

Fusion de doublons

Cet outil de Fusion des Doublons détecte les doublons et vous permet de les fusionner automatiquement ou manuellement s'il s'agit de vrais doublons, ou de les confirmer comme non doublons sinon.

Il existe trois façons de rechercher et de fusionner les doublons. Elles sont expliquées dans cette page.

- **Recherche globale** à travers toute la généalogie
- **Détection automatique** à chaque fois qu'une entité est modifiée
- **Action manuelle de fusion** par l'utilisateur pour forcer la fusion de deux entités sélectionnées
- **Glisser-Déposer** d'entités d'une généalogie à une autre

La fenêtre suivante est utilisée dans toutes ces situations.

Score : 129% - Doubleton 1 sur 8

Des doublons potentiels ont été identifiés. Vous pouvez les parcourir ici et décider de les fusionner ou de confirmer que ce ne sont pas des doublons.

Les propriétés en bleu et gris sont identiques pour les deux entités. Les bleues seront conservées en cas de fusion. Les propriétés en bleu des deux côtés sont des doublons probables.

Les propriétés en rouge sont différentes pour les deux entités. Cochez la case de celles à conserver en cas de fusion, ou sinon appuyer sur la baguette magique pour laisser Ancestris choisir et faire la fusion pour vous.

	I218 = DE PENTHIÈVRE, Adèle (*1035 +apr 1056)	I220 = de PENTHIÈVRE, Adèle (*env 1035 +apr 1056)
Individu n°	I218	I220
Nom	<input checked="" type="checkbox"/> DE PENTHIÈVRE, Adèle	<input type="checkbox"/> de PENTHIÈVRE, Adèle
Prénom	<input checked="" type="checkbox"/> Adèle	<input type="checkbox"/> Adèle
Préfixe	<input checked="" type="checkbox"/> DE	<input type="checkbox"/> de
Nom	<input checked="" type="checkbox"/> PENTHIÈVRE	<input type="checkbox"/> PENTHIÈVRE
Sexe	<input checked="" type="checkbox"/> Inconnu	<input type="checkbox"/> Féminin
Parents	<input checked="" type="checkbox"/> F107	<input type="checkbox"/> F109
Père	de PENTHIÈVRE, Éon (*+) (I217)	de PENTHIÈVRE, Éon (*+) (I217)
Mère	-	de CORNOUAILLES, Orguen ou Agnès (*+) (I216)
Mariage	-	Date: env 1035
Frère/Sœur [Les 2 entités]	DE PENTHIÈVRE, Adèle (*1035 +apr 1056) (I218)	de PENTHIÈVRE, Adèle (*env 1035 +apr 1056) (I220)
Frère	-	de PENTHIÈVRE, Alain (*1040 +4 août 1093) (I223)
Frère	-	de PENTHIÈVRE, Étienne (*env 1055 +1137) (I228)
Frère	-	PENTHIÈVRE, Alain (*+1098) (I227)
Frère	-	de PENTHIÈVRE, Brian (*+1084) (I222)
Frère	-	de PENTHIÈVRE, Geoffroy (*+1093) (I221)
Frère	-	de PENTHIÈVRE, Guillaume (*+apr 1055) (I224)
Frère	-	de PENTHIÈVRE, Richard (*+apr 1056) (I226)
Frère	-	de PENTHIÈVRE, Robert (*+apr 1055) (I225)
Changement	29 mars 2016, 00:16:46	29 mars 2016, 01:16:31
Naissance	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Date	<input checked="" type="checkbox"/> 1035	<input type="checkbox"/> env 1035
Décès	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Date	<input checked="" type="checkbox"/> apr 1056	<input type="checkbox"/> apr 1056
Numéro	<input checked="" type="checkbox"/> 4239384-a.1 G21	<input type="checkbox"/> 4239384-c.7 G21
Changement	4 fév 2018, 20:42:26	4 fév 2018, 20:42:26

Description

L'outil de Fusion des Doublons est composé de 3 éléments.

- **Fenêtre des Critères de Détection**: c'est ici que vous pouvez définir certains des critères qui détermineront si deux entités doivent être considérées comme des doublons ou non.
- **Fenêtre de Fusion**: c'est ici que vous verrez tous les doublons potentiels et que vous déciderez de les fusionner ou non.
- **Note spéciale des non doublons**: c'est ici qu'Ancestris stocke toutes vos confirmations que deux entités ne sont pas des doublons.

