Mapping the *Encyclopédie*: Working Towards an Early Modern Digital Gazetteer*

Katherine McDonough
Western Sydney University
Australia
kmcdono2@gmail.com

Matje van de Camp
Netherlands
matje@taalmonsters.nl

ABSTRACT

Historians investigating evidence with spatial significance increasingly rely on gazetteers to identify the location of geographical features/places. Existing digital gazetteers cater to twenty-first century or discrete historical geographies (the classical world, for example). Early modernists (ca. 1450-1750), particularly those who work on non-Anglophone cultures, represent a major scholarly community with no temporally-appropriate gazetteers available. In this paper, we introduce a project that fills this research infrastructure gap. Mapping place names in the canonical eighteenth-century *Encyclopédie* is a case study for semi-automating the identification, classification, and location of places and spatial relations in historical geographic reference works printed in French. We demonstrate the challenges of using existing geoparsers and introduce our plan for new tools and protocols for working with historical French texts.

KEYWORDS

geographic text analysis; early modern history; French; natural language processing; named entity recognition; gazetteer

ACM Reference format:

2017. Mapping the *Encyclopédie*: Working Towards an Early Modern Digital Gazetteer. In *Proceedings of GeoHumanities'17:1st ACM SIGSPATIAL Workshop on Geospatial Humanities, Los Angeles Area, CA, USA, November 2017 (GeoHumanities'17), 7 pages.*
DOI: 10.1145/3149858.3149861

1 INTRODUCTION

Currently available digital gazetteers fall short of research needs for early modern historical research. In the centuries before modern cartography and the rise of the nation-state, both political geographies and place name documentation were in flux. France presents a special case, where, because of the dramatic transformation of names and jurisdictions during the French Revolution, communities and administrative spaces before 1790 are not easy matches for their Revolutionary or post-Revolutionary counterparts [1]. Furthermore, people living in Europe during the early modern period (ca. 1450-1750) witnessed an explosion of printed dictionaries, encyclopedias, and other types of reference works. In these sources, we have an abundance of evidence for how early modern geographic horizons were changing and how spatial information was being publicized in new ways. Using these sources, we will 1) create a new digital gazetteer based on attestations of places in early modern geographic reference works and in future research 2) examine the way in which geographic information about France, Europe, and the world was produced and reproduced. How do cultures of geographic text reuse change? How is information about place structured in early modern reference texts? In this article, we focus on the preparation required to begin the first task, the digital infrastructure resource that will ease the way for future spatial histories that depend on interpretation of place-specific evidence.

Before we can develop a period-specific gazetteer for early modern European and global research, we face a range of challenges ranging from the lack of Natural Language Processing tools for the French language [2], [3] to limitations in the way that groundbreaking early Geographic Text Analysis (GTA) research has depended on geoparsing tools that are difficult to adapt to historical use cases [4]–[6]. These challenges motivate us to develop new Natural Language Processing tools for working with historical French texts. Our initial work uses the ARTFL version of the eighteenth-century *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* by Diderot and d'Alembert

GeoHumanities'17, November 7–10, 2017, Los Angeles Area, CA, USA

ACM ISBN 978-1-4503-5496-7/17/11 $15.00
https://doi.org/10.1145/3149858.3149861

[7]. Using only the "geography" classified articles (as defined by the eighteenth-century editors), we designed a gold-standard annotation exercise to test the effectiveness of the Edinburgh Geoparser against human annotators in identifying places mentioned in the articles and matching them to a gazetteer record (Geonames) with latitude and longitude coordinates. This gold standard experiment has helped us to pinpoint which parts of the geoparser pipeline are most problematic when working with non-English texts, evaluate what kind of spatial data model fits our evidence, and consider what we can learn about the form and content of the *Encyclopédie* geography articles from this sample. Our work led to the development of a new annotation interface, GeoViz, with greater metadata editing privileges than current platforms like Recogito currently permit. This interface was intentionally designed to account for early modern textual evidence about place.

2 REVIEW OF LITERATURE

Several overlapping literatures play a role in this project, inspiring the approach to early modern French texts that we borrow from recent NLP research. Drawing on the rich historiography in the history of science and intellectual history, we hope to take further recent work on spaces and places of the Enlightenment [8]–[11].

