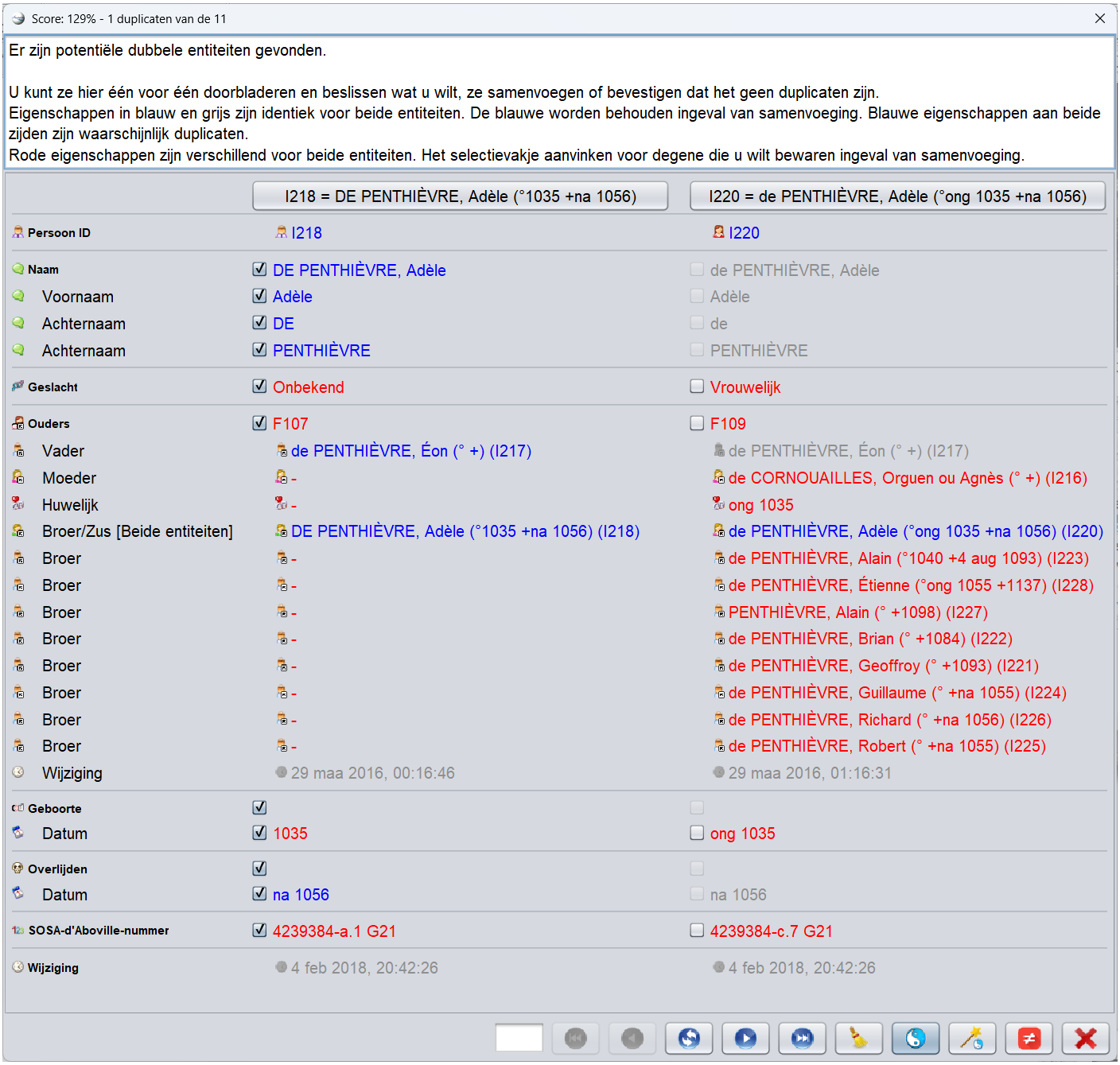
4.15. Duplicaten zoeken en samenvoegen

Deze tool detecteert dubbele entiteiten, analyseert ze en biedt de mogelijkheid om ze automatisch of handmatig samen te voegen als het echte duplicaten zijn, of om ze te bevestigen als niet-duplicaten als dat niet het geval is.

