



Duplicate Finder pro

Dubletten finden und auflösen



Copyright

Die hier enthaltenen Angaben und Daten können ohne vorherige Ankündigung geändert werden. Die in den Beispielen verwendeten Namen und Daten sind frei erfunden, soweit nichts anderes angegeben ist. Ohne ausdrückliche schriftliche Erlaubnis der CAS Software AG darf kein Teil dieser Unterlagen für irgendwelche Zwecke vervielfältigt oder übertragen werden, unabhängig davon, auf welche Art und Weise oder mit welchen Mitteln, elektronisch oder mechanisch, dies geschieht.

© 2015 CAS Software AG. Alle Rechte vorbehalten.

CAS-Weg 1 - 5, 76131 Karlsruhe, www.cas.de

Sämtliche erwähnten Kennzeichen stehen ausschließlich den jeweiligen Inhabern zu.

Einschränkung der Gewährleistung

Für die Richtigkeit des Inhalts wird keine Garantie übernommen. Für Hinweise auf Fehler sind wir jederzeit dankbar.

Stand: Januar 2015

Inhalt

1 Grundsätzliches	4
1.1 Lizenzen.....	4
1.2 Voraussetzungen	4
2 Duplicate Finder pro starten	5
3 Dubletten suchen und bereinigen	5
3.1 Dubletten auflösen.....	7
4 Duplicate Finder pro konfigurieren.....	8
4.1 Suchprofil bearbeiten.....	10

1 Grundsätzliches

Mit dem Modul Duplicate Finder pro finden und bereinigen Sie doppelt erfasste Daten.

Duplicate Finder pro findet drei Arten von Dubletten, die jeweils in einem eigenen Suchlauf ermittelt werden:

- Doppelt erfasste Firmenadressen sowie Einzelkontakte, die Duplikate von Firmen sind
- Einzelkontakte zu Firmen mit Ansprechpartnern
- Doppelt erfasste Ansprechpartner und Einzelkontakte

In der Grundeinstellung werden alle Adressen durchsucht. Die Auswahl können Sie über einen Filter einschränken. Über einen weiteren Filter können Sie eine Teilmenge von Adressen angeben, zu denen Dubletten gefunden werden sollen.

Duplicate Finder pro startet mit dem Standard-Suchprofil. Wenn Sie eigene Suchkriterien verwenden möchten, können Sie das Standard-Suchprofil duplizieren und Ihre eigenen Einstellungen festlegen.

- Jedes Feld wird mit einem Wert gewichtet, der bei Übereinstimmung des Felds bei zwei Adressen in die Gesamtgewichtung einfließt. Unterscheiden sich die Adressen in einem Feld, fließt für dieses der Wert 0 in die Gesamtgewichtung ein.

Nur wenn die Summe der Feldgewichtungen den angegebenen Schwellenwert übersteigt, werden zwei Adressen als Dubletten erkannt.

- Für jedes Feld legen Sie fest, ob identisch oder phonetisch verglichen wird. Beispielsweise wird bei einem phonetischen Vergleich im Feld **Name** die Adresse "Meier" als Dublette von "Maier" erkannt und das Feld **Name** entsprechend gewichtet.
- Für einzelne Felder können Sie festlegen, dass sie auch dann wie identische Felder behandelt werden, wenn sie bei einer von zwei verglichenen Adressen leer sind.

1.1 Lizenzen

Für Duplicate Finder pro ist eine Lizenz notwendig.

Die Lizenz wird vom Administrator in der Management Konsole im Bereich **Lizenzen** eingetragen und für die entsprechenden Benutzer freigeschaltet.


