**CUMMINGS — Hack**

<0 images>

**TEI HackAThon: Building Tools for TEI Collections**

Cummings, J. C.

The Guidelines of the Text Encoding Initiative Consortium (TEI) have been used throughout numerous disciplines of the field of digital humanities to mark up digital texts for many years. Doing so has produced huge numbers of TEI collections underlying leading digital editions, projects, and other resources. These digital texts are most often transformed for display as websites, camera-ready copy, or for import into other systems for processing, analysis, or visualization. While the TEI Consortium provides XSLT stylesheets for transformation to and from many formats, and both commercial and open-source software is TEI-aware, there is little standardisation across multiple projects in implementation of querying, searching, analysing, and visualising TEI-encoded texts. The proposals for documentation of intended processing models that forms part of the TEI Simple project should eventually help standardise the recording of this information. TEI encoding is taught frequently in the digital humanities, but the development of TEI processing systems is often approached by developers unfamiliar with the TEI either with trepidation or ignorance of its potential complications. This HackAThon is open to developers with very little experience with TEI (but with significant programming experience) and TEI also experts (with a little programming experience) as well as those who have experience in both. It is hoped that the developers and TEI experts will be able to share their expertise in a knowledge-exchange hacking event. It is intended to be both fun and fruitful.

**Participants and Organisation**

The HackAThon will be organised as an unconference-style event. Applicants will be asked for basic details of their experience and possible contribution to the HackAThon. Decisions will be made by an international programme committee of TEI experts based on the criteria of getting both a sufficient and variety of expertise (technical and TEI), interest in challenges similar to those proposed, as well as geographical, cultural, disciplinary, and gender balance.

A mailing list for HackAThon participants will be created to discuss in advance the possible challenges that might be undertaken. This will allow some preseeding of ideas and potentially people to start work or familiarise themselves with particular aspects of the TEI or technical solutions.

**Challenges**

Possible challenges that this HackAThon will focus on include:

• Development of a library of standard functions for the basic interrogation, analysis, or extracting of data from collections of TEI documents.

• A server platform that provides access to those functions through a RESTful interface.

• Integration of these with existing popular software (e.g., editors, processing infrastructures, etc.).

These are intended as a set of potential interlinking challenges suitable for a range of skill levels and familiarity with TEI. They also potentially coordinate well with developments in the TEI community such as the TEI Simple project and the TAPAS repository. The potential challenges and possible implementations of them will be openly discussed on the HackAThon’s mailing list. All outputs will be made freely available under open licenses.

**HackAThon Outline**

The HackAThon will be organised with an unconference-style pre-DH2015 mailing list discussion, followed by finalising on the workshop day the precise groups of participants and challenges they want to work on. Most of the rest of the day is spent working in these groups, with a review midway through of what the groups are working on. The day concludes with reporting back, demonstrating the work they have done, and discussing next steps (potentially with TEIC support).

**Morning**

• **09:30–10:30—**Introduction and coffee, finalise groups and challenges.

• **10:30–12:30—**Groups start work (break-out sessions).

**Lunch**

• **12:30–13:30—**Lunch.

**Afternoon**

• **13:30–14:00—**Groups briefly report on work done so far.

• **14:00–16:00—**Group work continues (break-out sessions).

• **16:00–17:00—**Regroup, report back and show work, plan for next steps, complete any DH2015 workshop evaluation form.

**Programme Committee**

The programme committee will decide on submitted applications (trying to err on the side of inclusion) and will include all of the TEI experts mentioned below as well as Hugh Cayless (as TEI Technical Council chair), Marjorie Burghart (as a representative of the TEI board), and Magdalena Turska (as a representative of the TEI Simple project).

**Organisers and TEI Experts**

The organisers of this proposal all have extensive experience in leading and coordinating hands-on workshops focused on TEI or related theories and technologies. They will be available during the workshop together with other TEI and DH experts who will be attending DH 2015 and confirmed their interest to take part in the HackAThon. These TEI experts (and developers in their own right) include:

*James Cummings*(james.cummings@it.ox.ac.uk), senior digital research specialist in Academic IT at University of Oxford’s IT Services, where he helps academics plan research projects with digital aspects, is the department’s liaison for digital humanities activities, and is a director of the annual Digital Humanities at Oxford Summer School. He has built many TEI-based websites, such as William Godwin’s Diary, which uses TEI XML served from an eXist XML database to power a datacentric view using jQuery DataTables of the information that this 48-year diary contains. He will be presenting a paper at DH2015 on the output of the TEI Simple project, which includes new proposals for documenting intended processing models. He is an elected member of the TEI Technical Council.

*Syd Bauman*(s.bauman@neu.edu) is the senior XML programmer analyst at Northeastern University Digital Scholarship Group, primarily focused on the Women Writers Project. He has been involved with structured text representation since before SGML, and TEI since 1990. Syd frequently teaches TEI encoding, TEI schema creation, and XSLT workshops. He is an elected member of the TEI Technical Council.

*Martin Holmes*(mholmes@uvic.ca) is a programmer and consultant for the Humanities Computing and Media Centre and the University of Victoria. He is an elected member of the TEI Technical Council.

*Conal Tuohy*(conal.tuohy@gmail.com) is an independent digital humanities software developer living in Brisbane, Australia. He has previously been an elected member of the TEI Technical Council.

**Target Audience**

DH practitioners and developers interested in experimenting with TEI, DH technical developers, experienced XML users, and TEI users with exciting ideas.

**Expected Number of Participants**

10–20; ~25 max, including four TEI experts.

**Rooms and Equipment**

• One room to host up to 25 people (ideally ability to break up into small groups with individual tables, or access to public space where small groups can work together quietly).

• Wifi.

• Projector.

• Lots of power sockets.

**Programme Committee**

• Hugh Cayless (TEI Technical Council chair).

• Marjorie Burghart (TEI board).

• James Cummings (TEI Technical Council).

• Martin Holmes (TEI Technical Council).

• Syd Bauman (TEI Technical Council).

• Conal Tuohy (independent consultant).

• Magdalena Turska (DiXiT Fellow; TEI Simple developer).