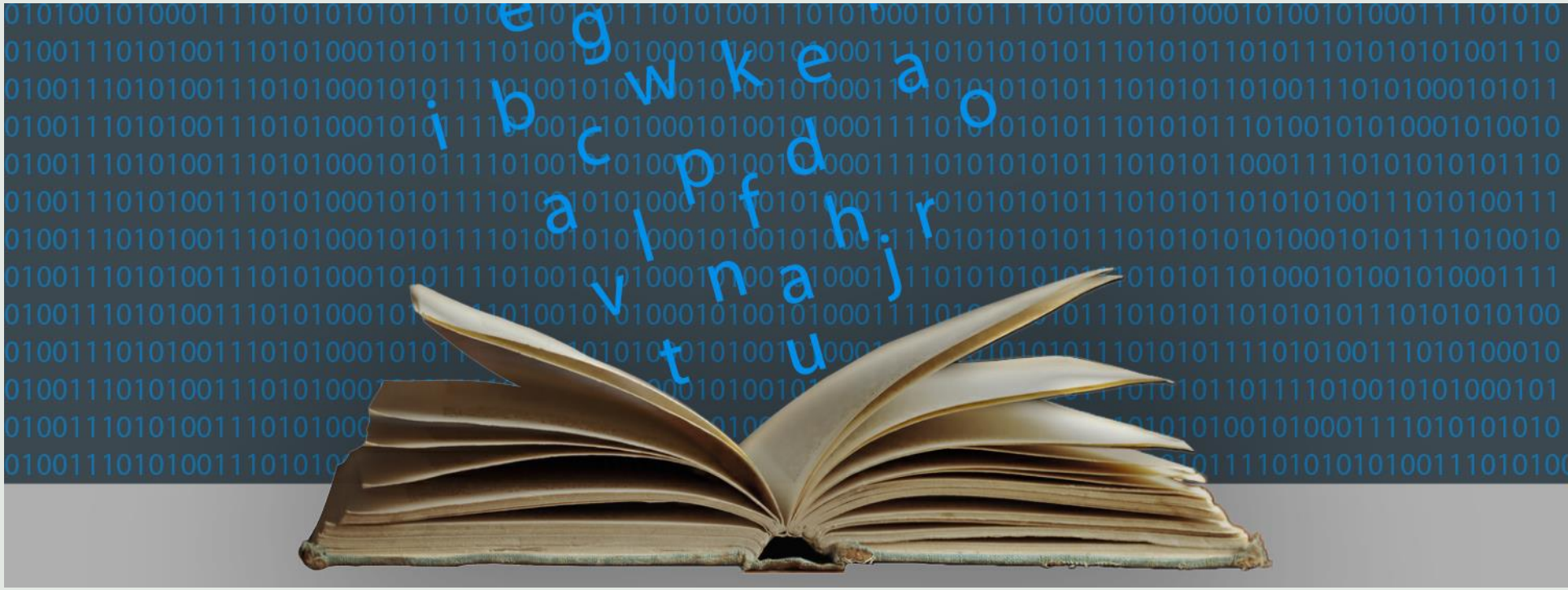
OCR qva!ity m@tters in Digital Humanities rese@rch

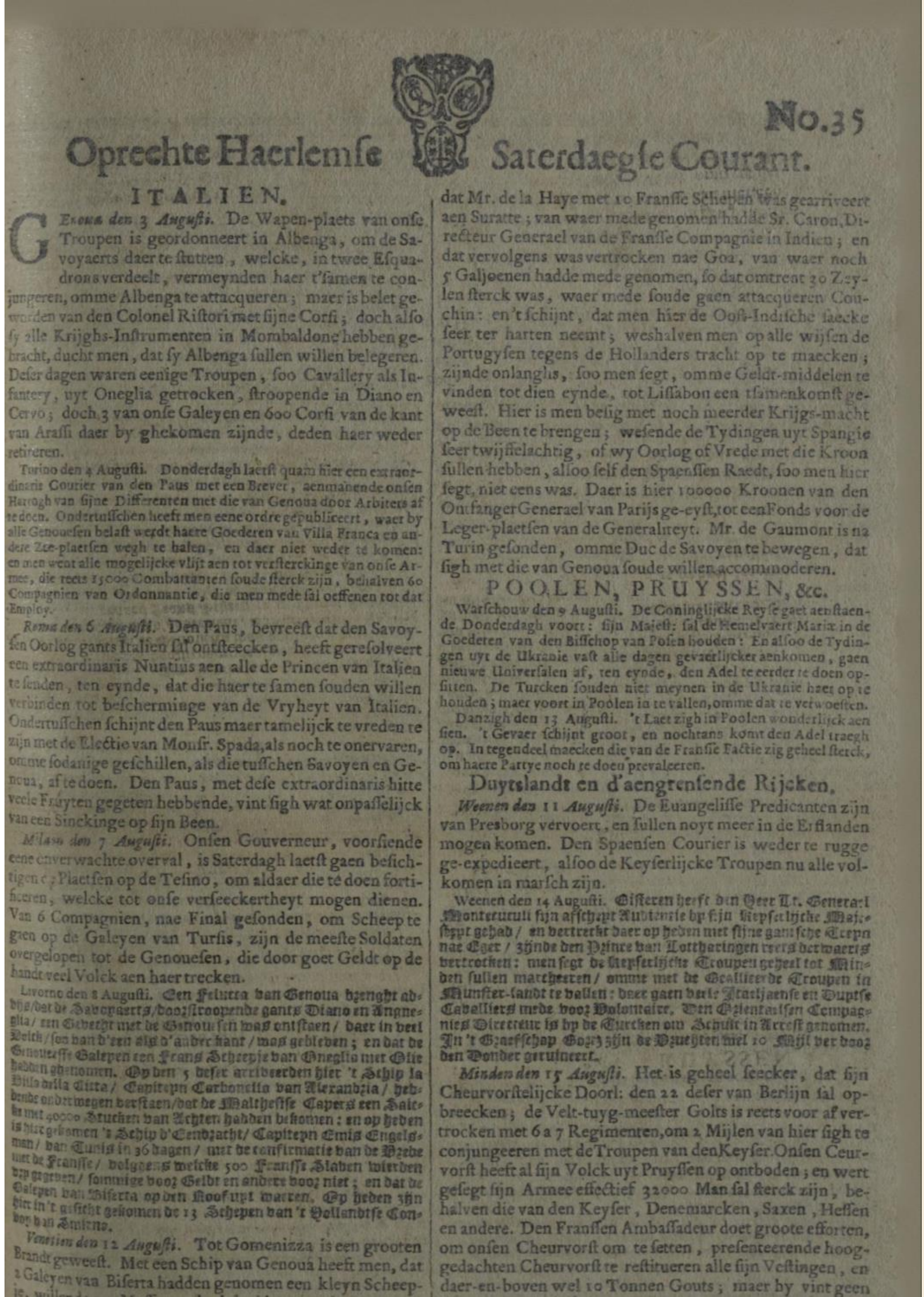


Mirjam Cuper, 1 maart 2021

KB national library of the netherlands

Research

- Poland
- 1672
- Oprechtse Haerlemse Courant



Oprechte Haerlemse Saterdaegse Courant. No. 35

ITALIEN.

Genoua den 3 Augusti. De Wapen-plaets van onse Troupen is geordonneert in Albenga, om de Savoyaerts daer te stutten, welcke, in twee Esquadrons verdeelt, vermeynden haer t'samen te conjungeren, omme Albenga te attacqueren; maer is belet geworden van den Colonel Ristori met sijne Corsi; doch also sy alle Krijghs-Instrumenten in Mombaldone hebben gebracht, ducht men, dat sy Albenga sullen willen belegeren. Deser dagen waren eenige Troupen, soo Cavallery als Infantery, uyt Oneglia getrocken, stroopende in Diano en Cervo; doch 3 van onse Galeyen en 600 Corsi van de kant van Arassi daer by ghekomen zijnde, deden haer weder retireren.

Turino den 4 Augusti. Donderdagh laetst quam hier een extraordinaris Courier van den Paus met een Brevet, aenmanende onsen Hartogh van sijne Differenten met die van Genoua door Arbiters af te doen. Ondertusschen heeft men eene ordre gepubliceert, waer by alle Genouesen belast werdt haere Goederen van Villa Franca en andere Zee-plaetsen wegh te halen, en daer niet weder te komen: en men went alle mogelijcke vlijt aen tot versterckinge van onse Armee, die reets 15000 Combattanten soude sterck zijn, behalven 60 Compagnien van Ordonnantie, die men mede sal oeffenen tot dat Employ.

Roma den 6 Augusti. Den Paus, bevreest dat den Savoysen Oorlog gants Italien sal ontsteecken, heeft geresolveert een extraordinaris Nuntius aen alle de Princen van Italien te senden, ten eynde, dat die haer te samen souden willen verbinden tot bescherminge van de Vryheyt van Italien. Ondertusschen schijnt den Paus maer tamelijck te vreden te zijn met de Electie van Monsr. Spada, als noch te onervaren, omme sodanige geschillen, als die tusschen Savoyen en Genoua, af te doen. Den Paus, met dese extraordinaris hitte veele Fruyten gegeten hebbende, vint sigh wat onpasselijck van een Sinckinge op sijn Been.

Milaen den 7 Augusti. Onsen Gouverneur, voorsiende eene onverwachte overval, is Saterdagh laetst gaen besichtigen de Plaetsen op de Tesino, om aldaer die te doen fortificeren, welcke tot onse verseeckertheyt mogen dienen. Van 6 Compagnien, nae Final gesonden, om Scheep te gaen op de Galeyen van Tursis, zijn de meeste Soldaten overgelopen tot de Genouesen, die door goet Geldt op de handt veel Volck aen haer trecken.

