

CLARIAH Media Suite Workshop: Automatic Speech Recognition (ASR)

Alec Badenoch, Mary-Joy van der Deure en Jasmijn Van Gorp

Dit is een verkorte versie gemaakt in 2025 waarin vooral aandacht is voor het zoeken in ASR. De Engelstalige, uitgebreidere versie uit 2024 staat [hier](#).

Inleiding

Automatic Speech Recognition (ASR) is een AI-technologie die gebruikt wordt om spraak te herkennen en om te zetten in tekst. Binnen de Media Suite biedt ASR de mogelijkheid om televisie- en radioprogramma's te doorzoeken. De ASR biedt een 'tekstueel' toegangspunt voor de gesproken inhoud van audio- of audiovisuele programma's. Hierdoor wordt het zoeken in deze bestanden vergelijkbaar met zoeken in geprinte bronnen zoals boeken en kranten.

De spraakherkenning in de Media Suite is ingesteld op de Nederlandse taal. Ook zit er een klein aantal niet-Nederlandse termen in het vocabulaire van de spraakherkenning. De ASR is momenteel alleen toegepast op een gedeelte van de collectie van Beeld en Geluid, namelijk Radio 1, Radio 5, de 'Source catalogues' (items uit de radioprogrammering, Weken Nederlandse Radio en de Hoorspelen collectie) en een gedeelte van de televisie-uitzendingen (voornamelijk nieuws en actualiteitenprogramma's).

In deze oefening leer je werken met de ASR laag in de Media Suite. Het staat je in deze opdracht **vrij** om je **eigen zoekopdracht** te formuleren rondom het **Europese Song Festival**.

Stap 0: Voorbereiding

Voordat je met de collecties aan de slag kan, moet je eerst **inloggen** en is het verstandig om een '**User project**' aan te maken.

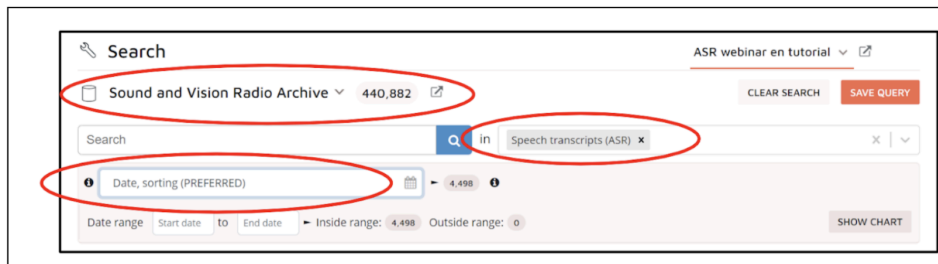
- Navigeer naar <https://mediasuite.clariah.nl/> en log in met je *Universiteit Utrecht* inloggegevens.

Maak ook een 'User Project' aan in de Workspace omgeving, om je zoekopdrachten en geselecteerde video's op te slaan.

Stap 1: Selecteer de Radio collectie en de ASR enrichment

Om te zoeken met ASR in de Media Suite zal je eerst **een collectie moeten selecteren** in de 'Search' omgeving. Het is belangrijk om na te gaan of de collectie relevant is voor je zoekopdracht.

- In deze opdracht zullen we ons focussen op discussies **over televisie, op de radio** (specifiek het Eurovisie Songfestival). We kiezen dus voor 'Sound and Vision Archive - Radio'.
- Om te zoeken in de ASR, selecteer je deze als 'search layer'. Ga naar het drop-down menu naast de zoekbalk, en selecteer hier '**Speech transcripts ASR**'. Je zoekt nu alleen in de spraak van de uitzendingen, niet meer in de overige metadata.
- Om een beeld te scheppen van je zoekresultaten is het verstandig je resultaten ook **per datum te sorteren**.
 - Ga hiervoor naar 'Select date field' onder de zoekbalk, selecteer '**Date, sorting (PREFERRED)**' en klik op 'Show chart'. Ondanks dat we nog geen zoekopdracht hebben ingevuld, krijgen we nu een visualisatie te zien van alle radioprogramma's die tot nu toe ASR hebben ontvangen. We hebben nu een beter beeld van de collectie waarin we gaan zoeken.



