

Historical Interpretation through Multiple Markup: The Case of Horatio Nelson Taft's Diary, 1861-62

Garfinkel, Susan

sgarfinkel@loc.gov

Library of Congress, Washington DC, USA

Heckscher, Jurretta Jordan

jhec@loc.gov

Library of Congress, Washington DC, USA

1. Background

Now fifteen years old, the still-growing American Memory Web site at the Library of Congress (<http://memory.loc.gov/>) offers more than 130 separate and diverse multimedia collections — comprising more than a million digitized library items — to a vast base of virtual patrons across the Internet. Yet despite radical changes in the ways that people have come to use the World Wide Web, the fundamental conception of American Memory and its collections remains much as envisaged fifteen years ago. What happens next to such established but largely static digital resources? Aside from implementing obvious upgrades such as higher-resolution image scans, cleaned-up OCR, fleshed-out metadata, or faceted search capabilities, how might cultural repositories more fundamentally enhance their existing online content? Can we fully (re)imagine a second iteration, a next generation, for already digitized historical materials?

Our demonstration project works to consider this question for a single American Memory collection, purposely setting aside issues of scale or interoperability in favor of exploring effective and compelling ways to convey to users the particular character of specific historical resources. Starting with the unpublished manuscript diary of Horatio Nelson Taft (1861-62), already digitized (as textual transcription with page images) at <http://memory.loc.gov/ammem/tafthtml/tafthome.html>, we have chosen to explore the implications of creating several alternate and

explicitly interpretive frameworks by applying multiple XML markup to this bounded but compellingly dense and historically significant text.

American Memory patrons bring unusually diverse research needs to the same body of historical materials: typical users include everyone from elementary school students and hobbyists to lawyers, librarians, college professors and members of Congress. Our project's concerns stem from this diversity as well as from our own role as scholars of American culture working with historical materials in a library setting. Traditional library practice treats sources as analytically separate from their interpretation, even while the most basic interventions of librarianship are fundamentally interpretive. Historians, by contrast, well recognize their own work with sources as interpretive, but still tend to view textual sources as fully fixed in meaning.

Textual analysis is an underdeveloped tool in American historical study, and history as a discipline has lagged behind literature in imagining computing as a threshold for innovative research. Historians' traditional use of computing has trended quantitative rather than qualitative, while recent innovations emphasize creating tools to assist the mechanics of research or scholarly interaction rather than transforming them. Explorations of how computing might fundamentally change the practices of history or the outcomes of historical interpretation are still needed.

Against this background the Taft Diary project takes on several related goals: (1) to visibly demonstrate a historical text's accessibility to multiple simultaneous interrogations enacted through digital scholarship; (2) to more fully explore the multidimensionality of a significant historical text; (3) to introduce the methods and questions of digital humanists more centrally into a library context (making them better known to library practitioners and more widely available to diverse library audiences); and (4) to meld the questions and methods of historians with the advances in digital textual scholarship arising among literary scholars. With our project we hope to establish a model for text-based scholarship — literary and ethnographic as well as historical — that foregrounds the individual user's interpretative needs while also conveying

that interpretation to the broad variety of the diary's potential users.

2. The artifact and its text

The Horatio Nelson Taft Diary commands wide historical interest. When it was introduced to the public on the Library's Web site in 2001, it offered the first new information about the events of Abraham Lincoln's death to come to light in half a century. In fact the diary emerges from the confluence of multiple transformative historical developments: not only the American Civil War and Lincoln's presidency, but profound long-term changes including the rise of the American middle class; the nation's industrial, technological, and consumer revolutions; the decisive expansion in the size and influence of the federal government; and the maturation of Washington, DC, as a complex urban community of national and international importance.

An exceptional historical record on many levels, the Taft Diary is also unusually well suited to an experiment in simultaneous multidimensional interpretive markup. Its contents are rich in the overlap of ordinary life and significant events, people, and places, thereby appealing both to specialized historians and to a broad general audience. Its limited size offers definite boundaries to the range, though not the depth, of interpretive markup. Moreover, the formulaic pattern underlying most of its daily entries invites ready comparison while ensuring an organizing degree of structural regularity.