Livorno den 8 Augusti. Een Felucca van Genoua brenght advijs/ dat de Savoyaerts/ doorstroopende gantsch Diano en Angnelia/ een Gevecht met de Genouesen was ontstaen/ daer in veel Volck/ soo van d'een als d'ander kant/ was gebleven; en dat de Genouesse Galeyen een Frans Scheepje van Oneglia met Olie hadden genomen. Op den 5 deser arriveerden hier 't Schip la Villa della Citta/ Capiteyn Carboncillo van Alexandria/ hebbende onderwegen verstaen/ dat de Maltheesse Capers een Saicke met 40000 Stucken van Achten hadden bekomen: en op heden is hier gekomen 't Schip d'Eendracht/ Capiteyn Emis Engelsman/ van Tunis in 36 dagen/ met de confirmatie van de Vrede met de Fransse/ volgens welcke 500 Fransse Slaven wierden vry gegeven/ sommige voor Geldt en andere voor niet; en dat de Galeyen van Biserta op den Roof uyt waren. Op heden zijn hier in 't gesicht gekomen de 13 Schepen van 't Hollandtse Convoy van Smirna.

Venetien den 12 Augusti. Tot Gomenizza is een grooten Brandt geweest. Met een Schip van Genoua heeft men, dat 2 Galeyen van Biserta hadden genomen een kleyn Scheepje, [illegible] dat Mr. de la Haye met 10 Fransse Schepen was gearriveert aen Suratte; van waer mede genomen hadde Sr. Caron, Directeur Generael van de Fransse Compagnie in Indien; en dat vervolgens was vertrocken nae Goa, van waer noch 5 Galjoenen hadde mede genomen, so dat omtrent 30 Zeylen sterck was, waer mede soude gaen attacqueren Couchin: en 't schijnt, dat men hier de Oost-Indische saecke seer ter harten neemt; weshalven men op alle wijsen de Portugysen tegens de Hollanders tracht op te maecken; zijnde onlanghs, soo men segt, omme Geldt-middelen te vinden tot dien eynde, tot Lissabon een tsamenkomst geweest. Hier is men besig met noch meerder Krijgs-macht op de Been te brengen; wesende de Tydingen uyt Spangie seer twijffelachtig, of wy Oorlog of Vrede met die Kroon sullen hebben, alsoo self den Spaensen Raedt, soo men hier segt, niet eens was. Daer is hier 100000 Kroonen van den Ontfanger Generael van Parijs ge-eyst, tot een Fonds voor de Leger-plaetsen van de Generaliteyt. Mr. de Gaumont is na Turin gesonden, omme Duc de Savoyen te bewegen, dat sigh met die van Genoua soude willen accommoderen.

POOLEN, PRUYSSEN, &c.

Warschouw den 9 Augusti. De Coninglijcke Reyse gaet aenstaende Donderdagh voort: sijn Majest: sal de Hemelvaert Mariæ in de Goederen van den Bisschop van Posen houden: En alsoo de Tydingen uyt de Ukranie vast alle dagen gevaerlijcker aenkomen, gaen nieuwe Universalen af, ten eynde, den Adel te eerder te doen opsitten. De Turcken souden niet meynen in de Ukranie haer op te houden; maer voort in Poolen in te vallen, omme dat te verwoesten.

Danzigh den 13 Augusti. 't Laet zigh in Poolen wonderlijck aensien. 't Gevaer schijnt groot, en nochtans komt den Adel traegh op. In tegendeel maecken die van de Fransse Factie zig geheel sterck, om haere Partye noch te doen prevaleeren.

Duytslandt en d'aengrensende Rijcken.

Weenen den 11 Augusti. De Euangelisse Predicanten zijn van Presborg vervoert, en sullen noyt meer in de Erflanden mogen komen. Den Spaensen Courier is weder te rugge ge-expedieert, alsoo de Keyserlijcke Troupen nu alle volkomen in marsch zijn.

Weenen den 14 Augusti. Gisteren heeft den Heer Lt. Generael Montecuculi sijn afscheyt Audientie by sijn Keyserlijcke Majesteyt gehad/ en vertreckt daer op heden met sijne gantsche Treyn nae Eger/ zijnde den Prince van Lotharingen reets derwaerts vertrocken: men segt de Keyserlijcke Troupen geheel tot Minden sullen marcheeren/ omme met de Geallieerde Troupen in Munster-landt te vallen: daer gaen veel Italjaense en Duytse Cavalliers mede voor Volontaire. Den Orientaelsen Compagnies Directeur is by de Turcken om Schult in Arrest genomen. In 't Graefschap Gorz zijn de Vruchten wel 10 Mijl ver door den Donder geruineert.

Minden den 15 Augusti. Het is geheel seecker, dat sijn Cheurvorstelijcke Doorl: den 22 deser van Berlijn sal opbreecken; de Velt-tuyg-meester Golts is reets voor af vertrocken met 6 a 7 Regimenten, om 2 Mijlen van hier sigh te conjungeeren met de Troupen van den Keyser. Onsen Ceurvorst heeft al sijn Volck uyt Pruyssen op ontboden; en wert gesegt sijn Armee effectief 32000 Man sal sterck zijn, behalven die van den Keyser, Denemarcken, Saxen, Hessen en andere. Den Franssen Ambassadeur doet groote efforten, om onsen Cheurvorst om te setten, presenteerende hooggedachten Cheurvorst te restitueren alle sijn Vestingen, en daer-en-boven wel 10 Tonnen Gouts; maer by vint geen

Research


J
tÃ©
â
t
n
t
t
I
btn
Ã¯
*
dÃ«
waer
nae
foude
fy
"
defe
vaa
â€
Ç
b
t
Den
Heer
fullen
dÃ©
bat
Armee

POOLEN,PRUYSSEN,&c.
â– iÃ¯AVarfÃ©J:Â«iuÂ»>,p-f-,',aij, Mr>'. 'â€¢, f ua-jctt teJ>!r&%FjÃ¯b â– 'Â» t^ptgaa aÃ©-i; â– yruy-ÃªÂ« Aijai a/bi*^;;'vi* je-Â« .]Â«%Â«)}'Â«fcWftim s,ijcii''(dat'",Ha* ata/ifia Â«tadifc PlaK.fcn..lijcf onder jijÂ» Mijcftcv-ji bip ;ftkMÂ»lÂ«Jp'Â»Â»i.^^ ; toegrvaJlwl'..f Da*!'iija defer dagi-ri bietbÃ³ek vctfrh-y-ie 'Taf'ie'Â» [â– â€¢ dÃ©Â» ColieUr.a s'tDgcfc'iitjeÃŸ j", aiWevtiÃ©im.'flÃ«worÃ©ftic !WÂ«ptfH*ei:krp''i'a,t,^ u'tenj^'Â»K*Site^^),epÃˆr;oÃ³d<Xi^ â– 'â– : <â– ,;. "Ã–i'iÃ¯'boV-l 'Â«.'â– ."â– a*Iri" '<Cbr-if frgft grnfsan â€¢jj*. ;rf*tt fflÃ¯i.b-'rf/it.aÃ¯ifie ',teÂ®i:sSÃ¯4te (jic oÂ»-Â»Â» â– 'Ã¯Â»n'iW|iteiiÂ

POOLEN, PRUYSSEN, &c.

' Wurfihotiw den 6, February.'VyÃ¯ dÃ¨ UkÃ¯faiiiÃ© kiÃ³men va-" rÃœblÃ¨Tydiagea,* dief iwÃ©fdt'gelchreveitj"dat'dÃ«nDorÃ¶fehsko ganti blbÃ¶tÃ©nalÂ£ iÃ«er k:ha<:rs';..Ã«ii de: Paddeni raÃ«eit weghViwaerorn de CÃ¶f hÃ«tri teVerlatca,"Ã©n haer nae" deÃŸ Hanenko'tÃ«begeyca ,'die defwÃ«genhydeiooÃ¶Ã¶ Man cflÃ«ftivÃ«' lÃ³udÃ© fterck"Ã¼jn C, waer vaa "inÃ«adÃ© confirmatie vÃ«nvacht.*i-:t':; '..;â–

POOLEN, PRUYSSEN, &c.

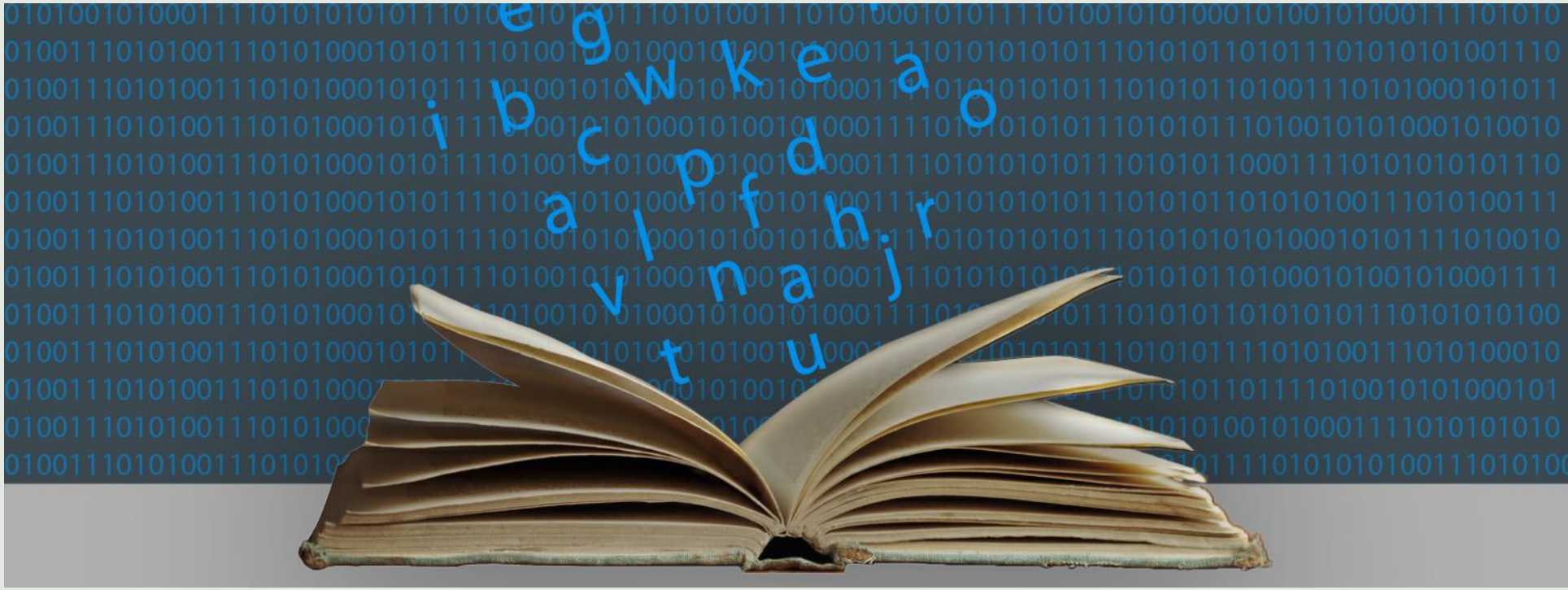
. S Warfchonm 'itol Den __t Â«cat dagen gsdutu-t heeft, Ã¼ eynA-ngg^aoldghteaÃªymds â– "V-â€¢â– ,<â– ,'..'" , ','â– ., â– '::â– â– -,'â– TT-,' ',; "r-r â– â– â– "-â– ': : â€¢'â– â– ' -\: .- .â– .'â€¢:.-gTdoopons; iaÃ¯g?HjckÂ«. heefti_oOj:fc dn, t.fatT)enkon.ft. VfanSaiq; domirfch toeOjjataw i^h^ewÃ«nfchtÃ©aeyn_bciomen..'Da TÃ¼rekft-liier btea Sm Â«ccl te weg? gciarâ€žfit<.dat."dÃ©n.. Adel;' % Ã¨cr.onaÃ¯nrghwasiâ€ž nitmar_illÃŸ.'andflr-Ã¯: rzydeo geÂ« ft* v*n-dÂ»VSdirland:BidefemiÃ¶ tÃ© ge*. li_j^fcÂ£lfcmme4ij_c&'lÃ¯ecftTviÃ¯a^ Voeten40000 l verrftsrektwerden /': Deri Heer Roi zfesky.P_iÂ£lijckcn ', en voordÃ©GoningitineÃ©ea RoiÃª mbd9llreaght,â€¢ Ã© tÃ³rcyts tot GratÃ¶aw'.aeAgebri:' zij m â– ~ Tcliein wil riiÃ«riyoÃ¶tgÃ©ldÃ³rbaÃ«rberichten., ditden cnKtiMiPoolenfilk'OJnea, en tot van Zywiec falVÃ©rtreckea," 't wÃ«kk en }â€¢ doch hier van vernsichtinÃ«n >. \y ;..-'"â– ;"â€¢ - -â– ; 7,7' 1 .-â– -

