Book Section:

[This is a postprint (final version, after review) chapter published in *Journalism and the Periodical Press in Nineteenth-Century Britain*, ed. by Joanne Shattock (Cambridge: Cambridge University Press, 2016). Please cite the published version]

Beyond the 'Great Index': Digital Resources and Actual Copies
James Mussell

In August 1886 W.T. Stead published an interview with Samuel Palmer, the compiler of *Palmer's Index to the Times Newspaper*, in the *Pall Mall Gazette*.[1] Stead, who had begun his own half-yearly index to the *Pall Mall Gazette* in 1884, was deeply interested in the archival status of old newspapers and periodicals. Stead recognized the value of the newspaper as record of the moment but also understood the difficulties in recovering the information in its pages. In the preface to the interview he acknowledged the importance of Palmer's ongoing index to *The Times*, remarking that 'the materials for a great part of our history lie embodied in its columns'. However, using an image he would later draw upon when setting out the case for both his *Review of Reviews* (1890-1936) and *Annual Index to the Periodical Literature of the World* (1891-1900), Stead claimed that without an index such as *Palmer's* the 'future historian […] would be as helpless as a Theseus wanting his clue to the Labyrinth'.

Like Palmer's *Index*, digital resources provide just such a clue. However, whereas previous attempts to address the field's bibliographical complexity, whether published in the nineteenth century like *Palmer's Index* and Stead's various endeavours or more recent projects such as the *Wellesley Index* (1966-79) and *Waterloo Directory* (1977; 1997; 2003),

---

[1] Anonymous, 'A Great Index', *Pall Mall Gazette*, 14 August 1886, 11.

have produced discrete publications that have stood outside of the archive to describe its contents, digitization has exerted an unprecedented degree of bibliographic control by changing paper and ink into bits and bytes. This transformation allows digital resources to serve as indices, where every word might be searched. With sufficient metadata, they can also become catalogues, enumerating articles, issues, volumes and publications.[2] Finally, they are also a kind of edition, reproducing aspects of the printed material in digital form.[3] Since at least the appearance of Gale Cengage's *Times Digital Archive* in 2003, the study of newspapers and periodicals has been structured by digital resources that combine all three of these functions at unprecedented scale.[4]

As I will go on to argue, searchable digital resources offering access to page facsimiles are not the only digital projects relevant to the study of the press; equally, not all resources are the same. Nonetheless, it is this type of digital resource – in which a textual transcript is used to index content – that has come to dominate research into historical newspapers and periodicals. Never before has the nineteenth-century press been so accessible: with a few taps on the keyboard, millions of pages of newspapers and periodicals can be viewed on the desktop over the web (depending, of course, on hardware, software, and the relevant subscriptions). The challenge is to place digital resources in longer traditions of indexing, cataloguing, and editing while recognizing what makes them distinctive. After explaining his

---

[2] For a nineteenth-century account of the distinction between indices and catalogues, see Andrea Crestadoro, *The Art of Making Catalogues of Libraries* (London: The Literary, Scientific and Artistic Reference Office, 1856).

[3] For editions and archives, see Kenneth M. Price, 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?', *Digital Humanities Quarterly*, 3 (2009), 1–40; James Mussell and Suzanne Paylor, 'Editions and Archives', in *The Nineteenth-Century Press in the Digital Age* (Basingstoke: Palgrave, 2012), pp. 114-148.

[4] Important predecessors are *Making of America* (1995), *Internet Library of Early Journals* (*ILEJ*, 1999), Heritage Newspapers' *newspaperARCHIVE* (1999), ProQuest's *Historical Newspapers* (1999), *Periodicals Contents Index* (1995) and *Periodicals Contents Index Full Text* (2000).

methods, Palmer took Stead's reporter upstairs, 'leading the way to a large, light room, surrounded by unvarnished deal shelves, giving shelter to a good many hundred volumes, while at one end, occupying the whole space, stood a great array of *Times* volumes, looking appallingly large and bulky beside the octavos and duodecimos surrounding them.'[5] Here, the forbidding bulky volumes represent the pure source, the raw authentic material of history. Palmer's indices leave these 'actual copies' untouched, referring readers back to the newspaper and so reinforcing the iconic solidity of *The Times's* bound volumes. Digital resources, however, transform the actual copies into something else, referring readers to new digital images rather than old printed pages. In what follows, I consider how digital resources stand in relation to the copies of newspapers and periodicals in the archive. Predicated on a radical transformation, digitization will always fail to reproduce print; however, such transformations make evident that what we think constitutes print and print culture is more complex than it appears. It is in the discursive space opened up by digitization that we can see print anew.

