

What makes digital humanities, digital?

James O’Sullivan

School of English, UCC

To be modern is to find ourselves in an environment that promises adventure, power, joy, growth, transformation of ourselves and the world — and, at the same time, that threatens to destroy everything we have, everything we know, everything we are.
(Marshall Berman)

A Myriad of Terminology

While not quite a neologism at this point, the term “digital humanities” for some still bears a significant measure of ambiguity. What separates digital humanities from the humanities? Throughout this article, I will attempt to offer some clarity on this separation, outlining what it is that makes digital humanities, digital.

The field of scholarship now recognised as the digital humanities has not always held this particular mantle. Initially, this emerging discipline was referred to as “humanities computing”, a term that gathered momentum as early as the late ‘70s, the evidence for which can be found in a quick n-gram of Google Books. N-grams offer an approach to probabilistic language modelling that can be used for a variety of purposes, in this case, to identify the frequency of a sequence of words in a set of texts.

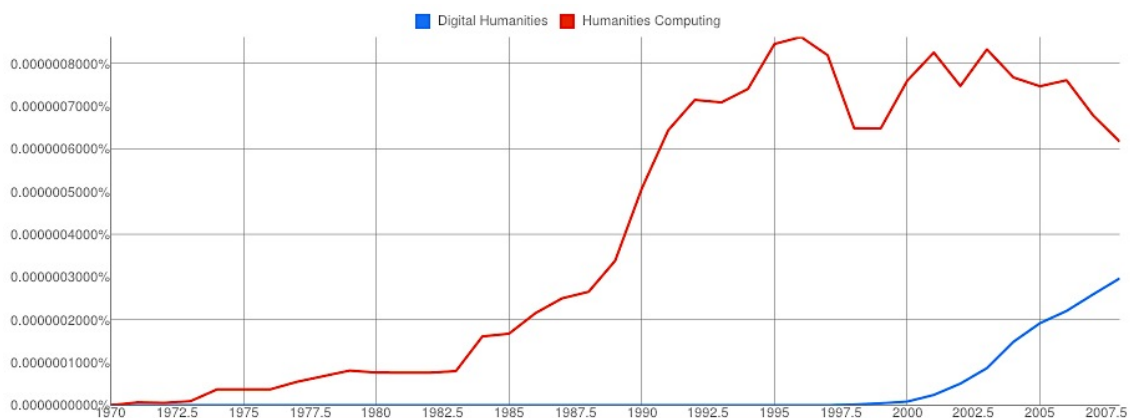


Figure 1: : Digital Humanities vs. Humanities Computing

Google Ngram Viewer is not a scholarly tool appropriate for research, but it is a good way of confirming any suspicions that you might have in relation to historical trends in public interest. One must take into consideration that Ngram Viewer currently only searches

up to 2008, and there has been a significant increase in the volume of literature within this particular field since then. Consequently, it is fair to say that this does not paint an entirely accurate picture. What it does show however, is that there was an explosion of interest in using technology to support humanist activities in the late 1980s, and that this continued right up until the turn of the century, at which point the growth in activity normalised somewhat. From the outset, the discipline was called humanities computing, and it wasn't until 2005 that the label of digital humanities emerged. In 2008, the former of the two terms was still very predominant, but the graph displayed in Figure 1 suggests that this trend was shifting; I'd very much doubt if there hasn't by now been a full reversal of those terms.

Many scholars and practitioners continue to use these two terms interchangeably, while others debate the legitimacy of applying the word "digital" across a broad spectrum of activities, many of which are focused on physical electronics and analogue technologies. As a result, those working within the discipline have devised and settled upon more specific terms that they feel reflect the particulars of their research with greater clarity. At any conference predominantly populated by digital humanists, you may find yourself traversing a myriad of terminologies: electronic scholarship, digital scholarship, computational analysis, electronic textual analysis, digital curation — and so forth. Essentially, all of these terms refer to academic pursuits that are concerned with the same thing: the place of technology in the arts and humanities.