Er zijn 4 manieren om duplicaten samen te voegen, elk manier wordt op deze pagina beschreven.

- **Globale zoekopdracht** in het gehele genealogiebestand
- **Automatische detectie** telkens wanneer een entiteit wordt gewijzigd
- **Handmatige samenvoeging** waarbij de gebruiker twee entiteiten selecteert om samen te voegen
- **Slepen-en-neerzetten** van entiteiten tussen genealogiebestanden

Bij elk van deze methoden wordt onderstaand venster gebruikt:



Beschrijving

De tool Duplicaten Samenvoegen bestaat uit drie onderdelen:

- **Selectievenster voor entiteiten**: hier kies je welke entiteiten je wilt controleren op duplicaten.
- **Samenvoegvenster**: hier worden alle potentiële duplicaten getoond en kun je analyseren of ze moeten worden samengevoegd.
- **Speciale lijst voor niet-duplicaten**: hier worden je bevestigingen opgeslagen dat twee entiteiten géén duplicaten zijn.

Voordat we deze drie componenten beschrijven, leggen we eerst uit hoe Ancestris de kans berekent dat twee entiteiten duplicaten zijn of niet. Ancestris berekent een Overeenkomstscore.

Overeenkomstscore (Resemblance Score)

Het is vaak moeilijk om met absolute zekerheid vast te stellen dat twee entiteiten duplicaten zijn of juist niet. Zelfs voor een mens kan het soms lastig zijn om met zekerheid te bepalen of twee individuen of entiteiten daadwerkelijk dezelfde zijn of juist niet hetzelfde zijn.

Natuurlijk zou het gemakkelijker zijn om de detectie te beperken door te zeggen dat twee individuen met exact dezelfde achternaam, voornaam en geboortedatum duplicaten zijn. In werkelijkheid kunnen datums ontbreken of bij benadering zijn, kunnen voornamen in een andere volgorde staan of onvolledig zijn, enzovoort. In deze gevallen wilt u nog steeds dat Ancestris iets kan detecteren.

Daarom gebruikt Ancestris een Overeenkomstscore. Hoe meer bepaalde informatie overeenkomt, hoe waarschijnlijker het is dat de entiteiten duplicaten zijn.

De berekende scores kunnen variëren van negatieve getallen tot hoge positieve getallen. Ze worden uitgedrukt als percentage, hoewel een score hoger dan 100% kan zijn.

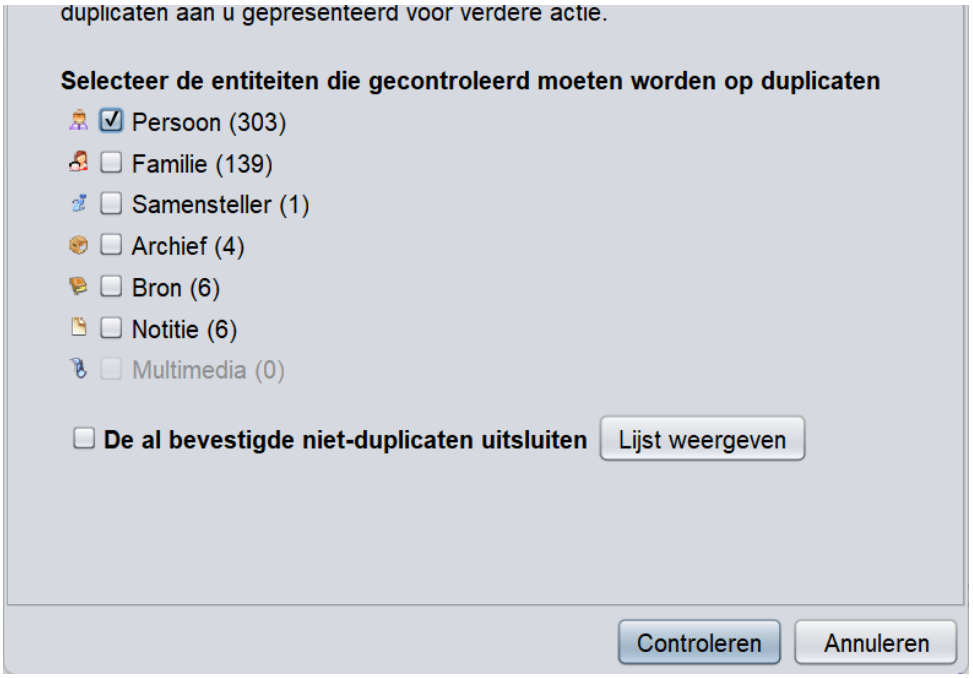
Ancestris zal de potentiële duplicaten volgens deze indicator in aflopende volgorde weergeven, met de bedoeling u te vertellen: "***Hoewel dit niet zeker is, gezien de gelijkenissen in de informatie tussen deze twee individuen, kunnen ze duplicaten zijn. En dit is de vertrouwensscore dat ze dat ook zijn.***"