Margins

Indices such as Palmer's were necessary because the very thing that made newspapers so valuable also made them difficult to use. Attuned to the moment with space to fill, newspapers combine disparate materials in timely fashion. Not only do they record major public affairs – the news as commonly understood – but set this alongside other content thought interesting to its readers. As serials, this content is always provisional, relevant only to the period before the appearance of the next number. As a result, old newspapers record

---

[5] 'A Great Index', 11.

moments intended to be superseded and forgotten, allowing readers in the present to revisit a past that has passed.

Stead, whose archival imagination complemented his journalistic instinct, recognized the value of an index to *The Times* but was reluctant to accord the newspaper any special status. Imagine, he asks, 'how many people would welcome an index to "Hansard" or the "Annual Register!"' A few months previously, Stead had argued that the newspaper's claim to be a democratic instrument rested on more than its coverage of politics or public affairs. In his well-known essay 'Government by Journalism', published in the *Contemporary Review* in May 1886, Stead argued the newspaper better expressed the public will because it was closer to the people. This was partly a function of its periodicity, with 'the editor's mandate' renewed 'day by day' and the franchise extended to all who can make 'a voluntary payment of the daily pence'. But it was also because the newspaper better represented the interests of its readers. The editor must, 'often sorely against his will, write on topics about which he cares nothing, because if he does not, the public will desert him for his rival across the street.' A successful newspaper 'must "palpitate with actuality"', a kind of realism that binds the publication into the lives of its readers.[6]

In many ways, the newspaper's importance is the result of its supplementarity. Palmer, certainly, understood the appeal of *The Times* lay in its compendious sweep of information as much as its lauded leading articles. In his interview with the *Pall Mall Gazette*, he recounted a visit from a policeman who needed to consult back issues of the paper to determine whether or not a prisoner in his custody had prior convictions. 'It will be a matter of surprise to many', remarks Stead's reporter, 'that *Palmer's Index* contains the only register of criminal

---

[6] Stead, W.T., 'Government by Journalism', *Contemporary Review*, 49 (1886), 653–74, 655.

cases in existence, the same applying to "Suicides", "Inquests", and "Accidents".[7] Coverage

of the courts was a staple of all sectors of the newspaper press in the period, guaranteeing

regular content that was reliably sensational. *Palmer's Index* was the only record because it

indexed a long-running newspaper, not *The Times* in particular. Newspapers needed to fill

white space, often already organized into predetermined departments, and they attuned

themselves to newsworthy sources, whether formal such as the courts or Parliament (when in

session), or informal such as society gossip or the queries of their readers. Not quite records

of the everyday, the need to provide copy of interest to its readers meant that the newspaper

nonetheless recorded information that escaped more formalised documentary practices, even

the otherwise impressive Victorian civic bureaucracy.

The Times might have supplemented more formal modes of record keeping, but this

supplementarity did not make it marginal. In his landmark essay 'Charting the Golden

Stream', Michael Wolff argued that the Victorian press – newspapers and periodicals –

constitute 'repositories of the general life of Victorian England'.[8] The word 'repository'

might make the newspaper appear passive, simply documenting what was going on, but

archiving is an intentional act (one must choose what gets put in the repository, and how) and

so too is the productions of newspapers and periodicals. While they do have a documentary

function, the content of newspapers and periodicals results from a number of authorial,

editorial, and publishing decisions calibrated for a complex and changing market. As Wolff

argues, it is the dynamism of that market that makes the press such an important resource.

