Digital Collection Development in English Literature

INFO 653: Digital Libraries

Vickie Marre
August 31, 2013
Abstract

Digital collections in English literature provide access to a myriad of full-text literary works, encompassing a wide array of genres, time periods, and specialized topics. With increasing numbers of collections available, however, quality and interoperability become crucial issues for English scholars when choosing digital resources. This paper examines common topics in English literature digital collection development, including content, metadata and text encoding standards, and interactive and collaborative elements in conjunction with Digital Humanities topics. Three collections are examined to exemplify the breadth and depth of English collections available to scholars: EEBO (Early English Books Online), NINES (Networked Infrastructure for Nineteenth Century Electronic Scholarship), and MONK (Metadata Offer New Knowledge) Workbench. These collections, diverse in both content and digital quality, create new implications for literary methods and scholarship, allowing English scholars to participate in activities such as text mining and distant reading. As developers and curators of these collections, librarians specializing in literary studies must be willing to experiment with digital collections and prepared to advise English scholars not only on collections’ content and quality but also on relevance for specific digital research.

Introduction

In 1988, Cynthia Selfe, distinguished English Professor at Ohio State University (OSU, 2013), championed the use of computing and digital scholarship in English departments. She wrote, “If our current use of computers in English studies is marked by any common theme at all, it is experimentation at the most basic level” (as cited in Kirschenbaum, 2010, p. 55). Studying the development of computer usage in English departments, and more specifically, digital collections in English literature, since Selfe’s time comprises a variety of topics, from the
nature of literary scholarship to collection content and organization to text encoding and
crossover topics in the developing Digital Humanities (“DH”) field. For literary scholars, digital
collections provide an opportunity to experiment with texts and language in innovative ways and
hold great promise for interactive learning, especially for students. With the advent of large-scale
digitization projects, such as Project Gutenberg, Google Books, and HathiTrust Digital Library,
both scholars and the general public can easily access countless works of English literature. Such
digital collections have changed the landscape of both literary studies and the types of resources
that librarians make known to scholars in English literature.

With increasing numbers of digital collections in English literature available, however,
quality, both in terms of content (images/texts) and digital architecture (metadata/encoding),
becomes an important issue for scholars interested in experimenting with age-old texts in
innovative ways. While digital collections that provide full-text access and high-quality images
may satisfy junior researchers, more advanced scholars seek sophisticated ways to manipulate
metadata for their research, for example, to graph how word usage changes over time or the
significance of certain terms in a period’s historical and cultural contexts. Nevertheless, digital
English collections allow scholars of all levels to participate in such original research and invite
the kinds of interactive and collaborative work that the DH field values.

This study will trace developments in English digital collections, through topics such as
collection content and topic diversity, metadata standards for encoded collections, and
implications for interactive and collaborative scholarship that digital collections invite. Through
an exploration of EEBO (Early English Books Online), NINES (Networked Infrastructure for
Nineteenth Century Electronic Scholarship), and MONK (Metadata Offer New Knowledge)
Workbench, quality of collections and usefulness to literary scholars will be evaluated. Changing
landscapes in digital literary studies will be discussed in light of developing digital collections. The role of the literature librarian will also be addressed, as English librarians must be well versed in the diversity of English digital collections and in issues of current literary scholarship, in order to assist scholars in using digital tools for their research.

**Literature Review**

At the turn of the twenty-first century, there was much speculation from those in English departments and libraries alike about the scholarly changes that digital literary texts would present to researchers. Discussing such a change in landscape, Day and Wortman (2000) coined a new format, “literary technology” (p. 11), meaning that while literature itself has remained the same, collections have become abundant in electronic form in addition to print. Further, in managing digital literary collections and “literary technology,” the librarian’s job not only involves preserving literary texts, but also, and more importantly, “renew[ing], remaking, and reassembling the literary cannon” (Day & Wortman, 2000). This work involves both technical and subject understanding as well as a willingness to experiment with collections (Willett, 2000).

