Het Radio Oranje Project

"Het was op zondagavond 28 juli 1940 dat voor het eerst in de aether de woorden weerklonken: ´Hier Radio-Oranje´, zulks ter inleiding van een uitzending, waarin Hare Majesteit Koningin Wilhelmina haar eerste grote toespraak zou houden tot haar verdrukte volk."

| PETER VD MAAS, WILLEMIJN HEEREN EN ARJAN VAN HESSEN |

Tijdens de oorlog werden er vanuit Londen radio-uitzendingen verzorgd voor de Nederlanders in het door Duitsers bezette Nederland onder de naam Radio Oranje. Naast allerlei informatie over de voortgang van de oorlog werden ook toespraken van hoogwaardigheidsbekleders uitgezonden. Voorbeelden hiervan zijn toespraken van Minister-president Gerbrandy, Prins Bernard en natuurlijk Koningin Wilhelmina. Omdat de toespraken van deze hoogwaardigheidsbekleders niet live voor de microfoon in de studio werden uitgesproken, maar vooraf werden opgenomen, is een deel van deze toespraken bewaard gebleven. De toespraken werden door de Engelse regering als "strategisch belangrijk" gezien en moesten vooraf aan de Engelse inlichtingendienst worden voorgelegd. Daardoor is ook de schriftelijke versie van de toespraken bewaard gebleven.

Het Radio Oranje Project, een samenwerking tussen de Universiteit Twente (oplijning), het NIOD (hosting en tekstdata) en Beeld en Geluid (geluidsbestanden), heeft als doel de schriftelijke en gesproken versie van de toespraken van Koningin Wilhelmina aan elkaar te koppelen zodat de geschreven tekst zo nauwkeurig mogelijk opgelijnd is met de gesproken tekst. Als deze oplijning eenmaal gedaan is, is het mogelijk om de toespraken van Wilhelmina te doorzoeken op woordniveau en de gevonden fragmenten direct af te luisteren. Bovendien is het mogelijk de tekst als ondertiteling te laten meelopen met het geluid. Hiervoor zal een web-interface worden ontwikkeld.

DATA

In totaal werden 32 toespraken opgelijnd, elk bestaande uit een gesproken en een geschreven document. De teksten, die in de oorlog in Londen op een typemachine waren uitgeschreven en waarvan soms hele regels met een rode stift door de Engelsen waren gecensureerd, werden na de oorlog in boekvorm gepubliceerd. Met behulp van OCR (Optical Character Recognition) werden deze pagina's omgezet in een digitaal bestand . Hoewel ook OCR fouten kan introduceren, bleek dat in minder dan 1% van de woorden het geval te zijn. De geluidsopnamen van de radiotoespraken zijn voor een groot gedeelte bewaard gebleven. Deze bestanden worden in het audioarchief van Nederlands instituut voor Beeld en Geluid gearchiveerd en werden recentelijk gedigitaliseerd.

OPLIJNEN

De tekst en het geluidsbestand van elke toespraak werden door de Twentse spraakherkenner opgelijnd. Aan de hand van de tekstuele versie bepaalt de computer hoe de zin moet klinken; de uitspraak van de woorden in de zin kan namelijk worden afgeleid met behulp van een "grafeem-naar-foneem" conversie. Daarin worden geschreven woorden

omgezet in de klanken waaruit ze zijn opgebouwd. Voor een aantal woorden dat in de oude spelling geschreven werd, zoals "mensch" of "landgenooten", klopte de grafeem-naar-foneem conversie niet en is een vertaling van oud naar nieuw Nederlands geïntroduceerd. Tijdens de oplijning werd de serie klanken op basis van de tekst gekoppeld aan de audio. Hiervoor is gebruik gemaakt van akoestische modellen die waren getraind op de stem van Koningin Wilhelmina.
Bij de oplijning van tekst met geluidssignaal geldt voor mooie, ruisvrije spraak dat de overeenkomst tussen de audio en de klankrepresentatie van de geschreven tekst zo groot is, dat het relatief eenvoudig is de woorden er precies onder te krijgen. Bij de spraak van Wilhelmina is dat anders. De opname-apparatuur van die tijd was primitief hetgeen resulteert in sterke ruis en tikken in de opname. Bovendien zijn de wasplaten waarop de spraak werd opgenomen in de loop der jaren verhard, waardoor de kwaliteit sterk achteruit gegaan is.
Het bleek niet mogelijk om een gemiddeld 10 minuten durende toespraak van Wilhelmina in één keer op te lijnen. De tekst werd daarom handmatig in fragmenten opgedeeld, en per fragment opgelijnd. Op deze manier was het begin van ieder fragment in ieder geval juist en nam de kans op fouten enorm af.

INFORMATIEBRONNEN TOEVOEGEN

Nu voor elke toespraak de audio en tekst zijn opgelijnd, willen we er automatisch extra informatie aan toevoegen. Deze extra informatie zal bestaan uit relevante foto´s die tijdens het afspelen van de audio getoond zullen worden en zo de ervaring van de luisteraar kunnen vergroten. De tekstuele inhoud van de zinnen wordt gebruikt om in het fotoarchief van het NIOD foto's te zoeken die qua omschrijving sterk op de zinnen lijken. Daarnaast wordt de uitzenddatum van de toespraken gebruikt om precies díe foto´s te vinden die uit dezelfde periode komen als de geluidsbestanden. Op deze manier worden er bij elke zin één of meerdere foto's gezocht die tijdens het afspelen van die zin zullen worden getoond.

CONCLUSIE

Hoewel oplijnen van oude en zeer ruizige teksten nog niet zó eenvoudig is, dat in één keer lange bestanden succesvol gedaan kunnen worden, is het resultaat na handmatige segmentatie toch zeer veelbelovend. Doordat er nu direct naar woorden en woordcombinaties in de audio gezocht kan worden, is het eenvoudig om dit soort audio-archieven via Internet voor iedereen toegankelijk te maken. In plaats van het afluisteren van talloze bestanden om net dat ene onderwerp te horen, kan er nu "op z'n Googles" gezocht worden en kan het resultaat eenvoudig via Internet worden afgespeeld.
De combinatie van beeld (de foto's), geluid (de radio-opnamen) en tekst (de transcripties) wordt op deze manier geheel automatisch gebundeld tot een multimedia-presentatie. Het NIOD is hier zo enthousiast over, dat ze besloten hebben de gehele presentatie binnen het thema "Wilhelmina in de oorlog" te hosten: vanaf volgend jaar is de presentatie dan via de website van het NIOD (www.niod.nl) te bekijken.

> "In plaats van het afluisteren van talloze bestanden, kan er nu 'op z'n Googles' gezocht worden."

PETER VD MAAS IS VERBONDEN AAN DE NIOD, WILLEMIJN HEEREN IS VERBONDEN AAN DE UNIVERSITEIT TWENTE / CHORAL PROJECT, ARJAN VAN HESSEN IS VERBONDEN AAN DE UNIVERSITEIT VAN TWENTE EN TELECATS