Technology and the Humanities

Delineating the place of technology within the very many realms of humanistic study is a complex task. Technology offers the potential for new methodologies, augmenting traditional humanist research methods with empirical analysis that can be used to assist interpretation. Beyond its contribution to new methodologies, technology has been the driver behind complex shifts in the transmission and reception of cultural artefacts; in the creation of new knowledge, and in the exploration of old meaning. Technology has not just presented new modes with which humanists can interrogate and re-problematise old questions, but has itself given rise to a whole range of emerging issues that warrant deliberation. Let us consider each of these in isolation, beginning with the new methodologies being deployed by digital humanists.

My own research is partly concerned with computational stylistics, a branch of electronic textual analysis that seeks to use computers to find trends in literature. Computational stylistics combines statistics with linguistics in an effort to form authorial signatures which can be used to establish such trends, from which the interpretation of a text may then emerge. The key difference between this methodology and more traditional approaches is that the subsequent interpretation is based on an analysis that is supported by empirical

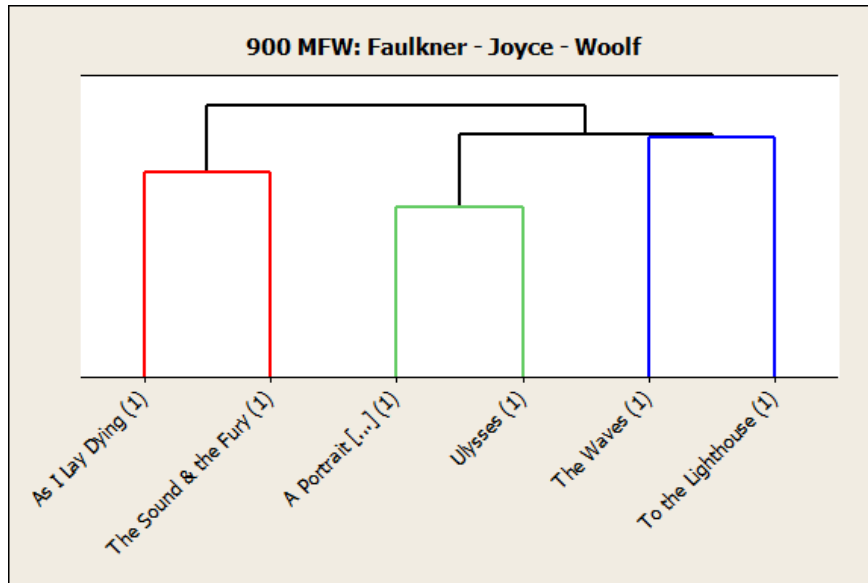


Figure 2: A cluster analysis of Faulkner, Joyce & Woolf

statistical findings. Conducting textual analysis in this fashion balances objectivity with subjectivity, offering additional evidence to interpretations based on close readings.

Electronic textual analysis can be used across all disciplines, but in computational stylistics, the focus is very much on literary texts, and the specific style used by an author in particular passages. One may be interested in authorship attribution, in which case a large sample from a writer's known corpus may be used to form an authorial signature to be tested against an unknown text, or a known text whose authorship is disputed. Alternatively, the focus may be on character dialogue, and the success with which an author is able to shape distinctive voices within their narrative. The limits are boundless, with some remarkable studies having already been conducted and published. Within computational stylistics, there are a range of statistically valid methods that can be used to examine a body of text, one of which is cluster analysis.

For illustrative purposes, I have produced a very simple cluster analysis, as shown in Figure 2, using a small sample of texts by the prominent modernist authors, William Faulkner, James Joyce and Virginia Woolf. The first step in this process was the creation of a most-frequent-wordlist, which can be produced using a variety of computational methods. Using the 900 most frequent words in each text, an authorial signature was formed, from which three distinct clusters were identified in a computer-generated dendrogram. Dendrograms illustrate hierarchical clustering, and thus can be used to measure the similarity between groups of texts. Reading the dendrogram in Figure 2 is relatively straightforward: the shorter the path between two titles, the more similar the authorial signature contained within these texts. For clarity, I instructed the system to colour-code the clusters. As can be seen, the computer has clustered the texts correctly, placing the two novels from each author within the same cluster, demonstrating that this method is entirely accurate. In-