The spatial turn in humanities research has supported a new community of scholars working to develop new technologies for spatial analysis of historical evidence. The deep maps and enriched gazetteers that have been called for [12], [13] depend on a rich metadata environment that operates as a foundation for evidence we wish to pin down on a map. Gazetteer-makers have invested substantial resources to create this infrastructure that both links humanities research together and to the growing world of linked open data [14]–[18]. As spatial data resources that are created by states (Geographic Names Information System) and volunteers (Geonames) are increasingly in competition with each other to operate as place-name authority records, scholars and institutions are working to clarify where spatial information comes from and how it is chronologically and culturally sensitive. In particular, recent efforts to leverage machine learning have resulted in new projects that aim to semi-automate place-name and other entity recognition processes on digitized texts and images that document knowledge about place in a specific moment in the past [19]–[21]. Many of these projects have worked closely with the Edinburgh Geoparser team which has been instrumental in adapting NLP tools for historical research [22]–[25]. Despite this growing interest in digitizing earlier print historical gazetteers, developing new ones, and creating platforms where users can query multiple gazetteers simultaneously (e.g. in Recogito, one can compare records from Geonames, Pleiades, and DARE, the latter two being ancient-specific gazetteers) a major gap exists for using NLP tools in place-name research with texts that are not in English.

Beyond these language challenges, some projects have begun to address data modeling issues that are also at the heart of this project and are applicable across languages. Historical texts and maps regularly represent information about imaginary places, places on other planets, places that no longer exist, places that are part of a formal name, institution, or event. The latter are particularly important in French, where noble and royal titles regularly incorporate the names of places. This example highlights the advantage of incorporating tree-structured NER, the automation of identifying structured entities (*roi de France*), instead of only recognizing "France" as a place independent of the person who is really being described [26]. Furthermore, rather than thinking of texts as having attestations only of real places that exist in one location, we have an opportunity to embrace the ambiguity of textual evidence about early modern place-related knowledge. Our early efforts to consider how we can use context in each geography article of the *Encyclopédie* to generate a likely location area for places with no clear match to current gazetteers grows out of ongoing research in information retrieval processes [27], [28] and spatial relation extraction [29]–[32] as well as other geographic text analysis research on imaginary places and literary mapping [33].

Finally, other researchers digitizing historical geographic sources in France are working towards a shared goal of improving infrastructure for spatial history. The GeoHistoricalData team at the EHESS has been digitizing the features of the Cassini map sheets by hand (this is the eighteenth-century map that became the foundation for the national cadastral map of the 19th century). They have done significant work to crowdsource historical map feature transcription, and compare current political geography data and earlier demographic research at the EHESS [34] with this Cassini map-derived data [35], [36].

3 METHODOLOGY

The first objective was to evaluate how well the Edinburgh Geoparser (EG) functioned with only minor changes. With the assistance of the EG development team, we received a modified software package that included a French Part-of-Speech Tagger, but not a French trained Named Entity Recognition tool. We received an XML file of all Geography articles from the *Encyclopédie* (14,445 articles, or extracts as we will refer to them). After removing XML formatting and splitting each article into a separate text file, we ran the set of text files through the EG. Each file thus represents one article about a place: the "head word" or article entry title is a place name.

The original Encyclopédie classifications for geography articles are broken down into géographie moderne (modern geography), géographie ancienne (ancient geography), géographie (geography, this includes feature types which skews the set slightly), and géographie sainte (religious geography). The articles are of widely varying length and sometimes include latitude and longitude measurements. They frequently discuss the multiple locations of one place, whether because the place no longer exists and there are multiple hypotheses or because the place has multiple competing attestations in the manuscript/print record available to the author (Lissus or Germanicopolis, for example).

Our version of the EG with the French POS Tagger was also adapted to call out only to the Geonames gazetteer. We had difficulties establishing the connection to the Pleiades+ gazetteer, which was unfortunate because of the high number of ancient place names in the Encyclopédie. We did not limit results from Geonames based on feature type. (By comparison, Recogito, the Pelagios Commons tool for annotating place names, does limit Geonames results to populated places over 5,000). It was important to us to see how French spellings of names in the alternate name field of Geonames records were treated in the EG search. We also disabled the option to have multiple results for geolocation of identified place names. Only the top match was saved.

We loaded the full text files with geoparser annotations into a database and built an interface, GeoViz, to visualize the results [37]. The interface also operates as an annotation interface, where researchers can verify, edit, and add new information to extract annotations. GeoViz is the interface where we performed the gold standard annotations, creating user accounts for three separate researchers (Katherine McDonough, Laure Philip, and Simon Burrows) who would independently annotate 100 randomly selected extracts from the corpus. The 100 extracts contained some original articles that did not meet the criteria of being articles about places (e.g. "rivière," "haut," "holm"). We replaced 6 of this type of extract with place-name extracts.

