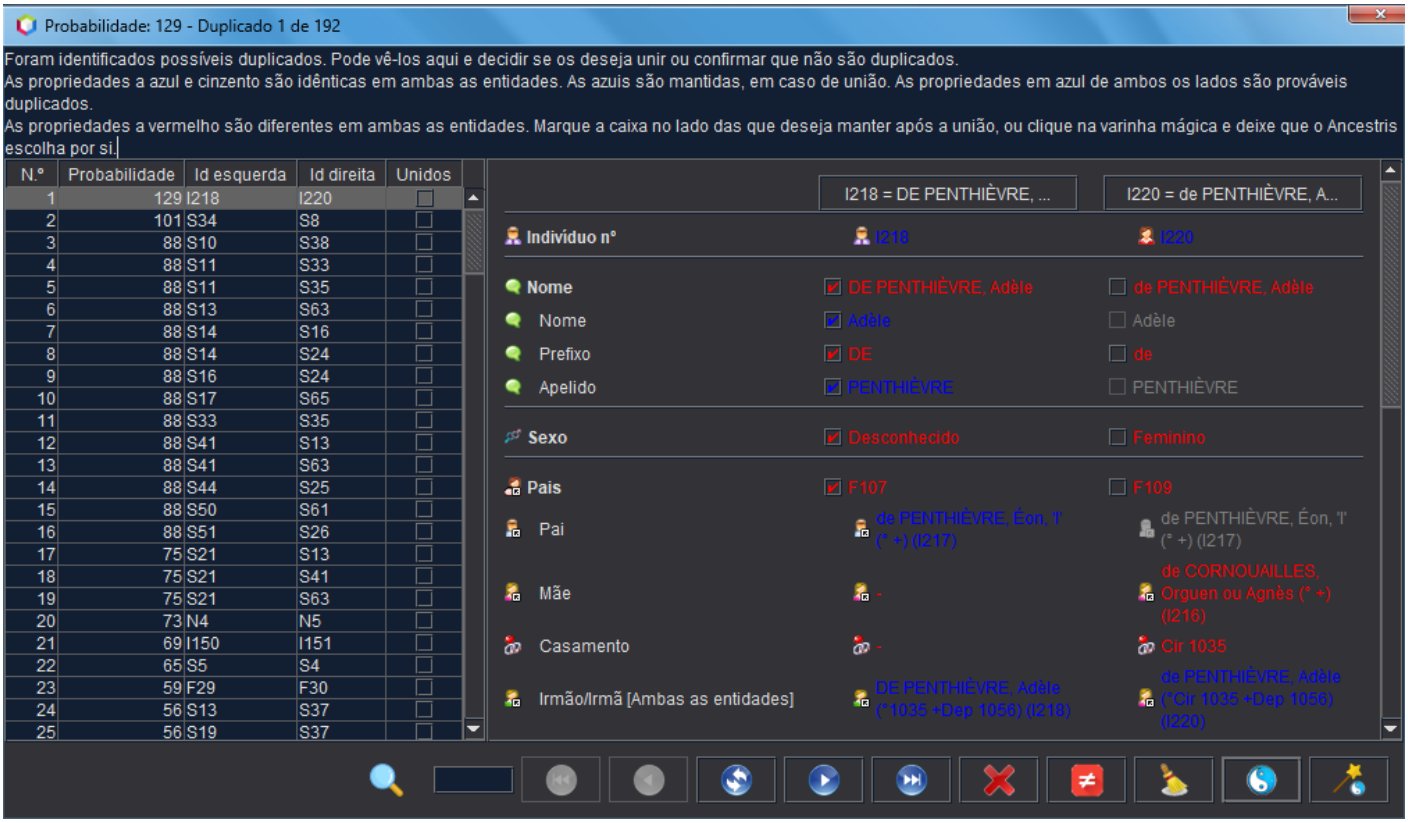
União de duplicados

Esta ferramenta de fusão de duplicados detecta duplicados e permite-lhe uni-los automática ou manualmente se forem verdadeiros duplicados, ou confirmá-los como não duplicados se não o forem.

Existem três formas de procurar e unir duplicados. São explicadas nesta página.

- **Pesquisa global** em toda a genealogia.
- **Detecção automática** sempre que uma entidade é modificada.
- **Acção manual de união** pelo utilizador sobre duas entidades seleccionadas.
- **Arrastar e largar** ou **copiar** entidades de uma genealogia para outra.

