The Provision of Digital Apparatus for Use in Experimental Interfaces

Stan Ruecker Illinois Institute of Technology Stéfan Sinclair McGill University Teresa Dobson University of British Columbia Geoffrey Rockwell University of Alberta Milena Radzikowska Mount Royal University The INKE Research Group

Abstract

In this article, we discuss the various ways in which the experiments we have been doing within the INKE Interface Design team and elsewhere are predicated on the availability of "digital apparatus" – various forms of metadata that can be made consistently available. These include structural, procedural, and semantic markup, digital indexes, textual variants, annotations, regularized citations, and taxonomies of references, to name a few. While some affordances are agnostic to the very existence of metadata, in some crucial instances the metadata is essential. The question we hope to address in these cases is the extent to which the new affordances are actually of sufficient potential benefit to the community to warrant being produced and maintained.

Keywords

INKE; Digital texts; Human-computer interaction; HCI; Digital humanities; Humanities computing

Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

Stan Ruecker is Associate Professor at the IIT Institute of Design. Email: sruecker@id.iit.edu

Stéfan Sinclair is Associate Professor of Digital Humanities in the Department of Languages, Literature and Culture, McGill University. Email: Stefan.sinclair@mcgill.ca

Teresa Dobson is Associate Professor at the University of British Columbia. Email: Teresa.Dobson@ubc.ca

Geoffrey Rockwell is Director of the Kule Institute for Advanced Study, University of Alberta. Email: geoffrey .rockwell@ualberta.ca

Milena Radzikowska is

Associate Professor at Mount Royal University. Email: mradzikowska@gmail.com

CCSP Press Scholarly and Research Communication Volume 5, Issue 4, Article ID 0401194, 11 pages Journal URL: www.src-online.ca Received August 1, 2014, Accepted August 25, 2014, Published December 17, 2014

Ruecker, Stan, Sinclair, Stéfan, Dobson, Teresa, Rockwell, Geoffrey, Radzikowska, Milena, & INKE Research Group. (2014). The Provision of Digital Apparatus for Use in Experimental Interfaces. *Scholarly and Research Communication*, *5*(4): 0401194, 11 pp.

© 2014 Alex Christie & INKE-MVP Research Team. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org /licenses/by-nc-nd/2.5/ca), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

Introduction

The development and maintenance of metadata has been a central concern of the library and information studies community since its inception. Ranging from books shelved according to the name of the historical bust sitting on top of the bookcase, as was often the case in early British libraries (Brown, 1898), on down to the Dewey Decimal system and faceted browsing, the combination of affordances for both searching and browsing have been given due attention.

However, with the advent of digital records and digital documents, many projects have been faced with decisions regarding how much and what kinds of metadata should be kept. On the one hand, we might say the more the better, since it is clear that new tools can emerge from the metadata available (Ruecker, Radzikowska, & Sinclair, 2011). In fact, since it is widely accepted that an XML encoding is a form of interpretation (Hockey, 2000), capturing and disseminating various encodings could become a central interest of the digital humanities.

Unfortunately, with metadata comes cost, both in terms of development and storage. Without clearly defined intended purposes for the information, it is difficult to justify the additional work and complexity. It is a catch-22 in some ways, since experimentation with new tools emergent from the metadata is contingent, at least to a certain extent, on having some metadata to consider. In the Implementing New Knowledge Environments (INKE) project, we have been privileged to work with a number of research partners in obtaining the kinds of metadata necessary for prototyping.

INKE was defined with two primary goals:

- Contribute to the development of new digital knowledge environments.
- Combine the traditional strengths of print with the advantages of digital environments in order to produce integrated digital reading tools.

In both cases, the emphasis is on the environment, the package of integrated tools. It focuses on the larger context, which goes beyond the individual experimental prototype and into the ways in which the scholarly community may benefit from having a larger system available.

INKE, however, remains primarily a research project, and as such, it is important for us not to lose sight of the future. We are working toward an image of a system that does not yet exist, but that might have the potential to allow us to do so much more than we can do now.

INKE therefore stands in contrast to projects like the Text Analysis Portal for Research (TAPoR) and the more recent Canadian Writing Research Collaboratory (CWRC), which were defined from the beginning as national infrastructure projects. Where the goal of INKE is to imagine, prototype, and test our way into a better understanding of what needs to be done, the goal of TAPoR, CWRC, and a variety of similar infrastructure projects around the world, is to take the current state of the art and make it available to the scholarly community.