POOLEN,PRUYSSEN,&c.

â– iÃ¯AVarfÃ©J:Â«iuÂ»>,p-f-,',aij, Mr>'. 'â€¢, f ua-jctt teJ>!r&%FjÃ¯b â– 'Â» t^ptgaa aÃ©-i; â– yruy-ÃªÂ« Aijai a/bi*^;;'vi* je-Â« .]Â«%Â«)}'Â«fcWftim s,ijcii''(dat'",Ha* ata/ifia Â«tadifc PlaK.fcn..lijcf onder jijÂ» Mijcftcv-ji bip ;ftkMÂ»lÂ«Jp'Â»Â»i.^^ ; toegrvaJlwl'..f Da*!'iija defer dagi-ri bietbÃ³ek vctfrh-y-ie 'Taf'ie'Â» [â– â€¢ dÃ©Â» ColieUr.a s'tDgcfc'iitjeÃŸ j", aiWevtiÃ©im.'flÃ«worÃ©ftic !WÂ«ptfH*ei:krp''i'a,t,^ u'tenj^'Â»K*Site^^),epÃˆr;oÃ³d<Xi^ â– 'â– : <â– ,;. "Ã–i'iÃ¯'boV-l 'Â«.'â– ."â– a*Iri" '<Cbr-if frgft grnfsan â€¢jj*. ;rf*tti'jÂ£'fcuteÂ«itf'te &^^ ; ftefsÂ»' teiX-S fÃ¶ fflÃ¯i.b-'rf/it.aÃ¯ifief fyoao Jfrari'fy f-Ã©fC;WtPa*t *epa*ji ',teÂ®i:sSÃ¯4te (jidf sa^teteintn;]nuÂ«bimiJ.Â«'Ã¼cÃ³jaant,-; oÂ»-Â»Â»

â– 'Ã¯Â»n'iW|iteiiÃ¯*swj.SÂ«'iSlÂ«B'tTf

OCR qva!ity m@tters in Digital Humanities rese@rch



Mirjam Cuper, 1 maart 2021

KB national library of the netherlands

Introduction

- Mirjam Cuper
- Data scientist
- KB - National Library of the Netherlands
- Digital heritage

DIGITAL HUMANITIES

Humanities

- Human society
- Human culture

- Heritage collections

Optical Character Recognition (OCR)

Cholera in Peru uitloper van 'pandemie'

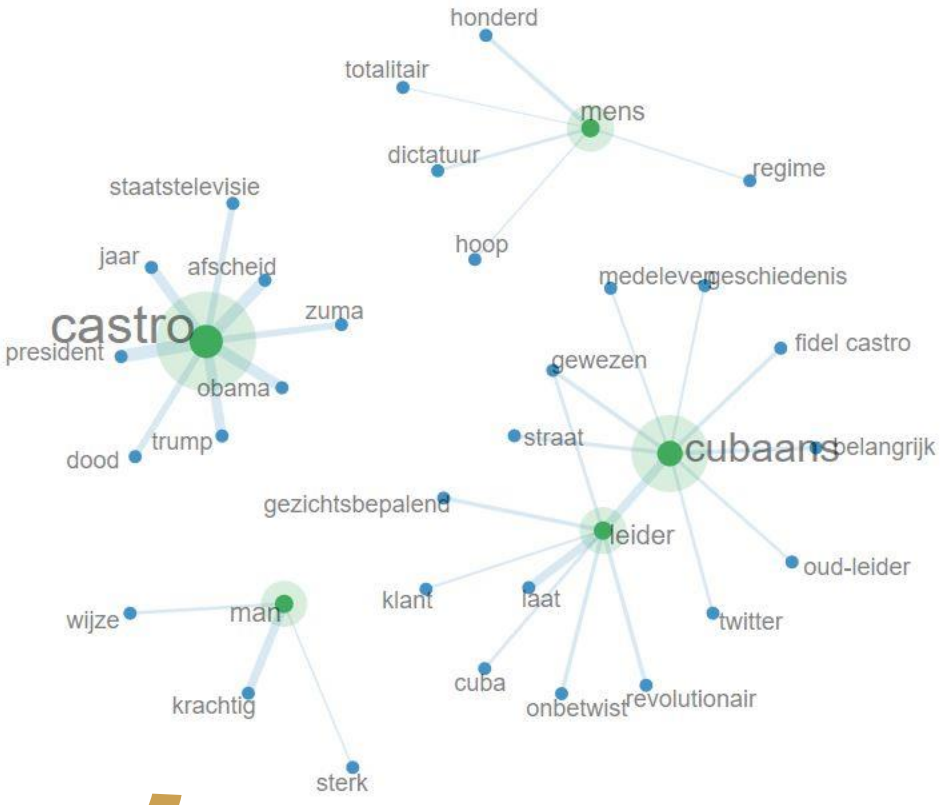
ATLANTA (Reuter) — De cholera-epidemie in Peru wordt veroorzaakt door de El Tor-stam van de ziekte, die voor het eerst in 1961 werd geïdentificeerd in Indonesië.

Dit heeft het Amerikaanse overheidscentrum voor ziektebeheersing CDC in Atlanta gisteren bekendgemaakt.

De El Tor-bacterie heeft zich in de afgelopen dertig jaar bijna over heel de wereld verspreid in wat het CDC een 'pandemie' noemt. Het instituut merkt op dat Zuid-Amerika tot afgelopen maand gespaard was gebleven.

Cholera in Peru uitloper van 'pandemie'

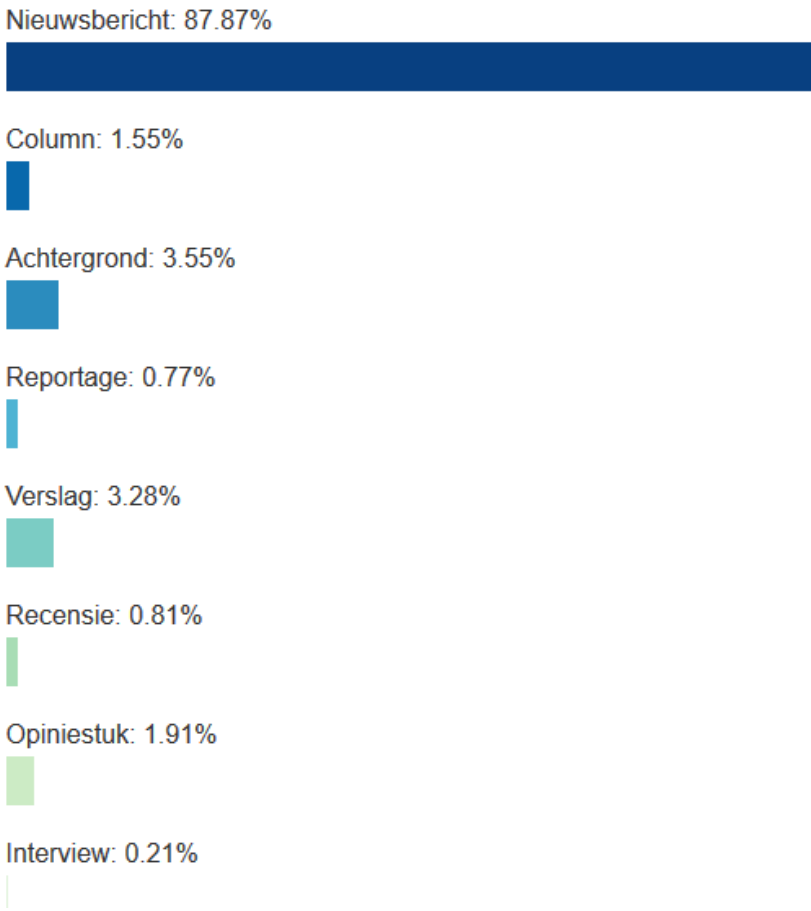
ATLANTA (Reuter) — De cholera-epidemie in Peru wordt veroorzaakt door de El Tor-stam van de ziekte, die voor het eerst in 1961 werd geïdentificeerd in Indonesië. Dit heeft het Amerikaanse oyerheidscentrum voor ziektebeheersing CDC in Atlanta gisteren bekendgemaakt. De El Tor-bacterie heeft zich in de afgelopen dertig jaar bijna over heel de wereld verspreid in wat het CDC een 'pandemie' noemt. Het instituut merkt op dat Zuid-Amerika tot afgelopen maand gespaard was gebleven.



