Textual Artifacts and their Digital Representations: Teaching Graduate Students to Build Online Archives

Deena Engel <deena_dot_engel_at_nyu_dot_edu>, Courant Institute of Mathematical Sciences, New York University
Marion Thain <marion_dot_thain_at_nyu_dot_edu>, New York University

Abstract

Co-teaching a digital archives course (ENGL-GA.2971) for graduate students in the English Department allowed us to bring together our expertise in both research and pedagogy from two fields: English Literature and Computer Science. The course built on a core pedagogical principle in Computer Science of teaching through projects rather than from unrelated one-off programming or web development assignments. Teaching the Text Encoding Initiative after students had completed hands-on projects (using xHTML, CSS, and a digital archive working in a standard content management system) enabled the building of technological skill sets in a logical and complementary manner. From a literary perspective, building a digital archive — and teaching text encoding — enabled an in-depth consideration of textual materiality, the processes through which literary scholarship must inform technological building decisions, and the ways in which the act of digitization can be used to ask new questions of the text (or to prompt the text to ask new questions of itself). This paper will survey our techniques and approaches to interdisciplinary teaching, culminating in our usage of text encoding for exploring issues of textuality through digital presentation.

Introduction

This article reflects on the practice of designing and implementing a course that aimed to teach graduate students in English the skills to build a scholarly online digital archive from primary source materials. A founding principle of our pedagogical practice is the integration of CS and DH methods. As a still fairly young pedagogical field, discussion and sharing on this topic might, we hope, be timely and useful to others. This course was part of an evolving response to the key current pedagogical challenge of teaching graduate students in English how to synthesize a new set of technological skills with literary expertise to discover new ways of working and new questions to ask of their texts.

The ultimate ambition of our course was for each student to digitize text otherwise inaccessible beyond the material archive. As such the focus was on dealing with manuscript materials, marginalia and other editorial markings on typescripts, and materials that had a particular material instantiation. The students were free to choose their focal materials to reflect their own interests, and had access to the manuscripts collections and Fales’ special collections in the Bobst Library at New York University. The resulting digital archives included, for example, collections of nineteenth- or twentieth-century letters by well-known writers; manuscript drafts of what went on to become well-known literary texts; type-script drafts of literary essays; and even a collection of contemporary hand-made ‘zines’. Each archive included images of the documents, transcriptions, TEI encodings, contextualization through further reading lists, and prose narratives about their materials and their digital archival practice.

New York University’s Department of English first offered a course on creating online archives from primary source materials in the fall, 2011 semester under the course number and title ENGL-GA.2957-001 Special Topics in Literary
Theory: Literary Archives and Web Development. (The course home page, hosted by the Department of Computer Science at New York University can be found at http://cs.nyu.edu/courses/fall11/ENGL-GA.2957-001/DT_openingPage.html). This course built on earlier experience in the Computer Science Department on teaching topics in the Digital Humanities through an undergraduate Special Topics course CSCI-UA.380 Computing in the Humanities and the Arts. (CSCI-UA.380 Computing in the Humanities and the Arts has been offered five times from the fall, 2007 semester through the fall, 2014. semester; see http://cs.nyu.edu/courses/fall14/CSCI-UA.0380-002/HC_index_fa14.php). The undergraduate Computer Science course is taught within the context of the Computer Science Department Web Programming minor (http://cs.nyu.edu/webapps/content/academic/undergrad/minors) and has two pre-requisites (one semester each in computer programming and in web design) [Engel 2013]. In each iteration of the undergraduate course, the capstone project included building an online digital archive from primary source materials; some of the results of this collaboration between the New York University Library and Archives and the Department of Computer Science are described in published essays. [Mitchell et al 2012]; [Bunde and Engel 2010a]; [Bunde and Engel 2010b].

The course under discussion in this article, ENGL-GA.2971 Practicum in Digital Humanities was given in the fall, 2013 semester in a new format in which it was designed and co-taught collaboratively by two faculty members, one each from the English and Computer Science departments. (The course home page, hosted by the Department of Computer Science at New York University can be found at http://cs.nyu.edu/courses/fall13/ENGL-GA.2971-001/DT_index_fa13.php) As computer systems become more integrated into and integral to studies in many fields, the development of courses that offer training in technology is becoming more and more relevant outside of Computer Science departments. [Guzdial and Forte 2005]. In addition, the role of web development in current Computer Science pedagogy is widely studied and discussed [Park and Wiedenbeck 2011]; [Wang and Zahadat 2009]; [Yue and Ding 2004]. From a pedagogical standpoint, this course offered the Computer Science department an opportunity to devise materials and a curriculum to best meet the needs of our growing student body in the Humanities who wish to augment their studies with computational and technological skills. As it ran in 2011 the course was very successful in bringing humanities computing skills to students in the English department. Building on this success, the introduction of a literature professor (with expertise in editing, and creating online archives of manuscript materials) as a co-teacher was designed to help expand the remit by forging greater connection with current debates within the humanities, and particularly within literary studies. The syllabus for the course can be found in the appendix.

Aims of the Course

From a Computer Science perspective this course sought to implement and contextualize four current goals and trends in Computer Science education. Following is a brief description of each of these goals with references for further reading.

(1) Project-Based Learning (“PBL”):

Project-based courses are common in computing education. As a method within the problem-based learning toolkit, projects can be used at any stage within a degree program to explore alternative and often more complete solutions to a given problem allowing the theory to emerge as necessary. [Richards 2009]

Students who drop out from a course do not have the opportunity to learn all the material and transfer it. If students understand the usefulness of what they are learning, they are less likely to give up. [Guzdial 2010]

There is a current trend in Computer Science pedagogy to encourage “project-based learning” (PBL) in place of “traditional” one-off programming or development assignments. [Chang and Lee 2010]; [Gülbahar and Tinmaz 2006];
[Richards 2009]. We found that all of the technical skills and technology infrastructure that we taught could be related to the main project in this course, which is the online digital archive. This approach appeared to incentivize the students to tackle topics outside of their areas of expertise as the results were immediately apparent (e.g. the ability to manipulate a variety of images of a manuscript within the context of a webpage) and moved them closer to their end goal of the course project.