GeoViz (Figure 1) was designed to allow researchers to simultaneously consult a full text, annotation metadata, and a map of any gazetteer record matches for the place-name entities identified in the text. Human annotators could check results from geolocated place names by selecting the place name in the text and viewing the metadata and map. If needed, they could edit the metadata and validate the new information. When place names were not identified by the EG, researchers could highlight the words of the place name in the text box and perform a gazetteer look-up on the fly for both Geonames and DBpedia.

It should be noted that the DBpedia look-up was a feature added after initial tests revealed several instances when Wikipedia entries contained better matches than Geonames records. DBpedia was selected because of previous experience with this knowledge graph, with the understanding that the switch to Wikidata would be made after the gold standard exercise. In the future, we will design a federated look-up of both DBpedia and Wikidata, to ensure multiple language coverage. (Unfortunately, it is not currently possible to query DBpedia without also querying Geonames, hence the breakdown of results in Table 3.) Wikidata maintains, for our purposes, a clear advantage in terms of returning results (from site links) that draw on Wikipedia pages written in different languages.

GeoViz also currently includes additional data fields that captured contextual information about place names (location type; if the place name is part of a person's name/title; if the place name is the same as the place the article is about [its headword], if it is "near" the main location, its position with regards to main [headword] location, and additional remarks). The positional contextual clues, added after the extracts were run through the EG, are our main source of data for testing how we can use spatial relations to indicate zones of relevance when exact coordinates are not possible to record.

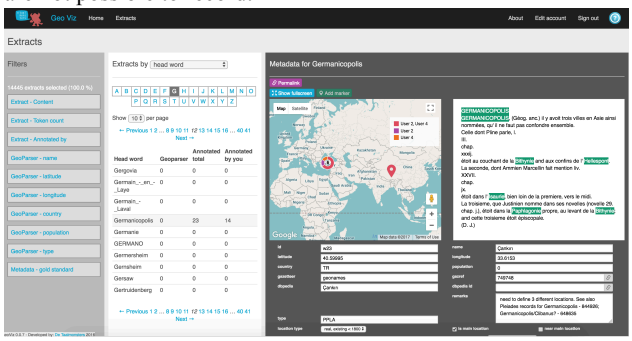

Figure 1. GeoViz annotation interface showing Germanicopolis article from the Encyclopédie.

The gold standard exercise fulfills three needs: it allowed us to evaluate the EG results when minimal changes had been made to the software; to familiarize ourselves with the content and structure of the extracts; and to establish a set of annotations that can later be divided into a training/development set and a test set. This final objective is part of the larger goal of training a new Named Entity Recognition (NER) classifier that will work best with this kind of historic French reference text. NER tools for French (generally, a low-resource language in NLP) are limited, and those that have been trained on historical texts even rarer. In addition to our own human annotations, we are also evaluating WikiNER data from French Wikipedia articles as an additional set of training data [38]–[41].

As we continue to develop GeoViz, we aim to make this service available to other scholars. For us, GeoViz filled a software gap between the EG output and Recogito. It allowed us to edit and add new metadata to the EG results and provided flexibility determining when current gazetteer options (in our case Geonames and DBpedia) did not meet our needs. In a future version, we will adapt GeoViz to serve as a gazetteer record validator. We have already incorporated some features that will be expanded upon when we move beyond the Encyclopédie text. For example, where latitude is indicated in the extracts, a horizontal green line appears on the map as an aid in making possible matches to gazetteer records (which, when selected, are pinned on the map for comparison purposes) (Figure 2).

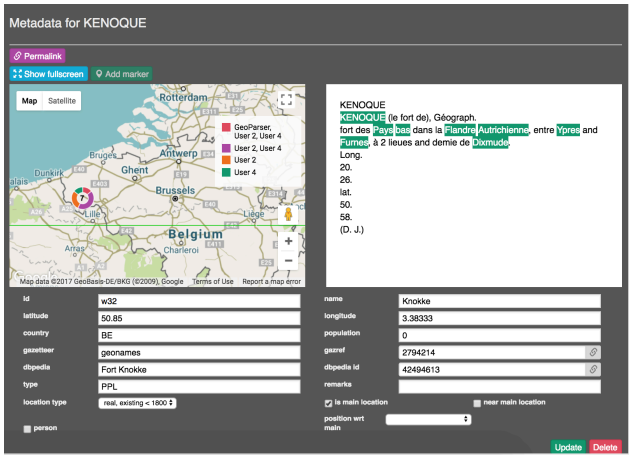

Figure 2. Detail of map and metadata for Kenoque extract.

