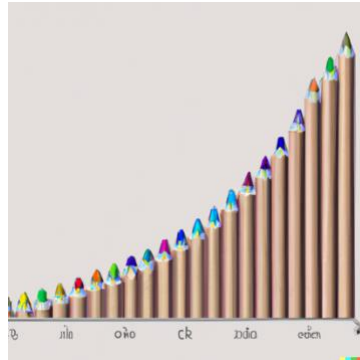


Digitale begripsgeschiedenis

Tutorial A/B/Thesis – 2023/24

Pim Huijnen



De groeiende digitalisering van cultureel erfgoed – denk aan de kranten, boeken, tijdschriften en archieven die steeds vaker online te vinden zijn – geeft veelbelovende nieuwe impulsen aan een bronnenintensieve vorm van geschiedschrijving: de begripsgeschiedenis. Traditioneel is de begripsgeschiedenis geïnteresseerd geweest in politieke cultuur en de vraag hoe veranderende invullingen van begrippen als ‘volk’, ‘vrede’, ‘vrijheid’ of ‘voortgang’ (evenals woorden die met andere beginletters dan v) daarmee verband hielden - met name in de scholen rondom Reinhart Koselleck in Bielefeld en Quentin Skinner en J.G.A. Pocock in Cambridge. De veranderende betekenis van deze concepten zochten historici in een veelheid aan bronnen, van boeken tot overheidsdocumenten, van wetteksten tot correspondentie en van kranten tot encyclopedieën.

Interessant aan de massadigitalisering waarmee veel erfgoedinstellingen bezig zijn, is niet alleen dat ze historici in staat stelt veel eenvoudiger in grote en diverse hoeveelheden bronnen te zoeken, maar ook dat ze de *aard* van het beschikbare bronnenmateriaal verbreedt. We krijgen steeds eenvoudiger toegang tot historische bronnen als kinderboeken, damestijdschriften, wetenschappelijke journals, opiniebladen en kookboeken die soms lange periodes beslaan en digitaal doorzoekbaar zijn. Deze bronnen maken een nieuw type begripsgeschiedenis mogelijk die nu eens naar manifestaties van *culturele* verandering kijkt: hoe, bijvoorbeeld, verandert de lading van ‘gezondheid’ in historische recepten? Wat definieert ‘beschaafd’ door de tijd heen in kinderboeken? Welke nieuwe invullingen geven kranten mettertijd aan woorden als ‘duurzaamheid’? Betekenisverandering van dit soort woorden – zoals de verschuiving van ‘trauma’ van een lichamelijke verwonding naar een mentale kwetsuur – kan interessante inzichten bieden in hoe wij naar de wereld kijken en wat wij belangrijk vinden.

In deze tutorial gaan jullie zelf aan de slag met culturele begripsgeschiedenis. Digitale bronnen zijn daarbij het vertrekpunt (kijk bijvoorbeeld eens [hier](#) voor inspiratie) en digitale technieken en methoden vormen de wijze van aanpak. Digitale technieken hebben voor het doen van begripshistorisch onderzoek het voordeel dat ze de onderzoeker in staat stellen 1) grote hoeveelheden bronnen ineens te analyseren, 2) daarmee onderzoek over de lange termijn mogelijk te maken en 3) specifiek de talige aspecten van historische begrippen te onderzoeken. Deze tutorial gaat hiermee over begripsgeschiedenis als een interessante tak van historisch onderzoek, maar óók over hoe digitale methoden en technieken in historisch onderzoek te gebruiken. Met de toenemende digitalisering worden deze technieken alleen

maar relevanter en dat maakt een kennismaking hiermee voor geschiedenisstudenten geen overbodige luxe.

In deze tutorial gaan we het in dit verband hebben over het (al of niet geautomatiseerd) verzamelen van digitale bronnen en het formatteren en/of opschonen van het gevonden materiaal, maar vooral over digitale analysemethoden. De historicus die digitaal zo flexibel mogelijk wil zijn, leert zelf [programmeren](#). Een introductie in de programmeertaal Python vormt daarom onderdeel van dit tutorial, om jullie een eerste zetje te geven richting interessante analysetechnieken. Die introductie zou voldoende moeten zijn om de code die jullie ter beschikking zal worden gesteld ter analyse van de bronnen, te gebruiken.