(2) The Role of Web Development in Computer Science Pedagogy:

Web development can provide a rich context for exploring computer science concepts and practicing computational creativity. [Park and Wiedenbeck 2011]

Web development has become a serious area of study for undergraduate Computer Science majors [Adams 2007]; [Connolly, 2011]; [Liu and Phelps 2011]; [Mendes et al 2012]; [Schaub 2009] and for undergraduate non-majors [Gousie 2006]; [Greenberg et al 2012]; [Guzdial and Forte 2005]; [Kurkovsky 2007]. Web development offers instructors and students both a granular approach to teaching technologies by focusing a lesson or assignment on a single technology (e.g. xHTML/CSS; PHP; using Photoshop to manipulate images, etc.) as well as the opportunity to study and work with the many-faceted technological structure of web implementation (e.g. the underlying relationship between xHTML, CSS, PHP and a database in the context of modifying and manipulating a content management system (CMS) site).

(3) Inclusion and Computer Science Pedagogy:

Supporting a workforce that can create, not simply consume, computing technology requires a shift in pedagogy toward problem solving in a gender neutral, culturally and ethnically diverse community. [Pulimood and Wolz 2008]

There is an acute awareness within the Computer Science pedagogy field that education in Computer Science and technology must consciously and deliberately address issues of inclusion so that women and minority students as well as disabled students should have full access and encouragement to participate on a “level playing field”. As female students are well represented in the discipline of English, the issues of gender and computer technology were particularly relevant for the course. The recent research into inclusivity within Computer Science education [Guzdial et al 2012]; [Pulimood and Wolz 2008]; [Stephenson et al 2007] helps to inform practices and approaches to teaching technology skills to Humanities students. Open discussion of discomfort with technology appeared to ameliorate some of the concern. Furthermore, imposing a rule in classroom discussion that “there is no such thing as a dumb question” and that “web technology is constantly evolving and the trick is to learn how to frame the question” helped to assuage the students’ fears.

(4) The Development of Computer Science Pedagogy Within Inter-disciplinary and Cross-disciplinary Areas of Education:

Computer science holds a unique position to craft multidisciplinary curricula for the new generation of faculty and students across the academy who increasingly rely on computing for their scholarship. [LeBlanc et al 2010]

Computer Science departments must continue to meet the needs of other departments and collaborate both on University pedagogy goals and with research [Bills and Canosa 2007]; [LeBlanc et al 2010]. Teaching collaboration between Computer Science and literature faculty aids the development of more radically interdisciplinary courses, which can offer something different from literature courses that use technologically-enhanced learning or web development courses that utilize literary texts.
Our course was structured so as to begin with connected but distinct issues in literature and technology that ran in tandem, but to work towards the integration of the two sets of considerations. This integration culminated in the final section of the course on the Text Encoding Initiative (http://www.tei-c.org/index.xml). From the perspective of the literature professor, then, the overall pedagogical aim was to provide a layer of commentary on the course activities that encouraged students to reflect on all the technical decisions they made in building their sites as themselves potentially acts of interpretation of the text they were representing. Through looking at recent debates around textual interpretation in literary studies, it was possible to make connections that enabled students to consider how their own practice in digital reproduction might be intersecting fundamentally and significantly with issues of textuality. This goal was one of teaching the students how to think in ways that confront the deepest issues of interdisciplinarity, and to avoid the natural compartmentalization of skill sets. We maintained the specificity of the issues within each disciplinary field while showing the points of intersection by role-modeling debates between the two teachers. With the main outcome of the course being the online digital archive, we decided not to ask the students to write a separate reflective essay-commentary on their building of those sites, but to ensure that this reflection was built into the site through their layout, the choices they made, and through the prose in their “About” sections.

Also from a literary perspective, it was crucial to the course that there was the potential not just for a growth in technological skills informed by considerations drawn from the humanities, but a growth in awareness of textual features through attending to the literature in a new way. In other words, the act of digitization became within the course a new way of being sensitive to the features of the text for literary, critical, and interpretative purposes. The potential for this process to work in two disciplinary directions simultaneously was particularly interesting pedagogically because the students came to the course with varied levels of expertise in technology and in literary study. The humanities skills the students developed through the course were those of the literary critic (particularly in relation to textual materiality), the literary-historical scholar, and the editor. Depending on each student’s previous experience and their comfort with literature and technology, respectively, the opportunities were different, but for every student the objective was a symbiotic relationship between the two, with the digitization process prompting new engagement with the texts they had chosen, and for the resulting considerations to then go on to inflect the way they used the technology within their site.

While this article focuses around the collaboration of the two professors who designed and taught the course, it should be noted from the outset that the course benefitted enormously from the involvement of a broader team. The librarians at the Bobst Library, and particularly those working within Fales (Charlotte Priddle, Lisa Darms, and Amanda Watson), provided a session for our students teaching them how to use special collections and how to handle its materials. Melitte Buchman from the Bobst Digital Studio (http://dlib.nyu.edu/dlts/) worked with our students both as a group and individually to help them photograph their materials and understand the principles and practices of imaging. The Systems Group within the Courant Institute of Mathematical Sciences (CIMS) provided the technological resources required to host and support the students’ projects. Working within this broader term not only helped better support our students but also offered something in return, as our students provided the library with high resolution tiff images of the materials they had scanned.