Vervolgens is het aan u om te beslissen of u de duplicaten wilt samenvoegen of de gelijkenis wilt negeren.

> Dit betekent dat Ancestris u mogelijk potentiële duplicaten laat zien die u als geen duplicaten beschouwt, en omgekeerd. Ancestris kan ook mogelijk duplicaten **niet** weergeven, die volgens u **wel** echte duplicaten kunnen zijn.
>
> Onze excuses als de detector niet perfect is en laat het ons alstublieft weten als u dergelijke gevallen vindt.

Venster voor entiteits-selectie

Wanneer een globale zoekopdracht naar duplicaten wordt gestart, wordt het volgende venster weergegeven. Dit toont de verschillende soorten entiteiten en hoeveel ervan in het GEDCOM-bestand kunnen worden gevonden.



Selecteer de vakjes van de entiteiten waarvoor u naar duplicaten wilt zoeken.

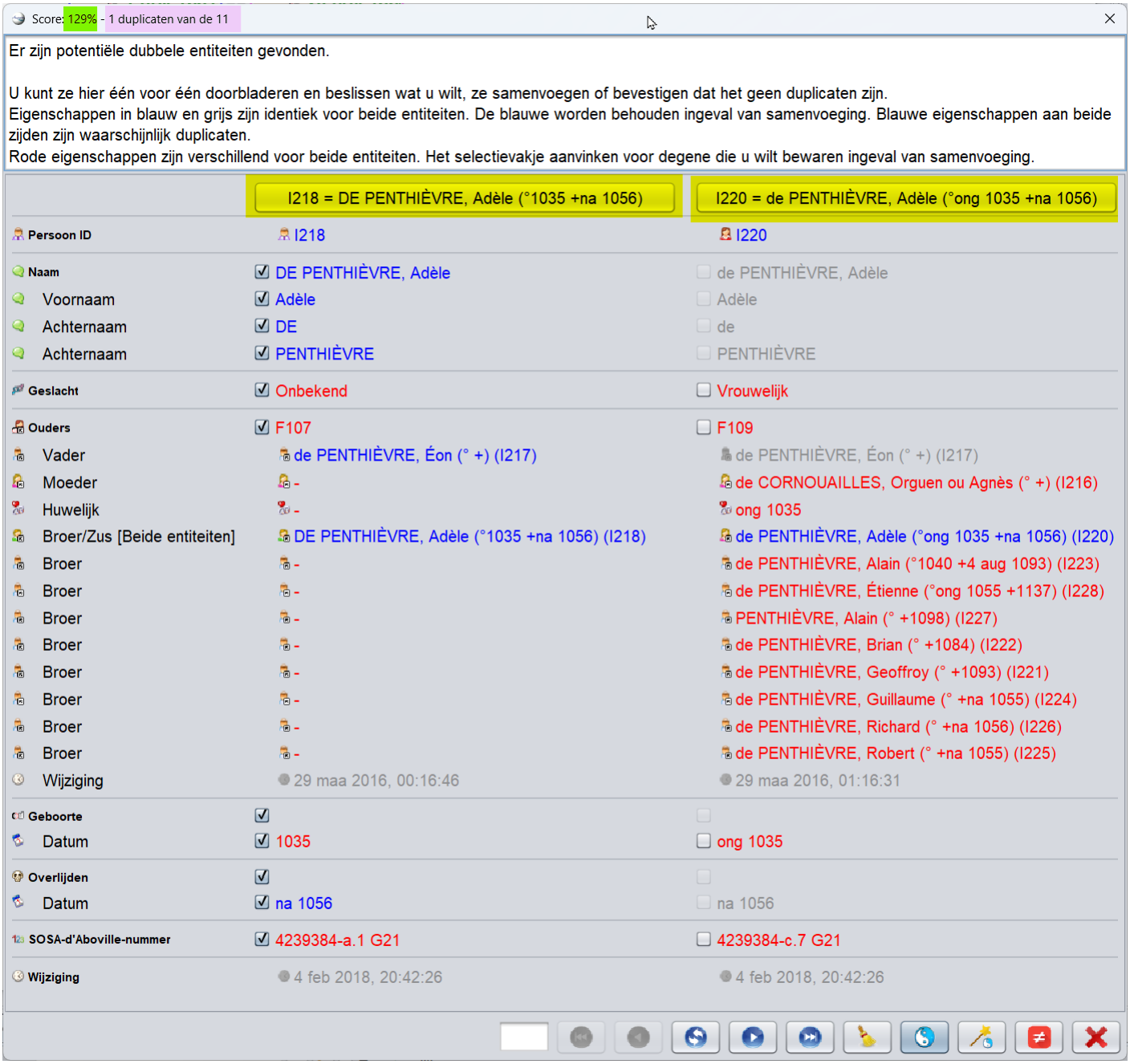
Alleen de vakjes van entiteiten die in het GEDCOM-bestand aanwezig zijn, zijn ingeschakeld. In het bovenstaande voorbeeld zijn er bijvoorbeeld geen *multimedia-entiteiten*.

Als u bekende niet-duplicaten wilt uitsluiten van de zoekopdracht, d.w.z. entiteiten die u al als niet-duplicaten hebt bevestigd, vink dan het overeenkomstige vakje aan.

Om de lijst van paren van entiteiten te zien die u als niet-duplicaten hebt bevestigd, drukt u op de knop **Lijst weergeven**. Dan verschijnt de lijst met duplicaten waarvan u bevestigd heeft dat ze **geen duplicaten** zijn.