De bijeenkomsten in dit tutorial hebben een tweeledig karakter. Enerzijds zullen we literatuur lezen en bespreken over klassieke en digitale begripsgeschiedenis en aan de hand daarvan gezamenlijk concreet maken hoe we dit type historisch onderzoek zelf in de praktijk kunnen brengen. Anderzijds komen praktische oefeningen aan bod over het verzamelen van digitaal materiaal, evenals het opschonen, analyseren en visualiseren ervan. Een gastspreker van de Koninklijke Bibliotheek in Den Haag, die internationaal voorop loopt als het gaat om digitalisering, zal ons vertellen hoe de ‘digitale wending’ het werk van erfgoedinstellingen verandert.

Cursusdoelen

(I) Inhoudelijke kennisverwerving van de klassieke, culturele en digitale begripsgeschiedenis. Deze kennis doe je op via de te lezen literatuur, door bespreking tijdens de colleges, en in verdiepte vorm door de casus die je uitwerkt voor je eindpaper.

(II) Het aanleren van een praktische, maar ook kritische omgang met digitale bronnen. Het gaat hierbij zowel om heuristiek – het vinden van de juiste bronnen via relevante repositories en zoektermen – als om bronnenkritiek – het kunnen bepalen van de waarde van en het productief inzetten van die bronnen. Je zult voor je presentatie je bronnencorpus presenteren en de kritische omgang ermee verder uitdiepen in de eindpaper.

(III) Het kennismaken met en toepassen van digitale methoden en technieken voor cultuurhistorisch onderzoek. Deze methoden en technieken komen aan de orde in de te lezen literatuur en via praktische oefeningen in de colleges en je zult ze voor het onderzoek dat uitmondt in de eindpaper zelf toepassen.

Toetsing

De toetsing voor dit vak bestaat uit twee onderdelen.

(I) Presentatie van je onderwerp (30% van het eindcijfer)

(IIa) Eindpaper (70% van het eindcijfer, zie hieronder)

(IIb) Is deze tutorial je Thesis tutorial? Dan leg je je onderwerp vast in overleg met de docent en schrijf je als eindpaper een uitgebreide opzet, inclusief werkplan, voor je thesis.

Meer uitleg over presentatie en eindpaper vind je verderop in deze cursushandleiding.

Weekschema

(I) 16 november: Wat is begripsgeschiedenis?

Locatie: JK15A - 201

Deze eerste week starten we met de vraag wat begripsgeschiedenis als school en methode nu eigenlijk kenmerkt. Welke vormen van begripsgeschiedenis zijn er? Wat is het doel van begripsgeschiedenis – en hoe onderscheidt ze zich van ideeëngeschiedenis, mentaliteitsgeschiedenis, sociale geschiedenis en andere vormen van geschiedenis? Jordheims tekst geeft gestructureerd antwoord op deze vragen. Lees deze tekst dan ook nauwkeurig door. Daarnaast staat een tekst van Koselleck op het programma. Onder ‘further reading’ vind je meer literatuur die relevante en interessante achtergrond bij het onderwerp van deze week biedt.

Te lezen:

- Helge Jordheim, ‘Conceptual History’, in: *Bloomsbury History: Theory and Method Articles* (London 2021), DOI: <http://dx.doi.org/10.5040/9781350970878.066>.
- Reinhart Koselleck, ‘Social history and conceptual history’, in: idem, *The practice of conceptual history: timing history, spacing concepts* (Stanford 2002) 20-37.

Further reading:

- Quentin Skinner, ‘Meaning and Understanding in the history of ideas’, *History and Theory* 8 (1969) 3-53.
- Melvin Richter, ‘Reconstructing the history of political language: Pocock, Skinner, and the Geschichtliche Grundbegriffe’, *History and Theory* 29 (1990) 38-70
- Hans Erich Bödecker, ‘Concept - meaning - discourse. Begriffsgeschichte reconsidered’, in: I. Hampsher-Monk et.al. (eds.), *History of concepts. comparative perspectives* (Amsterdam 1998) 51-64.
- Ifversen, Jan (2011). ‘About key concepts and how to study them’, *Contributions to the History of Concepts* 6:1, 65-88.
- J.L. Austin, *How to do things with words. The William James Lectures delivered at Harvard University in 1955* (Oxford 1962).
- Roland Barthes, *Mythologies* (Parijs 1957): ‘Myth today’ (p. 109ff)
- Reinhart Koselleck, ‘Die Geschichte der Begriffe und Begriffe der Geschichte’, in: idem, *Begriffsgeschichten. Studien zur Semantik und Pragmatik der politischen und sozialen Sprache* (Frankfurt a.M. 2006) 56-76.
- Reinhart Koselleck, ‘Stichwort: Begriffsgeschichte’, in: idem, *Begriffsgeschichten. Studien zur Semantik und Pragmatik der politischen und sozialen Sprache* (Frankfurt a.M. 2006) 99-102.
- Arthur O. Lovejoy, *The Great Chain of Being. A Study of the history of an Idea* (Harvard 1936): ‘[I Introduction. The study of the history of ideas](#)’ and ‘[IX The outcome of the history and its moral](#)’

