

1. Introduction

- The Nederlab project aims to bring together **all digitized texts relevant to Dutch national heritage** (c. 800 -present) consisting of terabytes of data in one **user-friendly and tool-enriched web interface**, allowing scholars to simultaneously **search and analyze textual data** in a **virtual research environment**.
- It has a **strong end user focus** aimed at supporting **data selection tasks** and offering multiple data perspectives using **advanced search features**. It is envisaged that from this, the user can be further supported in more **advanced data analysis tasks**. We make this possible using **visual analytics**.
- The visual analytics are also used for data selection in a **search engine**. The visuals need to be **rendered fast**, so we **pre-process and index the data for retrieval**. We show some of the analytics used.

2. List and Treemaps

Wit over zwart Beelden van Afrika en zwarten in de westerse populaire cultuur
non-fictie (non-fictie/sociologie, non-fictie/kolonien-reizen)
Publicatie: 1990
Auteur: Jan Nederveen Pieterse (1946-)

... aan te leggen van dit materiaal, zoals de Golliwig in Engeland en **Zwarte Piet** in Nederland, als een ...
 ... , net zomin als **racisme** in de westerse wereld tot het verleden behoort. Bovendien, het verleden is niet ...
 ... Europa heeft geen vergelijkbare ervaringen als de Verenigde Staten. Ze heeft nooit een zo grote **zwarte** ...
 ... Europees **racisme**: het waren immers de slaven van Europa die in West-Indië en Amerika, in Europese koloniën ...
 ... en op plantages aan het werk werden gezet. Het **racisme** dat hierbij te pas kwam is van oorsprong geen ...

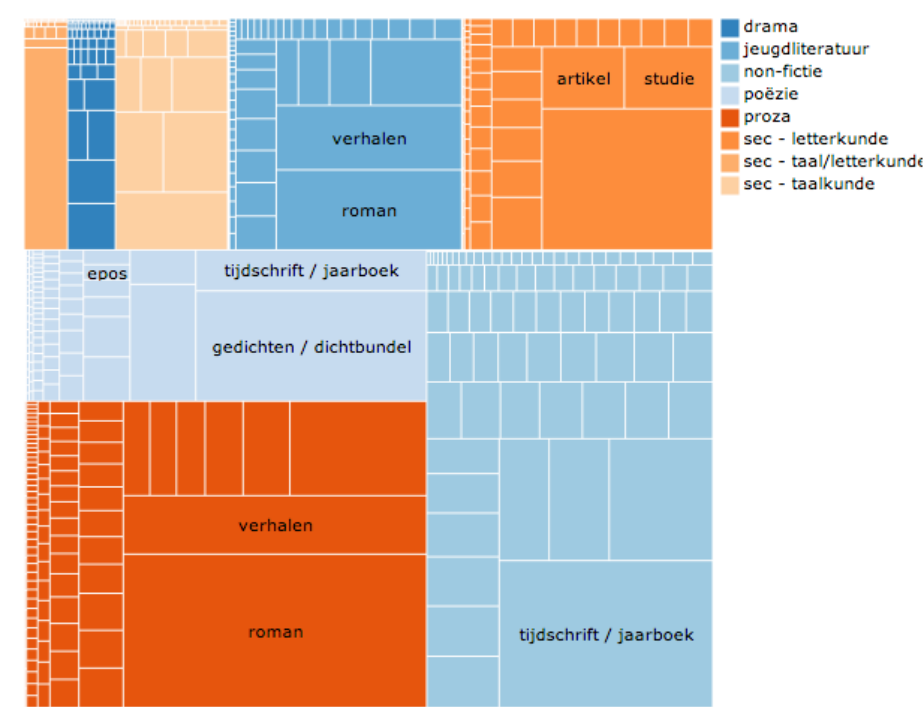
Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

De Rasoei-Komrij affaire. Een geval van elite-racisme
sec - letterkunde (non-fictie/pamflet-brochure, studie)
Publicatie: 2003
Auteur: T.A. van Dijk (Naaldwijk, 1943-)

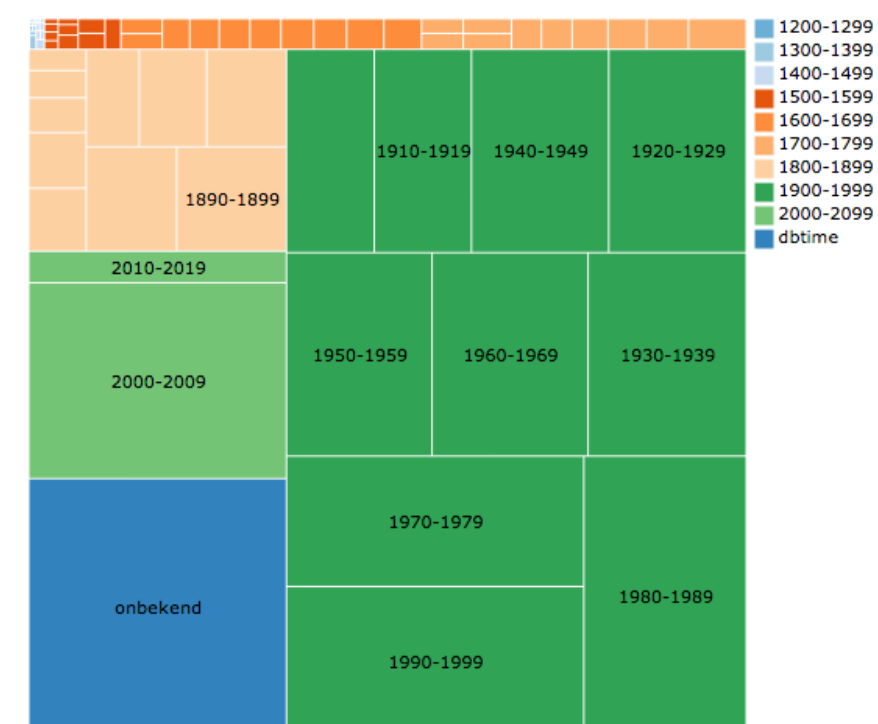
... van het 'elite-**racisme**' dat ik uitvoerig en jarenlang in mijn wetenschappelijk werk had bestudeerd ...
 ... elitair **racisme** in het algemeen, en in Nederland in het bijzonder. Juist het gigantische succes van de ...
 ... anti-islamitische ideeën van Fortuyn c.s., en de stelselmatige ontkenningen van **racisme** in dit verband ...
 ... buitenland steeds werd gehouden. Zowel door mijn eerdere werk over elite-**racisme** als door mijn ervaringen met ...
 ... teksten en **racisme**. In dat opzicht is het ook een populair-wetenschappelijk boek over elitair **racisme** en ...

Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

Figure 1: A few items from the result list with snippets with keyword highlighting.



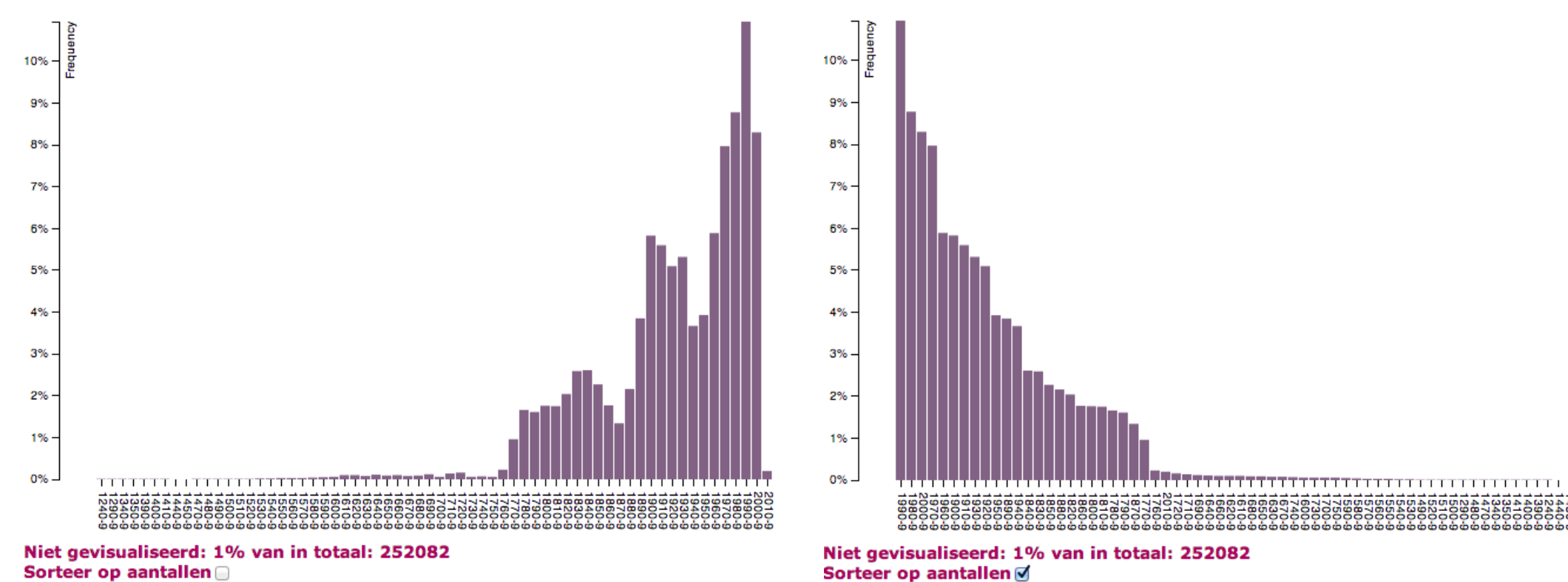
(a) Distribution of sub-genres by genres.



(b) Distribution of resources by time.

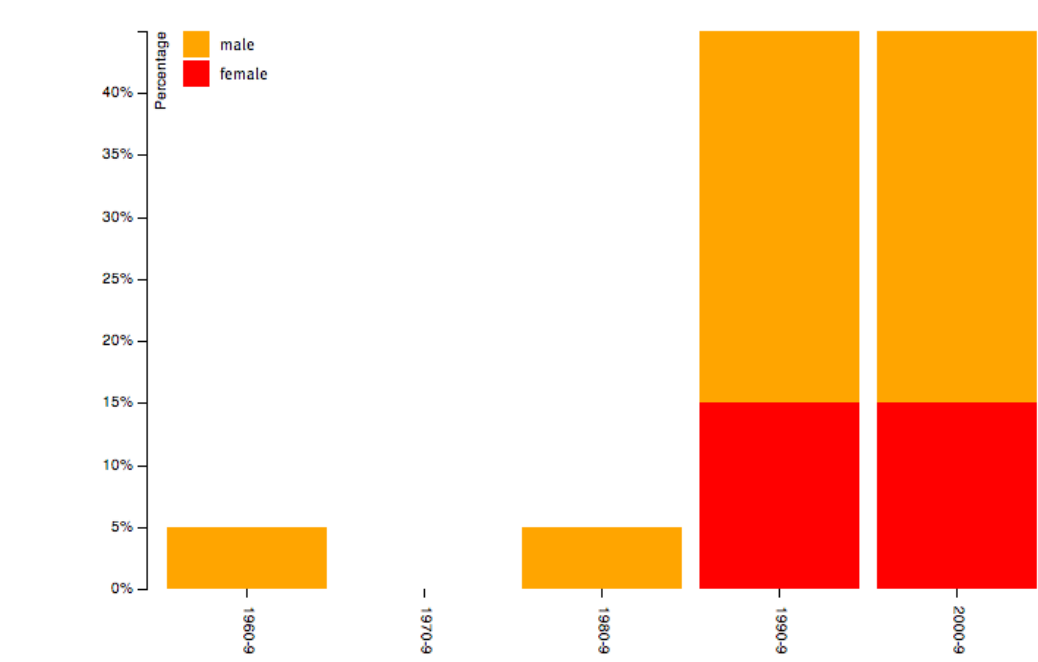
Figure 2: Use of treemaps for displaying the distributions of texts by genres and subgenres, and the time of publication. These are used as an alternative representation of the list of results.

3. Bar Charts

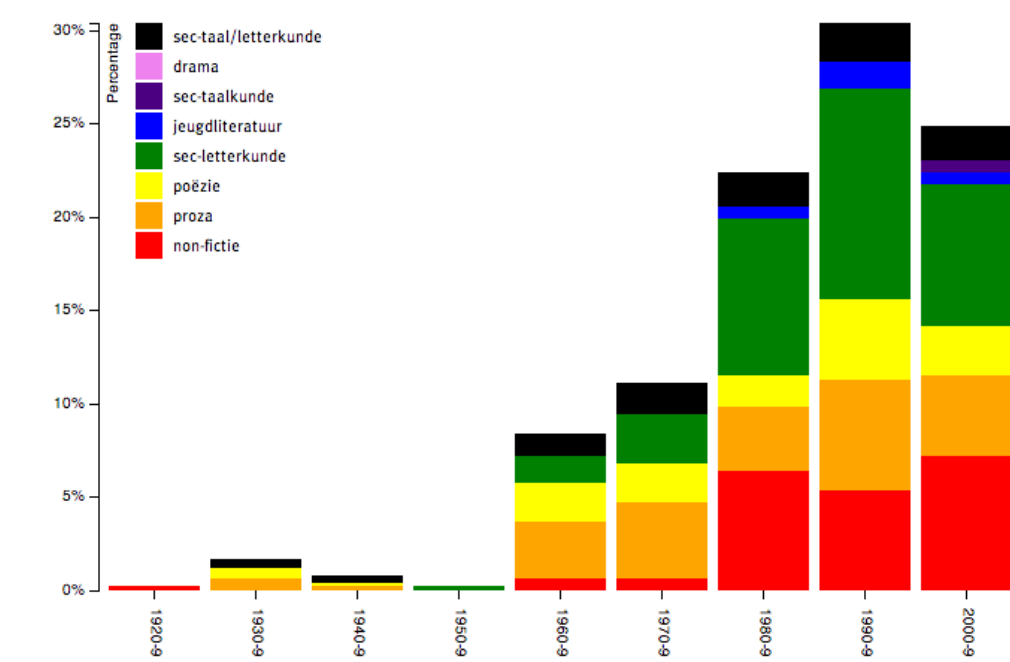


(a) Distribution of number of results by time of publication, sorted by time.

(b) Distribution of number of results by time of publication, re-sorted by count.



(c) Distribution of number of results by gender of author.



(d) Distribution of results by genres.

Figure 3: Use of bar charts as an alternative representation of the list of results.

4. Comparative Analytics

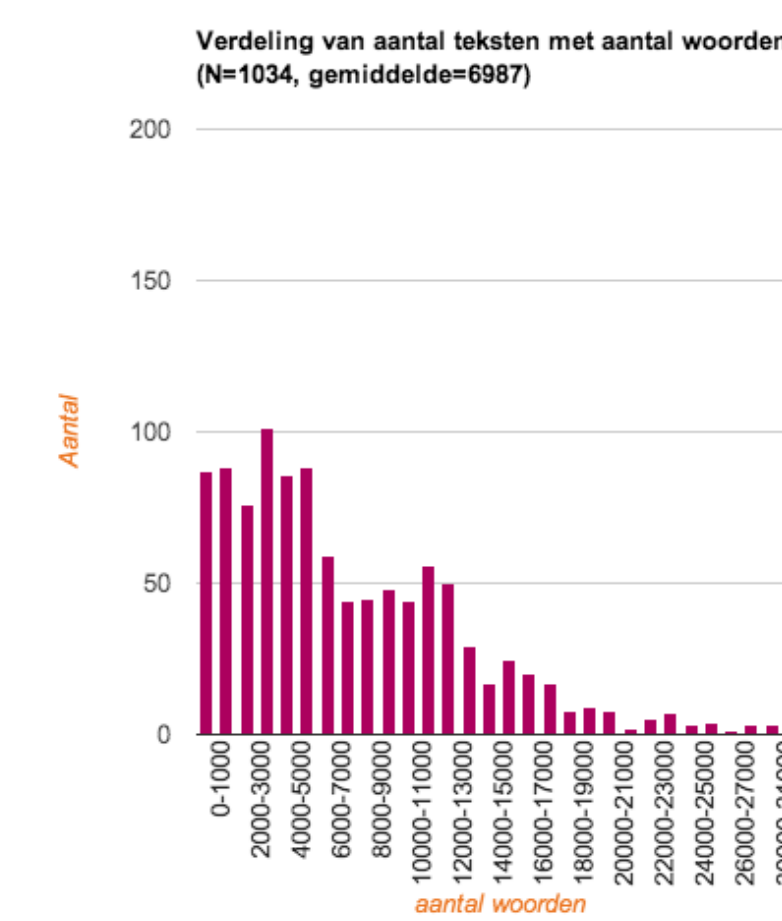
- After search and selection, the selected datasets are stored. We make **pair-wise comparison of datasets** possible for analysis. In this example, we compare the published texts of two journals in the same period for a meaningful comparison.
- Multiple dimensions of the **descriptive metadata** can be used for a comparison, such as the number of words and pages, the names and gender of authors, their place of birth, etc. **Many more metadata dimensions can be used for comparison!**

Vergelijk twee corpora

1. De Gids, 1885-1894

Totaal: 1034 documenten (berekend in 3.314 seconden)

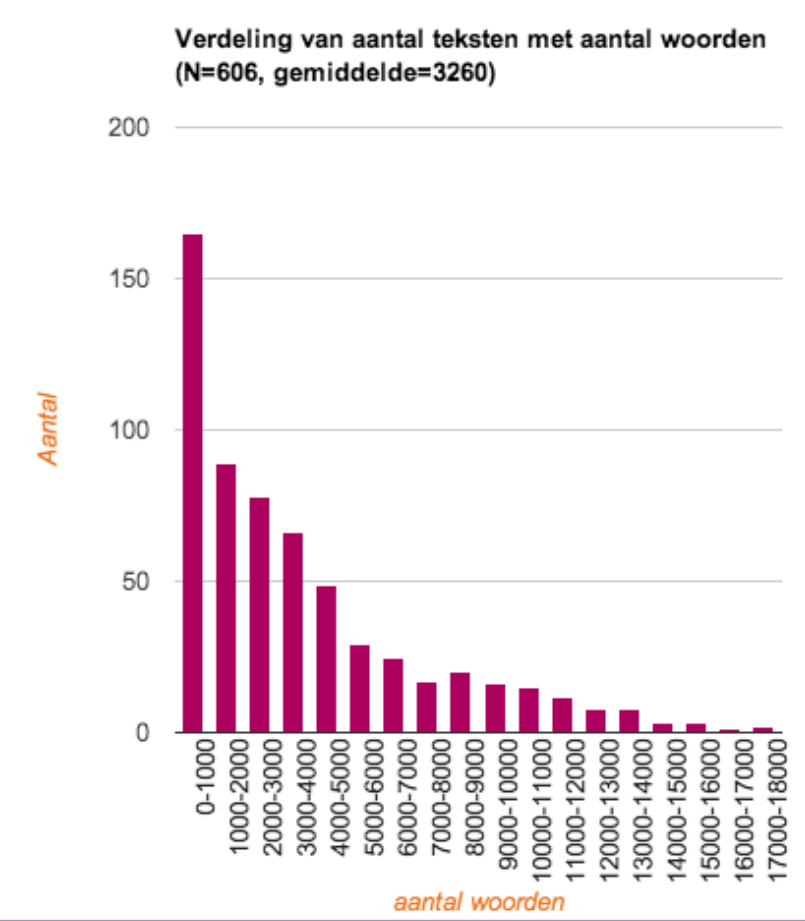
aantal woorden



2. De Nieuwe Gids, 1885-1894

Totaal: 606 documenten (berekend in 0.326 seconden)

aantal woorden



(a) Distribution of number of words in a text.

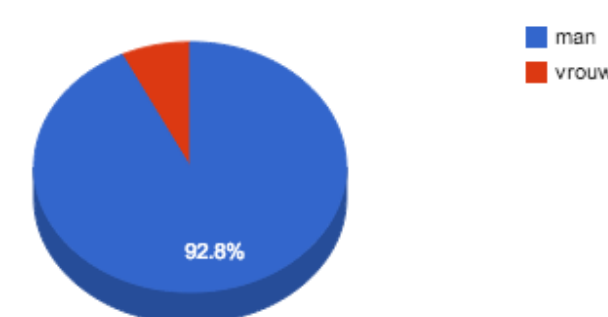
achternaam van auteur (N=1123, uniek=217)

"" (319) Hall (63) Swarth (29) Byvanck (26) Hamel (25) Pijzel (25) Hooyer (18) Quack (17) Mont (16) Viugt (16) Polak (14) Sillem (14) Buys (13) Hubrecht (13) Viotta (13) Muller (12) Beaufort (11) Boddart (11) Fruin (11) Rooses (11) Boissevain (10) Kielstra (10) Pierson (10) Spruyt (10) Couperus (9) Kalf (9) Sachse (8) Brill (7) Lith (7) Berckenhoff (6) Berg (6) Cort van der Linden (6) Dyserinck (6) Hofstijk (6) Holwerda (6) Jaeger (6) Veth (6) Carelsen (5) Goede (5) Muller frz. (5) Buysse (4) Chantepie de la Saussaye (4) Hoop Scheffer (4) Loman (4) Naber (4) Nieuvelt (4) Plaisius (4) Seipgens (4) Simons (4) Swart Abrahamsz (4) Beets (3) Boekennoogen (3) Borel (3) Busken Hunt (3) Deventer (3) Doodes (3) Duyl (3) Emants (3) Kluijver (3) Lapidoth (3) Loenen Martinet (3) Manen (3) Nierop (3) Oort (3) Sluithouwer (3) Snouck Hurgronje (3) Terburgh (3) Uhlenbeck (3) Wallis (3) Warren (3) Werumeus Buning (3) Wijck (3) Wilken (3) Wolters (3) Amtenius (2) Berg van Dussen Mulkerik (2) Boer (2) Bosch (2) Bredius (2) Brink (2) Fockema Andrae (2) Gallée (2) Goltstein (2) Groenewegen (2) Heckeren (2) Heemskerk Azn. (2) Heim (2) Honigh (2) Huizinga (2) Ising (2) Jonissen (2) Kate (2) Kips (2) Kok (2) Kollwijn (2) Koster (2) Koster Gzn. (2) Land (2) Logeman (2) Loveling (2) Meyboom (2) Moleschott (2) Moltzer (2) Mulder (2) Netscher (2) Nolthenius (2) Nouhuys (2) Ovinck (2) Penning (2) Rijckevorsel (2) Rogge (2) Schimmel (2) Sijmons (2) Snijder van Wissenkerke (2) Speyer (2) Valetton (2) Vissering (2) Vries (2) Zilcken (2) Adriaensen (1) Alberdingk Thijm (1) Andriessen (1) Asser (1) Aulinis de Bourouill (1) Bas (1) Beekman (1) Bemmelen (1) Beynen (1) Bilderdijk (1) Binger (1) Blok (1) Blüte (1) Boele van Hensbroek (1) Bohl (1) Bosdoo-Tousaint (1) Bothmans Brouwer (1) Brandt (1) Brunsma (1) Bussy (1) Coenen (1) Cosijn (1) Donders (1) Dorhout Mees (1) Drabbe (1) Engelmann (1) Enklaar (1) Eymael (1) Fabius (1) Farncombe Sanders (1) Feith (1) Franken (1) Frantzen (1) Geer van Juthaas (1) Geyter (1) Giltée (1) Gleichman (1) Gobée (1) Greven (1) Groot (1) Gunning Jr. (1) Harthoorn (1) Hartog (1) Hasebroek (1) Hecker (1) Heering (1) Heijermans (1) Helderling (1) Hertog (1) Imminck (1) Jong (1) Jura (1) Kampen (1) Kan (1) Kern (1) Knappert (1) Knuttel (1) Knuttel-Fabius (1) Krämer (1) Leopold (1) Leyssohn Norman (1) Lindenberg (1) Mallinson (1) Mannen (1) Matthes (1) Mees Azn. (1) Melj (1) Meijer (1) Meisma (1) Meurs (1) Müller (1) Niemeijer (1) Nijland (1) Oudemans (1) Pit (1) Priem (1) Pullé (1) Ridder metter swane (1) Riemsdijk (1) Serrurier (1) Seyffardt (1) Singels (1) Six (1) Spronck (1) Steinmetz (1) Stokvis (1) Straub (1) Telling (1) Tiele (1) Toorenbergen (1) Tours (1) Trenard (1) Treub (1) Vliet (1) Was (1) Winkler Prins (1) Zimmerman (1)

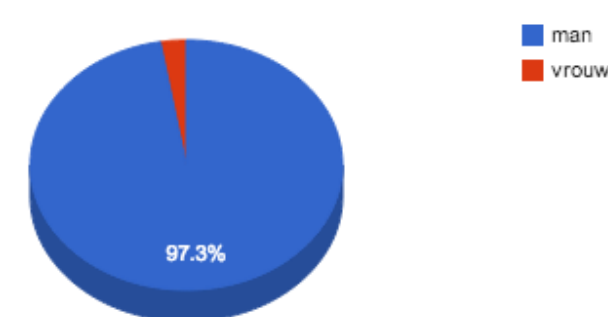
achternaam van auteur (N=722, uniek=68)