Not only was this, as Linda K. Hughes has argued, the 'first mass media era', but the market

was diverse and fluid, allowing publications to target particular niches or, more commonly, to

---

[7] 'A Great Index', 11.

[8] Michael Wolff, 'Charting the Golden Stream: Thoughts on a Directory of Victorian Periodicals', *Victorian Periodicals Newsletter*, 4 (1971), 23–38, 26.

try and maintain an appeal to multiple readerships simultaneously.[9] Wolff's claim that the individual issue is 'the basic unit for the study of Victorian cultural history' is quite correct – no other cultural documents from the period represent its diversity quite so fully – but such a history will always be inflected by that issue's mediating function.[10] The nineteenth-century press is not a simple reflection of nineteenth-century society, but rather one of its products.

Digitization brings the press in from the margins. As Stead's account of *Palmer's Index* makes clear, the bibliographical condition of the newspaper archive was unwieldy even in the nineteenth century. Today, it is even more so. Not only is the archive incomplete (some publications were more likely to be kept than others) but it is also widely and unevenly dispersed. Of what survives, both newspapers and periodicals are likely to be in the form of bound volumes, with supplementary material and, in the case of periodicals (and many weekly newspapers), advertisements removed. Although consolidated online library catalogues help readers to find runs, access is increasingly restricted to special collections rather than available on the shelves, and to readers one at a time during library opening hours.[11] The great bibliographic achievements of the twentieth century, the *Wellesley Index* and *Waterloo Directory*, serve as a checklist of periodical titles and a guide to some of their contents, and these can be supplemented with earlier works such as the press directories, Stead's *Annual Index to Periodical Literature*, and *Poole's Index*.[12] Digital resources, however, provide access to anyone, anywhere, with access to the web and the relevant

---

[9] Linda K. Hughes, 'SIDEWAYS!: Navigating the Material(ity) of Print Culture', *Victorian Periodicals Review*, 47 (2014), 1–30, 1, 3.

[10] Wolff, 'Charting the Golden Stream', 27.

[11] For more on access to periodicals, including the decline of material on open shelves, see Laurel Brake, 'London Letter: Researching the Historical Press, Now and Here', *Victorian Periodicals Review*, 48 (2015), 245–53.

[12] W.T. Stead and Eliza Hetherington, eds., *Index to the Periodical Literature of the World*. 11 vols (London: Review of Reviews, 1891-1900); William Frederick Poole et al., *Poole's Index to Periodical Literature*, 7 vols (London: Kegan Paul, Tench, Trübner and Co., 1882-1908).

credentials. They usually provide sufficient metadata to generate lists of titles and articles. Finally, the use of textual transcripts provides an index that makes content recoverable, regardless of the size of the archive. None of these processes are perfect and they frequently lack the scholarly rigour that characterizes earlier bibliographic projects, but whereas research into the newspaper and periodical press required access to well-stocked libraries, bibliographical tools (and the skills to use them), and sufficient experience to make sense of the engrossing but complex texts in their pages, digital resources bring newspapers and periodicals readily to hand.

Margins are important in another sense, too. The white space that surrounds the printed text, the margins mark the page rather than what is written upon it. If what makes newspapers and periodicals so valuable is their participation in a vibrant print culture, then their material facets are crucial for understanding how they were produced, distributed, and consumed. To make the sense of the press is to take materiality seriously, and this applies to encounters with newspapers and periodicals in digital resources as well as in print. The language of the virtual that still clings to digital objects elides their materiality; for digital representations of nondigital objects this is even more pronounced, with the nondigital appearing as the substantial source for its ethereal digital other. Yet this is not a case of the digital being more material than the nondigital: each has its own distinct materiality and, in the case of digital representations of nondigital objects, these different materialities inform one another, shaping the way each is experienced.

