

Preface

The digital humanities, putatively the intersection between the humanities disciplines and computation, was popularized in the early 1990s. In the intervening decades, the “digital humanities” has not yet settled on a defined self-identity. One indicator of this is that of the dozens of “DH” manifestos on the Web; they all have differing and competing visions for the field. Another indicator is the rich variety of work being done under these auspices that does not fall into simple summaries and descriptions. The Digital Humanities Manifesto 2.0, which originated from nine seminars co-taught by Jeffrey Schnapp and Todd Presner, and was released by the UCLA Mellon Seminar in Digital Humanities, reads in part:

Digital Humanities is not a unified field but an array of convergent practices that explore a universe in which: (a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and (b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences. The Digital Humanities seeks to play an inaugural role with respect to a world in which universities—no longer the sole producers, stewards, and disseminators of knowledge or culture—are called upon to shape natively digital models of scholarly discourse for the newly emergent public spheres of the present era (the www, the blogosphere, digital libraries, etc.), to model excellence and innovation in these domains, and to facilitate the formation of networks of knowledge production, exchange, and dissemination that are, at once, global and local” (“The Digital Humanities Manifesto 2.0,” 2009).

DH work is additive and contributes to learning and methods in a range of extant fields. It is about hybrid mash-ups that pull from analog and digital sources and that give back digitally. There are combinations of research and analytical methods, drawing from quantitative and qualitative methods, going Web scale and small scale, employing a range of frameworks, and combining various disciplines, often through collaborative and distributed teams.

Deconstructing “DH”

The “digital” part of the “digital humanities” involves a range of technologies: the Web and Internet, mobile devices, social media, databases, digital curation platforms, geographical mapping tools, social mapping tools, linguistic analysis software programs, data analytics tools, and others. The “humanities” focus points to a combination of fields: the classics, literature, philosophy, religious studies, psychology, modern and ancient languages, culture studies, art, history, music, philosophy, theater, film, political science, geography, anthropology, linguistics, social work, communications, and others. For all the technological focuses, the DH work is about people and for people through the ages and for the ages. Every work, no matter how technological, contains the human touch and is conceptualized for human benefit. The works are human-readable, in general, and are packaged as stories and understandings. The works are often aesthetically dazzling: think interactive webscapes, data visualizations, 3D virtual immersive spaces, and network diagrams.

Likewise, one can ask what “data” in the digital humanities is. Well, the general opinion is that everything has some informational value. The world is informatized, and DH researchers themselves are not only researchers and elicitors of insights, but they are in-world participants making the world and information simultaneously. The information is viewed through a subjective and interpretive lens, but that interpretation is backed up by more traditionally quantitative (statistical) and computational (machine learning) methods. The work is not seen as neutral nor theory-free not independent of its contributors.

The challenge is in how to turn a “thing” in the world into something with informational and social value. This data may seem at once a range of contradictions: ephemeral and fragile and yet permanent and perpetual, subjective and objective, created for the local but also the global, inclusive of machine-scale distant reading methods as well as close human reading.

The Spirit of DH

Often, the spirit of the digital humanities is countercultural; it fights against extant power structures and norms. The DH is often value-informed and socially active, in the pursuit of social justice. DH practitioners take a questioning stance because the world is never quite as it seems. There are elusive and latent interpreted truths to be discovered or (co)created.

While its core spirit is of changing the current order, at the same time, DH work requires ever higher levels of expertise in content fields and technological methods. The work is anti-Establishment but requires some of the expertise of Establishment and accrued skills from years of development. Virtuosity in the digital humanities simultaneously requires the fiery rebel spirit and the iciness of acquired expertise. The work is often inclusive of nonexperts and the online crowds, who are sourced

for their contributions and voices and concerns in rich ways. It seems fitting that in this hypersocial age that such broadscale collaborations exist.

The spirit of DH is of experimentation along the entire work chain: theorizing and conceptualization, research, data collection, content curation, data processing, data analytics, and often open publishing (of digital corpora and collections, of virtualized experiences, of publications, and of multimedial presentations). The order of the work is not necessarily linear but recursive; for some projects, conceptually and somewhat practically, the work is continuous, ongoing, and perpetual. While some methods and technologies are borrowed from a range of fields, digital humanists have applied experimental means to every step of the process as well, with new technological platforms and tools and fresh methods. Problem-solving is achieved on-the-fly, often with the direct support and help of Web developers, coders, server administrators, data scientists, computer scientists, and librarians. This is about an artist and artisan sensibility with code.

If one asks practitioners what the digital humanities is, there is a range of answers. A common response is to go to the work, to explore it in depth, and to acquire a more inductive sense of the answer. Certainly, as DH rose from a groundswell of practice and self-definition, it may be best to understand it from a bottom-up instead of a top-down way. While practitioners are aware of each other's work, the focus is not on conformance to common practice but rather diversity. The general consensus is that DH is still emerging as an approach. The feeling is that the work is exploratory and provisional and open to re-interpretations and re-visionings.

About This Book

Data Analytics in Digital Humanities was over a year in the making, with various individuals and authoring teams conducting their research and writing it up, going through double-blind peer reviews, revising, and finalizing their work. Here, the respective authors describe their data analytics work in the digital humanities.

The book is comprised of six parts:

Part I: Design of Representational Systems

Part II: Text Capture and Textual Exploration

Part III: Engaging Social Data

Part IV: Applied Technologies for Data Analytics

Part V: Sense-Making in the World

Part VI: Support for Digital Humanities Work

The book is not comprehensive by any means. There are so many live projects and endeavors that each author and team really only have a limited perspective on the whole.

In Part I, “Design of Representational Systems,” there are two works that describe data labeling. In “Semantic Web for Cultural Heritage Valorization,” a research team proposes a set of ontology modules (Cultural-ON) to model the

classification of cultural heritage data, given the heterogeneous nature of cultural artifacts, both analog and digital, tangible and intangible. The authors of Chap. 1 include Dr. Giorgia Lodi, Luigi Asprino, Andrea Giovanni Nuzzolese, Dr. Valentina Presutti, Dr. Aldo Gangemi, Dr. Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini.

Dr. Gian Piero Zarri offers an in-depth description of Narrative Knowledge Representation Language (NKRL) and its use to describe narratives as logical associations based on “elementary events.” The careful specifications enable its use conceptually and computationally to capture and represent the elements that comprise narratives. Zarri’s chapter is titled “Using the Formal Representations of ‘Elementary Events’ to Set Up Computational Models of Full ‘Narratives’” (Chap. 2).

The second part, “Text Capture and Textual Exploration,” features two works related to text analytics. Joshua L. Weese, Dr. William H. Hsu, Dr. Jessica C. Murphy, and Dr. Kim Brillante Knight describe a machine learning-based classifier they developed to identify parody, an elusive challenge. Their work is titled “Parody Detection: An Annotation, Feature Construction, and Classification Approach to the Web of Parody” (Chap. 3).

Dr. Michael Percillier, in Chap. 4 “Creating and Analyzing Literary Corpora,” describes a methodical process using Python and other tools to collect and process literary corpora based around a particular topic. He includes Python code to demonstrate efficient automated means to build text corpora.

Part III, “Engaging Social Data,” offers two works summarizing “social” data. The first is about information from online social networks, and the second about the social aspects of learning.

Davide Di Fatta and Roberto Musotto apply integrated sentiment analysis (iSA) to online social networks in Chap. 5, “Content and Sentiment Analysis on Online Social Networks (OSNs).” They consider various ways to apply their research insights to practical applications, like marketing through social media.

In Chap. 6: “The Role of Data in Evaluating the Effectiveness of Networked Learning: An Auto-Ethnographic Evaluation of Four Experiential Learning Projects,” Jonathan Bishop employs a light auto-ethnographic approach to study the role of data in evaluating networked learning across four experiential learning projects. One central question is whether there are more effective designed electronic stand-ins for educator-learner interactions to promote learning.

In the fourth part, “Applied Technologies for Data Analytics,” one work focuses on computational linguistic analysis based on psychological features in texts. Another focuses on capturing research insights from related tags networks.

In Chap. 7, “Psychological Text Analysis in the Digital Humanities,” Ryan L. Boyd describes insightful uses of the Linguistic Inquiry and Word Count (LIWC) tool for text analysis. His long-term uses of LIWC and direct contributions to the tool’s co-development (in LIWC2015) make him a singularly appropriate author for this excellent work. Boyd argues that computational linguistic exploration of psychological aspects in language is an untapped and highly promising area of research.