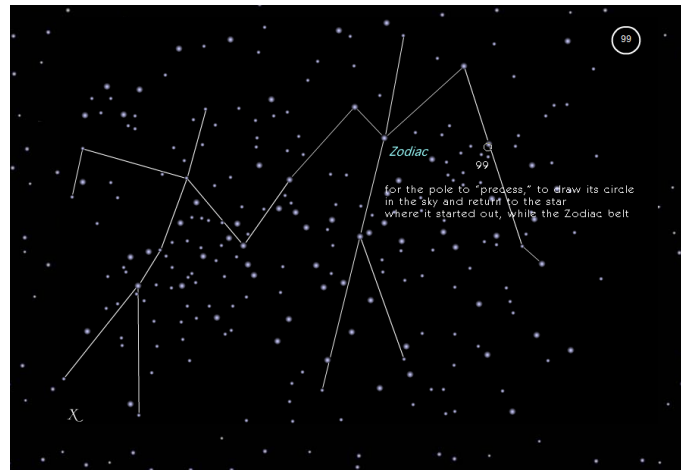


Figure 3: : Stephanie Strickland's *Vniverse*

identally, this data also shows that the novels by Joyce and Woolf are more similar in style than the Faulkner works. To reiterate, this is indicated by the proximity between the Joyce and Woolf clusters, although the difference is marginal. It is worth noting that, as this analysis was conducted for purely demonstrative purposes, the integrity of the textual data was not thoroughly checked.

As already noted, the study of digital humanities goes beyond the pursuit and use of new technology-based methodologies. Juxtapositions between the “old” and the “new” have caused complex shifts in scholarly consumption and artistic creation. Literature has experienced just such a cultural upheaval with many traditional literary practices having been merged with new media in an effort to achieve new forms of writing, style and textual construction. Consider Stephanie Strickland's *Vniverse*. Comprised of both a physical and electronic text, Strickland disregards the traditional separation between these differing media, using both forms in complement to generate a multiplex of reading possibilities.

Strickland's work is a textual construction that is not typically encountered in traditional literary practice: the reader's interaction determines the sequence of the piece. The text of the poem is either displayed based on the user's entry of a numerical value, which corresponds to a portion of the poem, or alternatively, on where the cursor is placed amongst the star-scape that appears on the screen, where constellations form, providing imagery alongside the text. Technology has provided authors with new aesthetics from which a new genre of electronic literature has emerged. The study of this literature is of particular interest to digital humanists, who hold an interest in both the electronic and literary elements that form such works.

What makes digital humanities, digital?

Let us return at last to the question at hand: what separates digital humanities from its traditional counterpart? Technology has long played an important role within the humanities. Scholars have been availing of increasingly feature-rich text processors, bibliographic software — even the printing press was once considered an advance in technology — the codex itself was once a form of new media. If such is the case, then why has this field of scholarship seemingly only recently emerged as a discipline? This might be attributed to ubiquity and sophistication. Technology has become ubiquitous, supporting the great majority of those activities that form our everyday lives. This exponential growth in the presence of electronic resources with rapidly progressing intricacies from which we can draw has led to change that requires assessment. It is on such explorations that the focus of digital humanists are centred. Digital humanities is more than the use of technology to display research findings in electronic form. Rather, it resides at the juncture between complex or novel uses of new media and traditional humanities research and artistic endeavour. It is concerned with the use of technology to reproblematised humanist questions, or oftentimes, the exploration of technology from humanist perspectives.

Digital humanities does not seek to replace the traditional fields that humanists have occupied for centuries, but rather, facilitate the study of how our new environment has impacted on such, and analyse the products of these transformations.

James O'Sullivan is a doctoral candidate on the PhD in Digital Arts & Humanities at UCC. James works in the school of English, under the supervision of Professor Graham Allen and Dr Orla Murphy. He would like to acknowledge the on-going contributions of his supervisors to his research, as well as the School of English and College of Arts, Celtic Studies and Social Sciences as a whole. This research is generously funded by the Higher Education Authority, under the Programme for Research in Third Level Institutions, Cycle 5. Further information on James' work and research interests can be found at josullivan.org.