ARCHAEOLOGY IN NEW ZEALAND

Data Mining New Zealand Newspapers for Historical Archaeology

Simon Bickler *(Bickler Consultants Ltd.)*

Introduction

I have previously written on the potential of using machine learning applications for looking at various archaeological datasets in New Zealand archaeology (see Bickler 2018a, 2018b, Jones and Bickler 2019). Here I briefly examine the use of "data mining" of historic newspapers (e.g., Lansdall-Welfare et al. 2017) as a way of broadening the types of information available about the artefacts routinely found in historic archaeological sites in New Zealand. While the National Library's PapersPast website (https://paperspast.natlib.govt.nz/) is now routinely used by historical researchers, most archaeologists use it to find specific details about events, people and places, or in cataloguing suppliers and distributers of products, their origin and duration of operation. These range from bottles (see e.g., Low 2005) and bottle top capsules (Petchey and Innanchai 2012), to clay pipes (e.g., White 2016) and velocipedes (Woods 2016). There is no question that such information continues to inform on the material culture recovered from excavations. That approach to the data tends to rely either on cataloguing occurrences of information regarding the culture or very limited sampling of the information, for example the start and end dates of references, rather than exploring the changing patterns within that data and how it is reflected in and by the material culture from archaeological sites.

Data Mining

Data mining from large quantities of textual data has led to the development of the "digital humanities" in research programmes around the world (see e.g., Schriebman et al. 2014). No attempt is offered here to summarise the many and varied ways in which that work is now undertaken, but given the increasing digitisation of historic data, its expanding availability and the rapidly evolving tools such as Natural Language Processing (NLP), Topic and Sentiment Analysis which can be used in analysing large sets of textual data including archaeological reports (see e.g., Richards et al. 2015, Kintigh 2015, Jeffrey et al. 2009), there are many new ways to explore data for archaeological research. The approach here focuses on exploring the use of the digitised historic newspaper libraries to illustrate retail and consumer trends during the latter half of the 19th and early 20th

centuries that relate to objects found in historic NZ archaeological sites. These trends provide the social context for those objects. The objects themselves act as the inspiration behind exploring the historic newspapers.

The primary access to the PapersPast newspapers is by using their website, but the web-based interface makes the analysis presented here laborious. Instead, the analysis was undertaken using the DigitalNZ API (1) which provides programmatic search access to the over three million pages of PapersPast newspapers. The search results can then be collated and statistically analysed. Searching the PapersPast newspapers is done using a specific word or phrase, just like the website interface, with additional abilities to sort and filter the results. Where possible those newspaper items catalogued as advertisements can be identified and analysed.

The search relies on the Optical Character Recognition (OCR) that has been undertaken on the newspaper articles, converting the digital images to digital text. This is problematic with the reliability of the text recognition variable across the newspaper corpus with earlier newspapers often very poorly recognised. Spelling mistakes in the material hamper the search facility which means that the DigitalNZ interface uses quite a flexible approach to reporting positive results, ignoring spacing and capitalisation, for example. This means that a broader bucket of data is returned but necessitates more extensive filtering to ensure that the results contain valid results.

The corpus of data being examined is also a moving target as new papers are being added to the library and the OCR continues to improve. This also means that analysis of the data continues to evolve, and future searches will differ from earlier ones. Therefore, the results needed to be treated with caution. Various approaches to improving the data and normalising it for consistent analysis are available but beyond the scope of this paper. This sort of approach relates in part to "Big Data" style analysis (see also Bickler 2019), balancing the difficulties inherent in the data (veracity) with the quantity (volume), variety and longitudinal nature of the data.

Data from Australian newspapers using the Trove API (http://trove.nla.gov.au) which currently provides access to 18.5 million pages from more than 1000 newspapers (http://help.nla.gov.au/trove/using-trove/digitised-newspapers) was undertaken for comparison. The extraction technique was like that undertaken for the PapersPast corpus, but the Trove data tends to be more rigorously classified with better delineation of advertising material without surrounding information and more reliable OCR.

Two examples, one specific relating to the recovery of old electric batteries from an historic house and a more general example looking at a common artefact type, fizzy water bottles, are used to illustrate the analytical potential of the approach.

Leclanché Cells

Two batteries were found under the floorboards at the Alfred Nathan House on Princes Street in central Auckland during archaeological work on the property (Mailhot *et al*. 2016). The batteries consisted of a glass jar, inside of which was a porous ceramic pot and a heavily corroded anode, probably zinc, which were heavily oxidised (Figure 1). A central carbon cathode was visible in both jars, but no fluid was present. The batteries were around 15.3cm high (13.5cm for the jar), a width of 9.3cm and top diameter of 9cm. Black paint covered the top of the jar where a pouring lip was also apparent.
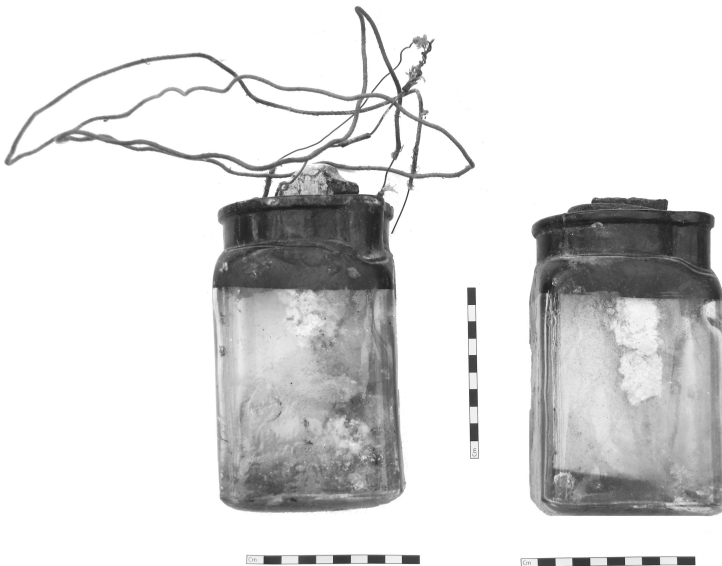

Figure 1. Leclanché cells (side view).

Identification as Leclanché-type batteries was relatively straightforward. Originally developed by Georges Leclanché in 1866, these batteries were a forerunner to the dry cell batteries. These batteries were common around the world and soon found a niche with the telegraph and telephone networks as they

developed a reasonable charge for short periods. They were also relatively easy to maintain (https://en.wikipedia.org/wiki/Leclanch%C3%A9_cell).

While international information provided a good overview for the batteries, nothing about their New Zealand use is described. It was not clear from the context of the batteries what they had been used for in the house, nor when. However, it is quite possible they were used for telephone or doorbell type arrangements in what was quite a well-to-do property. Creating some locally relevant information prompted the use of the search facility of PapersPast to see what else could be gleaned. Fortunately, the name Leclanché is identifiable so that searching for it was relatively straightforward.

The results (Figure 2) point to the arrival of the batteries in New Zealand in the late 1870s and early 1880s and sold with a range of general electrical supplies. Their use in the telegraph in New Zealand is also confirmed by newspapers dating from the 1890s (Figure 3), although often competing technologies and eventually electricity available by the growing national grid would make the technology obsolete.

Figure 2. Advertisement for electrical supplies
(Star, Issue 4563, 8 December 1882, Page 4, Papers Past)

The telegraphic apparatus at the local office is undergoing an extensive overhaul at the hands of Mr McKenzie, Linesman; Leclanche cells are being substituted entirely for the old Daniel's cells, the number being thus very largely reduced. A new copper wire to the 'earth' is being laid in place of the decaying galvanised iron one and other matters affecting the efficiency of the office attended to.

Figure 3. Use of the Leclanché batteries for telegraphic apparatus
(Bay of Plenty Times, Volume XXII, Issue 3409, 10 June 1896, p. 2, Papers Past)

Plotting the frequency of the results (Figure 4) for all mentions of the batteries pointed to a high number of mentions when they became available followed by a very rapid drop in mentions. This was probably because sales were quite specialised and no longer targeted to the general public. Later references in the papers relate more to the use and replacement of the batteries, with the advertisements relating mostly to the re-use of the jars for other purposes by the 1920s. The limited charge duration, increasing availability of the grid, and the improving technology saw the demise of this type of battery.

Comparison with the Australian results show that they had initially been exhibited as early as 1875 (2) but as in New Zealand gained their main popularity in the 1880s. The number of mentions however stayed much higher well into the early 20th century but with a similar brief rise in the 1910s. Although the number of results reflects the much larger corpus of newspapers in Trove, the popularity of the batteries probably lasted longer as they were likely to have had much more use in rural Australia. This was most likely associated with the difficulties of getting the national grid out into those areas. More research on this may uncover some interesting insights into how electricity was implemented in both countries.
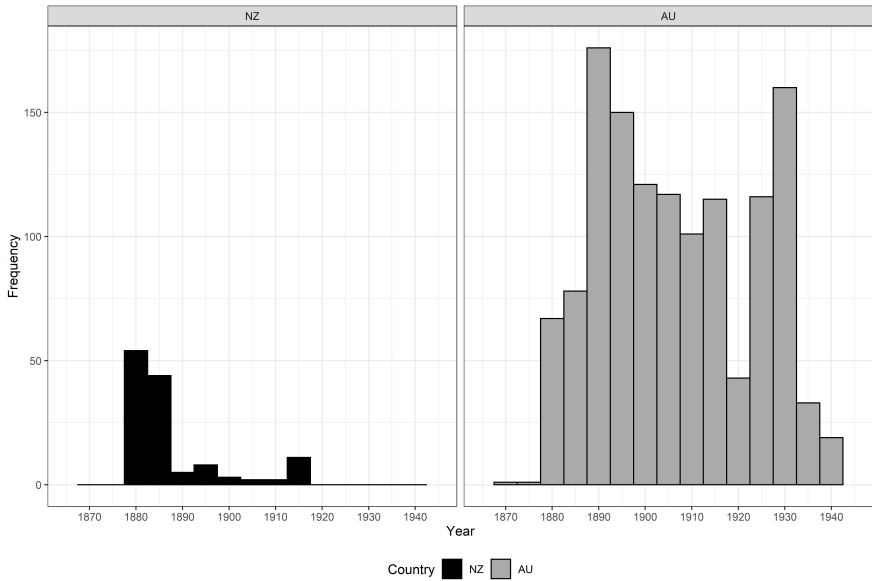
Figure 4. Number of references to Leclanché batteries by year in old newspapers from New Zealand (left) and Australia (right)

Fizzy Water

The Leclanché batteries were obviously quite specialised so it is useful to take a broader category of the artefact, soda water bottles, and explore the PapersPast results. Fizzy water bottles are frequently found in archaeological excavations around the country (see e.g., Garland 2013) and have a rich social role in New Zealand colonial history (e.g., Kouzminov 2017). The search for soda water was immediately more difficult as in fact several terms were probably used and the number of contexts associated with the use of the phrases, not just in advertising, is much greater. Four common phrases were therefore chosen: aerated water, soda water, seltzer and mineral water. The frequency of the term use in advertisements is shown in Figure 5.
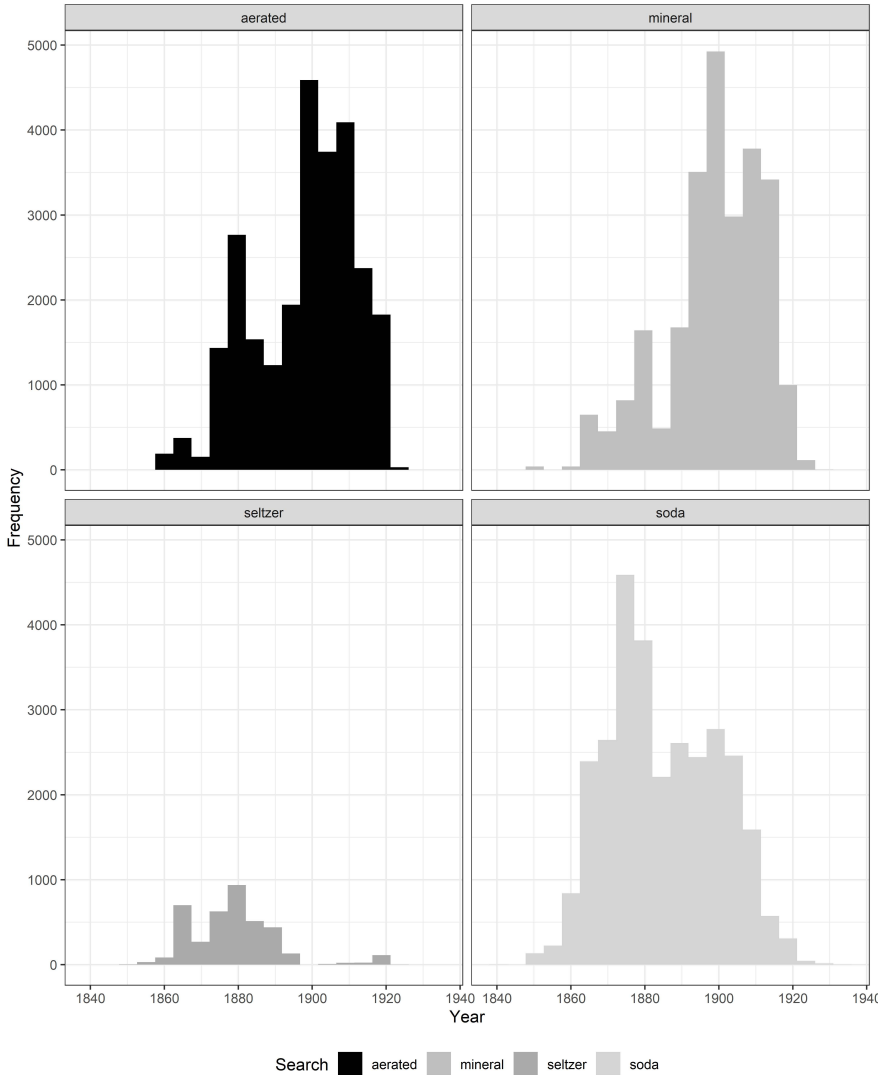
Figure 5. Frequency of fizzy and mineral water mentions in New Zealand newspapers
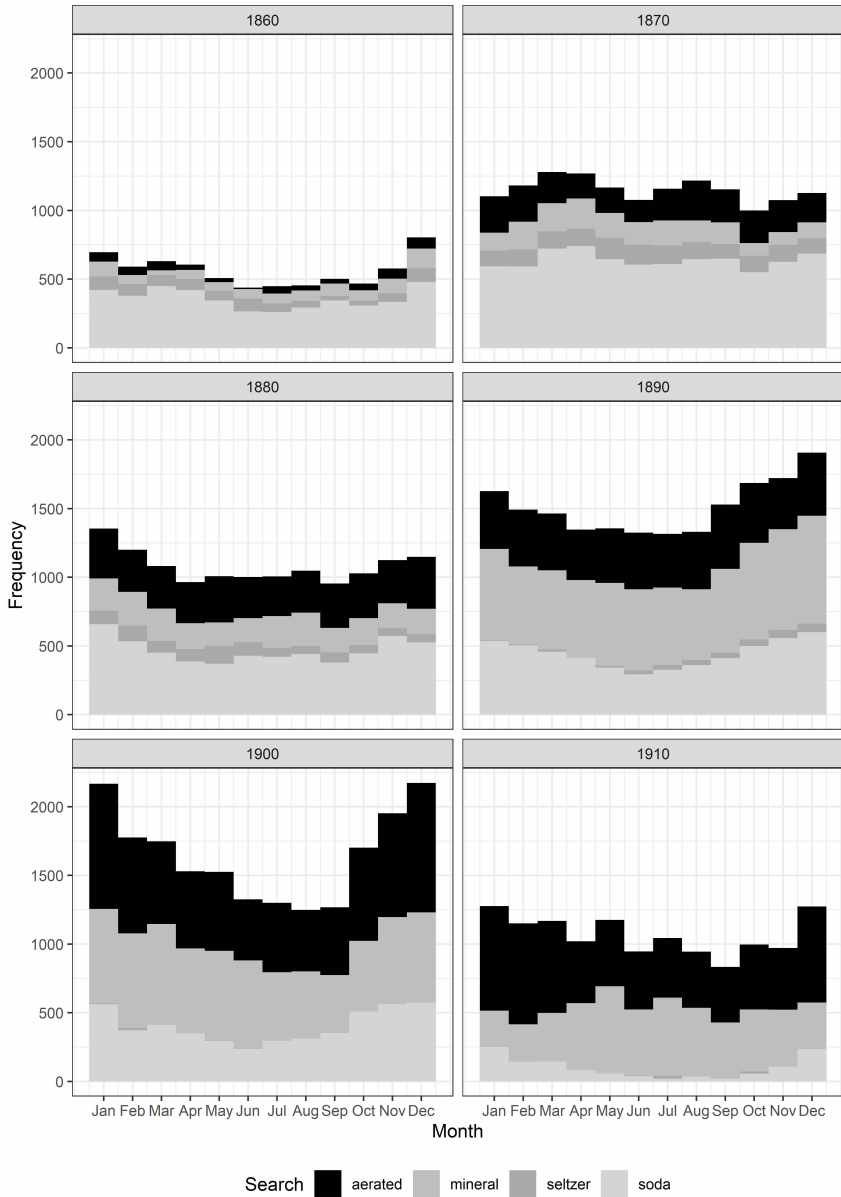
Figure 6. Seasonality in NZ newspaper records of fizzy water.