Het samenvoeg-venster

Na het starten van het zoeken naar duplicaten verschijnt het volgende venster.



Dit venster toont één voor één de volledige **lijst van alle potentiële duplicaten waarbij de waarschijnlijkheid hoger is dan 40%**.
De lijst is gesorteerd van **het meest zekere duplicatenpaar naar het minst zekere**, per entiteitscategorie. Voor elk paar vergelijkbare entiteiten geeft Ancestris de **score-indicator** bovenaan weer.

Grote genealogieën kunnen enkele duizenden duplicaten bevatten. Dit venster toont maximaal de eerste 10.000 duplicaten per entiteitstype. Als uw genealogie meer duplicaten bevat, moet u deze eerst samenvoegen voordat u de overige kunt bekijken.

De titel van het venster geeft het nummer aan van het duplicatenpaar in de lijst (**hier in roze**), samen met de score-indicator die aangeeft hoe groot de kans is dat de twee entiteiten daadwerkelijk hetzelfde zijn en dus samengevoegd moeten worden (**hier in groen**).

Er wordt een **algemene boodschap** weergegeven (in het witte veld) die afhangt van de manier waarop de tool werd gestart: via een globale zoekopdracht, automatische detectie of een handmatige actie. Dit bericht helpt ook bij het begrijpen van de kleurcodering die wordt gebruikt om de informatie weer te geven.

Elk duplicatenpaar, bestaande uit twee entiteiten, wordt weergegeven in twee kolommen.

In de titel van elke kolom kan met een knop elk van de entiteiten in de bewerkingsvensters geopend worden voor meer details. (**hier geel aangegeven**)

In elke kolom worden de eigenschappen van elke entiteit van het veronderstelde duplicaat weergegeven.

- **Waarden die verschillen**, worden in **rood** weergegeven.
- **Waarden die identiek zijn**, worden in **blauw** weergegeven voor de linker entiteit en in **grijs** voor de rechter entiteit.
- **Waarden die waarschijnlijk duplicaten zijn**, worden aan beide zijden in **blauw** weergegeven.

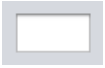
> **Het doel van de vergelijking is om de rechter entiteit < < < < in de linker entiteit samen te voegen als u bevestigt dat ze duplicaten zijn. Dus gegevens gaan van rechts < < < < naar links!**

Daarom is er voor elke eigenschap aan de rechterkant een selectievakje beschikbaar waarmee u Ancestris handmatig kunt laten weten welke gegevens na het samenvoegen behouden moeten blijven.

De knoppen helemaal onderaan het scherm, worden gebruikt om door de lijst met duplicatenparen te navigeren, de beslissing uit te stellen, ze onmiddellijk samen te voegen of ze direct als niet-duplicaat te bevestigen.

Werkbalk met knoppen



Veld 1: Zoek dubbele 

Met dit veld kan een dubbele gezocht worden in de lijst. Voer de tekst in en druk op Enter. gebruik vervolgens de knoppen **Vorige** en **Volgende** om naar de vorige of volgende overeenkomende duplicaat te gaan.

Knop 2: Ga naar de eerste dubbele 

Geeft de eerste dubbele weer in de volgorde van de betrouwbaarheidsindex, dwz de meest waarschijnlijke dubbele.

Knop 3: Ga naar de vorige dubbele 

Geeft het vorige duplicaat weer.

Knop 4: Wissel linker- en rechter- entiteiten 

Verwissel de linker en rechter entiteiten om de twee entiteiten aan de linkerkant samen te voegen. Dit is handig als de meeste informatie die moet worden bewaard na het samenvoegen zich aan de rechterkant bevindt.

Knop 5: Ga naar de volgende dubbele 

Geeft het volgende duplicaat weer.

Knop 6: Ga naar de laatste dubbele 

Geeft het laatste duplicaat in de betrouwbaarheidsindex weer, dus het minst waarschijnlijke duplicaat.

Knop 7: Verwijder dubbele 

Verwijdert de potentiële dubbele **uit de weergegeven lijst**.

Als de zoekopdracht naar dubbelen opnieuw wordt gestart, wordt deze (verwijderde) wel opnieuw weergegeven.

Knop 8: Voeg **handmatig** samen 

> Door op de knop **Handmatig Samenvoegen** te klikken, worden de entiteiten samengevoegd **op basis van de aangevinkte selectievakjes**.

- De entiteit aan de rechterkant wordt uit het GEDCOM-bestand verwijderd.
- De informatie waarvoor het selectievakje aan de rechterkant is aangevinkt, wordt **toegevoegd aan** of **vervangt** de informatie van de linker-entiteit.

- Voor informatie die slechts één keer kan voorkomen, is het alleen mogelijk om de gegevens van één van beide entiteiten te behouden.