In “Parsing Related Tags Networks from Flickr to Explore Crowd-Sourced Keyword Associations” (Chap. 8), Dr. Shalin Hai-Jew describes how crowd-based folk tags applied to digital imagery on a content-sharing site may be used for creating collective mental models of large-scale phenomena.

Part V is about “Sense-Making in the World.” In this part, Dr. Cobi Smith describes the use of technological methods to harness crowd-sourced imagery during natural disasters. The methods and tools described in this chapter, “A Case Study of Crowdsourcing Imagery Coding in Natural Disasters” (Chap. 9), may have broad applications in the digital humanities.

Dr. Glenda Alicia Leung’s “YouTube Comments as Metalanguage Data on Non-standard Languages: The Case of Trinidadian Creole English in Soca Music” (Chap. 10) suggests new analytical applications of interaction comment data related to shared, social videos. This work captures the energy and power of nonstandard lived language in connecting people on Google’s global video-sharing platform.

Chapter 11 is about “Creating Inheritable Digital Codebooks for Qualitative Research Data Analysis.” Here, Dr. Hai-Jew describes the research importance of sharing digital codebooks, particularly with the popularization of Computer Assisted Qualitative Data Analysis (CAQDAS) tools, and she offers firsthand insights on effective processes for developing and sharing such digital codebooks. If researchers are to leave a legacy of their unique coding “fists,” then computational means may offer a more convenient and efficient way of transfer.

In Part VI, “Support for Digital Humanities Work,” one author highlights supports for DH work. Hannah Lee, in “Is it Worth It? The Library and Information Science Degree in the Digital Humanities” (Chap. 12), argues for the importance of library and information sciences (LIS) to practically support the work in the digital humanities.

So what is the state of data analytics in the digital humanities? Based on these collected works, it is a rich and evolving one, driven by local research needs and some global ones. DH practitioners harness technologies to complement and augment the human abilities of perception, expression, analysis, and memory. The technologies used range from open-source and free to closed-source and proprietary, and these tools are cobbled in creative and complex sequences for understanding and analysis. The skill comes not only in the applied techniques but also in the insights surfaced and the sharing of the applied and innovative techniques. While some data analytics results are “reproducible” and “repeatable” based on computational means, the assumptions underlying the DH research and data analytics are very much drawn from qualitative data analytics:

that all human phenomena have potential informational value, depending on researcher perspective and context (and vision and skill);
 that human researchers are wrapped in their own subjectivities, for better and for worse, and benefit from deeper self-understandings through practiced reflection;
 that all data are filtered through subjective lenses and self-informed understandings;
 that data “measures” are limited, imprecise, conditional, and contested (and yet still are insightful);

that the researchers are part and parcel of the research and inform the research and the data findings and their applications;
that the communal context for the humanities is a critical part of the work (as research participants, data consumers, and researchers);
that DH research is subsumed to human needs, interests, and values;
and that social justice and practitioner ethics apply to every context.

There are inherent digital humanities truisms about data as well. For example, writings considered classically “fictional” contain the seeds of truth, and traditionally non-fictional works may not have as much truth as advertised. In this light, there are critical interrogations to be made, to understand the multihued shades between truths. Data are extracted in fresh ways, with manuscripts tagged and mapped for patterns (linguistic, spatial, psychological, cultural, and others), dialogs mined for insights, genders explored, and cultural practices probed, across and through time. All human residua contain substance that may be interpreted and informatized for new ways of seeing, feeling, and being.

DH cartographers describing data analytics in the digital humanities are in the early phases of defining and mapping this space. In theory and in practice, there are numerous other potentials that have yet to be explored, applied, and shared.

Data Analytics in Digital Humanities offers an early look at this topic, with hand-sketched DH data analytics “maps” of fine granular detail for particular defined needs but which does not yet have the cardinal directions defined or accepted compass roses.

Manhattan, KS, USA

Shalin Hai-Jew

Data Analytics in Digital Humanities

Hai-Jew, S. (Ed.)

2017, XXII, 295 p. 37 illus., 23 illus. in color., Hardcover

ISBN: 978-3-319-54498-4