Teaching TEI: The Need for TEI by Example

Melissa Terras, Ron van den Branden and Edward Vanhoutte

m.terras@ucl.ac.uk

Abstract

The TEI (Text Encoding Initiative)\(^1\) has provided a complex and comprehensive system of provisions for scholarly text encoding. Although a major focus of the “digital humanities” domain, and despite much teaching effort by the TEI community, there is a lack of teaching materials available which would encourage the adoption of the TEI’s recommendations and the widespread use of its text encoding guidelines in the wider academic community. This paper describes the background, plans, and aims of the TEI by Example project, and why we believe it is a necessary addition to the materials currently provided by the TEI itself. The teaching materials currently available are not suited to the needs of self directed learners, and the development of stand alone, online tutorials in the TEI are an essential addition to the extant resources, in order to encourage and facilitate the uptake of TEI by both individuals and institutions.

1. Introduction

Over the past 20 years, the TEI (Text Encoding Initiative)\(^2\) has developed comprehensive guidelines for scholarly text encoding (TEI, 2007a). In order to expand the user base of TEI, it is important that tutorial materials are made available to scholars new to textual encoding. However, there is a paucity of stand alone teaching materials available which support beginner’s level learning of TEI. Materials which are available are not in formats which would enable tutorials to be provided in classroom settings (such as through part of a University course), or allow individuals to work through graded examples in their own time: the common way of learning new computational techniques through self-directed learning. As a result, there is an urgent need for a suite of TEI tutorials for the self directed learner. The “TEI by Example”\(^2\) project is currently developing a range of freely available online tutorials which will walk individuals through the different stages in marking up a document in TEI. To do so, the development environment will need to be explained, documented, and links to freely available software provided to allow users to undertake TEI based markup themselves. In addition to this, the tutorials will provide annotated examples of a range of texts, indicating the editorial choices necessary when marking up a text in TEI. Linking to real examples from projects which utilise the TEI will reaffirm the advice given to learners.

In this paper, we discuss the current methods of teaching TEI, and why these do not cater for the lone scholar or self directed learner interested in learning TEI in their own environment and in their own time. We discuss the need for specifically designed

\(^1\) http://www.tei-c.org/

\(^2\) http://www.teibyexample.org
online tutorials in the TEI, and why it is important to incorporate example material of TEI code in these tutorials. Finally, an overview of the TEI by Example project is given, discussing its aims, structure, deliverables, and future work.

2. Teaching the TEI

The Text Encoding Initiative, “an international organization founded in 1987 to develop guidelines for encoding machine-readable texts in the humanities and social sciences” (TEI, 2007b), has produced a variety of guidelines for the encoding of scholarly texts. The rigorous intellectual endeavour to create the guidelines ensures that “The TEI is a very extensive encoding language and is intended to support very complex encoding of very complex documents” (TEI, 2007b). As a result, the TEI has the potential to be used in a variety of situations. Markup projects often train their workforce in the principles and theory of markup, with encoders learning on the job. Students in Literature and Language and other Humanities based subjects may have a need for TEI. Students in Library, Archives, and Electronic Communication and Publishing, and Librarians and Archivists, may benefit from understanding how best to encode, document, and ultimately preserve electronic textual data (and the widespread inclusion of information professionals in the TEI community would further the aims of the initiative, by encouraging the uptake of TEI as an aid to preserve electronic textual data). Academics wishing to join the “digital revolution” may have an introduction to the field of Digital Humanities through the discipline of textual markup. Fundamentally, many of these individuals who come across the TEI may go on to teach or inform others, and as a result, if we wish to expand the user community and the use of TEI, it is important to provide training, and teaching materials, which foster and build confidence, and demonstrate the use and usefulness of the TEI guidelines. However, individuals wishing to learn the TEI are currently faced with the lengthy and technically descriptive guidelines, which are hardly written with the absolute beginner in mind. An alternative is to attend a taught course regarding the TEI, or to consult the materials made available on the TEI website.