Digitisation draws attention to the importance of materiality even as it is apparently imperilled. When material aspects of print such as weight of the volume, the size of the page, or the colour of paper and ink are transformed we become reminded of their importance. The

print objects are unchanged, of course, but when compared with digital representations certain aspects of their material form become enhanced. The nondigital world also structures our experience of digital materiality. Digital objects are often designed to act like nondigital counterparts, allowing users to turn pages or open folders, but they can only do this because of our readiness to read depth, structure and texture into the two dimensional space of the screen. Materiality is not just a matter of fixed properties, associated with an object in advance of whatever is done with it; it also, as our experiences with digital media make clear, emerges in the context of a particular encounter.[13] Just as typography and layout lend print pages texture in addition to that bestowed by the impression of type on paper, so the behaviour of digital objects onscreen lends a further layer of materiality to that of drives and chips, screens and keyboards.[14] Only a richer understanding of materiality can make sense of the way we regularly interact with objects, treating articles as distinct entities on the page, for instance, or interacting with complex simulated worlds online. In other words, there is more to say about the 'hardness' of hard copy than the fact it is 'harder' than digital representations onscreen. It sounds counter-intuitive, but materiality is culturally contingent and situated in the moment: the weight of the volume might stay the same, but its cultural heft changes. The challenge for scholars of the nineteenth-century press is to recover the material meanings of the past.

The material form of newspapers and periodicals is the key to understanding the place of these publications in society. Designed primarily to be read, print publications had all kinds

---

[13] See, for instance, N. Katherine Hayles's arguments in *Writing Machines* (Cambridge, MA: MIT Press, 2002) and 'Translating Media: Why We Should Rethink Textuality', *The Yale Journal of Criticism*, 16 (2003), 263–90.
[14] For an account of this difference, see Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008).

of uses at different times of their lives.[15] One thing those that survive usually share is a large part of their cultural life: at some point they found themselves part of a library collection and were treated accordingly. As a result, the material forms that newspapers and periodicals took in the past can often only be imagined on the basis of the quite different forms in which they exist today. It is a commonplace in scholarship, for instance, to write about individual issues even if the scholar has only ever seen one copy in a bound volume. Just as we would be rightly sceptical of an account of the press that neglected aspects of material form, so too we should expect researchers to think about the materiality of the digital resources they use. Not only do the forms in which digital representations of nineteenth-century print affect the way we conceive the nondigital publications, but the material facets of the digital resource shape how it functions, determining what we see and how. Newspapers and periodicals are not passive bearers of textual content; neither are the digital resources that remediate them.[16]

By the Numbers

Palmer's method was to work through each article of *The Times* (except for advertisements) and assign relevant headings from the index. Some articles took longer to index than others: according to Palmer sometimes a short article '"requires as many entries in the Index as lines in the paragraph, where again perhaps an article of a couple of columns require but one entry."'[17] The transcripts that underpin digital resources are more like the former than the latter. Generated through a process called Optical Character Recognition (OCR) in which a programme attempts to identify verbal information on page images, the resulting transcripts

---

[15] See, for instance, Leah Price, *How to do Things with Books in Victorian Britain* (Princeton, NJ: Princeton University Press, 2012).

[16] For a fuller account, see James Mussell, *The Nineteenth-Century Press in the Digital Age* (Basingstoke: Palgrave, 2012).

[17] 'A Great Inedx', 11.

can then be processed further to create searchable indices. Unlike Palmer's team (his four sons and 'but three strangers'[18]), who scanned articles for content to be indexed, OCR programmes have to first identify which marks on an image constitute glyphs and then to ascribe them correct alphanumeric values. Although some OCR programmes work at the level of words and there are various ways in which the accuracy can be improved by using natural language processing, there is rarely a semantic element to the process. The transcripts might be machine-readable, but as far as the machine is concerned they are just strings of characters.