A janela seguinte é utilizada em todas estas situações.



Descrição

Esta ferramenta é composta de três elementos:

- a janela **Critérios de detecção**, onde pode definir alguns dos critérios que determinarão se duas entidades devem ser consideradas como duplicadas ou não;
- a janela **Unir**, onde verá todos os potenciais duplicados e decidirá se os deve fundir ou não;
- e a **Nota especial sobre não duplicados**, onde o Ancestris grava todas as suas confirmações de que duas entidades não são duplicadas.

Antes de descrever estes 3 elementos, vamos primeiro ver como o Ancestris mede a probabilidade de 2 entidades serem duplicadas ou não. O Ancestris calcula uma pontuação de semelhança.

Pontuação de semelhança

Muitas vezes é difícil afirmar com total certeza se duas entidades são duplicadas ou não. Mesmo um ser humano pode por vezes ter dificuldade em certificar que dois indivíduos ou entidades são definitivamente iguais, ou definitivamente diferentes.

É claro que seria mais fácil limitar a detecção dizendo que duas pessoas com exactamente o mesmo apelido, nome próprio e data de nascimento são duplicados. Na realidade, as datas podem estar em falta ou ser aproximadas, os nomes próprios podem estar numa ordem diferente ou incompletos, etc. Nestes casos, é desejável que o Ancestris seja capaz de detectar algo.

> É por esta razão que **o Ancestris utiliza uma pontuação de semelhança**. Quanto mais semelhantes forem determinadas informações, mais provável é que as entidades sejam duplicadas.

As pontuações calculadas podem variar de valores negativos a altos. São expressas como uma percentagem, embora a pontuação possa exceder 100%.

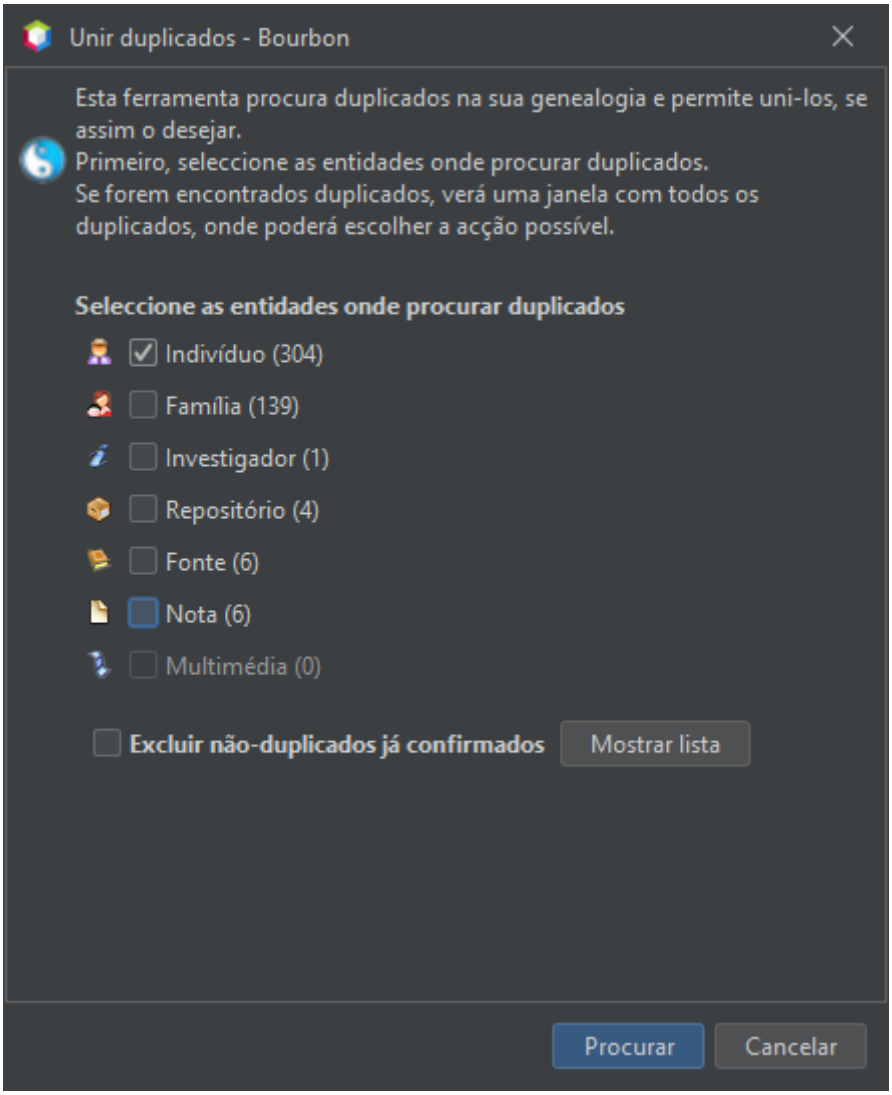
O Ancestris elabora então uma lista de potenciais duplicados com base nesta pontuação, com a intenção de lhe dizer: "**Embora não seja certo, dadas as semelhanças nas informações entre estas duas pessoas, é possível serem duplicados. E este é o nível de confiança entre elas**". Depois, cabe-lhe a si decidir se quer unir ou descartar a semelhança.

