
Koppeling van historische gegevens op naam

Gerrit Bloothoof

Inleiding

Nu er steeds meer historische gegevensbestanden worden gedigitaliseerd, lijkt het ook gemakkelijker te worden om in die bronnen bijvoorbeeld informatie over personen te vinden. Historici en genealogen zijn daarin zeer geïnteresseerd. Het vinden van die informatie blijkt in de praktijk helaas vaak tegen te vallen, omdat de spelling van namen in oude bronnen verre van consistent is. Laten we dat aan de hand van een voorbeeld bekijken. Van de naam van de Antwerpse koopman *Delbeecque* vinden we in bronnen uit het vroeg 17^e-eeuwse Amsterdam de vormen: *del*, *de la*, of *de le* in combinatie met *Beeck*, *Beecq*, *Beecque*, *Beke*, *Beque*, *Berque*, of aaneengeschreven *Delbeecque*, *Delbeque*, *Delbeke*. Dat deze namen gebruikt werden voor dezelfde persoon, kunnen we – ook in het algemeen – uit andere informatie afleiden, bijvoorbeeld uit de – bijzondere – voornaam, de combinatie van voornaam en patroniem, de naam van echtgenote, de handelspartner, het beroep, de herkomstplaats, etc. Een onderzoeker die over al deze informatie beschikt, zal het wel lukken om met de nodige moeite alle vermeldingen van de koopman te vinden. Als het echter om grootschalig prosopografisch onderzoek gaat waarbij informatie van veel personen uit diverse bronnen per persoon gegroepeerd moet worden, is dat handmatig al gauw niet meer doenlijk. Daarvoor zouden we de computer willen inschakelen. De computer moet dan eigenlijk de werkwijze van de onderzoeker kopiëren, en deze razendsnel uitvoeren. Het probleem is echter dat deze werkwijze zeer complex kan zijn. De onderzoeker brengt een enorme hoeveelheid kennis in, van namen, geschiedenis, logica, van intuïties over combinatorische waarschijnlijkheden, en niet te vergeten een algemene kennis van de (historische) wereld, die buitengewoon moeilijk in de computer over te brengen zijn. Alhoewel het probleem van gegevenskoppeling op naam in essentie een complex probleem is uit de kunstmatige intelligentie, kunnen we over de naamkundige inbreng daarin wel de nodige gedachten ontwik-

kelen. We zullen een aantal verschillende naamkundige aspecten de revue laten passeren, en aangeven welke kennis er al is om het probleem te lijf te gaan, maar ook welke kennis we nog ontberen.

Bronnen van variatie in een naam

Een persoonsnaam kan om vele redenen in verschillende vormen tot ons komen. Het begint al bij de opgave van de naam door de persoon zelf. Afhankelijk van het doel van de registratie kan de betreffende persoon uit verschillende naamvormen kiezen. Wanneer het om een kerkelijke gebeurtenis gaat, kan een doopnaam gebruikt worden die anders is dan de roepnaam uit het informele verkeer. En in bijvoorbeeld zakelijke transacties kunnen ook weer andere namen worden gekozen, wellicht met de vermelding van het beroep in plaats van de achternaam, maar misschien is het beroep inmiddels wel de erfelijke familienaam. Afhankelijk van de context kan alleen voor een patroniem of juist een achternaam gekozen worden, waarbij het lang niet altijd duidelijk is of een enkele patronymische vorm misschien wel al de familienaam is. Daarnaast kan ook sociale wenselijkheid van de naam een rol spelen: een Latijnse vorm maakt voor een arts of wetenschapper meer indruk dan de Nederlandse.

In de meeste gevallen schreef de persoon de naam niet zelf, maar sprak hem uit terwijl een klerk, notaris, een pastoor of dominee de naam op schrift stelde. De naamdrager kon lang niet altijd controleren hoe dat gebeurde: ofwel omdat men zelf niet lezen kon of omdat het schrift in de foliant aan het oog onttrokken werd. Door luisterfouten, interpretatiefouten en door eigen spellingconventies van de schrijver werd de nodige variatie geïntroduceerd. Er kunnen daarnaast schrijffouten zijn gemaakt, of er werd slordig geschreven, wat weer tot leesproblemen leidt.

