

Accessing the content of nineteenth-century periodicals: the Science in the Nineteenth-Century Periodical project (SciPer)

Nineteenth-century periodicals significantly outnumber books from that era, and present historians with an immensely valuable set of sources, but their use is constrained by the difficulty of identifying relevant material. For many periodicals, contents pages and volume indexes have been the only guide, and the few subject indexes that exist usually provide only an indication of the subjects mentioned in the article titles. By contrast, the Science in the Nineteenth-Century Periodical project (SciPer) indexed the science content of general-interest periodicals by skim-reading the entire text. The project's approach to indexing is described and the relative merits of indexing and digitization in aiding researchers to locate relevant material are discussed. The article concludes that, notwithstanding the more sophisticated search interfaces of more recent retro-digitization projects, human indexing still has an important role to play in providing access to the content of historic periodicals and in mapping their data structure.



JON TOPHAM

Lecturer in History of Science
University of Leeds

As they have strained to house the vast runs of dusty and rarely consulted quarto and octavo volumes, librarians have had good reason to be far more acutely aware than most academics of the immense bulk represented by nineteenth-century periodicals. Even so, the recent assertion by John North that 'periodicals and newspapers are more than 100 times the volume of printed books' might be greeted with surprise.¹ There is, nevertheless, clear evidence that, if not exceeding books by quite such a margin, nineteenth-century periodicals and newspapers did exceed them significantly, both in numbers of titles and in commercial value. North's own *Waterloo Directory of English Newspapers and Periodicals, 1800–1900* suggests that, by the end of the century, more than 20,000 periodical and newspaper titles were in print in England, while the *Nineteenth-Century Short-Title Catalogue* suggests that only a third that number of new books were issued each year in the main British publishing centres.² This disparity is also reflected in Simon Eliot's finding that, when measured by net commercial value of printed output, periodicals and

newspapers represented two-thirds as much again as books in 1907, a figure which rose to 250% more than the value of books by 1924.³

Such findings as these give ample justification for John North's claim that 'Periodical literature is the largest single source of Victorian material available to us'.⁴ However, despite the activities of a dedicated body of periodical researchers (the Research Society for Victorian Periodicals will shortly celebrate its fortieth anniversary), historical scholars have generally continued to be circumspect in their use of this vast and valuable corpus of material. Among the leading reasons for this have undoubtedly been the difficulties of access. With such an overwhelming weight of printed matter, how are historians to find material relevant to their particular researches? Even if they are aware of the relevant periodicals (an assumption it is far from safe to make), how are they to pinpoint the key references across many thousands of pages? In this article, I will begin by considering the traditional means of accessing the content of Victorian periodicals. I will then outline the approach taken

to the problem by the Science in the Nineteenth-Century Periodical project (SciPer), which involved manual indexing of periodical content. Finally, I will reflect on the relative merits of such an approach when compared with retro-digitization.

When I began periodical research some 20 years ago, much of the work had to be done by identifying relevant periodicals from existing historical literature and various periodical bibliographies and catalogues, before reading through innumerable contents pages (and volume indexes where they existed) and skim-reading the periodical texts, in order to track down relevant articles. (The first part of this exercise has certainly been eased in the interim by the advent of John North's monumental *Waterloo Directory*, which, by allowing subject-searching, has provided an invaluable means of identifying relevant periodicals.) In addition, however, there were in existence a relatively small number of printed indexes to the contents of periodicals which sometimes provided invaluable assistance.

The first class of index comprised author–title indexes usually based around particular periodicals or groups of periodicals. The most notable single-periodical index was *Palmer's Index to The Times*, which was issued separately in quarterly volumes, starting in 1868, although the advent of electronic publishing latterly provided a much more convenient, fully integrated version. Perhaps more common were author–title indexes of groups of periodicals, such as annuals or little magazines. Particularly significant here, of course, was the *Wellesley Index to Victorian Periodicals* (5 vols, 1966–89), which provided an index to articles in 43 leading magazines and reviews, identifying anonymous authors as far as possible. More recently, Chadwyk-Healey's online *Periodicals Content Index* has taken a similar approach to a far wider range of periodicals, although without the scholarly focus on the identification of anonymous authors.

A second class of index, which to some degree graduated into this first class, comprised periodical subject indexes. Often prompted by professional requirements, many of the early subject indexes focused on particular groups of specialist periodicals, such as the Royal Society's *Catalogue of Scientific Papers, 1800–1900* (19 vols, 1867–1925) and the *Index Medicus* (1879–2004). William Poole's *Index to Periodical Literature* (6 vols, 1882–1908) was notable in ranging widely over general interest periodicals,

including a total of 479 British and American periodical titles and more than half a million articles. This mammoth achievement remains unparalleled, although Helen Cushing and Adah Morris's retrospective *Nineteenth-Century Readers' Guide to Periodical Literature, 1890–1899* (2 vols, 1944) provides a valuable supplement.

Useful as they could be, both of these classes of index had only limited usefulness for the historian seeking to identify articles on a particular topic. Author–title indexes, of course, make no pretence to offer the reader a guide to the subjects of articles, although both title and author may incidentally do so. Even the subject indexes mentioned, however, provide only limited help. *Poole's Index*, for instance, takes as its subject terms words derived directly from article titles. As Scott Bennett has shown, this results in 26 articles on the Paris Commune appearing under 13 different index headings.⁵ A key part of the motivation for the *Nineteenth-Century Readers' Guide* was the desire to substitute a standard list of subject headings for Poole's more haphazard list, but there is no evidence to suggest that subjects were assigned on the basis of any information other than the title.

Article titles are often, however, a poor guide to their content. For instance, the title of the *Boy's Own Paper* article, 'Men Who Are Talked About', does little to alert scholars to the extensive material it contains on both Darwin and Edison (Figure 1). Neither does the title of the *Punch* illustration, 'Animal Magnetism; Sir Rhubarb Pill Mesmerizing the British Lion', immediately suggest a caricature of Sir Robert Peel (Figure 2). Consequently, a number of scholars in recent years have put together subject indexes and bibliographies of articles based on a more detailed engagement with the source material. Thus, for instance, Eugenia Palmegiano's *Health and British Magazines in the Nineteenth Century* (1998) provides brief characterizations of medical articles from nineteenth-century magazines, the titles of many of which give no clue to their medical content. Supremely, Ruth Richardson and Robert Thorne's *The Builder: Illustrations Index, 1843–1883* (1994) provides a fantastically detailed subject analysis of over 12,000 illustrations, based on personal inspection.

Such indexing is, however, extremely time consuming. When the Science in the Nineteenth-Century Periodical project was inaugurated at the Universities of Leeds and Sheffield eight years ago, the project staff seriously debated whether detailed

in one place, and had seen an encampment of Arabs in another, but had passed both at a distance without drawing rein. Towards evening the country through which they passed began to wear a different appearance, being more fertile and better watered. Flocks of sheep and goats were to be seen feeding by themselves; or, perhaps, under the care of a solitary Arab,

whose form might be detected doubled up in the angle of a wall, asleep in his benignant.

They scarcely knew whether to congratulate themselves on these tokens of civilisation or the contrary. They did not wish to risk meeting any of the natives while it was possible to keep out of their way, but it was evident that they could

not expect to travel much farther through the country without being discovered.

While they were discussing this subject with themselves, suddenly on mounting a hill they found themselves in the immediate neighbourhood of an Arab encampment.

(To be continued.)

MEN WHO ARE TALKED ABOUT.

WE give herewith portraits of four men who, in different walks of life, have achieved an eminence that must have made their names

occupation consisted in passing from carriage to carriage, through the connecting passage, while the train was *en route*, and offering newspapers,

success, and even obtained notice from the "Times," as being the only newspaper published and printed in a train.



