TEI By Example

TEI by Example. Module 7

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

2. Textual variation................................................................................................................ 1
2. Textual variation

Similar to all other TEI modules, the elements and attributes defined in the TEI textcrit module can be used for the encoding of existing source materials (be they in print or digital form), or the encoding of electronic documents from scratch. However, the use of this module in the context of electronic critical editing, adds another perspective to this traditional authorial/editorial angle (see ¹). Electronic critical editions can be created from scratch either by encoding different primary source materials straight as a critical edition, or by generating the edition from previously encoded electronic transcriptions of those materials as independent texts in their own right. Therefore, the tags defined in the TEI textcrit module can be used to:

- **digitise** an existing print edition
- **create** an electronic edition, e.g. by recording some or all of the known variations among different witnesses to the text in a critical apparatus of variants
- **generate** an electronic edition from encoded transcriptions of the documentary source material

In the examples in this TBE module, "critical editing with TEI" will be understood as the act of encoding material sources in a TEI representation that allows for the creation or generation of a digital edition in some form (using any output format in the digital medium, e.g. HTML pages, PDF, flash movies,...), rather than digitising an existing critical edition. In this sense, the authorial/editorial angle of this TBE module differs from that of the other modules (focusing on the digitisation of a material source text in a certain genre). However, the strategies discussed in this tutorial for representing textual variation can equally be applied to the digitisation of existing critical editions. Where there are differences, these will be pointed out explicitly.

For example, consider following texts:

Some of these images may look more or less familiar to you: they are facsimiles from the first page of chapter 2 of the printed TEI Guidelines throughout their different incarnations, from version P2 (1992) to the latest version, P5 (2009). As you can imagine, the technological evolutions of these 17 years have prompted considerable changes to this chapter that introduces the technological background of text encoding with TEI, ranging from rephrasing, addition or deletion of notes, changes in italicisation, restructuring of paragraphs, etc. One way of approaching this textual variation could consist of encoding these text versions as physically distinct TEI documents, in which corresponding text structures could be aligned by a common identification mechanism. For example, the first couple of paragraphs in these 4 text witnesses could be encoded in different TEI documents as follows:
The encoding scheme defined by these Guidelines is formulated as an application of a system known as the Standard Generalized Markup Language (SGML). Although widely said to be short for the surnames of its progenitors, the official expansion of this abbreviation is "Standard Generalized Markup Language." SGML is an international standard for the definition of device-independent, system-independent methods of representing texts in electronic form. This chapter presents a brief tutorial guide to its main features, for those readers who have not encountered it before. For a more technical account of TEI practice in using the SGML standard, see chapter 30, "Conformance," on page 727. For a more technical description of the subset of SGML used by the TEI encoding scheme, see chapter 39, "Formal Grammar for the TEI-Interchange-Format Subset of SGML," on page 1247.
Before going any further we should define these terms. 

Historically, the word markup has been used to describe annotation or other marks within a text intended to instruct a compositor or typist how a particular passage should be printed or laid out. Examples include wavy underlining to indicate boldface, special symbols for passages to be omitted or printed in a particular font and so forth. As the formatting and printing of texts was automated, the term was extended to cover all sorts of special markup codes inserted into electronic texts to govern formatting, printing, or other processing.
2 A Gentle Introduction to XML

As originally published in previous editions of the Guidelines, this chapter provided a gentle introduction to ‘just enough’ SGML for anyone to understand how the TEI used that standard. Since then, the Gentle Guide seems to have taken on a life of its own independent of the Guidelines, having been widely distributed (and flatteringly imitated) on the web. In revising it for the present draft, the editors have therefore felt free to reduce considerably its discussion of SGML-specific matters, in favour of a simple presentation of how the TEI uses XML.

The encoding scheme defined by these Guidelines may be formulated either as an application of the ISO Standard Generalized Markup Language (SGML) or of the more recently developed W3C Extensible Markup Language (XML). Both SGML and XML are widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web.

In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines.

Strictly speaking, XML is a metalanguage, that is, a language used to describe other languages, in this case, markup languages. Historically, the word markup has been used to describe annotation or other marks within a text intended to instruct a compositor or typist how a particular passage should be printed or laid out. Examples include wavy underlining to indicate boldface, special symbols for passages to be omitted or printed in a particular font, and so forth. As the formatting and printing of texts was automated, the term was extended to cover all sorts of special codes inserted into electronic texts to govern formatting, printing, or other processing.
system-independent methods of storing and processing texts in electronic form; XML being in fact a simplification or derivation of SGML. In the present chapter we introduce informally the basic concepts underlying such markup languages and attempt to explain to the reader encountering them for the first time how they are actually used in the TEI scheme. Except where the two are explicitly distinguished, references to XML in what follows may be understood to apply equally well to the TEI usage of SGML. For a more technical account of TEI practice see chapter 28 <hi>Conformance</hi>; for a more technical description of the subset of SGML used by the TEI encoding scheme, see chapter 39 <hi>Formal Grammar for the TEI-Interchange-Format Subset of SGML</hi>.<p>

XML is an extensible markup language used for the description of marked-up electronic text. More exactly, XML is a <hi>metalanguage</hi>, that is, a means of formally describing a language, in this case, a <hi>markup language</hi>. Historically, the word <hi>markup</hi> has been used to describe annotation or other marks within a text intended to instruct a compositor or typist how a particular passage should be printed or laid out. Examples include wavy underlining to indicate boldface, special symbols for passages to be omitted or printed in a particular font and so forth. As the formatting and printing of texts was automated, the term was extended to cover all sorts of special codes inserted into electronic texts to govern formatting, printing, or other processing.</p>

This would allow for maximal representation of the distinct material sources, and leave the identification of the actual variation either to further processing or human inspection. A variant of this approach could integrate the transcriptions of the text in all material witnesses in a single
TEI by Example. Module 7

2. Textual variation

TEI document, and make use of appropriate linking attributes to point out the alignment between the different text structures. In their naivety, such systems are both redundant and crude. While providing all text of all text witnesses, and aligning the corresponding text structures, they provide little insight in the places where the different witnesses actually differ.

In order to encode the actual textual variation between the different text versions in a meaningful way, the TEI Guidelines provide a specialised module of elements and attributes that allow you to encode textual variation at word level. This TBE tutorial will first discuss how to describe the different text witnesses represented in the critical edition; next deal with the encoding of textual variants between these witnesses in isolation; then treat different ways of integrating such records of variation within the encoding of the critical edition; and finally point out potential problems and pitfalls when creating a critical edition with TEI.