

Zoeken naar beeld en geluid

De *semantic gap* bedwongen?

Anders dan bij tekst, zeggen de bits en bytes waarmee beeld en geluid digitaal gerepresenteerd worden, niets over de betekenis ervan. Deze kloof tussen digitale representatie en betekenis wordt wel de ‘semantic gap’ genoemd. Die maakt dat zoeken naar beeld en geluid anders gaat dan zoeken naar tekstmateriaal. Dat een tekstueel zoekwoord toch relevant beeld of geluid kan opleveren is daardoor minder vanzelfsprekend dan bij het zoeken binnen teksten. Recente *machine learning*-technieken spelen intussen een belangrijke rol bij pogingen om die *semantic gap* te overbruggen.

Tekst **Eric Sieverts**

Beeldmateriaal

Bij digitaal beeldmateriaal kun je denken aan afbeeldingen die ter illustratie op webpagina’s staan of aan de inhoud van collecties met afbeeldingen. Afbeeldingen op webpagina’s zijn losse bestanden. Voor het zoeken daarnaar hebben zoekmachines als Google, Bing en Yahoo speciale ‘image’-versies. Die maken primair gebruik van gewone tekst-*search*. Daarvoor beschikbare tekst is:

- de bestandsnaam van de imagefile;
- tekst die binnen een afbeelding zelf voorkomt en door OCR voor de computer leesbaar gemaakt is;
- tekst die in een webpagina in de omgeving van een afbeelding voorkomt;
- semantische metadata die in de HTML-codering van webpagina’s verwerkt is.

Deze soorten tekst bieden geen garantie dat er een inhoudelijke relatie is met wat de afbeelding voorstelt. Zeker bij

webpagina’s staat tekst soms min of meer toevallig in de buurt van een bepaald plaatje. Bij toegevoegde metadata is een inhoudelijke relatie waarschijnlijker, zeker als de ontologie van schema.org wordt gebruikt om ook plaatjes in webpagina’s te karakteriseren. Daarnaast zijn zoekmachines als Google, dankzij *machine learning*-technieken, steeds beter in staat om in te schatten wat is afgebeeld op beeldmateriaal en of dat voldoende met de zoekvraag overeenkomt. Die kennis wordt dan bij de relevantiebepaling (de plaats op de ranglijst) van zoekresultaten betrokken, waardoor de eerste schermen met gevonden plaatjes niet meer zo vaak onbedoelde resultaten tonen.

Filteren

Vaak kun je de zoekresultaten ook filteren op formele kenmerken. Sommige, zoals afmetingen, kleur of zwart-wit en bestandsformaat, kan een zoekmachine direct uit de eigenschappen van het beeldbestand afleiden. Slimme software kan meer kenmerken bepalen, zoals de belangrijkste voorkomende kleuren, of het een foto of een tekening is en of er gezichten op een foto voorkomen. Deze eenvoudige vorm van zogenoemde ‘content based image retrieval’, wordt onder meer in Google Image toegepast. Digitale camera’s slaan bovendien technische gegevens van gemaakte foto’s in het imagebestand op, in de EXIF-metadata. In sommige gevallen zijn dat ook geografische coördinaten van de plek waarop de foto is genomen. Die gegevens kunnen soms getoond worden, maar worden vrij zelden in het zoekproces betrokken.

Semantische conceptdetectie

De meest geavanceerde vorm van *content based image retrieval* maakt gebruik van ‘semantische conceptdetectie’. Daarbij leert de computer, op basis van veel voorbeelden,

per individueel concept, hoe afbeeldingen daarvan te herkennen zijn. Google gebruikt dit om haar beeldzoeker te verbeteren en ook om daar bewaarde foto's automatisch te sorteren. Een belangrijk element daarbij is het herkennen van een veelheid aan karakteristieke 'landmarks' die vaak op foto's voorkomen (maar van die bezienswaardigheden zijn toch al veel foto's te vinden).

Een variant van 'content based image retrieval' is 'reverse image search'. Daarbij bestaat de zoekvraag uit een bestaande afbeelding, van internet of van je eigen computer, die als zoekvoorbeeld dient. Bij Google Image geeft deze zoektechniek soms nauwelijks gelijkende afbeeldingen; alleen de kleuren komen dan min of meer overeen. Bij Bing en de Chinese zoekmachine Baidu is dat vaak veel beter. TinEye is een hierin gespecialiseerde zoekstelsel dat alleen zeer sterk gelijkende afbeeldingen vindt, zodat je daarmee kunt controleren of een bepaalde afbeelding vaker gebruikt of bewerkt is. Speciaal voor *fact-checking* kan dat nuttig zijn. Voor dit doel bestaan ook browser-extensies, zoals RevEye, die via een rechtermuisklik op een afbeelding keuze bieden uit verschillende systemen om daarop *reverse image search* te doen.

Het aantal foto's dat gebruikers uploaden is inmiddels gigantisch. Van enkele miljoenen per dag op Flickr of Twitter, tot honderden miljoenen op Instagram, Facebook en Google. De aard van het materiaal op deze sites kan sterk uiteenlopen, van kiekjes op Facebook tot kunstzinnige producten op Flickr. Voor het zoeken ben je afhankelijk van de tekst die bij het uploaden is toegevoegd: tags, korte beschrijvingen of berichten, zoals bij Twitter. Maar weinig hiervan kun je met algemene zoekmachines vinden, ook al zijn foto's vrij toegankelijk. Helaas zijn de eigen zoeksystemen van deze systemen vaak nogal slecht. Uitzondering daarop vormt Flickr. Aan daar geüpload foto's kent Flickr bovendien automatisch inhoudelijke tags toe.

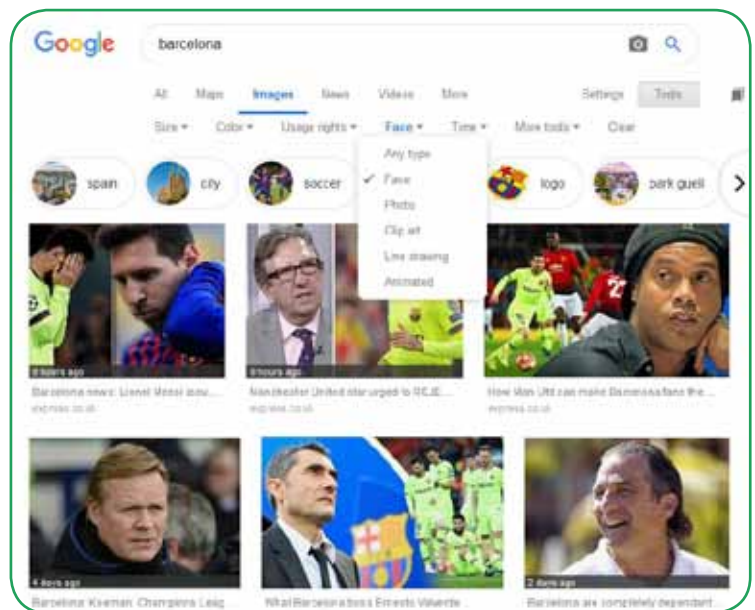
Voor geografisch zoeken kun je gebruikmaken van Google Earth, waarop mensen geüpload foto's aan een locatie kunnen koppelen. Anderzijds kan natuurlijk ook Google Streetview dienen als bron van gelokaliseerd beeldmateriaal.

Grote fotocollecties

Diverse organisaties met grote fotocollecties stellen die via internet beschikbaar. Voorbeelden zijn het Nationaal Archief (ook op Flickr) en Wikimedia Commons (de afbeeldingencollectie die ter illustratie van Wikipedia is aangelegd en ook elders gebruikt mag worden). Aan die laatste wordt ook digitaal materiaal bijgedragen door musea, archieven en bibliotheken. Voor vrij hergebruik van beeldmateriaal is er verder een algemene Creative Commons-zoekmachine (<https://ccsearch.creativecommons.org/>) die meerdere collecties doorzoekt. Je kunt daarin filteren op de CC-licenties waaronder afbeeldingen beschikbaar zijn. De oude versie hiervan (<https://oldsearch.creativecommons.org/>) biedt *metasearch* op nog andere systemen, waaronder Flickr.