Beginning with subject coverage and content, digital collections in English literature cover a wide array of genres, periods, specialized topics, and authors (Arlen, 2000). General collections that provide open access to primary texts with aims of long-term availability and preservation include those like HathiTrust, with millions of digitized materials maintained by the University of Michigan (HathiTrust, 2013); Project Guttenberg, with access to hundreds of thousands of e-books in multiple formats (Perrault, 2013b); and the Electronic Text Center at the University of Virginia, with thousands of full-text searchable materials (Perrault, 2013b). Genre-specific collections, including poetry and drama, include subscription databases, such as Columbia Granger’s World of Poetry Online maintained by Columbia University, and open
access sources, such as the Poetry Archive, with influential works of British poets from the Victorian period to the twenty-first century (Perrault, 2013a).

More diverse are English collections arranged by time period, with major projects in the Renaissance, Early Modern, Romantic, Victorian, and Modern Eras. Two of the largest subscription-based digital collections in Early Modern British literature include EEBO, or *Early English Books Online* offered through Chadwyck-Healy (a ProQuest product), and ECCO, or *Eighteenth Century Collections Online*, by Thomson-Gale (Lindquist & Wicht, 2007). Open access collections range in genre from the all-encompassing *Literary Resources on the Net*, maintained by Rutgers University English Professor Jack Lynch, and *Voice of the Shuttle* by UC Santa Barbara (Perrault, 2013a), to genre-specific and special collections, including: *Romantic Circles*, from the University of Maryland (Arlen, 2000); Brown University’s *Victorian Web* and *Women Writers Project*; and Yale University’s *Modernism Lab*. Specialized collections include: Brown and the University of Tulsa’s *Modernist Journals Project*; Indiana University’s *Victorian Women Writers Project*; and the University of Pennsylvania’s Schoenberg Center for Electronic Text & Image (SCETI) collection, with strengths in English Renaissance texts and Women’s Studies (Arlen, 2000). Digital collections on particular authors are diverse; some of the most influential include SCETI’s fully-searchable online collection from the Furness Shakespeare Library (Brown, 2006); The Dickens Project, by UC Santa Cruz; and Columbia University’s Digital Dante Project (Shipe, 2000).

More recent digital collections in English literature include those that are text-encoded, allowing for easy data manipulation and experimentation for scholars. For example, the MONK (*Metadata Offer New Knowledge*) Project includes a “Workbench” that provides access to 525 works of American literature and 42 works by Shakespeare; for scholars affiliated with member
institutions, thousands of works in British literature from the 16th to 19th centuries are available to manipulate. Each text in the collection has been “adorned” with sentence boundaries, spelling normalizations, and parts of speech, and has been encoded according to TEI standards (MONK Workbench, n.d.). As the preferred scheme in the humanities (Shipe, 2000), TEI makes e-texts ripe for data extraction and manipulation. Older collections, such as Indiana University’s *Victorian Women Writers Project*, which began in 1995, are now ensuring that new texts added to its collections are TEI encoded for “more sophisticated retrieval and analysis” (Indiana University, 2013), as the website explains.

Although increasing numbers of digital collections in English literature are developing with particular attention to standardized metadata and encoding schema, problems arise when schemas are not standardized, leading to much frustration for literary scholars. In their study of EEBO at the University of Colorado Boulder, Lindquist & Wicht (2007) discuss the difficulties that can arise when students work with full-text primary source English collections, including variations in historical word usage, spelling, and typography (p. 349). Cummings & Rahtz (2012) also study EEBO, analyzing problems with digital files created through TCP, the Text Creation Partnership. The authors discuss converting TCP files to the latest TEI P5 standards; challenges include misrepresented attributes and incompatible content models, new developments in both TCP and TEI standards since EEBO files were created, and issues in making XML files generally easier to manipulate for scholars (Cummings & Rahtz, 2012).

