

DHBenelux 2019 workshop

Advancing digital editing: how to work with graph models, multiple perspectives, and flexible workflows

Instructors:

Elli Bleeker, Bram Buitendijk, Ronald Haentjens Dekker, Astrid Kulsdom
Research and Development group
Humanities Cluster, Royal Academy for the Arts and Sciences

Introduction

Participants of this full-day workshop learn to work with TAG, an alternative (graph-based) data model for text, which allows them to explore a more flexible way of digital editing than is available with the single hierarchy afforded by XML. The goal of the workshop is not (only) to learn to work with a new text editing tool, but primarily to acquire a way of thinking about text modeling that is not constrained a priori by external models or by software assumptions built into software tools. During the workshop, participants will be introduced to the principles of a distributed architecture for digital scholarly editions as they learn to work with the TAG reference implementation *Alexandria*. This implementation allows editors to import, edit, and export complex texts in a straightforward manner. *Alexandria* accommodates the encoding of multiple, coexisting views on text and thus supports various approaches to scholarly editing.

Background

Over the past decades, the field of digital text editing has witnessed some major developments, both practical (the development of advanced tools and technologies for text editing) and conceptual (an increased awareness of the modeling process). There are, however, at least three important areas of digital text editing where we can identify room for further development:

- 1) *Models of text*. In general, we can assume that the practice of text modeling improves significantly when the data model is consistent with the conceptual model because the data model then supports the scholar's process instead of restricting it. However, prevailing data models for text not always correspond to a scholar's idea or understanding of text. An oft-cited example is the discrepancy between the understanding of text as an *Ordered Hierarchy of Content Objects*—the *OHCO* model that underlies XML—and the many textual phenomena that do not fit naturally within this model, such as (self-)overlapping, discontinuous and non-linear texts (Haentjens Dekker and Birnbaum 2017; Haentjens Dekker *et al.* 2018).
- 2) *Perspectives on text*. Another contrast can be seen between, on the one hand, the digital scholarly edition as a closed-off environment that offers one exclusive view of the text

and, on the other hand, the edition as a *digital workstation* that provides for multiple coexisting perspectives on a text. Because a perspective implies a certain hierarchical structuring of the textual data, expressing multiple coexisting perspectives in XML often produces overlapping hierarchies and requires workarounds.

- 3) *Workflow*. Over the years, many integrated editing environments with specifically designed graphical user interfaces (GUI) have been developed, yet these editing environments are rarely adopted by other editors outside the intended users. In fact, the creation of a unified model of the editorial workflow has proven to be unrealistic (Van Zundert 2018). One of the main reasons for this is that interfaces as well as workflows are highly project-specific and personal. A comprehensive GUI will inevitably constrain some editorial choices even as it facilitates others. Furthermore, it significantly hinders the sharing and reuse of scholarly tools and applications, and has resulted in multiple reinventions of the same wheel.

TAG is a new and flexible technology that addresses these three areas. The TAG data model, a hypergraph structure that can successfully address the various challenges posed by the modeling of complex texts, has been described as “both simple and brilliant” (Sperberg-McQueen 2017). Furthermore, the TAG data model doesn’t compromise sustainability or interoperability of textual data: texts encoded in TAGML can be converted to XML, although the down-conversion to a hierarchical format entails some decision-making on the side of the user. *Alexandria*, the reference implementation of TAG, offers its users an editorial workflow that can be freely configured and used within an editor of their preference (like oXygen Editor or Sublime Text). Within *Alexandria*, users can add multiple layers of markup to a transcription; the layers may overlap. A text transcription is not required to follow one hierarchical structure or adhere to a single perspective on text. In short: the TAG technology enables a versatile approach to editing, instead of forming a mold in which every project has to fit.

