DIXIT

TIJDSCHRIFT OVER TOEGEPASTE TAAL- EN SPRAAKTECHNOLOGIE

Eurospeech 2005

Ubiquitous Speech Processing

Intelligent Design

Over wetenschapdefinities en de zin van het leven

De stem van WFH ontrafeld

Audiovisuele archieven geïndexeerd

JAARGANG 3 NUMMER 3 OKTOBER 2005

Automatische indexering van audiovisuele archieven

De stem van Willem Frederik Hermans ontrafeld

‘Willem Frederik Hermans (1921-1995) wordt algemeen beschouwd als de belangrijkste Nederlandstalige schrijver van de twintigste eeuw. Zijn oeuvre is van een onovertroffen veelzijdigheid, geeft tot op de dag van vandaag aanleiding tot discussie en dwingt de nieuwe generaties lezers en auteurs steeds weer tot standpuntbepaling.’ Zo staat te lezen op de Willem Frederik Hermans web-portal van het gelijknamige instituut (WFHi) dat dit jaar, in het tiende sterfjaar van de schrijver, werd gelanceerd. De portal biedt een grote hoeveelheid informatie over Hermans en zijn schrijven en, uniek voor Nederland, ook de mogelijkheid om gedetailleerd te zoeken in Herman’s gesproken woord: in interviews, lezingen en voorgelezen verhalen.

| ROELAND ORDELMAN |

Audiovisuele opnameapparatuur en opslagruimte wordt steeds goedkoper en zowel het aantal als de omvang van audiovisuele archieven in Nederland neemt daardoor snel toe. Dankzij digitaliseringsinitiatieven bij instellingen en bedrijven die van oudsher beschikken over audiovisuele collecties (retrospectieve digitalisering) wordt het aanbod nog eens drastisch verbreed. Een aantal voorbeelden. Een aantal gemeenten in Nederland zijn gestart met het online brengen (‘webcasten’) van raadsvergaderingen (zie bijvoorbeeld http://www.bestuuronline.nl) in het kader van de ‘openbaarheid van bestuur’. Onderwijsinstellingen zijn aan het experimenteren met het audiovisueel archiveren van lessen, colleges, lezingen en conferentiepresentaties. Ook bedrijven nemen in toenemende mate vergaderingen en presentaties op ten behoeve van de interne informatievoorziening en natuurlijk kent iedereen die mensen die overal een videocamera naar toe zeulen om alles op te nemen wat ze niet willen vergeten of graag willen delen met anderen via weblogs.

STIEFKINDEREN VAN HET ARCHIEF

Een speciale categorie vormen instellingen zoals Beeld&Geluid (http://www.beeldengeluid.nl) die een van de grootste audiovisuele archieven in Europa aan het digitali-

seren is, het gemeentearchief van Rotterdam waar binnenkort een project start om haar gesproken-woord-archieven te ontsluiten, en instellingen zoals het eerdergenoemde WFHi. Deze categorie heeft in toenemende mate de belangstelling van historici omdat de betreffende archieven een belangrijk onderdeel vormen van het Nederlands cultureel erfgoed.
Het zijn allemaal prachtige en interessante collecties natuurlijk, maar zonder mogelijkheden om erin te zoeken verliezen ze veel van hun waarde, vooral wanneer de collecties omvangrijk zijn. Het is soms verbazingwekkend dat soms zelfs een minimale beschrijving van het materiaal ontbreekt. Archieven verworden daardoor tot een 'stiefkind van een archief', zoals Franciska de Jong, professor multimedia technologie aan de Universiteit Twente het onlangs uitdrukte, waar je weinig mee kunt.

