A Larger View of Digital American Studies

RYAN CORDELL

By abandoning our conception of the computer as merely a mechanical clerk suited mostly to repetitive routine operations, by learning to know its features, uses, limitations, and possibilities—its nature, in short—we shall be properly re-organizing our thinking for the new age. What the computer will enable us to do in our humanistic tasks has hardly been imagined yet. Even immoderate speculation tends to fall behind the new reality.


The digital humanities are large; they contain multitudes: a rewriting of Whitman that can read equally as commendation or condemnation. To cite a specific example, it might surprise that scholars who devote their work to the textually minute processes of editing and encoding digital scholarly editions rightly consider themselves members of the same field as scholars who develop algorithms for classifying data across millions of works. Yet both of these things are digital humanities (DH).

The most commonly cited genealogy of the DH field—both in the popular press and in scholarship, such as Dunst’s “Digital American Studies” article in this issue—begins with Father Roberto Busa’s indexes and moves through Humanities Computing, following an inexorable path toward Franco Moretti, distant reading, and macroanalysis. This chronology conflates “digital humanities” with particular kinds of DH research that happen largely in literary studies, leaving digital history, archeology, classics, art history, religious studies, and the many other fields that contribute to DH out of the narrative. Such genealogies also overlook conversations happening with the DH community about its diverse genealogies. As Adeline Koh argues,

We need to cast our formulation of the history of the digital humanities beyond the field of humanities computing to incorporate into its intellectual genealogy such fields as new media studies, diy (do-it-yourself) digital recovery projects from the 1990s, digital projects on postcolonial studies, and work taking place outside of the provenance of the United States and the United Kingdom.

Koh further proposes that these histories should include also “postcolonial and feminist science and technology studies [...] work on Afrofuturism and ‘and other varied scholarship on race and the Internet’” (102). Other scholars have argued forcefully for histories of DH that locate its genealogies in design or multimodal composition. Meanwhile, Tom Scheinfeldt writes digital public history as “a story that begins with people like Allan Nevins of the Columbia Oral History Office and Alan Lomax of the Library of Congress’s Archive of American Folk-Song, especially with the man on the street interviews Lomax coordinated in the aftermath of the Pearl Harbor attacks.” Such revised genealogies of the field offer important historical and theoretical contexts for DH research and teaching in the present moment, and in particular
promise a more robust foundation for digital American Studies. Any genealogy of DH and American Studies must note that #transformDH, a movement envisioning “a digital humanities that will center on the intersection of digital production and social transformation through research, pedagogy, and activism,” coalesced at the American Studies Association conference in 2011 (Bailey et al).

Considering other origin stories for DH also helps us avoid conflating “big data” or “distant reading” with the field entire, or of positing corpus-level text analysis as the primary activity of the field rather than one activity of the field. Such a view risks excluding DH research in encoding, archive building, mapping, sound studies, and digital publication, to name only a few branches of DH less prominently featured in journalistic or scholarly rumination on the field. My own DH research began encoding digital scholarly editions and drifted toward computational text analysis, a shift accompanied by a growing awareness of how frequently the latter is taken as a metonymy for the field. In the remainder of these brief remarks, I too will focus on computational text analysis, but with the aim of offering a more nuanced view of its role in the DH field and its potential within American Studies.

When computational text analysis is taken as the center of DH, its primary intervention is often framed as introducing empiricism into humanistic disciplines, allowing hypotheses to be offered and then either validated or falsified. Dunst summarizes this view, “For the first time, quantitative but also other empirical approaches allow for transparent validation or falsification, based on representative corpora and established tools that can be disclosed and shared…the results of empirical testing are better described as findings: what DH scholars initially observe are statistical patterns that must be explained to produce insights.” Certainly some scholars claim, explicitly or implicitly, that computation enables a new relationship to evidence and proof for humanities scholars. Consider Moretti’s praise of “operationalizing,” which in his account describes the process whereby concepts are transformed into a series of operations—which, in their turn, allow to measure all sorts of objects. Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts.

For Moretti, this principle of operationalizing helps demonstrate “how the unprecedented empirical power of digital tools and archives offers a unique chance to rethink the categories of literary study” (1, 13). A similar impulse could be traced in David L. Hoover’s claim that “Literary criticism’s problematic relationship to facts, claims, and evidence seems more like a bug than a feature,” and his work to identify a subset of literary questions that are “tractable,” “testable,” and “solvable” using computational methods (online). We might also consider Matthew L. Jockers’ recent posts about his Syuzhet Package for the R programming language, which attempts to be a “systematic way of extracting plot arcs from fiction.” This software and Jockers’ use of it to extract “six, or possibly seven, archetypal plot shapes” led to a heated debate among computational humanities scholars about technical topics such as Fourier transformations and low pass filters.¹

¹ The debate over Syuzhet unfolded largely through research blogs and social media and thus is difficult to cite as an event. My quotation comes from Matthew L. Jockers, “The Rest of
I cite these examples not to vilify these scholars or their work. By positing that humanistic questions might advance through hypotheses, testing, and validation, such research offers provocations for humanists resting perhaps too comfortably on implicit ideas about the boundaries of inquiry in their fields. Whether such work represents true empiricism or some kind of positivist scientism depends on whose account one trusts. As Stephen Ramsay notes, “There are many who think scatter plots filled with data points drawn from, say, English novels, are already a crime against the humanities […] For them, the problem is positivism in its properly technical sense. They fear an epistemology that does not merely value empirical data, but which (in its extreme philosophical forms) considers empirical data to be the only valid form of evidence.” With Ramsay, I would argue “these fears are not so much unwarranted as they are grossly overblown” (“Humane”).

The most vital critical voices in digital humanities today—and even within the subfield of computational text analysis—resist binary accounts of computation or its potential for humanistic research and promise to enliven theories of Digital American Studies. In a recent conference paper, Lauren Klein offers a compelling metaphor for an experimental, humanistic, even ludic mode of computational research that draws on the unique affordances of the digital medium without falling into a trap of scientism. Klein describes topic modeling—a statistical technique for identifying closely-associated words within a corpus—as “a technique that stirs the archive,” bringing unexpected patterns and connections into view. Klein reminds us that “for the model to be truly meaningful, domain experts […] must be able to probe the semantic associations that the model proposes, and seek out additional perspectives on the model, as well as on the archive itself” (“Carework”).

Computational methods can reorganize the archive in new and multiple ways, offering humanists novel perspectives on its contents and their relationships. “Big data” is not a product of digitization, though digitization perhaps makes the vastness of the archive more apparent. Digitization can enable scholars to sift through the archive differently than we have in the past, to surface features of which we were unaware, or to test ideas across wider sections of primary texts. In “The Image of Absence: Archival Silence, Data Visualization, and James Hemings,” Klein models such an approach, using network visualizations drawn from the Papers of Thomas Jefferson to read against the erasures of that same archive. By visualizing Jefferson’s many close connections to enslaved people he mentions but does not name in his letters, Klein “conjures a sense of the dependence, on the part of Jefferson, on the men and women he enslaved, even as it cannot recreate what these people said in their conversations, where they went in order to conduct their transactions, and how they truly lived their everyday lives” (673–74). Klein’s research moves actively between patterns across her corpus and close textual details, while her computational work pressures rather than reifies the biases of her source archive. We can note a similar impulse in mapping

the Story” (25 February 2015), http://www.matthewjockers.net/2015/02/25/the-rest-of-the-story/. Fortunately, Eileen Clancy compiled a Storify that usefully outlines the larger discussion of which this post is only one part. That Storify can be found at https://storify.com/clancynewyork/contretemps-a-syuzhet.
projects such as the multi-institutional project “Mapping Inequality: Redlining in New Deal America,” which uses “security maps,” produced by the Home Owners’ Loan Corporation between 1935-1940, to illustrate the ways “federal housing programs helped codify and expand practices of racial and class segregation” in the United States (Mapping).

Another way in which computational tools stir the archive is by aiding scholars who wish to reconstruct earlier material phenomena that leave traces, often quite subtle, in their digitized forms. The Photogrammar project at Yale University, for instance, offers a range of novel methods for organizing, searching, and visualizing the “170,000 photographs from 1935 to 1945 created by the United States Farm Security Administration and Office of War Information” (Wexler). Photogrammar reimagines this photography archive in a number of generative ways, such as by mapping the locations of photographs, allowing scholars to imagine the collection spatially. A forthcoming interface will enable exploration based on similarities of hue, saturation, and lightness across the photographs: an aesthetic mode of interaction that draws from art and design rather than the typical hierarchies of catalogs but which is made possible by computational image analysis. Perhaps the most surprising insight of the Photogrammar project is their discovery that buried in the Library of Congress’ MARC records for the FSA-OWI photographs were sequences of letters and numbers that referred back to each photograph’s place in an original roll of film. Using these features, the Photogrammar researchers can computationally reconstruct the order in which strips of film were shot by FSA-OWI photographers in the 1930s and 40s (Tilton).