Nuttige links:

- Instituut voor de Nederlandse Taal: <https://ivdnt.org/>
- Historische Woordenboeken: <https://gtb.ivdnt.org/search/>
- Etymologisch Woordenboek van het Nederlands: <https://www.etymologie.nl/>

(II) 23 november: Van klassieke naar digitale begripsgeschiedenis

Locatie: JK15A - 201

Deze week gaan we uitvoerig in op de voorzet die Jordheim vorige week al gaf: hoe digitale technieken een nieuwe impuls aan de klassieke begripsgeschiedenis hebben gegeven. We zullen in de les de specifieke focus leggen op de kansen die digitalisering specifiek voor de cultuurgeschiedenis biedt en enkele voorbeelden de revue laten passeren van digitale begripsgeschiedenis. Ook in de 'further reading' vind je voorbeelden van empirische digitale begripshistorische studies.

Te lezen:

- Michael Gavin et.al., 'Spaces of Meaning: Conceptual History, Vector Semantics, and Close Reading', Matthew K. Gold en Lauren F. Klein (eds.), *Debates in the Digital Humanities* (Minneapolis 2019), 243-267.
- D. Nguyen et.al., 'How we do things with words: analyzing text as social and cultural data', *Frontiers in Artificial Intelligence* 3 (2020), DOI: <https://doi.org/10.3389/frai.2020.00062>.

Further reading:

- Peter de Bolla, *The Architecture of Concepts. The Historical Formation of Human Rights* (New York 2013) 11-47.
- Pumfrey, Stephen, Paul Rayson and John Mariani (2012). '[Experiments in 17th century English: manual versus automatic conceptual history](#)', *Literary and Linguistic Computing* 27:4, 395-408.
- Melvin Wevers en Marijn Koolen, 'Digital Begriffsgeschichte: Tracing semantic change using word embeddings', *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53 (2020) 226-243, DOI: <https://doi.org/10.1080/01615440.2020.1760157>.
- Arianna Betti and Hein van den Berg, 'Modelling the History of Ideas', *British Journal for the History of Philosophy* 22 (2014) 812-835.
- Scharloth, Joachim, David Eugster, and Noah Bubenhofer (2013), 'Das Wuchern Der Rhizome. Linguistische Diskursanalyse Und Data-Driven Turn' In: Dietrich Busse and Wolfgang Teubert (eds.) *Linguistische Diskursanalyse: Neue Perspektiven* (Wiesbaden: Springer VS) 345-80.
- David Armitage, 'What's the Big Idea? Intellectual History and the Longue Durée'
- Joris van Eijnatten & Pim Huijnen (2021), 'Something Happened to the Future. Reconstructing Temporalities in Dutch Parliamentary Debate, 1814-2018', *Contributions to the History of Concepts* 16, pp. 52-82, DOI: 10.3167/choc.2021.160204.
- Marten Scheffer et.al., 'The rise and fall of rationality in language', *Proceedings of the National Academy of Sciences* 118, 51 (2021). DOI: <https://pnas.org/doi/full/10.1073/pnas.2107848118>.

(III) 30 november: Introductie in digitale tekstanalyse

Locatie: DH Workspace Drift 27 – E.032 (12-14 uur), JK15A – 201 (14-15 uur)

Deze week gaan we programmeren! Online zijn allerlei tools en programma's te vinden die je in staat stellen basale tekstanalyses uit te voeren. Hier kun je een heel eind mee komen. Veel flexibeler is evenwel de historicus/-a die zelf kan programmeren. Om de daad bij het woord te voegen is dat wat we deze week eens gaan proberen – en wel aan de hand van de online cursus [Python Programming for the Humanities](#) van UvA-collega Folgert Karsdorp.

Ons college vindt deze week plaats in de DH Workspace in de UB, dat speciaal uitgerust is om in groepsverband achter de computerschermen te kruipen.

Voorbereiding:

Probeer ter voorbereiding op het college van deze week les 1 van Python Programming for the Humanities te doen. We zullen het college besteden aan vragen en onduidelijkheden, dus geen nood als je niet direct alles begrijpt. Deze cursus wordt aangeboden als Jupyter Notebooks, een interactieve en gedocumenteerde presentatievorm van programmeercode. Om een Jupyter Notebook te openen, moet je Anaconda hebben geïnstalleerd – een platform dat alle noodzakelijke dingen om Python en Jupyter op je pc te draaien, installeert. Op [Karsdorps cursuswebsite](#) vind je een handleiding hoe je Anaconda installeert (en op [The Programming Historian](#) ook). Als je Anaconda na installatie opstart, kun je vanaf het beginscherm Jupyter Notebooks openen, die in je browser zal starten. Daarbinnen kun je Notebooks openen of creëren vanuit je eigen pc.

Je docent zal je Karsdorps tutorial tijdig beschikbaar stellen, evenals de notebooks die we in de rest van deze cursus gaan gebruiken.

(IV) 7 december: Een eigen digitale begripsgeschiedenis

Locatie: DH Workspace Drift 27 – E.032 (12-14 uur), JK15A – 201 (14-15 uur)

Dit is de week om ter zake te komen. Aan het einde van de week zul je het onderwerp moeten hebben gekozen dat je in de rest van de tutorial zult uitwerken. We zullen specifiek ingaan op het verzamelen van materiaal, ‘data’ dus. Zie voor vereisten omtrent de data en enkele suggesties aan het einde van de cursushandleiding. Daarnaast zullen we kennismaken met programmeercode om die data te analyseren.

De literatuur van deze week helpt je hierbij en gaat in op de specifieke vormen van bronnenkritiek die de omgang met digitale bronnen met zich meebrengt. Wie daarover meer wil weten, kan terecht op het platform voor digitale bronnenkritiek [Ranke.2](#) van het Luxembourg Centre for Contemporary and Digital History (C2DH).

Te lezen:

- Marijn Koolen, Jasmijn van Gorp en Jacco van Ossenbruggen, ‘Toward a model for digital tool criticism: Reflection as integrative practice’, *Digital Scholarship in the Humanities* 34 (2019) 368-385.

Voor, tijdens of meteen na afloop van dit college maak je een keuze voor een begrip – de rest van de cursus staat in het teken van het uitwerken van dat onderwerp. De deadline voor het meedelen van een onderwerp is: **vrijdag 15 december, 12u ('s middags)**.

(V) 14 december: Geen college i.v.m. excursies

Deze week hebben we geen college vanwege de excursies van de andere tutorials. Zorg dat je zelf voortgang boekt wat betreft het vinden van tekstbestanden en het werken met de Python-methodes.

(VI) 21 december: Gastles Mirjam Cuper (KB) en presentaties

Locatie: DH Workspace Drift 27 – E.032

Vandaag is Mirjam Cuper bij ons te gast. Mirjam werkt als data scientist voor de Koninklijke Bibliotheek waar ze o.a. onderzoek doet naar automatische tekstherkenning. De KB loopt internationaal voorop als het gaat om digitalisering van cultureel erfgoed en de computationele toepassing van gedigitaliseerd materiaal. Daarover zal Mirjam een presentatie geven.

Voorbereiden:

De helft van de studenten presenteert deze week zijn of haar onderwerp in een presentatie van ca. 10 minuten (zie uitleg hieronder). Besteed aandacht aan het historische begrip dat je wilt onderzoeken, de data die je daarvoor hebt verzameld en de technieken op basis waarvan je hoopt te vinden wat je zoekt. Elke presentatie zal worden voorzien van commentaar door een medestudent, maar het is nadrukkelijk de bedoeling dat ook de rest van de groep zich mengt in een productieve discussie over mogelijke onderzoekspaden en -technieken.

(VII) 18 januari: Gastles Ruben Ros (UL) en presentaties

Locatie: DH Workspace Drift 27 – E.032

Vandaag krijgen we Ruben Ros als gast op bezoek. Ruben is promovendus aan de Universiteit Leiden, nadat hij eerder geschiedenis studeerde aan de UU. Zijn promotieonderzoek is gebaseerd op computationele methoden, op basis van welke hij ‘technocratisch denken’ in het Nederlandse parlementaire debat in de twintigste eeuw onderzoekt. Hierover zal hij een presentatie geven.