Thomas A. Edison.



Sir John Lubbock, Bart., M.P., F.R.S.



Edward Whymper.

more or less familiar to all readers of the Boy's Own Paper.

We start with Mr. Edison, because, while at the outset possessing none of the social and educational advantages of such children of fortune as Sir John Lubbock and the late Dr. Darwin, he yet by sheer force of indomitable perseverance and grit has won his way to the front. He might not inaptly be described, indeed, as the "boy who determined to succeed," and affords a capital example of how often pluck and genius have overcome the obstacles of deficient education in childhood and poverty in youth. His past successes have been so remarkable, his industry is so unceasing, and his perseverance in overcoming failure so persistent, that if his life be spared he is likely to take a foremost place in the ranks of great inventors.

Mr. Edison was born in 1847, in the State of Ohio, his father being of Dutch and his mother of Scotch nationality. His youth was passed in the State of Michigan, and his school education was fragmentary and insufficient, as he only spent two months at a public school, and received most of his early training from his mother, who had been a teacher in Canada. His real education was self-obtained, and chiefly gained after his nominal school life ended.

He early displayed a decided taste for reading, and devoured every book of value to which he could obtain access. At the age of ten he had studied Gibbon and Henry, and began to take a great interest in chemistry. Owing to the poverty of his parents, and the number of their family, he was obliged early to seek a mode of earning his bread, and became a "train-boy" on the Grand Trunk Railway of Canada. His

periodicals, or fruit for sale to the passengers. Being very successful in this avocation, he formed the idea of printing a weekly paper in the



The late Charles Darwin.

train himself. He set up his printing press in a spare compartment, and gaining the co-operation of the conductors and others, who supplied matter for publication, he brought out a paper, sixteen inches by twelve, which he named the "Grand Trunk Herald," which was a pecuniary

He now entered with renewed zest upon his chemical studies, and set up a little laboratory in the same place as his printing press; but his scientific labours came to an untimely end, for having by some accident managed to set fire to the carriage, not only was all his apparatus destroyed, but he himself was personally castigated by the irate conductor. A second attempt at editorship also caused unpleasant results to poor young Edison; for having started a paper, called "Paul Fry," at Port Huron, an enraged reader, who imagined himself insulted, pitched the youth bodily into the river. We next find him learning the art of telegraphy, instructed by a grateful station-master, whose child he had saved from being run over by a train. Such good use did Edison make of this opportunity, that he became so efficient as to obtain the highest salary attainable as a telegraphist, and was able to repair the apparatus. He now began to search for new methods of economising expense and labour in the use of electricity, and by patient and persistent work, rather than brilliant conceptions, soon made his name known. He originated the plan of making a single wire do double work, and in 1874 devised quadruplex telegraphy, by which messages could be simultaneously transmitted from both ends of the wire, and four currents of different strength and direction utilised at the same time. He discovered, too, a plan by which Morse recording instruments could be worked with immense rapidity, without a magnet, and so numerous were the improvements he originated, that he obtained eighty-nine patents, eight of which were for variations in the Morse register, eight to the duplex and quadruplex transmission of messages, thirty-five to automatic and chemical

