Advanced Topics in TEI

Julia Flanders (*Julia_Flanders@brown.edu*)
Brown University

Syd Bauman (*Syd_Bauman@brown.edu*)
Brown University

Laurent Romary (*Laurent.Romary@loria.fr*)
INRIA Laboratoire Loria

David J. Birnbaum (*djbpitt+@pitt.edu*)
University of Pittsburgh

Matthew Zimmerman
(*Matthew.Zimmerman@nyu.edu*)
New York University

In the decade since the 1994 publication of the *TEI Guidelines*, this important text encoding standard has seen widespread use in a variety of research and digitization environments. In some contexts, its application has become routine: digital libraries now publish huge volumes of lightly encoded *TEI* documents through mechanisms which are well understood and thoroughly documented. However, in other quarters intensive research on the *TEI* continues unabated. Not only are the *Guidelines* themselves now being revised (with the publication of *P5* planned for 2005), but applications of the *TEI* to specific research areas continue to emerge, and new tools are continually being developed to support a variety of analytic and publication functions.

This panel session brings together several short presentations on advanced topics in the *TEI* research landscape, which reflect the breadth and depth of work currently being done in this community. The presentations include advanced markup issues, the design of the language in which the *TEI* itself is written and documented, and current *TEI* tools development. The panel chair will open the panel by giving a very brief contextual description of the current development context for the *TEI*: the goals of *P5*, the user community, and current trends in analytical use of *TEI* markup. Following the four short papers by panelists (described below), the chair and panelists will lead a discussion of advanced use of the *TEI* and future research directions. The goal of the panel is twofold: first, to provide an update to the humanities computing community on some important research efforts within the *TEI*; and second, to provide an opportunity for a discussion of the impact and value of this research and its direction for the future.

The first paper will discuss the perennial problem of overlapping markup, and will describe a *TEI* implementation of the *CLIX* solution, which has emerged from the work of the *TEI Special Interest Group (SIG)* on *Overlap*. The *CLIX* approach involves using two empty elements to indicate where each element in a subordinate hierarchy (or at least, each element which overlaps an element in another hierarchy) begins and ends. These empty elements have the same name as would have been used for the equivalent 'normal' element which has content, and use special attributes, `sID=` & `eID=`, to indicate that an empty element indicates the beginning or the end of a pseudo-element (see `<http://www.mulberrytech.com/Extreme/Proceedings/html/2004/DeRose01 EML2004DeRose01.html#t6>`). *RelaxNG*, the schema language underlying *TEI P5*, is perfectly capable of representing some of the constraints that would desired to validate this type of markup. However, *ODD*, the abstract literate encoding language in which *TEI P5* is written, cannot. A mechanism for permitting *TEI P5 RelaxNG* schemas to perform some *CLIX* validation without changing the *ODD* language itself, but rather by using a slightly more complex 'tangle' process to produce schemas from the *ODD* sources, will be presented.

The second paper will discuss analytical approaches to manuscript description and the use of this markup to support advanced research in quantitative codicology. Data-centric manuscript description has recently emerged as a topic of interest in light of the new opportunities provided by electronic text technology. While traditional printed manuscript descriptions have been substantially prose-like (a tendency reflected in more document-centric encoding approaches), the more analytical approach presented here (which will be adopted as part of the new *TEI* chapter on manuscript description) treats manuscript description as structured databases rendered in XML. Highly structured descriptions with rich markup of all descriptive details (using controlled vocabularies wherever possible) permit users to conduct much more advanced research, for instance on the correlation between specific watermarks and specific orthographic norms, or on the resemblance between manuscripts in a given set of features. These kinds of questions go well beyond the tradition of consulting indices or searching for access points, and enable scholars to envision manuscript transmission in ways that would otherwise be impossible. This presentation will illustrate both the provisions of the *TEI* MS description module and its application to these advanced research topics.

The third paper will focus on designing and extending document models with the *TEI*. It will present the main characteristics of the new *TEI* specification platform, which is being used to describe both the documentation and technical characteristics of the next edition of the *TEI guidelines (P5)*. The specification platform (also known as *ODD* for "One Document Does it all") allows one to describe elements and their attributes, through a

combination of prose and formal descriptions. It also allows document model designers to refer to classes of elements, when similarity of behaviour or semantics have to be taken into account. The presentation will illustrate the new *TEI* architecture by presenting the online environment (*Roma*) that allows anyone to design his or her own *TEI* subset and possibly extend the *TEI* capacities by adding or modifying elements and attributes. We will exemplify these mechanisms in the light of the new terminology chapter that is to appear in the *TEI P5* edition.

The final paper in this panel will present the current landscape of *TEI* tools development, and in particular the work of the *TEI Tools Special Interest Group (SIG)*. It will discuss the current challenges faced by developers of *TEI* tools, the genres of tools which are currently of greatest interest, the ways in which the *TEI* community can most effectively assist tool developers (for instance, by contributing to a library of sample documents for testing), and the support framework provided by the *SIG*.