"" (98) Eeden (50) Kloos (48) Goes (46) Veth (40) Tideman (36) Deyssel (32) Boeken (31) Tak (30) Looy (29) Verwey (29) Hofker (27) Deventer (23) Erens (20) Aletrino (15) Prins (14) Lotsij (12) Gorter (11) Jelgersma (11) Stoffel (11) Witsen (9) Bolland (8) Swarth (7) Broedelet (6) Canter (6) Diepenbroek (6) Netscher (5) Wilbau (5) Groeninger (4) Perk (4) Biederlack (3) Domela Nieuwenhuis (3) Roossdorp (3) Buysse (2) Leopold (2) Muller (2) Raap (2) Roland Holst (2) Barris (1) Bauer (1) Blink (1) Brunsma (1) Busken Hunt (1) Clercq (1) Doorenbos (1) Fiore della Neve (1) Gerhard (1) Gerritsen (1) Graaf (1) Houten (1) Huysmans (1) Keller (1) Loffelt (1) Mijnsen (1) Nawijn (1) Pierson (1) Plasschaert (1) Poelman (1) Renterghem (1) Roland Holst-van der Schaik (1) Santen Koiff (1) Smit Kleine (1) Treub (1) Uldriks (1) Unger (1) Valk (1) Vloten (1) Winkler Prins (1)

bijdragen mannen / vrouwen (N=792)



bijdragen mannen / vrouwen (N=561)

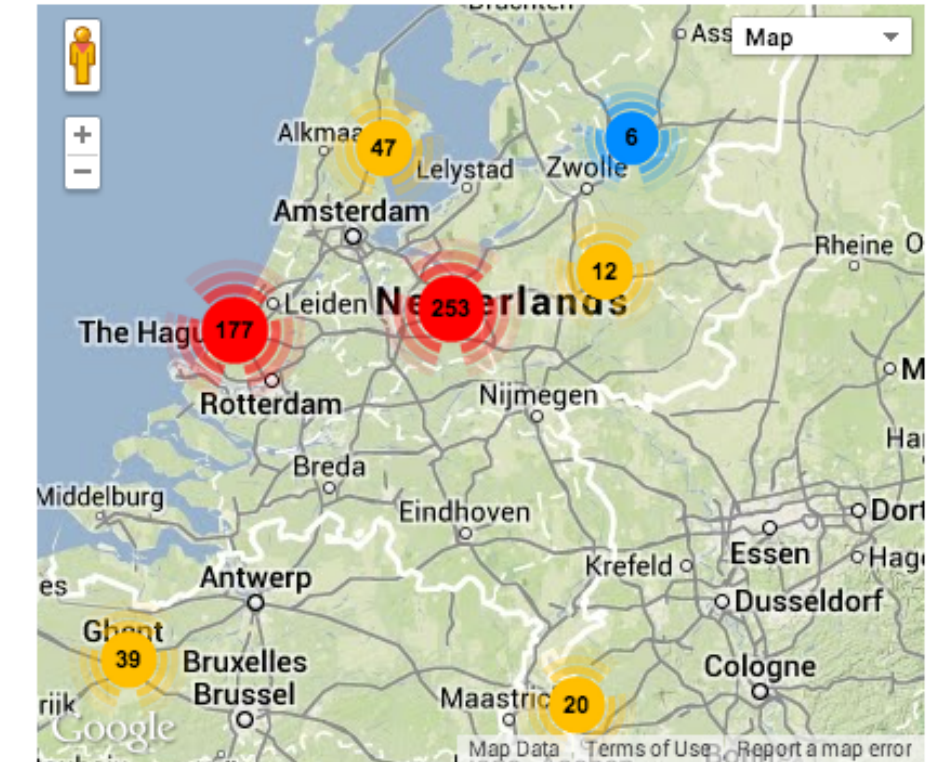


(b) 1) Word clouds with the last names of authors, and in orange, the names that appear in both datasets. 2) Pie charts to show the distribution of male vs. female authors.

geboorteplaats van auteur



geboorteplaats van auteur



(c) A map and clusters of markers to visualize the place of birth of authors.

Figure 4: Using visual analytics for pairwise comparison of datasets using metadata.

5. Acknowledgements

This work is supported by the Netherlands Organisation for Scientific Research (NWO), Royal Netherlands Academy of Arts and Sciences (KNAW), and Common Language Resources and Technology Infrastructure (CLARIN).