Taft's three-volume manuscript diary consists of daily entries from January 1, 1861, through April 11, 1862 (the end of vol. 1), and less frequent entries through May 30, 1865. The current phase of our project deals only with vol. 1, which contains in total 466 entries. Because Taft used a printed blank diary book for vol. 1, each 1861 entry is eleven handwritten lines long, or typically ninety to one hundred words (fig. 1). (The volume's 1862 entries, written into pages intended for back matter, vary more in length.) Beyond the consistency of the entries' length, certain content elements recur so frequently as to be almost formulaic: information about the weather, illnesses within and beyond the family circle, political and military news about the Civil

War, and the people and locations that populate Taft's daily life.

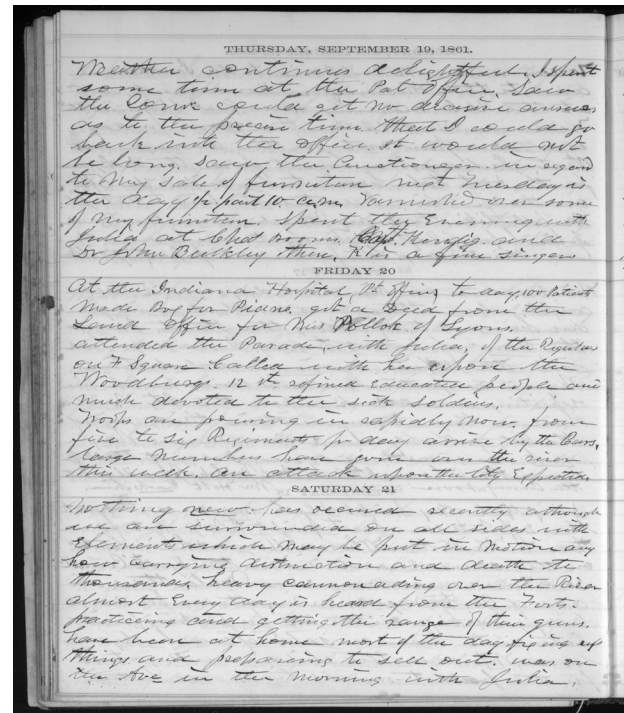


Figure 1: Taft Diary vol. 1, entries for Sept. 19-21, 1861; transcription of Sept. 20 entry

3. Implementation

In selecting markup schemas, we considered and discarded a number of interpretive typologies, including Library of Congress Subject Headings (LCSH) (<http://www.loc.gov/aba/cataloging/subject/>) and the Library of Congress Classification Outline (<http://www.loc.gov/catdir/cpsolcco/>), because they turn out to lack a consistent level of granularity across cultural and historical domains. Seeking to demonstrate applicability and variety, for phase one we selected three schemas for their breadth as well as depth, their flexibility, and their inherent relevance to the content of our text:

1. TEI, <http://www.tei-c.org/>
2. Outline of Cultural Materials (OCM), <http://www.yale.edu/hraf/outline.htm>
3. Scholar-defined "webs of significance" such as:
 - i. Taft's recurring concerns
 - a. weather
 - b. places

- c. illness
- ii. networks
 - a. of social relationships
 - b. how news is conveyed

These three schemas together provide for theoretical and well as methodological diversity, representing a variety of approaches to the analysis of historical texts. TEI is a widely used standard for the representation of texts in digital form, with well-developed editorial standards and an established community of users. The OCM is anthropological, deriving from attempts to develop comprehensive categorizations of human cultural phenomena. Scholar-defined “webs of significance” follow from our own close readings of the text. In the future we anticipate exploring additional markup typologie based in the Thesaurus for Graphic Materials (<http://www.loc.gov/pictures/collection/tgm/>), GIS, and data visualizations, so as to demonstrate the diversity and versatility of interpretation that simultaneous multiple markup can sustain.

Our baseline text for markup is the transcription of Taft’s vol. 1, which was originally provided in SGML mapped to the American Memory DTD. Our own markup reverses normalizations in transcription in favor of a diplomatic version, and strips out the older SGML tags.