University courses sometimes integrate TEI into their teaching (for example, the “Humanities Computing: Electronic Text” undergraduate course at the University of Antwerp, or the “Digital Resources in the Humanities” Masters level module, in the School of Library, Archive and Information Studies, University College London) although this is rare, and access is limited to a few interested students. More commonly, short courses are sometimes sponsored and provided by the TEI, or related organisations: for example the workshops organised by the Centre for Scholarly Editing and Document Studies, and those ran by the Brown Women Writers Project “offers periodic hands-on workshops on text encoding and the TEI Guidelines. These range from one to five days and cover a range of topics in basic and intermediate TEI encoding, TEI customization, basic XSLT, and issues in text encoding theory” (WWP, 2007)., for instance at the yearly Digital Humanities Summer Institute. An archive of a range of documentation from workshops,
including presentations, exercises, and handout materials, is maintained on the TEI website\(^7\). However, short term courses have their own pedagogical problems: they are rarely assessed, so it is difficult to know if the students have really learnt anything useful that will be retained. When the intensity of the course ends, students may go back to their old habits. When motivational tutors are no longer around to ask when things go wrong, it may be the case that students give up their attempt to learn the TEI. There is little room for a “holistic” approach to teaching (Bernold, n.d.) where what is learnt can be reinforced over a period of time through a variety of pedagogical methods such as evaluation, feedback, discussion, experimentation, and teamwork.

Although the materials emanating from the TEI workshops are available online, until recently the design of the TEI website dissuaded many potential new users from learning TEI. The old website comprised of a multitude of broken or misleading links, maze like structures, and dated tutorials (in early 2007, the latest introductory material available on the site presented outdated and therefore erroneous material to novices: Sperberg-McQueen and Burnard (2002a), and Sperberg-McQueen and Burnard (2002b)). The new website design, launched in October 2007, presents a cleaner and more modern face to the TEI. An up to date generic tutorial, which also features in the TEI guidelines itself, is now available on the TEI website (Sperberg-McQueen and Burnard 2007). However, this tutorial, and links to workshop materials, are still buried deep within misleadingly titled menu items, and unlikely to be found instantly by potential new users. Although there has been much time devoted to teaching TEI, and the preparation of teaching materials for lectures and workshops, the online presence of the TEI suggests that reaching out to new users is not high on the TEI’s agenda. This may not be the intention, but the design and focus of the website is not welcoming to those new to the concept of textual markup, or from outside the existing TEI community.

Additionally, there is the problem that the retrospective posting of workshop materials on a website is not the same, for users, as actually attending a workshop. The nuances of bullet points on PowerPoint slides are lost when the presenter is not there to explain their meaning. There is no room for feedback, or for any communication of any sort with the course leader. Exercises which may have been clear in the classroom, when a computing environment was provided, may be impossible for those attempting them alone, on a different system. As a result most of the workshop materials posted online are intimidating rather than illuminating, and serve more to act as an archive for the TEI for teaching activities than to provide learning materials for those wishing to learn TEI unaided. Although it makes sense to offer online materials for distance learners, these have to be tailored to the needs of online users.

Online materials need to take a different form from face to face teaching materials, as the online experience is different to that presented in the traditional classroom because: learners are different; the communication is via computer and World Wide Web; the social dynamic of the learning environment is changed; feedback mechanisms function differently; there is the potential to reach a much wider audience; and there is the potential for re-use in other learning environments. As a result, instructors wishing to provide such materials online should “master design and

\(^7\) http://www.tei-c.org/Support/Learn/tutorials.xml#body.1_div.3
delivery strategies, techniques, and methods for teaching online courses” (Yang and Cornelius 2005).

If the digital humanities community wants to promote the TEI markup framework as a serious tool for digital humanities, humanities computing, digital culture, or humanities informatics, to name just a few of the labels this archipelago of disciplines gets (McCarty 2006), and to expand the use of the TEI guidelines beyond the reaches of the TEI community as it stands, then there is an urgent need for an online TEI course which is less generic than the introduction published within the TEI guidelines and more user friendly, comprehensive, and interactive than the online workshop materials which are currently presented as stand alone teaching materials.

The demand for such introductory material can be illustrated by the popularity of a paper published in a special edition of LLC: “An Introduction to the TEI and the TEI Consortium” (Vanhoutte, 2004). This was reviewed by Matthew Driscoll, thus:

This is followed by a short introduction by Edward Vanhoutte to the Text Encoding Initiative (TEI) in general. There are many such introductions to the TEI available both in print and on the web and this one is fine so far as it goes, but one may wonder about its appropriateness here, given that few readers of LLC are likely to be so unfamiliar with the TEI as to require such an introduction. (Driscoll, 2005, p. 337)

However, this introductory paper to TEI was consistently amongst the top ten articles requested online from LLC in the three years following publication.

3. The Need for TEI by Example

The TEI by Example project was conceived after a difficult teaching session. (It should be noted that, in many cases at University level, those teaching a course may not be subject experts in all aspects of a field, and rely on appropriate resources and teaching materials to assist them in the areas they are weaker in. This is a fact of University teaching life, where academics are asked to teach broadly across a discipline whilst tending to focus on one aspect of the discipline as a research topic). In this case, TEI was being taught by a lecturer who has used TEI in the past, but not on a day to day basis. An intelligent and articulate Masters level student asked whether “TEI was a theoretical exercise on the principles and theory of textual markup” as, although many projects purported to be using the TEI, there are very few examples of source code which are available for those learning the TEI to consult.

