

Gale Primary Sources

Start at the source.

In Search of America: An Introduction to Digital Research Techniques

Bob Nicholson

Edge Hill University

Various source media, Punch Historical Archive 1841-1992



Introduction

The *Punch Historical Archive* is packed with articles, adverts and cartoons on a remarkable variety of topics. As the introductory guides on this website show, *Punch* magazine allows us to explore: the historical workings of British and international politics; the rise and fall of Empires; shifting gender relations; technological transformations; responses to major global conflicts; and more than 150 years of social and cultural change. Enter any of these topics into the archive's basic search box and you'll immediately be rewarded with a selection of fascinating results. But this is just the tip of the iceberg. Countless treasures are hidden deeper within the archive, but they can only be uncovered with the creative use of keywords, the help of advanced search tools, and an understanding of how *Punch* magazine was organised. The techniques required to conduct this research vary from topic to topic, but it is possible to identify a set of core methodological approaches that can be applied to a range of different research areas. This article demonstrates how to make the most of these digital research techniques by applying them to a case study - *Punch's* tumultuous relationship with the United States.

Researching the Representation of America

What follows is not a comprehensive survey of *Punch's* dealings with America, but a roadmap for how this (and countless other research projects) might be pursued now that the magazine has been opened up to digital search tools. The specific keywords and approaches detailed below will be useful to anybody interested in exploring *Punch's* coverage of the United States, but it is important to stress that the underpinning research

techniques can be applied to any research project. As we'll see, the United States offers a particularly useful case study because it requires us to overcome a series of common methodological challenges: (1) selecting, refining and combining appropriate keywords; (2) responding to social, cultural and political developments; (3) identifying images with a text-based search engine; (4) and narrowing down tens of thousands of results to a focused and manageable sample.

Learning how to approach these problems is vital training for any would-be digital researcher. In this respect, searching for America is an excellent teaching tool. Inviting students to puzzle through some of the problems outlined below will provide an opportunity for them to engage with the *Punch Historical Archive's* advanced search tools and reflect on the strengths and weaknesses of digital research platforms. Teaching activities might include:

- Challenging students to come up with a preliminary list of keywords relating to America, and then asking them to evaluate their usefulness by testing them out in the archive. What makes a good keyword, and why? Which keywords should be avoided? How can you find new keywords?
- Challenging students to locate images relating to America. What kind of keywords are most effective at this task, and why? How can you find them?
- Challenging students to exclude a particular set of articles from their search results. How, for example, might we identify articles about Abraham Lincoln and exclude articles about

the city in Lincolnshire? What factors should be considered when selecting exclusionary keywords? How do you avoid eliminating useful results?

This article does not provide definitive answers to these questions - I'm sure there are plenty of useful techniques that I have yet to discover - but it does make one thing very clear. In order to make the most of a digital archive like *Punch* it is necessary to think carefully and creatively about how we construct and refine our searches.

Selecting and Excluding Keywords

The selection of appropriate keywords is vital for any digital research project. At first glance, locating material related to the United States seems quite straightforward. A simple 'Entire Document' search for the term 'America' instantly returns more than 11,000 results. This, alone, is a formidable amount of material for any researcher to wade through. However, by refining our keywords we can significantly expand these results. Firstly, it is important to recognise that a search for 'America' will not return variations on the word, such as 'American', 'Americans', 'Americanism', or 'Americanisation'. These search terms could be entered individually, but the archive's wildcard tool offers a more effective way of doing things.

By appending an asterisk to the end of the word - 'America*' - the search will automatically return all words that begin with 'America'. This is an essential tool for effective searching and is particularly useful for locating both a word and its plural version (e.g. 'Yankee'

and 'Yankees'). In this case, a search for 'America*' nearly triples the number of results.

While most articles about America use the word (or one of its variations) at some point, there are many relevant sources that refer to the country in other ways. For example, a search for the term 'United States' returns 4,218 hits. Many of these articles also contain variations on 'America' and consequently would have been found using our first search. However, if we perform a search for 'United States NOT America*' (a search that locates instances of the term 'United States', but ignores any article that also features variations on 'America') we find that 1,503 of these hits are new. If you've already spent hours sifting through the results of one search, this is an excellent technique for conducting overlapping searches which ignore material that you've already seen.

Using the 'Not' field is also an excellent way to purge irrelevant results from your searches. For example, searching for the word 'America' will also locate all references to 'South America'. If we want to focus exclusively on the United States, then a search for 'America NOT South America' will cut out some of the irrelevant hits. Excluding results in this fashion requires careful consideration. In this case, our search might eliminate useful articles about the United States that only mention the phrase 'South America' in passing. Before using this technique it would be wise to perform a search for the term that you intend to exclude and to double check that the results do not contain too much relevant material.