4 RESULTS

Our results reflect data collected from the EG and three human annotators. Unfortunately, not all three annotators were able to complete the entire gold standard set. Results cover basic analysis of way that place-names were annotated by the EG and the human annotators and point to new possibilities for using contextual evidence to represent complex spatial relations. Despite our original sample of 100 extracts and the replacement of non-place-name records, our final results contain information from 103 extracts where only 96 extracts had an identified head word. Table 1 breaks down the totals for how many extracts included annotations by the Geoparser (61 no, 42 yes) and how many included annotations by the 3 human annotators (2 by none, 10 by 1 person, 81 by 2 people, and 19 by three people).

Table 1: **Extracts with Annotations by Geoparser (GP) or Humans (H)**

	0 H	1 H	2 H	3 H	total
No GP	0	7	51	3	61
GP	2	3	30	7	42
total	2	10	81	19	103

Note that while the EG found at least one location in 42 extracts, its recall is poor (Table 2). KM (Katherine McDonough) found an average of 10.44 places per extract compared to the 1.27 from the Geoparser. All of the human annotation average come in well above the EG result, providing evidence for how many places the EG does not even classify as place entities.

Table 2: **Average Number of Annotations per Extract**

GP	1.27
KM (100)	10.44
LP (92)	11.45
SB (10)	7.0

Table 3: **Gazetteers**

Gazetteer type	Places matching gazetteer record
geonames	485 (22.37 %)
dbpedia	0 (0.0 %)
both	1679 (77.44 %)
none	4 (0.18 %)

Table 4: **Human location types**

Location type	Places identified as type
real, existing < 1800	1710
real, historical	420
biblical	12
unknown	4
mythical	4
real, extra terrestrial	1

Taking EG and human annotations combined, we can see that while initial matches were always to Geonames only, it was possible for human annotators to find a DBpedia match 77% of the time. Table 3 is slightly misleading because GeoViz currently does not support manual entry of DBpedia metadata unless there is also a Geonames match. When there was a match, the DBpedia link was entered manually into the remarks field. Having human annotators run a DBpedia search to match was a useful exercise for revealing how common it was to find specific matches to historical places when Geonames had less relevant modern ones. These alignment decisions were particularly challenging for articles where the author referred to different names for the same place as they changed over time (for example, Zimara).

Location types as annotated by the human annotators are detailed in Table 4. Judging from this table, the EG can only be expected to identify max ~75% of the locations when using only modern gazetteers (real, existing <1800 category is just over 75% of locations).

The next set of results concerns annotation reliability. These are measured by calculating the Inter-Annotator Agreement (IAA). We use Cohen's kappa to compare each pair of annotators. It takes into account the chance that the observed annotations are generated by chance. Scores closer to 1.0 indicate more reliable results. Generally, a margin of 0.05 is considered acceptable, meaning that anything above 0.95 is deemed reliable enough, for instance, for use in a Machine Learning application.

Table 5: **Word ID agreement between Annotators (A1 & A2)**

A1	A2	extracts	words	annotated words A1	annotated words A2	kappa
GP	KM	100	30621	122	1214	0.9635
GP	LP	92	28578	117	1111	0.9646
GP	SB	10	1160	26	80	0.9413
KM	LP	91	28516	1107	1107	0.9849
KM	SB	10	1160	92	80	0.9836
LP	SB	10	1160	84	80	0.9775

Table 6: **Gazetteer reference agreement**

A1	A2	extracts	words	annotated words A1	annotated words A2	kappa
GP	KM	100	30621	122	1210	0.9616
GP	LP	92	28578	117	1111	0.9631
GP	SB	10	1160	26	80	0.9293
KM	LP	91	28516	1103	1107	0.9714
KM	SB	10	1160	89	80	0.9603
LP	SB	10	1160	84	80	0.9620

The Word ID agreement (Table 5) measures whether the words that chosen to mark location names agree sufficiently. It does not look at the location identification. Agreement between humans and geoparser is high (though lower than human-human), but this is mostly due to the fact that only a small percentage of the words are locations. The words that are not locations also weigh in on the score, so even though humans found far more locations than the geoparser did, the number is small compared to the total number of words. Thus, a small difference in the Kappa (agreement) score translates to a big difference in the annotations. For example, the higher Kappa score for human-human annotators (0.98 between KM and LP) reflects that human choices had more in common with each other than human-geoparser comparisons (0.96 between KM and the Geoparser).