Voorbereiden:

De tweede helft van de studenten presenteert deze week zijn of haar onderwerp in een presentatie van ca. 10 minuten (zie uitleg hieronder). Besteed aandacht aan het historische begrip dat je wilt onderzoeken, de data die je daarvoor hebt verzameld en de technieken op basis waarvan je hoopt te vinden wat je zoekt. Elke presentatie zal worden voorzien van commentaar door een medestudent, maar het is nadrukkelijk de bedoeling dat ook de rest van de groep zich mengt in een productieve discussie over mogelijke onderzoekspaden en -technieken.

Deadlines

22 januari: Deadline Thesis tutorial opzet

29 januari: Deadline tutorial eindopdracht

Presentaties

De presentaties zijn bedoeld om de rest van de groep deelgenoot te maken van je onderzoek, wat in deze tutorial mede betekent dat de presentator de anderen mobiliseert om mee te denken. Samenwerking is een kernwaarde van digitale geschiedenis, aangezien ‘it is rare for a single individual (be it instructor or student) to have all of the technical and interpretive skills that are needed to reach a learning outcome’ (Guiliano 2022). Hou de presentaties daarom kort. Beperk je slides tot de kern: één over je begrip, één over je methode en eindig met een slide met vragen aan de rest van de groep die je verder kunnen helpen. Het is de expliciete bedoeling van de presentaties discussie uit te lokken.

Elke presentatie heeft een referent, die de discussie zal aftrappen. De presentator zorgt ervoor dat de referent uiterlijk twee dagen van tevoren de teneur van de presentatie (en de vraag!) krijgt. De rest van de groep wordt uiteraard nadrukkelijk uitgenodigd aan de discussie deel te nemen.

De presentatie vormt 30% van het eindcijfer.

Eindpaper

Julie schrijven de eindpaper voor deze tutorial als een Jupyter notebook (dus een .ipynb bestand). Hiervoor maak je in Jupyter Notebooks een nieuw bestand aan (op het startscherm zie je rechtsboven een knop ‘New’, waar je op kunt klikken. Selecteer vervolgens ‘Python 3 (ipykernel)’). *Zie eventueel de al genoemde [tutorial op The Programming Historian](#) hierboven of die van Cuper en Den Boer hieronder voor een zachte introductie in het gebruik van Notebooks.*

De cellen in Jupyter Notebooks kun je zowel gebruiken voor code (‘Y’) als voor tekst (‘M’). Op deze manier kun je in je paper de lezer door jouw toepassing van Python heenleiden – en misschien zelfs interactief maken. Bedenk dat de tekst in Jupyter wordt opgemaakt in de formatteertaal [Markdown](#). Deze taal bevat speciale codes voor opmaak, waar je even aan zult moeten wennen (zoals een # voor kopjes of ** om tekst vet te drukken). Hier vind je de officiële [cheat sheet](#) met een lijst van de meestvoorkomende codes. Als je dat prettig vindt, kun je je tekst eerst in Word of andere tekstverwerker typen en later naar je Notebook kopiëren, maar let er dan dus op dat je de opmaak opnieuw moet doen (met andere woorden: het heeft geen zin je tekst in je standaard tekstverwerker mooi op te maken).

De formele eisen voor je eindpaper zijn niet anders dan voor ‘papieren’ papers. Belangrijk is dat een beredeneerde aanpak van begripsgeschiedenis te herkennen is, geworteld in de literatuur die we hebben gelezen. Ook moet je analyse voldoen aan de algemene criteria die passen bij een academisch historisch essay. Het moet, met andere woorden, kritisch zijn en van voldoende niveau en lengte (ca. 4000 woorden exl. voetnoten en literatuur); het moet een of meerdere cultuurhistorische perspectieven, theorieën, of methodes (uit de cursussen van

blok 1) toepassen; het moet, uiteraard, voldoen aan de normen en waarden inzake wetenschappelijke integriteit van de Universiteit Utrecht (zie [UU fraude en plagiaatregels](#)). Daarnaast schept de vorm waarin jullie deze eindopdracht schrijven enige verwachting aangaande creativiteit met zich mee, die in de beoordeling een rol zal spelen. Zie voor vereisten en suggesties betreffende het bronnenmateriaal hieronder.

