



Open Library of Humanities



Part of the Ubiquity
Partner Network

Digital Studies /
Le champ numérique

Research

How to Cite: Fan, Lai-Tze. 2018. "On the Value of Narratives in a Reflexive Digital Humanities." *Digital Studies/Le champ numérique* 8(1): 5, pp. 1–29, DOI: <https://doi.org/10.16995/dscn.285>

Published: 27 March 2018

Peer Review:

This is a peer-reviewed article in *Digital Studies/Le champ numérique*, a journal published by the Open Library of Humanities.

Copyright:

© 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Digital Studies/Le champ numérique is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

RESEARCH

On the Value of Narratives in a Reflexive Digital Humanities

Lai-Tze Fan

Lingnan University, HK

laitzefan@LN.edu.hk

This paper returns to the relationship of “narrative versus database” (an argument originally made by Lev Manovich in 2001) as one that can be further addressed. A specific issue persists in text analysis research in the digital humanities: the difficulty of representing the figurative meaning of narratives through digital tools. Towards an accommodation, this paper adopts a narratological framework in order to propose alternative models of content management and organization that more closely resemble figurative meaning making in human language. These alternative models therefore better allow for the computational representation of figurative elements that N. Katherine Hayles describes as “the inexplicable, the unspeakable, the ineffable” of narrative literature. This paper argues that the construction of figurative meaning through paradigmatic substitution (as part of an imaginary vocabulary that is drawn from in the process of meaning making) is difficult to account for in the relational database—arguably still the most culturally prominent database model. By focusing on NoSQL (“no” or “not only” Structured Query Language) databases, this paper explores how layers of figurative meaning can be represented together through these flexible and non-relational models. In particular, the ability of non-relational databases to group together multiple values—encouraging their association, comparison, and juxtaposition—can be analyzed as a computational albeit imprecise counterpart to the formation of paradigmatic and figurative meaning. Thus, towards accounting for a word, image, or idea’s layers of meaning as expressed in literature, this paper offers a study of the limitations of digital tools and their critical negotiation with humanities research and reflection.

Keywords: Narrative; database; reflexive; figurative meaning; paradigmatic; language; digital tools; relational; non-relational; NoSQL; narratology; humanities advocacy

Cet article reprend la relation de « narration par rapport à base de données » (un argument posé à l'origine par Lev Manovich en 2001) comme une relation qui pourrait recevoir une plus grande attention. Une question précise persiste dans la recherche de l'analyse de texte en humanités numériques: la difficulté de représenter la signification figurative des narrations par l'entremise d'outils numériques. Comme mesure d'accommodement, cet article adopte un cadre de travail narratologique afin de proposer d'autres modèles de gestion et d'organisation de contenu qui ressemblent plus étroitement à la recherche de signification figurative en langage humain. Par conséquent, ces autres modèles permettent une meilleure représentation informatique des éléments figuratifs que N. Katherine Hayles décrit comme « l'inexplicable, l'innommable, l'ineffable » de la littérature narrative.

Cet article soutient que la construction de la signification figurative par la substitution paradigmatique (dans le cadre d'un vocabulaire imaginaire qui en est tiré dans le processus de recherche de la signification) est difficile à représenter dans la base de données relationnelle —sans doute encore le modèle de base de données le plus prédominant sur le plan culturel. En mettant l'accent sur les bases de données NoSQL (« non » ou « non seulement » en langage de requête structuré), cet article explore comment les niveaux de signification figurative peuvent être représentés ensemble par l'entremise de ces modèles flexibles et non relationnels. De plus, la capacité des bases de données non relationnelles à regrouper des valeurs multiples —encourageant leur association, leur comparaison et leur juxtaposition — peut être analysée comme un homologue informatique quoique imprécis de la formation de la signification paradigmatique et figurative. Ainsi, en ce qui concerne la représentation des niveaux de signification d'un mot, d'une image ou d'une idée tels qu'ils sont exprimés en littérature, cet article offre une étude des limitations des outils numériques et de la négociation critique avec la recherche et la réflexion des humanités.

Mots-clés: Narration; base de données; réflexif; signification figurative; paradigmatique; langage; outils

Introduction

As a field of scholarship, the digital humanities are increasingly important to understand and develop, as they are uniquely attuned to the wide-ranging impact of digital media and culture. Yet, there remains a discrepancy between the epistemological underpinnings of the humanities and digital technologies and culture. On the one hand, we live in an information age that privileges technological progress and that is tasked with the creation, storage, and management of large amounts of data. On the

other, our (western) traditional methods of interpreting information are grounded in humanities philosophy—through theoretical, interpretive, and reflexive methods of understanding history, tradition, culture, and storytelling.

The epistemological differences between digital technologies and the humanities are in one way exemplified by the relationship between the database and the traditional narrative. There is a discursive history of thinking about the database and narrative in terms of opposition, most notably beginning with Lev Manovich's article "Database as Symbolic Form" (1999) and its expansion in the foundational *The Language of New Media* (2001), in which he calls database and narrative "natural enemies," even expressing "surprise" that narrative still exists in new media (2001, 225; 2001, 228).¹ Since such statements, the relationship between narrative and database has been examined to reveal more complexity. In the areas of digital humanities literary research and digital narratives in particular, narratives and databases are often analyzed in terms of their dynamicism, as digital tools can be used to store, manage, represent, share, and create narrative literature—for instance, through electronic literature.

Yet, many of these digital tools and methods have limitations that are at the crux of Manovich's original argument—namely the problem of juggling the ludic depth of literature with the qualities of precision, efficiency, and "knowing" that are dictated by rigid data management models and systems both on-screen and behind the screen. Given that machinic operations are designed to produce outcomes, quantify data, and otherwise offer answers, is it possible for methods of quantification to represent, for instance, the depth or affect of a metaphor?

Databases in particular, as computational structures of content management, may struggle to store let alone re-present figurative meaning in literature. As this paper will show, this difficulty stems from the broader limitations of digital tools for representing the semiotic depth that is foundational to paradigmatic meaning

¹ Manovich himself has since further explored the nuances of this relationship in his work on software studies, most recently in the 2012–2015 Mellon research project on big data called "Tools for the Analysis and Visualization of Large Image and Video Collections for the Humanities."

making through human language. In using digital tools and methods to represent literature, then, digital humanists must ask whether the methodological prowess and scope of digital tools risk any loss of literary- and humanistic-based reflection and interpretation.

I am not in this sense the first to inquire into the wheres and whens of the “H” in DH.² This does not imply that the digital humanities are not humanistic; rather, it refers to scholarship in literary studies, media studies, and the digital humanities that calls for investigative analysis that can account for more reflexive and interpretive ways of thinking. For instance, in response to Manovich’s 2001 statements, N. Katherine Hayles (2012) contends that any scientific and engineering research presented through data and facts requires narrative for “the interpretation of the relations revealed by database queries” (2012, 182). Narratives are necessary to articulate the contexts and implications of any data- or fact-based research, including: background information; relations between groups; examinations of patterns in statistics; possible applications and their outcomes; and alternative methodologies that had been or could be attempted. The use of explanation in these examples illustrates the praxis and necessity of narrative forms and training even for research that is grounded in data, presenting a significant case for the value of reflecting upon narratives and narrative representation. This includes digital humanities projects and texts that are digitized, born-digital, and digitally informed.

