Information extraction from Amsterdam Address Books

Leon van Wissen, Thunnis van Oort, Ivan Kisjes, Melvin Wevers (<u>DH2020 abstract</u>) University of Amsterdam - <u>createlab@uva.nl</u> <u>https://amsterdamtimemachine.nl/</u>





UNIVERSITY OF AMSTERDAM Faculty of Humanities



AMSTERDAM
TIME
MACHINE
MACHINE

translation: Liquor sellers and Cafe holders

Material Amsterdam Address books 1821-1940 (96 volumes)

Amsterdam City Archives

https://archief.amsterdam/inventarissen/details/30274/

Example from https://archief.amsterdam/archief/30274/52

1133

Slijters en Tappers.

Kooreman. (J.) Kattenburgergracht 9. Kooij. (A.) N. Achtergracht 16. Kooijman. (W.) Hoogte Kadijk 33. Koppen. (P. S.) Laagte v/d. Kadijk 2. Koppen. (H. van) Prinsengracht 341. Kortbeck. (J.) Singel 410. Korthals. (M. T. B.) de Ruijterkade 2a-2b. Korver. (P.) Haarlemmerplein 21. Krabbenbosch. (J.) Haarlemmerdijk 125. Kramer. (C. C.) Weesperstraat 88. Kreetz. (J.) O. Z. Voorburgwal 14. Kremer. (F. A.) Nes 128. Krieger. (C. H.) St. Nicolaasstraat 14. Kröne. (N. J.) Gr. Kaltenburgerstraat 18. Kron. (P. H.) Rozenstraat 27. Krot. (H.) Warmoesstraat 83. Kruiming, (A.) P. H. Kade 154. Kruisbrink. (H.) N. Z. Voorburgwal 175. Kuipers. (A.) Gr. Kattenburgerstraat 16. Kulker. (H. F.) Marnixstraat 46. Kuijper. (H. J.) Rozengracht 160. L'an. (C. van der) v. Beuningenstraat 4. Lackwijk. (J. van) P. H. Kade 28. Lader Jr. (F. H. C.) Brouwersgracht 137. Lam. (W. het) Govert Flinckstraat 261. Lamers. (A.) Buiksloterwey 45. Lange. (J. H. M.) Jac. v. Campenstr. 125. Langkemper. (J. F.) Jonge Roelensteeg 4. Lanzaat. (Wed.) Haarlemmerdijk 30. Lassing. (J. G.) Lijnbaansgracht 62a. Laugeman. (H. B.) Oude Looiersstr. 104. Legerman. (L. S.) Waterlooplein 60. Leek. (H. J. M. van) Rozengracht 84. Leenders. (Wed. W.) Prinsengracht 435. Leeuw. (W. de) Baren/splein 27. Leiting. (C. A.) Plantage Middenlaan 44. Lent. (Mej. M. van) Korte Keizerstraat 3. Leugers. (G. J.) van der Helststraal 31, Leuring. (Wed. J. A.) Blankenstraat 14. Libert. (H. T.) Leidschestraat 93. Linck. (J. G.) N. Leliestraat 196. Lobatto, (S.) Weesperstraat 59.

Slijters en Tappers.

Mazen. (G. v. d.) St. Janstraat 52. Meergart. (H.) St. Anthonisbreestraat 102. Meet len. (P. T. van) Browwersgracht 139. Meirink. (J. J.) Weteringstraat 28. Melchers. (J. T.) Quellijnstraat 127. Mellenbergh. (Ph. J.) Vondelkade 35. Mellink. (B.) Bloemstraat 116. Menks. (J. G.) N. Achtergracht 144. Meulen. (J. van der) Marnixstraat 263. Meurs. (F. C. van) Wittenburgergr. 27. Meij. (A. G. de) 1e Spaarndammerstraat 79. Meijer. (C.) Leeghvoaterstraat 8. Meijer. (G.) Blankenstraat 75. Meijer. (J. A. H.) Weesperzijde 30. Meijer. (J. P.) Ruysdaelkade 87. Meijer. (J. T.) N. Z. Voorburgwal 293. Meijer (M. S.) Nieuwendijk 81. Meijer. (T. W.) Stroomarkt 2. Meijer. (W. R.) 1e Spaarndammerstraat 69. Meyvis, (A.) Amstel 316. Meijzel. (P. J.) Rokin 110. Michel. (H.) Marnixstraat 82. Michels. (A. B.) Wagenaarstraat 70. Mier. (A. H.) Utrechtschestraat 117. Möhlmann. (J. F. C.) Oude Looiersstr. 37. Mol. (D. J.) Amstel 202. Moolen. (J. J. van der) Amstel 54. Monwerik. (W. van) Marnixstraat 173. Mulder. (F.) Prinsengracht 531. Mulder. (J. C.) Gelderschekade 123. Muller. (Mej. C. C.) Bloemstraat 155. Muijderman. (K. F.) Brouwersgracht 14. Nagel. (J. R.) Lindengracht 160. Nehr Jr. (J. B.) Warmoesstraat 24. Nie. (J. van) Prinsengracht 288. Nieberg. (Mej. W.) Lindenstraat 43. Nieland Jr. (P. A.) Kerkstraat 419. Nipper. (B. F. G.) Ferdinand Bolstraat 99. Nolte. (E.) Elandsstraat 195. Nooij. (A. J.) Nes 18. Nord. (J. S.) Zeedijk 122. Nort. (J.) Jodenbreestraat 79.



Method

- OCR (<u>Tesseract</u>)
- Comparing consecutive issues of periodicals to correct OCR-errors.
 - An address book is published every year
 - Information is repeated inside an issue (e.g. in an index)
- Identifying **person names**, **occupations**, and **location names** (Entity extraction)
 - Deduplicating person name references (including occupations) to distinguish **unique persons**
 - Identifying unique building units (**residences** and **businesses**) by street names and house numbers
- Reconciling against (geospatial) **thesauri**, such as <u>https://adamlink.nl/</u> (LOD, in RDF)



Result & embedding in the Amsterdam Time Machine

- Knowing where people:
 - Lived
 - Worked (incl. occupations)
 - Went out (Theatres, Cinemas, other Cultural Industries)
- Knowing where businesses:
 - Operated
 - Are clustering
- Connected to civil registries, trade registers:
 - Demographics
 - Social stratification

Pilot study: Known (pink dot) and newly found (blue dot) cafes in Amsterdam 1880-1940. (<u>Source</u>)