> Isto significa que o Ancestris pode mostrar-lhe potenciais duplicados que você considera não serem duplicações e, inversamente, pode não encontrar certas duplicados que você considera serem verdadeiras duplicações.
> Pedimos desculpa por o detector não ser perfeito e por favor, informe-nos se encontrar algum destes casos.

> Veja abaixo a configuração dos limites da pontuação de semelhança.

Janela de selecção de entidades

Quando é lançada uma pesquisa global de duplicados, a janela de critérios de detecção é mostrada, com uma série de critérios.



Marque as caixas das entidades para as quais pretende procurar duplicados.

Apenas as caixas para entidades presentes no ficheiro Gedcom estão activadas. No exemplo acima, não existem entidades multimédia.

Não-duplicados conhecidos

Se **pretender excluir da pesquisa os duplicados conhecidos**, ou seja, aqueles que já confirmou não serem duplicados, marque a caixa correspondente.

Para ver a lista de pares de entidades que já confirmou não serem duplicados, clique em **Mostrar lista**. Verá a **lista de entidades não duplicadas**.

Pontuação mínima

Foi criado um **cursor de deslizamento de pontuação mínima**, que pode utilizar para excluir as pontuações menos prováveis.

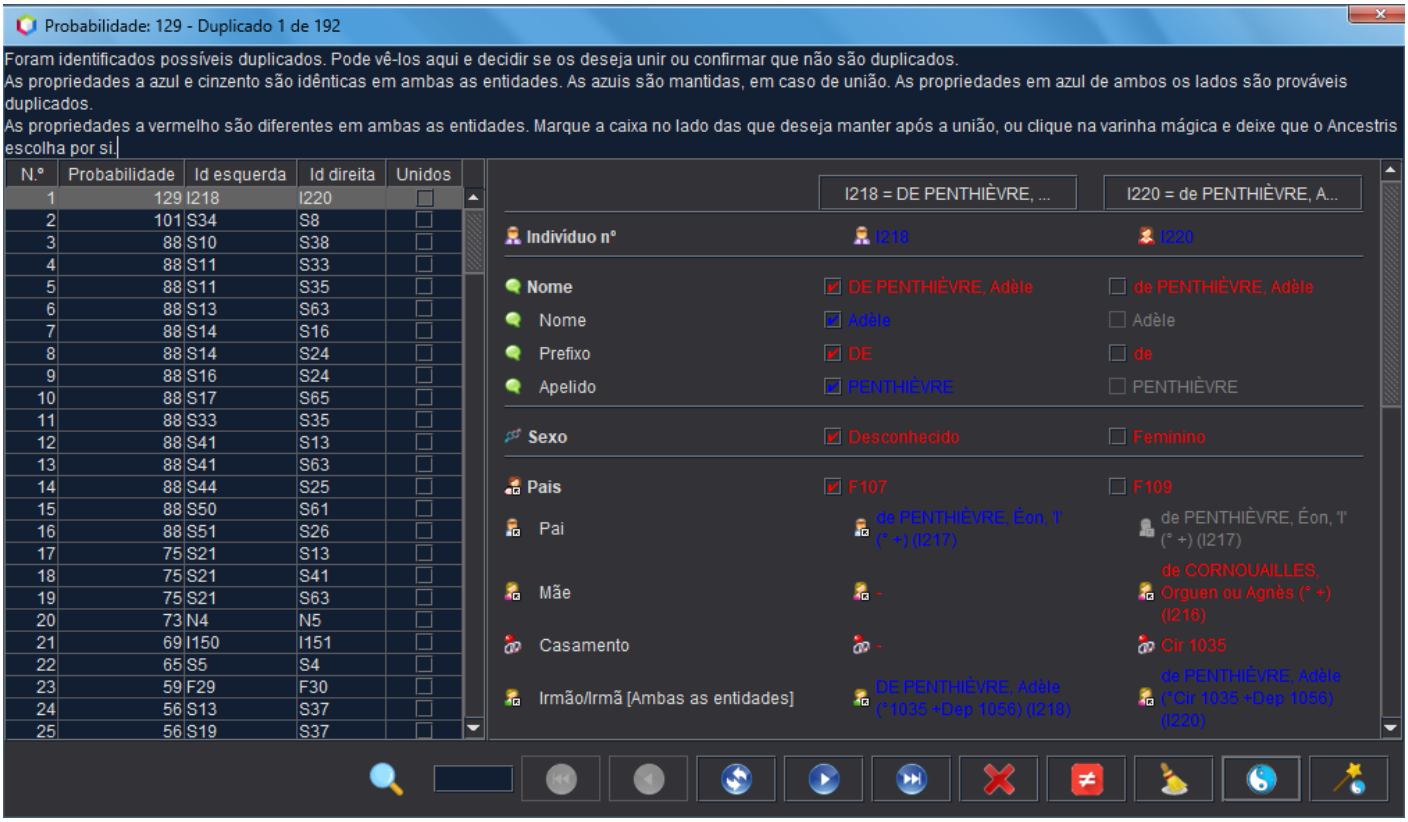
> Definir a sua própria pontuação mínima é útil quando são encontrados muitos potenciais duplicados, mas a principal razão para a sua existência é dar-lhe a garantia de que o Ancestris encontrou todos os potenciais duplicados e não deixou escapar nenhum.