The shortcomings of OCR transcripts are well-known. Despite the increasing sophistication of OCR technologies, producing transcripts from historical material will never be perfect and the closely printed columns of nineteenth-century newspapers, especially when scanned from tightly-bound volumes, are notoriously challenging to process. While even inaccurate transcripts can be used as the basis for search, accurate metadata is needed for browsing. Anyone who has tried to work with digitised periodicals published in resources designed for books – *Google Books* (2004-) for instance, or the *Internet Archive* (1996-) – can attest to the difficulties in locating the right volume in a list where each has the same title. The bibliographical complexity of periodicals makes such metadata quite complicated: titles change over the course of a run; volume sequences restart with a new series; and issues might appear in multiple editions. As metadata requires addition by hand it can be expensive to carry out at scale; however, as this metadata is used to label search results, helping users browse lists of hits, commercial publishers prefer to devote resources here rather than correct transcripts.

---

[18] 'A Great Index', 11.

The limitations of free-text searching are also well-known. As the transcripts lack semantic information, the assumption that underpins search is that if a particular search term appears in an article, then that article is likely to be about that search term. This works well for named entities (people, places, events, objects, etc), especially when terms are sufficiently distinct, but less well for concepts, themes, or topics. As Patrick Leary has argued, effective free-text searching requires a combination of precision and inclusiveness (to get just the right amount of detail in the such term) and 'a close prior acquaintance with nineteenth-century prose' (to be able to predict terms likely to appear).[19] If done well, such searches can be extremely powerful, allowing scholars to map people, texts, and events across the diverse terrain of nineteenth-century print. Familiar figures are encountered in new settings, predictable cultural narratives can be unsettled by alternative interpretations, and Victorian culture emerges in all its diversity.[20] However, if done with a lack of methodological rigour, free-text searching can reproduce many of the limitations of the archive. For instance, reading might usually be performed on page images, but free-text searching relies on verbal information, making form subsidiary to content. As search results return articles out of context, resources encourage what Mark Turner has called the 'smash-and-grab' approach where articles are studied in isolation while publications are forced back into the margins.[21] Finally, as most searches result in an abundance of hits and the diversity of the press makes it likely a range of views will be found, broad generalizations can be made to rest on individual articles, themselves part of an ongoing conversation. The uncritical use of these powerful resources can make scholarship little more than descriptive accounts of what people wrote in the past.

[19] Patrick Leary, 'Googling the Victorians', *Journal of Victorian Culture*, 10 (2005), 72–86.
[20] For more on search, see Catherine Robson, 'How We Search Now: New Ways of Digging Up Wolfe's "Sir John Moore"', in *Virtual Victorians*, edited by Veronica Alfano and Andrew Stauffer (Basingstoke: Palgrave Macmillan, 2015), pp. 11-28.
[21] Mark W. Turner, 'Time, Periodicals, and Literary Studies', *Victorian Periodicals Review*, 39 (2006), 309–16, 309.

It is easy to blame digital resources for encouraging methodological bad habits, but these resources are designed in such a way because this is what users know and want. Habituated to Google, we are used to navigating an unknown mass of material on the basis of a few speculative search terms, confident that there will nearly always be something of relevance. I have argued previously that whenever nondigital material is encountered in digital form researchers need to establish two things: firstly, what it is they are looking at (ie, what type of article is this? What is its print context? Who produced it? In what genres is it situated?); secondly, why it appears as it does (why does it behave this way? Why has it been returned in this search? How might its appearance be shaped by its production and display?).[22] Pursuing these lines of enquiry requires what Paul Fyfe has recently called 'curatorial intelligence', the need to 'assess and recontextualize digital objects' according to broader interpretive frameworks.[23] Both nondigital and digital object must be put in context, recognized as cultural products that are produced under specific conditions for specific purposes. Only then can the relationship between them – the way the digital resource mediates the printed material – be fully understood. Without this critical engagement, digital resources are reduced to delivery mechanisms for page scans, with access dependent on an incomplete index that is often undisclosed.

Despite the shortcomings of OCR-transcripts, there is considerable scope for doing other things with structured data than simply free-text searching. Over the past few years, projects have emerged that leverage the digital difference to ask new questions of archive material and what it represents. Many adopt what has become known as 'distant reading', using

---

[22] See Mussell, *Nineteenth-Century Press*, chapter one.