Avant de décrire ces 3 éléments, présentons d'abord comment Ancestris mesure la probabilité que 2 entités soient des doublons ou pas. Ancestris calcule un Score de ressemblance.

Score de ressemblance

Il est souvent difficile d'affirmer avec une complète certitude que deux entités sont des doublons ou ne sont pas des doublons. Même un être humain peut parfois avoir des difficultés à certifier que deux individus ou entités sont certainement les mêmes ou certainement pas les mêmes.

Bien sûr, il serait plus facile de limiter la détection en disant que deux personnes ayant exactement le même nom, le même prénom et la même date de naissance sont des doublons. En réalité, les dates peuvent être manquantes ou approximatives, les prénoms peuvent être dans un ordre différent ou incomplets, etc. Dans ces cas, vous voudriez quand même qu'Ancestris puisse détecter quelque chose.

C'est pourquoi **Ancestris utilise un Score de ressemblance**. Plus certaines informations sont similaires, plus il est probable que les entités soient des doublons.

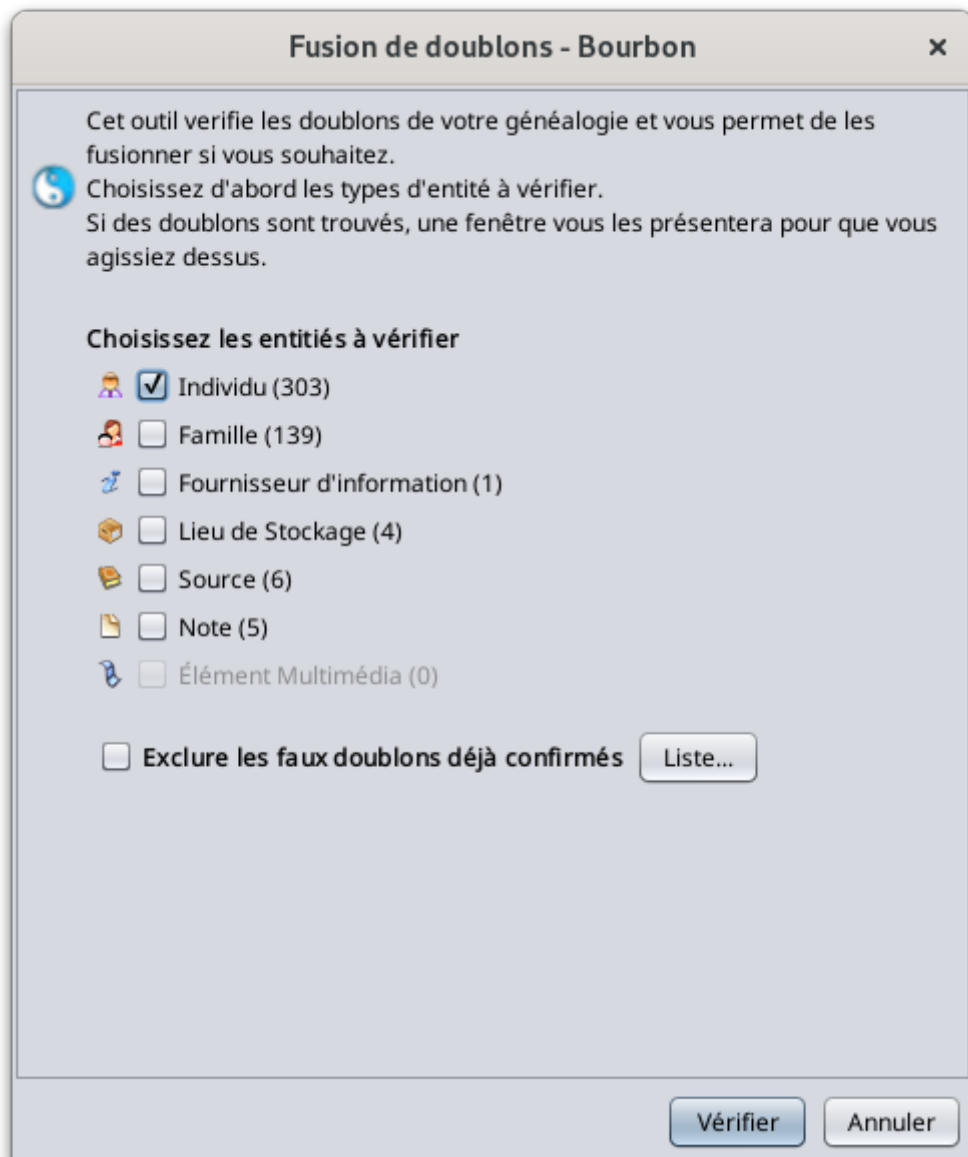
Les scores calculés peuvent avoir des valeurs négatives jusqu'à des valeurs élevées. Ils sont exprimés en pourcentage, même si ce score peut dépasser 100%.