Het is goed mogelijk om automatisch te detecteren waar er bijvoorbeeld spraak zit in een bestand en waar muziek

Op zichzelf is het niet zo vreemd dat collectiebeschrijvingen minimaal zijn, dan wel ontbreken. Het moet allemaal handmatig gebeuren en een hogere mate van detail vereist dat de betreffende audiovisuele bron van voor naar achter moet worden beluisterd of bekeken. Doorgaans zijn alleen instellingen met een professionele archivarissen hiertoe in staat. Maar dan nog, de zogenaamde 'meta-data' die voor met zorg aangelegde collecties beschikbaar is, beperkt zich meestal tot een titel, een datum en een korte beschrijving van de inhoud. Dit soort beschrijvingen kan gebruikt worden om *binnen een collectie* naar een specifiek bestand te zoeken, maar niet om binnen een ongestructureerd audiovisueel bestand zelf te zoeken. Om te weten te komen of een bestand misschien iets bevat waarnaar je op zoek bent, moet alsnog het hele bestand afgeluisterd of bekeken worden. Vooral als het bestand groot is, is dat vervelend. Bij voorbaat is de kans om op een potentieel interessant document uit te komen al niet zo hoog omdat de (specifieke) zoekvraag moet passen bij de in algemenere termen gestelde beschrijvingen.

AUTOMATISCHE INDEXERING

Om de geschetste problemen het hoofd te bieden, wordt er al een aantal jaren gekeken naar mogelijkheden om automatisch beschrijvingen toe te kennen aan collectiebestanden door gebruik te maken van audio- en videoanalysesoftware.
Het zoeken binnen een audiovisueel bestand zou al een stuk makkelijker gemaakt kunnen worden door het aanbrengen en zichtbaar maken van wat meer structuur. Het is goed mogelijk om automatisch te detecteren waar er bijvoorbeeld spraak zit in een bestand en waar muziek. Of waar een scène-overgang zit. De visualisatie hiervan in een zoekomgeving zou het zoeken al vereenvoudigen.

Een stap verder is dat individuele sprekers of zelfs het gesproken woord automatisch worden herkend en voorzien van een tijdlabel, zodat naar het voorkomen van sprekers of woorden binnen een bestand gezocht kan worden. Nog een stap verder is dat op basis van het herkende gesproken woord onderwerpen kunnen worden gedetecteerd. In het ideale geval ontstaat een goed gestructureerd audiovisueel document dat vanuit verschillende invalshoeken bevraagd kan worden.

Het gebruik van spraakherkenning voor het automatisch generen van meta-data, ofwel automatische indexering, ligt voor de hand. Geleid door professor Franciska de Jong doet de Human Media Interaction (HMI) groep van de Universiteit Twente al sinds de jaren '90 onderzoek naar het gebruik van spraakherkenning voor multimedia-ontsluiting. Een belangrijke stap was het ontwikkelen van een Nederlands spraakherkenningsysteem dat spraak één op één kan omzetten naar tekst. Een bestaand commercieel spraakherkenningssysteem van de plank halen was geen optie omdat het toepassingsgebied een flexibi-

liteit van de software vereist (aanpassen van akoestische modellen, vocabulaire, taalmodellen) die in commerciële systemen doorgaans niet of nauwelijks kunnen geven. Aanvankelijk richtte het onderzoek zich op het ontsluiten van Nederlandse nieuwsuitzendingen. De laatste jaren wordt steeds meer naar andere domeinen gekeken, zoals vergaderingen, lezingen, raadsvergaderingen en historische archieven.

Wanneer spraakherkenning wordt ingezet voor ontsluitingsdoeleinden is het niet zozeer van belang dat alle woorden in de spraak correct herkend worden, in tegenstelling tot wat bij dicteertaken het geval is. Internationaal onderzoek suggereert als minimale vereiste voor bruikbare herkenning voor ontsluitingsdoeleinden dat er zo'n 50% van de woorden correct moet zijn. Belangrijk is dat vooral inhoudswoorden en namen goed herkend worden. Dat zijn immers de woorden waarnaar gezocht gaat worden. De algemeen gangbare maat om de kwaliteit van een spraakherkenningsystemen mee uit te drukken, de 'word error rate' (WER), is dan ook niet erg veelzeggend als het er om gaat hoe bruikbaar een systeem is voor automatische indexering. Deze maat telt ook foutherkende functiewoorden of bijna-goedherkende woorden (enkelvoud in plaats van meervoud, fout in vervoeging, of een deel van een samenstelling fout) mee die voor het zoeken toch goed bruikbaar kunnen zijn.

HISTORISCH MATERIAAL

Met name bij historisch materiaal kan de 'word error rate' hoog zijn. De kwaliteit van de audio zelf kan daar debet aan zijn maar ook de manier van spreken ('gezwollen' taal) en het taalgebruik (ouderwetse woorden) dragen ertoe bij dat de spraakherkenning zich nogal eens verslikt. In spraakherkenningstermen is 'adaptatie' het sleutelwoord. De herkenner moet worden aangepast aan de akoestische karakteristieken en taal van het taakdomein met behulp van voorbeelddata: spraakdata waarbij handmatig is aangegeven welke woorden werden gezegd, en tekstdata die zoveel mogelijk lijken qua woord- en taalgebruik op de spraak in het taakdomein.

Voor het genereren van spraaktranscripties voor ontsluiten van de interviews en lezingen van Willem Frederik Hermans werd in eerste instantie een standaard spraakherkenningsconfiguratie gebruikt die normaal gesproken wordt toegepast voor het herkennen van nieuwsdata. Het was dan ook niet verwonderlijk dat zo'n 80% van de woorden verkeerd werd herkend. Dat is een resultaat dat zelfs voor zoekdoeleinden ongeschikt wordt geacht. Door de zeer kleine hoeveelheid voorbeelddata die voor dit domein beschikbaar was te gebruiken voor de adaptatie van de herkenner, werd uiteindelijk een 'word error rate' van rond de 65% gehaald. Hoewel dit nog ruim boven de eerdergenoemde kwaliteitseis voor zoekdoeleinden van 50% ligt, is het resultaat al best bevredigend. Wanneer je 'tranen der acacia's' intypt als zoekvraag, levert het zoeksysteem ook daadwerkelijk audiofragmenten op waarin de term voorkomt. Het is ook goed om te bedenken dat zonder deze vorm van automatische generatie van meta-data, zoeken in de bestanden al helemaal niet mogelijk was.

AUDIOVISUEEL MATERIAAL GOOGELEN

De Willem Frederik Hermans audio-zoekdemo laat zien dat door gebruik te maken van beschikbare Nederlandse spraaktechnologie het mogelijk is om in audiobestanden te zoeken. Dat er nog een hoop te verbeteren valt is duidelijk. Het verbeteren van de mogelijkheden om snel en doeltreffend een spraakherkenningssysteem aan te passen aan de eisen van een willekeurig taakdomein behoort hiertoe. Ook wat betreft zoekopties en de presentatie van de zoekresultaten kan nog het nodige werk worden verzet, bijvoorbeeld door automatisch gegenereerde segmentgrenzen (sprekerwisseling, pauzes, eventueel onderwepen) aan te bieden. Maar in ieder geval laat het zien dat de mogelijkheid om audiovisueel materiaal ook eenvoudigweg te googelen, langzaam dichterbij komt.

De Willem Frederik Hermans demo is te zien via http://www.willemfrederikhermans.nl (onder 'stem/beeld') of direct via http://wwwhome.cs.utwente.nl/~huijbreg/demopages/hermans/hermans.php. De demo is gemaakt door Marijn Huijbregts en Roeland Ordelman van de Human Media Interaction groep (http://hmi.ewi.utwente.nl) van de Universiteit Twente, in het kader van het MultimediaN project (http://www.multimedian.nl. door Contact: ordelman@ewi.utwente.nl of fdejong@ewi.utwente.nl