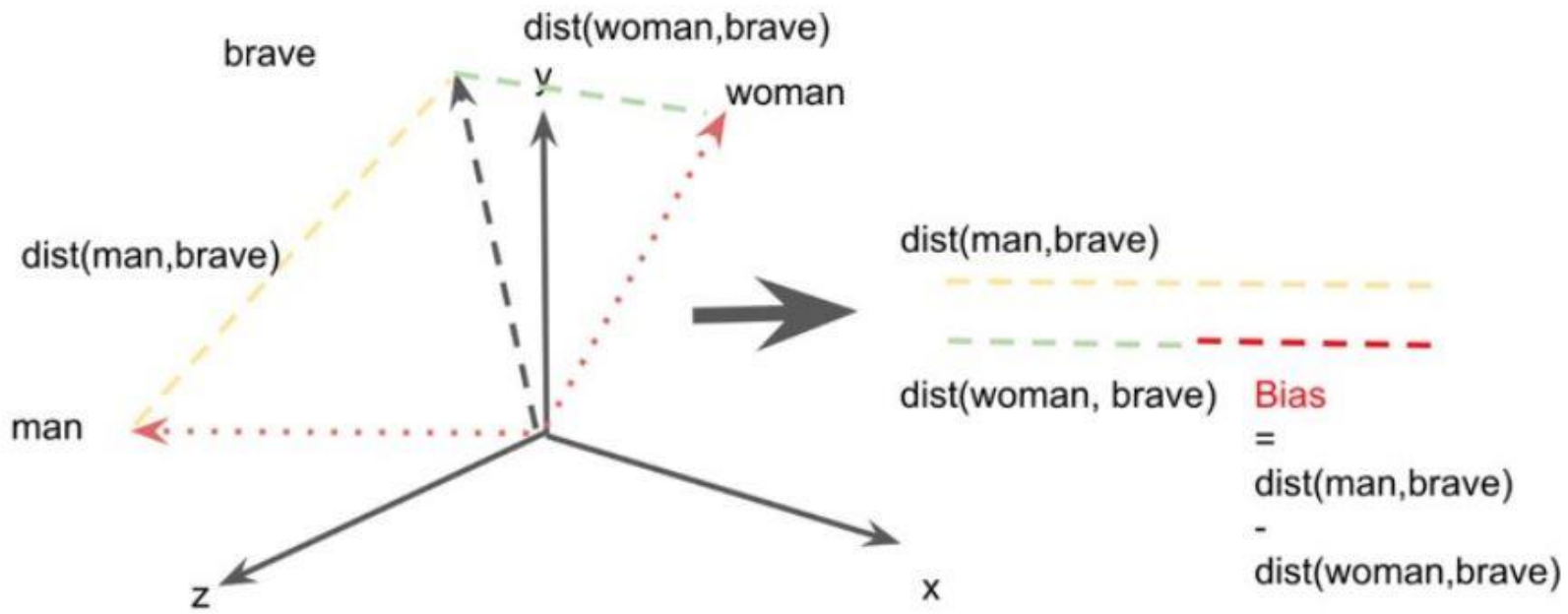
KB national library of the netherlands

Genre classifier

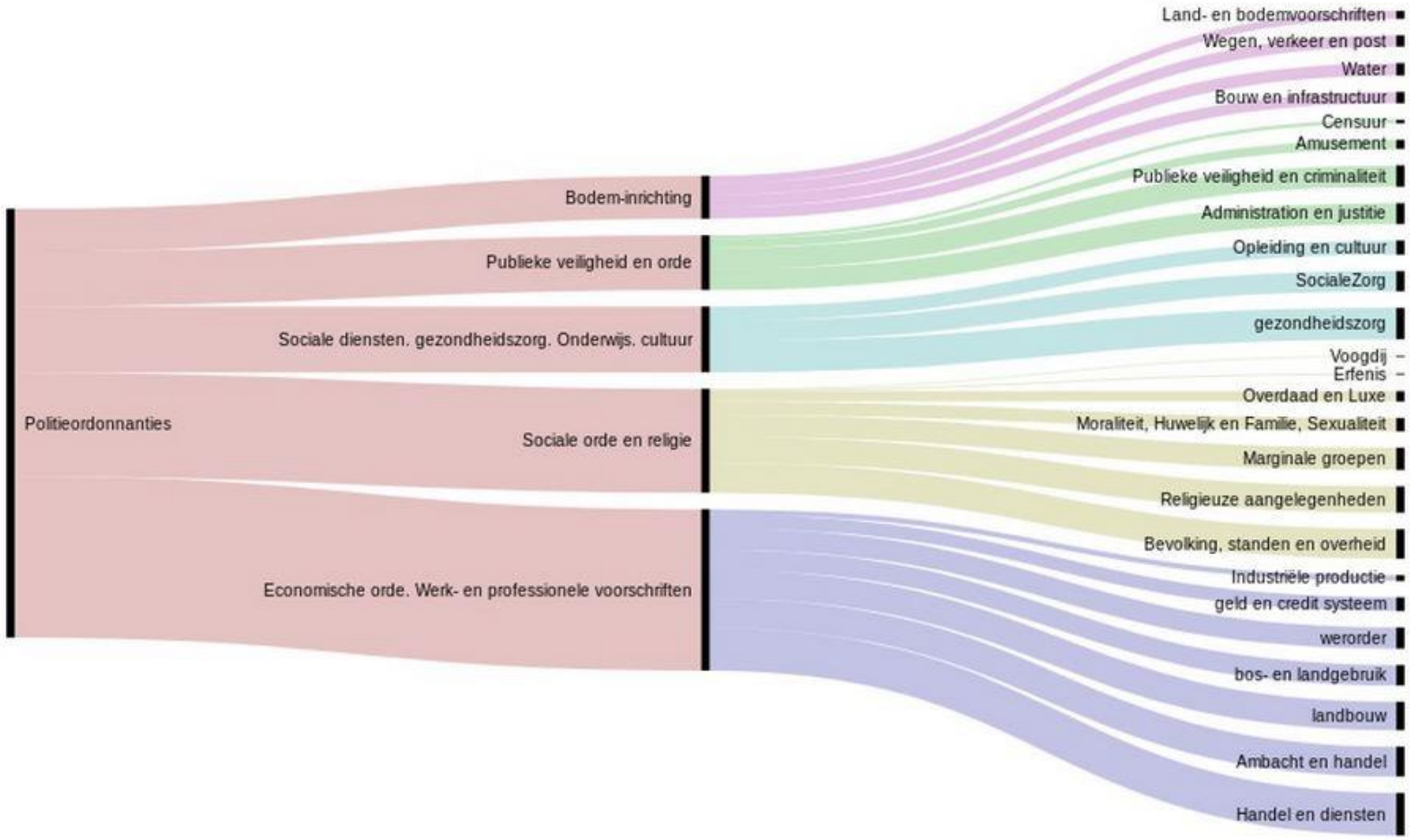
Churchill, die jarenlang de eerste plaats innam onder de meest bewonderde mannen, doch deze het vorige jaar moest afstaan aan dr Drees, heeft kans gezien de eerste plaats weer te halen met een stijging in populariteit van vijf procent, terwijl dr Drees met een daling van vier procent weer op de tweede plaats kwam te staan. Dit is een van de resultaten van het onderzoek van het Nederlands Instituut voor de Publieke Opinie, dat onlangs alleen aan mannen over het gehele land verspreid en uit alle lagen van de bevolking de vraag stelde: Welke van alle nu levende mannen, leden van de koninklijke familie niet meegerekend, bewondert U het meest? Churchill kreeg vijftien procent van de "stemmen", dr Drees 13, Eisenhower 6, Jan van Zutfen 6, Albert Schweitzer 3, Paus Pius XII 2, Adenauer 2, oud-rninifter Lieftinck 2, Einstein 1 en Abe Lenstra tenslotte ook 1 procent. Jan van Zutfen, Adenauer en Einstein komen dit jaar voor het eerst op de lijst van de meest bewonderde mannen voor. (ANP)



