

Het Nederlandse fotografische geheugen is in toenemende mate online beschikbaar. Hoe zorg je ervoor dat je daarin doeltreffend kunt zoeken? Artificial Intelligence-technieken kunnen helpen.

Mathilde Jansen

De meeste beeldherkenningsalgoritmes zijn getraind op modern fotomateriaal en presteren daarom minder goed op historische beelden. Door de algoritmes te trainen op historisch materiaal kan het fotografisch geheugen aanzienlijk worden verbeterd. Melvin Wevers van het KNAW Humanities Cluster werkt aan zo'n algoritme voor de twee miljoen foto's uit de periode 1945-2004 van Fotopersbureau De Boer. De foto's liggen in de depots van het Noord-Hollands Archief. Een deel is gedigitaliseerd en wordt vanaf deze zomer door vrijwilligers van velehanden.nl van labels voorzien. Het door Wevers ontwikkelde algoritme is daarbij leidend.

### Sneeuw of zand?

Bestaande beeldherkenningstechnieken werken soms al heel goed, legt Wevers uit, omdat de algoritmes getraind zijn op specifieke datasets voor specifieke taken. "Zelfrijdende auto's zijn bijvoorbeeld heel goed in het herkennen van verkeersborden. Maar veel objecten verschillen door de tijd heen te veel: een telefoon ziet er anders uit dan tien jaar geleden. Bovendien bevat



Het labelen van afbeeldingen gebeurt deels handmatig. Ook wordt gebruikgemaakt van labels die eerder zijn toegevoegd. In samenspraak met het Noord-Hollands Archief komt Wevers zo tot een lijst met labels.

Credits: Noord-Hollands Archief

## 2 miljoen foto's online doorzoekbaar

# Met AI zoeken in historisch fotomateriaal

ten veel datasets geen historisch materiaal. "En dan hebben we het nog niet over de kwaliteit van de afbeelding of de techniek. Veel oude foto's zijn in grijstinten; een sneeuwlandschap en een strand zijn moeilijk van elkaar te onderscheiden. Mogelijke oplossingen zijn om met AI oude afbeeldingen automatisch te laten inkleuren, of bestaand trainingmateriaal juist zwart-wit te maken."

Toch werkt Wevers met bestaande algoritmes voor een eerste indicatie.

"Een bestaand model is op heel veel afbeeldingen getraind. Sportwedstrijden worden bijvoorbeeld makkelijk herkend. Maar veel van die afbeeldingen komen van websites als Flickr, en hebben een sterke Amerikaanse bias." Het labelen van de afbeeldingen gebeurt daarom deels handmatig. Ook maakt de onderzoeker gebruik van labels die eerder zijn toegevoegd door de fotografen van persbureau De Boer. In samenspraak met het Noord-Hollands Archief komt Wevers zo tot een lijst met labels. Dat wordt de in-

put voor het Vele Handen-project: mensen kunnen straks kiezen uit de labels die het computeralgoritme genereert.

### Naar 90 procent

Het algoritme heeft nu een correcte score van 85 procent, Wevers streeft naar 90 procent. "Ik hoop uiteindelijk een algoritme te ontwikkelen dat niet alleen toepasbaar is op deze collectie, maar ook op andere collecties, zoals die van het Nationaal Archief."

velehanden.nl

## GELEZEN

### Recommendations for Services in a FAIR data ecosystem

D. Bangert, E. Hermans, R. van Horik, M. de Jong, H. Koers, M. Mokrane

FAIR data ontstaan en worden gebruikt in een complex wetenschappelijk landschap (FAIR ecosysteem); het omvat ICT-diensten, data stewardship en samenwerking binnen en buiten instellingen en disciplines. Daarbij is de vraag: hoe FAIR-aligned zijn de diensten? Deze vraag hebben de Europese initiatieven FAIRsFAIR, RDA Europe, OpenAIRE, EOSC-hub en FREYA gesteld in een reeks workshops. In het heldere eindrapport *Recommendations for services in a FAIR data ecosystem* springen vier aanbevelingen eruit: gebruik services voor persistent identifiers (ook voor organisaties en instrumenten); op standaarden gebaseerde metadata worden idealiter automatisch gegenereerd; maak FAIR-alignment en het delen van data onderdeel van onderzoeksevaluaties; werk internationaal samen aan de uitdagingen voor het uitvoeren van de FAIR-principes. Ook zijn een aantal aanbevelingen acties verbonden: bibliotheken willen de waardering voor FAIR-alignment en FAIR data aanwakkeren, onderzoeksinstellingen zetten in op data stewardship en service providers op domeinspecifieke ontologieën en tools voor repositories. "Work in progress", aldus de auteurs, en belangrijk om voort te zetten in beleid, infrastructuur en onderzoekspraktijk.