Seltzer was less common than the others and primarily used between the 1860s-1900s. Interestingly, aerated and mineral water peaked in the 1900s while soda water had peaked in the 1870s. Not all mineral waters were fizzy, so the similarity in advertising between aerated and mineral waters suggests they were being marketed together. Soda water lost popularity possibly as its medicinal origins became less important and perhaps as the range of ways the sparkling water was made increased.

Delving a little deeper into the results, it is possible to look at seasonality in the data by plotting the frequency of charts by month (Figure 6). A pattern of increasing advertising in spring and summer, diminishing again towards autumn is apparent during various decades. This may have been predicted but it largely occurs after the 1870s and not apparent in the 1910s where it seems to have diminished. Whether such changes reflected consumer or retail patterns or accessibility to refrigeration and ice are avenues of research worth pursuing.

Discussion

The examples of data mining of New Zealand and Australian newspapers are shown here only as an introduction to how such data can be used to inform historic archaeology projects in New Zealand. Petchey (2006: 59ff) identified some of the early hydroelectric plants in Central Otago which are just one example of the industrial heritage of New Zealand (Smith 2001), but there has been little work on the smaller artefacts associated with the arrival and use of electricity in New Zealand. The newspapers research provides an opportunity to explore other broader themes motivated by the chance find of the Leclanché batteries from the Alfred Nathan House in central Auckland. Digital humanities approaches to such datasets open up new opportunities to explore the social history of material culture found in New Zealand on a local, regional, national and international scale from the 1840s through to the present. Even relatively well-known objects such as aerated water bottles can be used to yield information regarding the New Zealand's colonial and post-colonial contexts with a focus not only on what was written but in what was left behind (c.f. Deetz 1977:161).