The distinction is important to keep in mind, in part, because the partnership roles vary considerably between the two approaches. For an infrastructure project, the partners work toward industry standards. They contribute data and metadata for improved access and use by the scholarly community. For a research project, on the other hand, the partner organizations are extending their reach into the future, combining their existing knowledge of best practices with the research goal of moving toward the next generation of best practices. This more speculative role can sit somewhat uncomfortably for some partners, who are conscious of limited resources. However, for the field to progress, both kinds of commitment are necessary.

Record headers/bibliographic information

We deal more completely with possible uses of citations below; however, document description metadata, such as appears in Text Encoding Initiative (TEI) headers or library catalogue records, has a number of potential uses in interactive visualization experiments. First is the browsing environment, where the goal of the reader is to get a sense of what is in the collection. An example of this kind of use is the TextTiles project (see Figure 1), where metadata providing document descriptions is used to populate an array of small boxes that can be interactively changed both in terms of how much metadata they display and how they are organized on the screen.

Figure 1: The TextTiles browser shows some of the bloggers from A Day in the Life of the Digital Humanities, 2009.

Texttiles Day in the Life of the Digital Humanities Master			Julianne Nyhan author	Merrilee Proffitt author	Alastair Dunning author	Richard Cunningham author	Matthew Steven Carlos author	Elaine Murphy author
Site Feed Pages Total Tiles [95]			Brett Bobley author	Joris J. van Zundert author	Willard McCarty author	Laboratoire NT2 author	Claire Warwick author	Bill Turkel author
Controls display group			K Faith Lawrence author	Dot Porter author	Shawn Day author	Paul Youngman author	Amanda Cash author	Matthew K. Gold author
title author organization	~		Susan Hesemeier author	Andrew Mactavish author	Rafael Alvarado author	Vika Zafrin author	Rudy McDaniel author	Malka Čeh author
city country			Trevor Owens author	Amanda French author	Kevin Kee author	William Allon author	Diane Neary author	Katherine Harris author
url year month			Martin Holmes author	Miran Hladnik author	Susan Schreibman author	Sara Steger author	Ray Siemens author	Stefan Sinclair author
date hour			Richard Cunningham author	Michael Eberle-Sinatra author	Maureen Engel author	Matthew Bouchard author	Martin Muellor author	Marcel OGorman author
minute alphabet category			Jon Bath author	John Newman author	James Chartrand author	Chris Meister author	Domenico Fiormonte author	Daniel ODonnel author
content clear all	back fo	rward	Alan Galey author	Aimee Morrison author	Sander Westerhout author	Kathryn Eccles author	Elaine Murphy author	Ingrid Mason author
Auto-Resize Saved Tiles	+	Save	Roger Osborne author	Ashton Nichols author	Frederic Clavert author	Gary Stringer author	Steven Hayes author	Craig Bellamy author
clear group: organization		ear Kai-Christian Bruhn author	Patrick Sahle author	Patrik Svensson author	Federico Meschini author	Alejandro Giacometti author	John Bradley author	
group: city group: organization			Teresa Dobson author	Megan Meredith-Lobay author	Stan Ruecker author	Geoffrey Rockwell author	Peter Organisciak author	Kamal Ranaweera author
			Deselected Tiles [0] Click to	switch Selection				

Ruecker, Stan, Sinclair, Stéfan, Dobson, Teresa, Rockwell, Geoffrey, Radzikowska, Milena, & INKE Research Group. (2014). The Provision of Digital Apparatus for Use in Experimental Interfaces. *Scholarly and Research Communication*, 5(4): 0401194, 11 pp. Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

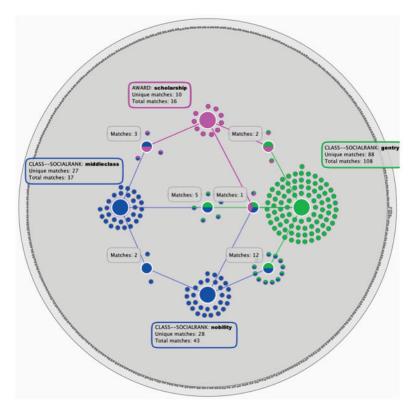
Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

As with many prototypes in the digital humanities (DH), TextTiles also includes a reading panel that can be used to examine the original file. Although there is some utility in having document descriptions separated from their contents, it has become recognized as a DH best practice to keep the contents ready to hand whenever subscription rights or other copyright concerns make it possible.

