Logiciel

DuplicateContactsManager : Gestionnaire de contacts en double pour Thunderbird

Pré-requis

Installation

Configuration

Utilisation

Pour démarrer ce module :

- menu **Outils → Duplicate Contacts Manager…**
- ou, dans le carnet d'adresses, cliquez sur le bouton `Rechercher les doublons dans votre carnet d'adresses`.

duplicateContactsManager recherche dans les carnets d'adresses des paires d'entrées de contact (appelées **fiches**).

Il peut supprimer automatiquement les fiches qui ont les mêmes informations ou qui ont moins d'informations correspondantes.

Les paires de fiches correspondantes restantes sont présentées comme doublons pour un traitement manuel.

Chaque paire est présentée côte à côte avec une comparaison de tous les champs contenant des données, y compris une photo.

Certains champs importants sont toujours affichés pour être remplis s'ils sont vides.

Pour les paires de doublons présentées, des informations sont données dans la colonne qui les sépare, sous forme de symboles :

'≡'
: champs non vides avec des valeurs identiques
: les valeurs non identiques sont mises en évidence par la couleur.

'≃'
: noms, adresses e-mail ou numéros de téléphone correspondants.

'≅'
: fiches équivalentes, champs équivalents ou ensembles égaux (après abstraction).

'⊆' et '⊇'
: relation de sous-ensemble/sur-ensemble sur l'adhésion à la liste de diffusion, les adresses e-mail et les numéros de téléphone.

'⊊' et '⊋'
: un champ ou une carte entière contient moins ou plus d'informations que l'autre.

'<' et '>'
: comparaison sur des valeurs numériques ou relation sous-chaîne/super chaîne sur des noms et d'autres textes.

Pendant le traitement manuel d'une paire de fiches correspondantes, l'utilisateur peut les ignorer, modifier l'une ou les deux et décider de supprimer l'une d'entre elles.

Si une carte est supprimée, que son adresse e-mail principale est dans une ou plusieurs listes de diffusion et que l'autre carte n'a pas la même adresse e-mail principale, l'adresse est également supprimée des listes de diffusion respectives.

Pour empêcher des paires de fiches similaires d'être présentées à plusieurs reprises pour un traitement manuel, on peut leur attribuer des noms AIM différents, de sorte qu'elles n'apparaissent plus dans les résultats de recherche.

Entrées de contact semblables

La recherche de fiches semblables peut se faire :

1. sur **un seul carnet d'adresses** de **n** fiches : chaque fiche est comparée à toutes les autres → $n\frac{(n\text{-}1)}{2}$ comparaisons.
2. sur **deux carnets d'adresses différents**, respectivement de **m** et **n** fiches : chaque fiche du premier carnet est comparée avec chaque fiche du second → $n \times m$ comparaisons.

Les paires de doublons potentiels s'affichent.

Deux fiches sont considérées comme semblables si :

- Les **noms des fiches, adresses e-mail ou numéros de téléphone** correspondent
- ou que **aucun nom, adresse e-mail ou numéro de téléphone** ne correspond entre les deux fiches.

Mais les fiches dont le **pseudo AIM** diffère ne correspondent jamais, ce qui est pratique pour empêcher que les fiches soient présentées plusieurs fois pour un traitement manuel.

Avant la recherche de correspondance, les noms sont complétés et leur ordre est normalisé - par exemple :

- si deux parties de nom sont détectées dans le **Nom à**

afficher (par exemple, “John Doe”) ou dans une adresse e-mail (par exemple, “John.Doe@company.com”), ils sont pris comme prénom et nom.

- Dans une fiche, plusieurs adresses e-mail ou plusieurs numéros de téléphone sont traités comme des ensembles, c'est-à-dire que leur ordre et leurs types sont ignorés.
- Les noms ne différant que par la casse sont considérés comme identiques.

- Les **noms** de deux fiches correspondent si:
  - leur **Nom à afficher** est non vide et équivalent
  - ou les paires **prénom** et **nom** sont non vides et équivalentes
  - ou la paire **Nom à afficher** est vide mais les paires **prénom** et **nom** sont non vides et équivalentes
  - dans une fiche, le **Nom à afficher** est vide et le **prénom** ou le **nom** sont non vides et équivalents au **Nom à afficher** de l'autre fiche
  - leur **pseudo AIM** est non vide et équivalent

2. Les **adresses e-mail** de deux fiches correspondent si l'une de leurs adresses e-mail principale ou secondaire est équivalente.
3. Les **numéros de téléphone** de deux fiches correspondent si l'un de leurs numéros **Mobile**, **Work** ou **Pager** est équivalent. Les champs **Home** et **Fax** ne sont pas pris en compte car ils sont souvent partagés par un groupe de personnes.

Normalisation des valeurs de champ

Avant de comparer les champs de la fiche, leurs valeurs sont normalisées comme suit :

1. **Élagage** (suppression du contenu non pertinent pour la comparaison) :
   1. **ignorer les valeurs de certains types de champs** (liste des champs ignorés configurable ; par défaut : UID, UUID, CardUID, groupDavKey, groupDavVersion, groupDavVersionPrev, RecordKey, DbRowID, PhotoType, PhotoName, LowercasePrimaryEmail, LowercaseSecondEmail, unprocessed:rev, unprocessed:x-ablabel)
   2. **nettoyage des numéros de téléphone** (suppression des espaces blancs et des caractères non numériques)
   3. **suppression des doublons erronés d'adresses électroniques**, insérés comme **nom par défaut** par certains clients de messagerie
   4. **remplacement de @googlemail.com par @gmail.com** dans les adresses e-mail
2. **Transformation** (réorganisation des informations pour une meilleure comparaison)
   1. **correction l'ordre du prénom et du nom** de famille (par exemple, réorganiser “Doe, John”)
   2. **déplacement du nom de famille vers le prénom** les initiales du milieu telles que “M”
   3. **déplacement vers le nom de famille des préfixes** tels que “von”.