[23] Paul Fyfe, 'Technologies of Serendipity', *Victorian Periodicals Review*, 48 (2015), 261–66, 262. Fyfe credits the phrase to an audience member at the annual conference of the Research Society for Victorian Periodicals in Delaware 2014.

techniques from corpus linguistics to map cultural trends.[24] Perhaps the most well-known

example of this is in Victorian studies is Dan Cohen and Fred Gibbs's project to test the

conclusions of Walter Houghton's *The Victorian Frame of Mind* using the titles of the

1,681161 books within *Google Books*.[25] Although such projects look beyond the specificities

of individual texts to detect patterns within the wider corpus, there are ways to make print

culture the object of analysis. Anne DeWitt's recent work, for instance, uses network

analysis to explore the way reviews published in periodicals associated novels with one

another: focusing on an important facet of the periodical, the review, DeWitt can challenge

genealogical and evolutionary constructions of literary genre by presenting those delineated

by reviewers instead.[26] Network analysis can also reveal the relationships between those

clustered around particular publications. Susan Brown, for instance, has shown how the data

in the long-running *Orlando Project* can used to visualize the personal, social and political

connections between contributors to particular publications. By mapping such relationships,

it becomes easy to see what certain contributors might have in common, or otherwise

unsuspected points of connection between people and publications.[27] Ryan Cordell's *Viral*

Texts project also uses network analysis, but this time to trace content as it is reprinted across

the press. As newspapers are date-stamped, the project shows lines of influence between

---

[24] 'Distant reading' was coined by Franco Moretti in 'Conjectures on World Literature', New Left Review, 1 (2000), 57–58. See also Franco Moretti, *Graphs, Maps, Trees* (London: Verso, 2005). For a good account of the possibilities of distant reading the nineteenth-century press, see Bob Nicholson, 'Counting Culture; Or, How to Read Victorian Newspapers from a Distance', *Journal of Victorian Culture*, 17 (2012), 238–46.

[25] See Dan Cohen, 'Searching for the Victorians', *Dan Cohen* (4 October 2010) <http://www.dancohen.org/2010/10/04/searching-for-the-victorians/> [accessed 30 July 2015]; Frederick W. Gibbs and Daniel J. Cohen, 'A Conversation with Data: Prospecting Victorian Words and Ideas', *Victorian Studies*, 54 (2011), 69–77. Further details about the project and access to the underlying data are at *Victorian Books* (2010) <www.victorianbook.org> [accessed 30 July 2015].

[26] Anne Dewitt, 'Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press', *Victorian Periodicals Review*, 48 (2015), 161–82.

[27] See Susan Brown, 'Networking Feminist Literary History: Recovering Elize Meteyard's Web', *Virtual Victorians*, edited by Veronica Alfano and Andrew Stauffer (Basingstoke: Palgrave Macmillan, 2015), pp. 57-82.

publications; but it also reveals the longevity of certain articles, as they reappear again and again over time, subject to different degrees of modification.[28] There are also projects that use metadata to study form. Tim Sherratt's *The Front Page*, for instance, uses genre labels from the National Library of Australia's digital library, *Trove*, to graph the changing types of articles on the front pages of Australian newspapers from the nineteenth century onwards.[29] Dallas Liddle's work on *The Times* focuses on technical metadata, using the file sizes to work out the number of characters in a particular issue. Such calculations allow Liddle to establish the density of the newspaper over time, linking digital information to decisions made in the print room.[30] Finally, Natalie Houston and Neal Audenaert's project, *The Visual Page as Interface* explores various ways of representing the graphic dimensions of the printed page. Taking layout information from the OCR process, rather than the character strings, the project directs attention to the poetics of the page, allowing researchers to systematically examine presentational trends both within and between publications at scales that would be impossible in print.[31]

What unites these disparate projects is the recognition that digital objects have properties of their own. Although there are difficulties in obtaining raw data, especially for British

---

[28] *Viral Texts* (2013-) <http://viraltexts.org/> [accessed 30 July 2015]. See Ryan Cordell, 'Viral Textuality in Nineteenth-Cenutry US Newspaper Exchanges', *Virtual Victorians*, edited by Veronica Alfano and Andrew Stauffer (Basingstoke: Palgrave Macmillan, 2015), pp. 29-56; Ryan Cordell, 'Reprinting, Circulation, and the Network Author in Antebellum Newspapers', *American Literary History* 27 (2015) and David Smith, Ryan Cordell, and Abby Mullen, 'Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers', *American Literary History* 27 (2015).