Structural markup

Primarily of value in formatting XML for reading purposes, structural markup can also have a role to play in interactive visualizations. In the Mandala Browser (see Figure 2), for example (Orlando Project, 2006), we rely on the structural markup as a means of deciding how best to subdivide the file into the "dots" that constitute the core unit of analysis. It is important in structuring an interview, for instance, that the question be somehow associated with the answer, so that the answers are not showing up *ex nihilo*. Similarly, in studying plays, it is useful to have the name of the character associated with the speech.

Figure 2: The Mandala Browser with dots indicating biographies of British women writers extracted from the structural markup of the Orlando Project.



Without structural markup, it is not possible to make decisions about how best to divide a file. Instead, it is necessary to accept a default division that is ready to hand, such as paragraph breaks.

Semantic markup

Having some higher-level intelligence available in the metadata opens up opportunities for new affordances. That is, if we know more when we are designing a prototype, we can think of more things that can be done. In the Bubblelines visualization (see Figure 3), for example, the reader can see comparative search results across multiple documents, or across multiple parts (e.g., chapters) of the same document. Of general use for anyone doing string searches, the tool becomes more powerful when working with material that has been previously prepared using XML encoding to address some particular research question. Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

Figure 3: This Bubblelines screenshot shows the first few letters in Richardson's epistolary novel *Clarissa*, for which the researcher Susan Liepert has developed and encoded an XML schema that captures various indications of emotion and emotional interactions.



By providing encoding that is relevant to the interpretation being pursued, the researcher is able to identify patterns across the document that would not otherwise be easy to see.

Digital indexes

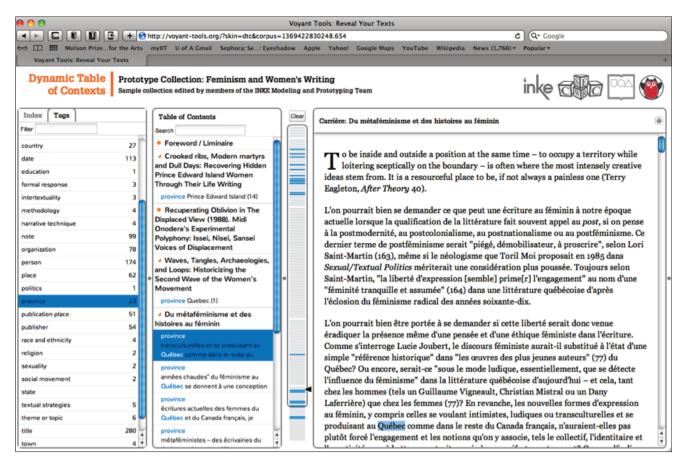
Although indexes have become less important in digital texts, which can be searched easily, it is still recognized that they serve a role distinct from both string searching and semantic encoding. The former will not identify every place where a relevant concept has been discussed without using the particular words used for searching. The latter is primarily of use where repeated instances of a higher-level concept or function category can be used to constrain a search to terms within that category. For example, someone might look for mentions of other titles generally in a document, or constrain that search to only titles that appear within an <intertextuality> element.

Scholarly and Research Communication

VOLUME 5 / ISSUE 4 / 2014

In terms of providing an affordance for visualization, we have used digital indexes as a way of extending the functionality of the Dynamic Table of Contexts (DToC) (see Figure 4), so that the reader is able to interactively add and subtract index items into the table of contents.

Figure 4: The Dynamic Table of Contexts allows readers to choose items from the semantic markup, the index, or free-text searches to add to the table of contents.



While some believe that free-text searches are adequate for most purposes, both semantic markup and digital indexes provide the benefit of bringing additional human intelligence to the text.

Textual variants

For canonical or otherwise privileged source texts, the record of textual variants has been of significant concern to the scholarly community for many centuries. The current industry standard for preserving knowledge of textual variants is the variorum edition, where the goal is to record within a single volume all of the most important variations of the text that have ever appeared in print. As the basis for these editions, it is becoming increasingly common to use a custom XML schema.

The prototyping project where we began experimenting with this schema was called the MultiTouch Variorum (MtV) (see Figure 5). Our goal was to produce an interface on a large touch surface that would allow real-time collaborative editing of a new edition, created by a small group of people working with all the previous editions. In

Figure 5: The MultiTouch Variorum provides a reading and editing system on a touch surface for multiple simultaneous users.

our scenario, they would be physically present at the same time and working together on the table. We believe that this is not the usual current method of scholarly editing, and so hoped to offer a new affordance.