Acknowledgements

Endnotes

1. The API is Accessible Programming Interface from DigitalNZ, https://www.digitalnz.org/, which allows researchers to get access to PapersPast and other databases using a programming interface, in this case the R Statistics package, https://cran.r-project.org. Additional documentation available at those locations.
2. The Argus Account of the Victorian Exhibition September 3, 1875 (http://trove.nla.gov.au)

References

Bickler, S.H. (2017). Layers of Learning: Archaeology at the University of Auckland. Paper presented to the symposium - The Ground Beneath Our Feet: A Place Pedagogy Project at the University of Auckland/Te Whare Wananga o Tamaki Makaurau. 12-13 September 2017.

Bickler, S.H. (2018a). Machine learning identification and classification of historic ceramics. *Archaeology in New Zealand*, 61 (1): 20-32.

Bickler, S.H. (2018b). Prospects for machine learning for shell midden analysis. *Archaeology in New Zealand* 61 (1): 48-58.

Bickler, S.H. (2019). Sustaining New Zealand archaeology's digital future. *Archaeology in New Zealand* 62 (4): 45-57.

Deetz, J. (1977). In *Small Things Forgotten: The Archaeology of Early American Life*. Doubleday: New York.

Garland, J. 2013. Fizz, bang, pop!: Christchurch's early soda water industry. https://blog.underoverarch.co.nz/2013/08/fizz-bang-pop-christchurchs-early-soda-water-industry/

Jeffrey, S., J. Richards, F. Ciravegna, S. Waller, S. Chapman and Z. Zhang. (2009). The Archaeotools Project: Faceted classification and natural language processing in an archaeological context. *Philosophical Transactions of the Royal Society* A 367: 2507–2519.

Jones, B. and S.H. Bickler. (2019). Multi-scalar and semi-automatic approaches to detect archaeological features in NZ using Airborne LiDAR data. *Archaeology in New Zealand* 62 (3): 10-24.

Kintigh, K. (2015). Extracting information from archaeological texts. *Open Archaeology* 1: 96–101.

Kouzminov, A. (2017). Waters of New Zealand: Inside the bottle. *NZToday* 73: 84-91.

Lansdall-Welfare, T., S. Sudhahar, J. Thompson, J. Lewis, *FindMyPast* Newspaper Team and N. Cristianini. (2017). Analysis of 150 years of

British periodicals. *Proceedings of the National Academy of Sciences* 114 (4): E457-E465.

Low, J. (2005). Health In A Bottle? The Archaeology of Self-Medication in 19th Century New Zealand. Unpublished MA thesis, University of Auckland.

Mailhot, C., C. Judge and R. Clough. (2016). Extension to Alfred Nathan House, 24 Princes Street, University of Auckland: Final Archaeological Monitoring Report. Unpublished report for University of Auckland Property Services.

Petchey, P. (2006). *Gold and Electricity: Archaeological Survey of Bullendale, Otago*. Science & Technical Publishing, Department of Conservation: Wellington.

Petchey, P. and J. Innanchai. (2012). Bottle top capsules in New Zealand historic archaeological Sites. *Journal of Pacific Archaeology* 3: 1-16.

Richards, J., D. Tudhope and A. Vlachidis. (2015). Text mining in archaeology: Extracting Information from archaeological reports. In J. Barcelo and I. Bogdanovic (eds), *Mathematics and Archaeology*. CRC Press: Boca Raton. pp. 240-254.

Schreibman, S., R. Siemens and J. Unsworth (eds). (2014). *Companion to Digital Humanities*. Blackwell: Oxford.

Smith, N. (2001). *Heritage of industry. Discovering New Zealand's Industrial History*. Reed Books: Auckland.

White, S. (2016). The McPhees, New Zealand's first clay pipemakers. *Archaeology in New Zealand* 59 (3): 10-28.

Woods, N. (2016). Velocipede mania hits Whanganui. *Archaeology in New Zealand* 59 (1): 5-17.