Gazetteer reference alignment in Table 6 measures agreement between identified locations. (The difference in the number of annotations between Word ID and Gazetteer agreement indicates the places in the text that do not have matches in the gazetteers.) Again, scores were highest between KM and LP, suggesting that two independent annotators who are fluent in French and knowledgeable in French geography and early modern history could reach substantial agreement in making annotations. But a fairly high score between the Geoparser and KM also highlights that when the EG identifies a place name, it does usually succeed in making the same match that the expert annotator did. This high rate of precision emphasized to us the importance of refining the NER part of the Geoparser pipeline rather than the georesolution side. The EG can match an identified place name to a location, but the major problem for non-English texts is identifying the place name.

5 DISCUSSION

Based on previous GTA work and conversations with the EG developers, we knew in advance that the EG would have mixed results in identifying place names written in languages other than English and finding the correct locations for those place names. The gold standard exercise confirmed our concerns for the first issue. If we are able to improve identification rates, then we will need to return to the georesolution phase.

This experiment highlighted a range of challenges that we hope to take up in future research. First, conflicting protocols for establishing alignments to existing gazetteer metadata should be foregrounded. What are the consequences of assigning pre-existing metadata from Geonames to and eighteenth-century attestation? How are these a “match”? The sources of geospatial information are historical, and our metadata choices should reflect that. Our goal is eventually to document place-related knowledge across different early modern reference works. Geonames is embedded in the history of the chronology of geographic knowledge (re-)production and our data models should reflect this.

We will need to define a data model that balances disambiguation, fuzziness, and documentation of attestation provenance (something that Pleiades does well). Our documentation of location types (real in 1800, historical, biblical,

etc.), for example, begs the question of whether Paris has the same metadata as Lutèce in the context of the Encyclopédie. The rampant recourse to conflicting attestations for locations of mythical Greek places, the problematic use of exonyms to name places in the Ottoman Empire, Egypt, and beyond: these are all regular features in early modern geographic texts that do not neatly fit into a one place=one location model. One promising pathway towards addressing locational ambiguity is to improve automation of contextual spatial relations. If we can track fuzzy spatial relationships to identify a "seismic activity" area for places with no fixed or multiple locations, we can use this to create a heat-map like visual for the places that digital gazetteers have left behind.

The Encyclopédie extracts are echoes of the textual tradition of sharing geographic knowledge. As a whole, the Encyclopédie holds many texts within it, perhaps a gazetteer of gazetteers. The depth of knowledge about place in these volumes defies a simple 1:1 or even 1:3 place-name to location identification. Moving beyond the "match" to modeling other kinds of relations can bolster our ability to deal with available early modern geographical information. Indeed, working with the Encyclopédie has raised many issues with regards to what counts as an attestation. The use of geography articles in the Encyclopédie to raise doubts and document competing attestations in older literature about place names means that any transformation of this evidence into machine-readable data will need to account for this uncertainty. If we have embraced Web 2.0 technologies like Geonames in our enriched digital history infrastructure, we also need to approach these resources critically as inheritors of multi-layered spatial information that is more than a basic tool for geolocation. By developing new methods for digitally extracting and representing spatial information from early modern texts and critically analyzing the same in modern resources, we set both of these along a continuum of knowledge production that can be an object of study.

6 CONCLUSION

Annotating the Encyclopédie geography articles demonstrates the messiness of canonical reference sources. If we want to locate early modern places in digital research, we have to contend with the challenges of attestations of places that do not conform to simple expressions of unique latitude/longitude coordinates. To date, the gazetteer community has talked a lot about gazetteer alignment and what this means in designing our data models and expressing these as linked open data [42]. But working with the early modern material brings home how much we have left to discuss about just what "alignment" means and how we can express complex relationships between pieces of evidence in digital ways.

REFERENCES

[1] K. McDonough, "Putting the Eighteenth Century on the Map: French Geospatial Data for Digital Humanities Research," in *Digitizing Enlightenment*, S. Burrows and G. Roe, eds. Oxford Studies in the Enlightenment, 2019.