As the Cummings & Rahtz (2012) article displays, a large consideration when working with digital collections in English involves vendor-created digital files versus open-source files, created through initiatives such as TEI. Dino Franco Felluga (2006) discusses text-encoding barriers in 19th-century literary collections, mainly comparing vendor-produced databases like
EEBO and EECO, with open-source, open-access sites, namely the NINES (*Networked Infrastructure for Nineteenth Century Electronic Scholarship*). Felluga (2006) notes the ease with which vendor databases allow modifications to TCP and TEI standards; however, scholars remain generally either unable or unwilling to “retrofit” their work to current XML and TEI standards to produce open access alternatives (p. 312). Felluga (2006) argues that sites like the NINES give authors control over their own text creation by educating users on principles of XML and RDF (resource description framework), which the NINES employs for more reliable and precise searches. In this way, the NINES offers one example of an alternative to vendor databases, particularly for nineteenth-century literary studies.

Such encoding challenges in English digital collections are not only significant for scholars who work with texts from a variety of digital platforms, but also for cataloging and authority control in accordance with Library of Congress authorities and subject headings (Shipe, 2000). Walsh (2010) notes that while the Library of Congress Subject Heading (LCSH) is the most popular choice for subject access in digital libraries and collections in North America (p. 329), LCSH is often criticized for its difficulty of use and poor retrieval. One major advantage of using LCSH in digital collections is interoperability among collections. However, Walsh (2010) discusses disadvantages to LCSH in digital collections, namely that users do not tend to search by subject but rather by keywords and phrases, in hopes that metadata will match. In contrast, other scholars note the successes of folksonomies in digital English collections, as opposed to controlled vocabulary. In *William Blake and the Digital Humanities* (2013), for example, Whitson and Whittaker study the effects of folksonomies in Blake studies, allowing students to craft their own collections within collections. User tagging is especially useful in digital platforms, such as Flickr and YouTube, allowing researchers to examine how authors like
William Blake, in Whitson & Whittaker’s (2013) study, are imagined and culturally represented in online communities. In spite of challenges including metadata, encoding, and organization that scholars mention in studies of digital English collections, such collections have produced great strides toward more interactive scholarship and new discoveries from old texts and materials. As Whitson and Whittaker’s (2013) project on William Blake displays, digital collections, enhanced by various social media platforms, have opened up new avenues for scholars and students to make literary studies “collective, multimodal, and participatory” (p. 2). Stephen Clarke (2000) makes a similar argument in digital Shakespeare studies; while not experimenting with social media, Clarke (2000) argues that digital texts and images of Shakespeare’s works invite an involvement, discussion, and exchange that print collections do not always enable (p. 105). Overall, the literature shows that digital collections in English have been changing the shape of literary studies. Digital collections not only allow scholars to work creatively with online tools to make texts come alive, but they also instruct researchers on how to actively participate in critical analysis and thus further inform cultural studies.

**Examination of Digital Collections: EEBO, NINES, and MONK Workbench**

To assess topics presented in the literature review, three prominent digital collections in English literature were chosen to evaluate how different levels of collections can serve English scholars. Factors considered included: content (research and discovery), digital quality (metadata and encoding standards), and interactivity (collaboration with other users and manipulation of metadata). The below examples will demonstrate how each collection was explored and how it met these criteria.
The EEBO database, created in 1999 through a partnership between ProQuest, the University of Michigan, and Oxford University, contains 125,000 titles from the Early Modern period of British literature. Through the Text Creation Partnership (TCP), EEBO houses fully searchable, TEI-compliant SGML/XML texts (EEBO, 2013). As the Lindquist & Wicht (2007) study displayed, EEBO offers a wide corpus of texts and is easily searchable. Undergraduates whom the authors studied generally used EEBO for upper-level research papers, while graduate students and faculty used the database for more complex research (Lindquist & Wicht, 2007).