Ancestris dresse ensuite la liste des doublons potentiels en fonction de ce score dans l'intention de vous dire : "***Bien que cela ne soit pas certain, étant donné les similitudes dans les informations entre ces deux personnes, il est possible qu'elles soient des doublons. Et voici le niveau de confiance qu'il y a entre eux***". Ensuite, c'est à vous de décider de fusionner ou d'écarter la ressemblance.

Cela signifie qu'Ancestris peut vous montrer des doublons potentiels que vous considérerez comme non doublons, et inversement, Ancestris peut ne pas trouver certains doublons qui pourraient en fait être de vrais doublons selon vous.
Veuillez accepter nos excuses si le détecteur n'est pas parfait et veuillez nous faire savoir si vous trouvez de tels cas.

Fenêtre de sélection des entités

Lorsqu'une recherche globale de doublons est lancée, la fenêtre des critères de détection s'affiche avec un certain nombre de critères.



Cochez les cases des entités pour lesquelles vous souhaitez rechercher des doublons.

Seules les cases des entités présentes dans le fichier Gedcom sont activées. Dans l'exemple ci-dessus, comme il n'y a pas d'entités médias.

Si **vous souhaitez exclure de la recherche les doublons connus**, c'est-à-dire ceux que vous avez déjà confirmés comme n'étant pas des doublons, cochez la case correspondante.

Pour voir la liste des paires d'entités que vous avez confirmées comme n'étant pas des doublons, **cliquez sur le bouton "Liste..."**.

Si vous appuyez sur le bouton "Liste...", la **liste des entités non doublons** s'affiche.

Fenêtre de Fusion

Pour les entités comparées en tant que doublons, la fenêtre suivante est utilisée.

Probabilité : 75% - Doubleton 1 sur 1

Des doublons potentiels ont été identifiés.

Vous pouvez les parcourir ici et décider de les fusionner ou de confirmer que ce ne sont pas des doublons.

I218 = DE PENTHIÈVRE, Adèle (°1035 +apr 1056)

I220 = de PENTHIÈVRE, Adèle (°env 1035 +apr 1056)

Individu n°	I218	I220
Nom	<input checked="" type="checkbox"/> DE PENTHIÈVRE, Adèle	<input type="checkbox"/> de PENTHIÈVRE, Adèle
Prénom	<input checked="" type="checkbox"/> Adèle	<input type="checkbox"/> Adèle
Préfixe	<input checked="" type="checkbox"/> DE	<input type="checkbox"/> de
Nom	<input checked="" type="checkbox"/> PENTHIÈVRE	<input type="checkbox"/> PENTHIÈVRE
Sexe	<input checked="" type="checkbox"/> Inconnu	<input type="checkbox"/> Féminin
Parents	<input checked="" type="checkbox"/> de PENTHIÈVRE, Éon (° +) (I217) x (F107)	<input type="checkbox"/> de PENTHIÈVRE, Éon (° +) (I217) x de CORNOUAILLES, Orguen ou Agnès (° +) (I216) (F109)
Mari	<input type="checkbox"/> de PENTHIÈVRE, Éon (° +) (I217)	<input type="checkbox"/> de PENTHIÈVRE, Éon (° +) (I217)
Enfant	<input type="checkbox"/> DE PENTHIÈVRE, Adèle (°1035 +apr 1056) (I218)	<input type="checkbox"/> de PENTHIÈVRE, Adèle (°env 1035 +apr 1056) (I220)
Changement	<input type="checkbox"/> 29 mars 2016, 00:16:46	<input type="checkbox"/> 29 mars 2016, 01:16:31
Epouse	<input type="checkbox"/> -	<input type="checkbox"/> de CORNOUAILLES, Orguen ou Agnès (° +) (I216)
Mariage	<input type="checkbox"/> -	<input type="checkbox"/> Date: env 1035
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Geoffroy (° +1093) (I221)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Brian (° +1084) (I222)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Alain (°1040 +4 août 1093) (I223)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Guillaume (° +apr 1055) (I224)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Robert (° +apr 1055) (I225)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Richard (° +apr 1056) (I226)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> PENTHIÈVRE, Alain (° +1098) (I227)
Enfant	<input type="checkbox"/> -	<input type="checkbox"/> de PENTHIÈVRE, Étienne (°env 1055 +1137) (I228)
Naissance	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Date	<input checked="" type="checkbox"/> 1035	<input type="checkbox"/> env 1035