Zo lag een naam eeuwenlang vast in de boeken, tot die overgenomen wordt in een elektronisch gegevensbestand. En dan kan er weer het nodige verkeerd gaan. Er kunnen leesfouten worden gemaakt, bijvoorbeeld door verwarring van gelijkende letters, of door het invoegen of weglaten van letters door onjuiste interpretaties van het schrift. Daarnaast treden al eerder genoemde ambiguïteiten op die zonder nadere kennis moeilijk direct op te lossen zijn, zoals in *Dirk Albertsen*: is *Albertsen* een patroniem of de familienaam? En in *Jan Jansen Coeckebakker*, is de achternaam daarin *Coeckebakker* en *Jansen* een patroniem, of is *Coeckebakker* alleen de beroeps-

aanduiding? Wanneer in een gegevensbestand direct bij invoer hierover een beslissing moet worden genomen omdat de beschikbare gegevensvelden dat vereisen, kunnen fouten worden gemaakt. En ten slotte zijn typefouten onvermijdelijk.

Zo zijn er tal van bronnen van variatie in de naam van dezelfde persoon, vaak overigens met een eigen karakter. En laten we niet denken dat deze problematiek alleen bestaat voor historische bronnen, want een belangrijk deel van het bovenstaande is vandaag nog steeds aan de orde.

Invalshoeken

Concentreren we ons op de schrijfwijzen van een naam, en willen we proberen de variatie daarin te verkleinen, dan is een tweetal oplossingen denkbaar. De eerste is om voor elke naam een standaardschrijfwijze voor te stellen. Het doet er niet veel toe welke variant als standaard wordt gekozen, zolang die maar uniek is. Het zou de grondvorm van de naam kunnen zijn, maar dat hoeft niet. We kiezen bijvoorbeeld de meest voorkomende vorm *Dirk* als standaard voor *Dirck*, *Dierick*, maar ook *Diederik* etc. Het voordeel van deze aanpak is dat we vrij snel veel varianten onder dak hebben. Wat betreft voornamen zouden we dat bijvoorbeeld kunnen doen op basis van het Prisma voornamenboek van Van der Schaar, met ca. 20.000 namen. Maar echt succesvol is dat niet meteen. In de poorterboeken van Amsterdam (1531-1611) staan ca. 24.000 personen die 1725 verschillende voornamen hebben. 38% hiervan wordt genoemd in Van der Schaar. In de doop-, trouwen en begraafboeken van Amsterdam (1776-1811) staan ca. 343.000 personen, met 11.826 verschillende voornamen, waarvan er slechts 19% in het voornamenboek staan. En om dicht bij huis te blijven: tussen 1983 en 2003 werden in Nederland 4,2 miljoen kinderen geboren die 177.000 verschillende voornamen kregen (Bloothoofd et al, 2004), waarvan slechts 3% te vinden is in het standaardwerk. Daarnaast zijn er veel naamvormen waarbij het niet eenduidig is welke standaard gekozen moet worden. Moet *Theodorus* ook de standaard *Dirk* krijgen of is het zelf een standaardvorm met een verkortingsvariant zoals *Theo*? Is *Berque* een variant van *Berk* of is het, zoals boven, een verschrijving van *Beek*? Een tussenoplossing is om per naam niet één standaard toe te wijzen, maar eventueel meerdere. Als we dit zouden kunnen uitbreiden met een zekere waarschijnlijkheid van het behoren tot de ene of de andere standaard, dan zijn we al dicht in de buurt van de

tweede aanpak, namelijk om helemaal geen naamstandaardisering toe te passen, maar om aan elk paar namen een kans toe te kennen dat de namen door *dezelfde persoon in een bepaalde tijd* gedragen zijn. Het lastige is om die kans te bepalen. Daarvoor is veel informatie nodig over de varianten van de naam van dezelfde persoon. Toch is deze aanpak intuïtief sterk. Ook een historicus zal zich niet bij voorbaat vastleggen op vaste interpretaties van losse informatie, maar gaandeweg aan de hand van alle beschikbare gegevens een argumentatie opbouwen voor bepaalde keuzen. De meest waarschijnlijke argumentatie zal dan de doorslag geven (Bloothoof, 1994a).