[2] A. Azpeitia, M. Cuadros, S. Gaines, and G. Rigau, "NERC-fr: Supervised Named Entity Recognition for French," in *Text, Speech and Dialogue*, P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds. Springer International Publishing, 2014, pp. 158–165.

[3] S. Rosset, C. Grouin, K. Fort, O. Galibert, J. Kahn, and P. Zweigenbaum, "Structured Named Entities in Two Distinct Press Corpora: Contemporary Broadcast News and Old Newspapers," in *Proceedings of the Sixth Linguistic Annotation Workshop*, Stroudsburg, PA, USA, 2012, pp. 40–48.

[4] C. J. Rupp *et al.*, "Customising Geoparsing and Georeferencing for Historical Texts," *IEEE Int. Conf. Big Data*, 2013, pp. 59–62.

[5] P. Murrieta-Flores and I. Gregory, "Further Frontiers in GIS: Extending Spatial Analysis to Textual Sources in Archaeology," *Open Archaeol.*, vol. 1, no. 1, 2015.

[6] J. Clifford, B. Alex, C. M. Coates, E. Klein, and A. Watson, "Geoparsing history: Locating commodities in ten million pages of nineteenth-century sources," *Hist. Methods J. Quant. Interdiscip. Hist.*, vol. 49, no. 3, pp. 115–131, Jul. 2016.

[7] R. Morrissey and G. Roe, eds., "Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers, etc., eds. Denis Diderot and Jean le Rond d'Alembert." University of Chicago: ARTFL Encyclopédie Project (Spring 2016 Edition).

[8] R. J. Mayhew, "Geography as the Eye of Enlightenment Historiography," *Mod. Intellect. Hist.*, vol. 7, no. 3, pp. 611–627, Nov. 2010.

[9] C. W. J. Withers and R. J. Mayhew, "Geography: Space, Place and Intellectual History in the Eighteenth Century1," *J. Eighteenth-Century Stud.*, vol. 34, no. 4, pp. 445–452, Dec. 2011.

[10] N. Safier, "The Tenacious Travels of the Torrid Zone and the Global Dimensions of Geographical Knowledge in the Eighteenth Century," *J. Early Mod. Hist.*, vol. 18, no. 1–2, pp. 141–172, Feb. 2014.

[11] C. W. J. Withers, *Placing the Enlightenment: Thinking Geographically about the Age of Reason*. Chicago, US: University of Chicago Press, 2008.

[12] D. J. Bodenhamer, J. Corrigan, and T. M. Harris, *The spatial humanities: GIS and the future of humanities scholarship*. 2010.

[13] R. Mostern, H. Southall, and M. L. Berman, *Placing Names: Enriching and Integrating Gazetteers*. Indiana University Press, 2016.

[14] R. Mostern and I. Johnson, "From Named Place to Naming Event: Creating Gazetteers for History," *Int J Geogr Inf Sci*, vol. 22, no. 10, pp. 1091–1108, Jan. 2008.

[15] R. Mostern, "Historical gazetteers: An experiential perspective, with examples from Chinese history," *Hist. Methods J. Quant. Interdiscip. Hist.*, vol. 41, no. 1, pp. 39–46, 2008.

[16] P. Manning and R. Mostern, "World-Historical Gazetteer," 2015.

[17] M. L. Berman, J. Ahlfeldt, and M. Wick, *Historical Gazetteer System Integration: CHGIS, Regnum Francorum, and GeoNames*. 2012.

[18] H. Southall, R. Mostern, and M. L. Berman, "On historical gazetteers," *Int. J. Humanit. Arts Comput.*, vol. 5, no. 2, pp. 127–145, Oct. 2011.

[19] U. Hinrichs *et al.*, "Trading Consequences: A Case Study of Combining Text Mining and Visualization to Facilitate Document Exploration," *Digit. Scholarsh. Humanit.*, p. fqv046, Oct. 2015.

[20] R. Simon, E. Barker, L. Isaksen, and P. de Soto Cañamares, "Linking early geospatial documents, one place at a time: annotation of geographic documents with Recogito," *E-Perimetron*, vol. 10, no. 2, pp. 49–59, 2015.

[21] D. Cooper, D. of P. D. Cooper, C. Donaldson, and P. Murrieta-Flores, *Literary Mapping in the Digital Age*. Routledge, 2016.

