

Lora Aroyo, professor in de computerwetenschappen:

# ‘Laat gebruikers actief vertellen en collecties verrijken’

“Erfgoedinstellingen beschikken over een goudmijn aan data waarmee ze hun maatschappelijke relevantie kunnen vergroten.” E-data interviewt Lora Aroyo, professor in de computerwetenschappen.

Maarten Heerlien

Begin oktober was Lora Aroyo keynotespreker tijdens de International Conference of Art Libraries. In haar kernboodschap richting bibliotheken, archieven en musea is Aroyo stellig: “gebruik je publiek om meer kennis en inzichten uit je collectie te halen. Zo kunnen we erfgoed interpreteren vanuit steeds meer perspectieven”. Samen met het Rijksmuseum, Beeld & Geluid en andere erfgoedinstellingen werkt Aroyo samen om collecties en objecten te verrijken met behulp van semantische technieken en machine learning.

## User data als sleutel

Aroyo: “Amazon zette door het gebruik van user data de retailmarkt op zijn kop. Erfgoedinstellingen hebben de sleutel in handen om datzelfde te doen. Door de userlogs van hun website, collectieportaal en andere online uitingen structureel te analyseren en trends daarin te identificeren, kunnen instellingen voortdurend anticiperen op de behoeften en voorkeuren van hun bezoekers. Zoektermen die online bezoekers invoeren bijvoorbeeld, geven inzicht in de diversiteit aan terminologie die ze daarbij bezigen. Met zulke inzichten kunnen instellingen hun eigen thesauri verrijken en zo hun zoekingen en aanbevelingen beter afstemmen op de vocabulaires van hun gebruikers.”

## Bredere context

Veel bezoekers zijn bovendien kundige erfgoedliefhebbers die staan te springen om actief en vrijwillig hun eigen kennis toe te voegen aan collecties en objecten waarvoor zij een passie hebben. Aroyo: “Het is voor musea essentieel om doelgroepen, of dat nou professionals, niche-experts of leken zijn, de ruimte en de middelen te bieden om collecties te verrijken met hun eigen annotaties en te gebruiken om hun verhaal te vertellen. Elk van die doelgroepen doet dat vanuit zijn eigen perspectief. Het vastleggen en leren begrijpen van al die verschillende perspectieven is waardevol want daardoor ontdek je wat elk van die groepen kan meebrengen om je collectie in een bredere context te plaatsen. Om zo meer en meer gestructureerde informatie te verzamelen over objecten en deze in te zetten om gebruikers continu nieuwe en rijkere ingangen op collecties te bieden. De uiteenlopende thematische experimenten die we deden rondom vogels, mode en bijbels, toonden aan hoeveel potentie dit heeft. Het bewerkstelligt inclusiviteit en creativiteit, belangrijke thema's voor erfgoedinstellingen.”

Crowdsourcing en nichesourcing, termen



In haar kernboodschap richting bibliotheken, archieven en musea is Aroyo stellig: ‘gebruik je publiek om meer kennis en inzichten uit je collectie te halen’. Credits: Nacho Guevara

waarmee Aroyo een combinatie van input vanuit de grote menigte met specialistische input van niche-experts aanduidt, zijn veelbelovende methoden om verschillende perspectieven op collecties te verzamelen. “Je zult over veel onderwerpen heel uiteenlopende uitspraken terugkrijgen van de crowd. Daaronder zullen zich onherroepelijk een aantal, in de ogen van de meerderheid, vreemde zienswijzen bevinden, waarvan je je niet kunt voorstellen dat er mensen zijn die zo tegen een onderwerp of object aankijken. Die afwijkingen moet je koesteren, want dan ga je inzien hoe controversieel een onderwerp of object eigenlijk is. En controversie zet mensen aan het denken, het bewustzijn van het bestaan van een tegenbewering is het begin van nieuwe kennis!”

## Patronen in Crowdtruth

De kans bestaat dat erfgoedinstellingen verdrienvoudigen in de zee van controversiële zienswijzen die dat mogelijk gaat opleveren. Aroyo: “Met dat gegeven in het achterhoofd hebben we Crowdtruth.org ontwikkeld. Ik doe onderzoek naar human-in-the-loop machine learning. Dat houdt kortweg in dat je computers traint om patronen te herkennen in grote hoeveelheden data op basis van verzamelingen door mensen geannoteerde voorbeelden, zogeheten ‘ground truths’. Samen met Chris Welty, één van de ontwikkelaars van IBM’s supercomputer Watson, onderzocht ik de methoden die onderzoekers doorgaans gebruiken om zulke ground truths te genereren. We constateerden dat dit vaak gebeurt op basis van een zevental misvattingen, bijvoorbeeld de aannames dat er op elke vraag slechts één

## INTERVIEW

### ‘Plaats je collectie in een bredere context’

juist antwoord mogelijk is en dat de input van de professional meer waard is dan die van de leek. Ook zijn het doorgaans momentopnames en worden veranderende inzichten over langere periodes niet ingecalculleerd. In CrowdTruth vangen we het hele spectrum aan mogelijke antwoorden op een vraag, ook door de tijd heen. We wegen hoeveel steun elk van die antwoorden geniet en stellen die weging

voortdurend bij op basis van nieuwe input. Voor erfgoedinstellingen kan deze methodiek doorlopend inzichtelijk maken hoe verschillende doelgroepen tegen collecties aankijken, hoe die zienswijzen zich tot elkaar verhouden en in hoeverre dat door de tijd heen verandert. Als ze dat in de infrastructuur van de organisatie inbedden, hebben instellingen een krachtig hulpmiddel om op innovatieve manieren de interactie met hun publiek aan te gaan.” Aroyo erkent overigens dat erfgoedinstellingen vaak onvoldoende expertise in huis hebben om dit soort datagedreven processen te kunnen beheren. “Datawetenschappers zoeken voortdurend naar partners om hun theorieën en methoden in praktijk te testen. Veel interessanter wordt het wanneer een erfgoedinstelling op óns afstapt met hun onderzoeksvragen. Daarmee wordt het voor beide partijen relevanter en dat vergroot de kans dat onderzoeksresultaten terugvloeiën naar de maatschappij. En dat is voor iedereen een gunstig perspectief.”

[lora-aroyo.org](http://lora-aroyo.org)

#### Lora Aroyo

Lora Aroyo studeerde computerwetenschap aan de universiteit van Sofia te Bulgarije en specialiseerde zich in kunstmatige intelligentie. In 2001 promoveerde ze aan de Universiteit Twente op het onderwerp *Task-based Approach to Information Handling Support for Web-based Education*. In de peri-

ode 2001-2006 bekleedde ze posities aan de Universiteit van Osaka, het CWI, de Universiteit Eindhoven. Vanaf 2006 was Aroyo werkzaam aan de Computer science faculteit van de Vrije Universiteit Amsterdam. Aldaar werd ze in 2016 benoemd tot professor en gaf ze leiding aan de onderzoeksgroep User-Centric Data Science.

Daarnaast is Aroyo onder andere als visiting researcher betrokken bij Columbia Data Science en CornellTech in New York en bij het Center for Advanced Studies van IBM Nederland en is ze Track Leader Creatieve Industrie bij Amsterdam Data Science. Ook is ze hoofdonderzoeker bij de New Yorkse startup Tagasauris Inc.