Geluidsmateriaal

Voor het zoeken en vinden van geluidsmateriaal gelden vergelijkbare factoren als voor beeld, zij het dat gewone web-





zoekmachines geen mogelijkheden bieden specifiek naar audio te zoeken. Hoogstens kun je zoeken naar in URL's voorkomende audio-extensies zoals .mp3, in combinatie met inhoudelijke zoekwoorden. Je hoopt daarmee dat zo gevonden webpagina's geluidsbestanden bevatten. Er zijn wel een paar geluidsdatabanken. Soungle bevat een variëteit aan korte geluidsfragmenten die met trefwoorden beschre-

ven zijn. Findsounds is een soortgelijk systeem dat geluidsfragmenten uit webpagina's haalt.

Er bestaat ook 'content based sound recognition', dat geluiden automatisch herkent. Meestal is dat alleen voor specifieke categorieën, zoals de app Shazam, die gebruikers kan vertellen naar welk muziknummer ze luisteren, of de app Birdnet die vogelzang kan determineren. Maar die technieken worden nog nauwelijks voor zoektoepassingen gebruikt. Voor spraak ligt dit anders. Automatische spraakherkenning wordt steeds beter. Hoewel zij vooral wordt toegepast voor spraakaansturing van software en in *enterprise search*, zijn er nu ook toepassingen om op internet in gesproken tekst te zoeken. Een voorbeeld is de zoekmachine Voxlead voor het full-text zoeken in nieuwsprogramma's op internet, zowel in puur audio- als in videomateriaal. Voordeel is dat dit zich niet beperkt tot Engelstalige spraakherkenning, maar dat Voxlead ook acht andere talen kent, waaronder Nederlands. Nadeel is dat het vooral een demosysteem is, met maar een beperkte hoeveelheid en niet heel recent materiaal. Andere voorbeelden zijn de zoekmachine van BNR Nieuwsradio (<https://www.bnr.nl/search>) en de vooral Engelstalige systemen Audioburst en FluidData, waarin ook podcasts doorzoekbaar gemaakt worden.

Daarnaast wordt gesproken tekst ook wel handmatig omgezet, zoals bij de TED-lezingen en bij ondertiteling van



film en tv. Jammer genoeg wordt dat nog vrijwel niet gebruikt om in te zoeken.

Video

De bekendste bronnen van videomateriaal zijn die waar je zelf materiaal kunt uploaden, en dan vooral YouTube. Naar YouTube wordt circa vijfhonderd uur video per minuut geupload. YouTube-video's kunnen onder meer nuttig zijn voor praktische uitleg hoe bepaalde reparaties, installaties en dergelijke uit te voeren. Andere bronnen voor uploads van vooral serieus videomateriaal zijn onder meer Vimeo en de fotosite Flickr.

Hoewel upload-sites ook wel materiaal van tv-programma's bevatten, zitten nieuws en documentaires systematischer in de videoversies van de zoekmachines Bing, Google en Yahoo. Het is niet altijd duidelijk of daarbij ook in beschikbare ondertiteling of eventueel aanwezige transcripties gezocht wordt.

Soms bestaat videomateriaal uit opeenvolgingen van afzonderlijke, al dan niet samenhangende scènes, die eigenlijk afzonderlijke zoek-ingangen behoeven. Het eerder genoemde Voxlead is een van de weinige systemen dat daarmee overweg kan, omdat het via spraakherkenning full-text zoekmogelijkheden biedt. Daardoor kan het meteen naar specifieke scènes springen waarin de zoekwoorden worden uitgesproken, zonder dat je de hele voorgaande video hoeft te bekijken.



Een archief van Nederlands omroepmateriaal is te vinden bij Beeld en Geluid (<https://zoeken.beeldengeluid.nl/>), maar hun zoekstelsel biedt geen links naar de video's zelf. Een beperkt deel daarvan wordt wel aangeboden op Openbeelden.nl.

Voor *factchecking*-doeleinden bestaat voor video niet zo'n eenvoudige verificatiemethode als *reverse image search*. Amnesty heeft daarom een speciale YouTube DataViewer ontwikkeld waarmee iets dergelijks wel mogelijk is (<https://citizenevidence.amnestyusa.org/>). ●