Using the 'Advanced Search' window it is also possible to exclude a list of keywords, or to ignore articles with a particular title or author. For example, I once attempted to track usage of the American slang term 'skedaddle' across a range of digital newspaper archives. Searching for such a distinctive phrase is usually straightforward, but in this case it turned out that 'skedaddle' was also the name of a fairly successful racehorse. My results were flooded with hundreds of sports reports, which made it awkward to pick out any other articles using the word. In this case, the solution was to identify a series of unusual words that appeared regularly in race reports and then exclude them when searching for the term 'Skedaddle'. Again, it is important to select these words carefully. Eliminating the term 'race', for example, might have excluded material on American race relations as well as columns of horseracing results. Similarly, a 'horse' might skedaddle in the Grand Canyon as well as Grand National. Words like 'handicap' or 'steeplechase', on the other hand, rarely appeared outside of horseracing columns and could therefore be used to filter this material out of my searches.

Agonising over these possibilities might seem like an overly elaborate way to avoid picking through irrelevant articles, but generating a focused set of results ultimately saves time and opens up the possibility of conducting interesting forms of quantitative analysis.

Refining Results

A search for 'America* OR United States NOT South America' returns 14,224 results. Some dedicated researchers may feel compelled to wade through these

articles one-by-one, but most of us will prefer to narrow things down a bit first. The menu that appears to the left of the search results screen provides four useful options for filtering results.

Firstly, it allows you to focus on specific types of articles. For example, it is possible to limit our search purely to *Punch*'s famous cartoons. If you'd like to exclude a particular article type (such as advertisements) from your search then this can be accomplished by ticking the relevant boxes in the 'Limit Results By' section of the 'Advanced Search' menu. Secondly, it is possible to filter your results by 'Issue Type'; this makes it possible to focus on, or exclude, material from *Punch*'s Almanacs and Seasonal Specials. Thirdly, you can refine your results by date. The archive allows you to: (1) focus on the period before or after a specific date; (2) specify a specific date range to search within. While it's helpful to focus on specific time periods, it's worth remembering that *Punch* continued to reference some events and historical figures long after they ceased to be topical.

Arguably the most useful way to refine your search results is to introduce new keywords. While 'United States' and 'America' were generally used as synonyms by *Punch*, other terms used to describe the country and its inhabitants can be used to home in on specific kinds of content. For example, a search for the term 'Yankee*' (1,355 results) generally finds *Punch* in a teasing mood (e.g. 'Yankee Big-Drum-Taps', 30 Oct 1869: 176). In order to focus on political content, it is wise to search for terms such as 'President', 'Congress', 'Senator', 'Washington', or the surname of a

sitting President. President Lincoln, it should be noted, was occasionally referred to simply as 'Abe'. *Punch's* (often uncomplimentary) thoughts on American literature and culture can be found by searching for the surnames of the country's most prominent writers. Major cities like New York, Chicago, Washington, Philadelphia, Detroit and San Francisco all return results, as do most of the states (once again it is worth using wildcards to capture both California and Californian). Vaguer geographical references such as 'Wild West', 'Far West', and 'Backwoods' also return results, as do items and characters associated with these locations such as 'cowboys' and 'revolvers'. Finally, and more uncomfortably, *Punch's* thoughts on slavery and the racial politics of America require us to make use of search terms such as 'nigger*', 'negro*', and 'slave*'.

Punch took great pleasure in mocking the excesses of the American press, and these pieces can usually be found by using terms like 'American paper' and 'American press', or by searching directly for the names of prominent newspapers such as the New York Tribune or the New York Herald. Like most British newspapers *Punch* relied upon the New York press for much of its American news, and these papers tend to be overrepresented. Searching for these items can sometimes be tricky. A search for 'America* AND press' will return all articles featuring these words, no matter how far apart they appear on the page - this results in a lot of irrelevant content. Conversely, a search for the precise phrase "American press" will not return articles with phrases like "The press has been very agitated of late in America". Proximity operators - one of the *Punch* Historical Archive's most advanced

search tools - can be used to find some middle ground. A search for 'press n10 America*' will find all articles in which these two terms appear within 10 words of one another. This number can be tweaked in order to strike a good balance between precision and flexibility - any number followed by the letter 'n' will do the trick, but a higher number will usually lead to more irrelevant results.

It is also important to take note of the language surrounding key events in American politics and transatlantic relations. During the 1860s and 70s, for example, the term 'Alabama' emerges as a particularly useful search term due to the Alabama Claims - an acrimonious dispute between the two countries over Britain's intervention in the American Civil War. Much of *Punch's* coverage of the dispute fails to mention obvious keywords like 'America' or 'United States', and therefore will not be captured by general searches (e.g. 'The Case of the Alabama', 21 March 1868: 128). In order to pinpoint these topical keywords it is necessary to combine background reading on your topic with the close reading of articles captured by more general search terms. I maintain an extensive list of keywords for each of my research projects and add to them each time I spot a viable new term.

Searching for Images

The *Punch Historical Archive* contains a wealth of fascinating cartoons, illustrations and photographs. While it is easy to focus your searches on these materials, locating cartoons on specific subjects can sometimes be difficult. The archive's search engine can only examine the small passages of text that appeared

beneath most of the magazine's illustrations. A search for 'America', in other words, will only find cartoons that include the word 'America' in their caption. While this will capture some results, it is necessary to use different kinds of keywords in order to uncover many of the archive's most fascinating images.