DOI: 10.5281/zenodo.3585742

23:30 - 07:00

Slapen, met partner

07:00 - 07:10

Snoozen, alleen

07:10 - 07:20

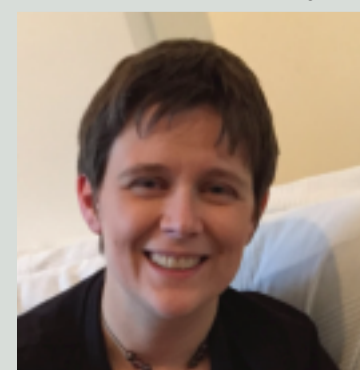
Douchen, alleen; nevenactiviteit: zingen.

Voor het tijdsbestedingsonderzoek (TBO) dat het CBS samen met het SCP eens per 5 jaar uitvoert, vullen respondenten naast een algemene vragenlijst ook 7 dagen lang een papieren dagboek in. Per 10 minuten geven zij aan welke activiteit zij hebben gedaan, of zij daarnaast nog iets deden en met wie ze dat deden. Hoe mensen hun tijd besteden, zegt iets over hun interesses, mogelijkheden en beperkingen. Het TBO geeft daarmee ook een mooi zicht op maatschappelijke verschijnselen en hoe die veranderen in de tijd. Zo komt het, vergeleken met tien jaar eerder, bijvoorbeeld vaker voor dat mensen meerdere activiteiten op hetzelfde moment uitvoeren, het 'multitasken'. SCP heeft een mooi overzicht gemaakt in de vorm van

een 'card stack', een verzameling digitale kaarten die ingaan op een aspect van de tijdsbesteding van Nederlanders. We zijn onze respondenten dankbaar dat ze de tijd (!) nemen om een hele week hun activiteiten bij te houden, want dat is heel veel werk.

Maar ook het verwerken van al die papieren boekjes op dusdanig wijze dat de privacy wordt geborgd, kost de nodige uren: inscannen, coderen, opschonen, het maken van een databestand bruikbaar voor onderzoekers. Redenen om samen met het SCP een app te ontwikkelen die digitaal veilig is, en die het de respondent gemakkelijker maakt zijn of haar tijdsbesteding bij te houden. Eerder voerde het SCP ook al experimenten met CentERdata op

dit terrein uit. Het bijkomend voordeel van een app is dat het de kwaliteit van de invoer verbetert (je kunt controleregels inbouwen als tijdslots worden vergeten) en de verwerking vereenvoudigt (je hebt immers alles al digitaal binnen). Het CBS is al aan het experimenteren met een app die verplaatsingen meet voor het onderzoek Onderweg in Nederland, en ook wordt er gewerkt aan een app die het inscannen van bonnetjes



van je boodschappen mogelijk maakt, voor het Budgetonderzoek. Beide zijn onderzoeken die met slimme apps het voor respondenten minder tijdrovend kunnen maken en zo hopelijk de respons op het onderzoek kunnen verhogen.

Natuurlijk zijn er mitsen en maren, zoals: 'Willen mensen wel een app installeren?' of 'Is het niet een hele selectieve groep die meedoet?'. Om dit soort vragen te beantwoorden, wordt de app eerst getest in een veldtest. Voor het TBO hopen we dit eind dit jaar te kunnen doen. Als de test positief is, wordt de app verder ontwikkeld. In eerste instantie voor het TBO, en met wat aanpassingen ook voor andere onderzoeken die tijd meten, zoals onderzoeken naar vrije tijd en naar het

gebruik van social media. Er zijn genoeg mogelijkheden, bijvoorbeeld door slim gebruik te maken van de sensoren die standaard in deze apparaten aanwezig zijn. Een walhalla voor een onderzoeker, uiteraard met een uiterst scherp oog voor privacy, doelmatigheid en informatiebeveiliging.

22:50 - 23:00

Tandenpoetsen, alleen; nevenactiviteit: whatsappen

23:00 - 23:20

Nog even lezen, alleen

23:20 - 07:00

Slapen, met partner

### Marieke Houben-Van Hertem

Marieke is projectmanager/statistisch onderzoeker bij het CBS. Als clustercoördinator *Leefsituatie* heeft zij zicht op alle persoons-enquêtes die het CBS doet op het gebied van leefsituatie, sociale samenhang en welzijn. Daarnaast werkt Marieke mee aan de gezondheidsstatistiek van het CBS. Marieke geeft de pen door aan Karin Hagoort.

## COLUMN

## Tijd voor een nieuw tijdsbestedingsonderzoek