In examining EEBO’s basic features, the collection interface proved user-friendly, with explicit instructions on search terms and parameters. When searching, users are given the option to select variant spellings and word forms, for which EEBO provides extensive explanations for users such as undergraduates, who may be unfamiliar with different spellings and historical usages. Searches can also be conducted by keyword, full-text result, results with images, title, subject, date, and bibliographic number, or the number that EEBO assigns to each text (see Appendix A, Fig. 1). To examine EEBO’s search functions, a simple author search for author and clergyman “John Donne” was performed; results included Done, John; Donne, John (1572-1631); and Donne, John (1604-1662). The second Donne is the author for whom the search was meant; however, this could be confusing to a user who did not have this information beforehand. This result retrieved 72 records, with 281 hits for John Donne, which can be determined since the collection is full-text searchable and considers each instance of the author’s name. EEBO provides the option to view documents as a record, document image, illustration, full text and thumbnail image. The illustration feature was selected, which provided rather clear reproductions of Donne’s *Fifty Sermons*, published posthumously in 1649 (see Appendix A, Fig. 2).
default image size views the page at 41%; however, image size can be increased to 400%. A sample page from Donne’s *Fifty Sermons* was viewed at 150%, which enabled easy reading of the material (see Appendix A, Fig. 3).

Overall, EEBO proved effective for scholarly research purposes; the site functions well for those interested in accessing a wide variety of Early Modern texts. Because EEBO is a subscription-based product, however, not all scholars will have access to it through their institution.

**NINES (Networked Infrastructure for Nineteenth Century Electronic Scholarship)**

As Felluga (2006) describes, NINES is an open-source, open-access, peer-reviewed repository for nineteenth-century British and American literary collections. As a digital repository, NINES uses the Collex interface (allowing sharing, tagging, collecting, and exhibiting works and images) to gather quality nineteenth-century resources that are fully searchable, interoperable, and available for scholars to use for creation and publication of works (NINES, n.d.). The NINES home page features user tags, digital collections, news, and a featured object (see Appendix B, Fig. 1).

To explore the site, the tag for “cats” was clicked, which retrieved four images: three from the British Library and one from the University of Florida’s Historical Children’s Literature Collection, called *Animal Life All the World Over*, by H. W. Dulcken (1880s) (see Appendix B, Fig. 2). Before clicking on the resource and being redirected to the digital text on the University of Florida’s site, NINES provides options to “collect,” “discuss,” or “exhibit” a work if a user already has a collection. In order to choose these options, the user must sign up for an account, which merely requires a valid email address.

A very useful feature of the NINES is the “Classroom” section, which allows instructors
to share collections and exhibits that students have created. In fact, once a user logs in, certain coursework is publicly available and can be searched by institution or by course name. A sample collection chosen was entitled “ENGL 227: Survey of American Literature,” a course at Texas A & M University (see Appendix B, Fig. 3). The collection then shows each student’s work and gives students the option to create discussion threads; one collection examined is entitled “American Independence” (see Appendix B, Fig. 4).

With options for tagging, discussing, and creating collections, NINES provides a rich platform for students and researchers to collaborate on and interact with nineteenth-century literature collections. As an open access collection, NINES allows for more user collaboration and manipulation of materials and metadata than a proprietary database like EEBO.

**MONK (Metadata Offer New Knowledge) Workbench**

The MONK Project was funded by the Andrew W. Mellon Foundation from 2007-2009 with the goal of providing open-source, TEI-encoded text collections for humanities researchers to mine and analyze. Texts have been contributed from Indiana University, the University of North Carolina at Chapel Hill, the University of Virginia, and Northwestern University (MONK, n.d.a). The MONK Workbench in particular provides access to 525 works from 18th- and 19th-century American literature, in addition to 42 of Shakespeare’s plays and poems (MONK Workbench, n.d.). The Workbench offers documentation and tutorials for how to use analytic features, explaining that its tools are ideal for humanities scholars with little background in statistics (MONK, n.d.c). The site offers easy access to certain collected Shakespeare works in addition to many other users’ test materials that have been uploaded and saved as “worksets.” If users do not want to create an account, they can simply log in as “guest” and gain access to all of these tools.
To explore the site, a user selects a project from those saved by MONK and other users. The project “Shakespeare Plays” was selected and analyzed. The simplest form of visualizing a work is called “Classification,” in which the user can rate items in a collection based on frequency and previously tested analytics. The Classification feature provides the full text of each item and also indicates the TEI standards used to encode the text (see Appendix C, Fig. 1). To perform comparison analytics within a certain workset (continuing to use the “Shakespeare Plays” example) (see Appendix C, Fig. 2), a user can choose among specific types of plays, for example, “comedy” and “history,” in order to compare the frequency of adjectives’ spelling in each work and combined between the two works (see Appendix C, Fig. 3). Another analysis the Workbench will perform is called “Dunning’s” (the Dunning’s log-likelihood statistic invented in the early 1990s by Ted Dunning) (MONK, n.d.b), which presents the adjectival frequency comparison between comedies and historical plays analyzed in the previous example. The results are presented as a word cloud juxtaposed with the raw data (see Appendix C, Fig. 4).