Cette fenêtre affiche un par un la **liste totale de tous les doublons potentiels dont la probabilité est supérieure à 50%**.

La liste est triée **de la paire de doublons la plus certaine à la paire de doublons la**

moins certaine, par catégorie d'entité. Pour chaque paire d'entités similaires, Ancestris vous donne un pourcentage de probabilité.

Les grandes généalogies peuvent comporter plusieurs milliers de doublons. Cette fenêtre n'affichera que les 10 000 premiers doublons pour chaque type d'entité. Si votre généalogie comporte plus de doublons, vous devrez les fusionner avant de voir les autres.

Le **titre de** la fenêtre indique le numéro de la paire de doublons affichée et le score de ressemblance que les deux entités de cette paire sont bien les mêmes, et donc à fusionner.

Un **message général** est affiché et dépend de la situation dans laquelle l'outil a été lancé : recherche globale, détection automatique ou action manuelle.

Chaque paire de doublons constituée de deux entités est affichée dans les deux colonnes.

En titre de chaque colonne, un bouton permet de sélectionner chacune des entités dans les éditeurs pour plus de détails.

Dans chaque colonne sont affichées les propriétés de chacune des entités du doublon supposé.

- Les valeurs qui sont différentes sont affichées en rouge.
- Les valeurs identiques sont affichées en bleu pour l'entité de gauche, en gris pour l'entité de droite.
- Les valeurs qui se ressemblent sont affichées en bleu dans les deux colonnes.
-

L'objectif de la comparaison est de fusionner l'entité de droite dans celle de gauche si vous confirmez qu'elles sont dupliquées.

Par conséquent, une case à cocher est disponible pour chaque propriété sur le côté droit pour indiquer à Ancestris de conserver les informations sélectionnées de chaque entité après les avoir fusionnées.

Les boutons permettent de naviguer dans la liste des doublons potentiels, de reporter la décision, de les fusionner maintenant ou de confirmer maintenant qu'elles ne sont pas en double.

Barre de boutons



Champ Recherche de doublon

Ce champ permet de rechercher un doublon dans la liste. Saisissez le texte et appuyez sur Entrée. Utiliser ensuite les boutons Précédent et Suivant pour se positionner sur le suivant ou le prochain doublon correspondant.

Bouton aller au début



Affiche le premier doublon de la liste dans l'ordre de l'indice de confiance, c'est-à-dire le doublon le plus probable, ou le doublon qui se trouve 100 positions avant le doublon actuel s'il y a plus de 100 doublons dans la liste.

Bouton aller au précédent



Affiche le doublon précédent. Si un texte de recherche existe dans le champ Rechercher de doublon, il affichera le doublon précédent correspondant à ce critère de recherche.

Bouton retourner les entités



Intervertit les entités de gauche et de droite afin de fusionner ensuite les deux entités sur celle de droite. Cette opération est utile si la plupart des informations à conserver après la fusion se trouvent sur le côté droit.

Bouton aller au suivant



Affiche le doublon suivant. Si un texte de recherche existe dans le champ Rechercher de doublon, il affichera le doublon suivant correspondant à ce critère de recherche.

Bouton aller au dernier



Affiche le dernier doublon de la liste dans l'indice de confiance, donc le doublon le moins probable, ou le doublon qui se trouve 50 positions après le doublon actuel s'il y a plus de 100 doublons dans la liste.

Bouton retirer ce doublon de la liste



Supprime le doublon potentiel de la liste affichée.

Cette option est utile si vous ne savez pas encore si les deux entités sont des doublons ou non, et si vous souhaitez **reporter la décision**.

Si une nouvelle recherche globale de doublons est lancée, le doublon réapparaîtra.

Bouton Fusionner à la main



En cliquant sur le bouton Fusionner, les **entités seront fusionnées en utilisant les cases à cocher indiquées**.

L'entité de droite est supprimée du fichier Gedcom et les informations dont la case est cochée à droite sont ajoutées à, ou bien remplacent, celles de l'entité de gauche.

Pour les informations qui ne peuvent exister qu'une seule fois, il n'est possible de conserver les informations que dans l'une des deux entités.

Dès que la fusion est effectuée, la fenêtre affiche le même doublon avec le résultat de la fusion afin que vous puissiez vérifier que tout a été conservé comme vous le souhaitez.

Vous pouvez ensuite passer au doublon suivant.

Bouton Fusionner automatiquement



En cliquant sur le bouton Fusionner, les **entités seront fusionnées automatiquement par Ancestris sans utiliser les cases à cocher**.

Ancestris déterminera quelles informations de l'entité à supprimer doivent être conservées et enrichira les informations de l'entité à conserver.

Ancestris vous permet d'effectuer cette fusion automatique pour 3 périmètres différents. Lorsque vous appuyez sur le bouton Fusion automatique, les choix suivants apparaissent.

Fusion automatique [X]

Choisir les entités à fusionner

- ☒ Entités de ce doublon uniquement
- ☐ Entités de la recherche effectuée seulement
- ☐ Toute la généalogie

Saisir le seuil de score au delà duquel fusionner les entités

OK Annuler

Les choix sont les suivants

- **Entités de ce doublon uniquement :**
 - Seul le doublon actuellement affiché sera fusionné.
 - Ancestris détectera les informations à conserver sur l'entité de droite pour enrichir l'entité de gauche.

- Ensuite, le doublon sera mis à jour et n'affichera que l'entité de gauche en texte vert avec les informations résultantes.
- La fenêtre de sortie (Ctrl+T) affiche les entités fusionnées et le score. Vous pouvez analyser ce fichier et le sauvegarder comme référence.

- **Entités de la recherche effectuée uniquement :**

- La fusion automatique décrite ci-dessus sera effectuée pour tous les doublons de la liste dont le score est supérieur à un certain seuil, que vous devrez spécifier dans le champ en dessous.
- Seuls les 10 000 premiers doublons de chaque entité seront pris en compte. Utilisez le choix décrit ci-dessous si vous souhaitez fusionner toute la généalogie, et pas seulement les 10 000 premiers doublons trouvés.
- La fenêtre de résultat comprendra tous les doublons fusionnés au-dessus du score indiqué, ainsi que tous les autres doublons non fusionnés.
- Tous les doublons comprenant des entités supprimées seront supprimés de la liste.
- La fenêtre de sortie (Ctrl+T) répertorie toutes les entités fusionnées. Vous pouvez analyser et enregistrer ce fichier de sortie comme référence.

- **Toute la généalogie :**

- La fusion automatique décrite ci-dessus sera effectuée pour tous les doublons de la généalogie au-delà d'un score donné, et pas seulement pour les 10 000 premiers. Vous devrez spécifier ce score dans le champ en dessous.
- La fenêtre de résultat disparaîtra car il se peut qu'elle n'ait pas inclus tous les doublons trouvés.
- La fenêtre de sortie (Ctrl+T) affichera la liste de tous les doublons trouvés.

Bouton Non doublon



Ce bouton **exclut la paire d'entités des doublons potentiels**.

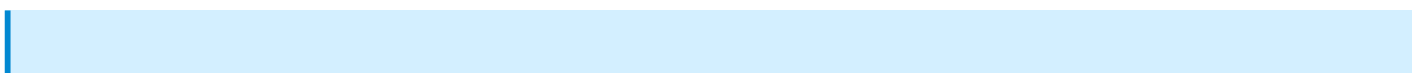
Il marque la paire d'entités comme n'étant pas en double et enregistre cette confirmation dans la **note spéciale "non duplicates"**.

Bouton Fermer



Ferme la fenêtre.