One of the challenges of the MtV editing environment was to create a system that did not privilege any of the edges of the table, while nonetheless making the working materials available to everyone. A previous version showed a row of items that was duplicated across all the edges; the current iteration has a single set of items placed on a carousel in the middle.

Digital annotations

It has been widely recognized in the DH community that annotations have an important role to play historically and are gaining increased importance in the growing availability of social editions (e.g., Marshall & Bernheim Brush, 2004). One of the most common contemporary forms of digital annotation is in the comment threads associated with various primary source materials, such as a video, blog post, or tweet. When they are serving as a specialized form of conversation, annotations can provide the basis for visualizations that attempt to provide a structure for recalling, discussing, and analyzing conversational patterns and topics. Alternatively, annotations may take a more scholarly form (see Figure 6), where their purpose is to add information to a digital object.

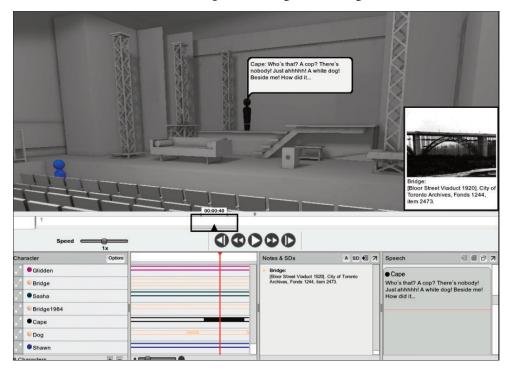
For example, in the Simulated Environment for Theatre (SET) project, we developed a number of ways of supporting people interested in watching stylized versions of plays. The ability to turn on annotations, whether in the form of text, images, or video, significantly enriches the system from the perspective of a theatre historian interested in examining the stage model, character blocking, and other features in a context that also makes scholarly annotations readily available.

VOLUME 5 / ISSUE 4 / 2014

Scholarly and Research Communication

VOLUME 5 / ISSUE 4 / 2014

Figure 6: This prototype shows the Simulated Environment for Theatre with an annotation open on the right of the stage.



Regularized citations

From the beginning of INKE, when our topic was interdisciplinary citation, we have been experimenting with the productive visualization of citations. Our first prototype, called the Paper Drill (see Figure 7), was intended to allow researchers to begin with a seed article, and use the system to generate a report of which authors were most frequently appearing, following a chain of citations originating from that article. Citation chaining of this kind is a venerable approach to information identification, surpassed only in the last few decades by digital searching of keywords.

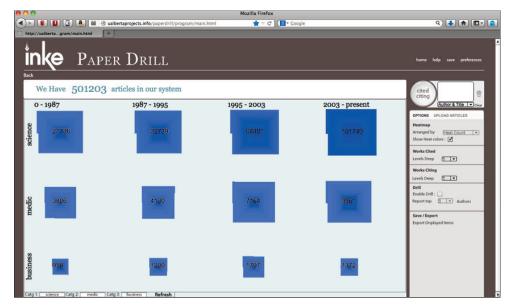


Figure 7: The Paper Drill homepage.

The main difficulty with the Paper Drill prototype is that it requires bibliographic information for every citation in every paper, down to the level of detail of article title and author names. Moreover, it will work best if the citation metadata has been regularized in an intelligent way, so the system can recognize, for example, that Bill Buxton, William Buxton, and William Arthur Stewart Buxton are all the same author, but William J. Buxton is a different author working in a different field entirely.

The current state of the art of citation metadata in the humanities, however, is that it largely does not exist. Citations are typically embedded in PDFs. We therefore spent some time in INKE looking at ways of parsing PDF files in order to get citation lists, but the results were mixed at best. In the end, we opted for testing using regularized citation metadata that was readily obtainable for papers indexed by the ISI (now Thomson Reuters) Web of Science.

There are a number of possible reasons for the lack of citation information in the humanities. One is that it requires time, effort, and perhaps even some text-mining capabilities. Second is that, unlike in the sciences, the humanities tend not to place much, if any, weight on citation indexes, since citations are used differently in the humanities than in the sciences, where there is a greater emphasis on identifying recent key papers. Furthermore, a list of top citations in the humanities will almost always begin with names of "primary source" authors such as Shakespeare and Milton.