Stap 2: Het formuleren en verfijnen van een zoekopdracht

2.1 Formuleren van een zoekopdracht

Nu we een beeld hebben van de ASR-collectie kunnen we een zoekopdracht gaan formuleren,. Voor deze opdracht gaan we kijken naar een specifiek onderwerp: Het Eurovisie Songfestival. Het ESF heeft een relatief consistente naam gehad gedurende haar bestaan, hoewel er wel variaties zijn in de verschillende talen: In het Nederlands is het 'Eurovisie Songfestival', in het Engels 'Eurovision Song Contest' en in vele andere talen is het festival vaak simpelweg 'Eurovision' genoemd.

- Vanwege de verschillende taalvariaties zijn '**Eurovisie**' en '**Eurovision**' de meest logische zoekopdrachten om mee te beginnen. Neem een moment om deze verschillende variaties uit te proberen. Je kan ook **Eurovis*** gebruiken om programma's met beide schrijfwijzes te vinden.
- Combineer vervolgens Eurovisie met namen van Eurovision Songcontest deelnemers die je kent met landen, of met woorden uit liedjes. Bijvoorbeeld **Eurovisie AND Abba**, of **Eurovisie AND Australië**. Let wel: het taalmodel is getraind op Nederlands en ook de metadata zijn Nederlandstalig, dus gebruik voor landen de Nederlandstalige schrijfwijze.
- Kijk ook even wat je vindt als je **Eurovisie AND douze** invult. En vervolgens andere **cijfers in andere talen** dan het Nederlands. Kan je een reden bedenken waarom de Media Suite meerdere talen in de programma-beschrijvingen heeft als het cijfers betreft?

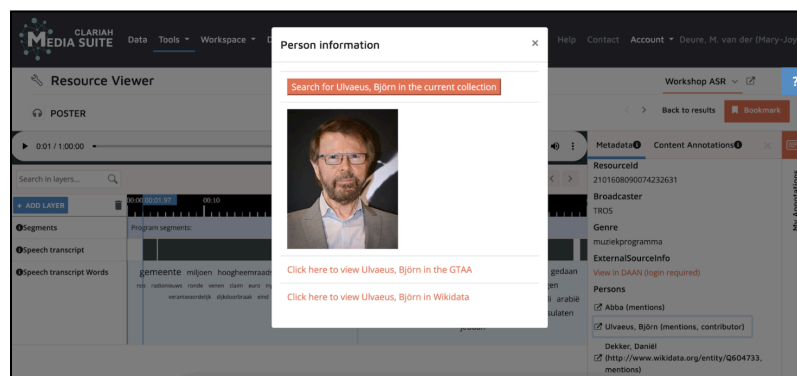
Antwoord:

- Voeg resultaten die je interessant vindt toe aan je User Project via de 'Bookmark' functie.

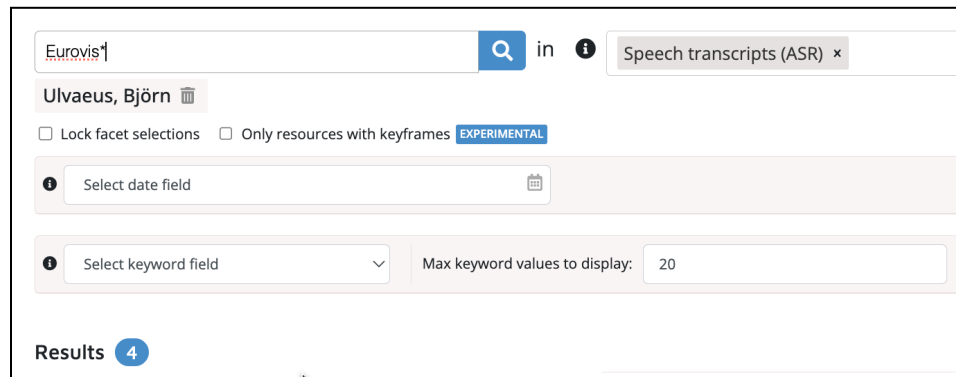
3.2 De zoekopdracht verfijnen: Linked Data

Een manier om je zoekopdracht te verfijnen is met behulp van '**linked data**'. Hiermee is het mogelijk om naar specifieke mensen te zoeken in programma's, of om meer informatie over hen te verkrijgen.

- Om met Linked Data te zoeken verwijder je eerst alle eerdere zoekopdrachten. Dit doe je door op '**Clear search**' te klikken rechts in het scherm.
- Klik nu op het metadata drop-down menu naast de zoekbalk. Kies hier voor '**Persons - All**'.
- Je kunt nu een naam invullen in de zoekbalk. Denk bijvoorbeeld aan **Björn Ulvaeus**, bandlid van ABBA. Het drop-down menu zal je verschillende opties geven, en na het selecteren van een naam zal deze nu als filter onder je zoekbalk staan.
 - Vergeet niet, linked data werkt alleen met **namen van personen**. De term ABBA, bijvoorbeeld, zal hier dus niet werken.