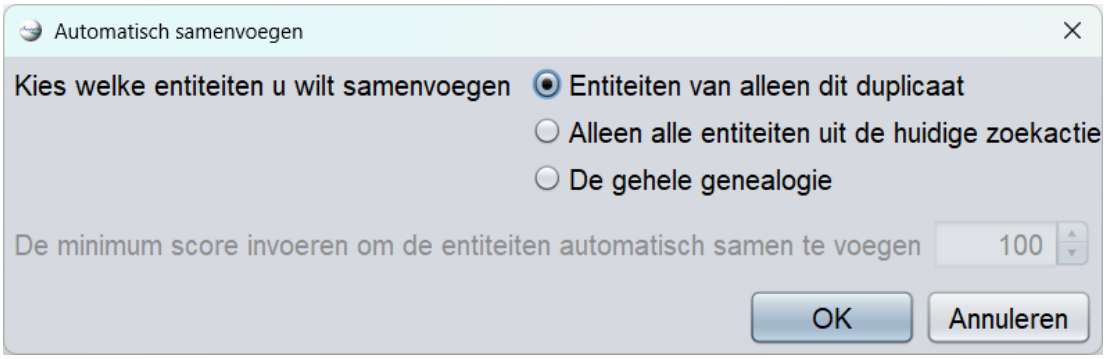
Zodra de samenvoeging is voltooid, wordt het resultaat weergegeven in groen, zodat je kunt controleren of alles correct is overgenomen. Daarna kun je verder gaan met het volgende duplicaat.

Mocht je denken dat er toch iets niet goed is gegaan, dan is er altijd nog de Undo-knop in de buttonbar bovenaan het scherm.

Knop 9: Voeg **automatisch** samen 

> Door op de knop **Automatisch Samenvoegen** te klikken, worden de entiteiten **automatisch** samengevoegd **zonder gebruik te maken van de selectievakjes**.

- Ancestris bepaalt zelf welke informatie van de te verwijderen entiteit behouden blijft en voegt deze toe aan de te behouden entiteit.
- Bij het indrukken van de knop verschijnen er drie keuzemogelijkheden:



1. **Alleen deze duplicaat:**
   - Alleen het weergegeven duplicaat wordt samengevoegd.
   - Ancestris detecteert welke informatie van de rechterentiteit behouden blijft om de linkerentiteit aan te vullen.
   - Het resultaat wordt in groen weergegeven.
   - In het **Uitvoervenster** (Ctrl+T) worden de samengevoegde entiteiten en de scores vermeld, zodat je deze kunt analyseren en opslaan.
2. **Alle entiteiten van de huidige zoekopdracht:**
   - De automatische samenvoeging wordt toegepast op alle duplicaten in de lijst met een score boven een bepaalde drempel (deze moet je zelf opgeven).
   - Alleen de eerste 10.000 duplicaten per entiteit worden samengevoegd. Wil je de hele genealogie samenvoegen, kies dan de volgende optie.
   - Het **resultaatvenster** toont alle samengevoegde duplicaten en de overgebleven duplicaten.
   - Het **Uitvoervenster** (Ctrl+T) vermeldt alle samengevoegde entiteiten en kan worden opgeslagen ter referentie.

3. **De hele genealogie:**
   - Alle duplicaten in de genealogie worden samengevoegd boven een bepaalde score (niet beperkt tot de eerste 10.000).
   - Het **resultaatvenster** verdwijnt, omdat niet alle duplicaten erin zouden passen.
   - Het **Uitvoervenster** (Ctrl+T) vermeldt alle samengevoegde entiteiten en kan worden opgeslagen als referentie.

Knop 10: Dit is geen duplicaat 

- Deze knop verwijderd het paar entiteiten uit de lijst met potentiële duplicaten.
- Het paar wordt als "**geen duplicaat**" gemarkeerd en deze bevestiging wordt opgeslagen in een speciale notitie, gemerkt als "**niet-duplicaat**".

Knop 11: Sluit het venster 

Sluit het venster.

De speciale Notitie voor Niet-Duplicaten.

In Ancestris wordt een speciale notitie aangemaakt en bijgewerkt om bevestigingen van **niet-duplicaten** op te slaan.

- Deze notitie bevat de bevestigingen van de gebruiker dat bepaalde entiteiten, ondanks overeenkomsten, **geen duplicaten** zijn.
- Hierdoor voorkomt Ancestris dat deze entiteiten opnieuw als mogelijke duplicaten worden gedetecteerd bij een **globale zoekopdracht** of **automatische detectie**.
- De notitie krijgt de referentienaam **"Non_Duplicates"**.