Most projects marking texts up in TEI deliver their texts via the Internet: which means their code is transformed, via XSLT, into HTML or XHTML. Interested users can generally only see this transformed version, and so cannot inspect and learn from the underlying TEI code.8 At time of writing, the TEI wiki page which hosts sample texts from those utilising the TEI framework features only eight projects willing to make some of their marked up texts available to the general public9. The Oxford Text Archive10, which collects, catalogues, and preserves electronic text for use by the

---

10 http://ota.ahds.ac.uk/
research community, has a few TEI marked up texts (of any useful granularity, excluding those with just a TEI Header which is added to all texts by the OTA themselves) available, but it is impossible to find these through searching the website, and these texts are only available by contacting the OTA and requesting TEI marked up texts (Cummings, 2007). This is a loss to users, who would benefit from seeing both good and bad examples of markup, to learn to encode by example. It is understood that much intellectual and temporal effort goes into marking up textual material with suitable granularity to facilitate in depth analysis and manipulation of textual material, and that projects may not wish to make this investment public. However, being able to view the markup approaches of established scholars and projects in the field is an essential tool for TEI teaching which is currently not utilised.

Learning a computing language (especially through self-directed learning) is usually accomplished through examining and working through examples. Learning by example is effectively an implementation of problem based learning, an efficient and useful approach to teaching skills to individuals in order for them to undertake similar tasks themselves, successfully. The literature on this is wide and varied\(^\text{11}\). The paucity of TEI examples currently available to learners can be contrasted with the teaching literature for computing: at time of writing, there were 837 titles available on Amazon.co.uk with the words “by example” in the title: most were featured in the Computers and Internet section. 756 computing books had “case studies” in the title. There has been particular consideration as to the effectiveness of example and problem based learning when learning computer programming (for example, see Mayer, 1981, 1988; Kelleher and Pausch, 2005). Even the fictional LOLcode programming language, constructed as a joke after the popularity of the internet meme LOLcats, where instant messaging English is used to caption cute pictures of cats, has a variety of examples of code available which users can scrutinise to learn LOLOcode for themselves\(^\text{12}\). If learning by example is such a fundamental approach to learning a computing language, where is TEI by Example?

Additionally, the development of any online teaching course would need to understand how to develop online materials successfully, and how this may differ from more traditional teaching and learning environments (Stephenson, 2001; Jochems \textit{et al.}, 2003). Understanding the nature of online tutorials, and grappling with the pedagogical issues these technologies offer, is a core issue when beginning to consider the construction of a TEI by Example online course.

The need for introductory training materials regarding text encoding within the Text Encoding Initiative framework, and the present lack of appropriate teaching resources, spurred us to create TEI by Example ourselves. We aim to produce an online TEI course by example which will introduce novice users to text encoding within the TEI framework, and serve as an introductory teaching package for instructors in the classroom, while presenting the user with real, annotated examples from encoding projects. Additionally, we will need to make a software toolkit available for teaching text encoding, to support interested trainers and learners. Investigating the affordances

\(^{11}\) For seminal literature regarding the effectiveness of problem based learning as a pedagogic approach see Norman and Schmidt (1992), Garrison (1997), and Savin-Baden and Wilkie (2006).

\(^{12}\) \url{http://lolcode.com/}
of online teaching tools (such as quizzes and interactive feedback) will also aid in the creation of useful learning materials for those who wish to undertake textual markup using the TEI guidelines.

4. TEI by Example Development

The “TEI by Example” project is currently developing a range of freely available online tutorials walking individuals through the different stages in marking up a document in TEI. Project partners are The Centre for Scholarly Editing and Document Studies (CTB)\(^\text{13}\) of the Royal Academy of Dutch Language and Literature, the Centre for Computing in the Humanities (CCH)\(^\text{14}\) of King’s College London, and the School for Library, Archive, and Information Studies (SLAIS)\(^\text{15}\) of University College London, with an international advisory board consisting of experts in textual encoding and markup\(^\text{16}\). The development team consists of the project leaders, Melissa Terras and Edward Vanhoutte, and the executive project officer, Ron Van den Branden. The deliverables will be published and hosted by CCH (King’s College London) under endorsement by the Association of Literary and Linguistic Computing (ALLC)\(^\text{17}\). A small amount of funding has been procured from the CCH, the ALLC, and the CTB which allows for a few days of development time to construct the tutorials.