Our argument for the development of citation metadata is to help students and researchers who may be new to an area to identify central articles and authors. Unfortunately, it is difficult to estimate how much time is spent in citation chaining by scholars each year. We therefore extended our work into further potential uses of citation metadata.

Figure 8: CiteLens is a multitouch application that allows readers to explore several elements of citations that are not typically available, such as how the author has used them in constructing an argument.

			Mozilla Firefox ★ T C S + Google	
	labs.fluxo.art.br/Cit	telens/	Mark Coogie	۹ 🖡 👘
http://labs.fluxo.art.br/CiteLens/				
cite lens	I would like to compare	ontextualize) citations of The Ca	re of the Dead in Late Antiquity. Éric Rebillard. 2009.	
march	Add comparison set	Add comparison set +	approaches religion on the basis of beliefs. By adopting a point of view based instead on social pra	ctice, we shift the emphasis to the
sort by: author 100 date	language	language	development of religious pluralism. ² Without attempting to go back to the origins of this phenomeno	enomenon, we can say with confidence that in ews, Christians, and pagars competing with onopolistic position of civic religion. Justing the impact of these groups in traditional
1 Delehaye, Hippolyte. Les Origines du cuite des martyrs. 1933.	DE EN ES FR HE IT LA	DE EN ES FR HE IT LA	the third century the Roman Empire was a true "marketplace of religions." Not only were Jews, Chri each other but, within paganism, a plurality of religious groups appeared, weakening the monopolist	
1 Lieu, Judith, North, John, Rajak,	country	country		
Tessa. The Jews among Pagans and Christians in the Roman	*** AUT BEL DEU	+++ AUT BEL DEU	From this standpoint, John North has noted that one of the most sensitive criteria for evaluating the	
World. 1992.	FRA GBR ISR ITA	FRA GBR ISR ITA	Greco-Roman society involves tracing areas of conflict with members' families. ³ Here, the study of fune	
1 North, John. The Development of	NLD USA VAT	NLD USA VAT	studies regarding burtal inscriptions of civil populations in the Western Roman Empire have shown t mentioned, 80 percent of commemorators were wives, parents, children, or cousins of the decease	Contraction of the second states of the second stat
Religious Pluralism.1992.	publication type	publication type	Interfactories, so percent of control of the fourth century, but the samples used for Late Antiguty were all Christian and therefore less representative. Even if commemoration with an epitaph did not extend to every level of Roman society, the numbers allow us to conclude that it was traditional in the Roman Empire for the family, in fact the nuclear family, to maintain the tombs of its members. Did the appearance of n cults and subsequent religious groups lead to tensions between a family and a religious group over the choice of a grave? Mystery Cults, Oriental Cults, and New Cults	
2 Saller, Richard, Shaw, Brent D. Tombstones and Roman Family	BSEC BOOK CODX	BSEC BOOK CODX		
Relations in the Principate:	COMM CONF CORP DISS EXCV JART	COMM CONF CORP DISS EXCV JART		members. Did the appearance of new
Civilians, Soldiers and Slaves. 1984.	DISS EXCV JART MNSC OTHR REVW	MNSC OTHR REVW		the choice of a grave?
1 Shaw, Brent D., Latin Funerary	SCHE	SCHE		
Epigraphy and Family Life in the Later Roman Empire1984.	period	period		
1 Martin, Dale B., The Construction	year to year o	year to year o	The success of the oriental cutts in the Roman Empire has often been viewed as preparation for the rise of Christianity, Because of the common origin in the East, of the mystery surrounding their rites, and of the initiation that separated their members from the rest of so the comparison was appealing. The cutts and their beliefs, as well as their organization, have long been analyzed through the reference system of Christianity. Water Burkert, who constantly emphasizes the discontinuity between oriental cutts and Christianity, has thus attacked the causal relationship between eschatological beliefs and collective burkial grounds postulated byFranz Curronitat the beginn the twentieth century. ⁴	ise of Christianity. Because of their
of the Ancient Family:	function	function		
Methodological Considerations. 1996.	Primary Source	Primary Source		
1 Perry, Jonathan S. A Death in	Secondary Source	Secondary Source		
the 'Familia': The Funerary	Further Reading	Further Reading		
Colleges of the Roman Empire. 1996.	author	author		
2 Burkert, Walter. Ancient Mystery Cults. 1987.	author name	(juftocrame) O	1. Hippolyte Delehaye, Les Origines du culte des martyrs, Subsidia hagiographica 20(Brussels: Société der	EALTHING SANDARY NO. 2019 (ANALY SANDARY)
1 Curnont, Franz, Les mistères de			 This is the point of view adopted in The Jews among Pagans and Christians in the Roman World, ed. Juc (London: Routledge, 1992). United Kingdom 	Ith Lieu, John North, and Tessa Rajak

