
Renaissance and Reformation Renaissance et Réforme



Mueller, Martin, and Bill Parod, project leads. *WordHoard*.
App.

Michael Ullyot

Volume 41, Number 4, Fall 2018

URI: <https://id.erudit.org/iderudit/1061922ar>

DOI: <https://doi.org/10.7202/1061922ar>

[See table of contents](#)

Publisher(s)

Iter Press

ISSN

0034-429X (print)

2293-7374 (digital)

[Explore this journal](#)

Cite this review

Ullyot, M. (2018). Review of [Mueller, Martin, and Bill Parod, project leads. *WordHoard*. App.] *Renaissance and Reformation / Renaissance et Réforme*, 41(4), 186–188. <https://doi.org/10.7202/1061922ar>

All Rights Reserved © Canadian Society for Renaissance Studies / Société canadienne d'études de la Renaissance, Pacific Northwest Renaissance Society, Toronto Renaissance and Reformation Colloquium and Victoria University Centre for Renaissance and Reformation Studies, 2019

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

Mueller, Martin, and Bill Parod, project leads.***WordHoard*. App.**

Evanston, IL: Northwestern University, 2004. Accessed 9 February 2018. wordhoard.northwestern.edu/userman/index.html.

If digital philology needs a founding father, my candidate is Martin Mueller. An Emeritus Professor of English and Classics at Northwestern University, Mueller has done more than anyone else to promote computational research on the language of classical, medieval, and early modern texts. *WordHoard*, a Java application he released in 2004 “for the close reading and scholarly analysis of deeply tagged texts” (“Index”), offers these tools to “scholarly but non-technical users [to] explore the [...] query potential of textual data” (“What is *WordHoard*?”). And despite its lack of updates since 2013, the program remains viable in 2018.

Words are the core of *WordHoard*: namely, the works of Homer, Hesiod, and the Homeric Hymns (in Greek and in English and/or German), Chaucer, Spenser, and Shakespeare. Every word of these texts has been tagged with its morphological data (lemma, part of speech, word class, major word class, and spelling); its narrative data (whether words are narrated or spoken; and for some texts, the speaker’s name, gender, and whether they are mortal or supernatural); its form (in prose or in verse); its metrical shape (for Greek texts); its date of publication; and its definition (for Chaucer’s texts).

All of this word-level metadata is as useful as the researcher needs it to be, but *WordHoard*’s “Analysis” menu makes it straightforward to do processes like comparing frequencies, comparing texts, and tracking forms over time. These preset tools are one of *WordHoard*’s greatest strengths. Because they have a gentle learning curve, they make many of the capabilities of natural-language processing available to novice users.

For instance, if you want to compare the vocabulary of a single Shakespeare play with that of his entire corpus, it is a simple matter of selecting a menu item and ticking a few boxes to generate a table that reveals which words (or rather, lemmas) he uses most or least frequently in a given play, compared to the reference population of his complete works. This “log-likelihood analysis” reveals far more than lists of lemmas most frequently used in a text; it reveals which lemmas are used more and less often *than would be expected* in that text, compared to that reference: e.g., one play compared to all of his other plays. You

can just as easily compare his histories with his comedies, or his plays with his sonnets, or his male to his female speakers, or his texts published before 1603 to after. If these preset divisions are insufficient for the user's inquiries, they can create a "work set" for further analysis: say, to compare parts of a text with other parts; or everything published by Spenser and Shakespeare in 1590 with 1591; or they can create a "word set" to compare, say, words spoken in Shakespeare's histories in prose with those spoken in comedies in prose.

Yet one need not have a research question in mind, or deform the texts into statistical tables, to find *WordHoard* valuable and interesting. Users who simply want to analyze texts by reading and exploring queries as they arise, will find a customizable interface that allows them to move nimbly through texts. Information about each word—e.g., its part of speech, its lemma, its frequency, and its variant forms—is a keystroke away. Keyword-in-context (KWIC) searches are a familiar and powerful capability; but *WordHoard* enables users to search the four text corpora using twenty criteria, each with its own controls. For instance, you can list all the adverbs spoken by mortals in Spenser's *Faerie Queene*, or all the occurrences of words in Chaucer that begin with the letter "z". Results are displayed in a concordance window that groups KWIC lines by work, date, or other customizable criteria; the bottom half of the window displays the lines in question. The result is an elegantly designed, flexible workbench capable of displaying a reading copy of a work (the analysis text) alongside this concordance window (compiling the reference texts).

It's also worth mentioning that *WordHoard* is open to developers—not through an API, but by offering up all of its source code, its raw data files, its MySQL binary database files, and other files. They are not in a GitHub repository, but are downloadable with a little work. Comprehensive documentation is available, and support is an e-mail away.

A final caveat: *WordHoard* was released before the Folger Digital Texts edition of Shakespeare was published, so it is based instead on the 1866 *Globe Shakespeare*, the best machine-readable text available in 2003. This is hardly a fair criticism; in a few years the electronic text of the *Oxford Edition of the Collected Works of Edmund Spenser* will render *WordHoard's* use of the 1932–49 Johns Hopkins edition of Spenser obsolete, too.³ No doubt someday *WordHoard's* use of the 1987 *Riverside Chaucer* (3rd edition) will be superseded. I am less

3. Patrick Cheney, "The Oxford Spenser Edition: An Update," *Spenser Review* 48.1.5 (2018).

qualified to judge the classical editions and translations, whose dates range from 1793 to 2005. Ultimately, my caveat is an endorsement of the project's future. Because scholars need *WordHoard* to serve us well beyond its fifteenth birthday, it will need to revisit and periodically update its editions.

In a footnote in 2008, Mueller quotes a response from Harold Bloom to Mueller's letter announcing *WordHoard*: "I am a throwback and rely entirely on memory in all my teaching and writing." The rest of us will have to rely on what Mueller calls "digital surrogates" to see features of these texts "that you could not discover by just reading [them]" ("Metadata"). Those who are humbler than Bloom are all the wiser.

MICHAEL ULLYOT

University of Calgary

Stringer, Gary A., gen. ed.

DigitalDonne: The Online Variorum, vol. 6.

Greenville, NC: East Carolina University, 1995. Accessed 31 January 2018. donnevariorum.tamu.edu/.

The text on the first page of this major digital resource for studying Donne—and for studying early modern poetry and manuscript culture more generally—begins "*DigitalDonne* constitutes volume 1 of *The Variorum Edition of the Poetry of John Donne*." This defines quite exactly what *DigitalDonne* seeks to do, if you know the *Donne Variorum* well. But its significance will not be evident otherwise, and many readers may find this resource odd in not being focused on Donne's writing in the way they might expect; it is certainly not intended as an online edition of Donne's poetry. Though the volume 1 program is most significant as a component of the *Donne Variorum*, this resource is still important on its own; at this point in its development, it is also doing something rather different from the original plan for that volume, something that makes very productive use of the digital medium to extend that plan.

Volume 1 of the *Donne Variorum* was always described (in the volumes published to date, somewhat more than half of the projected total) as an introduction, with no texts, and two major parts: a general commentary on Donne reception and a general textual introduction. The other volumes in the