Figure 1. Men Who Are Talked About, 'Boy's Own Paper', 1881/2, 4(189), 764

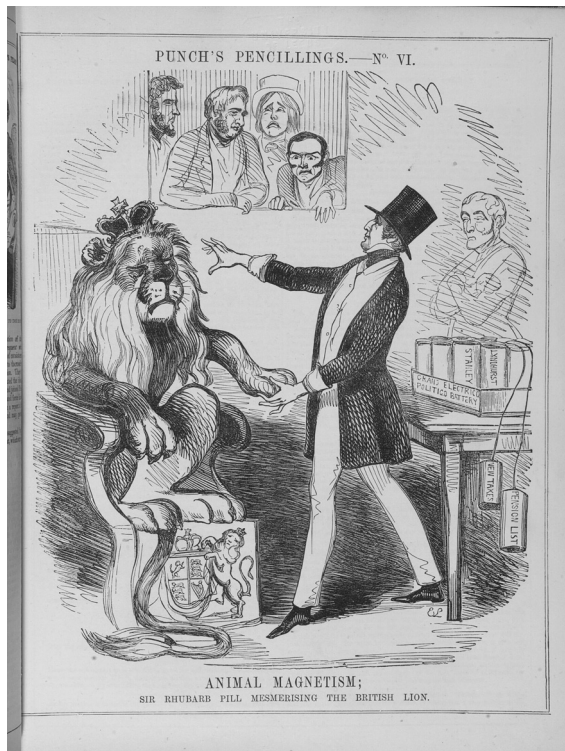


Figure 2. John Leech, *Animal Magnetism; Sir Rhubarb Pill Mesmerising the British Lion*, 'Punch', 1841, 1(6), 67

manual indexing of this sort was warranted in an increasingly digital age. Would the advance of digitization rapidly supersede manual indexing? Our conclusion at the time was that it would not. Experience with existing retro-digitization projects, like the *Internet Library of Early Journals*, suggested that searches could return huge numbers of hits, many of which were irrelevant or trivial. For instance, searching a 20-year run of *Blackwood's Edinburgh Magazine* returned 3,247 hits for the word 'science', including multiple references to 'con-science'.⁶ Employing trained historians to index the material would, we believed, add considerable value to the data that could only be added by such means.

The aim of the SciPer project was to identify and analyse the representation of science, technology and medicine, as well as the interpenetration of science and literature, in the general periodical press in Britain between 1800 and 1900. The approach taken was consequently to skim-read each periodical from cover to cover, however unlikely it seemed that individual articles would contain relevant material, in order to locate significant scientific references incorporated in fiction, news commentaries, poetry and even sermons. Articles were included in the index if references were either of

relevance to the current interests of historians, or revealing of the science as it was conceived in its original historical context. However, while articles containing longer references of scientific relevance were catalogued however conventional their contents, those containing brief or passing references were handled with more selectivity. Thus, where a passing reference of scientific relevance was similar in content to many much longer references appearing in the same journal, it was omitted, while a passing reference of a more unusual or novel nature was generally included.

All articles selected for indexing were assigned a standard bibliographical record, but the project's larger aspirations demanded that additional interpretative information be provided, most notably genre classifiers (drawn from a restricted thesaurus of 65 terms) and subject classifiers (from a similar thesaurus of 350 terms). The subject classifiers are of two types, describing either scientific fields, like 'economic geology', 'matter theory', or 'physiology', or historical themes, like 'amateurism', 'experiment', or 'medical practitioners'. In addition, the index provides regularized records of people, institutions and publications mentioned in significant ways in the articles, with special records to indicate when publications were reviewed, extracted, abstracted, or merely noticed. Most of the entries also include a descriptive paragraph, ranging from 15 to 1,500 words, giving further information about the article content and sometimes containing quotations (Figure 3).

The SciPer Index contains entries for approximately 15,000 articles indexed from runs of sixteen different periodicals, chosen to reflect a wide variety of genres across the century.⁷ These range from the *Wesleyan-Methodist Magazine* to the *Review of Reviews*, and from the *Youth's Magazine* to *Punch*. The index can be accessed in both browsable and searchable interfaces. By allowing historians to browse through the index entries within a framework of periodical volumes, issues, sections and subsections, the main browsable interface to some extent replicates the original structure of the periodical, providing the context that may be crucial for historical interpretation. In addition, historians can browse through static indexes of the authors and illustrators of articles, and of people, books, periodicals and institutions that are mentioned in them, linking through to all relevant articles. These relatively conventional elements of the periodical index are supplemented



Figure 3. Screenshot from 'Science in the Nineteenth-Century Periodical: An Electronic Index' <http://www.sciper.org>

by a complex search facility, using the full range of indexed fields to permit highly sophisticated searches. Thus, the researcher can search for all fictional articles relating to botany written by women born between 1800 and 1850 – and indeed there are some!

It is easy to see that the SciPer Index circumvents the problems identified above with both the title-based printed subject indexes of the past and the first generation of digitized periodicals. The index identifies relevant references within articles, while avoiding the indiscriminately large number of hits encountered with full-text searching. As one reviewer has observed, 'If the scanner is the combine harvester of electronic scholarship, then SciPer (in contrast) should be seen as a quality organic producer'.⁸ Of course, the main drawback of this approach is that it is immensely time consuming, and in three years the project team indexed 160 volumes. In conclusion, therefore, I will reflect on the advance of retro-digitization in the eight years since the SciPer project began, and re-examine the question of its relation to manual indexing.

The progress in retro-digitization of nineteenth-century periodicals since 1999 has been considerable. At the latest count, 60 of the titles digitized in JSTOR

include nineteenth-century runs, and back-files from many commercial publishers have added dozens more. Two significant JISC-funded projects – the Medical Journals Backfiles Digitisation project and the British Library Nineteenth-Century Newspapers project – are in the process of adding further titles.⁹ Moreover, Chadwyck-Healey have developed their *PCI Full Text* product (recently renamed *Periodicals Archive Online*) to incorporate an increasing range of nineteenth-century titles. Particularly notable in this regard is the launch in 2006 of the first part of their *British Periodicals Online*, containing 160 titles from the UMI microfilm collection *Early British Periodicals*. Furthermore, Thomson Gale are preparing to launch their impressive full-text collection of nineteenth-century periodicals in the near future, which will make a further significant impact on the field.

The advent of these resources marks an epoch in historical scholarship. The quantity of often difficult-to-locate periodical material which they make widely and easily available is staggering. Moreover, the capacity they provide for locating references to individuals, institutions, publications, events and concepts can circumvent weeks, months, or even years of painstaking research, and thus make possible research projects that were previously

unthinkable. Do they, therefore, render the retro-indexing of SciPer redundant? My view is that they do not. The quality of OCR has improved immeasurably since the *ILEJ* project, leading to fewer spurious search results, and publishers have sought to provide increasingly sophisticated search interfaces using complex metadata and results clustering. Nevertheless, searches still frequently produce prohibitively large and inchoate data sets, which threaten to submerge the historian. Of course, one aspect of the problem is that historians need to develop new techniques to maximize both the efficiency of their research and the historical reliability of their findings. However, there are limits to the extent to which electronic searching of these relatively unstructured texts can be made to answer the historian's requirements. Thus, to take a single example, a search for 'music and science' in *British Periodicals Online* does not return a hit for the anecdote in the *Mirror of Literature* concerning the geologist Jean Deluc's love of music, whereas the SciPer Index identifies it by the subject keywords, 'Music' and 'Scientific Practitioners'.

The continuing value of human indexing is clearly recognized by those involved in retro-digitization. As I have already observed, the capture of metadata to allow more structured searching is, to a greater or lesser extent, a standard feature of the leading digitization projects. Moreover, Chadwyck-Healey has been quick to exploit existing indexes to provide enhanced searching – both with their Palmer's Full-Text Online and, more recently, with their cross-linking between electronic versions of the Wellesley and Poole's indexes and full-text periodicals in Periodicals Archive Online. This kind of amalgamation of digitization and traditional indexing clearly has the potential to provide the best of both worlds, and suggests a continuing role for retro-indexing. In addition, academic indexing provides an important conceptual resource for large-scale digitization projects. By exploring and analysing the data structure of the periodical in order to produce a useful index, the historical or literary scholar provides conceptual tools of importance in further developing the structuring of digital texts.