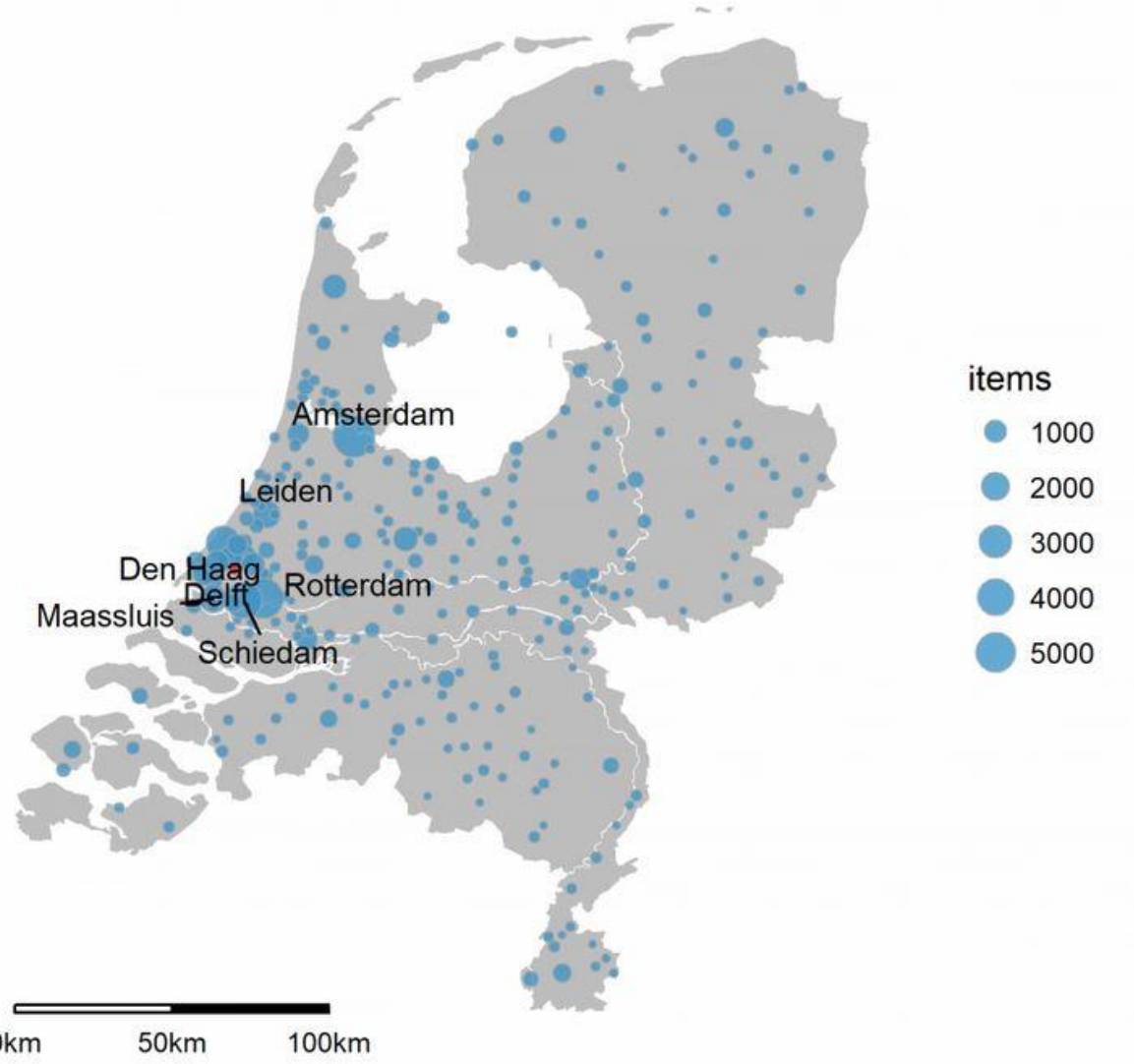
Gender bias in historical newspapers



Classify books of ordinances



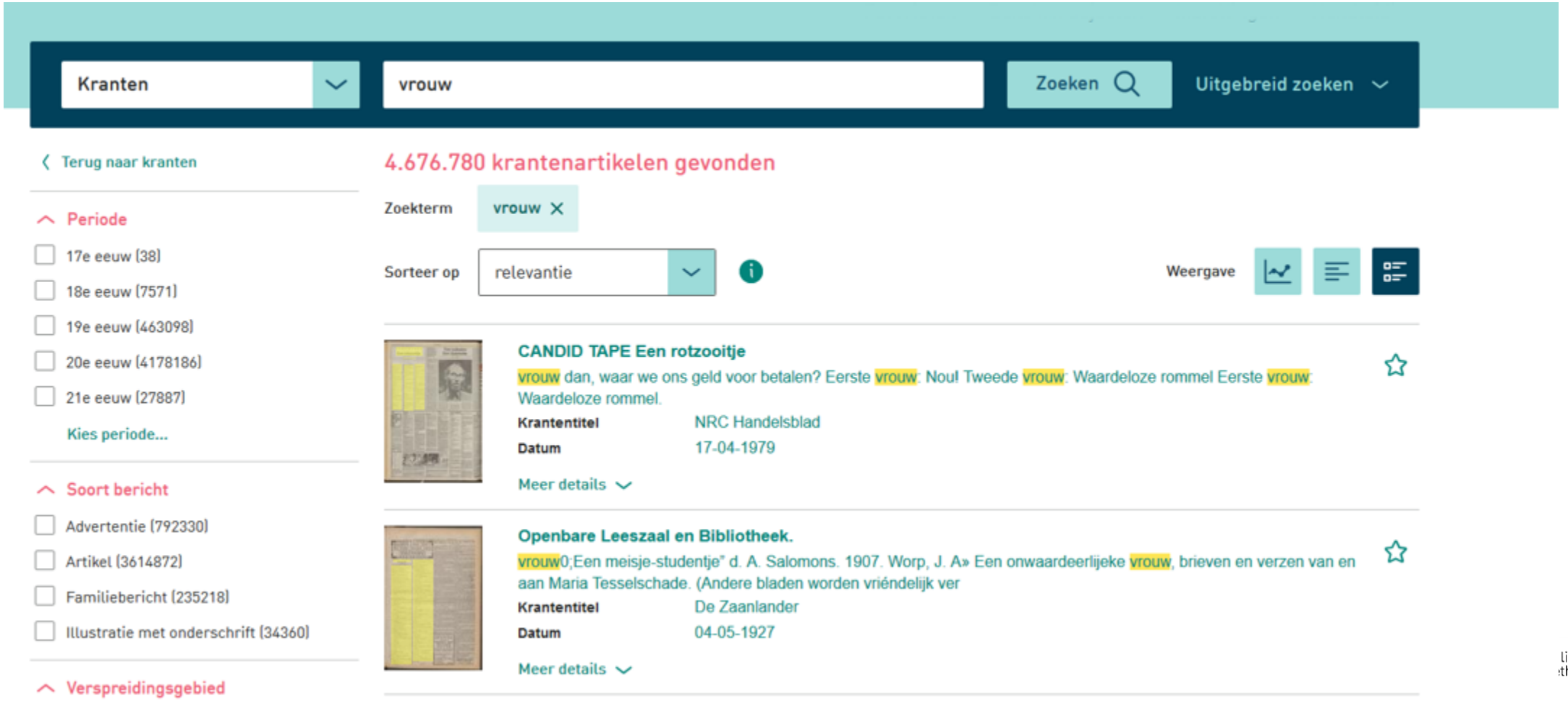
Cities in historical newspapers



OCR PROBLEMS

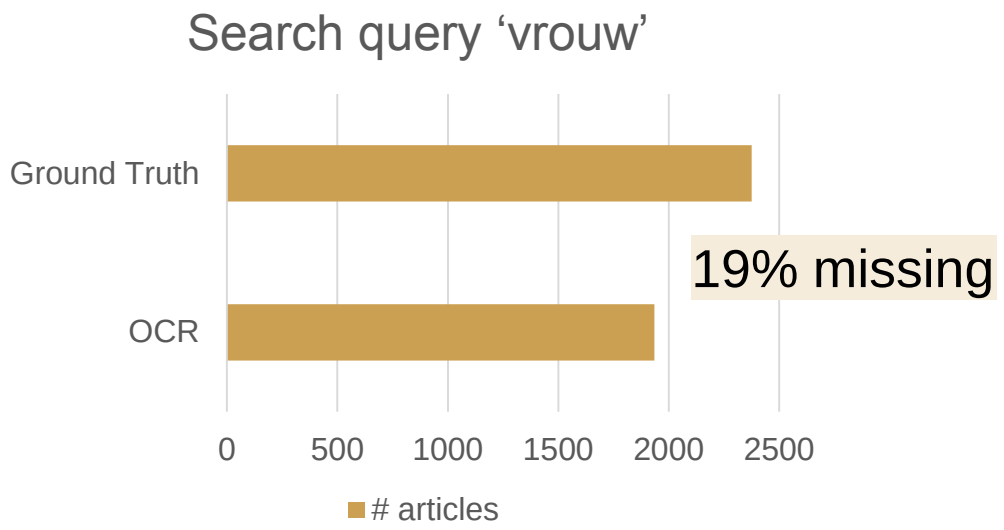
INFORMATION RETRIEVAL

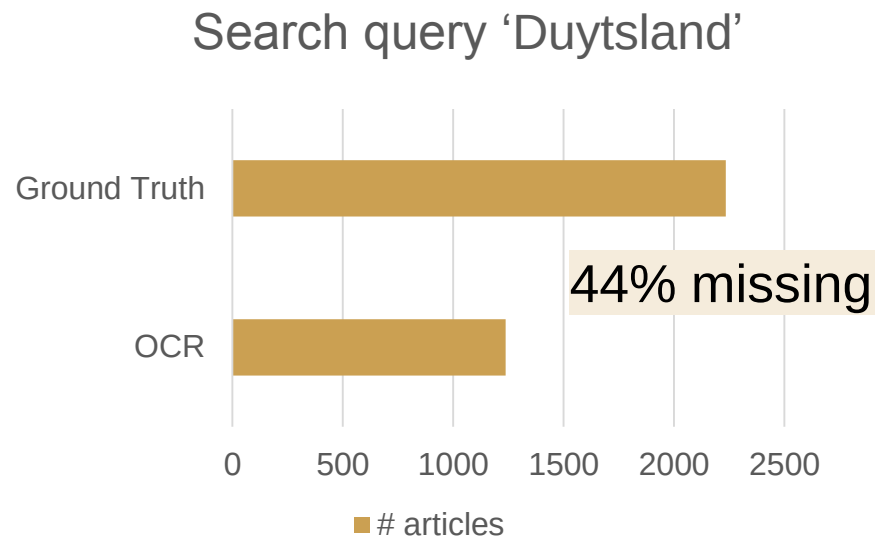
Information retrieval

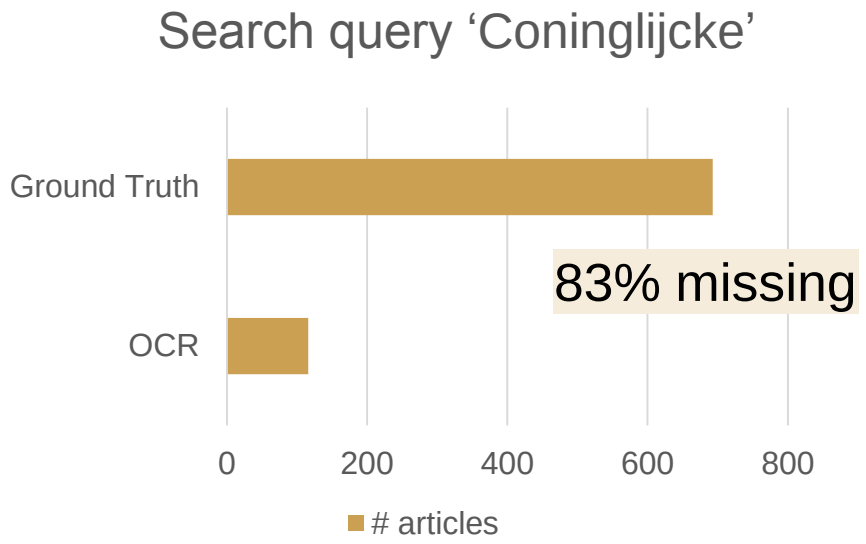


library
etherlands

Information retrieval