Implementation raises methodological and technical challenges. In our process we must move from theoretical issues to the development of standards for each tag category, exploring the dilemmas that intensive and self-created markup entails: for example, whether phrases or sentences should be the unit of markup with OCM tags (fig. 2). Resolving such questions requires us to confront the practical challenges of creating markup conventions originally based in theory rather than established practice. Theoretical language about multiplicity is inspiring — and even starts to get at the truth of lived experience — but in practice we must also make sound choices about how to complete the markup with reasonable consistency so that others may use and rely on it. Further, do we mark the same text multiple times, or invest ourselves in some version of a standoff markup (see the XStandoff toolkit, for example)? How do we best present multiple interpretive options to

our audience? At this level, the Taft Diary project is still a work in progress.

```

FRIDAY 20 [September 20, 1861]
758                                     423
727                                     425
476 743                               335 534 676 675
<At the <Indiana Hospital (PI office)> today, 100 Patients>. <Made Box> for <Piano>. <Got a <Deed from the

647                                     131 186 541 593 704 367
<Land Office> for Mr Pollok of <Lyons>. <Attended the <Parade> with <Julia> of the <Regulars> on <F[ranklin]

574                                     363 554 708
Square>. <Called with her upon the Woodburys>. <12th St>. <refined educated people> and <much

758                                     704
727                                     752
devoted to the <sick soldiers>. <Troops are pouring in rapidly now, from five to six <Regiments> pr day

493                                     505? 805 726
<arrive by the Cars>. Large numbers have gone <over the river> <this week.> <An attack on the City

expected.>

key:
131 Location 555 Talent Mobility
186 Cultural Identity and Pride 574 Visiting and Hospitality
335 Carpentry 593 Family Relationships
363 Streets and Traffic 647 Administrative Agencies
367 Parks 675 Contracts
423 Real Property 676 Agency
425 Acquisition and Relinquishment of Property 704 Ground Combat Forces
476 Mutual Aid 708 Auxiliary Corps
484 Travel 726 Warfare
493 Vehicles 727 Aftermath of Combat
505 Water Transport 743 Hospitals and Clinics
534 Musical Instruments 752 Bodily Injuries
541 Spectacles 758 Medical Care
554 Status, Role and Prestige 805 Ordering of Time

```

Figure 2: Text of Taft Diary vol. 1, Sept. 20 entry showing conceptual phrase-level OCM markup

4. Conclusion

We anticipate that simultaneous multiple markup will render the text dynamic, surfacing its performative and ritualized aspects; and that it will invite new attention to its literary and linguistic aspects. Markup can also foreground key aspects of worldview of which the writer himself was unaware. In these and other ways, simultaneous multiple markup becomes a new instrument of historical interpretation, reimagining analysis beyond the realm of narrative prose.

While our technical questions are not yet fully answered, and the project itself is ongoing, we seek to share important methodological and theoretical questions, and our conclusions so far, with our colleagues in digital humanities.

References

- Cohen, Daniel J., Rosenzweig, Roy** (2006). *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press. <http://chnm.gmu.edu/digitalhistory/>.

Geertz, Clifford (1973). 'Thick Description'. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books, pp. 3-30.

McGann, Jerome (2004). 'Marking Texts of Many Dimensions'. *A Companion to Digital Humanities*. Schreibman, Susan, Siemens, Ray, Unsworth, John (eds.). Oxford: Blackwell, pp. 198-217. <http://www.digitalhumanities.org/companion/>.

Milligan, Frank D. (2007). 'A City in Crisis: The Wartime Diaries of Horatio Nelson Taft'. *President Lincoln's Cottage. entry of August 1*, blog. <http://lincolncottage.wordpress.com/2007/08/01/a-city-in-crisis-the-wartime-diaries-of-horatio-nelson-taft/>.

Sellers, John (February 2002). 'Washington in Crisis, 1861-1865: Library Acquires the Diary of Horatio Nelson Taft'. *Library of Congress Information Bulletin* 61, no. 2. <http://www.loc.gov/loc/lcib/0202/cw-diarist.html>.

Stührenberg, Maik, Jettka, Daniel (2009). 'A Toolkit for Multi-Dimensional Markup: The Development of SGF to XStandoff'. *Proceedings of Balisage: The Markup Conference 2009*. Balisage Series on Markup Technologies. 3 vols. <http://www.balisage.net/Proceedings/vol13/html/Stuhrenberg01/BalisageVol13-Stuhrenberg01.html>.