

Meer data in CLARIAH's Media Suite dankzij nieuwe archiefovereenkomst

Archief Publieke Omroep toegankelijk voor wetenschappers

Een lang gekoesterde wens van mediawetenschappers is uitgekomen: sinds kort kunnen zij gearchiveerde radio- en televisie-uitzendingen van de Publieke Omroep online raadplegen. *Erica Renckens*

Zo'n dertig geesteswetenschappers kwamen de eerste week van juli bijeen in het Nederlands Instituut voor Beeld en Geluid voor de CLARIAH Media Studies Summer School. In teams werkten zij aan onderzoeksvragen waarin media centraal stonden. Hierbij gebruikten zij de Media Suite, een platform waarin wetenschappers onderzoek kunnen doen met audiovisueel materiaal en tekstuele mediabronnen. Zo analyseerden zij wie er wel en niet aan het woord komt in het vluchtelingendebat en of de verhouding mannen en vrouwen in het tv-programma Buitenhof is

veranderd. Voor het eerst konden zij ook op afstand de gearchiveerde uitzendingen van de Nederlandse Publieke Omroep raadplegen, dankzij een nieuwe archiefovereenkomst die online toegang exclusief voor wetenschappelijk gebruik toestaat. "De data op het platform zijn alleen toegankelijk met een account van een Nederlandse universiteit of onderzoeksinstituut", vertelt Eva Baaren, Adviseur Digital Humanities bij Beeld en Geluid.

Delen van data

Naast de collecties bij Beeld en Geluid biedt de door CLARIAH ontwikkelde Media Suite ook de mogelijkheid om met kranten van de Koninklijke Bibliotheek, de Desmet-collectie van EYE Filmmuseum en de Oral History-collecties van DANS onderzoek te doen, al dan niet gecombineerd. "De Media Suite is gebouwd op de groei, zodat ook collecties van

andere instellingen kunnen worden toegevoegd", aldus Baaren. "Onderzoekers kunnen straks ook hun eigen data importeren. Iemand kan dan zelf Twitter-data scrapen en die in samenhang met andere media analyseren."

Volop in ontwikkeling

De Media Suite beschikt over mogelijkheden data te doorzoeken, visualiseren, annoteren en analyseren. Daarnaast is er een persoonlijke gebruikersomgeving om resultaten te bewaren en annotaties te bewerken. Aan de tools wordt nog verder gebouwd. Baaren: "Onlangs hebben we een groot deel van het radio- en televisiemateriaal door ons spraakherkenningsysteem gehaald en de transcripties als data in de Media Suite toegevoegd. Uiteindelijk willen we dat gebruikers dit ook zelf met hun data kunnen doen."

De Media Suite is dus nog volop in ontwik-

keling – kan de gemiddelde geesteswetenschapper er wel al mee uit de voeten? "Het is zeker gebruiksvriendelijk", stelt Baaren. "We hebben ingezet op transparantie en duurzaamheid. Maar zoals met elke onderzoeksmethode moet je leren wat er kan en wat niet. Voor iemand met ervaring in de digitale geesteswetenschappen is het zeer goed te doen. Nieuwe gebruikers zullen we met onze kennis ondersteunen."

Vierde online versie

Eind 2018 wordt de vierde versie van de Media Suite opgeleverd, de laatste binnen CLARIAH-Core. Hiermee komt de voortgang van het project echter niet tot stilstand: ook in het vervolproject CLARIAH PLUS zal het platform verder doorontwikkeld worden.

mediasuite.clariah.nl/

Hoe zorgen semantische verbanden ervoor dat gebruikers online museumcollecties beter kunnen verkennen? Over deze vraag boog promovendus Chris Dijkshoorn zich de afgelopen jaren bij het Rijksmuseum.

Maarten Heerlien

Dijkshoorn, die zijn onderzoek deed in het kader van het COMMIT/-project SEALINC-Media (Socially Enriched Access to Cultural Media), analyseerde de zoektermen van gebruikers van Rijksstudio, het online collectieplatform van het museum en onderzocht hoe semantische links naar gestandaardiseerde vocabulaires bijdragen aan de diversiteit van zoekresultaten. Hij werd daarbij begeleid door Guus Schreiber en Lora Aroyo van de onderzoeksgroep User-centric Data Science aan de Vrije Universiteit (VU).

Dijkshoorn: "Het resultaat was een verbreding en verdieping van zoekresultaten door de relaties tussen collectieobjecten en thesaurustermen enerzijds en tussen thesaurustermen onderling anderzijds. Het succes van deze relaties hing sterk af van de rijkheid van de objectdata bij het Rijksmuseum zelf. Om deze te vergroten, lanceerden we Accurator, een online platform om specialistische kennis bij niche-experts te verzamelen en hiermee objectdata te verrijken."

Schat aan informatie

Terwijl Dijkshoorn de laatste hand legt aan zijn proefschrift, werkt hij al als databasebeheerder bij de afdeling Research Services van het Rijksmuseum. De resultaten uit zijn onderzoek gaat hij nu vertalen naar de praktijk. "Het Rijksmuseum is een Linked Data-uitdaging op zichzelf. Bij de collectie heeft het museum nog zoveel andere bronnen die relevant zijn bij de interpretatie ervan. Honderden meters documentatie, een enorme kunstbibliotheek, terabytes onderzoeksdata, enzovoort. Hoe verbind je dat allemaal aan elkaar op een betekenisvolle wijze die interne en

zoekresultaten toegankelijker

'Het Rijksmuseum is een Linked Data-uitdaging op zichzelf'



"We lanceerden Accurator, een online platform voor niche-experts" foto Maarten Heerlien

JONG TALENT

externe gebruikers zo volledig mogelijk ondersteunt in hun zoektocht naar informatie?"

Twee benen

Omdat Research Services nog niet de capaciteit heeft om intern grootschalige R&D-activiteiten te ontplooiën op het gebied van

datamanagement en content search, helpt het dat Chris met één been in de onderzoekswereld staat en met het andere in het museum. Het eerste onderzoeksvoorstel is dan ook al de deur uit én toegekend. "Samen met computerwetenschapper Victor de Boer en kunsthistorica Ingrid Vermeulen, beide van

'Hoe kun je bronnen zodanig aan elkaar verbinden, dat het gebruikers ondersteunt in hun zoektocht naar informatie?'

de VU, gaan we van een negentiende-eeuwse schilderijencollectie, die inmiddels is uitgewaaid over museale collecties wereldwijd, onderzoeken of we de samenstelling en herkomst kunnen reconstrueren. Een ideale Linked Data-uitdaging, want daarmee kunnen we de vele heterogene datasets die we hierbij nodig hebben normaliseren om ze veel efficiënter te bevragen."

chrisdijkshoorn.nl