KB national library of the netherlands

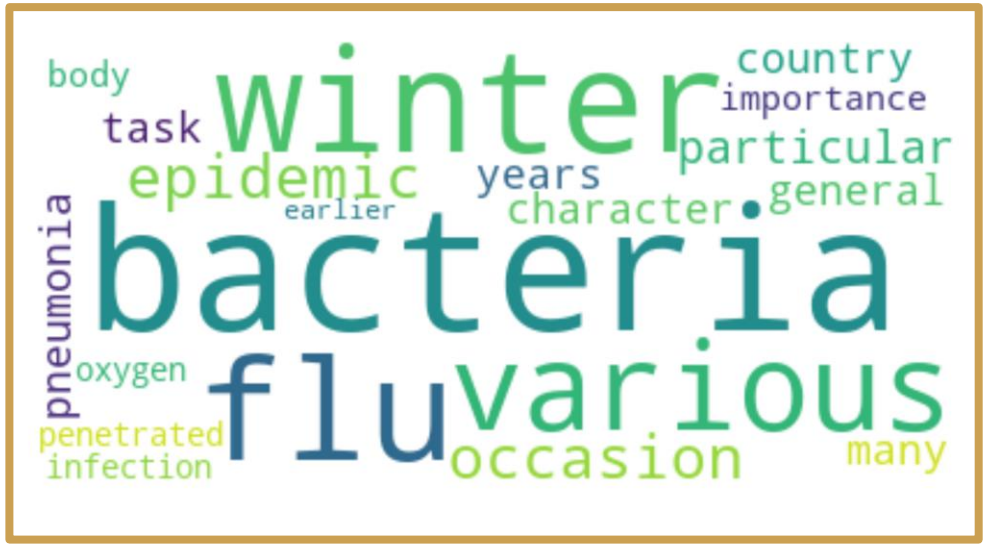
ANALYSES

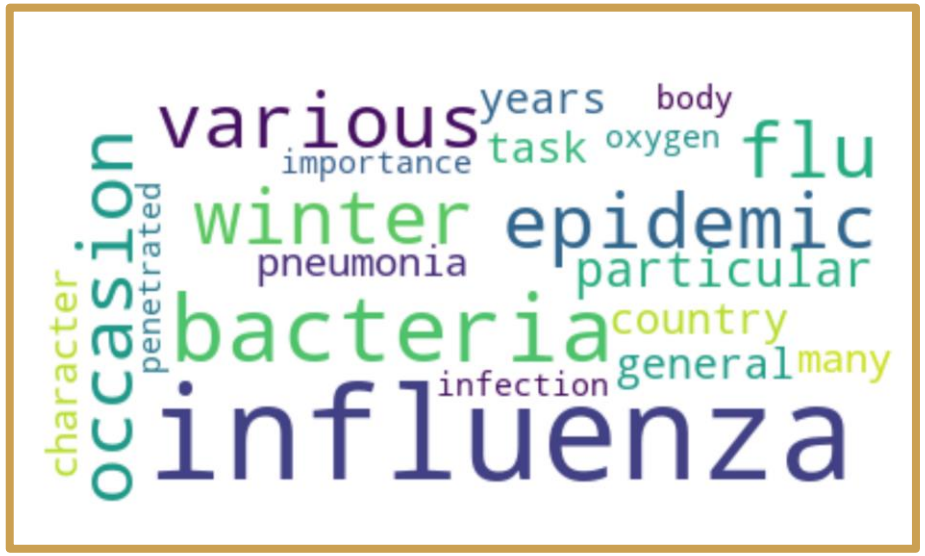
Analyses - wordclusters

Topic: reports about Poland in 1672

Analyses - wordclusters

Topic: relation between ‘disease’ and ‘influenza’ in newsarticles

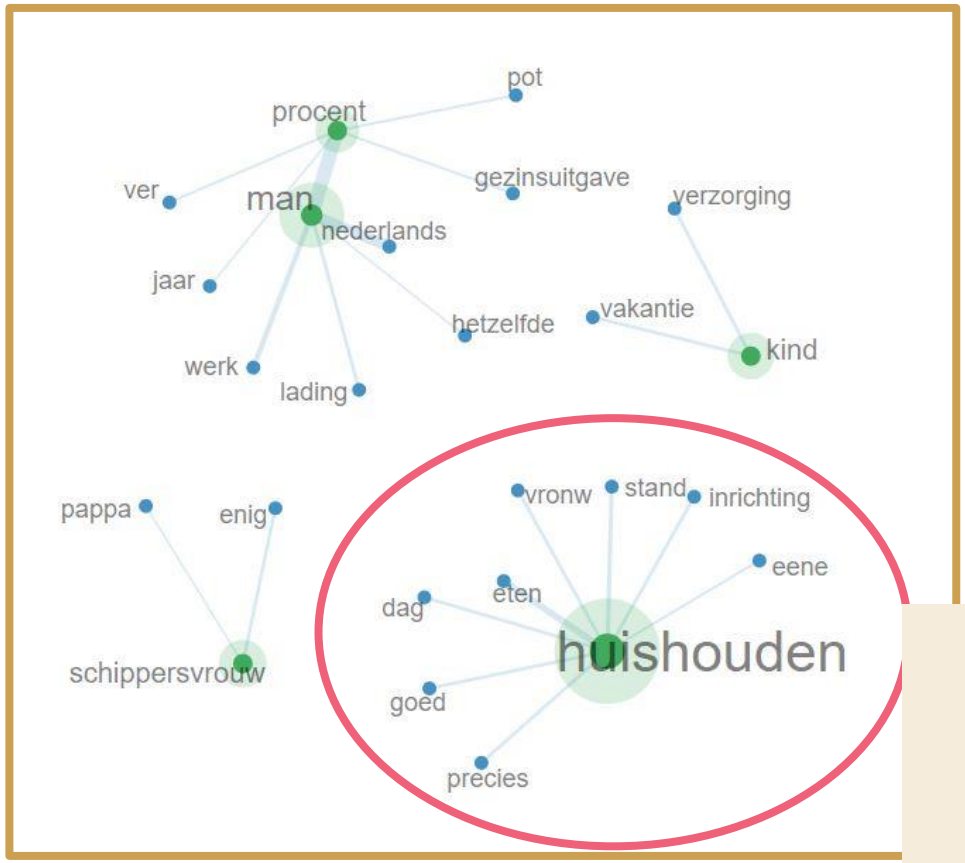


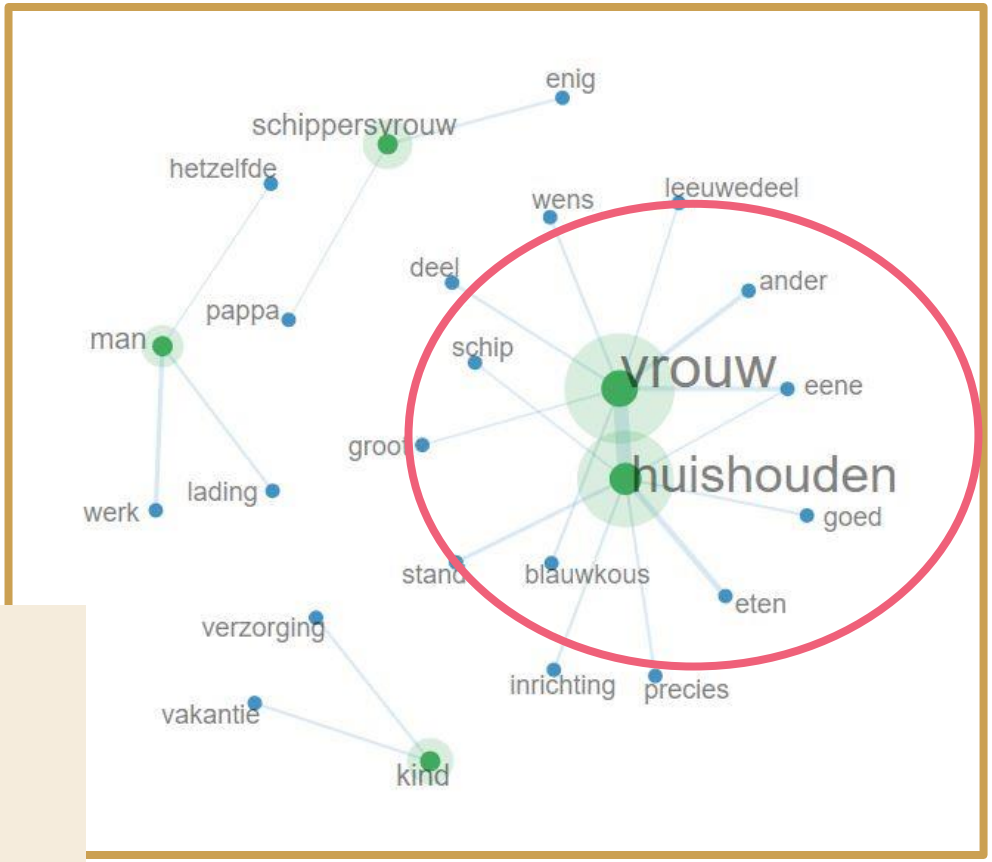


influensa
influemza
infleunza
imfluenza
influanza

Analyses – keyword assocation