I argue that the identification of the limitations and affordances of digital tools and methodologies for literary analysis only remind us of the value of two modes of inquiry in a humanistic digital humanities:

- Humanistic thinking: reflexive and interpretive modes of inquiry in which humanities scholars and students are trained. These modes uniquely position us to ask whether the use of quantitative digital methodologies and tools (which participate in a discourse of “efficient” and “precise” meth-

² In addition, for whom is the “H” in DH? Further inquiry into this question can address critical issues surrounding (the politics of) representation in the digital humanities, as explored by scholars such as Adeline Koh and Anne Cong-Huyen.

odological prowess) risks any priorities and responsibilities of the larger humanities project.

- Narratological thinking: an understanding of the linguistic play and semiotic depth of language as it is used to construct works of narrative literature. Narratological thinking requires a consideration of literary elements such as plot, theme, imagery, poetics, medium/media, and intertext. Narratological thinking is, in this sense, a mode of inquiry that is necessary to understanding how figurative meaning functions as a unique and vital quality of meaning making in general, including how we communicate with each other by offering information in the form of stories.

Together, these modes of inquiry as applied to the digital humanities encourage the critical comparison, juxtaposition, interpretation, and reflection of digital tools and research—a critique that is a *necessarily ongoing* endeavour in the still-nascent stages of development for the digital humanities as an academic field.

Applying these two modes of inquiry to the analysis of specific database models that are popular for structuring, managing, and representing data reveals that the discussion of “narrative versus database” is not over. In fact, these database models point to an issue that continues to be a topic of inquiry and even skepticism in digital humanities text analysis projects: that, whether qualitative or quantitative, digital tools are not always capable of capturing the essence of makes literary texts “literary” in the first place—including the elements of figurative meaning that Hayles describes as “the inexplicable, the unspeakable, the ineffable” of narrative literature (2012, 179).

What humanistic and narratological modes of inquiry reveal, then, is the need for alternative models of content management that better accommodate for the literary. Towards such an accommodation, this paper proposes that digital text analysis projects can utilize NoSQL or non-relational database models—an approach to content (as data) management that more closely resembles the paradigmatic dimensions of meaning making in human language and that therefore begins to address elements of figurative meaning that carry so much literary “weight” and semiotic depth through

their imagery, metaphor, and depth in human language. This alternative content management model is especially pertinent, I show, to address contemporary forms of narrative literature that mediate the impact of digital structures and representation on how we read, write, and think of literature itself today.

Seeking methods of representing figurative meaning is only one way that humanistic and narratological thinking can encourage reflexivity and interpretation in the digital humanities. In this sense, I offer this paper that explicitly focuses on figurative meaning as only the start of a broader study on the dynamic between digital and narratological meaning making. Terms that I map throughout as a part of this ongoing comparison, juxtaposition, interpretation, and reflection may be aligned with my earlier descriptions of the epistemological underpinnings of a computational information age and (western) humanistic philosophies, whereby “database” and “narrative” as network nodes may branch out to include the quantitative and qualitative, data and interpretation, and the literal and the figurative. These terms, much like their nodal roots, are not to be considered in opposition, but rather, as in connection and thus conversation along with other existing epistemological modes of knowledge. The main difference I wish to illustrate is the wait and weight of the humanities: its position to inquire beyond that which is “known” and its critical negotiation with that which claims to know.

Database versus Narrative: The Known and the Unknown/ Indeterminate

The “narrative versus database” discussion emerges from Manovich’s description in *The Language of New Media* of the rise of a “computerization of culture,” in which the database plays a key role as a symbolic form and significant cultural form (2001, 43). While many scholars have sought to reframe the relationship between narrative and database to reveal more complexity (discussed further below), it remains the case that there are aspects of literary narrative that are not accounted for or represented by all digital tools, simply because of the ways in which these tools are designed to manage content.

Some database models are rigid in their parameterization of content and others are more flexible. It is therefore necessary to distinguish that while Manovich and

Hayles identify many models of databases and content management, each focuses on the *relational database*, which has been and arguably predominantly still is the database form of cultural choice (Dourish 2014, n.p.). Relational databases resemble the format of spreadsheets such as those seen in MS Excel, as both resemble print-based forms such as the index, their table structures remediating analogue methods of information organization that existed long before the digital computer was invented. It is perhaps this transferability of and therefore established literacy in more familiar cultural forms that attribute to the relational database's continued popularity as a database form.

For her own choice, Hayles explains in 2012 that the relational database has “almost entirely replaced the older hierarchical, tree, and network models” and also object-oriented database models (176). The relational database is composed of one or more tables (with rows and columns) that are drawn from for their data, a structure that is dictated by its programming language, SQL (Structured Query Language). SQL offers a rigid form of data organization through which content is dictated by the model of the table: if one requests data from a relational database, one must specify its database location; in reverse, any changes to the database structure or hierarchy of organization are also expressed in the code.

It can be said that this rigidity is unavoidable because of why the general database was developed in the first place. The influx of digital devices offers a bounty of data that has become our blessing and our curse, as we try to find the “best” ways to manage and access data, typically through the methods of structured languages, programs, and databases. This leads to the creation of databases as “collections of items on which the user can perform various operations: view, navigate, search” (Manovich 2001, 234), and many computer operations function through the operations of requesting, adding, deleting, and updating data.

Two types of potential incommensurability between the narrative and database emerge: the structural/formal and the semiotic. Manovich's initial observations of the discrepancy between narrative and database involve a consideration of how content is maintained and managed differently among distinct cultural forms. Specifically, it is the amount of information collated in digital culture that presents a conundrum

of *structure and form*: data storage and management in computational devices allows for a massive amount of content to be stored, often resulting in efforts to mass archive and digitize that began in the early 1990s as a trend Manovich describes as “storage mania” (2001, 234).

In contrast, narrative cannot nor does it traditionally try to contain all information. As defined by narratology scholars such as Mieke Bal (2009) and David Herman (2009), a literary narrative is defined by its dynamic movement between markers of time (the beginning and end of a trajectory), composed of what Joseph Tabbi and Michael Wutz (1997) describe as “the progression of a central protagonist from a beginning through a middle toward an end that progressively diminishes possibilities and so represents that character’s fate” (14). The traditional narrative follows a cause-and-effect model, certainly a model of meaning making in which a linear pathway is developed in the mind of the reader and in which not all trajectories are mapped. For these reasons, Manovich argues of database and narrative that “each claims an exclusive right to make meaning out of the world” (2001, 225). As the database is a dynamic body of information with no beginning or end, he asks, “how can one keep a coherent narrative or any other developmental trajectory through the material if it keeps changing?” (2001, 221).

