TEI By Example

TEI by Example. Module 0

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

1. Introduction........................................................................................................................ 1
1. Introduction

Computers can only process texts whose characters are represented by a system that relates to the binary system computers can interpret. This is called character encoding. One such character encoding scheme based on the English alphabet is ASCII (American Standard Code for Information Interchange). Character encoding facilitates the storage of text in computers and the transmission of text through telecommunication networks. Character encoding, however, does not say anything about the semantics, interpretation or structure of a text. Such information on a text is called meta-information. If we want to add any meta-information to a text so that it can be processed by computers, we need to encode or markup texts. We can do this by inserting natural language expressions (or codes representing them) in the text with the same character encoding the text is using, but separated from the text by specific markers. One such an expression, we call a tag. All of the tags used to encode a text together constitute a markup language. The application of a markup language to a text, we call text encoding.

The Text Encoding Initiative (TEI) is a standard for the representation of textual material in digital form through the means of text encoding. This standard is the collaborative product of a community of scholars, chiefly from the humanities, social sciences, and linguistics who are organized in the TEI Consortium (TEI-C http://www.tei-c.org). The TEI Consortium is a non-profit membership organization and governs a wide variety of activities such as the development, publication, and maintenance of the text encoding standard documented in the TEI Guidelines, the discussion and development of the standard on the TEI mailing list (TEI-L) and in Special Interest Groups (SIG), the gathering of the TEI community on yearly members meetings, and the promotion of the standard in publications, on workshops, training courses, colloquia, and conferences. These activities are generally open to non-members as well.

By ‘TEI Guidelines’ one may refer both to the markup language and tag set proposed by the TEI Consortium and to its documentation online or in print. Informally ‘TEI Guidelines’ is often abbreviated to ‘TEI’. In this article ‘TEI Guidelines’ is used as the general term for the encoding standard. The TEI Guidelines are widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. Since the TEI is expressed in terms of the eXtensible Markup Language (XML) and since it provides procedures and mechanics to adapt to one’s own project needs, the TEI Guidelines define an open standard that is generally applicable to any text and purpose.

This introductory module first introduces the concepts of text encoding and markup languages in the humanities and then introduces the TEI encoding principles. Next, the article provides a brief historical survey of the TEI Guidelines and ends with a presentation of the Consortium's organization.