Roeland Ordelman

Europese projecten (2)

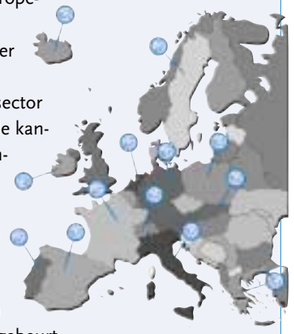
Zoeken in historisch videomateriaal

De expertise van de Parlevink Groep van de Universiteit Twente op het gebied van Spoken Document Retrieval wordt ingezet voor het ontsluiten van de Nederlandstalige filmarchieven bij het Nederlands Audiovisueel Archief. Dit in het kader van ECHO, een Europees project ter ontwikkeling van een digitale bibliotheekservice voor historische films van grote nationale audiovisuele archieven.

Stel, een journalist bij een televisieomroep maakt een programma over studenten in de jaren zestig. Natuurlijk wil die journalist dan wat shots van provo's, witte fietsen, nozems of andere typische gebeurtenissen die in deze periode op film werden vastgelegd. Het vinden van deze stukjes film uit het archief betekende tot voor kort urenlang stapels videobanden doorworstelen. Nieuwe technieken maken het mogelijk om vanaf de eigen werkplek per computer te zoeken naar specifieke fragmenten in videoarchieven. De journalist stelt een vraag aan een multimediazoeksysteem, bekijkt de videofragmenten die het systeem oplevert en knipt en plakt bij wijze van spreken zo het stukje televisie in elkaar. Niet alleen de journalist, ook een gebruiker thuis zou op deze manier natuurlijk graag in een videoarchief willen grasduinen. Het Europese project ECHO (European CHronicles On-line) levert een belangrijke bijdrage aan het verwezenlijken hiervan.

Europa!

Na het inleidend artikel van Peter van Poortvliet in Informatie Professional nr 11, 2000, vervolgen we de serie 'Europese projecten' met de beschrijving van het project ECHO.
Door middel van een groot aantal subsidies stimuleert de Europese Unie innovatie en onderzoek binnen Europa. Voor de periode 1998-2002 is hiervoor een budget van ongeveer 28 miljard gulden beschikbaar.
In opdracht van het ministerie van Economische Zaken heeft Bureau EG-Liaison in Den Haag onlangs onderzoek gedaan naar het rendement van de Europese technologiesubsidies voor Nederland. Daaruit bleek onder andere dat Nederlandse instellingen uit de informatiesector niet optimaal profiteren van de kansen die deze subsidieprogramma's bieden. De komende maanden wijdt Informatie Professional daarom een aantal artikelen aan lopende en recent afgesloten projecten, om zo een beeld te geven van wat er op het Europese front gebeurt.



Audiovisuele geschiedenis

ECHO wordt gefinancierd in het vijfde kaderprogramma van de Europese Commissie en moet de enorme hoeveelheden historisch filmmateriaal, die zich in de televisiearchieven van Nederland, Italië, Frankrijk, Duitsland en Zwitserland bevinden, makkelijker toegankelijk maken voor de Europese burger. Gezien de enorme hoeveelheden materiaal, plus het feit dat het om filmmateriaal gaat (audio en video in plaats van alleen tekst), moeten er echter nieuwe technieken ontwikkeld worden om dat mogelijk te maken. Voor het Nederlandse taalgebied wordt het onderzoek naar deze nieuwe technieken gedaan door de Parlevink Groep van de Universiteit Twente.

Het Nederlands Audiovisueel Archief (NAA) beheert de Nederlandse archieven die binnen ECHO worden gebruikt. Zo heeft het NAA het hele Polygoon Journaal, met de Stem des Vaderlands *Phillip Bloemendaal,* 'in de kast staan'. Maar ook documentaires over de meest uiteenlopende onderwerpen uit de Nederlandse samenleving van 1920 tot op heden liggen op geïnteresseerden te wachten. *Annemieke de Jong,* beleidsmedewerker Informatisering en Digitalisering van NAA en namens NAA projectleider van ECHO: 'De beeld- en geluidscollecties die de deelnemende archieven uit Nederland, Frankrijk, Italië en Zwitserland gezamenlijk beheren, vertegenwoordigen een zeer belangrijk deel van de Europese "audiovisuele" geschiedenis. In het digitale ECHO-videoarchief komt dat straks allemaal samen en kan het door een breed – professioneel en algemeen – publiek worden geraadpleegd en hergebruikt.' Het bijzondere van de ECHO-collectie zal zijn, dat kan worden bekeken en vergeleken hoe twintigste-eeuwse gebeurtenissen werden gefilmd vanuit verschillende nationale perspectieven. 'Hierbij streven we niet naar een archief met alleen beelden van de belangrijke politieke en maatschappelijke zaken,' merkt De Jong op. 'We willen ook opnamen bieden van het dagelijks leven in de verschillende Europese landen.'

Om het filmmateriaal toegankelijk te maken moet echter nog wel het nodige gebeuren. Op dit moment zijn de

Foto: Nederlands Audiovisueel archief Hilversum

Filmploeg Polygoonjournaal 1930

ECHO (European Chronicles Online)

Looptijd: februari 2000 tot september 2002.
Doelstelling: het ontwikkelen van een digitale bibliotheekservice voor historische films van grote nationale audiovisuele archieven. Het project zal een open-architectuurbenadering volgen zodat nieuwe diensten eenvoudig kunnen worden toegevoegd.

Deelnemers:
- CNR-IEI (Italië)
- ITC-irst (Italië)
- Instituto Luce (Italië)
- INA (Frankrijk)
- CNRS/LIMSI (Frankrijk)
- Tecmath (Duitsland)
- Universiteit van Mannheim (Duitsland)
- Carnegie Mellon University (US)
- Memoriav (Zwitserland)
- Eurospider Information Technology (Zwitserland)
- Universiteit Twente

archieven ontsloten in de vorm van een catalogusbeschrijving. De informatie die over de video's beschikbaar is, is vaak niet erg gedetailleerd. Titel, namen van de makers en uitzenddatum zijn natuurlijk wel beschikbaar, maar het blijft toch raden waar het precies over gaat in elk stukje video. De enige mogelijkheid om meer informatie over het materiaal te verkrijgen, is simpelweg documentalisten inschakelen en die naar het materiaal laten kijken en luisteren. Maar dat kost natuurlijk veel tijd en geld. Het zou een uitkomst zijn als dit kijken en luisteren geautomatiseerd zou kunnen worden en als de aldus opgedane informatie zou kunnen worden opgeslagen om in te zoeken.

Kijken en vooral... luisteren

Dat kijken en luisteren *kan* geautomatiseerd worden. 'Luisteren' kan immers gesimuleerd worden met behulp van automatische spraakherkenning, 'kijken' met behulp van beeldherkenning. Helaas is het onderzoek naar beeldherkenning nog lang niet zover, dat bijvoorbeeld een hippie in een video zou kunnen worden 'herkend'. Beeldherkenningsonderzoekers zijn al heel blij als ze een mens in een video kunnen herkennen. Hoewel in de toekomst beeldherkenning meer en meer gebruikt zal gaan worden om automatisch beeldinformatie uit video's te halen, staat het onderzoek nog te veel in de kinderschoenen om nu al in te zetten binnen een project als ECHO. Zie de lijst met websites over beeldherkenning om een beeld te krijgen van het onderzoek in binnen- en buitenland.

Automatisch luisteren met behulp van spraakherkenning kunnen we echter al een stuk beter. Goed genoeg in ieder geval om succesvol in te kunnen zetten bij een project als ECHO. Aangezien in 'het gesproken woord' doorgaans genoeg informatie zit om erachter te komen waar het in een stukje video nu precies over gaat, heeft spraakherkenning zich bewezen als een belangrijk hulpmiddel bij het zoeken in audiovisuele archieven. Of beter gezegd, als een onmisbaar gereedschap bij het creëren van databases met gegevens ('indexing') over de video, waarin vervolgens gezocht ('retrieval') kan worden. Deze vorm van Information Retrieval wordt 'Spoken Document Retrieval' (SDR) genoemd.