Topic: relation between ‘household’ and ‘woman’ in newsarticles



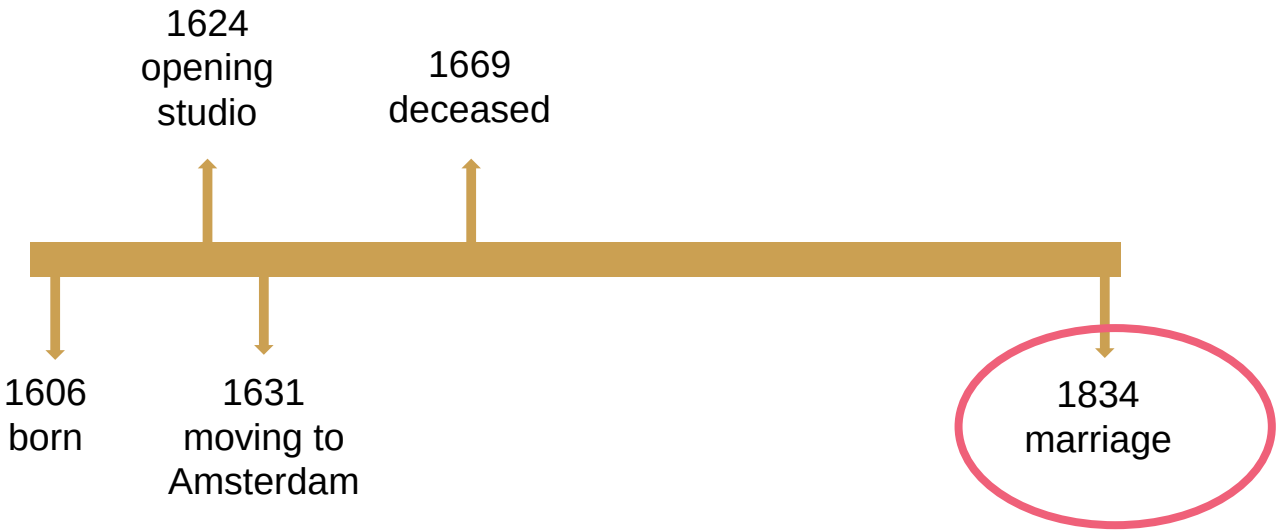


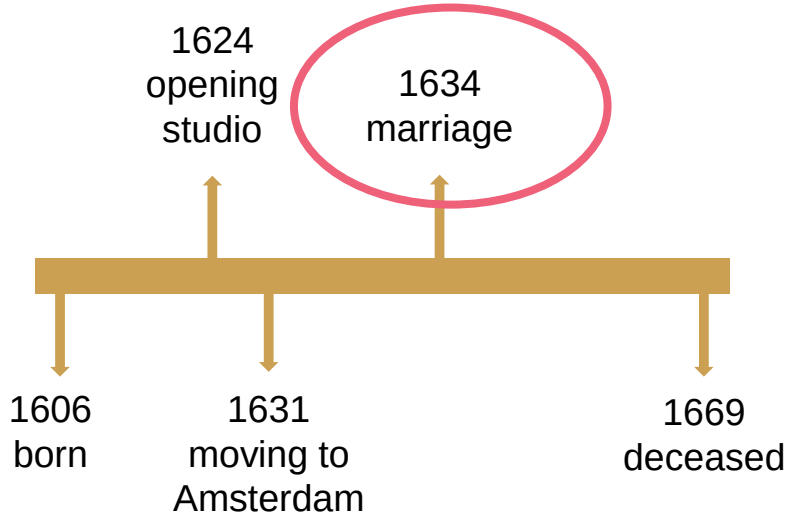
vronw
vrowe
vruow
vrcuw
viouw

household = huishouden
woman = vrouw

Analyses – timelines

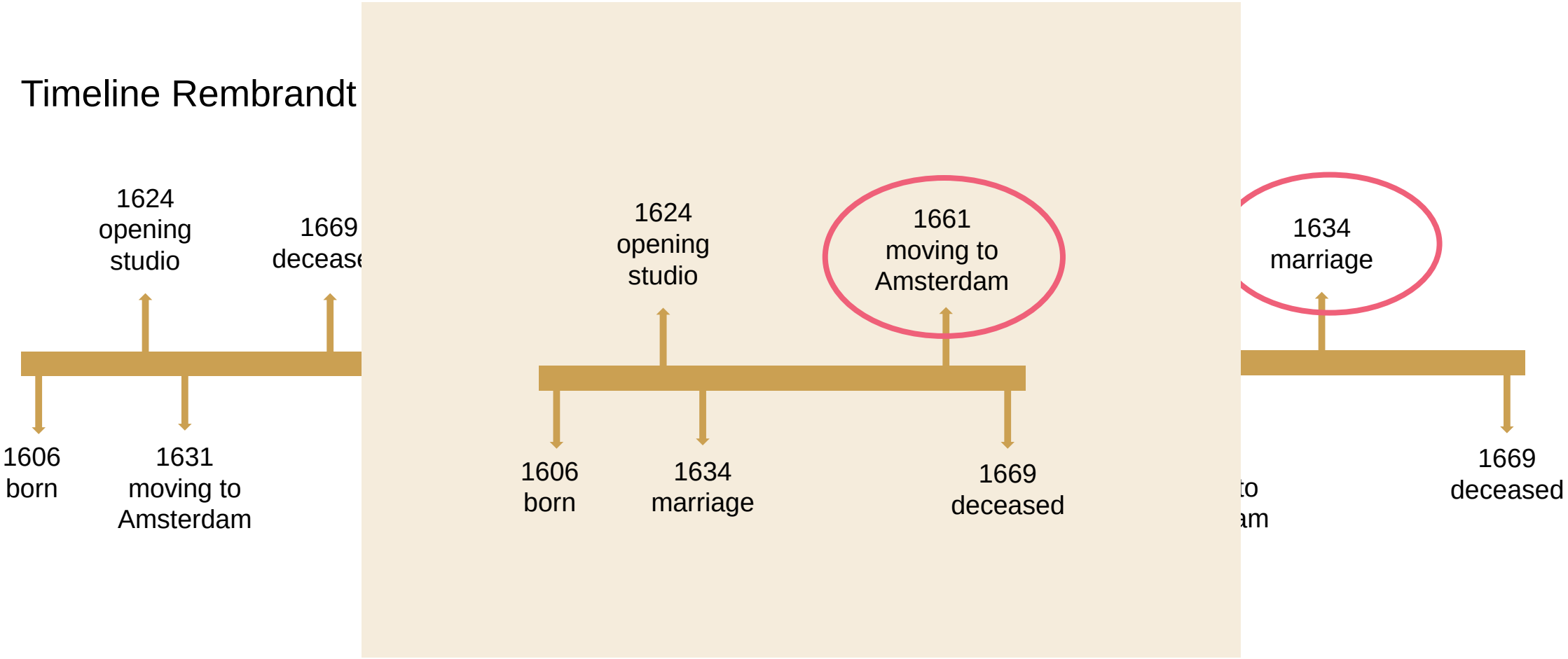
Timeline Rembrandt van Rijn – Dutch painter





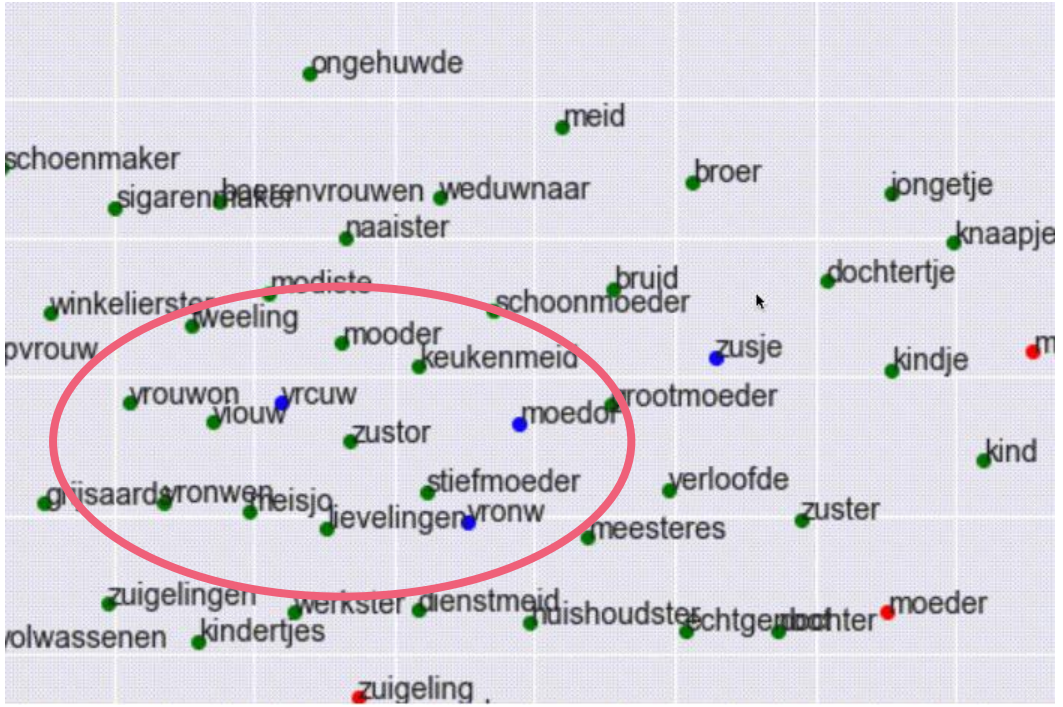
Analyses – timelines

Timeline Rembrandt



Is all OCR useless?