1.2 Voraussetzungen

Benutzer benötigen für die Verwendung von Duplicate Finder pro folgende Rechte:

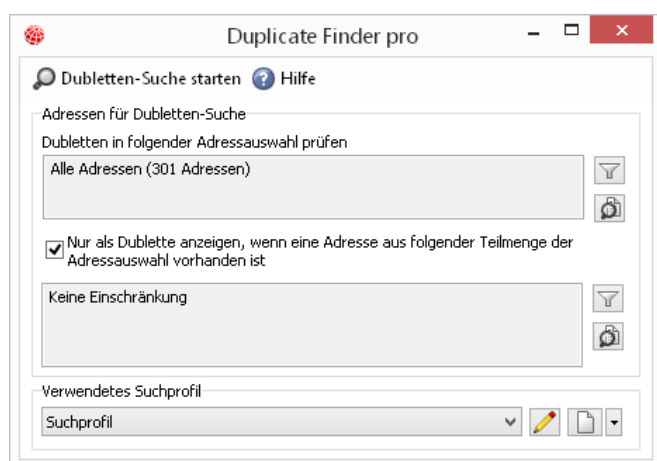
- Das weitere Recht **Dubletten auflösen**, damit die Funktion aufgerufen werden kann.

- Leserechte** auf alle zu prüfenden Adressfelder, damit die Dubletten ermittelt werden können.
- Bearbeitungsrechte** auf alle zu prüfenden Adressfelder sowie **Löschrechte** auf Adressen, damit die Dubletten aufgelöst werden können.

2 Duplicate Finder pro starten

-  Klicken Sie im Hauptfenster von CAS genesisWorld auf **Extras** und wählen Sie **Duplicate Finder pro**.

Das Fenster **Duplicate Finder pro** wird geöffnet.



In der Grundeinstellung werden alle Adressen durchsucht. Sie können die Auswahl über einen Filter einschränken, siehe Kapitel "Duplicate Finder pro konfigurieren" auf Seite 8.

3 Dubletten suchen und bereinigen

-  Klicken Sie auf **Dubletten-Suche starten**, um Dubletten zu finden.

Ablauf der Prüfung

Bei der Dubletten-Suche werden 3 Suchläufe durchgeführt. Die jeweiligen Suchläufe werden nur dann angezeigt, wenn Dubletten gefunden wurden.

Die Reihenfolge ist immer dieselbe. Zuerst kommt der Firmen-Suchlauf, dann der Einzelkontakte-Suchlauf und zum Schluss der Ansprechpartner-Suchlauf.

- Firmen-Suchlauf**

Beim Firmen-Suchlauf werden Firmen untereinander verglichen.

Außerdem werden Firmen mit den Einzelkontakten verglichen, bei denen das Feld **Name** leer und das Feld **Firma** ausgefüllt ist.

Einzelkontakte-Suchlauf

Beim Einzelkontakte-Suchlauf werden Firmen mit den Einzelkontakten verglichen, bei denen die Felder **Firma** und **Name** ausgefüllt sind.

Wenn zu einem Einzelkontakt eine Firma mit identischem Firmennamen gefunden wird, dann wird der Einzelkontakt als Dublette der Firma angezeigt. Sie können den Einzelkontakt dann über den Kontextmenü-Eintrag **Firmenbildung** zu einem Ansprechpartner der entsprechenden Firma umwandeln.

Ansprechpartner-Suchlauf

Beim Ansprechpartner-Suchlauf werden Ansprechpartner und Einzelkontakte untereinander verglichen.

Adressen, bei denen das Feld **Name** leer ist, werden nicht verglichen.

Gewichtung

Bei jedem Suchlauf wird die Übereinstimmung in Prozent angezeigt. Der Wert wird anhand der Feldgewichtungen ermittelt. Dabei werden die erreichten Punkte durch die erreichbaren Punkte geteilt und dann mit 100 multipliziert. Das Ergebnis ist die Übereinstimmung in Prozent.

Beispiel: Im Standard-Suchprofil sind maximal 840 Punkte erreichbar. Bei 420 erreichten Punkten beträgt die Übereinstimmung 50 %, die sich so berechnen:

$$420/840 = 0,5$$
$$0,5*100 = 50 \%$$

Für jedes Feld, das bei beiden Adressen übereinstimmt oder leer ist, gibt es die volle Punktzahl.

Felder, die nur bei einer Adresse ausgefüllt sind und bei denen im Feld **Leerbehandlung** die Option **Dublette erkennen** eingestellt ist, erhalten ebenfalls die volle Punktzahl.

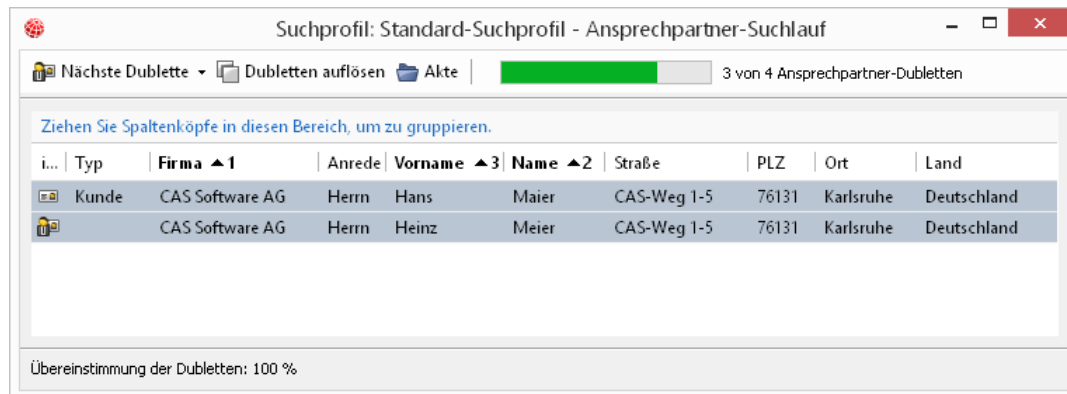
Felder mit unterschiedlichen und nicht leeren Werten erhalten keine Punkte.

Besonderheiten bei der Prüfung

- Deaktivierte Adressen werden nicht geprüft.
- Zu den Modulen Easy Invoice und ERP connect gehörende Adressen werden geprüft, siehe Kapitel "Duplicate Finder pro konfigurieren" auf Seite 8.
- Adressen, die von der Dublettenprüfung ausgeschlossen wurden, werden trotzdem gefunden.

Beim Speichern einer Adresse wird diese üblicherweise sofort auf Dubletten geprüft. Bei dieser Prüfung können Sie die Adresse von künftigen Dublettenprüfungen ausschließen. Diese Option bezieht sich allerdings nur auf die Sofortprüfung, über den Duplicate Finder pro werden diese Dubletten weiterhin gefunden.

Funktionen der Symbolleiste



- Beim Firmen-Suchlauf wechseln Sie mit **Nächste Dublette** zum nächsten Firmen-Vergleich oder zum Einzelkontakte-Suchlauf.
- Beim Einzelkontakte-Suchlauf wechseln Sie mit **Nächste Dublette** zum nächsten Einzelkontakte-Vergleich oder zum Ansprechpartner-Suchlauf.
- Beim Ansprechpartner-Suchlauf wechseln Sie mit **Nächste Dublette** zum nächsten Ansprechpartner-Vergleich.
- Mit **Dubletten auflösen** führen Sie die doppelten Adressen zusammen.
- Mit **Akte** blenden Sie die Akte ein oder aus.

3.1 Dubletten auflösen

- Klicken Sie auf **Dubletten auflösen**, um doppelte Adressen zusammenzuführen. Wenn keine Konflikte vorliegen, werden die Dubletten automatisch aufgelöst.

Die Felder der beiden Adressen werden zu einer Adresse zusammengefügt. Diese Zusammenführung schließt Verknüpfungen und eventuelle Teilnehmer mit ein.

Sollten Konflikte vorliegen, dann öffnet sich das Fenster **Dubletten auflösen**. In diesem Fenster sehen Sie alle Konflikte.

Konflikte sind z. B. unterschiedliche Schreibweisen von Namen, die im Rahmen einer phonetischen Suche als Duplikat erkannt wurden.



- Prüfen Sie die Feldwerte und übernehmen die jeweils richtigen Werte mit einem Klick auf **Wert übernehmen**.
- Wenn Sie alle Werte von einem Datensatz übernehmen möchten, dann klicken Sie auf **Alle übernehmen**.
- Klicken Sie auf **OK**.

Das Ergebnis wird im originalen Datensatz gespeichert und die Dublette wird gelöscht.

Bei der Zusammenführung der Adressen werden die Verknüpfungen beider Adressen ebenfalls zusammengeführt. Falls Adressen nicht öffentlich sind, werden auch die Teilnehmer und Rechte zusammengefasst.

4 Duplicate Finder pro konfigurieren

Im Fenster **Duplicate Finder pro** konfigurieren Sie die Dublettensuche über optionale Filter und das Suchprofil.

Adressauswahl einschränken

In der Grundeinstellung wird Ihr gesamter Adressbestand durchsucht. Sie können die Auswahl über einen Adressfilter einschränken.

- Klicken Sie hinter dem Feld **Dubletten in folgender Adressauswahl prüfen** auf die Schaltfläche **Filter** und legen Sie die entsprechenden Einstellungen fest.
- Klicken Sie auf die Schaltfläche mit dem Lupensymbol, um die gefilterten Datensätze zu prüfen.

Hinweis zu den Modulen Easy Invoice und ERP connect

Die folgenden Informationen sind für Sie wichtig, wenn Sie das Modul Easy Invoice oder das Modul ERP connect einsetzen.

In den Standard-Einstellungen werden auch Adressen geprüft, die zu den Modulen Easy Invoice und ERP connect gehören. Diese Dubletten sollten allerdings nicht aufgelöst werden, weil sonst inkonsistente Daten entstehen können.

Idealerweise schließen Sie Adressen von Easy Invoice und ERP connect per Filter aus, so dass die Duplikate gar nicht erst gefunden werden. Stellen Sie dazu im Feld **Dubletten in folgender Adressauswahl prüfen** folgende Filterkriterien ein:

- Easy Invoice: Kundennummer ist leer
- ERP Connect: für Abgleich definiertes Nummernfeld ist leer

Ergebnisliste einschränken

Wählen Sie die Option **Nur als Dublette anzeigen, wenn eine Adresse aus folgender Teilmenge in der Adressauswahl vorhanden ist**, um nur bestimmte Adressen als Dubletten anzuzeigen.

☑ Klicken Sie auf die Schaltfläche **Filter** und wählen Sie die Adressen, zu denen Dubletten erkannt werden sollen.

Auf diese Weise können Sie z. B. importierte Adressen mit den bestehenden Adressen vergleichen, ohne dass alle Adressen in CAS genesisWorld miteinander verglichen werden.

🔍 Klicken Sie auf die Schaltfläche mit dem Lupensymbol, um die gefilterten Datensätze zu prüfen.

Suchprofil einstellen

Wählen Sie das **Suchprofil** für die Dublettensuche.

In der Grundeinstellung wird das **Standard-Suchprofil** verwendet. Das Standard-Suchprofil basiert auf den Einstellungen der Dublettenprüfung in der Management Konsole und kann in CAS genesisWorld nicht bearbeitet werden.

Die Einstellungen in der Management Konsole kann ein Administrator unter **Sonstiges/Adressen/Dublettenprüfung** anpassen.

✎ Klicken Sie auf **Bearbeiten**, um ein Suchprofil zu bearbeiten.


📄 Klicken Sie auf die Schaltfläche **Neu**, um ein neues Suchprofil anzulegen.