3. **Normalisation** (harmonisation des variantes d'écriture) :
    1. **conversion en minuscules** (sauf pour la partie nom des adresses e-mail AOL)
    2. transcription des umlauts et des ligatures
    3. **remplacement, si configuré, dans les numéros de téléphone le préfixe** de l'appel international (IDD) (tel que '00') par '+' et le préfixe interurbain national (tel que '0') par le code d'appel par défaut du pays (tel que '+49')
4. **Simplification** (suppression des informations moins pertinentes)
    1. **suppression des accents et de la ponctuation**
    2. **suppression des chiffres et lettres isolés** (comme les initiales)

Les champs correspondants de deux fiches sont considérés comme équivalents si leurs valeurs normalisées sont égales.

Les adaptations ci-dessus ne sont calculées que pour la comparaison : les champs réels de la fiche ne sont pas modifiés

Si la suppression automatique est activée, seules les fiches proposées pour la suppression (informations équivalentes ou inférieures à celles d'une autre fiche) sont supprimées.

Quand une paire de fiches est présentée pour inspection manuelle, la fiche signalée par une couleur rouge pour retrait est :

- celle préférée pour la suppression
- ou celle utilisée le moins souvent (plus petit indice de popularité), si les fiches ne sont pas comparables,
- ou celle modifiée ou créée plus tôt (plus petit LastModifiedDate)
- ou celle du deuxième carnet d'adresses ou de celui trouvé plus tard si les deux carnets d'adresses sont identiques.

Équivalence des informations

Une fiche est considérée comme ayant des informations équivalentes ou inférieures à une autre si pour chaque champ non ignoré :

- le champ est équivalent au champ correspondant de l'autre fiche,
- sinon c'est un ensemble et sa valeur est un sous-ensemble de la valeur du champ correspondant de l'autre fiche,
- ou il s'agit du Prénom, Nom ou Nom à afficher et sa valeur est une sous-chaîne de la valeur de champ correspondant de l'autre fiche,
- ou il s'agit du PopularitéIndex ou de LastModifiedDate (ignorés ici), ou
- il a la valeur par défaut (vide pour les champs de texte, 0 pour les champs numériques ou false pour les champs booléens).

Pour la comparaison par champ ci-dessus, les adresses e-mail d'une fiche sont traitées comme un ensemble, les numéros de téléphone d'une carte sont également traités comme un ensemble et l'ensemble des noms des listes de diffusion auxquelles appartient une fiche est considéré comme un champ supplémentaire.

Une fiche contenant des informations équivalentes ou inférieures à une autre est préférée pour la suppression si :

- tous les champs non ignorés ne sont pas équivalents (ils contiennent moins d'informations),
- sinon le poids des caractères de la fiche est plus petit, c'est-à-dire que ses valeurs de champ élaguées et transformées (non ignorées) ont un nombre total égal ou inférieur de lettres majuscules et de caractères spéciaux que l'autre fiche, ou bien le poids des caractères est égal
- et son indice de popularité est plus petit,
- sinon LastModifiedDate est plus petit.

Voici un exemple.

La fiche de droite sera préférée pour la suppression car elle contient moins d'informations.

NickName	“Péte”	“ pete ! ”	accent, ponctuation, majuscule et espace ignorés
FirstName	“Peter”	“Peter Y van”	préfixe du nom “van” déplacé vers le nom de famille, initiale “Y” au milieu ignorée
LastName	“van Müller”	“Mueller”	préfixe du nom “van” déplacé vers le nom de famille, umlauts transcrits
DisplayName	“Hans Peter van Müller”	“van Müller, Peter”	prénom déplacé vers l'avant, le nom est une sous-chaîne
PreferDisplayName	'yes'	'yes'	même valeur true
AimScreenName	“”	“”	même nom AIM
PreferMailFormat	'HTML'	'unknown'	valeur par défaut («inconnu») considéré comme moins d'informations

PrimaryEmail	“Peter.vanMueller@company.com”	“P.van.Mueller@gmx.de”	e-mails traités comme des ensembles, casse de lettres ignorée
SecondaryEmail	“p.van.mueller@gmx.de”	“”	e-mails traités comme des ensembles, casse de lettres ignorée
WorkPhone	“089/1234-5678”	“+49 89 12345678”	préfixe national normalisé et caractères non numériques
PopularityIndex	5	3	champ ignoré pour la comparaison des informations
LastModifiedDate	2018-02-25 07:51:28	2018-02-25 08:30:37	champ ignoré pour la comparaison des informations
UUID	“”	“903a61be-64d5-4844-802a”	champ ignoré

Variables de configuration

Les préférences **options/configuration/préférences** utilisées par cette extension de Thunderbird sont enregistrées dans des clés de configuration commençant par **extensions.DuplicateContactsManager**. - par exemple, la liste des champs ignorés est stockée dans la variable **ignoreFields**.

Désinstallation

Problèmes connus

Voir aussi

- **(en)** https://addons.thunderbird.net/fr/thunderbird/addon/duplicate-contact-manager/

Basé sur « Duplicate Contact Manager » par David von Oheimb.

From:
https://doc.wikis.frapp.fr/ - **doc**

Permanent link:
https://doc.wikis.frapp.fr/doku.php?id=logiciel:reseau:mail:thunderbird:adresses:doublons:start

Last update: **2024/09/21 10:47**