- Research by dr. Giovanni Colavizza
- When OCR is not to bad:
  - Topic modelling
  - Document classification
  - Word embeddings



Analyses – Named Entity Recognition

Correct

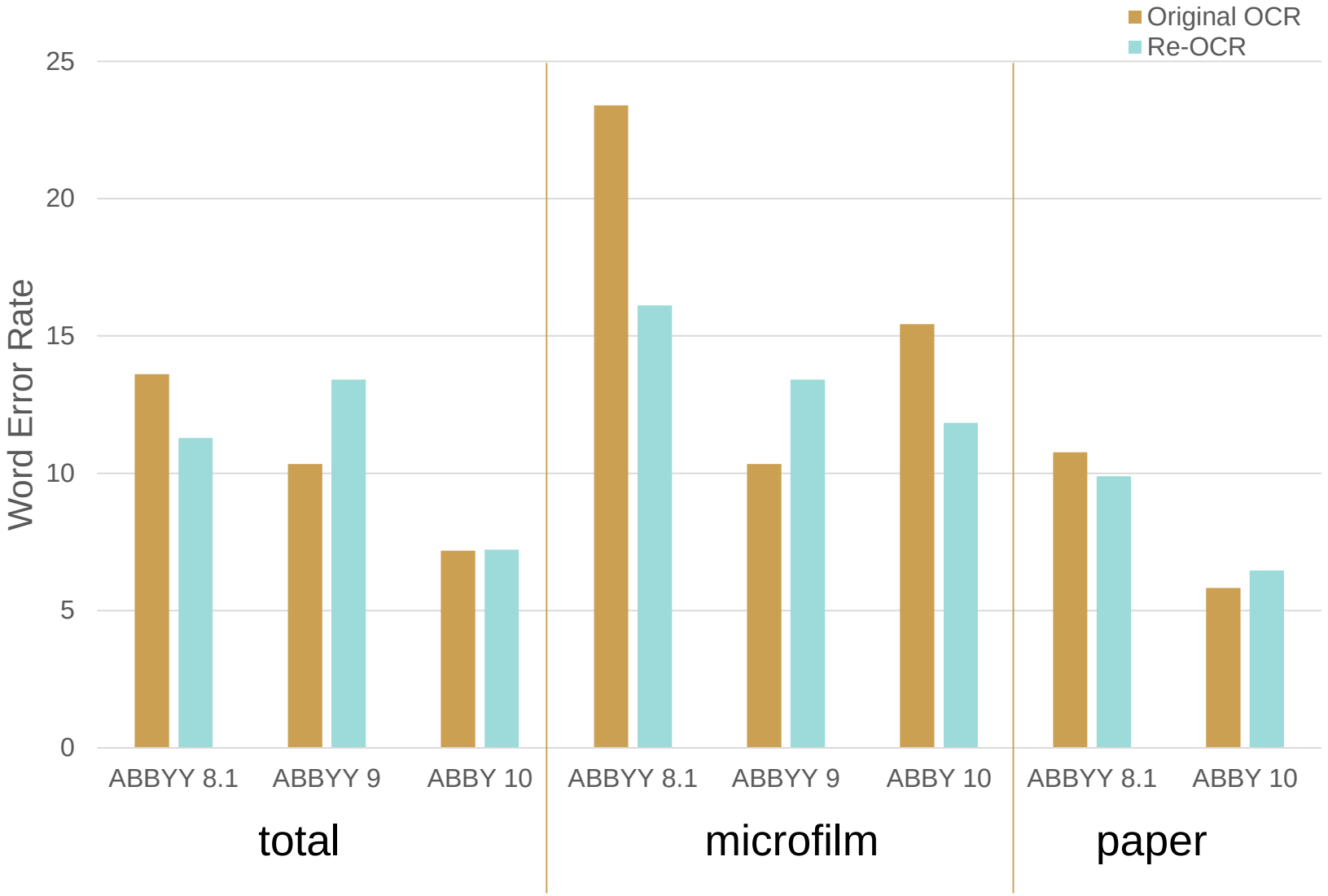
Errors

WHAT TO DO?

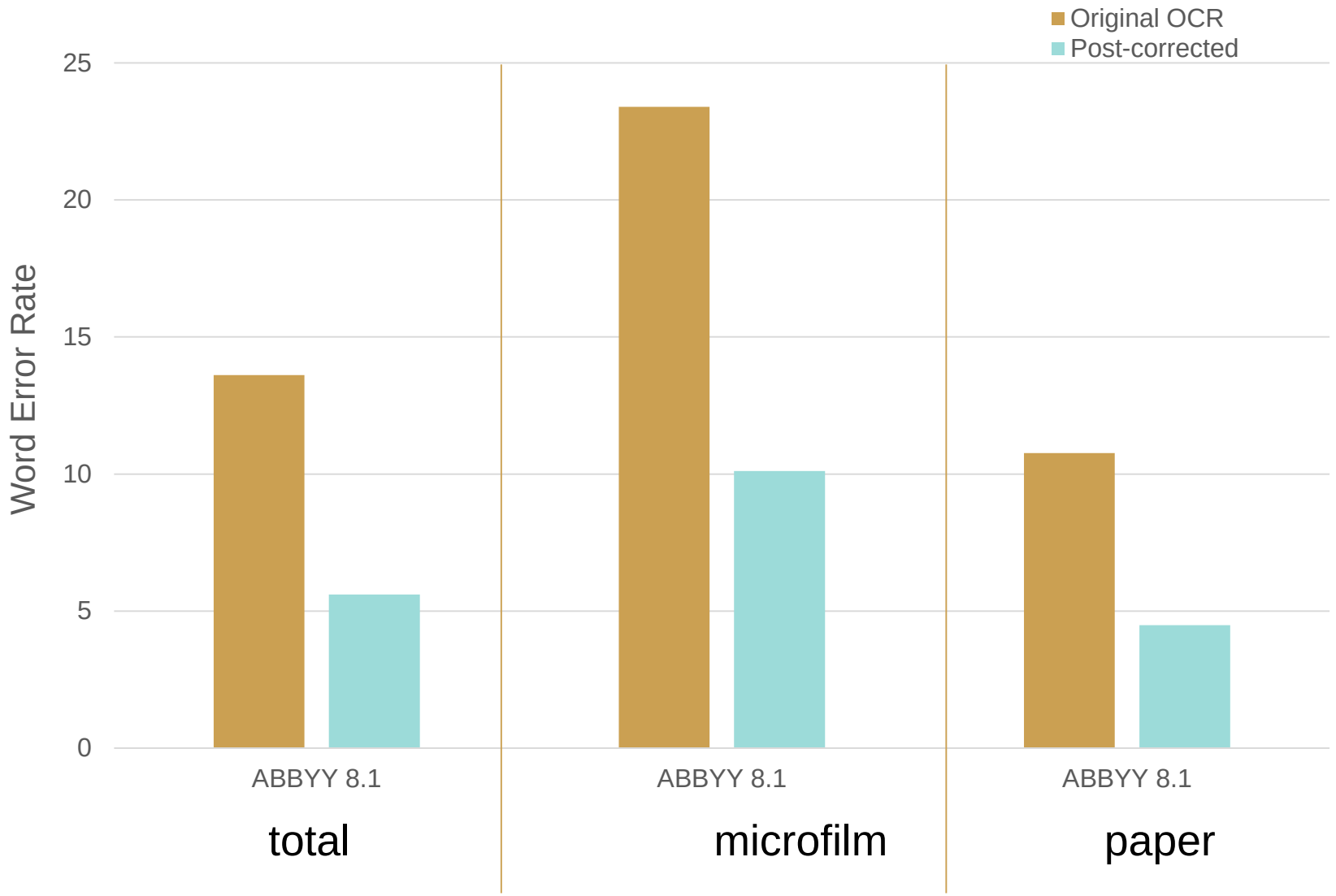
Optimize digitisation process

- Pre-select material
- Software based on document characteristics
- Most recent software

Re-ocr old material



Post-correction



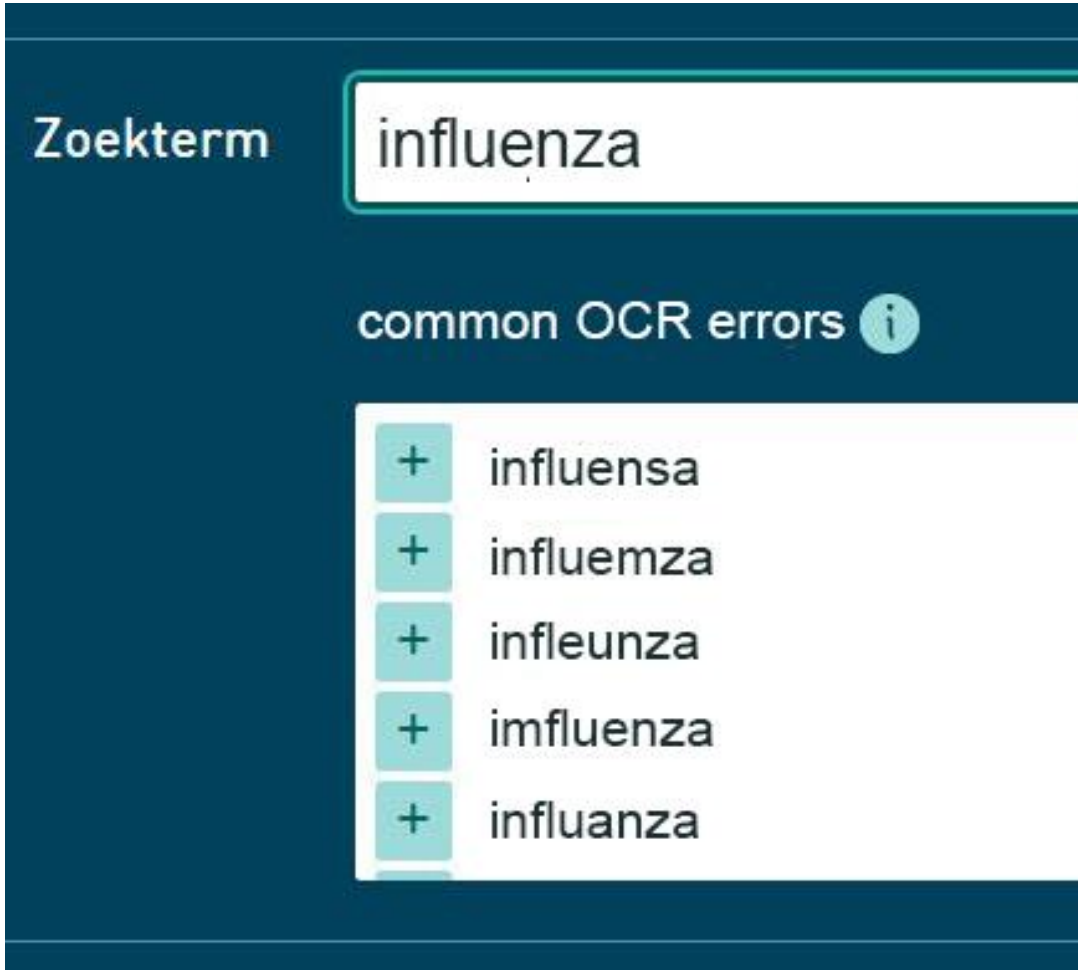
Transparancy

- OCR software
- Post-corrected or not
- Data source (microfilm, paper)

- Estimated quality

Data accessibility

- Other interfaces
- Expand queries



Improve OCR quality,

improve Digital Humanities research!

KB national library of the netherlands

Questions?

KB national library of the netherlands