The Course Syllabus and Structure: Technology

This course met twice each week: once in a traditional classroom lecture and discussion format; and a second time in a multi-media lab for hands-on work. Both instructors attended all of the lecture classes; in addition, the Computer Science faculty member provided lab sessions, and the English faculty member offered consultation hours immediately following each lab session (in practice, there was much collaboration between the two professors in each of these additional strands).

Every student was given an account on a production webserver (http://cims.nyu.edu/webapps/content/systems/resources/i6). The Courant Institute provides two servers dedicated to teaching purposes that are used throughout the Computer Science department: a web-server for students to host their
sites and a database server running MySQL so that students learn to create and manage their own databases. Students also worked in the Digital Studio (http://nyu.libguides.com/digitalstudio) and ITS Multi-media lab which provided scanning equipment and appropriate software for image manipulation and web development.

The course was divided into four units. A list of the technical skills which were taught in each unit is followed by an overview of the pedagogical practices. The first three units each entailed a project-based assignment which is described below as well.

**Unit I - Basic web development skills**

In this unit of the course, we covered skills and topics typically taught in a standard introductory course to web design. These topics include:

- Mark-up languages: Students were introduced to mark-up languages using HTML5 and CSS3 for rendering basic webpages.
- Working with images: Students learned to scan objects to specification; with discussion on image resolution and image file types (.tiff, .psd, .jpg, and .gif in particular) in order to understand how to best implement the process from scanning a manuscript through rendering that image for web viewing at appropriate sizes and resolution. Melitte Buchman, the Digital Content Manager at NYU’s Division of Libraries, provided a guest lecture for the class on meta-data standards for image files as well as on accepted scholarly standards for images used for scholarly purposes.
- File management skills for working on a Unix web server: Students learned to upload files for publication; manage file folders; download files; assess file sizes and permissions; set up .htaccess files to restrict access to portions of their sites when appropriate; and other file management skills.

The project assignment for this unit consisted of a website of a minimum of two pages rendered in HTML5 and CSS3 to describe an author of the student’s choice and publish two or three selections of his or her work.

**Unit II - Content Management System (CMS)**

In this unit of the course, each student installed his/her own WordPress (http://wordpress.org) site to the teaching webserver. The Computer Science faculty member, in consultation with the English Department, decided to use WordPress in this case as Omeka was not supported on the server we used and we believed Drupal is too complex for an introductory class taught within one semester for this student population. As each student installed his or her own WordPress site, they had the opportunity to learn about WordPress as an example of a CMS and will be able to carry this knowledge to future CMS projects.

Specific topics in this unit included:

- Themes (http://wordpress.org/themes/) and child-themes (http://codex.wordpress.org/Child_Themes) for customization: Each student was expected to select and install a WordPress theme of his or her choice based on the content needs of the site. In addition, each student was required to create a child-theme for further customization; one result is that none of the original themes were recognizable in the final sites.
- The students explored additional topics in xHTML and CSS to facilitate their site development including the debugging process such as with Firefox Firebug (https://getfirebug.com/); implementing external style sheets such as Google Fonts (https://www.google.com/fonts) and understanding and implementing responsive design (the way that a site might change configuration depending on whether the user is viewing the site on a desktop, tablet or mobile phone for example).
- We taught a brief introduction to PHP where all students were required to modify the source code in at least one PHP script (e.g. all students were required to modify the default footnote to credit the Fales Archives...
The assignment for this unit of the course consisted of their first draft of an online digital archive based on primary source materials; content included text and images.

Unit III: – XML and the Text Encoding Initiative

Based on the principle in Computer Science pedagogy of building skills on a strong foundation, we introduced XML in the final development unit of the course. In this unit of the course, we sought to build on the students’ understanding and experience with HTML5 as a mark-up language in order to contextualize and introduce XML [MacKellar 2012]; [Paterson et al 2005] in general and the TEI application of XML in particular.

Students were introduced to the syntax and structure of XML as it is used in its narrative form for descriptive meta-data; students in this course did not use XML to generate stand-alone datasets. There are significant differences between XML and HTML; the significantly greater flexibility of XML was presented to the students as a means to better annotate and analyze their literary texts. Students already understood that HTML and CSS separate the content of a text from the infinite variety of possible output formats. In this course, we discussed XML implementation to further describe and define the underlying literary structure of a text (e.g. a play consists of acts; and each act may consist of one or more scenes while a novel containing chapters would be encoded differently in order to capture the underlying structure of prose). We further introduced the flexibility and variability available in XML languages to capture meta-data; specifically using XML tags standardized in the literary scholarly community through the international efforts of the TEI Consortium and related projects. The two professors worked closely together to pick a series of examples for the first lectures on TEI/XML in order for the students to make the transition and see the value of XML over plain text or HTML.

For web presentation of the TEI documents to accompany their websites, students were offered the choice to use CSS or TEI-Boilerplate (http://dcl.slis.indiana.edu/teibp/index.html). XSLT was not taught formally in this class due to the concern that this is a more advanced topic which could intimidate students at this level, especially in light of the complex literary texts that they selected. However, the Computer Science instructor introduced XSLT briefly at this time in the course so that the students would understand the importance of it in web publishing and its wide usage in the TEI community. Several students expressed an interest in XSLT workshops such as those conducted by the Women Writers Project (http://www.wwp.northeastern.edu/outreach/seminars/).

Unit IV: Site Preservation

Preservation is a very important topic with respect to digital archives from both a literary and a technical perspective. In the final class, we discussed technological aspects of the students’ sites which would require modification for permanent online publication; and how to render a finished CMS site into a static HTML/CSS site for permanent publication (this discussion centered on a plug-in currently available at http://wordpress.org/plugins/static-html-output-plugin/. Note that this plug-in required modification by the instructor in order for the students to use it.). The students understood that such a transition would eliminate any future issues of upgrades to PHP running on the webserver in general and to future versions of WordPress in particular; as well as the value of rendering the site easier for the Systems administrators to...
support as MySQL would no longer be needed to keep their sites running after such a transition.