[29] Tim Sherratt, *The Front Page* (2012) <http://dhistory.org/frontpages/>. For more information see '4 Million Articles Later...' *Discontents: Working for the Triumph of Content over Form, Ideas over Control, People over Systems* (29 June 2012) <http://discontents.com.au/shed/experiments/4-million-articles-later/> [accessed 30 July 2015].

[30] See Dallas Liddle, 'Reflections on 20,000 Victorian Newspapers: "Distant Reading" *The Times* Using the *Times Digital Archive*', *Journal of Victorian Culture*, 17 (2012), 230–37.

[31] Natalie M. Houston, 'Toward a Computational Analysis of Victorian Poetics', *Victorian Studies*, 56 (2014), 498–510.

publications, most of which have been digitized as part of commercial resources, such projects make clear that digital research can do more for print than simply broaden access.[32] These data-driven projects are not a step towards a more objective analysis of the press, but a series of creative transformations that need interpretation. The graphs and visualizations produced by projects are neither objective representations of Victorian print culture nor the final word about them; rather, they are provocations that force us to return to the archive and look at its contents more closely.[33]

Clockwork

When Stead's reporter asked how long it took to index an issue of *The Times*, Palmer answered "'The best part of a day'"[34]. Given that Palmer was working backwards through the published volumes of *The Times* while, at the same, trying to keep up with the paper as published, this rate of work – one day per issue of a daily newspaper – suggests a Sisyphean nightmare of eternal indexing. Palmer told Stead's reporter it would take him forty years to reach the beginning of *The Times* and this turned to be more or less correct. The volume for 1790 appeared in 1925, the final five years completed under the supervision of *The Times* newspaper after they took over the index in 1941.[35]

---

[32] For more on the availability of data, see Bob Nicholson, 'Tweeting the Victorians', *Victorian Periodicals Review*, 48 (2015), 254–60, 258-9.

[33] There is plenty of scope for more traditional quantitative work, too. Data could be harvested from periodicals such as the *Publishers' Circular* or the newspaper press directories. Equally, secondary sources like the *Wellesley*, *Waterloo* and *Dictionary of Nineteenth-Century Journalism* contain information that, if fully-encoded and made available, could support much important quantitative work.

[34] 'A Great Index', 11.

[35] See Barbara James, 'Indexing The Times', *The Indexer*, 11 (1979), 209–11.

A serial work indexing an ongoing serial, the material form of *Palmer's Index* is marked by its temporality. The same is true for digital resources. There remains an iterative component, as resources move from closed to open beta and then to a full launch, but most resources tend to be cumulative, aggregating more data within a fairly stable interface.[36] Resources and their contents have a cultural life: interfaces are refreshed, new features added, and sometimes they are withdrawn or cease functioning, their contents repurposed as part of another resource. *Palmer's Index* has itself been caught up in the temporality imposed by digital form. Computers began to assist compilation in 1973, but the first digital publication of the index was as a CD-ROM published by Chadwyck-Healey in 1995 then online as part of *Historical Newspapers* in 1999. Its data has been incorporated as part of *C19: The Nineteenth-Century Index* (2005-) and volumes have been incorporated in the large, free-to-access collections *Internet Archive*, *Google Books* and *Hathi Trust Digital Library* (2008-). These digital remediations, whether of data or full-text facsimiles, are more like different editions of the index: even if its contents are unchanged, the way they are situated within new digital contexts changes the terms under which they can be discovered and used. Even so-called new media can have complex bibliographical lives.