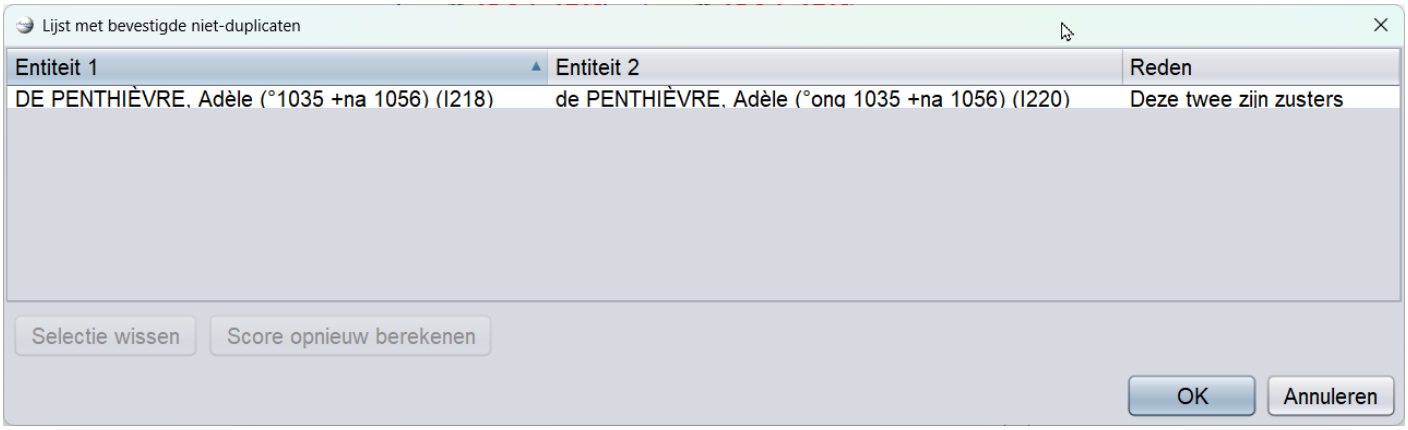
Deze notitie wordt bijgewerkt telkens wanneer je:

- Op de knop **Niet-duplicaat** drukt in het **samenvoegvenster**.
- De lijst met niet-duplicaten bijwerkt via het **Criteria-venster**.

> Ancestris slaat deze informatie **in het GEDCOM-bestand zelf** op, omdat de analyse en bevestiging van niet-duplicaten een**waardevol onderdeel** van je genealogische gegevens vormt.
>
> Omdat de **GEDCOM-standaard** hier momenteel geen specifieke oplossing voor biedt, heeft Ancestris ervoor gekozen om deze informatie op te slaan in **één enkele notitie**.
>
> Mochten er in de toekomst wijzigingen in de GEDCOM-standaard komen, zoals een **NOALIAS-tag**, dan kan Ancestris de manier waarop deze informatie wordt beheerd aanpassen.

Lijst met niet-duplicaten (bekijken)

- Je kunt de lijst met bevestigde **niet-duplicaten** bekijken via het **Venster Duplicaten samenvoegen door op de knop "Lijst weergeven"** te drukken.
- De lijst kan worden **gesorteerd** om specifieke entiteiten sneller te vinden.
- Je kunt één of meerdere regels **selecteren en verwijderen** als je bepaalde niet-duplicaten uit de lijst wilt halen.



Gebruik

Zoals hierboven vermeld, zijn er 4 manieren om entiteiten samen te voegen.

- **Uitgebreide zoekopdracht** in het hele genealogiebestand.
- **Automatische detectie** telkens wanneer een entiteit wordt gewijzigd.
- **Handmatige samenvoegactie** door de gebruiker om twee geselecteerde entiteiten te dwingen samen te voegen.
- **Slepen en neerzetten** van entiteiten van het ene genealogiebestand naar het andere.