Tussen standaardisatie en statistiek liggen een aantal mogelijkheden die zinvol benut kunnen worden om variatie in namen te reduceren zonder veel verlies van informatie. Dit zijn spellingvereenvoudiging en suffix-afsplitsing.

Spellingvereenvoudiging

Een belangrijk deel van spellingvariatie in namen ontstaat door een verschillende interpretatie van de manier waarop de gesproken vorm neergeschreven moet worden. Daarom is er een goede kans op minder variatie wanneer de naam in fonetische vorm wordt weergegeven. Daarbij is het van belang om niet alle fonetische details te willen weergeven. Het verschil tussen stemloze medeklinkers zoals /f/ en /s/ en hun stemhebbende tegenhangers /v/ en /z/ bijvoorbeeld, kunnen we uit het oogpunt van vereenvoudiging beter achterwege laten. Hieronder staan een serie fonetisch geïnspireerde vereenvoudigingen die leiden tot een flinke beperking in vormvariatie:

1. Alle diakritische tekens kunnen het beste worden genegeerd. Het gebruik ervan is vaak niet systematisch.
2. Een j zonder voorafgaande i wordt in het midden of aan de het einde van een naam herschreven als een i of wordt weggelaten: *Ravenswaaj*, *Rajmund*, *Benesj*.
3. De naaminitiële I en Y worden, behalve als ze gevolgd worden door een e, als J geschreven.
4. Combinaties van twee of meer klinkers worden door één symbool vervangen. Dit is fonetisch lang niet altijd correct, maar het blijkt wel effectief te zijn om variatie in de schrijfwijze van die klinkerparen op

- te vangen. Bijeen komen combinatie zoals {i,ie,y,ei,ey,ij,eij}, {ea,ae}, {ou,au,aa}, {oa,ao}, {eu}, {oe}, {ai,ay,aij}, {ui,uij,uy}, {oi,oy,oij}.
5. De ph-combinatie wordt in namen meestal probleemloos een f, behalve op de grens van samenstellingen, zoals *Schiphol*.
 6. In de woordinitiële Tj wordt ook wel eens een s ingevoegd, *Tsjetske*. Daarom wordt die s steeds weggelaten.
 7. De stemhebbende z en v worden als stemloze s en f herschreven.
 8. De q wordt als k of kw herschreven.
 9. De x wordt s (begin naam: *Xander* -> *Sander*), ks, of weggelaten in Franse uitgangen (*Margaux*).
 10. De c is een gecompliceerde letter. De fonetische realisatie is echter steeds doorslaggevend. Het leidt tot een serie regels die uitmonden in een s of een k, of in de combinatie ch als g, sj (*Charles*) of weglating (*Aschwin*).
 11. De h kan buiten het begin van de naam of in samenstellingen, veelal weggelaten worden.
 12. De d wordt volgens de fonetische realisatie aan het einde van een naam een t.
 13. Weer fonetisch geïnspireerd: alle dubbele letters worden tot een enkele gereduceerd. Dit brengt bijvoorbeeld de vormen *Abraham*, *Abraaham*, *Abraham*, *Aabraham* bijeen.