Note spéciale pour les non doublons

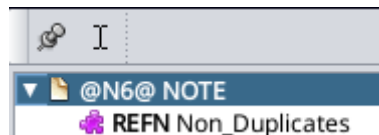


Une note spéciale est créée et mise à jour dans Ancestris pour stocker les confirmations non dupliquées.

Cette note stocke vos confirmations de paires similaires qui ne sont en fait pas des doublons selon vous.

Cela évite à Ancestris de les détecter encore et encore à chaque fois que la recherche globale ou la détection automatique est lancée.

Cette note a un nom de référence appelé "Non_Duplicates".



Cette note est mise à jour chaque fois que vous appuyez sur le bouton "Non Doublon" dans la **fenêtre Fusion** ou que vous mettez à jour la liste à l'aide de la **fenêtre Critères**.

Nous avons choisi de stocker ces informations utilisateur dans le fichier Gedcom lui-même parce que nous voulions que vos efforts pour analyser les entités et décider qu'elles ne sont pas des doublons constituent une information généalogique précieuse qui doit être conservée et transférée en tant que faisant partie du fichier Gedcom. La norme Gedcom ne répond pas à ce besoin, d'où le choix d'Ancestris de la stocker de cette manière, dans une seule note. Si la norme Gedcom venait à évoluer sur ce point, par exemple avec une balise NOALIAS, nous pourrions alors changer la façon dont ces informations sont gérées.

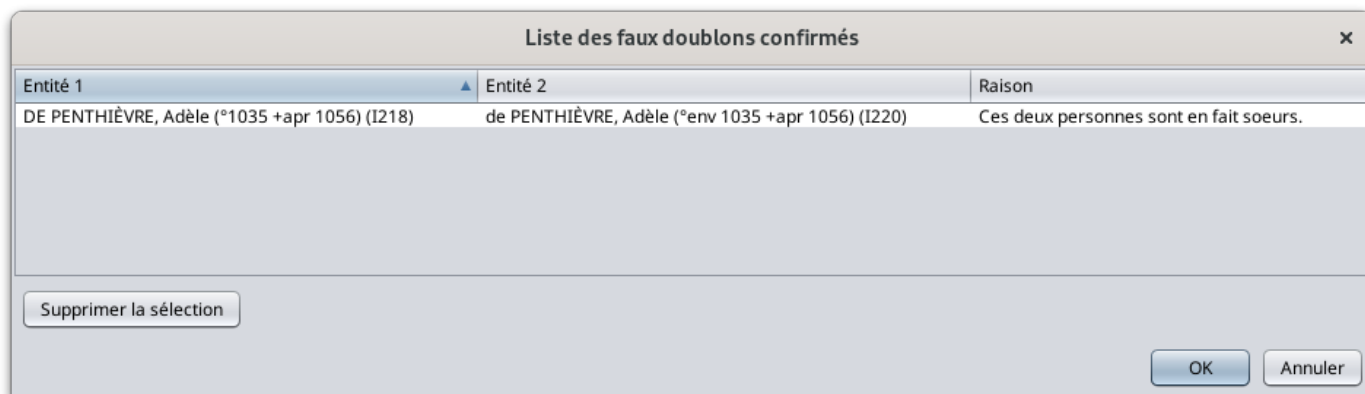
Vous pouvez consulter la liste des non doublons confirmés dans la **fenêtre des critères**.

Liste des Non Doublons

Vous pouvez voir la liste des entités non doublon en cliquant sur le bouton **"Afficher la liste" dans la fenêtre Critères de détection**.

Vous pouvez trier les lignes pour trouver les entités qui vous intéressent.

Vous pouvez sélectionner une ou plusieurs lignes et les supprimer de la liste si nécessaire.



Usage

Comme indiqué plus haut, il existe trois façons d'utiliser cet outil.

- **Recherche globale** sur l'ensemble du fichier généalogique
- **Détection automatique** à chaque modification d'une entité
- **Action de fusion manuelle** par l'utilisateur pour forcer la fusion de deux entités sélectionnées
- **Glisser-Déposer** d'entités d'une généalogie à une autre

Recherche globale

Le but de la recherche globale est à la fois d'**identifier les doublons dans l'ensemble de la généalogie et d'agir sur eux**, c'est-à-dire de décider un par un ce que vous voulez faire sur chacun d'eux, ou laisser Ancestris effectuer une fusion d'ensemble pour vous au delà d'un score donné de ressemblance.