Algumas considerações a ter ao definir a pontuação mínima:

- o limiar mínimo ideal é aquele que separa os duplicados confirmados dos não confirmados;
- não existe na realidade, pois alguns pares de duplicados não reais podem obter uma pontuação mais alta do que os duplicados reais próximos desse limite;
- esse limiar depende da genealogia de cada utilizador e situa-se geralmente entre 40 e 50;
- deve experimentar vários limiares para a sua genealogia até ficar satisfeito com o facto de todos os duplicados mostrados serem, na sua maioria, verdadeiros duplicados, mas ainda mostrando alguns que não o são, apenas para se certificar de que os tem todos.

Janela Unir

Para entidades comparadas como duplicados, é utilizada a seguinte janela.



> Esta janela exibe uma a uma a **lista total de todas as duplicatas potenciais cuja probabilidade é maior que 50%**.
> A lista é ordenada **do par de duplicados mais provável para o par menos provável**, por categoria de entidade. Para cada par de entidades semelhantes, o Ancestris dá-lhe uma percentagem de probabilidade.

> As grandes genealogias podem conter vários milhares de duplicados. Esta janela apenas apresentará os primeiros 10.000 duplicados para cada tipo de entidade. Se a sua genealogia contiver mais duplicados, terá de os unir antes de poder ver os outros.

Título

O **título** da janela mostra o número do par de duplicados apresentado e a pontuação de semelhança, que indica que as duas entidades neste par são de facto as mesmas e devem, portanto, ser unidas.

Mensagem

É mostrada uma **mensagem geral**, consoante a situação em que a ferramenta foi lançada: pesquisa global, detecção automática ou acção manual.

Lista de selecção ordenável

No lado esquerdo, encontra-se uma lista ordenável de todos os pares identificados como possíveis duplicados.

Pode ordenar a lista clicando nos títulos das colunas.

Ao seleccionar a linha de um possível duplicado, o par correspondente é mostrado com os detalhes na parte direita da janela.

A caixa "Unidos" indica os pares de duplicados que já uniu usando um dos botões de união localizados abaixo.

Detalhes comparativos dos duplicados

Cada par de duplicados composto por duas entidades é mostrado em duas as colunas.

No cabeçalho de cada coluna, um botão permite-lhe selecionar cada uma das entidades nos editores para obter mais detalhes.

Em cada coluna, são mostradas as propriedades de cada uma das entidades no presumível duplicado.

- Os valores que são diferentes são mostrados a vermelho.
- Os valores idênticos são apresentados a azul para a entidade à esquerda e a cinzento para a entidade à direita.
- Os valores semelhantes são apresentados a azul em ambas as colunas.

> O objectivo da comparação é unir a entidade da direita com a entidade da esquerda, se se confirmar que são duplicados.

Como resultado, está disponível uma caixa de verificação para cada propriedade no lado direito para dizer ao Ancestris para manter a informação seleccionada para cada entidade depois de as unir.

Os botões permitem-lhe consultar a lista de potenciais duplicados, adiar a decisão, unir agora, ou confirmar agora que não são duplicados.

Barra de botões



Campo de procura de duplicados 

Utilize este campo para procurar um duplicado na lista. Introduza o texto e prima Enter.

Botão Primeiro duplicado 

Mostra o primeiro duplicado da lista por ordem de índice de confiança, ou seja, o duplicado mais provável, ou o duplicado 100 posições antes do duplicado actual, se existirem mais de 100 duplicados na lista.

Botão Duplicado anterior 

Mostra o duplicado anterior. Se existir um texto de pesquisa no campo Procurar duplicado, será mostrado o duplicado anterior correspondente a esse critério de pesquisa.

Botão Trocar duplicados 

Inverte as entidades esquerda e direita para que as duas entidades possam ser unidas na direita. Esta operação é útil se a maior parte da informação a ser retida após a união estiver no lado direito.

Botão Duplicado seguinte 

Mostra o duplicado seguinte. Se existir um texto de pesquisa no campo Procurar duplicado, será mostrado o duplicado seguinte correspondente a esse critério de pesquisa.

Botão Último duplicado 

Mostra o último duplicado da lista por ordem de índice de confiança, ou seja, o duplicado menos provável, ou o duplicado 50 posições depois do duplicado actual, se existirem mais de 100 duplicados na lista.

Botão Fechar 

Fecha a janela.

Botão Não-duplicados 

Este botão **exclui o par de entidades da lista de potenciais duplicados**.

Marca o par de entidades como não sendo duplicado e regista esta confirmação na **nota especial "não duplicados"**.

Botão Remover duplicado 

Remove o duplicado potencial da lista mostrada.

Esta opção é útil quando ainda não se sabe se as duas entidades são duplicadas ou não, e se deseja adiar a tomada de decisão.

Se for lançada uma nova pesquisa global de duplicados, o duplicado voltará a aparecer.

Botão União manual 

Ao clicar em Unir, **as entidades serão unidas utilizando as caixas de verificação marcadas**.

A entidade da direita é eliminada do ficheiro Gedcom e a informação cuja caixa está marcada à direita é acrescentada ou substitui a da entidade da esquerda.

Para informações que só podem existir uma vez, só é possível manter as informações numa das duas entidades.

Assim que a união estiver concluída, a janela mostra o mesmo duplicado com o resultado da união, para que possa verificar se tudo foi mantido como pretendido.

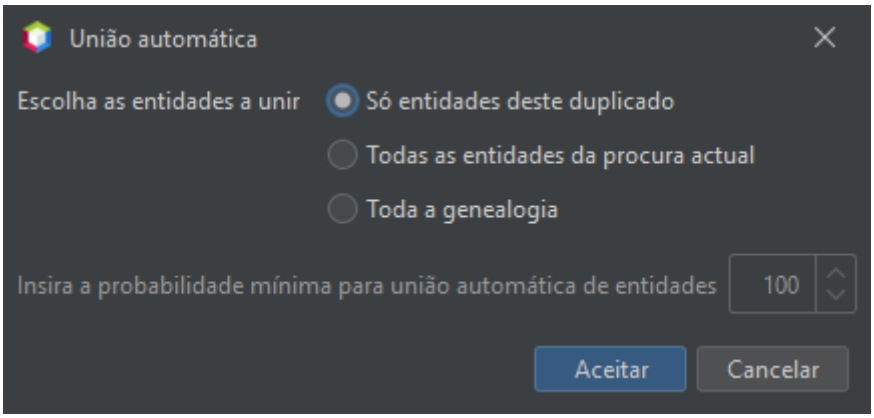
Pode então passar para o duplicado seguinte.

Botão Unir automaticamente 

Ao clicar em Unir automaticamente, as entidades serão unidas automaticamente pelo Ancestris sem utilizar as caixas de verificação.