But more complicated is the question of language-based *semiotic* content when it is stored and represented as data. Given the imposed rigid structure of relational databases in addition to its ability to edit content, the question of narrative versus database requires that we—and by this I mean digital humanists, but also computer scientists who work in linguistic-informed areas such as NLP (Natural Language Processing)—further negotiate computational semantics of content organization (computer-specific meaning making) in relation to human languages and the semiotic construction of meaning through language. It is in part such a negotiation upon which Manovich draws in order to anchor some of his juxtapositions between narrative and database, particularly through a delineation of the differing functions of paradigmatic and syntagmatic dimensions of each. These dimensions are important because they function as a core aspect of human language—the logic of syntagmatic grammar and paradigmatic substitution through which we form

semiotic meaning in sentences. Here, Manovich argues that “in the case of a written sentence, the words that comprise it materially exist on a piece of paper, while the paradigmatic sets to which these words belong only exist in the writer and reader’s minds,” and in contrast, in the database, the “paradigmatic dimension” has “material existence” (2001, 230–1). He thus imagines, Hayles describes, that “the paradigmatic possibilities are actually present in the columns and rows, while the syntagmatic progress of choices concatenated into linear sequences by SQL commands is only virtually present” (2012, 180).

Hayles disagrees with the idea that databases possess much less relay paradigmatic meaning in this way. As content management tools such as the relational database may abstract content (such as text in words and clauses) into individual rows and columns, they force content (and any generation of content through a “transition” across rows of cells) to follow the organizational schema dictated by the database’s structure and organization. So, while all content is materially present in the relational database, Hayles stresses that “in neither rows nor columns does [the paradigmatic dimension’s] logic of substitution obtain; the terms are not synonyms or sets of alternative terms but different data values” (2012, 180).

Her observation of the limits of this model of content management for paradigmatic and syntagmatic meaning making reveal that the way relational databases encourage us to interpret data is not how human language works, how humans make meaning out of language or narrative, or how humans construct meaning through narrative. Hayles’ distinction matters to a discussion of figurative meaning because of the formation of figurative meaning through a paradigmatic set of associated meanings.

Figurative meaning, which can be described as the association of a signifier (as a word, image, or idea) with potential metaphors, similes, analogies, tropes, and metonymies, is constructed through the paradigmatic dimension—an imaginary set of affiliations that are shaped through composition of, and encounter and practice with, cultural texts and objects. Figurative meaning can therefore only be constructed through a logic of substitution such that a subject can associate a signifier with a set of related meanings—a process of exploratory and imaginary substitutions that

I will describe as creating a *depth* of meaning and therefore as possessing a “deep movement” through the paradigmatic set. Paradigmatic sets of literal and figurative meaning are thus different albeit related: for example, a paradigmatic set of literal meanings for the word “red” may include the synonyms “crimson,” “rose,” “carmine,” “cherry,” “scarlet,” and “vermilion,” while a paradigmatic set of figurative meanings may include “passion,” “lust,” “rage,” “fever,” and “violence.”

The limits of the relational database for representing paradigmatic meaning in literature can be narrowed down to an aspect of what makes literature “literary” in the first place—and one characteristic is its depth of meaning beyond the literal and through the figurative that necessitates qualitative and reflexive analytical methods rooted in literary study, such as close reading. In particular, Hayles proposes that the epistemological differences between database and narrative are rooted in their differing “worldviews” through the element of *indeterminacy*, as narratives reach for it and databases are designed to avoid it.

The element of indeterminacy is attributed as a quality of the literary character of narratives and also encourages close reading for an interpretive exploration of a text’s layers of meaning. Hayles juxtaposes narratives and databases through the indeterminate in this way, arguing that:

Narratives gesture toward the inexplicable, the unspeakable, the ineffable, whereas databases rely on enumeration, requiring explicit articulation of attributes and data values ... databases in themselves can only speak that which can explicitly be spoken. Narratives, by contrast, invite in the unknown, taking us to the brink signified by Henry James’s figure in the carpet, Kurtz’s ‘The horror, the horror,’ Gatsby’s green light at pier’s end, Kerouac’s beatitude, Pynchon’s crying of Lot 49. (2012, 179)

In this string of examples, the figurative is indeterminate insofar as it provokes imagination and a depth of possible meanings: the single image of Jay Gatsby’s green light captures (at the same time that it overwhelms) the character’s yearning for a system of ideals that are epitomized in the character Daisy. His yearning is metaphorized in the unreachable light, the hue of which also represents envy. If

there is a way to quantify the depth and affect of these layers of the indeterminate through figurative meaning, we have not necessarily yet found it.

On Limits and the Value of Humanistic and Narratological Thinking

Digital humanists have actively attempted to ameliorate such differences by drawing upon both digital and humanistic methodologies and philosophies. For instance, Hayles proposes to “locate digital work within print traditions, and print traditions within digital media, without obscuring or failing to account for the differences between them” (7). She has sought to address larger-scale ideas of difference between the logics of meaning making in the humanities and the digital through the specific media with which they are associated and through which they often work. This is the basis of her proposal of a “media-specific analysis” in 2004’s “Print is Flat, Code is Deep: The Importance of Media-Specific Analysis.”

Building on the need for media-specific analysis, one of her central arguments in *How We Think* (2012) is that we require three modes of reading in an era in which “print is no longer the default medium of communication,” naming these modes as close reading, hyper reading, and machine reading (2012, 249). As the identification of literary studies with the practice of close reading risks pushing digital reading “to the margins as not ‘really’ reading or at least not compelling or interesting reading,” Hayles examines the value of hyper reading as a necessary method for today’s scholar to engage with all the materials and resources that are made available today (2012, 60).³ Drawing upon James Sosnoski, she also offers examples of hyper reading texts through search queries, filtering with keywords, skimming, hyperlinking, fragmenting, “pecking” (“pulling out a few items from a longer text”), and juxtaposing (a comparative method of reading across, for instance, several open browser tabs and windows) (2012, 61).

³ In regard to the number of books that can hypothetically be read in a single lifetime, Hayles cites Gregory Crane’s argument “that the upward bound for the number of books anyone can read in a lifetime is twenty-five thousand (assuming one reads a book a day from age fifteen to eighty-five)” (2012, 27).

A “synergy” or “recursive feedback loop” among close, hyper, and machine reading is thus necessary in an era in which our understanding of communication must take into consideration the specific “affordances and limitations” of individual media systems, as Marie-Laure Ryan describes (2004, n.p.).⁴ We need methods of reading that are specific to interpreting and scrutinizing the minute of individual texts (close reading), methods of reading that can account for enormous collections of digitized texts (hyper reading), and methods of reading that can process computer code of varying degrees of abstraction (machine reading) (Hayles 2012, 58-72). Hayles’ tripartite model of reading, then, shows that hyper reading and machine reading, which are digitally informed, can also be applied to methods of interpretation and by extension to reading narrative. In this sense, digital methods of information and content engagement can make room or account for narrative forms and narratological thinking.

