us that it has no less than 53 volumes, with different titles per volume.
One way of encoding this majestic work would be to treat *La Comédie Humaine* as a kind of ‘supertext’ containing all different works. This can be done in TEI by treating the whole as a `<teiCorpus>`, containing each separate work in its own `<TEI>` text. As each of these `<TEI>` texts needs its own TEI header, you can imagine the amount of meta-information, much of which will have to be repeated. This can be avoided by placing the common meta-information in the `<teiHeader>` element of the `<teiCorpus>` element, while retaining all work-specific meta-information in the TEI header section of the respective `<TEI>` text. This mechanism allows you to be maximally expressive in the description of all texts in a TEI corpus, *and* maximally efficient in the reduction of the individual TEI headers.

Following example gives an impression of what the TEI header for an electronic edition of *La Comédie Humaine* might look like:
<teiCorpus>
  <!-- general TEI header information or the entire corpus -->
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>La Comédie Humaine</title>
        <author xml:id="HdB">Honoré de Balzac</author>
        <editor role="editor" xml:id="TBEcrew">The TBE crew</editor>
        <respStmt>
          <name xml:id="RvdB">Ron Van den Branden</name>
          <resp>transcription</resp>
          <resp>annotation</resp>
        </respStmt>
        <sponsor>Association for Literary and Linguistic Computing (ALLC)</sponsor>
        <sponsor>Centre for Computing in the Humanities (CCH) - King's College London</sponsor>
        <sponsor>University College London (UCL)</sponsor>
      </titleStmt>
      <extent>0.5 Gb</extent>
      <publicationStmt>
        <publisher>Centre for Scholarly Editing and Document Studies (CTB)</publisher>
        <distributor>Centre for Computing in the Humanities (CCH) - King's College London</distributor>
        <pubPlace>Gent</pubPlace>
        <address>
          <name type="institution">Centre for Scholarly Editing and Document Studies (CTB)</name>
          <name type="institution">Royal Academy of Dutch Language and Literature</name>
          <street>Koningstraat 18</street>
          <postCode>9000</postCode>
          <name type="city">Gent</name>
          <name type="country">Belgium</name>
        </address>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
</teiCorpus>
<idno type="ISBN">0-00-000000-9</idno>
<availability status="free">
<p>Published under a <ref target="http://creativecommons.org/licenses/by-sa/3.0/">Creative Commons Attribution ShareAlike 3.0 License</ref>.</p>
</availability>
<date when="2110-01-01">1 January 2110</date>

<title>The TBE collection: sample texts encoded with TEI</title>
<idno type="ISSN">0000-0001</idno>
<respStmt>
<name>Edward Vanhoutte</name>
<resp>compiler</resp>
</respStmt>

<note>OCR scanning done at KANTL, Gent.</note>

<sourceDesc>
<bibl>
<title>La Comédie Humaine</title> by <author>Honoré de Balzac</author>.
<pubPlace>London</pubPlace> : <publisher>Caxton</publisher>, <date from="1895" to="1900">1895-1900</date>. <extent>53 v.</extent>.</bibl>
</sourceDesc>

<encodingDesc>
<editorialDecl>
<correction method="markup">
<p>Apparent errors have been corrected using the <gi>sic</gi> <gi>corr</gi> elements, wrapped in a <gi>choice</gi> element.</p>
</correction>

<normalization method="markup" source="http://www.oed.com/">
<p>Spelling has been modernised using the <gi>orig</gi> / <gi>reg</gi> elements, wrapped in a <gi>choice</gi> element.</p>
</normalization>

<quotation marks="all">
<p>Diplomatic transcription, all original quotation marks have been retained and normalised to double quotation marks.</p>
</quotation>

<hyphenation eol="none">
<p>End-of-line hyphenation has been removed. All other hyphenation has been retained.</p>
</hyphenation>
</editorialDecl>

</encodingDesc>

4. The Header of a Complex Text
Thematic analysis added, studying the main motifs.

Names and dates are marked.

Text encoded for the TBE collection: sample texts encoded with TEI, aiming at providing a collection of prime exemplar TEI encoded materials.

The paragraphs in the text are numbered with the attribute. Each number consists of four digits; numbering is consecutive throughout the book. For example: numbers the 203th paragraph throughout the book.

Each chapter is identified with a formal identification code inside the attribute. Chapters are numbered using arabic numerals. The codes are composed by concatenating the identification codes for all ancestor text divisions down to the chapter level, with the dot as separation marker. For example: identifies the third chapter of the second book of the first volume.

The paragraphs in the text are numbered with the attribute. Each number consists of four digits; numbering is consecutive throughout the book. For example: numbers the 203th paragraph throughout the book.

Each chapter is identified with a formal identification code inside the attribute. Chapters are numbered using arabic numerals. The codes are composed by concatenating the identification codes for all ancestor text divisions down to the chapter level, with the dot as separation marker. For example: identifies the third chapter of the second book of the first volume.

The paragraphs in the text are numbered with the attribute. Each number consists of four digits; numbering is consecutive throughout the book. For example: numbers the 203th paragraph throughout the book.

Each chapter is identified with a formal identification code inside the attribute. Chapters are numbered using arabic numerals. The codes are composed by concatenating the identification codes for all ancestor text divisions down to the chapter level, with the dot as separation marker. For example: identifies the third chapter of the second book of the first volume.
<language ident="en" usage="98">English</language>
<language ident="fr" usage="1">French</language>
<language ident="ar" usage="1">Arabic</language>
</langUsage>
</profileDesc>
<revisionDesc>
<change when="2109-12-20" who="#RvdB">final proofing</change>
<change when="2009-03-08" who="#MT">addition of thematic analysis</change>
<change when="2009-03-08" who="#RvdB">addition of explanatory notes</change>
<change when="2008-10-25" who="#RvdB">spell check</change>
<change when="2008-08-25" who="#RvdB">addition of phrase level markup</change>
<change when="2008-08-20" who="#RvdB">file creation</change>
</revisionDesc>
</teiHeader>
<!-- ... -->
<TEI>
<!-- specific TEI header information for the distinct TEI documents -->
<teiHeader>
<fileDesc>
<titleStmt>
<title>The Wild Ass's Skin: an electronic edition</title>
<editor role="translator" xml:id="EM">Ellen Marriage</editor>
<editor role="preface" xml:id="GS">George Saintsbury</editor>
</titleStmt>
<editionStmt>
<edition n="2.0">
<title>Version 2.0, enriched with thematic annotations.</title>
<date when="2010">2010</date>
</edition>
<respStmt>
<name>Melissa Terras</name>
<resp>Added thematic annotations.</resp>
</respStmt>
</editionStmt>
</fileDesc>
<extent>572 Kb</extent>
<publicationStmt>
<idno type="ISBN">0-00-000000-0</idno>
<date when="2010-01-01">1 January 2010</date>
</publicationStmt>
<sourceDesc>
<bibl>
title>The Wild Ass's Skin</bibl> by <author>Honoré de Balzac</author>.
<pubPlace>London</pubPlace> : <publisher>Dent</publisher>, <date when="1906">1906</date>.
4. The Header of a Complex Text
Summary

A complex text encoded as a `<teiCorpus>` should have a `<teiHeader>` in its own right. This TEI header on the corpus level can contain the general descriptive information about all corpus texts embedded as `<TEI>` documents. Each corpus text then should have its own `<teiHeader>`, describing only those aspects that are specific to that text.

1 The TEI Guidelines provide even more refined ways of associating contextual information with specific (parts of) texts. See section 15.3 Associating Contextual Information with a Text for more information.