This is one of the aspirations of the Nineteenth-Century Serials Edition project (NCSE), currently underway at Birkbeck College. The project is preparing an electronic 'exemplary edition' of six nineteenth-century newspaper and periodical

titles, combining full-text images with sophisticated searching and indexing functions.¹⁰ Seeking to address the concern that digitization can readily flatten out many elements of the data structure of the original periodical – particularly as regards its textual and illustrative content and its physical form – they have developed data and concept maps of extraordinary complexity. Aspects of this mapping will inform the data architecture of their own edition, but, in addition, they consider this an opportunity to 'stretch and develop current perspectives and ways of thinking about these materials'.¹¹ Moreover, the project is combining its exhaustive conceptual analysis with pioneering use of automated indexing techniques, so that, while the assignment of metadata is dependent on the initial academic analysis, academics do not have to spend their time assigning the metadata manually. Such techniques include automated indexical mark-up (e.g. from a name index), and data-mining (pattern-searching the dataset and relating the findings to a concept map), which might make possible the automated generation of thematic metadata. The combination of conceptual and technical analysis undertaken by projects like the NCSE promises significant dividends for those engaged in much more extensive retro-digitization projects. I, for one, await the outcome of their researches with great interest. Thus, my general conclusion remains, that academic interest in indexing nineteenth-century periodicals will continue to provide important resources for those involved in digitization, as we strive to achieve increasingly sophisticated means of accessing this vast and under-used resource.

Reference

1. Howsam, L, Award for Distinguished Achievement, *SHARP News*, 2006, 15(4), 3.
2. Cantor, G, et al., *Science in the Nineteenth-Century Periodical: Reading the Magazine of Nature*, 2004, Cambridge University Press.
3. Eliot, S, Some Trends in British Book Production. In: *Literature in the Marketplace: Nineteenth-Century British Publishing & Reading Practices*, Ed. Jordan, J O and Patten, R L, 1995, Cambridge, Cambridge University Press.
4. North, J, The Rationale – Why Read Victorian Periodicals. In *Victorian Periodicals: A Guide to*

- Research*, Ed. Vann, J D and VanArsdel, R T, 1978, New York, Modern Language Association of America.
5. See Bennett, S, The Bibliographical Control of Victorian Periodicals. In *Victorian Periodicals: A Guide to Research*, Ed. Vann, J D and VanArsdel, R T, 1978, New York, Modern Language Association of America.
 6. *Internet Library of Early Journals*:
<http://www.bodley.ox.ac.uk/ilej>.
 See also Field, C, The Internet Library of Early Journals: Preserving and Disseminating the British Journal Heritage, *Serials*, 1997, 10(2), 84–90.
 7. *Science in the Nineteenth-Century Periodical: An Electronic Index*:
<http://www.sciper.org>. The last portion of the data is currently being checked, and will be added to the online index in coming months.
 8. Secord, J, The Electronic Harvest, *British Journal for the History of Science*, 2005, 38(4), 463–67.
 9. On JSTOR see
<http://www.jstor.org>
 On the Medical Journals Backfiles Digitisation Project see
<http://library.wellcome.ac.uk/node280.html> and
 Kiley, R, The Medical Journals Back-Files Digitization Project and Open Access, *Serials*, 2006, 19(2), 95–102.
 10. *Nineteenth-Century Serials Edition*:
<http://www.ncse.kcl.ac.uk>.
 11. Mussell, J and Paylor, S, Mapping the “Mighty Maze:” The Nineteenth-Century Serials Edition, 19: *Interdisciplinary Studies in the Long Nineteenth Century*, 2005, 1
<http://www.19.bbk.ac.uk>.
- On the Nineteenth-Century Newspapers project see
<http://www.bl.uk/collections/britishnewspapers1800to1900.html>

Article © Jon Topham

Illustrations reproduced by kind permission of Leeds University Library

■ **Dr Jon Topham**
 Lecturer in History of Science
 University of Leeds
 E-mail: J.R.Topham@leeds.ac.uk

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=1&spage=20>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=1>