The development of the digital humanities has also seen a surge in literary text analysis projects that take quantitative, data-based, or algorithmic approaches to literary research, representation, and analysis. Twelve years ago, Franco Moretti published *Graphs, Maps, Trees: Abstract Models for Literary History* (2005), a fascinating re-approach to literary study whereby the data visualization of hundreds of literary texts’ narrative content (through graphs, maps, and trees) allows us to grasp larger trends in literary history through a method he calls “distant reading.” Six years ago, Google Books and Harvard physicists attempted to quantify the English language through a database: drawing upon millions of digitized literary texts, they mapped patterns in the literary usage of words through a method called “culturomics,” whereby language is proven to reflect cultural atmospheres and change (Michel et al. 2011). In the past five years, and with increasing urgency and interest, digital humanists and literary scholars have expanded methods of database

⁴ For an excellent example of recursive and comparative reading in action, see *Reading Project: A Collaborative Analysis of William Poundstone’s Project for Tachistoscope (Bottomless Pit)* (2015) by Jessica Pressman, Mark C. Marino, and Jeremy Douglass.

analysis to consider the computational representation and potential quantification of narrative.

Yet, if literature possesses a quality of the indeterminate and if the objective of the database is to avoid the indeterminate, we must question of limits of digital representation itself for analyzing aspects of the literary. The identification of these limitations occurs through two crucial modes of inquiry I describe in the introduction: humanistic thinking and narratological thinking.

While relational databases are useful for counting instances, exploring degrees of relationships, visualizing patterns and shifts, and so forth, in the data itself there is little reflexive meaning; as Hayles notes, it needs to be formed through interpretation (2012, 179). When examining databases, meaning and humanistic reflection come in at another layer, in part through additional information and in part through the interpretation of data. A narratological approach to digital text analysis may allow us to expand upon approaches to literary intertext as paratext that is significant to a work's larger corpus—in write-ups, commentary, footnotes, endnotes, appendices, forwards, afterwards, glosses, and so forth—and to think of extensive metadata itself as an accompanying narrative about a text and its contexts. In particular, by examining descriptive metadata that articulates examples of data content and application, we may construct comprehensive narratives of the processes of content production, management, access, and reception, shaping narratives about the trajectory (the cause-and-effect) of digital humanities projects, tools, and research.

To see how humanistic and narratological thinking aid in the identification of the limitations of digital tools for representing literary text, I will discuss a text analysis project that reflects upon these limitations: *Network Theory, Plot Analysis* (2011), which comes out of Stanford University's Literary Lab. The Literary Lab, co-directed by Franco Moretti and Mark Algee-Hewitt, houses several collaborative projects that, upon completion, are published on the Lab's website as research pamphlets. In the project pamphlet of *Network Theory, Plot Analysis*, Moretti analyzes narratives through the quantification of literary elements and variables. These and similar projects re-visit key ideas and principles of narrative hermeneutics through

their mediations of narrative data. At the same time, Moretti's use of data storage and visualization methods to represent literary features is paired with his reflexive mode of literary criticism, which observes that his methodology may be unable to capture what he calls the "weight" of narrative (2011, 2).

The pamphlet's write-up, which is itself a form of metadata and an integral part of the project's paratext, is thus revealed to be necessary to understanding the data visualizations. It takes up a storytelling mode to speak to the project's struggle to negotiate narrative factors with network diagrams. Also, it considers this struggle in a way that retains and captures the humanistic inquiry of a digital humanities that is critically reflexive of its own tools and methodologies.⁵ The pamphlet utilizes network theory in order to visualize relationships between narrative characters, including in

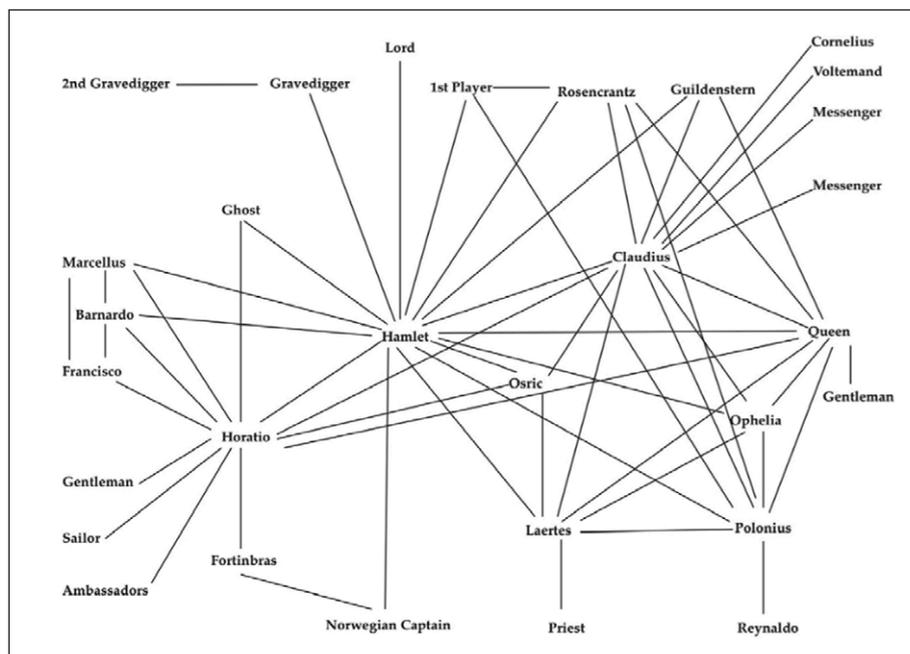


Figure 1: *Network Theory, Plot Analysis*. Literary Lab, Stanford University. 2011.

⁵ This article was written, reviewed, and revised for publication before allegations of intimidation and sexual assault were made against Moretti by several former students (Liu and Knowles 2017). According to these reports, the unproven allegations are under investigation by Stanford University as this article goes to press (editorial and authorial note).

Shakespeare's *Hamlet*, the analysis on which I will focus. The research question for a network data visualization such as **Figure 1** may be "who speaks to whom and how often?" whereby the most loquacious characters (here, Hamlet, Claudius, and Horatio) occupy more central positions in the network and minor speaking parts are on the outskirts. The data that corresponds to and generates this network could (but does not necessarily) take the form of a relational database, as it is an excellent tool for methodological tasks such as counting the frequency of something. We may say that the parameters of this relational database are also defined by the same research question, "who speaks to whom and how often?" such that characters' names could be charted on both X and Y axes of a relational database, and their direct encounters could be ticked off.

