Linking historical streets: a semi-automatic approach to place identification

DH Benelux 2023 | short form paper

Vincent Ducatteeuw[1,2]
0000-0003-4493-6268
@VDucatteeuw

[1] Ghent Centre for Digital Humanities (GhentCDH) & Department of History, Ghent University, Ghent (BE)
[2] Antwerp Cultural Heritage Sciences (ARCHES), University of Antwerp, Antwerp (BE)

Keywords
city directories, digital gazetteers, place identity, data reconciliation, historical streets

Abstract
City directories, which became wide-spread in the late 18th century, contain a wealth of micro-geographic information in the form of household, industry, and street listings (Shaw, 1984; Shaw & Coles, 2019). In the past, the processing of this information was difficult due resource constraints and technological limitations. This has improved in recent decades due to technological advances in semi-automatic data extraction (Berenbaum et al., 2018; di Lenardo et al., 2019; Kappner & Albers, 2022). As a result, the creation of digital gazetteers storing spatiotemporal information on historical places has become less resource-intensive (Bol, 2011; Southall et al., 2011). Yet, to efficiently and unambiguously link records to one another, a methodology is required that addresses the question "Is place A identical to place B?". This formalization of "place" remains a challenge for digital gazetteers. (Shaw, 2016; Merschdorf and Blaschke, 2018; Garbacz et al., 2021). This paper evaluates the efficacy of the identification methodology developed by Garbacz et al. (2018, 2021) to link diachronic historical street observations. The ambition is to create an authority file of historical urban spaces for cultural heritage cataloging. (Danniau et al., 2022; Ducatteeuw, 2021).

For database management purposes, spatial objects are often defined by explicit identity criteria (e.g. name, type, location) to determine if records are related to the same spatial object (Hill, 2000; Zedlitz & Luttenberger, 2014). Garbacz et al. (2021) proposed a set of properties that help to determine the identity of historical localities (i.e. human settlements) over time. These identity criteria consist of 1) proper name, 2) geographical location, 3) place type, and 4) l-mereology, which refers to the part-whole relationship between localities (e.g. locality A is a suburb of locality B). Their methodology states that if a number of these properties change over time, a locality loses its identity if:

"(i) both its name and geographic location are changed or (ii) both its name and type are changed or (iii) both its name and l-mereology are changed, then it is destroyed within this period (and is not identical to any locality that exists at the end of this period)." (Garbacz et al., 2021, 66)

In the present study, the general applicability of this methodology is evaluated for other place types, specifically streets, squares and bridges observations derived from a digitized collection of Ghentian city directories (Delbaere, 1962). These directories were published from 1770 to 1932 and contain a yearly alphabetical street index. Ten of these street listings were sampled, OCR'd and parsed (Ducatteeuw et al., 2022). The resulting datasets total 9,766 place observations, providing information on the following properties for streets, squares and bridges in Ghent:

- Dutch toponym
- French toponym
- Description of the geographical location
- Police and judicial district number[1]

Since places can have multiple toponyms during their existence, it is not sufficient to use place name similarity to determine whether rows in the datasets relate to the same place. As an alternative, all information related to one place observation (Figure 1) was mapped to the properties proposed by Garbacz et al. Following their methodology, specific property changes determined whether a place ceased to exist and was succeeded by a new place (e.g. the division of street A into street B and C).

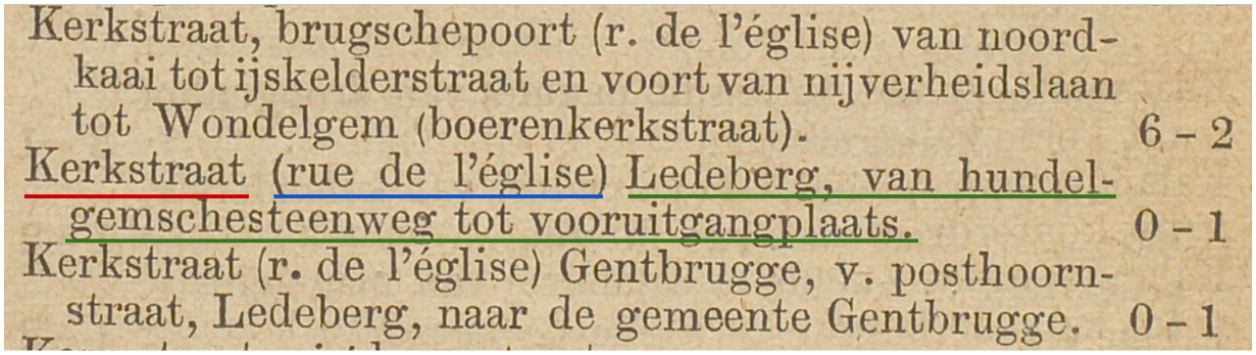
Kerkstraat, brugschepoort (r. de l'église) van noord-
kaai tot ijskelderstraat en voort van nijverheidslaan
tot Wondelgem (boerenkerkstraat). 6 - 2
Kerkstraat (rue de l'église) Ledeberg, van hundel-
gemschesteenweg tot vooruitgangplaats. 0 - 1
Kerkstraat (r. de l'église) Gentbrugge, v. posthoorn-
straat, Ledeberg, naar de gemeente Gentbrugge. 0 - 1

Figure 1: Excerpt from the 1880 city directory listing streets with the name "Kerkstraat". Dutch toponym (red), French toponym (blue), description of the geographical location and mereology (green) are used to differentiate between streets.
Source: Vanderhaeghen (1880), *Dubbele Wegwyzer Der Stad Gent En Der Provincie Oost-Vlaenderen*. Gent. Provided by Ghent University Library (License: non-commercial use only).