O Ancestris determinará quais as informações da entidade a eliminar que devem ser mantidas e enriquecerá as informações da entidade a manter.

O Ancestris permite-lhe efectuar esta união automática em três âmbitos diferentes. Ao premir o botão Unir automaticamente, aparecem as seguintes opções:



- **Só entidades deste duplicado**:
  - Apenas o duplicado actualmente mostrado será unido.
  - O Ancestris detecta a informação a conservar na entidade direita para enriquecer a entidade esquerda.
  - Em seguida, o duplicado é actualizado e mostrada apenas a entidade esquerda em verde com a informação resultante.
  - A janela de saída (Ctrl+T) mostra as entidades unidas e a pontuação. Pode analisar este ficheiro e gravá-lo como referência.
- **Todas as entidades da procura actual**:
  - A união automática descrita acima será efectuada para todos os duplicados da lista cuja pontuação seja superior a um determinado limiar, que terá de especificar no campo abaixo.
  - Apenas os primeiros 10.000 duplicados de cada entidade serão tidos em conta. Utilize a opção descrita abaixo se pretender unir toda a genealogia e não apenas os primeiros 10.000 duplicados encontrados.
  - A janela de resultados incluirá todos os duplicados unidos acima da pontuação especificada, bem como todos os outros duplicados não unidos.
  - Quaisquer duplicados que contenham entidades eliminadas serão removidos da lista.
  - A janela de saída (Ctrl+T) lista todas as entidades unidas. Pode analisar e gravar este ficheiro de saída como referência.
- **Toda a genealogia**:
  - A união automática descrita acima será efectuada para todos os duplicados na genealogia acima de uma determinada pontuação, e não apenas para os primeiros 10.000. Terá de especificar esta pontuação no campo abaixo.
  - A janela de resultados desaparecerá, uma vez que pode não ter incluído todos os duplicados encontrados.
  - A janela de saída (Ctrl+T) mostrará uma lista de todos os duplicados encontrados.

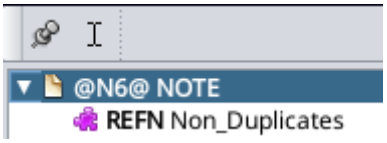
Nota especial para não duplicados

> É criada e actualizada uma nota especial no Ancestris para guardar as confirmações de não-duplicados.

Esta nota guarda as suas confirmações de pares semelhantes que não considera duplicados.

Isto evita que o Ancestris os detecte repetidamente cada vez que a pesquisa global ou a detecção automática é lançada.

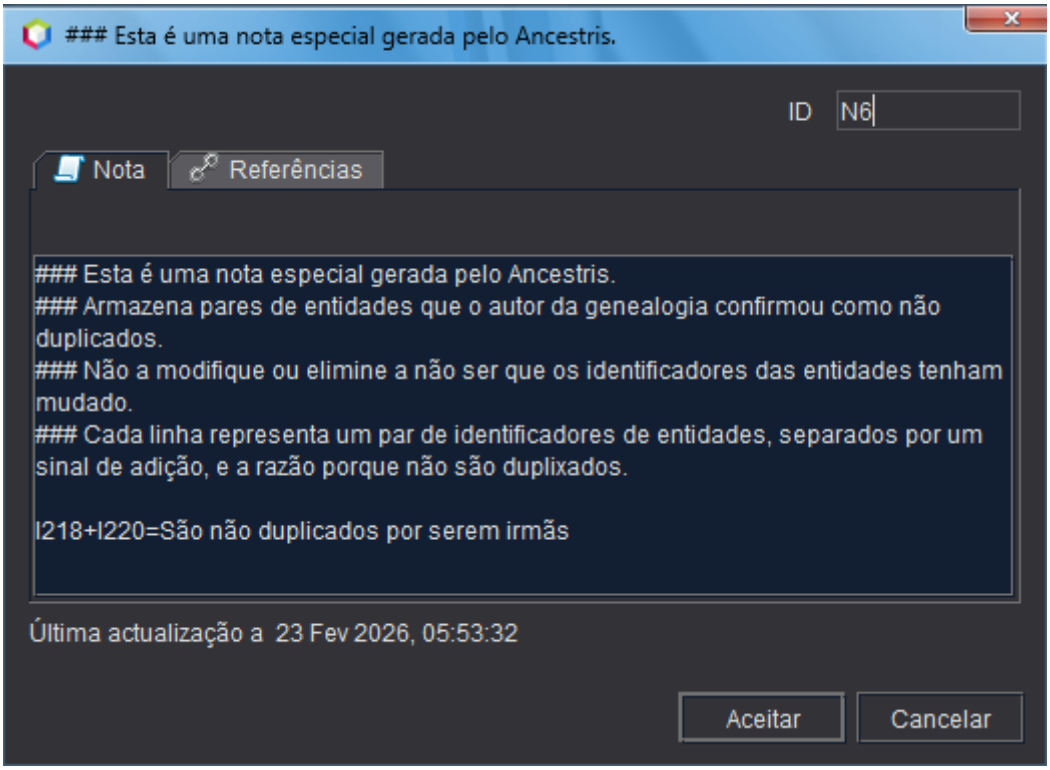
Esta nota tem um nome de referência chamado "Não-Duplicados".