[22] C. Grover *et al.*, "Use of the Edinburgh geoparser for georeferencing digitized historical collections," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 368, no. 1925, pp. 3875–3889, Aug. 2010.

[23] R. Tobin, C. Grover, K. Byrne, J. Reid, and J. Walsh, "Evaluation of Georeferencing," School of Informatics and EDINA, University of Edinburgh, 18-Feb-2010.

[24] C. Grover and R. Tobin, "A Gazetteer and Georeferencing for Historical English Documents," in *Proceedings of LaTeCH 2014 at EACL 2014. Gothenburg, Sweden*, Association for Computational Linguistics, 2014, pp. 119–127.

[25] B. Alex, K. Byrne, C. Grover, and R. Tobin, "Adapting the Edinburgh Geoparser for Historical Georeferencing," *Int. J. Humanit. Arts Comput.*, vol. 9, no. 1, pp. 15–35, Mar. 2015.

[26] M. Dinarelli and S. Rosset, "Tree-Structured Named Entity Recognition on OCR Data: Analysis, Processing and Results," in *Language Resources Evaluation Conference (LREC)*, Istanbul, Turkey, 2012.

[27] B. Adams and K. Janowicz, "On the Geo-Indicativeness of Non-Georeferenced Text," in *ICWSM*, 2012, pp. 375–378.

[28] B. Adams and G. McKenzie, "Inferring thematic places from spatially referenced natural language descriptions," in *Crowdsourcing geographic knowledge*, Springer Netherlands, 2013, pp. 201–221.

[29] P. Kordjamshidi, M. Van Otterlo, and M.-F. Moens, "Spatial role labeling: Towards extraction of spatial relations from natural language," *ACM Trans. Speech Lang. Process. TSLP*, vol. 8, no. 3, p. 4, 2011.
[30] P. Kordjamshidi, P. Frasconi, M. Van Otterlo, M.-F. Moens, and L. De Raedt, "Relational learning for spatial relation extraction from natural language," in *International Conference on Inductive Logic Programming*, 2011, pp. 204–220.
[31] M. Snoussi, J. Gensel, and P.-A. Davoine, "Extending TimeML and SpatialML languages to handle imperfect spatio-temporal information in the context of natural hazards studies," in *Proceedings of AGILE'2012 conference. Springer, Avignon (France)*, 2012, pp. 117–122.
[32] I. Mani *et al.*, "SpatialML: annotation scheme, resources, and evaluation," *Lang. Resour. Eval.*, vol. 44, no. 3, pp. 263–280, 2010.
[33] P. Murrieta-Flores and N. Howell, "Towards the Spatial Analysis of Vague and Imaginary Place and Space: Evolving the Spatial Humanities through Medieval Romance," *J. Map Geogr. Libr.*, vol. 13, no. 1, pp. 29–57, Jan. 2017.
[34] J.-P. Pélissier and C. Motte, *Géonomenclature historique des lieux habités: rapport*. Paris: Direction des Archives de France, 2003.
[35] J. Perret, M. Gribaudi, and M. Barthelemy, "Roads and cities of 18th century France," *Sci. Data*, vol. 2, 2015.

Price:$15.00

[36] R. Cura, B. Dumenieu, J. Perret, and M. Gribaudi, "Historical collaborative geocoding," *ArXiv170307138 Cs*, Mar. 2017.
[37] M. van de Camp, *GeoViz*. http://geoviz.taalmonsters.nl/.
[38] J. Nothman, J. R. Curran, and T. Murphy, "Transforming Wikipedia into named entity training data," in *Proceedings of the Australian Language Technology Workshop*, 2008, pp. 124–132.
[39] J. Nothman, T. Murphy, and J. R. Curran, "Analysing Wikipedia and gold-standard corpora for NER training," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 2009, pp. 612–620.
[40] J. Nothman, N. Ringland, W. Radford, T. Murphy, and J. R. Curran, "Learning multilingual named entity recognition from Wikipedia," *Artif. Intell.*, vol. 194, pp. 151–175, 2013.
[41] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. R. Curran, "Evaluating entity linking with Wikipedia," *Artif. Intell.*, vol. 194, pp. 130–150, 2013.
[42] T. Elliott, "Pleiades Data Model," *Pleiades: a gazetteer of past places*, 09-Apr-2017. [Online]. Available: https://pleiades.stoa.org/help/pleiades-data-model. [Accessed: 08-Aug-2017].