Workshop program

The workshop would run for a single day, divided into morning and afternoon sessions of approximately 180 minutes each. The morning session concentrates on the theoretical background of data models for text, including a discussion of several textual phenomena and other features that are difficult to express in the tree structure of XML. We then introduce the graph model of TAG and the related markup language TAGML. The session concludes with hands-on work: participants will make a brief transcription of a text fragment in TAGML, with special attention to those textual features that pose challenges to XML as a model. For pedagogical reasons we provide a data set during the workshop, but participants will perform their own document analysis and will markup the text from their own perspective(s).

In the afternoon session we dive more deeply into practice in the context of the entire editorial workflow. Participants work with *Alexandria* to process the marked-up files created during the morning session. Because *Alexandria* is operated via shell commands, we will allocate some time to ensure all participants will be able to perform the necessary operations on the command line. It is however not imperative to have prior command line experience as *Alexandria* is designed to be intuitive and straightforward to use. The last part of the afternoon is devoted to publishing: we convert the markup files to XML (which facilitates their accessibility outside our toolkit) and then publish them as HTML.

To sum up: the TAG workshop will cover the entire editing process, from transcription to publication, paying attention to each step along the way. This “pipeline” approach provides participants with a deeper awareness of the many conceptual and practical transformations that textual data undergoes. It enables both the production of an actual edition and a final high-level, abstract reflection on the importance of choosing an appropriate data model to express, process and analyze textual information. Over the course of the workshop, participants work with *Alexandria* and thus experience first-hand TAG’s hypergraph model for text, the properties of which have significant benefits for the editorial process and for the way we conceptualize our texts. They will learn that expressing information about text, creating a digital edition, and shaping editorial practices doesn't have to be bound to a specific tool or technology.

Target audience

The target audience of this workshop is textual scholars, scholarly editors, digital humanists and scholars interested in text modeling. Experience with TEI (or other XML), text transcription, and textual editing is a plus. Experience with working on the command line, HTML and XSLT is also an advantage, but not required. The tutorial has room for a maximum of 20 participants.

References

- *Alexandria* version 2.2. [computer software]. 2019. Retrieved from: [<https://huygensing.github.io/alexandria/>](https://huygensing.github.io/alexandria/).
- TAG (Text as Graph) [data model]. 2019. Retrieved from: [<https://huygensing.github.io/TAG/>](https://huygensing.github.io/TAG/).
- Haentjens Dekker, Ronald, and David J. Birnbaum. 2017. “It’s more than just overlap: Text As Graph.” Presented at Balisage: The Markup Conference 2017, Washington, DC, August 1 - 4, 2017. In *Proceedings of Balisage: The Markup Conference 2017*. Balisage Series on Markup Technologies, vol. 19, DOI: [<https://doi.org/10.4242/BalisageVol19.Dekker01/>](https://doi.org/10.4242/BalisageVol19.Dekker01/).
- Haentjens Dekker, Ronald, Elli Bleeker, Bram Buitendijk, Astrid Kulsdom and David J. Birnbaum. 2018. “TAGML: A markup language of many dimensions.” Presented at Balisage: The Markup Conference 2018, Washington, DC, July 31 - August 3, 2018. In

Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies, vol. 21. DOI

<https://doi.org/10.4242/BalisageVol21.HaentjensDekker01>>.

- McCarty, Willard. 2004. “Modeling: a study in words and meanings”. Chapter 19 of *A companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004. <http://www.digitalhumanities.org/companion/>>.
- McCarty, Willard. 2005. *Humanities computing*. Basingstoke and New York: Palgrave Macmillan.
- Sperberg-McQueen, C. M. 2017. “Text. You keep using that word” Presented at Balisage: The Markup Conference 2017, Washington, DC, August 1–4, 2017. In *Proceedings of Balisage: The Markup Conference 2017. Balisage Series on Markup Technologies*, vol. 19. <https://doi.org/10.4242/BalisageVol19.Sperberg-McQueen02>>.
- Van Zundert, Joris. 2018. “On not writing a review about Mirador: Mirador, IIF, and the epistemological gains of distributed digital scholarly resources.” In *Digital medievalist*, vol 11, no. 1. DOI: <http://doi.org/10.16995/dm.78>>.