Spraakherkenning

De Parlevink Groep van de Universiteit Twente heeft de afgelopen jaren veel expertise opgebouwd op het gebied van Information Retrieval (IR). Binnen een aantal grote projecten (Twenty-One, Pop-Eye, Olive, DRUID) is

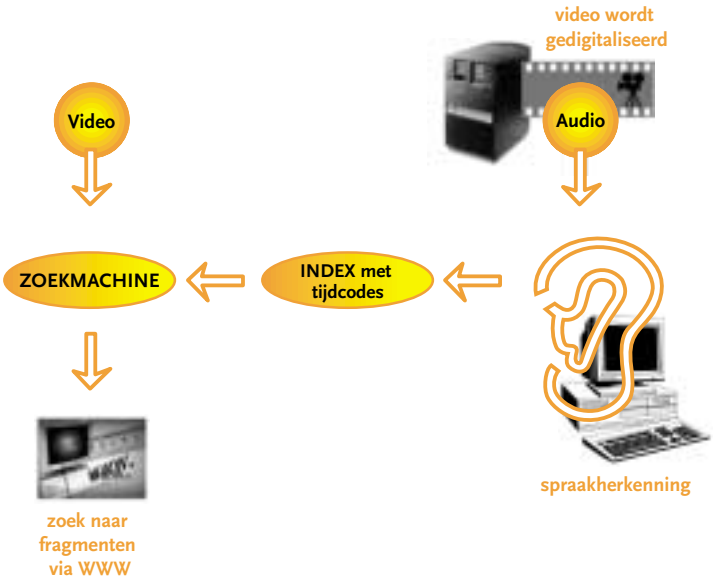
```
<sentence id="2" starttim="time=0.1280" endtime=
"time=2.0480">
    <words>
        <word time="0.1280" >
        gaan
        </word>
        <word time="0.2240">
        witte
        </word>
        <word time="0.7680">
        fietsen
        </word>
        <word time="0.9760">
        op
        </word>
        <word time="1.2320">
        straat
        </word>
    </words>
    <sentence_info>
        <totamprob>13.931030</totamprob>
        <totlmprob>4.803101</totlmprob>
        <totpathprob>4.803101</totpathprob>
        <text>gaan witte fietsen op straat/text>
    </sentence_info>
</sentence>
```

Voorbeeld 1 XML-output van de herkenner. De uitgesproken zin was 'Er staan witte fietsen op straat', 'gaan witte fietsen op straat' werd herkend.

gewerkt aan diverse IR-onderwerpen en er wordt elk jaar meegedaan met de belangrijke internationale Text Retrieval Conferentie (TREC) in de Verenigde Staten. Binnen ECHO wordt vooral de expertise van Parlevink op het gebied van Spoken Document Retrieval ingezet voor het ontsluiten van de Nederlandstalige filmarchieven bij het NAA.

Het inzetten van spraakherkenning binnen Information Retrieval is relatief nieuw. In principe werkt het zo, dat de gesproken audio in een video door de herkenner wordt omgezet in tekst en opgeslagen in een tekstbestand. Belangrijk hierbij is dat de tijdcodes (wáár werd wat gezegd) die de herkenner levert, ook worden opgeslagen (zie voorbeeld 1). De tekst met de tijdcodes worden vervolgens geïndexeerd en de zoekmachine zoekt uiteindelijk in deze index. Zo kan bepaald worden in welk van de video's en ook wáár precies in elke video, gesproken werd over bijvoorbeeld 'witte fietsen'. In een speciale browser kunnen de zoekresultaten dan worden weergegeven. Afhankelijk van het soort browser kan de gebruiker de gevonden fragmenten bekijken, heen en weer spoelen en eventueel direct laden in een video-editingprogramma.

Het meest cruciale element in dit proces is natuurlijk de spraakherkenner, want hoe slechter de herkenner, hoe slechter de retrieval-resultaten in principe zullen zijn. In



Spoken Document Retrieval geldt, dat wanneer een woord fout wordt herkend (bijvoorbeeld 'vuurwerk op de Dam' wordt herkend als 'uurwerk op de Dam'), dit eigenlijk twee fouten oplevert: het correcte woord kan niet worden teruggevonden ('vuurwerk') en er kan een woord wel worden gevonden dat helemaal niet voorkwam ('uurwerk'). Goede herkenning is dus vooral belangrijk voor woorden die van belang zijn in een zoekproces, de inhoudswoorden! Dit is een belangrijk gegeven. Want dit betekent dat de herkenner niet moet worden gezien als een dicteermachine die elk woord dat werd uitgesproken correct moet kunnen herkennen. Hij mag best fouten maken, als hij de inhoudswoorden maar goed herkent. Een bijkomend voordeel is, dat inhoudswoorden over het algemeen langer zijn dan functiewoorden en daarom in principe makkelijker te herkennen zijn. Door de bank genomen is daarom een herkenningsgraad van 65 procent (één op de drie woorden wordt fout herkend) wel voldoende voor goede retrieval-

Websites

- De officiële ECHO-website: http://pcerato2.iei.pi.cnr.it/echo/
- De website van de Parlevink Groep: http://wwwseti.cs.utwente.nl/Parlevink/
- Website van het NAA: http://www.naa.nl
- Dit is een voorbeeld van een Amerikaanse SDR Search-engine die is ontwikkeld door Carnegie Mellon University, een van de partners binnen ECHO: http://search.mediasite.net/
- DRUID-website: http://dis.tpd.tno.nl/druid/
- TREC-website: http://trec.nist.gov/

Beeldherkenning

- Een voorbeeld van Beeld Retrieval, hier wordt het QBIC-systeem gebruikt: www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English
- Een ander voorbeeld: het virage systeem wordt gebruikt bij AltaVista's image search: http://www.altavista.com/cgi-bin/query?pg=q&stype=simage
- Beeldherkenning op de Universiteit van Amsterdam: http://zomax.wins.uva.nl:5345/zomax/HTML/zomax.html
- Leuke voorbeelden zijn photobook en four eyes van het MIT media lab: http://vismod.www.media.mit.edu/vismod/demos/photobook/ en http://vismod.www.media.mit.edu/~tpminka/photobook/foureyes/ Bij beide is geen demo aanwezig, maar wel goede uitleg met veel voorbeelden. Photobook gaat over texture retrieval en gezichtsherkenning. FourEyes segmenteert en annoteert delen van plaatjes (lucht, gras, water, gebouw).

resultaten. Maar een herkenningsgraad van 65 procent halen is al moeilijk genoeg!

Audiokwaliteit

In 1999 begon Parlevink, in samenwerking met TNO Technische Menskunde in Soesterberg, in het kader van het DRUID-project met de ontwikkeling van een spraakherkenner voor het Nederlands die een groot aantal woorden (65.000) in continue spraak onafhankelijk van de spreker moet kunnen herkennen. Zo'n herkenner was er op dat moment nog niet in Nederland. Commerciële Nederlandse systemen die beschikbaar waren, konden óf alleen overweg met een beperkt aantal woorden (bijvoorbeeld maximaal 5000) óf waren sprekerafhankelijk (de herkenner moet dan speciaal op één spreker getraind worden). Het te ontwikkelen systeem moest in staat zijn om alle Nederlandse spraak in bijvoorbeeld het NOS journaal in tekst om te zetten; alles wat de nieuwslezer zegt, alles wat de weerman of -vrouw zegt, en alles wat de verslaggevers en geïnterviewden op locatie zeggen. Dat is een hele klus voor een herkenner, want hij krijgt dus te maken met goed opgenomen studiospraak van de nieuwslezer, maar ook met de veel lagere audiokwaliteit van de verslaggevers op locatie (achtergrondlawaai, telefoonverbinding). Ook slordig of met een accent sprekende geïnterviewden maken het de herkenner lastig. Binnen het ECHO-project komen daar nog extra moeilijkheden bij. Omdat het hier kan gaan om *historisch materiaal*, is het geluid van de opnamen soms extra slecht en maakt de ouderwetse manier van spreken het de herkenner ook niet makkelijker.