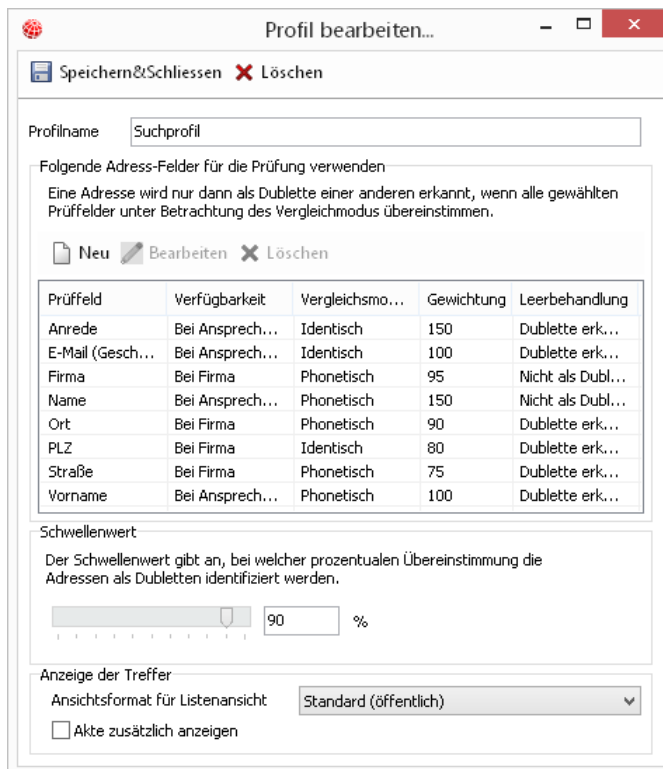
▾ Sie können das ausgewählte Suchprofil duplizieren, indem Sie auf den Pfeil hinter der Schaltfläche **Neu** klicken und **Duplizieren** wählen.

4.1 Suchprofil bearbeiten

Suchprofile werden öffentlich gespeichert. Jeder Benutzer mit dem Recht **Dubletten auflösen** kann alle Suchprofile sehen, neu anlegen, bearbeiten und löschen. Die einzige Ausnahme stellt das Standard-Suchprofil dar, das weder bearbeitet noch gelöscht werden kann.

Sie können beliebig viele Suchprofile speichern, in denen jeweils die gewünschten Suchkriterien genau festgelegt sind.

 Markieren Sie ein Suchprofil und klicken Sie auf **Bearbeiten**.



Profilname: Suchprofil

Folgende Adress-Felder für die Prüfung verwenden
Eine Adresse wird nur dann als Dublette einer anderen erkannt, wenn alle gewählten Prüffelder unter Betrachtung des Vergleichmodus übereinstimmen.

Neu Bearbeiten Löschen






Prüffeld	Verfügbarkeit	Vergleichsmo...	Gewichtung	Leerbehandlung
Anrede	Bei Ansprech...	Identisch	150	Dublette erk...
E-Mail (Gesch...	Bei Ansprech...	Identisch	100	Dublette erk...
Firma	Bei Firma	Phonetisch	95	Nicht als Dubl...
Name	Bei Ansprech...	Phonetisch	150	Nicht als Dubl...
Ort	Bei Firma	Phonetisch	90	Dublette erk...
PLZ	Bei Firma	Identisch	80	Dublette erk...
Straße	Bei Firma	Phonetisch	75	Dublette erk...
Vorname	Bei Ansprech...	Phonetisch	100	Dublette erk...

Schwellenwert
Der Schwellenwert gibt an, bei welcher prozentualen Übereinstimmung die Adressen als Dubletten identifiziert werden.

90 %

Anzeige der Treffer
Ansichtsformat für Listenansicht: Standard (öffentlich)
 Akte zusätzlich anzeigen

Funktionen im Fenster Profil bearbeiten

-  Klicken Sie auf **Neu**, um ein Feld hinzuzufügen.
-  Doppelklicken Sie auf ein Feld, um die Einstellungen des Feldes zu bearbeiten.
-  Alternativ können Sie das Feld auswählen und auf die Schaltfläche **Bearbeiten** klicken.
-  Markieren Sie ein Feld und klicken Sie auf **Löschen**, um das Feld zu entfernen.
-  Der **Schwellenwert** gibt an, bei welcher prozentualen Übereinstimmung die Adressen als Dubletten identifiziert werden.