If the research question is intent on studying the frequency and relations of dialogue, a content analysis through the relational database is most apt; however, if the research question inquires more deeply into character relations and dynamics in the plot, then a relational database's corresponding visualization is not as

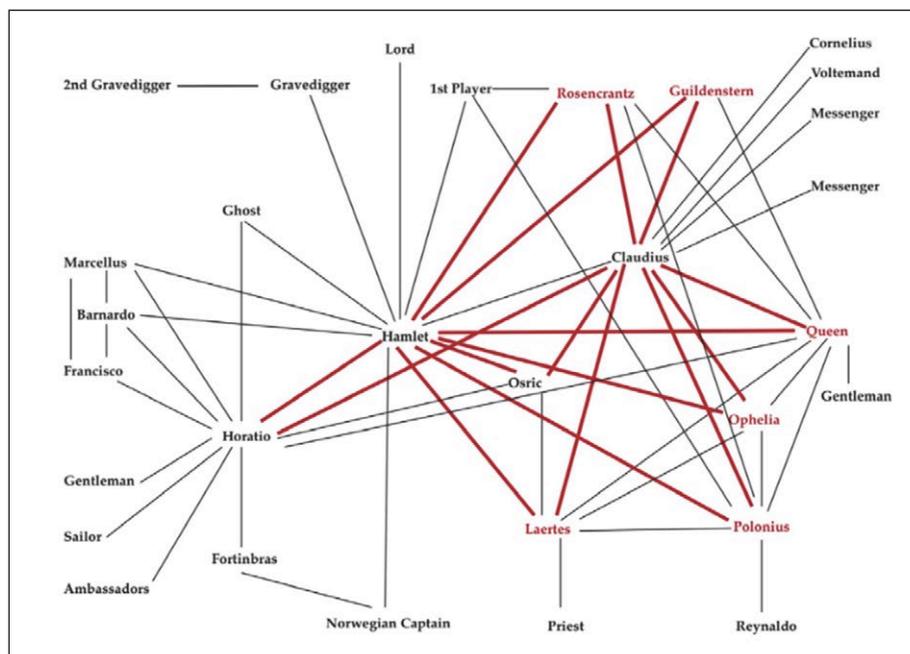


Figure 2: *Network Theory, Plot Analysis.* Literary Lab, Stanford University. 2011.

clear. For example, **Figure 2** is a visualization of deaths in *Hamlet*. The “region of death” in red illustrates the group of people who kill each other off at the end of *Hamlet*. The research question may be: “who dies and what is their relationship to other people who also die?” and the relational database parameters that produce the corresponding network may involve two layers: one table for character encounters and one for character deaths. The resulting data visualization allows researchers to compare characters’ encounters in dialogue to the actual people who die, and by doing so, we arrive at an analysis of the significance of certain relationships. The data visualizations developed by the network theory framework noticeably sway us toward the idea that the characters who engage more frequently in dialogue are also the ones that die, especially the ones that die together in the play’s final scene; yet, such a observation might discount the significance of characters such as Polonius and Ophelia, who die earlier in the play.

While the pamphlet possesses 57 data visualizations that are derived from the data of corresponding databases, some narrative thinking appears to be necessary to define the parameters of the databases. The pamphlet is also necessary to reflect upon how data visualizations encourage us to analyze narrative aspects of *Hamlet*, especially compared to traditional narrative hermeneutic techniques. The visualizations may be more analytically interesting than charts of data alone, but it is arguable that their main function in the *Network Theory, Plot Analysis* pamphlet is to complement Moretti’s explanation of the processes of the research. We may view this explanation as a narrative itself in the following structure: an original text was studied for features of plot (a narrative of text analysis); multiple narrative views were negotiated according to specific research questions that were derived from narratological thinking (a narrative of data analysis); and this exploration revealed discrepancies between narratological thinking and representing narrative through digital tools (a narrative of quantified text analysis).

It is these discrepancies that allow Moretti to identify and ruminate upon a significant idea: the uniqueness and complexity of what he calls narratological “weight” appears to elude his network visualizations. The “weight” of certain events for plot development, for instance, can be difficult to quantify, especially in a way that

is easily encoded for data management, and graphing and visualization purposes. To show the possible implications of attempting to quantify this literary weight, Moretti discusses the clustering and positioning of character encounters (see **Figure 3**). The significance of Hamlet, Claudius, and Horatio is spatially represented by the fact that they occupy central positions in the data visualization. In comparison, the ghost has few lines of dialogue and is therefore on the outskirts of the diagram, equated in spatial significance with characters such as “Gravedigger” or “Norwegian Captain.” In fact, the scene between Hamlet and the ghost is of fundamental importance to the rest of the narrative, as it is the ghost who inspires Hamlet’s theory that Claudius killed his father and thus his revenge plot. Yet, as the data visualization is unable to represent this weight, in this way, network theory risks reducing and abstracting the plot (Moretti 2011, 3). Matthew Jockers’ work in text analysis and plot visualization in *Macroanalysis: Digital Methods and Literary History* (2013) mollifies this specific issue, building on early work by Kurt Vonnegut in plot diagramming to capture the significance of chronological plot events in a linear series of crests and dips. His work

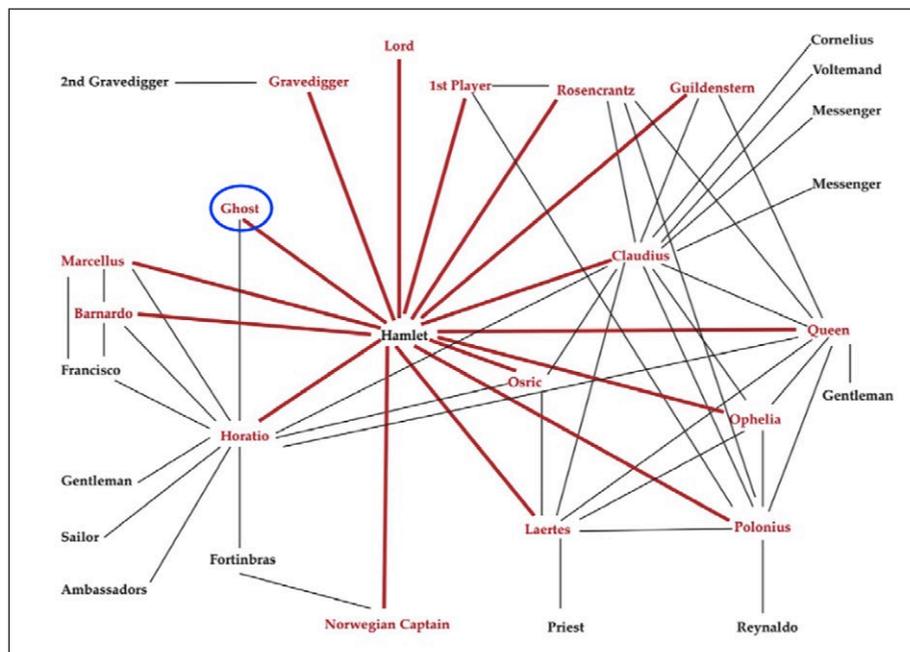


Figure 3: *Network Theory, Plot Analysis*. Literary Lab, Stanford University. 2011.

is related to that of the larger research team of *Novel TM* (to which Jockers belongs), a transnational research project on text mining the novel that is led by Andrew Piper.

While the issues of representation that *Network Theory, Plot Analysis* identifies are being actively tackled, I find equal value in reflections on certain limitations of data-based digital representations and analyses of literature. Reflexive inquires into digital humanities analysis, tools, and research production produce a text that is able to weave between this reflexive critique and the media-specific analytical affordances offered by digital media. In particular, what the observations reveal is that in the processes of both close and hyper reading, we should not make generalizations about content based on the data visualizations or their corresponding databases, as additional knowledge is often needed.

In the case of *Hamlet*, additional knowledge about the play's specific narrative is helpful to effectively analyze the visualizations. And with regards to databases and data visualizations for plots that are not so well known or accessible as *Hamlet* (for instance, rare texts, texts that are hard to access, or texts that are subject to copyright), it is through additional information and commentary that many of these difficulties are fleshed out. The metadata, here as a formal write-up, is crucial to clarifying where database forms and digital tools can fail, especially through their discrepancies with literary form, content, and hermeneutics. There is critical value in surprises, hiccups, obstacles, and failures, towards "failing better" so to speak.