A major point of attention at the start of the project was the status of the TEI model. Since early 2002, the TEI Consortium has been engaged in a major (backward-incompatible) revision of the TEI specification, migrating it from version P4 (2002, see TEI, 2004) to P5 (see Burnard and Bauman, 2007). Featuring more than just changes in the markup model and the content of the guidelines, P5 entails an overhaul of the complete production process of the standard. Apart from the innovations regarding the content of the TEI markup scheme, adoption of P5 involves coping with peripheral technical innovations. The TEI Pizza Chef software for deriving P4 TEI DTDs has been superseded by the Roma system, allowing users to derive TEI customisations in a number of formal expressions, from the (innate) Relax NG Scheme to DTDs or W3C XML Schemas. By developing P5 as a sourceforge project\(^\text{18}\), early adopters could prepare for adoption of this revision via public access to the latest source code, and more or less stable intermediate snapshot code releases. Of course, the inherent instability of a long (public) transition period mortgages any teaching material covering its changing subject matter. It seems that the timing of the TEI by Example project coincides with a turning point in the transition of TEI P4 to P5: the advantages of P5 adoption for this project seemed to outweigh the disadvantages of P4. When undertaking the preliminary investigations into instigating the project, the most recent snapshot suggested that stability would soon be at hand (Van den Branden, 2006). As a result, the project began developing materials in P5.

\(^\text{13}\) http://www.kantl.be/ctb/
\(^\text{14}\) http://www.kcl.ac.uk/schools/humanities/cch/
\(^\text{15}\) http://www.slais.ucl.ac.uk/
\(^\text{16}\) http://www.kantl.be/ctb/project/2006/tei-ex.htm#t4
\(^\text{17}\) http://www.allc.org/
\(^\text{18}\) http://sourceforge.net/projects/tei/
Indeed, in 2007, the P5 Guidelines were released (Burnard and Bauman, 2007), indicating that this was a prudent choice taken at the outset of the project.

The deliverables of the project are: online “TEI by example” tutorials, a printable PDF version of the “TEI by example” tutorials, an online software toolkit for text encoding, a downloadable CD-ROM image for burning off-line toolkits for use by course participants, and adequate documentation to enable the tutorials to be used elsewhere if needed. Development of the tutorials began in October 2006 and has continued throughout 2007 and into 2008. It is conceded that development time has been slow: however, this is due to the fact that the project is being undertaken with very little funding, and on top of full time academic and research projects by the development team. At present, the technical infrastructure of the project has been agreed and implemented. Work on the individual tutorials has begun, with an aim for a full project launch in summer 2008.

Figure 1: The current TEI by Example home page, providing the user with an overview of the structure and contents of tutorials, exercises and quizzes.

Eight tutorials are under construction. The first, an introduction to text encoding and the TEI, encourages the user to explore textual encoding and markup to foster an understanding of why this is useful, or even necessary, to allow texts to be processed automatically and used and understood by others. The TEI header tutorial covers the type of information and metadata captured in the header element. Three tutorials focus on examples of individual types of text: Prose, Poetry, and Drama, and a further two tutorials deal with examples of Manuscript Transcription and Scholarly Editing. The final tutorial will investigate how the TEI can be customized, and the use of ODD and Roma.

The TEI by Example tutorials aim to provide examples of markup for users of all levels. Examples will be provided of different document types, with varying degrees in the granularity of markup, to provide a useful teaching and reference aid for those involved in the marking up of texts. Likewise, the availability of a software toolkit for teaching text encoding will support the potential trainers to take up the challenge to teach TEI on several occasions.

The first tutorial to be fully developed was the Poetry module. This was chosen as it was a relatively self-contained module, and it could be used to test the various options available for development. There were many editorial, technological, and pedagogical choices the authors had to make. The team had to understand the technical possibilities and limitations afforded by the online environment, and decide how best to integrate these into the tutorial materials. By juxtaposing static (pages, articles) and dynamic (quizzes, validation) functionality, the project aims to provide a holistic learning environment for those new to the TEI. Further linking to other markup examples, provided by the community and the project, extends the remits of the project into another, alternative viewpoint by which to start learning the TEI, aside from the TEI guidelines themselves (TEI, 2007a). Additionally, the role of user testing will be explored to feature feedback and comments from the TEI user community, to aid in the development of intuitive tutorial materials. The completed poetry module has been circulated to the project board, and potential users, for
comment, and user requirements will inform the design and implementation of the remaining modules over the coming months.

Figure 2: The Poetry validation exercise. The user is presented with a poem and given a set of tasks to carry out. The online validator checks whether they have carried out those tasks correctly.