The Course Syllabus and Structure: Literature

Interwoven with this technology syllabus was a set of considerations led by the literature professor. This process began at the start of the semester through a discussion about why we might want to digitize texts and what might be gained; what our priorities should be in deciding which texts to digitize; which texts might be most at risk of being lost to us or inaccessible if not digitized; and what might be the consequences of digitizing text. This class discussion generated a humanities agenda (of aims and concerns) to which we returned throughout the course in relation to our practice. The following two weeks were devoted to analyzing existing online literary archives: first a few chosen for collective group attention,[1] and then those chosen by the students which they analyzed in ten minute presentations to the class. Students were given a list of criteria for the scholarly evaluation of such sites, but also asked to think more generally about what the online presentation of the material adds (what it might enable the user to do with the text that they couldn’t otherwise do) and how it might change our reading of the material presented. It is through this discussion that one of the key practices of the course, from a humanities perspective, was introduced: analyzing how site-building and presentational decisions are also in key part interpretative decisions that determine how the material is read. The lessons learned from this analysis fed into their own project development in various ways, particularly in relation to their design of site architecture and navigations. For one student working with a nineteenth-century manuscript of a novella, for example, a key opportunity arose from the author’s practice of dating his work on the manuscript. The student designed menu options that allowed the user to access the manuscript according either to date of composition or chapter structure. This enabled the correlation of the work the author was doing on the manuscript on a particular day with the concerns he expressed in letters written on or around the same day.

For the literary component of this first section of the course, essays such as Kenneth M. Price’s piece on “Electronic Scholarly Editions” were used to introduce many of the key concerns [Price 2013]. However, students were also asked to engage with thoughts about the practice of textual editing more generally and to consider the kinds of benefits the online environment might offer over the traditional paper-based environment. The issues discussed here ranged from the ability to rectify mistakes relatively easily in digital editions as opposed to runs of printed texts, to the kind of high-level interactivity and hypertextuality that online archives might offer.

Teaching the students about metadata was a crucial part of the first half of this course and the topic was introduced, from the literary-studies perspective, through the scholarly debate around textual materiality. The technical information about metadata standards was provided through the online pamphlet — “Understanding Metadata” — published by the National Information Standards Organization (http://www.niso.org/publications/press/UnderstandingMetadata.pdf), but that was really secondary to getting students to first consider why it might be important to record and represent certain types of information about the texts they were digitizing, and what they would want to record in order to capture various different aspects of the textual object. To bring this issue alive from the perspective of literary scholarship, the students were given some relevant literary scholarship in advance of the class. For example, Nicholas Frankel has written about the significance of reading text within the context of the book as a material object, using as his example Oscar Wilde’s Poems — a beautiful and elaborately-designed book shown to them in the University’s special collections the following week [Frankel 2000]. This scholarly work gave an agenda to thinking about textual materiality and provided a powerful example of the kinds of interests humanities researchers might have in a book’s extra-textual features, and how those features might be considered necessary for the interpretation of a text. For the class itself, the literature professor brought in some examples of books that raised interesting issues of materiality, and discussed with the students which aspects they would want to represent in a digital rendition and how they would go about giving that information in the form of metadata. These examples included not only texts that are elaborate or unusual in their material instantiation but, for example, a fairly plain mid-twentieth-century book that had an old photo tucked inside the front cover; it was not clear the photo had any relevance to the book other than through the life of a previous owner, and the students were
asked to discuss what they would do with it in their digitization process. The students were also asked to bring in any textual-object of their choice and talk briefly — and in light of the kinds of issues they had seen explored in the set reading — about what and how they would want to represent it in addition to its text.

The benefit of approaching the issue of metadata through literary scholarship around textual materiality was that it enabled a much deeper engagement with the issues of creating a digital representation. These would not have emerged had we simply instructed the students of the need to record, for example, the particularities of the place, date, and publication of the particular edition they were digitizing, or the physical location and call number of the manuscript they might be working with. The discussion around the importance of issues of materiality to the interpretation of text brought to light all kinds of other aspects of the physical text that might be important in representing it in a digital format: both through metadata and through the photographs and scans they were to create. For example, we discussed the importance of giving the online reader images of the cover and front-pages and end-pages of a text, and considered the ways in which sites that had not done so might limit the questions scholars could ask of the texts they represent. In light of recent scholarship on the history of paper, we also considered the potential importance of giving information such as the type and weight of paper used as part of the metadata, and the significance of such information to certain kinds of scholarly questions (see [Price 2012]).

Issues of materiality central to the course were given an interesting twist through one student’s focus on contemporary “zines”. Often created by hand rather than on a computer, and exploiting a rawness that appeared to oppose the aesthetics of new technologies, the students were encouraged to think about how these pieces posed interesting theoretical, and sometimes practical, questions for the process of digitization. One zine in particular, which was folded in such a way as to give the potential for many different views of its content depending on how it was unfolded, exploited the three-dimensional possibilities of paper in space. Although with enough separate images from different angles (or within a video or animation), these possibilities could be captured and uploaded to the digital archive, the zine was clearly designed to challenge the two-dimensional textuality of both the traditional codex technology and digital textuality. The introduction of literary-theoretical frames for thinking about the insistent materiality of these items led the student to write to the authors with questions both about their intentions and to ask permission to digitize the zines. The replies were often revealing of a reaction against digital culture and can be summarized with reference to the “8 ball” collective website, where they insist on “MORE ZINING, LESS BLOGGING” (http://8ballzinefair.com/ABOUT). By encouraging the group to think about what it means to digitize items that were created (in part, at least) in opposition to digital culture, we were able to provide a particularly interesting and challenging perspective on the processes the students were undertaking.

This discussion about the material instantiation of text (in relation to very varied examples from the nineteenth to the twenty-first century) led by the literature professor was part of the preparation for students to go into the university library’s special collections department to choose the materials they wanted to digitize. Attending a session organized by the library staff, the students learned about the collection, and how to handle its materials. The library staff also covered copyright issues that continued to play an important role in class discussions throughout the course as permissions were clarified, sought and granted. The following weeks of the course focused, from the literature perspective, on helping the students to choose their texts; and culminated in using one of the classes to present their materials and discuss them with the other students in the class. At the presentation the students were asked to give a rationale for their choices, including thoughts on the textual, critical/historical and technical significance of their selection. This discussion, in a roundtable format, was particularly useful because the students had chosen very different kinds of materials to digitize: sharing their work with each other enabled them to learn from the very different challenges they were each facing, and to broaden their understanding beyond the issues their own materials posed. Over the next couple of weeks, when the class time was devoted primarily to the technical issues of building their sites, it was possible to ensure the “literary” considerations stayed with them by asking them all the time to think about and develop that connection between the building decisions they were making in their archives and the issues of textuality.
and materiality introduced earlier. This was something the literature professor took up with students particularly during the lab sessions when there was the opportunity to address these issues individually and to encourage them to think about the particular problems and possibilities their own materials offered.

We brought all of this back into class again in week eight, when almost the whole session was devoted to another literature-led roundtable with each student presenting their site under construction and explaining in ten minutes why they had chosen the proposed site architecture and to give a rationale, from a literary and scholarly perspective, for the technical choices they had made. They were asked to consider not just how their materials could best be presented, but also who their intended audience was and how a desire to be useful and relevant to a broad audience might inflect the design of their sites. At this point the students were asked to consider what kind of interest their texts might hold within current scholarly fields of enquiry and how to connect with and aid those users, but also how they might make their sites sustainable in a literary, rather than a technological sense, by keeping their materials open as far as possible to the kinds of questions future users might want to ask of them that we cannot now necessarily predict. As an example, introducing archives that had not digitized the advertisements at the back of the nineteenth-century periodicals they represented, we discussed how such materials were now considered of great interest to scholars both in their own right and as context for the articles that had been digitized. All the students reported making changes to their sites as a result of the reflective process they undertook in presenting their rationale to others. Listening to each others’ presentations and discussing them afterwards also helped significantly in sharing good practice and introducing new possibilities for each project. At the end of each presentation the student was asked what they might enable through their digitization of the text that could not be done before. Here they were asked to go beyond the more obvious answers that related to preservation and access to think about how their digitization strategy enabled the texts to ask new questions of themselves; the ensuing discussion was particularly formative for shaping the next phase of site building the students undertook in the course.

The second half of this session was devoted to a round-up of the challenges students were facing with their projects from a literary, rather than technological, perspective. The agenda here was determined in key part by the problems, and solutions, that had arisen with students individually in the lab sessions (which often highlighted more general issues they should all consider). The agenda included a discussion of issues around the principles of manuscript transcription. Here the students were invited to consider the well-worked out transcription statement represented in a site such as the Bentham letters (given there in full to aid crowd-sourced transcription) and how it might help them develop principles for their own transcription (http://blogs.ucl.ac.uk/transcribe-bentham/). Editorial experience was also used to encourage the students to think about the significance of recording uncertainty, introducing the need to balance a natural desire for definitive and interpretative transcriptions with a strategy that left open possibilities for other meanings that might become relevant in the future. Also raised here was the general, and related, issue of fidelity versus clarity/usability, and clutter versus functionality, to help the students think about the challenges they were all facing in deciding how much detail to present and how many different functions to build into their sites. The possibility of offering different views of the material, so a reader could choose either a simple or more complex view and functionality, was a particularly useful point around which the professors could role-model the interplay of technical and scholarly considerations. Another major topic at this point was the use of visualization and mapping to enable the students to present data in new ways. With one student using a map to locate manuscripts geographically, we were able to introduce the idea of using visualizing more generally in other, non-geographical, ways to help get more out of their data. For example, another student had been wondering which of the many and varied published editions of her novella to provide alongside the manuscript version she was digitizing; one solution discussed with her was to use the software tool Juxta (http://www.juxtasoftware.org/) to show the relationship between the manuscript and various different published editions, showing visually the areas of greatest similarity and difference between them.

This process of reflecting on the literary issues at stake in their building procedures ended with the students being asked to imagine what they might enable their users to do with their materials in an ideal world — and what the literary
or scholarly value of those digital experiments might be. This was the basis for a discussion of how the course instructors might be able to help the students implement some of those ideas in a realistic form within the time and resources available. For example, a student who was interested in large-scale crowd-sourced transcription created a game-inspired transcription quiz in the form of a WordPress blog. She used this to address some of the challenges of legibility posed by the author’s hand-written marginal comments in the type-written manuscript she was digitizing. Another student who was digitizing an archive of mid-twentieth-century letters written from New York City plotted the addresses from which they were sent on a historically accurate map of the city. She used KML (a mark-up language based on XML for use with Google Maps) to identify the geographical positions, and then superimposed a map from the archives of the New York Public Library over the Google map to make the mapping historically relevant. Adding a geographical component to her work enabled her to analyze the importance of the location of writing to its content — a process that revealed interesting connections.

Text Encoding: Further Integrating Technological and Literary Considerations

The course culminated in teaching encoding to the students so that they could add a further layer of information to their sites, experimenting with what the TEI might be able to offer their projects. First we used the “The Red Wheelbarrow” by William Carlos Williams in order to introduce mark-up of basic metadata and poetic structure from a technical perspective. Then we turned to a passage from Hamlet and a chapter from Huckleberry Finn to teach the TEI conventions for encoding drama and prose. We taught this collaboratively, going through the literary passages with the students identifying which features one might want to encode, and showing how those features would be rendered in the XML. The Shakespeare example enabled us initially to introduce the idea of the TEI header and the kind of metadata that could be recorded there: both in relation to the creation of the electronic edition, and in relation to the specificities of the edition drawn on for digitization. It also enabled us to show some of the key tags and attributes for working with dramatic text. Chapter twenty-one from the Mark Twain novel was chosen to enable a more complex engagement with TEI through its playfulness with genre: here the characters engage in a muddled and half-remembered enactment of a Shakespeare play. The play-acting within the text enabled us to show how one might mix prose and drama tags to render the text in interesting and useful ways. In addition, the text invited a series of notes in the encoding disentangling the lines from different Shakespeare plays that were muddled together in a single soliloquy. The chapter also has instances of southern dialect that could be translated in the text encoding (thus rendering it searchable in standard English), and other textual features (such as a playbill centered in the middle of a page) that raised interesting questions about encoding practice.

Following on from this class teaching the language of the TEI, the literature faculty member led another session focusing on the literary and critical significance of encoding. The chapter The Text Encoding Initiative and the Study of Literature by James Cummings provided a good starting point for the students to think about the history of the initiative and its more theoretical and critical implications [Cummings 2013]. The students had read this chapter in advance of the class, and had also prepared their own notes towards the encoding of the first textual example we looked at in this session: Shakespeare’s Sonnet no. 116. Working through the sonnet with them gave a chance to think about a different genre through TEI. Starting with a literary critical reading of the sonnet, the students were asked to think about the rhyme scheme and meter. Noting the possibility of rendering the overall meter and rhyme scheme in the header, we also discussed, for example, the implications of the fact that the rhyme scheme does not in fact appear to comply to this scheme in modern English, and how we might reflect on that in the TEI. Similarly the opening lines of the poem (which invoke the liturgy of the marriage ceremony) are not in the standard iambic meter, and this enabled us to discuss what kind of metrical mark-up and interpretation might be noted through the TEI. The goal here was to begin to think about the variety of purposes for which encoding might be used, and to start to show the variety of ways one can note a particular feature of the text through the TEI depending on what one wants to achieve.
Having worked through the sonnet in detail, we then gave them the poem “Love Sonnet” by John Updike and asked them how they would encode it [Updike 1969, 66]. The poem has one complete first line (“In Love’s rubber armor I come to you”) but the subsequent thirteen “lines” of the sonnet are indicated purely by a letter marking the position on the page where each line would end. The letter given to mark each line ending crucially also represents the strict rhyme scheme of the sonnet. From a technological perspective, the aim of this exercise was to get the student to produce an XML encoding, but from a literary perspective it was also to ask students to think about the act of encoding and to recognize the poem as itself in some ways a challenge to some of the structuralist assumptions upon which TEI is founded. The poem is a post-modern parody of the conventional nature of love poetry, gesturing towards the generic, iterable and predictable nature of the romantic sonnet. As such, the poem is a gesture towards a sonnet rather than a sonnet itself, and in this way it raises questions about the kinds of classifications on which encoding is based. The poem has fourteen lines, and the rhyme scheme one would expect from a Shakespearian sonnet, but does that make it a sonnet, and should it be identified in that way in the TEI? What happens if we try to encode this poem in the way we tagged the Shakespeare sonnet? What questions does this ask of the poem, and what might it miss or misrepresent? To encode this poem as a sonnet is to explain its own joke in a rather heavy-handed way, but the joke is also in some sense on the process of encoding, which the playfulness of the poem seems to evade. Looking at this poem with the students aimed, then, to explore and question in practice some of the issues that Jerome McGann identified when he wrote that because the TEI “treats the humanities corpus – typically works of imagination – as informational structures, it ipso facto violates some of the most basic reading practices of the humanities community, scholarly as well as popular” [McGann 2001, 139]. Overall, then, the seminar introduced a theoretical engagement with the TEI as well as teaching students how to use XML.

It was in light of this reflective engagement with the TEI — on its possible limitations and problems as well as its benefits — that we asked the students to consider TEI in relation to their own projects. The key question was what problems could TEI solve for them, and what could it offer the user of the site that was currently not otherwise possible. The work we did in class on encoding various genres of literary text complemented the work we did with the students in the lab session which focused primarily around their own project work with manuscripts. One key area for discussion here was the potential for the TEI to render searchable the textual-content captured and presented primarily through images. The students’ work with hand-written manuscripts presented particularly interesting possibilities here, in relation to providing options for the interpretation of complex and problematic script, and for presenting issues of manuscript materiality. For example, a project digitizing nineteenth-century hand-written letters, in which additions were often placed in the gaps on the page, prompted class debate on how best to render this text and useful discussion about the complementarity of manuscript image and text encoding. In addition to transcription, the students also used the TEI to note additional contextual information as well as the usual metadata. For example, the student who digitized letters written from various locations in New York City used the TEI to encode the geographical data. Another used the TEI to add information about the hand-drawn images that appeared in her text, thus enabling a further explanatory and interpretative commentary that altered the way the text would be read.

Assessment

The structure of the assessment reflected the structure of the course more generally in that we began with smaller separate technical and literary assignments and then brought the two together in the main project. The first two assignments (building a small website for the purposes of studying HTML and CSS; and evaluating existing on-line text-archive sites) were assessed by the Computer Science and English faculty members respectively. The final online digital archive and the TEI-encoded literary texts were assessed by the two faculty members together. The course ended with the students formally presenting their sites to the group for the final assessment in order to maintain the group dynamic and the sense of peer-sharing which had been so important throughout the course.

Conclusions
Just as the Digital Humanities is an emerging field, so is the pedagogy associated with it. Case studies such as those described in *Digital Humanities Pedagogy: Practices, Principles, and Politics* [Hirsch 2012] show a variety of pedagogical goals both within and between different disciplinary areas in the Humanities. Lists of competencies such as those cited in *Specification 3 in Digital_Humanities* [Burdick et al 2012, 132–133] offer a different approach: documenting specific technical skills required for humanities students and scholars to successfully create and implement computational and digitally-based research and publications. Our collaborative approach — of building on principles of Computer Science pedagogy in combination with teaching principles of digital literary scholarship — aimed to provide our students with the best of both worlds.

The principles of our collaboration might be summarized under four points. The first being the preparation we did before the course: planning how we would interconnect the concerns of our respective disciplines conceptually and practically in relation to the structure of each class. The second being the weekly meetings we had to monitor and adjust the course content and pedagogical approach. The third category relates to the differences in disciplinary teaching methods. Our collaboration involved an integration, or alternation, of the seminar — or roundtable — discursive style of the humanities with the more lecture-driven style of Computer Science complemented with lab sessions. Fourth, and finally, we sought to role-model interdisciplinary collaboration in various ways throughout the class, showing the students that the same question might have two different answers: one technical and one from a scholarly humanities perspective; we then pursued the connections through discussion with the class, and through discussion with each other in front of the class. While at the start of the course we presented the introductory technological skills and the key humanities questions separately, the goal was to achieve convergence of these perspectives over the length of the course by teaching the students to see every technical decision they made in building their sites as an interpretative gesture (and to see their interpretative and critical analyses as having a technical embodiment).

Various problems and challenges arose as we taught the course. For example, the imaging of the primary materials occurred early on in the semester and generated a great deal of enthusiasm that resulted in students imaging more material than they could reasonably work with; in the future we would ask students to think more carefully at this stage to limit the focus and give more scope for the qualitative rather than quantitative work. We also ran into problems with the ever-changing technological environment. For example, Google was phasing out support for KML, which required a new solution to be found for one student’s mapping work; and TEI-Boilerplate turned out to serve our purposes less well than expected in the CMS environment. In terms of pedagogical approaches, we adapted the course as it progressed to fit the needs of the students; in particular, we worked to build in greater opportunities for discussion and sharing information between the students and their respective projects. In addition, once we began working with the students on their individual projects, we decided to use the particular challenges they were facing with their materials as our examples in the class lectures.

We have engaged with other departments to see how the course might be modified to become suitable for a broader range of humanities students. Our initial assessment suggests there would be ways to develop the course to cater to students from a range of disciplines in Literature and History. For example, introducing EAD (Encoded Archival Description: http://www.loc.gov/ead/) along with TEI as examples of XML languages would be useful for history students and complement work done by literature students and history students using TEI with manuscripts and documents. The addition of an historian to our teaching team would be essential in enabling us to expand our interdisciplinary conceptualization and theorization of the online archive in conjunction with this expanded technological syllabus. Having taught the course once we are also considering the following changes for the second iteration: the addition of a lecture on XSLT formatting the TEI examples used earlier in the class; and the addition of more reading and discussion on the theory of TEI.

Finally, we have found that teaching collaboratively enabled us to think about new research practices. This connection has not always been obvious:
From the outset, it was clear that teaching together in connected courses would be good for our students. What was not so clear was how good it would be for the faculty’s professional development and research output. [LeBlanc et al 2010]

Yet, as we worked together, we better appreciated the overlay of our fields of study, and ideas for new research collaborations emerged. Digital Humanities is an area in which collaboration can be particularly valuable to ensure a depth and breadth of both technical and conceptual knowledge, and through our collaboration we felt able to explore profound methodological intersections between the fields of technology and the humanities.

Appendix I: Course Description and Syllabus

(http://cs.nyu.edu/courses/fall13/ENGL-GA.2971-001/DT_index_fa13.php)

The interface of technology and the humanities represents a key to the future, yet many students feel they lack the skills to access this potential. This course offers an introduction to web development and digital publication especially created for students in the Humanities, with a view to equipping you with knowledge foundational for reflective engagement with the new media of literary creation and dissemination. Students will survey the principles of current technologies and apply them through practice as they learn the skills and techniques for formatting and publishing archival materials in a web-based environment. The course builds towards the creation of a digital edition, giving you the opportunity to work with primary source materials available through NYU’s rich archival collections (these include a wide variety of printed texts, manuscripts and images from which to select according to your interests).

The course will consist of a traditional classroom lecture and discussion format as well as computer lab sessions to promote and assist students as they work on each of their three projects in this course. Each student will have an account on a production webserver to post their work and learn to install and configure a WordPress site specifically tailored to his or her primary source materials. Topics and assigned projects will begin with an introduction to mark-up languages and building a site of related web pages followed by a project centered on encoding and annotating digital texts for scholarly purposes. The final project involves photographing or scanning, transcribing, and encoding digital texts to build an online archive.