Ruecker, Stan, Sinclair, Stéfan, Dobson, Teresa, Rockwell, Geoffrey, Radzikowska, Milena, & INKE Research Group. (2014). The Provision of Digital Apparatus for Use in Experimental Interfaces. *Scholarly and Research Communication*, 5(4): 0401194, 11 pp. Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

Scholarly and Research Communication VOLUME 5 / ISSUE 4 / 2014

Taxonomies of references

The prototype that relies on semantic encoding in references is called CiteLens (see Figure 8). It is a system intended to help scholars examine the use of citations, primarily not between academic papers, but instead within the context of a single monograph. We believed that if it were possible to produce the right combination of tags, we could develop an environment where the reader could look at the references from other perspectives beyond the alphabetical listing by first author's last name.

In addition to the standardized citation metadata needed by the Paper Drill, CiteLens also leverages more semantic tagging, specifically involving the context of use of a citation and its intended purpose as a form of evidence for the argument being made. That is, citations play a rhetorical role in scholarly articles, and CiteLens is a tool to help people trace that rhetorical trajectory.

Conclusion

Some metadata is produced during the creation of a resource, some is added after creation, some is provided manually, and some can be automatically generated. The least expensive of these in terms of production is metadata that is automatically generated at the time of creation. Currently we can expect, for instance, that our computers will keep track of the date that we first created a file and the date that it was last modified. It can do a pretty good job of recognizing the file type, since we specify those in the file extensions that are typically attached by the creating software.

It is not hard to envision, however, a time when more metadata could be automatically associated with a file, and in particular with the kind of files we are typically working with in projects like INKE: text files. It could perhaps also be associated with other media, but primarily text. We might, for example, expect to have parts of speech attached to every word. Named entity recognition could become standardized to the point that any name mentioned would be associated with information about the person or place. Further, we might find contextualizing information, such as how common a particular name was in the year that our named entity was born or created.

Each addition of this kind provides the opportunity to design new affordances for people working with digital text. The benefit of doing this on a large scale is that automation and standardization can come to play a role, hopefully making the choice of experimental dataset more interesting and the data itself more robust.

Websites

Bubblelines, Montreal, QC, http://docs.voyant-tools.org/tools/bubblelines/
Canadian Writing Research Collaboratory, Edmonton, AB, http://www.cwrc.ca/en/
CiteLens, Edmonton, AB, http://www.artsrn.ualberta.ca/inke/projects/citelens/
Dynamic Table of Contexts, Montreal, QC, http://voyant-tools.org/?skin=dtc&corpus=139153265 8789.4848&docId=d1391532101107.od5601c8-bb2d-4bbb-179e-d4769947167a&inkeTags=true
Mandala Browser, Montreal, QC, http://mandala.humviz.org/
Multitouch Variorum, Edmonton, AB, http://vimeo.com/70527973
Simulated Environment for Theatre, Montreal, QC, http://humviz.org/set/
Text Analysis Portal for Research, Edmonton, AB, http://www.tapor.ca/docs?name=about_tapor
TextTiles. London, UK, http://dev.giacometti.me/textTiles/trunk/

References

Brown, J.D. (1898). *Manual of library classification and shelf arrangement*. London: Library Supply Company. https://archive.org/details/manualoflibrarycoobrowrich

Hockey, S. (2000). Electronic texts in the humanities. Oxford, UK: Oxford University Press.

- Marshall, C.C., & Bernheim Brush, A.J. (2004). Exploring the relationship between personal and public annotations (pp. 349-357). New York, NY: ACM Press.
- Orlando Project (2006). *The Orlando project: An online history of women's writing in the British Isles*. Cambridge, UK: Cambridge University Press.
- Ruecker, S., Radzikowska, M., & Sinclair, S. (2011). *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*. Farnham, UK: Ashgate Publishing.

Scholarly and Research Communication

VOLUME 5 / ISSUE 4 / 2014