This study argues that the methodology set-out by Garbacz et al. to disambiguate historical settlements is useful for the identification of historical streets, squares and bridges provided that

[1] These numbers indicate which police precinct and judicial district had jurisdiction over the mentioned location. Note that this information is not available for every sample year and every type of place.

there are adjustments to the identity criteria. While the proper name, type and geographical footprint can be (in)directly derived from directory data, the part-whole relationship is less applicable for the disambiguation of streets, squares and bridges. This is because streets, squares and bridges - as spatial concepts - are arguably not part of one another, while this can be the case with settlements. The talk will feature case studies that exemplify this issue. Nevertheless, in a majority of the cases the methodology provides an adequate theoretical framework to determine place identity. Additionally, the city directories provided unique challenges that need to be considered during the automatisation of the identification process. For example, determining the geographical location of a place based on a textual description requires the identification of the historical sites mentioned therein. As a result, automatic matching of city directories' indexes is less feasible due to the need for prior interpretation of the place description by a domain expert (i.e. historian, geographer). This research therefore proposes a similarity-based, semi-automatic reconciliation method to reduce the time needed to link historical place observations.

The resulting geographical thesaurus or gazetteer can be used as a starting point for research in historical toponymy and urban morphology of Ghent (Figure 2). Due to the use of explicit identity criteria for streets, squares and bridges the gazetteer can function as a knowledge organization system for cultural heritage institutions (cfr. Shaw, 2016). In this particular function, the database is being used in the Ghent Mapped project to enrich digital heritage metadata.[2]

Figure 2: Geometries of contemporary streets in Ghent are displayed (blue) on the georeferenced Ferraris map (1777) using data from the historical gazetteer. Selecting a geometry provides additional information on the historical Dutch and French toponyms linked to the contemporary street. Georeferenced map provided by Agentschap Digital Vlaanderen. Image made with GeoSPARQL and the Linked Open Data platform Druid.

[2] https://gentgemapt.be/ghent-mapped/, accessed 06/02/2023.

Bibliography

Berenbaum, D., Deighan, D., Marlow, T., Lee, A., Frickel, S., & Howison, M. (2017). Mining Spatio-temporal Data on Industrialization from Historical Registries. *Journal of Environmental Informatics, 34*(1), 28-34. https://doi.org/10.3808/jei.201700381

Bol, P. (2011) "What do humanists want? What do humanists need? What might humanists get?." in Dear, M., Ketchum, J. Luria, S. & Richardson, D. (Eds.). Geohumanities: Art, history, text at the edge of place. London: Routledge.

Danniau, F., Ducatteeuw, V., Lamsens, F., & Verbruggen, C. (2022). *Ghent Mapped: Linking heritage to places and histories of Ghent*. [paper presentation]. Spatial Humanities 2022, Ghent, Belgium. https://doi.org/10.5281/ZENODO.7023290

Delbaere, R. (1962). Bronnen voor de sociale geschiedenis van Gent in de XIXde eeuw (1794-1914). *Handelingen der Maatschappij voor Geschiedenis en Oudheidkunde te Gent*, *16*(1). https://doi.org/10.21825/hmgog.v16i1.80

di Lenardo, I., Barman, R., Descombes, A., & Kaplan, F. (2019). *Repopulating Paris: Massive extraction of 4 Million addresses from city directories between 1839 and 1922.* [Data set]. DataverseNL. https://doi.org/10.34894/MNF5VQ

Ducatteeuw, V. (2021). Developing an Urban Gazetteer: A Semantic Web Database for Humanities Data. *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities*, 36–39. https://doi.org/10.1145/3486187.3490204

Ducatteeuw, V., Lamsens, F., Van Leynseele, S. (2022). *Building a historical gazetteer: Extracting place information from city directories*. [poster presentation]. Spatial Humanities 2022, Ghent, Belgium. https://doi.org/10.5281/ZENODO.6684047

Garbacz, P., Ławrynowicz, A., & Szady, B. (2018). Identity Criteria for Localities. *Formal Ontology in Information Systems*, 47–54. https://doi.org/10.3233/978-1-61499-910-2-47

Garbacz, P., Szady, B., & Ławrynowicz, A. (2021). Identity of historical localities in information systems. *Applied Ontology*, *16*, 55–86. https://doi.org/10.3233/AO-200235

Hill, L.L. (2000). Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In: Borbinha, J., Baker, T. (eds) *Research and Advanced Technology for Digital Libraries*. ECDL 2000. Lecture Notes in Computer Science, vol 1923. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45268-0_26

Kappner, K., & Albers, T. (2022). *Perks and Pitfalls of City Directories as a Micro-Geographic Data Source* (SSRN Scholarly Paper No. 4075520). https://doi.org/10.2139/ssrn.4075520

Shaw, G. (1984). Directories as sources in urban history: A review of British and Canadian material. *Urban History Yearbook*, *[11]*, 36–44.

Shaw, G., & Coles, T. (2019). *A Guide to European Town Directories: Germany, Austria, Switzerland and Scandinavia* (1st ed.). Routledge. https://doi.org/10.4324/9780429460319

Shaw, R. (2016). Gazetteers enriched: A conceptual basis for linking gazetteers with other kinds of information. In M.L. Berman, R. Mostern and H. Southall (Eds.), *Placing Names:*

Enriching and Integrating gazetteers (pp. 51–64). Bloomington, IN: Indiana University Press. https://doi.org/10.2307/j.ctt2005zq7

Southall, H, Monster, R., & Berman, M. (2011) On Historical Gazetteers. *International Journal Of Humanities And Arts Computing* 5, 2 (2011), 127-145. https://doi.org/10.3366/ijhac.2011.0028

Zedlitz, J., & Luttenberger, N. (2014). A survey on modelling historical administrative information on the semantic web. *International Journal on Advances in Internet Technology*, 7(3), 218–231.