<table>
<thead>
<tr>
<th>Week</th>
<th>Topics Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 56</td>
<td>1st hour – Literature: introduction to digitization / literary projects</td>
</tr>
<tr>
<td></td>
<td>2nd hour – Technology: Introduction to web design, mark-up languages and related technologies. Lab - Hands-on: Project #1 - Unix; HTML, xHTML, CSS</td>
</tr>
<tr>
<td>2 60</td>
<td>1st hour – Technology: Advanced CSS; Web-authoring software; Integration of text and graphics on the web</td>
</tr>
<tr>
<td></td>
<td>2nd hour – Literature: Introduction to selected scholarly literary archival sites and critical discussion. Lab: Hands-on: Project #1 continued</td>
</tr>
<tr>
<td>3 64</td>
<td>1st hour – Literature: Students present the scholarly literary sites to the class which they have prepared in assignment #1 followed by discussion.</td>
</tr>
<tr>
<td>Week</td>
<td>1st hour – Technology</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>4</td>
<td>1st hour – Technology: Working with image files: file types and how to use them (GIF, JPG, PNG for use on the web as well as RAW, TIFF, PSD for archival purposes); discussion on image compression, digitization standards, and related topics; instructions on scanning images using a flat-bed scanner as well as working with cameras to capture images of archival materials.</td>
</tr>
<tr>
<td></td>
<td>1st hour – Technology: Guest Lecture in the digital imaging lab.</td>
</tr>
<tr>
<td>5</td>
<td>1st hour – Technology: Guest Lecture in the digital imaging lab.</td>
</tr>
<tr>
<td></td>
<td>The students worked in the imaging lab with both professors present to accomplish the following in the second hour:</td>
</tr>
<tr>
<td>6</td>
<td>1st hour – Technology: Guest Lecture in the digital imaging lab.</td>
</tr>
<tr>
<td></td>
<td>Part 1 (1.5 hours): Technology – Introduction to working with a CMS; in-class demonstration of a WordPress installation, configuration and implementation; as well as themes modification and downloading themes appropriate the students’ selected archival materials.</td>
</tr>
<tr>
<td>7</td>
<td>Part 1 (1.5 hours): Technology – Introduction to working with a CMS; in-class demonstration of a WordPress installation, configuration and implementation; as well as themes modification and downloading themes appropriate the students’ selected archival materials.</td>
</tr>
<tr>
<td></td>
<td>Part 2 – Literature: ½ hour: Round table discussion as each student describes his/her content selection and literary/scholarly focus.</td>
</tr>
<tr>
<td>8</td>
<td>Technology Class – 2 hours: WordPress and CMS concepts; in-class demonstration of site customization; integration of text and images.</td>
</tr>
<tr>
<td></td>
<td>Lab: Project #2 Students begin building the cataloguing section of the site.</td>
</tr>
<tr>
<td></td>
<td>Literature Class – 2 hours: discussion on problems and possibilities that students encounter with their selected texts (including discussion of the principles of manuscript transcription).</td>
</tr>
</tbody>
</table>

http://www.digitalhumanities.org/dhq/vol/9/1/000199/000199.html
Table 1.

<table>
<thead>
<tr>
<th>Page</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Lab: Project #2 – Students continue work on their project.</td>
</tr>
<tr>
<td>88</td>
<td>Technology – 1st hour: ½ hour each on technical problems and questions students have about their individual sites; and a summary on the contents expected for the technical &quot;about&quot; page on the site.</td>
</tr>
<tr>
<td></td>
<td>Technology 2nd hour: Introduction to XML, encoding text and XML-TEI.</td>
</tr>
<tr>
<td></td>
<td>Literature: poems by William Carlos Williams.</td>
</tr>
<tr>
<td></td>
<td>Lab: Hands-on: Completion of Project #2; Students begin work on Project #3 (XML-TEI)</td>
</tr>
<tr>
<td>93</td>
<td>1st hour: Technology: Class lecture and discussion on XML-TEI.</td>
</tr>
<tr>
<td>10</td>
<td>2nd hour: Technology/Literature: Encoding Hamlet’s Soliloquy and the Soliloquy from Huck Finn (Chapter 21, excerpt) in XML-TEI. Students are asked to encode Shakespeare's Sonnet 116 for the following week.</td>
</tr>
<tr>
<td></td>
<td>Lab: Project #3 – XML-TEI, continued</td>
</tr>
<tr>
<td>97</td>
<td>1.5 hours: Literature - Case study on sonnets.</td>
</tr>
<tr>
<td>11</td>
<td>Last ½ hour: Technology - introduction to XSL/ XSLT</td>
</tr>
<tr>
<td></td>
<td>Lab: Project #3 – XML-TEI, continued</td>
</tr>
<tr>
<td>101</td>
<td>1st hour: Technology Site sustainability (e.g. creating a static HTML/CSS site from a CMS site for preservation purposes and permanent installations and technical questions that came up.</td>
</tr>
<tr>
<td>12</td>
<td>2nd hour: Literature – Discussion on the “about” statements from a literary perspective.</td>
</tr>
<tr>
<td></td>
<td>Lab: Completion of materials and preparation for class presentations.</td>
</tr>
<tr>
<td>105</td>
<td>1st hour: Student Presentations in lieu of a final exam</td>
</tr>
<tr>
<td>13</td>
<td>2nd hour: Technology – Additional requested topics included WordPress for blogging and considerations in crowd-sourcing; using Oxygen for editing XML; and other technical questions and topics that arose as students worked on their sites.</td>
</tr>
<tr>
<td>108</td>
<td>Student Presentations in lieu of a final exam.</td>
</tr>
<tr>
<td>14</td>
<td>Discussion on future projects and where to go from here.</td>
</tr>
</tbody>
</table>

http://www.digitalhumanities.org/dhq/vol/9/1/000199/000199.html
Assessment and Grade Distribution:

- Assignment #1 (Humanities site evaluation) – 10% of the final grade
- Assignment #2 (Proposal for project #2 – selection of Fales materials, including rationale) – included in the project #2 grade
- Project #1 – Build a unified site on an author of the student’s choice - 20% of the final grade
- Project #2 – Online digital archive based on primary source materials using WordPress - 40% of the final grade.
- Project #3 – Use TEI to encode several representative samples of literary texts in the student’s archival site - 30% of the final grade

Notes


Works Cited


Engel 2013 Engel, Deena. Embracing the Digital Humanities: A Course on Computing in the Humanities for Undergraduate Computer Science Minor Students
http://db.grinnell.edu/sigcse/sigcse2013/Program/viewAcceptedProposal.pdf?sessionType=poster&sessionNumber=2 SIGCSE 2013.


Gousie 2006 Gousie, Michael B. “A robust web programming and graphics course for non-majors.” SIGCSE Bull. 38, 1 (March 2006), 72-76.

Guzdial 2010 Guzdial, Mark. “Does contextualized computing education help?” *ACM Inroads* 1, 4 (December 2010), 4-6.


104.