Die prozentuale Übereinstimmung zweier verglichener Adressen ist der Anteil der Summe aller Feldgewichtungen an der maximal erreichbaren Gesamtgewichtung.

Die Gesamtgewichtung berechnet sich aus der Gewichtung der einzelnen Felder. Als Dubletten erkannte Felder fließen mit der angegebenen Gewichtung in die Gesamtgewichtung ein. Felder mit verschiedenen Inhalten werden mit dem Wert 0 gewichtet, siehe Kapitel "Dubletten suchen und bereinigen" auf Seite 5.

- Im Feld **Ansichtsformat für die Listenansicht** legen Sie fest, wie die gefundenen Dubletten angezeigt werden sollen.
- Aktivieren Sie die Option **Akte zusätzlich anzeigen**, wenn die Akte direkt angezeigt werden soll. Sie können die Akte außerdem im Fenster mit den Ergebnissen mit der Schaltfläche **Akte** ein- und ausblenden.

Feld hinzufügen oder bearbeiten

Wenn Sie ein Feld bearbeiten oder neu hinzufügen, können Sie die folgenden Einstellungen vornehmen.



- Im Feld **Prüffeld** wählen Sie das zu prüfende Feld.
- Das Feld **Verfügbarkeit** zeigt, bei welchen Adressen das Prüffeld verfügbar ist.
Das Feld **Verfügbarkeit** können Sie nicht direkt in CAS genesisWorld bearbeiten. Stattdessen nimmt ein Administrator diese Einstellungen in der Management Konsole unter **Sonstiges/Adressen/Feld-Synchronisation** vor.

- Im Feld **Vergleichsmodus** stellen Sie ein, nach welchem Modus die Feldinhalte verglichen werden. Verfügbar sind die Vergleichsmodi **Identisch** und **Phonetisch**, wobei **Identisch** die Voreinstellung für neue Felder ist.

Im Modus **Identisch** muss das Feld bei zwei Adressen zu 100% übereinstimmen, damit es mit der angegebenen Gewichtung einfließt. Die Groß- und Kleinschreibung spielt dabei keine Rolle. Beispielsweise ist die Schreibweise "Maier" identisch mit "MAIER" und mit "maier".

Im Modus **Phonetisch** muss die nach phonetischen Regeln ermittelte Aussprache des Felds bei zwei Adressen zu 100% übereinstimmen, damit es mit der angegebenen Gewichtung einfließt. Beispielsweise ist "Meyer" identisch mit "Maier".

- Im Feld **Gewichtung** legen Sie fest, wie das Feld gewichtet wird.

Jedes Feld wird mit dem angegebenen Wert gewichtet. Dieser Wert fließt bei Übereinstimmung der Felder bei zwei Adressen in die Gesamtgewichtung ein.

Unterscheiden sich die Adressen in einem Feld, fließt für dieses der Wert 0 in die Gesamtgewichtung ein.

Nur wenn die Summe der Feldgewichtungen den angegebenen Schwellenwert übersteigt, werden zwei Adressen als Dubletten erkannt.

- Im Feld **Leerbehandlung** geben Sie an, wie ein leeres Feld beim Vergleich mit einem ausgefüllten Feld gewertet wird.

Wenn Sie **Dublette erkennen** wählen, wird ein leeres Feld bei einem Vergleich wie ein identisches Feld behandelt und gewichtet.

Wenn Sie **Nicht als Dublette erkennen** wählen, wird ein leeres Feld als verschieden von dem Feld mit einem Wert gewertet.

Beispiel: Sie wählen für das Feld **E-Mail (Geschäftlich)** die Option **Dublette erkennen**. Somit werden Ansprechpartner auch bei einem leeren E-Mail-Feld als Dubletten erkannt, wenn die anderen Felder übereinstimmen.