Esta nota é actualizada sempre que premir o botão "Não-duplicados" na **janela Unir** ou actualizar a lista utilizando a **janela Critérios**.

> Optámos por armazenar esta informação do utilizador no próprio ficheiro Gedcom porque queríamos que os seus esforços para analisar as entidades e decidir que não são duplicados constituíssem informação genealógica valiosa que deveria ser retida e transferida como parte do ficheiro Gedcom.
> A norma Gedcom não responde a esta necessidade, pelo que o Ancestris optou por armazená-la desta forma, numa única nota.
> Se a norma Gedcom evoluísse neste ponto, por exemplo com uma etiqueta NOALIAS, poderíamos então alterar a forma como esta informação é gerida.

Esta nota aparece como nota isolada. Recomendamos que não a elimine. O texto da nota explica a sua função.



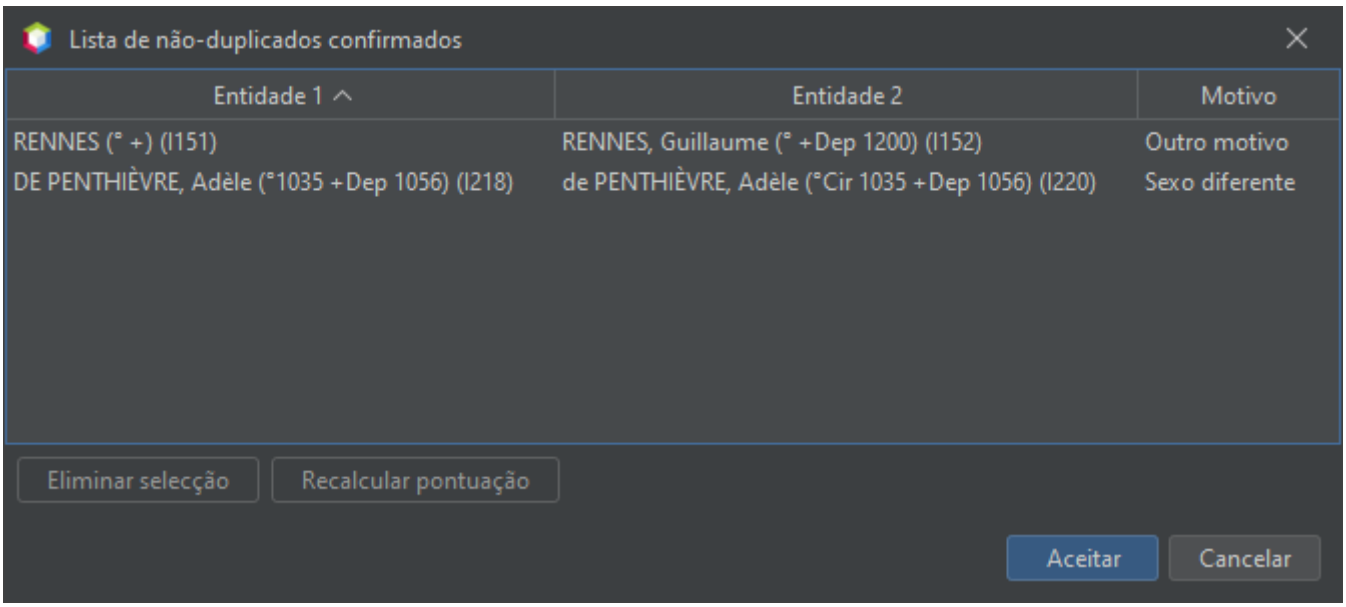
É possível consultar a lista de não-duplicados confirmados na **janela de critérios**.