Representing Paradigmatic Meaning in Non-Relational Databases

Identifying the limitations of these forms and tools critically gesture towards seeking alternative methods of content management that better accommodate for and represent on-screen: literary weight, figurative meaning, narrative forms, and linguistic play. In this sense, greater commensurability between the database and narrative (and between the unique cultures or "worldviews" of meaning making to which each belongs) may be met through the design or at least imagining of digital tools that can represent the indeterminate in figurative meaning. There is no "one-size-fits-all" computational tool for content management and representation, requiring that a digital humanities researcher, teacher, or student who is offered

multiple possibilities for content mediality and mediation weigh the pros and cons, the affordances and limitations, of various digital tools.

Returning to the earlier differentiation between paradigmatic sets of literal and figurative meaning—where a paradigmatic set of literal meanings for the word “red” differs from a paradigmatic set of its figurative meanings—a digital tool that offers a paradigmatic approach to content management would better accommodate and account for both these literal and figurative paradigmatic sets, particularly if it is flexible enough to allow for set editing and expansion. As with when a subject mentally searches for synonyms and metaphors out of their existing vocabularies and can also expand those vocabularies through training and reference, in the same way, a database model with a paradigmatic approach to content could store sets of meanings as imaginary possibilities that could be expanded. To be clear: the relational database also allows for this expansion, because one can theoretically add to it forever. However, the difference between the relational database and a database model with a paradigmatic approach is in its structure: the latter offers a non-relational—that is, non-rigid—schema for storing, organizing, receiving, and engaging with content-as-data.

Relational databases are useful for when one chooses parameters and variables that are likely to be content-rich, or when one has the time, reason, and occasion to go through different possible transitions across rows of cells. These methods work best with small amounts of data; however, if a researcher or teacher is trying to organize or interpret an enormous amount of cells and transitions, and if many of these cells do not have values, then the relational database fails on accounts of digital scalability, memory, and speed. What happens when we move away from the relational database or if we at least incorporate other forms of database?

In this regard, I do not refer to other traditional databases with SQL encoding models such as attribute-value, network, or hierarchal databases, but to a more recent paradigm of information organization: publicly introduced in 2009, and with particular significance starting around 2012, computer scientists have begun to embrace the NoSQL (“no” or “not only” SQL) movement, pushing for non-relational databases (Dourish 2014, n.p.). NoSQL is a database model that takes on several

formats, including a document style that allows data to be organized in groups and a graph model that can resemble a network.⁶

NoSQL was developed for programmers to code and alter data more easily through less rigid schema. As one aspect of this effort, content is less abstract and isolated in individual cells, often offering textual context in a way that can be read as metadata. NoSQL organizes data into a flow-chart form, with keys that can be defined by any consistent variable, such as a list of course codes or a series of dates. The particular trait and thus particular format of NoSQL databases that I want to focus on is the document-oriented database's ability to group together multiple values for each key (called a key-value or attribute-value store). Whereas in the relational database, each individual cell contains one value, for a NoSQL database, each key can contain a group of values (see **Figure 4**). For instance, for the database "ENG_101" (a course called "ENG 101"), each student could have associated values such as "name," "major," and "student id."

Rather than being structured in relational tables, the key-value model of data management, especially through document-oriented databases, can organize content to more closely resemble the paradigmatic dimension of language. Having associated values grouped together would allow multiple values to be read together as a set of paradigmatic words or associations, so that the values of the key "red" can contain "passion," "anger," "fever" and so forth, thereby offering an embodied version of a paradigmatic logic of substitution (see **Figure 5**). Additional values can also be added to the group of values through client reading and writing (user engagement).

Applying Non-Relational Databases to Narratives: Towards Representing Figurative Meaning

The layers of meaning in a figurative text and the weight that they carry to tell a story find a computational albeit imprecise counterpart in the paradigmatic approaches of content in non-relational databases. Such alternative content management models

⁶ It is entirely possible that the graph format for the Network Theory, Plot Analysis pamphlet was used because of the objective of generating network visualizations. I have analyzed these visualizations as if they are constructed through relational databases only because the accompanying observations about the potential limitations in network models for text analysis occur through—and therefore serve to underline—issues of rigidity in relational databases.

are important as we consider emerging forms of literature that increasingly trouble how we think of “narrative” and that highlight the difficulty that many digital tools have with capturing or supporting elements of figurative meaning.

One particularly important shift in re-thinking “narrative” that has also reframed its relationship with the database as a dynamic is the cultural practice and advent of narratives that are digitally composed and informed—through hypertexts, born-digital narratives, and online texts that do not necessarily embody a linear cause-and-effect form. Through the introduction of intermedial and transmedial (and thus trans-spatial and trans-temporal) qualities to cultural texts, digital narratives foster conversations about what it means to read and write creatively in and across various

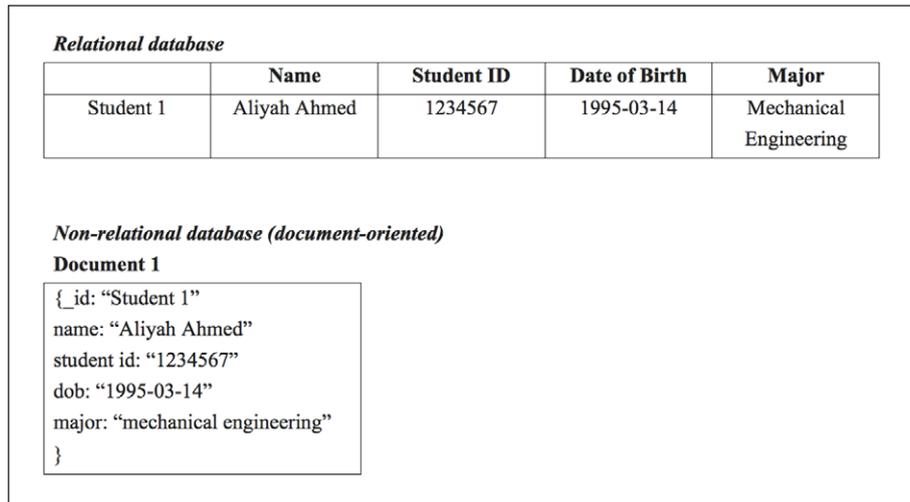


Figure 4: A relational database compared to a document-oriented non-relational database.

