



# Duplicate Finder pro

Finding and merging duplicates



Duplicate Finder pro was developed by itdesign GmbH, Tübingen - [www.itdesign.de](http://www.itdesign.de)

## **Copyright**

The information contained herein may be altered without prior notice. The names and data used in the examples are fictitious, except where otherwise stated. No part of this document may be reproduced or transmitted for any purposes whatsoever without the express written consent of CAS Software AG, irrespective of the manner or the means, electronic or mechanical, by which this occurs.

© 2015 CAS Software AG. All rights reserved.

CAS-Weg 1 - 5, 76131 Karlsruhe, Germany, [www.cas.de/en](http://www.cas.de/en)

All trademarks are the property of their respective owners.

## **Disclaimer**

No guarantee can be made for the accuracy of the content. Notification of errors would be appreciated.

**January 2015**

# Contents

<b>1</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Licenses.....	4
1.2	Requirements .....	5
<b>2</b>	<b>Starting Duplicate Finder pro .....</b>	<b>5</b>
<b>3</b>	<b>Searching and merging of duplicates .....</b>	<b>6</b>
3.1	Merging duplicates .....	8
<b>4</b>	<b>Configuring Duplicate Finder pro.....</b>	<b>9</b>
4.1	Editing search profiles.....	11

# 1 Introduction

---

The Duplicate Finder pro module helps you to find and merge redundant data which has been entered twice.

Duplicate Finder pro finds three types of duplicates which are checked in three separate search runs:

- Company addresses that were entered twice and individual contacts which are company duplicates.
- Individual contacts for companies with contact persons.
- Contact persons and individual contacts that were entered twice.

Using the default settings you search all addresses. Use a filter to restrict this selection. With another filter you can specify a subset of addresses for which you want to run a duplicate search.

Duplicate Finder pro starts its search with the default search profile. If you like to use your personal search criteria, you can duplicate the default search profile and define your personal settings.

- Each field is weighted with a value that is included in the total weight if one field of two addresses is the same. If one field value of the addresses differs, the value 0 is included in the total weight.

Only if the total of the field weightings exceeds the defined threshold value for duplication, the two addresses are detected as duplicates.

- You can define for each field whether the value should be compared identically or phonetically. For example, a phonetic comparison of the **Name** field results in the address for Smith being recognized as a duplicate of Smyth and the **Name** field is weighted accordingly.
- For individual fields, you can define that they are treated like identical fields if one field of the two compared addresses is empty.

## 1.1 Licenses

---

The Duplicate Finder pro module requires a separate license.

The license is entered in the Management Console, in the **Licenses** area, by the administrator and can be activated for specific users.

## 1.2 Requirements

---

In order to use Duplicate Finder pro users require the following rights:

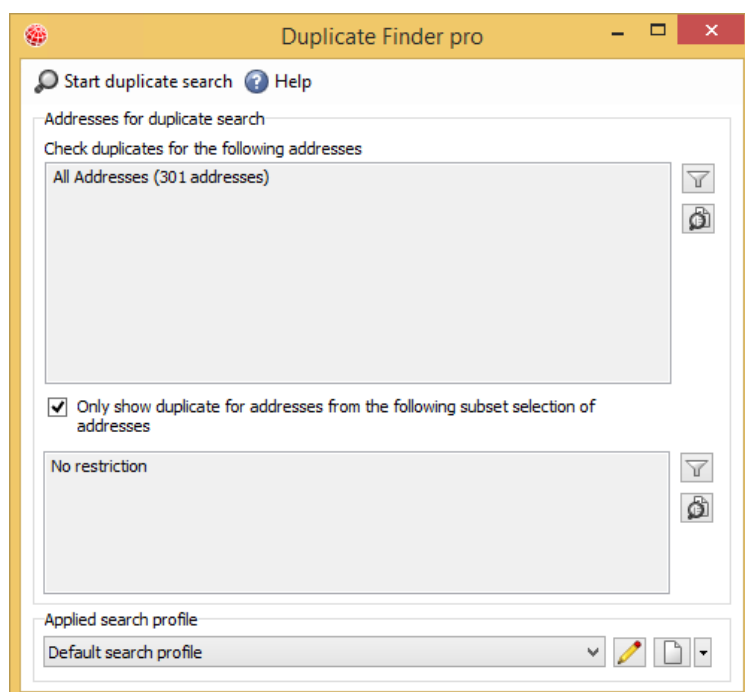
- The **Merge duplicates** right to be able to use the function.
- Read rights** on all address fields to find duplicates.
- Editing rights** on all address fields that have to be tested as well as **Deleting rights** on addresses to merge duplicates.

## 2 Starting Duplicate Finder pro

---

Click **Tools** in the main window of CAS genesisWorld and choose **Duplicate Finder pro**.