Lista de não-duplicados

Pode ver a lista de entidades não duplicadas clicando em **Mostrar lista**, na **janela Criterios de detecção**.

Pode ordenar as linhas para encontrar as entidades que lhe interessam.

É possível seleccionar uma ou mais linhas e removê-las da lista, se necessário.



Utilização

Como indicado acima, existem três formas de procurar e unir duplicados:

- **Pesquisa global** em toda a genealogia.
- **Detecção automática** sempre que uma entidade é modificada.
- **Acção manual de união** pelo utilizador sobre duas entidades seleccionadas.
- **Arrastar e largar** ou **copiar** entidades de uma genealogia para outra.

Pesquisa global

> O objectivo da pesquisa global é **identificar duplicados na genealogia como um todo e actuar sobre eles**, ou seja, decidir um a um o que pretende fazer em relação a cada um deles, ou deixar que o Ancestris efectue uma união global por si para além de uma determinada pontuação de semelhança.

Para cada duplicado, pode então decidir:

1. **unir** o duplicado;
2. declará-lo como **não duplicado**;
3. **adiar** a decisão.

Pode iniciar a pesquisa global a partir do menu **Ferramentas** do Ancestris.

A ferramenta de união de duplicados funciona em duas fases:

- primeiro, especifica os **critérios de deteção** na janela correspondente;
- em seguida, selecciona como unir os duplicados na janela **Unir**.

Quando a ferramenta é utilizada, a genealogia é modificada em conformidade:

1. As entidades que decidir unir são unidas com as informações que especificou manter;
2. As entidades declaradas como não duplicadas são gravadas na **nota especial**.

Detecção automática

> O objectivo da detecção automática é alertá-lo se a entidade que está a criar ou a modificar é uma potencial duplicação de outra entidade que já existe na sua genealogia.

A detecção automática de duplicados é activada por defeito nas **preferências do Ancestris**

Assim que validar a sua entrada num dos editores, e se a caixa de preferências correspondente estiver assinalada, a detecção procura automaticamente potenciais duplicados da entidade que está a ser editada.

Todos os potenciais duplicados são então mostrados na janela **Unir**, para poder decidir o que fazer com eles..

> No caso do editor Cygnus, onde várias entidades podem ser editadas ao mesmo tempo, todas as entidades modificadas são verificadas e o Ancestris listará todos os potenciais duplicados de todas as entidades modificadas na janela **Unir**.

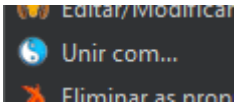
Acção de união manual

> O objectivo da acção Unir manualmente é uir duas entidades, quer o Ancestris as tenha detectado ou não.

>

> Outro objectivo é identificar todos os duplicados de uma determinada entidade.

Esta acção pode ser acedida a partir do **menu contextual** da entidade actual que se pretende unir com outra.



Quando esta acção é seleccionada, o Ancestris pergunta-lhe com que outra entidade a entidade actual deve ser unida.

- **Para unir efectivamente com outra entidade**, escolha a que considera ser um duplicado.
- **Para saber se a entidade actual tem duplicados na genealogia**, basta escolher qualquer entidade da lista de entidades.

O Ancestris mostra então a janela **Unir** com uma lista de potenciais duplicados, incluindo o par de duas entidades que escolheu e a sua correspondente probabilidade de serem a mesma entidade, que pode ser 0%.

A lista será ordenada por probabilidade decrescente e o par de duplicados mostrado será o par de entidades que escolheu.

Se a entidade a partir da qual começou não tiver outros duplicados encontrados, apenas o par de entidades escolhido será mostrado na lista.

Pode então decidir unir a entidade actual que deixou com a outra entidade que escolheu, ou com qualquer outra entidade da lista.

Arrastar e largar ou copiar de uma genealogia para outra

> O objectivo do **arrastar e largar**/**copiar** entidades entre genealogias é copiar entidades de uma genealogia para outra utilizando o rato ou através do menu **Ferramentas**.

Personalização

Há dois parâmetros de personalização para esta ferramenta:

- O **indicador de detecção automática**, armazenado nas **preferências do utilizador**.
- Os pares não duplicados, armazenados na **nota especial**.

Revision #17
Created 5 September 2020 07:30:39 by pedro

Updated 23 February 2026 06:03:55 by pedro