Pour chaque doublon, vous pourrez alors décider de

1. **fusionner** le doublon
2. déclarer qu'**il ne s'agit pas d'un doublon**,
3. ou de **reporter** la décision à plus tard.

Vous pouvez lancer la recherche globale à partir du menu Outils d'Ancestris.

L'outil de fusion des doublons fonctionne en deux étapes.

- Tout d'abord, vous spécifiez les **critères de détection** dans la fenêtre correspondante,
- Ensuite, vous choisissez comment **fusionner les doublons** dans la fenêtre Fusionner.

Lors de l'utilisation de l'outil, la généalogie est modifiée en conséquence

1. Les entités que vous décidez de fusionner sont fusionnées avec les informations que vous avez spécifiées de conserver.
2. Les entités que vous déclarez non dupliquées sont enregistrées dans la **note spéciale**.

Détection automatique

Le but de la détection automatique est de vous alerter au cas où l'entité que vous êtes en train de créer ou de modifier est un doublon potentiel d'une autre entité déjà existante dans votre généalogie.

La détection automatique des doublons est activée par défaut dans les **préférences d'Ancestris**.

Dès que vous validez votre saisie dans l'un des éditeurs, et si la case de préférence correspondante est cochée, la détection recherche automatiquement les doublons potentiels de l'entité en cours d'édition.

Tous les doublons potentiels sont alors présentés dans la **fenêtre Fusionner**, pour que vous puissiez décider de ce que vous voulez faire de ces doublons.

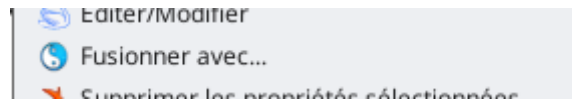
Dans le cas de l'éditeur Cygnus où plusieurs entités peuvent être éditées en même temps, toutes les entités modifiées sont vérifiées et Ancestris listera donc dans la fenêtre Fusion tous les doublons potentiels de toutes les entités modifiées.

Action de fusion manuelle

L'objectif de l'action Fusion manuelle est de fusionner deux entités, qu'Ancestris les ait détectées ou non.

Un autre objectif est d'identifier tous les doublons d'une entité donnée.

Cette action est accessible depuis le [menu contextuel](#) de l'entité en cours et que vous souhaitez fusionner avec une autre.



Lorsque cette action du menu contextuel est sélectionnée, Ancestris vous demande avec quelle autre entité l'entité actuelle doit être fusionnée.

- **Pour fusionner réellement avec une autre entité**, choisissez celle que vous pensez être un doublon.
- **Pour seulement savoir si l'entité courante a des doublons dans la généalogie**, il suffit de choisir n'importe quelle entité dans la liste des entités.

Ancestris affiche alors la fenêtre Fusionner avec une liste de doublons potentiels parmi lesquels se trouve la paire de deux entités que vous avez choisie et sa probabilité correspondante d'être la même entité, qui peut être de 0%.

La liste sera triée par probabilité décroissante et la paire de doublons affichée sera la paire d'entités que vous avez choisie.

Si l'entité dont vous êtes parti n'a pas d'autres doublons trouvés, seule la paire d'entités choisie sera affichée dans la liste.

Vous pouvez alors décider de fusionner l'entité actuelle dont vous êtes parti avec l'autre entité choisie, ou avec toute autre entité de la liste.

Glisser-Déposer d'une généalogie à une autre

Le but du glisser-déposer est de copier des entités d'une généalogie à une autre à l'aide de la souris. Après avoir déposé l'entité sur l'autre généalogie, Ancestris vous demandera comment vous voulez attacher ou ajouter l'entité et, dans le cas du choix Fusionner, la fenêtre de fusion apparaîtra. Voir la page correspondante [ici](#).

Personnalisation

Il existe 2 paramètres de personnalisation pour l'outil de fusion.

- L'**indicateur de détection automatique**, stocké dans les **préférences du répertoire utilisateurs**.
- Les **paires non doublon**, stockées dans la note spéciale.

Revision #18

Created 29 March 2020 19:23:18 by frederic

Updated 3 December 2024 07:56:52 by frederic