It has always been the TEI by Example project’s aim to integrate real examples from the TEI community within the modules. In December 2006, a call for examples was sent out to the TEI community via the TEI-L email discussion list\textsuperscript{19}. Specific projects that were also known to be using TEI to markup interesting and complex texts were also contacted to ascertain whether they would be able to provide examples of specific encoding approaches and to contextualise encoding theory with real world examples. However, the response from the community so far has been disappointing. It is understood that there is much intellectual effort placed into marking up a text, and that the creators of markup up texts may not want to make their TEI based markup available. TEI is often used as a production standard, and as a result, users can be hesitant in letting others glimpse into the internal workings of a system or project. It is also clear that there is some concern that markup approaches would be criticised, and projects are not keen to “air their dirty laundry in public”: even though showing users real life examples of markup can be more instructive than perfect specimen cases. Additionally, learning good techniques from the observation of bad techniques is a well used pedagogical approach which has some benefit: “An understanding of practical rhetoric as conduct … provides what a teacher cannot: a locus for questioning, for criticism, for distinguishing good practice from bad” (Miller, 1989, p. 23). It would be useful for projects to be able to provide examples where they do not feel the markup was well executed, and comment why this is the case. However, it is understood that individuals and projects do not wish to be open to criticism.

It may also be that projects and individuals do not wish to contribute to the development of a resource which is operating outside the safe bounds of the established, limited TEI community: as yet, the TEI by Example project has no official relationship with the Textual Encoding Initiative itself (until the tutorial development has been undertaken it was felt that it was better to keep the team small and focussed, although this may be revised in future. Certain established members in the TEI community have been critical of the efforts of TEI by Example, perhaps because they do not like the implied criticism that their approaches and their teaching methods are not reaching a wide audience). Additionally, due to the paucity of examples from the community available, the TEI by Example tutorials have been written with examples which were created by the project itself, which has come under some criticism (although this has the added benefit of allowing all learners to start from the same carefully chosen point). Finally, providing TEI by Example with marked up XML files is not enough for examples to be of use: it is important that the examples are accompanied with a brief introductory commentary to the editorial approaches used within each document markup, so contribution to TBE requires investment of time and effort for a project, which is an additional task for already hard

\textsuperscript{19} \url{http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L}
pressed people to add to their to-do lists. The TBE project continues to attempt to outreach to individuals and projects within the community and to encourage them to submit examples for use in the tutorials. Until this happens, a selection of texts are being marked up by the project itself, for use in the tutorials, providing textbook examples of markup approaches. It is acknowledged that this is second best to building up a library of real-life examples of markup.

A poster presented at Digital Humanities 2007 (Van den Branden et al., 2007) encouraged feedback regarding the implementation and design of the tutorials from the Digital Humanities community. A poster presented at Digital Humanities 2008 will encourage user testing from the Digital Humanities community and allow us to integrate feedback into the design process, before launching the tutorials online.

5. Future Work

The response of the TEI community to TEI by Example has been muted: although many examples of TEI markup have been promised, few have been provided to the project. Any projects working on texts that they think may be of interest to a learner, or those who would like their texts to be considered, should get in touch with the TEI by Example project: teibyexample@kantl.be.

In order to support multilingualism in the text encoding community, the online tutorials are being considered for translation into a number of languages from their original English. The translations proper, however, are outside the scope of the initial stages of the project, but the problems presented by internationalisation are important and pressing ones. Issues such as how to provide relevant examples of various text types in different languages must be addressed, as is how to reach as wide an audience as possible through the translation of the tutorial teaching materials into various languages.

Further user testing needs to be undertaken once the next phase of development of the projects begins. Students from both University College London and the University of Antwerp will be used to give feedback on the TEI by Example materials. Additionally, at some stage the tutorials will be open to feedback from the TEI community itself: we brace ourselves for the reaction.

6. Conclusion

TEI by Example is a modest, but important project which aims to produce stand alone tutorials in the of use the Text Encoding Initiative’s guidelines for document markup, which should be of use to the Digital Humanities audience, and beyond. It is hoped that the project results will be relevant to the trainers of TEI, the students of TEI, the text encoding community, and the humanities computing community in general. To do this, it is important to involve the TEI community both in the design and testing of the tutorials, but also in the provision of real world examples of markup materials which can be used as an alternative inroad for interested individuals wishing to learn and understand the aims of the TEI. By making use of the possibilities afforded by the online teaching environment, and creating and tailoring TEI based teaching materials
which can be used by both individuals and in classroom training sessions, the TEI by Example project aims to expand the user community of TEI by providing teaching materials which cater completely towards learners, rather than materials provided for the small community of TEI experts who have little requirement for introductory materials.

References