In an important recent article, Linda K. Hughes has argued that scholars need to move 'sideways', making 'lateral moves' that include 'analysis across genres; texts opening out into each other dialogically in and out of periodicals; sequential rather than "data mining" approaches to reading periodicals; and spatio-temporal convergences in print culture.'[37] Hughes offers two analogies to help conceptualize 'the mass circulation of Victorian print':

---

[36] A good example of this is the *British Newspaper Archive* which launched in November 2011 with 4 million pages; in July 2015 it had over 11 million. See Anonymous, 'The British Newspaper Archive - Launch Press Release', *The British Newspaper Archive* (29 November 2011) <http://blog.britishnewspaperarchive.co.uk/2011/11/29/221/> [accessed 30 July 2015].

[37] Hughes, 'SIDEWAYS!', 1-2.

the city, which captures the concurrent heterogeneity of print culture, a complex system

known only through local knottings and connections, and the discursive web, the way texts

responded to and perpetuated common ways of knowing.[38] Hughes does not argue for one

analogy over the other, but stresses how each illuminates a dimension of print culture. The

urban metaphor 'attunes us to the materiality of print, the need to find routes or pathways

through it, and the meaning to be found within its local formations, its "neighbourhoods" so

to speak.' The discursive web, on the other hand, makes evident 'what is produced by

materiality but often cannot be seen on a given page.'[39] Moving sideways allows the

researcher to keep both in play, recognizing the correspondences, whether continuous or

contiguous, that constitute print culture.

Hughes recognizes that moving sideways can be done online as well as with the surviving

print archive. After all, even when 'drilling down' into search results, we are really following

the recurrence of search terms across the corpus. For Hughes, free text searching makes

evident the discursive web but, for now, turning the pages of printed newspapers and

periodicals is necessary to understand 'how print organizes itself locally, materially, and

temporally.'[40] There is no substitute for turning pages: even the richest digital simulation

cannot capture the haptic experience of an encounter with printed paper. Yet we should be as

cautious of ossifying the current dominant trends in digitization as we are of privileging

existing research methodologies. For Hughes, page turning through successive issues

'constructs horizons of expectations that makes legible an array of temporal and material

cruxes in print culture'.[41] The way the majority of digitized newspapers and periodicals are

presented makes this next to impossible, but this genre of resource is not the only way of

---

[38] Hughes, 'SIDEWAYS!', 2-4.
[39] Hughes, 'SIDEWAYS!', 5.
[40] Hughes, 'SIDEWAYS!', 20.
[41] Hughes, 'SIDEWAYS!', 21.

doing digital research. Using some of the techniques currently being developed, it is possible to imagine resources that are attuned to the emergence of unexpected features or that are capable to detecting new configurations of content. And, because they leverage computational power, such resources can work at much larger scale than the lone researcher, turning pages one-by-one. This is not to argue machines are better readers, or that more can be done working with data than with print; rather, we might learn more when we turn to print if we complement our reading with that of the machine.

In conclusion, I recommend one further analogy for the nineteenth-century press. In addition to the city and the discursive web, I would add the heart. Not only does this remind us that encounters with print necessarily involve feeling, but it also provides a way to think about repetition. Newspapers and periodicals are defined by their open-ended seriality, every issue published with the expectation of another to follow. Repetition is built into the logic of print and no genres more fully embody this logic than newspapers and periodicals. Yet repetition also captures the various remediations that occur when historical print is digitized. Palmer's was one type of object while it was in progress, another once the sequence was completed, then something else again when digitized and incorporated within larger resources. As readers, the repetitive becomes tiresome or, worse, invisible, our attention drawn to the novel or exceptional. When material is remediated in digital form, we tend to only notice the differences that are introduced through digitization, overlooking the discontinuities upon which such processes depend. If digital research is, as Natalie Houston has argued, about moving 'beyond human limitations of vision, memory, and attention', then computers can help us see print in ways that otherwise escape our recognition.[42] The paradox of periodicals scholarship in the digital age is that although the print objects are closer to the past, it is by

---

[42] Houston, 'Toward a Computational Analysis of Victorian Poetics', 499.

doing things to them – reading them, of course, but also transforming, translating, processing and reformatting them – that we bring the past closer to us.