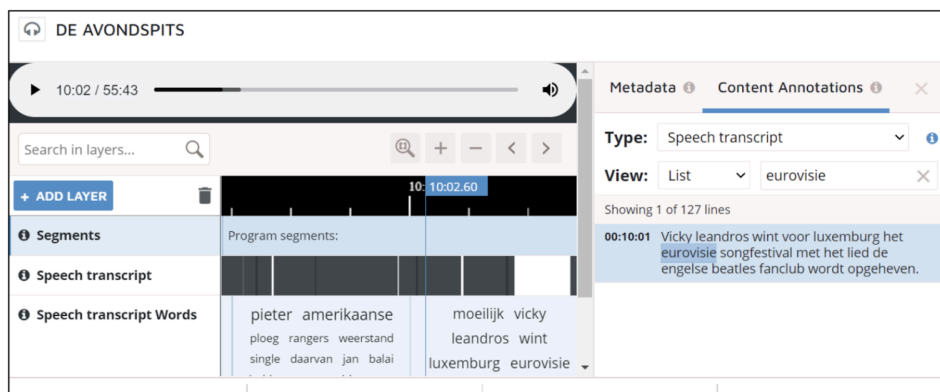
- Het is nu ook mogelijk om deze persoon te koppelen aan een andere zoekterm.
 - Zoek bijvoorbeeld, met **Björn Ulvaeus geselecteerd**, op 'Eurovisi*' in enkel de **spraaktranscripten**. Je hebt nu een zeer verfijnde zoekopdracht die, in dit geval, resulteert in 4 radioprogramma's waarin het ABBA-bandlid wordt besproken in de context van het ESF.



- Probeer een aantal opties uit en voeg interessante items toe aan je User Project. Vergeet ook niet je gefilterde zoekopdracht weer **op te slaan**.
- Je weet nu hoe je kan **zoeken in de ASR laag**, en dit kan combineren met **linked data en Boolean operators** in een zoekopdracht.

Stap 3: Close reading van het programma met ASR

3.1 Oriëntatie



Ga nu naar je User Project, en kies een van je opgeslagen programma's uit om te analyseren. Open dit programma, en oriënteer je eerst op de beschikbare informatie. Onder de audiospeler zie je **drie informatielagen**: Segments, Speech transcript en Speech transcript Words. Deze laatste twee zijn gerelateerd aan de ASR. De 'Speech transcript' laag toont lichte of donkere blokken. De 'Speech transcript Words' laag geeft een automatisch gegenereerde woordenwolk van elk segment aan.

- Speel de audio af en kijk hoe de twee lagen zich tot elkaar verhouden.
 - **Reflectievraag**: Wat hoor je tijdens de lichte en de donkere blokken?

Antwoord:

- **Reflectievraag:** Hoe relateren de woorden in de woordwolk tot wat je hoort? (Hoe) kan dit overzicht gebruikt worden om door het segment te navigeren?

Antwoord:

3.2 Analyse van het speech transcript

- Ga nu naar het volledige speech transcript. Hiervoor navigeer je naar de column aan de rechterkant van de audio, en klik je op 'Content Annotations'. Vanuit hier ga je naar het drop-down menu 'type' en selecteer je 'Speech transcript'. Zorg dat in het 'view' drop-down menu 'list' is geselecteerd.
- Hieronder vind je vervolgens **het transcript van het gehele programma**.
 - Let op, als je vanuit een nieuw resultaat werkt en niet vanuit een 'saved query', zal je zoekterm in de zoekbalk staan. Je kan hier echter ook op andere termen zoeken, of deze in zijn geheel weghalen.
- Navigeer nu door het audiobestand heen door in het spraak transcript te zoeken, of hierin op specifieke fragmenten te klikken. Je springt dan automatisch naar dat moment in de audio.
- **Reflectievraag:** Scan nu de rest van het spraak transcript van je gekozen uitzending.
 - Klopt het transcript helemaal? Kan je ook fouten vinden?
 - Hoe betrouwbaar lijkt het transcript? Wanneer is het minder betrouwbaar?
 - Wat betekent dit voor de conclusies die je kan trekken uit je ASR zoekopdrachten?

Antwoord: