

COMMENTARY

Searching maps by words: how machine learning changes the way we explore map collections

Valeria Vitale¹¹ University of Sheffield

Keywords: machine learning, digital collections, maps, cultural heritage

<https://doi.org/10.22148/001c.74293>

Journal of Cultural AnalyticsVol. 8, Issue 1, 2023

Large numbers of maps have always been difficult to examine in detail, even now that they are being digitized around the world. But imagine searching digitised map collections by their text content: moving beyond titles or other catalogue fields, you could search every single word that appears on map sheets, as if they were book pages in any of the well-known, full-text-search enabled collections. This experience is now a reality. This piece is a data-driven journey across such experimental “text on maps” searching in the online interface for one of the largest and best-known digital libraries of maps, the David Rumsey Map Collection. Starting from the search for a single placename, the author discusses potential, as well as the limitations, of this approach, and suggests ways in which this new interface, which brings together the power of machine learning, the beauty of data visualisation, and the interactivity of annotation, can fuel scientific curiosity as well as playful exploration

Maps are peculiar artefacts. We are surrounded by them, and yet they still fascinate us. Their role as witnesses (more or less reliable) of other places and times makes them invaluable as cultural documents, but also as travelling devices for the imagination. The relevance of maps as historical sources is getting increasingly recognised in humanities research, but it is currently very difficult, if not impossible, to use them efficiently and extensively, even if they have been digitised at scale.

Now, imagine how different it would be if we could search digitised map collections *by word*. Not just in the titles or metadata, but *every single word that appears on the map*, as if maps were pages of a book. We would have access to a completely new kind of complex corpora that could be explored and analysed with a multiplicity of approaches, from Natural Language Processing to the Spatial Humanities, and we’d be able to interrogate this information in entirely novel ways. Searching the text on maps would also change the way we use map collections and interact with them at all levels, enabling incredibly wider, richer and more stimulating search results that would feed scientific curiosity as well as playful exploration.

If you find this idea exciting, you may be happy to discover that soon these kind of experimental features will be made available on one of the largest and best known digital libraries of maps, the [David Rumsey Map Collection](#) (DRMC), home to more than 150,000 digitised maps from all over the world, all publicly available. Thanks to a collaboration with the [Machines Reading Maps](#) project, from this May all users of the DRMC will be able to discover the



[Roux, J., Pl. 57. Tarante \(Taranto\). 1764](#)

new functionalities that have been implemented bringing together the power of machine learning, the beauty of data visualisation, and the interactivity of annotation.

As a member of the team, I had the opportunity to experiment with the new tools while they were being developed. As soon as I had the chance, I did what most people would: I searched for the name of my hometown. I am from a lesser known place in the South of Italy: Taranto. It's not particularly small; on the contrary, it's actually a large city. But it's not a place you often hear about, and it's not a popular travel destination. Unless you are into antiquities, of course. In that case, you probably know about our stunning archaeological museum, and the fact that we used to be a proud Spartan colony and the capital of Magna Grecia.

Usually, when I search for Taranto in any digital collection, I don't get a very long list of results. Even in the David Rumsey Map Collection, a traditional query for "Taranto" returns 13 maps. The number includes some gems in at least three different languages, in which Taranto is featured prominently. But this is also the heart of the problem: I will only find the maps where the word "Taranto" appears in the title, or in some other metadata field.

These new features I'm testing, though, have made things very different. The machine learning pipeline ([mapKurator](#)) developed by Yao-Yi Chiang and students at the Knowledge Computing Lab at University of Minnesota has collected the text on about 60,000 georeferenced maps in the DRMC and stored it in geojson files, one per map sheet. What happens behind the scenes? Each digitised and georeferenced map gets divided into small patches, to facilitate the processing. Then, a text spotter algorithm ([TESTR](#)) is used to detect and recognise text labels on these smaller portions of the maps. This



First page of the results for the query “Taranto” in the text on maps in the DRMC. Masonry View

is not an easy task, as text on maps shows a wide variety of fonts, sizes, and orientations, especially in a large and diverse collection like David Rumsey’s. To make things even more challenging, there isn’t a lot of training data specific for maps to “teach” the algorithm how to make more and more reliable “readings”. So Chiang and his team have generated a large number of [synthetic maps](#) with contemporary data but with “fake” historical styles to train the model. Then the patches are recomposed, geo-coordinates are assigned to the labels, and the initial results are matched, where possible, against an Open Street Map dictionary to strengthen the accuracy of the machine-generated predictions.

In other words, I can now find (almost) all the maps in the collection where a specific name appears as text on the map. For example, that of my hometown. I am curious to compare the experience of searching catalogue metadata versus all the text on maps. I type “Taranto” in the new search field, and I see the name materialising on the page in many colours, orientations, and fonts, forming a mesmerising collage, almost a calligram. I am a migrant; I can’t help smiling with tenderness

Like playing a marimba, I slide the pointer across the bricks of what the [Luna Imaging](#) Team has called a “masonry view.” A preview of the full map unfolds for each of the little rectangles, putting the place name in context.

Many are maps of Italy. That was expected. These are the kind of searches I would have tried if I were to find manually the most likely candidates to feature Taranto. But maps are less predictable than we sometimes think. They don’t merely follow national boundaries, they make cultural statements.



[Lowry, J.W., & Sharpe, J. North Africa. 1848](#)



[Colton, G.W. North Eastern Africa. 1869](#)

I spot Taranto on a map of North Africa, on a map of Turkey, on a map of Mediterranean port cities. And I am reminded of our multi-layered history and mixed identities. Italians, sure, but also Mediterraneans, Southerners, Levantine.



[Johnston, A. K., Mediterranean Basin. 1861](#)

The dialects in my region are a blend of Arabic, Greek, Turkish, Spanish, French. These unanticipated inclusions and exclusions make me think. I savour the beauty of these invisible labels I didn't know I was carrying in the eyes of the world, now made tangible by this array of maps.

As a researcher, I work with historical artefacts, so my attention is caught by the fonts and styles of older maps. I can see my city represented at different moments in time, I can relive her changing fortunes. I see her as a prominent Greek port, and then declining during the Roman Empire. I stumble upon maps featuring Taranto at the time the enlightened reign of Frederick the Second, the *Stupor Mundi*, the king who loved poetry and falconry, and under whom Christians, Muslims and Jews thrived together in the blessed island of Sicily. I find her appearing in several maps that trace the ever changing political landscape of Italy prior to national unification; the South, though, quite firmly under the rule of the Bourbon dynasty. I click on an Austrian map of 18th-century Europe, and I am struck by some thick black lines, one just around mount Vesuvius.

It took me a moment to realise that this map is featuring the fledgling, and still disconnected, railways of Europe. Not a network yet, but still a wonder to behold. That black circle is the *Circumvesuviana*, the pioneering railway built to transport a growing number of tourists to see the very first public archaeological sites: the buried cities of Pompeii and Herculaneum. The map reminds me of a (brief) time when the Kingdom of Naples was a progressive place to be, and the South of Italy was not plagued yet by a chronic absence or inadequacy of infrastructure. It also takes me back to the times of my



[Mitchell, S. A. Jr., Map of the Austrian Empire, Italian States, Turkey in Europe, and Greece. 1865](#)

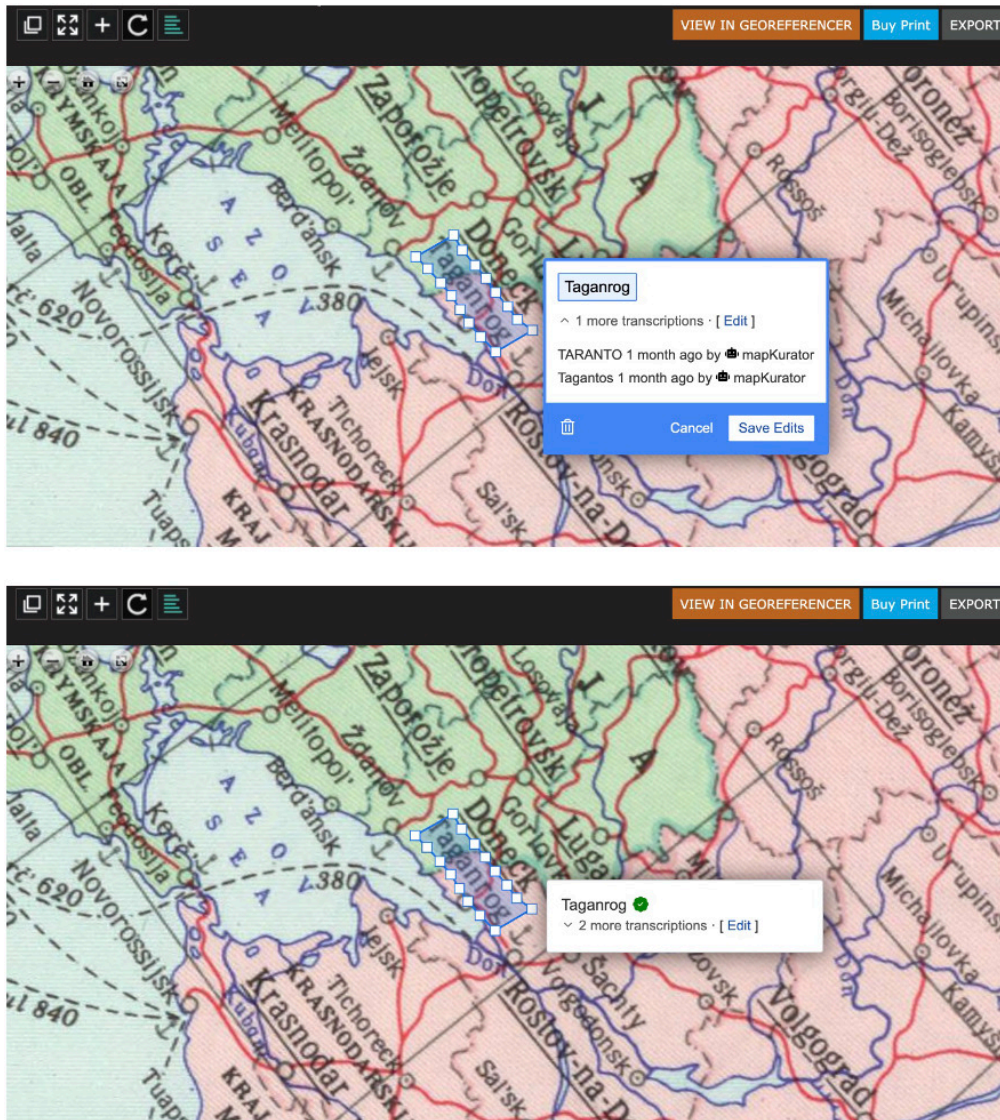
doctoral research, when I was studying the *Grand Tour* and interpretations of ancient buildings near Vesuvio. Maybe “Pompeii” will be my next search in the Rumsey Map Collection.

Back on the search results page, glancing at the bricks of this *opus incertum*, I can spot some mistakes. Of course, the machine is not perfect. It does not recognise semantics: it has never seen a fiery sunset in my hometown, nor has it ever cried with anger witnessing industrial pollution destroying the landscape and quality of life. Some names are odd, out of place. Quite literally. These are false positives: places that the algorithm has mistaken for Taranto.

I decide I should help the machine and correct the results, using the new annotation feature, developed by Rainer Simon. It’s simple and minimal, almost like taking your pen out of your breast pocket. *Here: it’s done already.*

My correction will now appear as the latest transcription, while all the history of user- and machine-generated edits still remains available for transparency. With time, the false positives will become fewer and fewer. For the moment, though, I am almost endeared by some recurring mistakes in the results.

I wonder where these accidental twins of my birth city are. Is Tarantou also a port city? Do people in Taganrog also have a dolphin on their flag? I wonder if I’ll ever visit these places, if I should one day plan a pilgrimage to all the mistaken Tarantos of the world. I know it won’t be long before (some) recurring errors in machine generated data will become insightful research questions instead of noise.



[Screenshots showing the annotation features in action. The map is Union of Soviet Socialist Republics. The World Atlas. 1967](#)

Other times, though, the algorithm has been surprisingly successful. It deals efficiently with the many background colours, on land and sea, with the old fashioned serif fonts, with the hachuring and contour marks. Even with the shape of the old town in Taranto, a small island only attached to land by two bridges. The floating piece of earth divides the two internal bays, or *i due mari* (the two seas) as we call them: the Big Sea and the Small Sea. On maps, it often hosts an “r” or a “a”, confusing the waters, making it hard for a machine to distinguish the words from the curves of the landscape. But the algorithm finds its way amidst this intricacy of lines, unravelling the tapestry of letters and unique geography.

Another word catches my attention. Another odd brick. It’s not the canonical name of my city, but it’s not a mistake either. It’s a variant spelling of the place name, “Tarento”, reminiscent of its latin form, *Tarentum*. It was common in the past, and traces of its former use can be found not only in old maps, but



[Bartholomew, J. G., Composite: Europe - communications. 1922](#)



[Thomas Bros. Map of San Diego, National City & La Mesa. 1938](#)

also in the deprecated adjective “Tarentini”. The name of the city and the one for the people living in it have long been replaced by the current “Taranto” and “Tarantini”, a choice more consistent with the Greek original place-name, *Taras*, which is also the name of the city’s mythic founder, who highlighted on our shores riding a dolphin.

We always had a complicated relationship with the Romans, our allegiance to Hannibal and his elephants never forgotten. And we still feel a connection to Greece, as you can tell from many of our street names: *Leonida*, *Temenide*, *Liside*, *Magna Grecia*. Street names can tell a lot about the story, and the spirit, of a city. So I'm delighted when, clicking on the unusual and obsolete spelling "Tarento" I discover it is not an old map of Italy, or even Europe. It's a street name in San Diego, California. I am now curious, and I teleport myself (digitally, of course) there.

I spot other names of small Italian and European cities. I wonder if some nostalgic immigrants named the places in the neighbourhood some decades ago. And if this is why they used a spelling that is now gone everywhere else. The past crystallised in toponymy like an ammonite.

I refresh the search page. Like in a fantasy tale, the word-bricks in the masonry rearrange each time, suggesting new travels, new itineraries of memory and imagination. I haven't moved from my chair, but my hometown, somehow, feels a little bit closer.

Acknowledgement

Many thanks to Katie McDonough for the many useful comments and suggestions, and for being such a generous and patient editor.

Funding

Machines Reading Maps is a UK-US collaboration, funded by the AHRC (award ref. AH/V009400/1) and the NEH (award ref. HC-278125-21). The team includes researchers from The Alan Turing Institute, the University of Minnesota, The Austrian Institute of Technology and the University of Southern California Library.

The work on the *David Rumsey Map Collection* was developed in collaboration with *Luna Imaging*, and generously funded by David Rumsey.

Submitted: March 02, 2022 EST, Accepted: March 16, 2023 EST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.