Uitgebreide zoekopdracht

Het doel van de uitgebreide zoekopdracht is om duplicaten in de hele genealogie te identificeren en hierop te handelen. Dit betekent dat je per duplicaat de keuze hebt om:

- het duplicaat **samen te voegen**,
- het als **niet-duplicaat** te verklaren,
- of **de beslissing uit te stellen**. tot later

Je kunt de uitgebreide zoekopdracht starten via het Gereedschaps-menu in Ancestris.

Het samenvoegproces verloopt in twee stappen:

- Eerst kies je welk type entiteit Ancestris moet doorzoeken op duplicaten.
- Daarna kies je of en hoe je de **duplicaten wilt samenvoegen** in het Samenvoegvenster.

Tijdens het gebruik van dit gereedschap worden wijzigingen in de genealogie aangebracht:

1. Samengevoegde entiteiten worden gecombineerd met de informatie die je hebt gekozen te behouden, of die Ancestris heeft geselecteerd.
2. Niet-duplicaten worden opgeslagen in een **speciale notitie**.

Automatische detectie

> Het doel van automatische detectie is om je te waarschuwen als de entiteit die je op dat moment maakt of wijzigt een mogelijk duplicaat is van een bestaande entiteit.

Automatische detectie is standaard ingeschakeld in de **Ancestris-voorkeuren**.
Zodra je een invoer in een editor bevestigt en de detectie-instelling is ingeschakeld, zal Ancestris automatisch zoeken naar mogelijke duplicaten.

Alle potentiële duplicaten worden weergegeven in het **Samenvoegvenster**, waar je kunt beslissen wat ermee moet gebeuren.

> Bij de Cygnus-editor, waar meerdere entiteiten tegelijk kunnen worden bewerkt, worden alle gewijzigde entiteiten gecontroleerd. Daarom zal het samenvoegvenster alle mogelijke duplicaten van alle gewijzigde entiteiten tonen.

Handmatige samenvoeging

> Het doel van de handmatige samenvoegactie is om twee entiteiten samen te voegen, ongeacht of Ancestris ze heeft gedetecteerd als duplicaten.

>

> Een andere doel is om duplicaten voor een specifieke entiteit te identificeren.

Deze actie is toegankelijk via het **contextmenu** van de entiteit die je wilt samenvoegen.



Selecteer de optie om een entiteit samen te voegen.
Ancestris vraagt je daarna met welke andere entiteit deze moet worden samengevoegd.

- **Om daadwerkelijk samen te voegen**, kies je een entiteit die je als duplicaat beschouwt.
- **Om alleen te controleren of er duplicaten zijn**, kies je willekeurig een entiteit uit de lijst.

Ancestris toont het Samenvoegvenster met een lijst van mogelijke duplicaten, inclusief de twee geselecteerde entiteiten en hun overeenkomstscore (die 0% of zelfs negatief kan zijn).

De lijst wordt gesorteerd op afnemende waarschijnlijkheid en toont eerst het gekozen paar.

Als de geselecteerde entiteit geen andere duplicaten heeft, wordt alleen het gekozen paar weergegeven.

Je kunt dan besluiten de huidige entiteit samen te voegen met de geselecteerde entiteit of een andere entiteit uit de lijst.

Slepen en neerzetten van de ene stamboom naar de andere

> Het doel van de slepen-en-neerzetten-functionaliteit is om entiteiten met de muis van de ene genealogie naar de andere te kopiëren. Zodra de entiteit in de andere genealogie wordt neergezet, vraagt Ancestris hoe je de entiteit wilt koppelen of toevoegen. Als je kiest voor samenvoegen, verschijnt het samenvoegvenster. **Zie de bijbehorende pagina hier**.

Aanpassingen

Er zijn twee instellingen waarmee je het Samenvoeghulpmiddel kunt aanpassen:

- **De vlag voor automatische detectie**, opgeslagen in de **Voorkeuren in de gebruikersmap (user-directory)**.
- **Niet-duplicaten**, opgeslagen in een speciale notitie.

Revision #15
Created 13 April 2023 08:29:38 by Tineke
Updated 9 May 2025 06:57:10 by Tineke