Er wordt geen poging gedaan om fonetische verfijningen, zoals de uitspraak van de letter e als korte /ɛ/, lange /e/, of sjwa, te beschrijven. Dit is in foutloos modern Nederlands al lastig, maar in namen is de context zo variabel dat er geen betrouwbare omzetting kan worden gerealiseerd: er wordt meer nieuwe variatie gecreëerd dan er wordt opgelost. In het algemeen kan er beter naar overgeneralisatie worden gestreefd. Dat leidt ertoe dat meer namen dezelfde vereenvoudigde spellingsvorm krijgen dan een 'ideale' vereenvoudiging, en dat sommige significante verschillen verloren gaan. In heel veel toepassingen is het echter helemaal niet erg wanneer een gebruiker wat meer alternatieven voorgeschoteld krijgt. Dan kan zelf een keuze worden gemaakt.

De naamvorm die resulteert zal normaliter als extra veld aan een bestand worden toegevoegd en niet zichtbaar zijn voor de gebruiker. Maar namen die dezelfde vorm hebben kunnen er door de computer wel mee bijeen gebracht worden, zoals *Dierk*, *Dirk*, *Dijrk*, *Dirck*, *Dierc*, *Dirc*, *Dierck*,

Dijrck, Dirrk, Diercq, Dircq. De winst die met deze spellingvereenvoudiging wordt behaald is ongeveer 30% minder naamvormen.

Decompositie

We kunnen niet zeker weten of met de namen *Fem, Femme, Femmegien, Femmegiena, Femmeje, Femmeke, Femmeken, Femmelien, Femmelina, Femmete, Femmetgen, Femmetje, Femmetjen, Femmetie, Femmetien, Femmetiena, Femje, Femjo, Femke, Femkje, Fempje, Femtje, Femtien, Femmy, Femmia, Femmije, Femmigje, Femmigjen, Femmigien, Femmina, Femmiete, Femmitgen, Femmitje, Femmitie, Femmitien* dezelfde persoon wordt aangeduid of dat de verschillen onderscheidend zijn. Willen we in staat zijn om ze onder één noemer te brengen, dan zal het (gestapelde) suffix moeten worden gescheiden van de stam. Het is niet eenvoudig om daar een complete regelverzameling voor te ontwikkelen. Om een indruk te krijgen van het probleem is de stam-suffix scheiding handmatig uitgevoerd op een bestand van 40.413 verschillende voornamen (20.000 namen uit het Prisma voornamenboek plus ruim 20.000 uit diverse historische bestanden, in oorspronkelijke spelling). Dat leidde tot de grote aantallen van 541 en 1337 suffixen voor respectievelijk mannen- en vrouwenamen. Vervolgens is onderzocht of het mogelijk is om het handmatige resultaat ook automatisch te bereiken door een computer daarvoor een regelset te laten ontwikkelen. De regels hadden allemaal de vorm van een vereist naameinde waarin het stam-suffix scheidingspunt kon worden aangebracht. Bijvoorbeeld een regel *tse# -> t-se#* zegt dat in alle namen die eindigen op *tse, se* het suffix is, zoals *Aetse -> Aet-se*. Een regel die de meeste context nodig heeft, dat wil zeggen: het meeste van de stam moet weten, gaat vóór een regel die minder eisen stelt. Het meest extreem zijn namen die in hun geheel nodig zijn om een besluit te nemen over het suffix. Als voorbeeld onderzoeken we namen die allemaal eindigen op *-bina*. De meeste namen zijn met de regel *bina# -> b-ina#* te beschrijven. Vier namen, zoals *Fabin-a*, zijn een uitzondering en krijgen een eigen regel *abina# -> abin-a#*. Deze regel kent op zichzelf weer uitzonderingen met namen die eindigen op *abina* waarin het suffix toch weer *-ina* is, zoals *Gab-ina*. Daar blijkt een regel voor nodig waar de volledige naam in staat: *Gabina# -> Gab-ina#*. Op eenzelfde manier zijn er nog 10 regels nodig om uitzonderingen te beschrijven. In totaal zijn er 15 regels nodig om 46 namen die eindigen op *-bina* van een juiste stam-suffix