My own research with the Viral Text Project adopts an exploratory method of “bottom-up bibliography,” beginning from computational observations of text reuse in order to reconstruct bibliographies of popular nineteenth century newspaper literature. Our work operationalizes the question of reprinting, using the pattern-finding capabilities of the computer to identify groups of text strings—what computer scientists would call Ngrams, or sequences of words of N length—that closely align with other groups of text strings across our source archives. We seek to redress one of the key limitations of keyword search: that it only surfaces patterns scholars are already primed to look for and thus merely amplifies our presuppositions rather than challenging them. Keyword search can help scholars find copies of texts they know to search for—in other words, familiar, canonical works—but it cannot make inferences about widely reprinted texts lost to scholarly bibliographies. By attending to a formal feature of the digitized archive—overlapping strings of words that appear near each other in multiple newspaper issues—we identify reprints irregardless of topic, author, or genre, including previously unknown but popular texts and those that do not explicitly identify themselves as reprints. The aim of such modeling is to “stir the archive” of nineteenth-century newspapers and surface widely circulated and read texts that shaped nineteenth-century cultural conversations but escaped the slower processes of canon formation. For instance, the poems surfaced by Viral Texts are most often anonymous,

\footnote{For more about the Viral Texts Project, including project data and publications, see the research website at http://viraltexts.org.}
or penned by authors virtually unknown in anthologies of the period. Moreover, in our data these poems intermix with an incredible variety of popular writing scholars rarely address: e.g. recipes, advice columns, jokes, popular science, even listicles. The point of this work is not to construct a definitive, empirical solution to the problem of nineteenth-century newspaper reprinting, but to facilitate an iterative conversation between the large-scale, quantitative output generated by a corpus analysis algorithm and qualitative, literary-historical readings of the surprising texts that algorithm brings into focus.

Indeed, the most vital computational work both stirs the archive and interrogates the structures of computation, using seeming exceptions, mistakes, or false positives in the “codework” as intellectual pivots toward unexpected new insight. In Hoyt Long and Richard Jean So’s computational exploration of modernist haiku poetry, for example, their algorithm identified a small but significant number of poems as haiku that they, as domain experts, would not classify as such. Rather than simply dismiss these false positives, however, Long and So engage these anomalies critically and historically, asking what features they do share with haiku that led to their misclassification, and identifying the subtle, diffuse influence of an Orientalist poetic style across the work of a range of modernist poets. As Long and So argue, humanistic computational work “will require a method of reading that oscillates or pivots between human and machine interpretation, each providing feedback to the other in the critic’s effort to extract meaning from texts. Literary pattern recognition, then, brings together close reading, cultural history, and machine learning so that they supplement one another” (266-67). Here rigid operations of the algorithm reorganizes texts across the genre boundaries established by previous scholarship, pressuring conventional readings of individual poems but also exposing the logic of computational categorization. A seeming glitch in the software generates humanistic insight.

I join Dunst’s final exhortation for humanists—and Americanists in particular—to view computation as a cultural and technical domain that requires our participation. My understanding of the line between “humanistic argument and computational logic,” however, is less polarized than Marche’s or Hall’s (with whom Dunst ends his article). To begin, our colleagues in computer science and related disciplines are more interested in complex data and hard, theoretical questions than we typically countenance. Corporate computing is ubiquitous in daily life, but does not describe the boundaries of computer science research, which often seeks precisely challenges that seem intractable, such as those rooted in language. As computation waxes both politically and sociologically, we need genuine, interdisciplinary conversation and, dare I say, intellectual exchange between our fields. Next, the long history of media makes clear that narratives of new media replacement rarely capture the nuances of real historical change. As Alan Liu re-

---

3 An article about these “Fugitive Verses” is forthcoming in American Periodicals in 2017. A preprint of that article is available at http://viraltexts.org/2016/04/08/fugitive-verses/.

4 In general, I suspect most humanist have very little idea what computer scientists research (and vice versa), a quandary well outlined in Stephen Ramsay’s article “DH and CS,” http://stephenramsay.us/2013/04/30/dh-and-cs/.
minds us, “new media encounters are messy,” meaning “that right-angled historical, socio-political, or psychological distinctions between old and new media typically do not survive concrete acts of narration. Instead, binary distinctions open out into overlapping, contradictory, or otherwise thick affordances between media regimes” (online). Both apocalyptic pronouncements and millennialist announcements likely miss the reality of our new scholarly ecology, which can foster both ingenious computational-humanistic praxis and rich critique of digital culture. We find ourselves in a “thick, unpredictable zone of contact—more borderland than border line” between computation and humanistic inquiry (Liu). Through creative, theoretically-informed engagements with and through the digital, American Studies can help shape the narrative of our new media rather than being reshaped from outside. Like DH, a digitally-informed American Studies must be large; it must contain multitudes, if it is to thrive in the twenty-first century.

Works Cited


Moretti, Franco. “‘Operationalizing’: or, the Function of Measurement in Modern Literary Theory.” *Stanford Literary Lab Pamphlet* 6 (2013). Web.


