Future Developments for TEI ODD

James Cummings, Sebastian Rahtz, Lou Burnard, Syd Bauman, Betrand Gaiffe, Laurent Romary, Piotr Banski

HAL Id: hal-00770623
https://inria.hal.science/hal-00770623

Submitted on 7 Jan 2013

Future Developments for TEI ODD

Cummings, James, University of Oxford, UK, James.Cummings@oucs.ox.ac.uk

Rahtz, Sebastian, University of Oxford, UK, Sebastian.Rahtz@oucs.ox.ac.uk

Burnard, Lou, TGE-Adonis, France, lou.burnard@tge-adonis.fr

Bauman, Syd, Brown University, USA, Syd_Bauman@Brown.edu

Gaiffe, Bertrand, ATILF, France, bertrand.gaiffe@atilf.fr

Romary, Laurent, INRIA, France, laurent.romary@inria.fr

Bański, Piotr, HUB & IDS, Poland, bansp@o2.pl

The purpose of this panel is to look at the application and future development of the literate programming system known as ODD which was developed for the Text Encoding Initiative (TEI) and underlies every single use of the TEI.

Though strongly influenced by data modelling techniques characteristic of markup languages such as SGML and XML, the conceptual model of the Text Encoding Initiative is defined independently of any particular representation or implementation. The objects in this model, their properties, and their relationships are all defined using a special TEI vocabulary called ODD (for One Document Does-it-all); in this way, the TEI model is used to define itself and a TEI specification using that model is, formally, just like any other kind of resource defined using the TEI. An application selects the parts of the TEI model it wishes to use, and any modifications it wishes to make of them, by writing a TEI specification (a TEI ODD document), which can then be processed by appropriate software to generate instance documents relevant to the given application. Typically, these instance documents will consist of both user documentation, such as project manuals for human use, and system documentation, such as XML schemas or DTDs, sets of Schematron constraints etc. for machine use. In this respect ODD is a sophisticated re-implementation of the 'literate programming' paradigm developed by Don Knuth in the 1970s reimagined as 'literate encoding'.

One of the requirements for TEI Conformance is that the TEI file 'is documented by means of a TEI Conformant ODD file which refers to the TEI Guidelines'. In many cases users employ pre-generated schemas from exemplar customizations, but they are usually better served if they use the TEI ODD markup language, possibly through the Roma web application, to constrain what is available to them thus customizing the TEI model to reflect more precisely their encoding needs.

Some of the mechanisms supporting this extensibility are relatively new in the TEI Guidelines, and users are only now beginning to recognize their potential. We believe that there is considerable potential for take-up of the TEI system beyond its original core constituencies in language engineering, traditional philology, digital libraries, and digital humanities in general. Recent additions to the TEI provide encoding schemes for database-like information about persons and places to complement their existing detailed recommendations for names and linguistic structures; they have always provided recommendations for software-independent means of authoring scientific documentation, but the ODD framework makes it easier for TEI documents

to coexist with other specialist XML vocabularies as well as expanding it to encompass the needs of new specialised kinds of text. It has been successfully used for describing other XML schemas, notably the W3C ITS (Internationalisation Tagset) and ISO TC37 SC4 standards documents; more recently its facilities have greatly simplified the task of extending the TEI model to address the needs of other research communities, such as musicologists and genetic editors.

We believe that the current ODD system could be further enhanced to provide robust support for new tools and services; an important step is to compare and contrast its features with those of other 'meta-encoding' schemes and consider its relationship to ontological description languages such as OWL. The potential role of ODD in the development of the semantic web is an intriguing topic for investigation.

This panel brings together some of the world's most knowledgeable users and architects of the TEI ODD language, including several who have been responsible for its design and evolution over the years. We will debate its strengths, limitations, and future development. Each speaker will focus on one aspect, problem, or possible development relating to TEI ODD before responding to each others suggestions and answering questions from the audience.

Lou Burnard will introduce the history and practical use of ODD in TEI, and describe its relevance as a means of introducing new users to the complexity of the TEI. Sebastian Rahtz will talk about the processing model for ODD, and the changes required to the language to model genuinely symmetric, and chainable, specifications. Bertrand Gaiffe will look at some of the core mechanisms within ODD, and suggest that model classes that gather elements (or other model classes) for their use into content models could be better as underspecified bags instead of sets. Syd Bauman will discuss co-occurrence constraints, pointing out that ODD's lack of support for this feature is a significant limitation, but also that it can often be worked around by adding Schematron to an ODD. Laurent Romary and Piotr Banski will describe issues in drafting ODD documents from scratch, in particular in the context of ISO standardisation work, introducing proposals to make ODD evolve towards a generic specification environment.

We believe that the DH2012 conference offers a useful outreach opportunity for the TEI to engage more closely with the wider DH community, and to promote a greater awareness of the power and potential of the TEI ODD language. We also see this as an invaluable opportunity to obtain feedback about the best ways of developing ODD in the future, thereby contributing to the TEI Technical Council's ongoing responsibility to maintain and enhance the language.

Organization

The 5 speakers will each give a 15-minute introduction to a problem or possible development that relates to TEI ODD. After this the organizer will moderate discussion between members of the panel on a number of questions before opening the discussion to questions from the audience.

Speakers

- Lou Burnard, lou.burnard@tge-adonis.fr, TGE-Adonis

- Syd Bauman, syd_bauman@brown.edu, Brown University Center for Digital Scholarship

- Bertrand Gaiffe, bertrand.gaiffe@atilf.fr, ATILF

- Sebastian Rahtz, sebastian.rahtz@oucs.ox.ac.uk, University of Oxford

- Laurent Romary & Piotr Bański (Piotr speaking), laurent.romary@inria.fr, bansp@o2.pl, Inria, HUB & IDS

Organizer James Cummings, james.cummings@oucs.ox.ac.uk, University of Oxford

references

Burnard, L., and R. Rahtz (2004). RelaxNG with Son of ODD. *Extreme Markup Languages 2004*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.7139&rep=rep1&type=pdf

Romary, L. (2009). ODD as a generic specification platform, *Text encoding in the era of mass digitization – Conference and Members' Meeting of the TEI Consortium* http://hal.inria.fr/inria-00433433

TEI By Example (2010). Customising TEI, ODD, Roma. *TEI By Example*, http://tbe.kantl.be/TBE/modules/TBED08v00.htm

TEI Consortium, eds. (2007). TEI P5 Guidelines Chapter on "Documentation Elements". *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html

TEI Consortium, eds. (2007). TEI P5 Guidelines Section on "Implementation of an ODD System". *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#IM

TEI Consortium, eds. (2007). Getting Started with TEI P5 ODDs. *TEI-C Website*. http://www.tei-c.org/Guidelines/Customization/odds.xml