scheiding te voorzien. Daarvan zijn er echter 12 die op een enkele naam betrekking hebben. Die uitzonderingen leiden tot een enorm aantal regels, 2516 regels voor de mannennamen en 5496 regels voor de vrouwennamen. Er waren in totaal echter 320 krachtige regels die ieder meer dan 10 verschillende voornamen voor hun rekening namen. Wanneer we de stam-suffix regels toepassen op een nieuwe verzameling namen, dan moeten we hopen dat de krachtiger regels het meeste werk zullen doen, en dat er niet al te veel nieuwe uitzonderingen opdoemen. Zo'n aanpak kan ons werk uit handen nemen, door de resultaten van zo'n algoritme als een eerste voorstel te zien. Controle achteraf is wel altijd nodig. En ook hier geldt weer, dat namen bijeenbrengen op grond van gelijke stam zeker een overgeneralisatie is. Maar het laat de gebruiker wel de keus om dan weer namen te schrappen.

Wat over voornamen is gezegd geldt op eenzelfde manier voor **patroniemen**. Patroniemen van Jan zijn bijvoorbeeld *Jans.*, *Jans*, *Janse*, *Jansen*, *Janson*, *Jansone*, *Janseune*, *Jansoon*, *Janz.*, *Janz*, *Janze*, *Janzen*, *Janzon*, *Janzoon*, *Janss.*, *Janss*, *Jansz.*, *Jansz*, *Janzs*, *Jansse*, *Jansze*, *Janssen.*, *Janssen*, *Janszn*, *Janzn.*, *Janszen*, *Janszenne*, *Janzzen*, *Jansson*, *Janszon*, *Janssoon*, *Janszoon*, *Janssone*, *Jansens*, *Janssz*, *Janssens*, *Janssenz*, *Jani* en voor vrouwen *Jandr*, *Jansd*, *Jansdr*, *Jansdr.*, *Janszdr* en *Jansdochter*. We constateerden in de historische bestanden 220 verschillende patronienvormen voor mannen en 46 voor vrouwen, met in de top uiteraard *-s*, *-sz*, *-sen*, *-se*, *-en*, *-z*, *-es*, *-ns*, *-e*, *-ze* en *-sdr*, *-dr*, *-endr*, *-dr.*, *-sendr*, *-d*, *-sdr.*, *-xdr*, *-szdr*, *--zdr*. Ook hier zijn weer veel regels nodig om automatisch tot afsplitsing van deze patronymische suffixen te komen. Patronymische uitgangen zoals *-ing(a)*, *-ink*, *-ma(n)* werden hierbij niet betrokken. Binnen de context van koppeling van gegevens vanaf de 16^e eeuw, is het twijfelachtig of deze vormen voor dezelfde persoon uitwisselbaar gebruikt werden met de eerdergenoemde *-zoon* vormen; in de meeste gevallen was voor de *-ing* en *-ma* uitgangen de status van familienaam waarschijnlijk al bereikt.

Bij **familienamen** geven voorvoegsels extra problemen. In het Nederlands Repertorium van Familienamen zijn 135 verschillende typen te vinden, met als top-20: *van*, {*van de*, *van der*, *van den*}, *de*, *den*, *ten*, *ter*, *van 't*, *te*, *'t*, *in 't*, *le*, *du*, *la*, *von*, *d'*, *de la*, *op den*, *van het*, *op 't*, *op de*. De moeilijkheid is dat deze voorvoegsels ook wel geschreven werden aan de kern van de naam, terwijl bij dat aaneenschrijven ook samentrekkingen kunnen optreden die niet altijd eenvoudig te herkennen zijn, zoals de 17^e-eeuwse varianten (voor dezelfde persoon) *Lermite*, *Lheremite* en *L'Hermite*, of *D'Admiraal* en

Damarael, De Fonseca en Daffonseka. Andere, eenvoudiger op te lossen voorbeelden zijn *Verheijst en Van Heijst, Le Bruijn en De Bruijn, Van der Burch en Verburgh, Da Costa en Dacosta, De Pas en Depas, Du Pre en Dupre, Le Maire en Lemaire, De Nijs en Denys, Van der Does en Verdoes, Van der Mersch en Vermersch*. In veel gevallen gaat het hier om niet-Nederlandse namen of om het - nog steeds bestaande - verschil tussen de noordelijke en zuidelijke Nederlanden in al dan niet aaneenschrijven van onderdelen van familienamen. Deze samentrekkingen zijn echter relatief zo laagfrequent dat het ontwikkelen van een automatische procedure niet de hoogste prioriteit heeft.

Context

Waar we hiervoor de afzonderlijke naamelementen onder de loep namen, zullen deze in historische bronnen vaak gezamenlijk optreden. Daarmee komt gegevenskoppeling op naam een stuk dichterbij, vooral als er zeldzame naamonderdelen en/of extra-onomastische gegevens beschikbaar zijn die de identificatie vergemakkelijken. Naast voornaam, patroniem en achternaam zijn vaak vermeldingen van beroep en plaats van herkomst behulpzaam, terwijl ook de datum van de vermelding kan helpen. In het voorbeeld van de familienaam *L'Hermite* was de voornaam *Jacques* een aanwijzing want die is in Amsterdamse bronnen zeldzaam. Als *Jacques* niet ook afwisselend geschreven wordt als *Jacob* – wat niet het geval was –, dan kunnen de spellingseisen aan de achternaam veel losser worden gemaakt. Aan de andere kant zijn er toch ook weer nieuwe interpretatieproblemen wanneer meerdere gegevens deel uitmaken van de beschrijving van een persoon. In het onderstaande voorbeeld volgen we de interpretatie van degenen die gegevens uit zes verschillende bronnen invoerden in een bestand:

Voornaam	Patroniem	Achternaam
Jan	Jacopsen	Benning
Jan, de jongen		Benningh
Jan, den jonge	Benningh	

Jan	Benning	Coeckebakker
Jan	Jacobsz. Benninck	Coeckebakker (koopman)
Jan	Jacobsz	Banning Couckebakker (koopman)

Het is duidelijk dat hier geworsteld is met de interpretaties van patroniem en een dubbele achternaam waarin een (vermeend) beroep optreedt. Alleen in de laatste twee bronnen stond alle potentiële informatie er expliciet bij, maar zelfs dat leidde tot interpretatieverschillen. Met dit soort “fouten” moet bij de analyse rekening worden gehouden en dat leidt tot complexe vergelijkingen. Zeker wanneer ook een herkomstplaats gegeven wordt die verward kan worden met een achternaam.

In het algemeen geldt wel dat hoe meer informatie er in één bronvermelding beschikbaar is, hoe sterker koppelingen met andere vermeldingen kunnen zijn. Het gaat dan meestal om de vermelding van meerdere personen met aanduiding van familierelaties. Als in bronnen de man *Hendrik Veuger* en als zijn vrouw *Cornelia Aaije*, *Cornelia Aaje*, *Kornelia Haye*, en *Cornelia Haaijen* wordt genoemd, dan is er weinig twijfel over dat het om dezelfde vrouw gaat. En wanneer in huwelijksakten de wederzijdse ouders worden genoemd is er al helemaal een stevige kern van familierelaties. Bij zulke rijke genealogische bronnen – waartoe in ieder geval de burgerlijke stand behoort – is het mogelijk een expertsysteem te ontwikkelen dat bij voldoende informatie automatisch een genealogie kan afleiden, waarbij onzekerheden en spellingvariaties in namen gecompenseerd worden door kennis van familierelaties (Bloothoof, 1994b, 1995).

Ontwikkelingen

In het interdisciplinaire probleemveld van koppeling van historische gegevens op naam is de naamkunde onontbeerlijk. Ze kan – in samenwerking met historici en genealogen – bijdragen met kennis over naam-

ontwikkeling en kennis over contemporaine naamvarianten. Daarnaast is de taalwetenschap van belang voor onderzoek naar de invloed van uitspraak op de spelling van namen, en naar methoden voor morfologische decompositie. Omdat daarvoor regels met een statistische basis kansrijk zijn, kan met vrucht gebruik gemaakt worden van methoden uit de computationele taal-kunde. Wanneer het om meerdere typen informatie gaat, en wanneer er relaties zijn tussen beschreven personen, dan wordt het van belang om al die kennis op een zinvolle manier in het hele koppelingsproces te integreren. Daarbij moet het vakgebied van de kunstmatige intelligentie ingeschakeld worden. Een voorwaarde voor succes is ten slotte een intelligente en efficiënte opslag van informatie, en het ontwikkelen van inzichtelijke zoekstrategieën voor gebruikers. Die willen op een eenvoudige manier zo snel mogelijk een compleet resultaat bij hun speurtocht naar informatie rond een persoon. Dat vereist kennis van gebruikersinterfaces en databasetechnologie (Bloothoofd, 1998).

Naast samenwerking tussen disciplines is ook de ontwikkeling van naamkundige basisbestanden noodzakelijk. Dat betreft verrijkte bestanden van voornamen en patroniemen, achternamen, en van beroepen en topniemen waarvan bekend is dat ze als achternaam voorkomen. De verrijking ligt vooral op het vlak van varianten van naamelementen waarvan bewezen is dat ze voor dezelfde persoon zijn gebruikt, met daarnaast plaats en jaar van optreden. Ook is het, ondanks de eerder genoemde beperkingen, nuttig om per naam tot een standaardvorm te komen, en daar op Nederlands/Belgisch niveau afspraken over te maken. Veel instanties ontwikkelen deze standaarden nu zelf (met alle verschillen van dien) en het dient de efficiëntie om de handen ineen te slaan. Deze basisbestanden zijn dan tevens een uitstekend startpunt voor het trainen van statistische technieken om niet eerder ontdekte naamvarianten te duiden.

Het belang van deze ontwikkeling is tweeledig. Ten eerste zitten we midden in een periode waarin een grote hoeveelheid historisch materiaal elektronisch beschikbaar komt, en waarbij computationele hulpmiddelen onontbeerlijk zijn om vele duizenden gebruikers te helpen dit materiaal via het Internet te exploreren. Ten tweede is de problematiek niet alleen relevant voor historische informatie. Moderne databases kennen vrijwel identieke toegankelijkheids- en koppelingsproblemen, zij het dat die vaker een internationaal namengebied zullen betreffen, met alle bijkomende problemen van dien.

Referenties

- Bloothoof, G. (1994a). 'Corpus-Based Name Standardization', *History and Computing*, **6.3**, 153-167.
- Bloothoof, G. (1994b). 'Automatische familieconstructie', *Cahiers VGI (Vereniging voor Geschiedenis en Informatica)*, **7**, 9-54.
- Bloothoof, G. (1995). 'Multi-Source Family Reconstruction', *History and Computing*, **7.3**, 90-103.
- Bloothoof, G. (1998). 'Assessment of Systems of Nominal Retrieval and Historical Record Linkage', *Computing and the Humanities*, **32**, 39-56.
- Bloothoof, G., van Nifterick, E. en Gerritzen, D. (2004). *Over voornamen*. Het Spectrum, Utrecht.
- Nederlands Repertorium van Familienamen*, onder redactie van P.J. Meertens, H. Buitenhuis en E.N. Palmboom. I-XIV. Van Gorcum, Assen / De Walburg Pers, Zutphen (1963-1988).
- Van der Schaar, J. (2002). *Prisma Voornamen*. Het Spectrum, Utrecht.

Veel voorbeelden in dit artikel zijn gebaseerd op materiaal dat verzameld is voor het proefschrift: Gelderblom, O. (2000). *Zuid-Nederlandse kooplieden en de opkomst van de Amsterdamse stapelmarkt (1578-1630)*. Verloren, Hilversum.