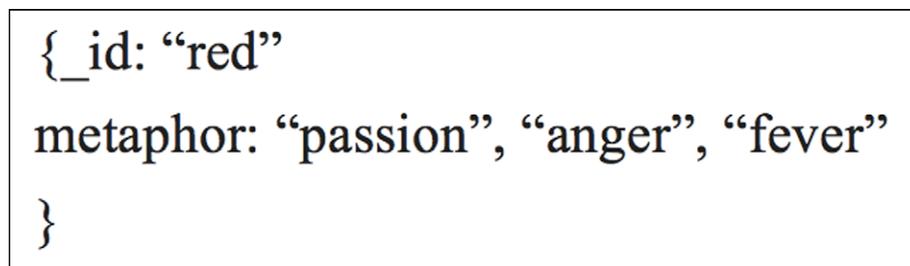


Figure 5: Document-oriented database containing the key “red”.

media platforms. For instance, variances have been identified between analogue and digital forms of textuality towards a complexity of their interrelations (see Ryan 2003; Hayles 2004; Morris 2006; Hayles 2008; and so forth). One premise in studies of new media, electronic literature, and literary studies is that in a so-called digital era, narrative persists, and often, that narrative *resists* by mediating their medium-specificity and unique material circumstances of meaning making (see Fitzpatrick 2002; Goldsmith 2011).

Digital narratives are especially unique because of their representational duality: on-screen narrative content always corresponds to what is behind the screen—computational methods of storing and managing content as data. The implications of this duality for digital writing lead Alan Liu (2004a) to identify a deviation of rigid database, markup schema, and encoding formats from the textual practices of older communication forms such as print. He argues that such rigid factors can confine writing and today's creative writer to the structure and content dictated by the database format and its behavioural parameters; this effect potentially leads to the author's "surrender[ing] the act of writing to that of parameterization" (59).

These literary shifts and the attempts to grasp them in interpretive analysis and digital tools prompt a humanistic and narratological inquiry into the place of the digital relative to older literary genres and styles that are difficult to represent for their layers upon layers of meaning. How, for instance, can we use a database to represent the relationship between image and text in the graphic novel? How can we represent the technique of literary stream-of-consciousness and the thematics of disorientation and fragmentation that might provoke it? In terms of ontology, how would we visualize temporality versus duration, or how could we visualize reference and memory?

Towards addressing some of these questions, I will briefly discuss a NoSQL approach to representing the properties of literary weight and semiotic depth in the construction of a world and its specifically defined ontology—its properties of time, space, and of narrative movement that I describe as an "imaginary ontology." Databases could be described as imaginary ontologies insofar as they create defining

parameters of being within which things are and through which events—which is to say, the phenomenological reception and mediation of content—can emerge.

I am specifically interested in literary texts that trouble what we mean by “narrative” or “book” through their mediation of the computerization of culture and representation, such as Mark Z. Danielewski’s *House of Leaves* (2000). *House of Leaves* is a novel that plays with the cultural form of the print-based narrative at the same time that it composes this narrative through a collection of technological mediations. In doing so, it reflects the changing face (and body) of literature in a cultural era when narrative and database engage in recursive dynamicism.

In fact, it is the *structure* and *format* of the novel’s mediations that attribute to *House of Leaves* its digital character and that characterize it as a kind of database itself. Others have described the labyrinthian and networked structure of the novel through its text, intertext, and paratext, as the novel offers multiple perspectives on the same events (see Hamilton 2008; Pressman 2006). It has at least four narrative trajectories, it employs various literary and mediated techniques to represent them, and it offers associated ideas, characters, and other information in a multi-linear way. This structure is comparable to the way content (as data) is stored in a database, as data must be hashed together from discrete locations to create digital objects and images.

The scattering of texts, images, and symbols about the pages of *House of Leaves* follows that the narrative is only formed through the compositing of fragments of information—and this action is also what makes *House of Leaves* inherently literary. The reader is told at the beginning of the novel that the central object of the plot (the house) cannot physically exist; however, the plot revolves around the mystery of the house and its mediations by other characters. That is: the content of the novel itself does not exist without mediations (Hansen 2004, 628). Expanding this series of mediations further, the novel’s narrative and meanings are constructed by the reader’s mediations of textual fragments into a formed “story” through their narratological thinking. The reader’s reflection upon and engagement with such fragments—with the semiotic depth of texts and images as well as with layers of mediation—draw out

their paradigmatic as well as literary meanings. The agential subject's mediations and narratological thinking, *House of Leaves* shows, remain central to the construction of meaning, whether the content is stored in a database or presented as a book of literature. To this effect, the novel offers a recursive feedback loop between narrative and database that is intent on encompassing the reader's mediations of the novel itself—what Mark B.N. Hansen describes as “copies with a difference” (2004, 618)—and that legitimizes them as part of the novel's (para)textual corpus.

House of Leaves is notably difficult to represent as data. To represent the intermediality, multimodality, and multi-linearity of the text in a relational database would result in a large collection of tables, many of which would be filled with empty values. A researcher may then have to compare the data in dozens of different visualizations while also addressing the database tables that such visualizations refer to. A digital text analysis project on *House of Leaves* through a non-relational database would ideally:

1. represent the layers of meaning in its narrative content, including through literary elements such as metaphor and trope;
2. represent the novel's discrete methods and instances of mediating the same narrative idea, space, or event; and
3. represent the various ways in which each of these methods and instances overlap and interact with each other.

For example, consider that the narrative event when Will Navidson and his friends enter a labyrinthian hall occurs as a documentary scene, and that the reader does not have access to this footage, but instead, to the character Zampanò's textual mediation of the cinematic moment. Zampanò's text is also accompanied by 1) his footnotes on this labyrinth event, 2) the character Johnny Truant's footnotes on Zampanò's text, and 3) a comic book depiction of the scene in the appendices of the novel. To organize and encode the labyrinth event with a document-oriented NoSQL database, one possibility would be to list all of these mediations under the key of “labyrinth” and also their associations with the novel's figurative themes and metaphors (see **Figure 6**). This database could thus be set up so that a search

for “labyrinth metaphor” could return the values “haunting,” “monstrosity,” and “uncanny” so that a reader can piece together these literary associations. The reader could also search for “labyrinth text” to discover the other ways in which the hallway scene is represented in the novel: “Zampano’s manuscript,” “Zampano’s footnotes,” “Johnny’s footnotes,” and “comic.”

This is only one example of the way that a paradigmatic approach to content management and representation can better account for figurative meaning in literary narrative, especially vis-à-vis digitally informed shifts to how we think of narrative creation, creativity, and engagement. Mark Z. Danielewski’s current endeavour *The Familiar* is a proposed 27-volume experimental book project that functions as an imagined layering of different characters over different times. It would be very difficult to represent this feat of temporal relativism in a digital humanities project or arguably even in traditional literary hermeneutics without further consideration of how it has been influenced by contemporary models and computational models of intertextual and networked content organization.

Feedback Loops: Between the “Known” and the “Unknown”

In academic discourse as with digital tools, variously flexible methods of representing shifting concepts of narrative play and negotiating them with our expectations of literary form, genre, and convention can be posed as alternative modes of creative and critical inquiry, particularly in the next steps of the (digital) humanities’ project.

```
{_id: "labyrinth"  
event: "documentary scene"  
text : "Zampano’s manuscript", "Zampano’s footnotes", "Johnny’s footnotes", "comic",  
reference: "labyrinth myth", "minotaur", "Ariadne", "twine",  
figurative: {  
  "theme": "non-linearity", "intertextuality", "multi-dimensionality", "disorientation",  
  "haunting", "uncanny",  
  "metaphor": "haunting", "monstrosity", "uncanny"  
}  
}
```

Figure 6: Document-oriented database containing the key “labyrinth”.