Overall, the MONK Workbench is a very powerful tool for researchers with little knowledge of data mining and statistical analysis but who wish compare and analyze word usage among various collections. The 15 CIC members (Committee on Institutional Cooperation, http://www.cic.net/home) have access to texts from databases such as EEBO, EECO, and Nineteenth-Century Fiction, while non-members can access the American literary works and Shakespearean plays provided by other libraries (MONK Workbench, n.d.).

Discussion

Although the digital collections in English literature examined above represent a small sample of available online sources, they reflect not only the variety of genre and period collections in English literature, but also differing levels of complexity based on a researcher’s
skill level and interest in manipulating texts. Each collection tested was highly usable in terms of design and interface. EEBO and NINES were easily searchable, with clear search boxes and options for advanced searches. As a vendor database, EEBO offered a variety of search options and controlled terms, whereas NINES provided a federated search of all its collections, in addition to displaying user tags prominently underneath the search bar to encourage users to explore in this fashion. The MONK Workbench site was not as initially intuitive. The interface did not display obvious “search” features and required reading various tutorials to understand how the site’s analytics work and the types of data comparisons that are available.

While all three collections provided powerful retrieval capabilities in terms of sophisticated searching and data manipulation, the collections were chosen purposefully to demonstrate the level of sophistication that different English digital collections can hold depending on a user’s needs. For close study of texts, the EEBO collection offers high-quality images of Early Modern texts, which are fully searchable. For example, an undergraduate student searching for patterns in John Donne’s attitudes toward the “divine” can easily create search terms that will lead him/her to specific texts that exemplify Donne’s thoughts. A graduate student can do more sophisticated searching to determine how Donne’s word usage changes over time through his texts, and what this means in the context of his Early Modern contemporaries. The NINES also allows these capabilities, by linking to sites that contain high-quality digital images of texts from libraries all over the world. However, the NINES goes one step further in inviting scholarly opinion, annotated texts, peer-reviewed collections, tagging, and collection building within the site. The NINES invites scholars to upload their own encoded works to share collections with others and promote further scholarship (Felluga, 2006). MONK Workbench also invites this kind of collaborative scholarship and contains more powerful analytic tools that
enable scholars to draw new conclusions from English texts in the collections. MONK mines data for scholars who provide certain parameters; it will create charts and graphs of word usage, allowing researchers to apply new computational methods to corpuses of texts (Little, 2011).

By providing quality content, encoded texts, and site interaction, each collection studied invites English scholars to experiment with texts in intriguing ways. In this way, digital collections have developed from merely making texts electronically available, as Project Guttenberg sought to do in the 1970s (Perrault, 2013b), to using word frequencies to map characters’ journeys around villages in English fiction, for example (Moretti, 2005). Further, such interactive studies as Whitson and Whitaker’s (2013) on William Blake, for example, espouse the principles of Digital Humanities scholarship, which any discussion of English digital collections must address. Although a somewhat elusive field, scholars define DH as the intersection of “humanities ideas and digital methods” (Perrault, 2013c, p. 18), examining scholarship across media, or “networked digital information” (Unsworth, 2000). DH studies espouse principles of open source/resource and collaborative knowledge that the field values, in addition to championing field standards, such as construction of text collections and metadata schemes (UCLA, 2009; Felluga, 2006; Walsh, 2010; Shipe, 2000; Cummings & Rahtz, 2012). Matthew Kirschenbaum (2010) addresses particular reasons why DH scholarship finds itself in English departments, including the tractable nature of textual data, the foundation of literary studies; the openness of English departments to larger cultural studies; and the prevalence of e-reading in today’s culture (p. 60). All of these points, as well as the topics discussed thus far, enhance the relevance of digital collections to the study of English literature, by enabling new discoveries from text that is generally widely available online and encoded for manipulation.
Conclusion