Training

De afgelopen twee jaar is door de Parlevink Groep en TNO hard gewerkt aan het trainen van de herkenner met Nederlandse spraak- en tekstdata. Tot nog toe werd zo'n 50 uur aan audiodata en 60 miljoen woorden aan tekstdata verzameld en geschikt gemaakt voor het trainen. Het lijkt veel, maar dat valt eigenlijk best tegen. Voor elk type spraak (voorgelezen, spontaan, in de studio of op locatie, vrouwen- of mannenspraak) heb je 'voorbeeld' data nodig om het systeem mee te trainen en hoe meer data je hebt, des te beter je het systeem kunt trainen. Wanneer de herkenner alleen getraind is met mannenspraak, zal hij het dus slecht doen op vrouwenspraak.
Binnen DRUID en ECHO werd met alle beschikbare audiodata, één algemeen akoestisch model getraind dat elk type spraak redelijk kan herkennen. De herkenner heeft een herkenningsgraad van tussen de 70 procent (voor studiospraak) en 50 procent (voor spraak met een lage audiokwaliteit). Een andere mogelijkheid zou zijn om allemaal losse modellen te trainen, één voor studiospraak, één voor telefoonspraak, één voor vrouwen, één voor mannenspraak, etcetera. Elk apart getraind 'submodelletje' werkt dan heel goed op het type spraak waar het op is getraind. Door automatisch te detecteren wat voor type spraak er herkend moet gaan worden, kan de prestatie van het systeem een stuk verbeteren. Je moet dan wel voor elk type spraak genoeg trainingsdata hebben verzameld.

Tekstdata zijn nodig om taalmodellen te trainen. Een taalmodel probeert te voorspellen welk woord de meeste kans heeft om te volgen, gegeven het zojuist herkende woord of woorden. Wanneer is herkend 'op het station staan witte' en de herkenner moet kiezen tussen de woorden 'fietsen', 'nietsen', 'ietsje', 'niettemin' of 'auto's', dan zal het taalmodel moeten aangeven dat 'fietsen' en 'auto's' hier de meest waarschijnlijke opties zijn.
Om een goed taalmodel te kunnen trainen is heel veel tekst nodig, het liefst tekst die erg lijkt op dat wat herkend moet gaan worden. Zo wordt voor journaaluitzendingen krantenmateriaal, of liever nog, letterlijke transcripties van andere journaaluitzendingen, gebruikt. Voor het historisch materiaal dat binnen ECHO herkend moet gaan worden, is echter nauwelijks tekstmateriaal digitaal beschikbaar. Dat is jammer want de manier van spreken, het woordgebruik en de grammatica uit vroegere tijden kunnen nogal verschillen van die van tegenwoordig. Er zijn wel teksten op papier beschikbaar, maar het zou te lang duren om al die teksten met de hand in te voeren. Een poging om de teksten met behulp van Optical Character Recognition (OCR) te digitaliseren is mislukt: de teksten waren kopieën van doorslagen en te onduidelijk om succesvol te kunnen worden gescand.

Onderzoek

Voor ECHO worden het komend jaar de eerste herkenningsevaluaties uitgevoerd. Daarnaast wordt er een begin gemaakt met het integreren van alle losse componenten die door de verschillende partners zijn gebouwd in een groot ECHO-systeem. De Parlevink Groep zal naast het werk aan de herkenner verder onderzoek doen naar taalmodellering voor het Nederlands binnen Spoken Document Retrieval. Ook wordt in samenwerking met het IPO in Eindhoven onderzoek gedaan naar het gebruik van prosodische informatie (zoals pauze en intonatie) binnen SDR. Bijvoorbeeld voor automatische segmentatie (knippen van het audiosignaal in kleinere stukjes) en topicdetectie.

Roeland Ordelman is medewerker Onderzoek bij de Parlevink Groep van de Universiteit Twente. Specialisatie: Spraakherkenning & Information Retrieval.