By this, I do not mean to imply that the digital can match or account for all aspects of the literary; the separate togetherness of media-specific analysis and also of media-specific analytical affordances initiated my inquiry into the narrative and database in the first place, and also encouraged me to draw upon Hayles' proposal of more synergistic approaches that resemble a "recursive feedback loop" between the digital humanities and the traditional humanities (2012, 32).

Where we might move on from here is a return to a question posed in the introduction—a consideration of what aspects of the humanities may be at risk in the use of tools and text forms with distinct epistemological worldviews. After pondering over this by focusing on figurative meaning, I also find value in inverting the question. What aspects of the humanities and the literary might be upheld through the dynamic of a recursive feedback loop between the humanities and the digital, the narrative and the database?

For one, reflexive observations reveal that research questions, surprises, and limitations are an origin or catalyst for a recursive feedback loop, as they necessitate a back-and-forth negotiation between what functions well (what we "know") and what asks us to pause and think (what we do not "know"). The ultimate gesture of this negotiation might be understood in the vein of what Alan Liu calls "the ethos of the unknown"—a political mode that is rooted in humanistic philosophy (and also advocacy for this philosophy) by way of varying degrees of the critical infiltration, hacking, and implosion of systems of post-industrial information culture (2004b, 9; 2004b, 294). It is such post-industrial systems that help to shape the expansive scope and rigidity of computational data management in the name of efficiency, function, and performance. Machines and databases prefer axioms, answers, and the determinate over reflections, interpretations, and the indeterminate, bringing this paper full circle in its consideration of the place of humanistic and narratological thinking amongst the digital.

Machines are designed to function as asked, such that they are meant to resist the indeterminate. This paper has sought to study their limitations in a way that is not meant to move in this direction of quantifying the figurative and determining the indeterminate; rather, it moves towards database forms (and content

representation forms) that are more ludic like language and like all the things we can say in a single word, a phrase, a look, a light. I have sought to highlight the value of the indeterminate and the unknown in necessitating an ongoing comparison, juxtaposition, interpretation, and reflection of tools and work. In between the known and unknown—or, what is in between that which functions as defined, rigid, and expected and that which requires us to ponder, interpret, critique, remember, and return—is the weight and wait of the question, where even that which is determined can start to fall apart.

Acknowledgements

Thank you to computer scientist Andrew Qu Yang for his suggestions on alternative database forms and content management in the research stage of this paper.

Competing Interests

The author has no competing interests to declare.

References

- Bal, Mieke.** 2009. *Narratology: Introduction to the Theory of Narrative*. Toronto; Buffalo, NY: University of Toronto Press.
- Danielewski, Mark Z.** 2000. *House of Leaves*. New York, NY: Pantheon Books.
- Dourish, Paul.** 2014. "No SQL: The Shifting Materialities of Database Technology." *Computational Culture: A Journal of Software Studies* 4: n.p. Accessed April 3, 2015. <http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology>.
- Fitzpatrick, Kathleen.** 2002. "The Exhaustion of Literature: Novels, Computers, and the Threat of Obsolescence." *Contemporary Literature* 43(3): 518–559. DOI: <https://doi.org/10.2307/1209111>
- Goldsmith, Kenneth.** 2011. "Revenge of the Text." *Uncreative Writing*. New York, NY: Columbia University Press.
- Hayles, N. Katherine.** 2004. "Print is Flat, Code is Deep: The Importance of Media-specific Analysis." *Poetics Today* 25(1): 67–90. DOI: <https://doi.org/10.1215/03335372-25-1-67>

- Hayles, N. Katherine.** 2008. "Future of Literature: Print Novels and the Mark of the Digital." In *Electronic Literature: New Horizons for the Literary*. Notre Dame, IN: University of Notre Dame.
- Hayles, N. Katherine.** 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago, IL: University of Chicago Press.
- Herman, David.** 2009. *Basic Elements of Narrative*. Chichester, UK; Malden, MA: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444305920>
- Jockers, Matthew.** 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Liu, Alan.** 2004a. "Transcendental Data: Toward a Cultural History and Aesthetics of the New Encoded Discourse." *Critical Inquiry* 31: 49–84. DOI: <https://doi.org/10.1086/427302>
- Liu, Alan.** 2004b. *The Laws of Cool: Knowledge Work and the Culture of Information*. Chicago, IL: University of Chicago Press. DOI: <https://doi.org/10.7208/chicago/9780226487007.001.0001>
- Liu, Fangzhou, and Hannah Knowles.** 2017. "Harassment, Assault Allegations against Moretti Span Three Campuses." *Stanford Daily*. November 16, 2017. <https://www.stanforddaily.com/2017/11/16/harassment-assault-allegations-against-moretti-span-three-campuses/>.
- Manovich, Lev.** 1999. "Database as a Symbolic Form." *Convergence: The Journal of Research into New Media Technologies* 5(2): 80–99. DOI: <https://doi.org/10.1177/135485659900500206>
- Manovich, Lev.** 2001. *The Language of New Media*. Cambridge, MA: MIT Press.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden.** 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331: 176–182. DOI: <https://doi.org/10.1126/science.1199644>
- Moretti, Franco.** 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London, UK; New York, NY: Verso.

- Moretti, Franco.** 2011. *Network Theory, Plot Analysis*. Literary Lab, Stanford University. Accessed March 7, 2015. <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Morris, Adalaide.** 2006. "New Media Poetics: As We May Think/How to Write." In *New Media Poetics: Contexts, Technotexts, and Theories*, edited by Adalaide Morris, and Thomas Swiss. Cambridge, MA: MIT Press.
- Pressman, Jessica.** 2006. "House of Leaves: Reading the Networked Novel." *Studies in American Fiction* 34(1): 107–128. DOI: <https://doi.org/10.1353/saf.2006.0015>
- Pressman, Jessica, Mark C. Marino, and Jeremy Douglass.** 2015. *Reading Project: A Collaborative Analysis of William Poundstone's Project for Tachistoscope {Bottomless Pit}*. Iowa City, IA: University of Iowa Press.
- Ryan, Marie-Laure.** 2003. "On Defining Narrative Media." *Image and Narrative* 6: n.p. Accessed March 27, 2015. <http://www.imageandnarrative.be/inarchive/mediumtheory/marielauryan.htm>.
- Tabbi, Joseph, and Michael Wutz.** 1997. "Introduction." In *Reading Matters: Narrative in the New Media Ecology*, edited by Joseph Tabbi, and Michael Wutz. Ithaca, NY: Cornell University Press.

How to cite this article: Fan, Lai-Tze. 2018. "On the Value of Narratives in a Reflexive Digital Humanities." *Digital Studies/Le champ numérique* 8(1): 5, pp. 1–29, DOI: <https://doi.org/10.16995/dscn.285>

Submitted: 29 October 2017 **Accepted:** 29 October 2017 **Published:** 27 March 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Digital Studies/Le champ numérique is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS The Open Access icon, a stylized 'O' with a person inside, representing open access.