In 2000, when Andrew Goodwyn edited the book *English in the Digital Age*, he argued that we “have only begun to glimpse the possibilities offered by technology for bringing all kinds of texts closer to students” (p. 18). He discussed how digital texts will allow teachers to better understand students’ reading patterns, for example, which will enhance teaching and learning. Thirteen years later, digital collections in English have made such predictions realities and have altered the landscape of literary studies. English digital collections and possibilities for data manipulation have also, however, inspired many debates over traditional literary theories. For example, text and data mining have invited what Franco Moretti, Stanford literary scholar, has termed “distant reading,” which involves studying large corpuses of literature across time and space to draw unique conclusions (Moretti, 2004). Other scholars, however, challenge the full-text search capabilities of databases and digital collections. Garrett (2006) argues that keyword searching eliminates the critical thinking and analysis that research involves, and that terms become “decontextualized” and more difficult to relate back to the meaning of the original text (p. 5). Flanders (2002), in writing about Brown University’s *Women Writers Project*, argues that sheer unfamiliarity with a topic, including many of the little-known writers in WWP, is ameliorated by a friendly search interface and highly encoded data that allows users to make new textual and critical connections. Despite these debates, digital collections allow literary scholars to open new dialogues and to work with each other, in conjunction with digital texts, to create their own meaningful collections and conclusions.

To meet the changes and challenges that digital collections in English literature pose to literary scholars, librarians must be prepared to not only recognize quality digital collections but also to advise scholars on exploring the most appropriate collections for their research.
Evaluating quality digital collections in English literature not only depends on community needs, (i.e. undergraduate versus graduate students), but also involves examining content, digital architecture, usability, and services that collections offer users (Borgman, 1999). Many communities of scholars and organizations exist to assist librarians in these endeavors, including ALA, whose Principles of Digital Content (2007) reiterate that “[d]igital collections are as important to libraries as all other types of collections” (p. 1). ALA (2007) further states that digital content should be handled with the same care as print, and that collaboration enables scholars to build collections relevant to their research.

Librarians play a large role in instructing scholars on using digital collections and tools. For example, English scholars may rely upon librarians to build institutional digital collections using platforms such as Omeka and WordPress and to advise scholars on the best host sites for showcasing theirs and students’ works. Librarians also prove valuable in connecting scholars to communities that provide examples of literary projects, such as the Electronic Literature Organization (“ELO”) supported by the Library of Congress, which brings networks of scholars together to navigate issues of digital publishing and circulation of collections (ELO, n.d.). Overall, the diversity of digital collections in English literature and the various subject- and technology-related issues that such collections harbor provide significant considerations for librarians, who continue to develop digital resources to benefit increasingly interdisciplinary and participatory communities of English scholars.
References

files/content/oitp/PDFs/Principlesfinalfinal.pdf

B. H. Day & W. A. Wortman (Eds.), Literature in English: A guide for librarians in the
digital age (pp. 160-186). Chicago: ACRL.

and Management, 35(3), 227-243. Retrieved from
http://dx.doi.org.ezproxy2.library.drexel.edu/10.1016/S0306-4573(98)00059-4

Reviews, 20(2), 33-34. doi: 10.1108/09504120647438

Clarke, S. (2000). Changing technology, changing Shakespeare, or our daughter is a misprint. In
A. Goodwyn (Ed.), English in the Digital Age: Information and Communications
Technology and the Teaching of English (pp. 103-114). New York: Cassell.