Punch's cartoonists often made sense of international relations by boiling things down to an encounter between national personifications. Britain was almost invariably represented by the character of John Bull, while different aspects of America were varyingly depicted as Columbia (a female personification of the nation), Uncle Sam (the American government), and Brother/Cousin Jonathan (the American people). All of these characters appeared in *Punch*, but the most commonly used figure during the nineteenth century was Jonathan. In appearance he was typically taller and more slender than the rotund John Bull, and was usually to be found wearing striped trousers, a star-spangled waistcoat, a stove-pipe hat, a black tailcoat, and a pointed beard reminiscent of President Lincoln. An image search for the term 'Jonathan' finds many of these encounters.

This references an earlier *Punch* cartoon 'Am I not a Man and a Brother' (1 June 1844: 235) showing a black slave in chains. *Punch* also signalled the presence of an American character by using an exaggerated version of American slang. By picking out a set of commonly recurring phrases it is possible to locate these items of dialect humour using the *Punch Historical Archive*'s search engine. The phrase "I guess" was considered by the magazine to be distinctively American and was

regularly used by its Yankee speakers. However, its effectiveness as a search term declines towards the end of the nineteenth century as the phrase entered British speech. Other useful phrases include "I reckon", "I calculate", "aint", "lick", and "wal", though the usefulness of these terms also fluctuates over time. The most effective strategy is to find one good passage of American slang in the period that you wish to explore and then search for all of the slang terms it uses in the hope of finding them in other articles. Alternatively, searching for the phrase "as the Americans say" or "as the Yankees say" can sometimes reveal which transatlantic slang terms were currently on *Punch*'s radar. Similar approaches might be used to identify speakers of other distinctive dialects.

Unreliable Search Terms

Finally, it is important to recognise that many potentially useful search terms can be problematic. The United States was sometimes abbreviated to 'U.S.', 'U.S.A.', or variations without the periods. The *Punch Historical Archive*'s search engine struggles to deal with such short search terms. Because it is not case-sensitive, the search engine reads 'us' and 'US' as the same - this makes it difficult to pick out references to the country. Similarly, while the archive does recognise periods as punctuation, it struggles to make sense of the search "U.S.A" and often returns rather confusing results. The usefulness and accuracy of these searches improves for the twentieth century, when the abbreviation USA became more commonly used.

Some of the techniques described above can be used in combination. The Advanced Search interface allows us

to stack up to 10 searches together. The default interface provides three search boxes, but seven more can be added using the 'Add Row' button. In order to stack this many searches together it is vital to change the boxes on the left-hand side to 'Or' rather than 'And'. If you don't make this alteration, the search will look for an article that includes all of these search terms rather than any of them. When it's configured correctly, this search returns over 40,000 results. Switching to a 'keyword' search rather than an 'entire document' search (which should, theoretically, produce more focused results) cuts this down to 10,000 articles, though some of the most interesting passing references to America will be lost.

Conclusions

No combination of searches will find everything to do with America in a single stroke; *Punch*'s subtle satirical references and complex use of allegory require a human eye to pick out, particularly when it comes to illustrations. However, the *Punch Historical Archive* does allow us to explore and interrogate the magazine in powerful new ways. A simple search for 'America' will return plenty of results, but a more sophisticated range of search terms allows us to probe the archive with more power and precision. By searching for distinctively American people, places, slang terms and characters we can identify thousands of articles that basic searches would have missed. By making use of wildcards ('America*') and proximity operators ('press n10 America*'), we can greatly expand the range of our searches and eliminate a large number of irrelevant results. The results leave us with a remarkable amount of new information to process; a set of sources that will allow us to map the turbulent relationship between

John Bull and Brother Jonathan, and the role that *Punch* played in shaping it, in new detail.

Crucially, this process of exploring and refining keywords is essential for any kind of digital research project. Not every topic will require you to manage tens of thousands of results, but even smaller projects demand careful attention to search terms in order to find most of the relevant sources. As we've seen, this can be a complicated process and often involves the combination of close reading and advanced digital search techniques. Fortunately, if you venture beyond the basic search box, you'll soon discover that the *Punch Historical Archive* provides a set of powerful advanced search tools. While these tools do not necessarily make digital research easy, they do allow us to interrogate the magazine in ways that were hitherto impossible. The pages of *Punch* have been carefully scrutinised by successive generations of historians, but digitisation of the magazine now allows us to explore it in exciting new ways. Who knows what we might find?

CITATION

Nicholson, Bob: "In Search of America: An Introduction to Digital Research Techniques." *Punch Historical Archive 1841-1992*: Cengage Learning 2014

© Cengage Learning 2014

**REQUEST A
FREE TRIAL**

**CONTACT YOUR
REPRESENTATIVE**

**FIND OUT MORE
ON OUR WEBSITE**



GALE DIGITAL SCHOLAR LAB

DISCOVER THE POSSIBILITIES