Voor studenten van de educatieve richting is er de ruimte om het onderzoek een didactische insteek te geven.

Vrijdag 15 december is deadline voor de onderwerpskeuze.

N.B. Hoewel we het gebruik van ChatGPT als hulp bij het programmeren aanmoedigen, wordt het doen voorkomen van door ChatGPT gegenereerde tekst als eigen werk in het eindpaper beschouwd als fraude. Jullie mogen ChatGPT wel als 'sparringspartner' inzetten bij het schrijven van het eindpaper. Neem in dit geval een paragraaf op met uitputtende uitleg over de wijze waarop je de tool hebt gebruikt.

Inspiratie & verdieping

Peter Verhaar, Text and Data Mining Tutorial Universiteit Leiden:
<https://cdsleiden.github.io/tdm-tutorial/Welcome.html>

Een collectie notebooks die vooral taalkundig georiënteerd zijn: tokeniseren, lemmatiseren, type-token-ratio, part of speech tagging, named entity recognition, sentiment analysis, etc.

Karsdorp, F., Kestemont, M., & Riddell, A. (2021). Humanities Data Analysis: Case Studies with Python. Princeton University Press: <https://www.humanitiesdataanalysis.org/index.html>

Een handboek voor het gebruik van data in de geesteswetenschappen, als open access versie in notebooks geschreven. De bijbehorende code is te downloaden via de website. Veronderstelt wel enige voorkennis in Python en gaat flink de diepte in.

Melanie Althage, Martin Dröge, Torsten Hiltmann en Philipp Schneider, Python für Historiker:innen. Ein anwendungsorientierter und interaktiver Einstieg: <https://digital-history-berlin.github.io/Python-fuer-Historiker-innen/home.html>

Een introductie in Python voor historici in het Duits. Geen eigenlijk analyses, maar een grondige kennismaking met de basics van Python. Net als het boek hierboven in notebooks geschreven en als zodanig zelf te doorlopen.

Mirjam Cuper en Ethan den Boer, Automatically extract XML content with Python:
<https://kbnlresearch.github.io/xml-workshop/intro.html>

Een tutorial specifiek bedoeld om xml bestanden (van de Koninklijke Bibliotheek) te openen en gebruiken in Python. Geschreven in notebooks en met een aardige introductie in het gebruik van Jupyter Notebooks.

The Programming Historian Python lessons:
<https://programminghistorian.org/en/lessons/?topic=python>

The Programming Historian is een collectie lessen en tutorials over computationele methoden van en voor historici. Op de website is van alles te vinden, waaronder 29 lessen in specifieke toepassingen van Python – van beginnersniveau tot toepassingen voor gevorderden. Absoluut het checken waard!

Data

Er zitten geen beperkingen aan de onderwerpen die je in deze tutorial begripshistorisch mag bestuderen, maar er wel zijn een paar vereisten voor de gegevens – de data - die je zult gebruiken. Uiteraard moet je kunnen beargumenteren in welke mate de gebruikte data jouw onderwerp representeert. Daarnaast gelden er enkele formele vereisten:

- Je data moet een **comparatief element** bevatten. Je zou bijvoorbeeld handboeken over psychiatrie uit 1900 kunnen vergelijken met die uit 2000 om te kijken hoe bepaalde psychiatrische ideeën (concepten) zijn veranderd. Je zou de rol en betekenis van ‘vrijheid’ in linkse media met rechtse media kunnen vergelijken. Of je zou het complete oeuvre van Dostojevski kunnen nemen om te zien hoe zijn begrip van schuld en boete door zijn werken heen is verschoven. Deze (willekeurige) voorbeelden tonen aan dat de vergelijking zowel over tijd als tussen domeinen of geografische regio's kan gaan, en kan plaatsvinden tussen enkele of meerdere teksten
- Je moet in staat zijn om de data als leesbare tekstbestanden op je eigen computer te **downloaden**, bij voorkeur legaal. De bovenstaande voorbeelden klinken misschien aantrekkelijk, maar kunnen onhaalbaar blijken vanwege het ontbreken van direct beschikbare gedigitaliseerde gegevens en/of vanwege auteursrechtelijke beperkingen. Het verkrijgen van relevante of interessante data kan moeilijk zijn, omdat het letterlijk overal op het web verborgen kan zijn. Als je moeite hebt iets interessants te vinden, kijk dan eens bij de suggesties hieronder. Als laatste redmiddel werken krantenartikelen altijd goed voor de doelen van deze module (Nexis Uni, teruggaand tot het midden van de jaren 1990).
- Dit brengt ons bij het derde vereiste: het verkrijgen van een werkbare **hoeveelheid** data. Mijn vuistregel is dat digitale tekstanalyse de meeste toegevoegde waarde heeft wanneer je meer gegevens hebt dan je redelijkerwijs voor je onderzoek kunt lezen. Wat betekent dat? Als je data bestaan uit boeken, zouden twee tekstbestanden (dus twee boeken) al voldoende kunnen zijn. Als je van plan bent om kranten te bestuderen, moet je eerder denken in termen van honderden of duizenden artikelen. Over het algemeen werkt digitale tekstanalyse beter naarmate je meer gegevens hebt. In dit opzicht is er geen bovengrens. Maar er zijn ook praktische omstandigheden en niet elke laptop of pc zal het hier mee eens zijn. Daarom is het belangrijk dat je een balans vindt tussen voldoende data om robuuste resultaten op te leveren, maar toch klein genoeg om niet uren of dagen te hoeven wachten om ze te genereren. Ik raad je altijd aan om klein te beginnen. Wat werkt voor kleine hoeveelheden gegevens zal ook werken voor grotere hoeveelheden, dus je kunt altijd uitbreiden als de tijd het toelaat - hoewel een minimum van 6 documenten (teksten, bestanden) voor de technieken die we gebruiken wel goed isom te hebben.
- Ten slotte, in welk **formaat** moet je je data opslaan en gebruiken? De notebooks die we in de colleges zullen gebruiken werken over het algemeen met het platte tekstformaat (.txt) of het platte spreadsheetformaat (.csv). Online kom je het laatste niet erg vaak tegen. Het eerste des te meer. Dit is het standaardformaat dat de meeste

online archieven die het downloaden van digitale of gedigitaliseerde gegevens toestaan, aanbieden. Als dat het geval is, kies dan altijd voor .txt, omdat het gegevensformaat klein is en de notebooks er toch mee werken. Pdf's van tekst of andere tekstformaten (zoals .doc of .rtf) zijn ook bruikbare formaten, omdat we deze kunnen converteren naar .txt of .csv. Als je onderzoeksproject betrekking heeft op vergelijking over tijd, is het standaardformaat om je gegevens lokaal op te slaan in bestanden per jaar - idealiter, met het jaar als bestandsnaam (dus als '1982.txt', '1899.txt', etc.). Als alternatief kun je (ook) meer metadata in de bestandsnaam opnemen (zoals auteur of land). Als je bekend bent met spreadsheetformaten, ben je natuurlijk welkom om alle metadata in de kolomstructuur op te nemen.

Corpora (suggesties)

- Architectuurtijdschriften: <https://tresor.tudelft.nl/tijdschrift/architectuurtijdschriften/>
- Delpher Tijdschriften: <https://www.delpher.nl/nl/tijdschriften> (o.a. Libelle, Margriet, Vrij Nederland, e.v.a.)
- Delpher Boeken: <https://www.delpher.nl/nl/boeken> (o.a. veel (academische) leerboeken)
- Digitale Bibliotheek der Nederlandse Letteren: www.dbnl.nl
- Hathi Trust: <https://www.hathitrust.org/> ("De grootste gedigitaliseerde collectie copyright-vrije kennis", 18 miljoen items)
- (Digitale) bronnencollecties Huygens Instituut: <https://resources.huygens.knaw.nl/>
- Natuurtijdschriften: www.natuurtijdschriften.nl
- Nederlands Tijdschrift voor Geneeskunde: www.ntvg.nl (via UU-proxy, historisch archief raadpleegbaar via zoekfunctie)
- Project Gutenberg: www.gutenberg.org
- Bij je docent: digitale edities van o.a.
 - Allerhande (1954-1973, 1979-2022)
 - BMGN (ca. 1970-2020)
 - jaarverslagen BVD/AIVD (1946-2019)
 - DSM I-V
 - Economisch Statistische Berichten (1916-2016)
 - De Groene Amsterdammer (1877-1940)
 - Militaire Spectator (1832-2007)
 - Staten Generaal Digitaal (1814-2018)