Cummings, J., & Rahtz, S. (2012). Kicking and Screaming: Challenges and advantages of
bringing TCP texts into line with Text Encoding Initiative. Retrieved from
http://ora.ox.ac.uk/objects/uuid:f9667884-220b-4ec9-4bbf-c79044302399

W. A. Wortman (Eds.), Literature in English: A guide for librarians in the digital age
(pp. 1-19). Chicago: ACRL.

EEBO (Early English Books Online). (2013). About EEBO. Retrieved from
http://eebo.chadwyck.com/marketing/about.htm
http://eliterature.org/what-is-e-lit/


http://muse.jhu.edu.ezproxy2.library.drexel.edu/journals/pedagogy/v002/2.1flanders.html

http://quod.lib.umich.edu/j/jep/3336451.0009.106?rgn=main;view=fulltext


http://webapp1.dlib.indiana.edu/vwwp/welcome.do?jsessionid=09391F5FB633CFA76F9B36BCBCB49C3F


Appendix A

Screen Shots from EEBO Examination

Figure 1: EEBO Advanced Search Parameters
(image retrieved from http://proxy.library.upenn.edu:2395/search)

Figure 2: Cover page and first pages of John Donne’s Fifty Sermons (1649)
(images retrieved from http://proxy.library.upenn.edu:2395/search)
My Lord,

A season so tempestuous, it is a great encouragement to see your Lordship called to the Helme, who (in your publick negociations) having spent so many years in that so famed Common-wealth of Venice, must of necessity have brought home such excellent Principles of Government, that if our Fate doe not withstand your Directions, we may reasonably, at last, expect to see our new British Lady, excel that ancient Adriaticque Queene.

Figure 3: A page from Donne’s *Fifty Sermons*, magnified at 150% (image retrieved from http://proxy.library.upenn.edu:2395/search)
Appendix B

Screen Shots from NINES Examination

Figure 1: NINES homepage, with user tags, digital collections, news, and “featured object” (image retrieved from http://www.nines.org/)

Figure 2: Results for user tag “cats,” including H. W. Dulcken’s *Animal Life All the World Over* (1880s) (image retrieved from http://www.nines.org/tag/results?tag=cats&view=tag)
**Figure 3:** Collection entitled “ENGL 227: Survey of American Literature,” from Texas A & M University (image retrieved from [http://www.nines.org/groups/2](http://www.nines.org/groups/2))

**Figure 4:** Exhibit from ENGL 227 collection, entitled “American Independence” (image retrieved from [http://www.nines.org/exhibits/American_Independence](http://www.nines.org/exhibits/American_Independence))

Thomas Paine’s *Common Sense* is a great example of the mindset that Americans had concerning British rule. This particular edition is an explanation of Government in general with a focus on the English constitution. Paine points out that the colonies all have special needs and that as the colonies grow they will encounter special circumstances that Parliament will not understand because we are separated by 3000 miles of ocean. This leads Paine to conclude a need for a government run by American colonists who understand the circumstances and can make legal and binding decisions that are good for the colonies, rather than an English Lord across the sea. Paine also goes on to show the deficiencies of the English Constitution and that a Constitution of the people is more effective than a Constitution of the government.

Albert Mordell's book *American Literature* briefly discusses some of the influences that are seen in many of our founding documents. Mordell points out that some of the main concepts of democracy have been around since the time of the Greeks and Romans but other ideas, like separation of church and state, were indigenous to America. Mordell also explains that there are many factors that affected the democratic view of America, as he put it, "Social, political, and economic conditions in the seventeenth and eighteenth centuries in America, along with the average intelligence of the common man, went into the formulating of our democratic views and our republican government."
Appendix C
Screen Shots from MONK Workbench

Figure 1: Classification feature, with full text of Shakespeare’s *A Midsummer Night’s Dream*, including standard TEI header (image retrieved from http://monkpublic.library.illinois.edu/monkmiddleware/apps/workflow/index.jsp)

Figure 2: “Shakespeare Plays” workset from MONK data (image retrieved from http://monkpublic.library.illinois.edu/monkmiddleware/apps/workflow/index.jsp)
Figure 3: Frequency comparison of adjectives from 16 “comedy” (in blue) and 10 “history” (in green) Shakespearean plays (image retrieved from http://monkpublic.library.illinois.edu/monkmiddleware/apps/workflow/index.jsp)

Figure 4: Dunning’s analysis, with adjective frequency from “comedy” and “history” plays represented through a word cloud with raw data juxtaposed (image retrieved from http://monkpublic.library.illinois.edu/monkmiddleware/apps/workflow/index.jsp)