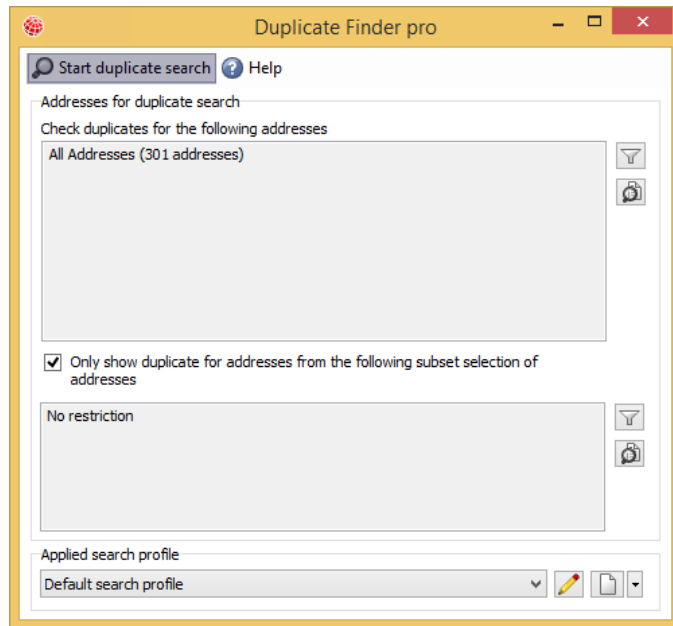
The **Duplicate Finder pro** window will open.



Using the default settings you search all addresses. Use a filter to restrict this selection, see "Configuring Duplicate Finder pro" on page 9.

### 3 Searching and merging of duplicates

Click **Start duplicate search** in order to detect duplicates.



#### How does the duplicate check work?

The duplicate search performs three search runs. The respective search runs will only be displayed if duplicates are detected.

The sequence is always the same. Starting with the company search run followed by the individual contact search run and at the end the contact person search run.

#### **Company search run**

The company search run compares companies.

Companies are compared with those individual contacts in which the **Name** field is empty and the **Company** field is filled in.

#### **Individual contacts search run**

The individual contacts search run compares companies and those individual contacts in which the **Company** and **Name** fields are filled in.

If the name of an individual contact is identical to a company name, the individual contact is displayed as a duplicate of the company. You can convert the individual contact into the contact person of the respective company via the **Company creation** shortcut menu.

**Contact person search run**

The contact person search run compares contact persons with individual contacts.

If the **Name** fields of the addresses are empty they will not be compared.

## Weighting

The matches after every search run will be displayed as percentage value. The value will be calculated by including the field weightings. The points scored will be divided by the available points and multiplied by 100. The result is accordance match in percent.

Example: In the default search profile, you can reach a maximum of 840 points. 420 points scored means a match of 50% that is calculated as follows:

$$420/840 = 0.5$$

$$0.5*100 = 50\%.$$

For each field that is matched in both addresses or that is empty, the maximum number of points is given.

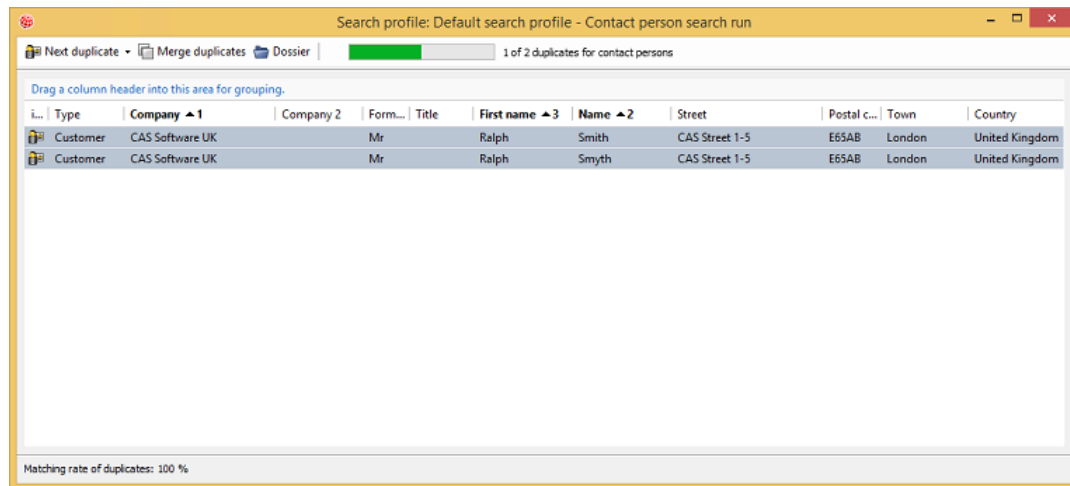
Fields that are only filled in in one address do also receive the maximum number of points if the **Identify duplicate** option in the **If empty** field is set.






Fields with different and not empty values are not given any points.

- Important information for the duplicate check Deactivated addresses will not be checked.
- The addresses belonging to the Easy Invoice and ERP connect modules will be checked, see "Configuring Duplicate Finder pro" on page 9.
- Addresses that were excluded of the duplicate search earlier will still be detected.

Usually, if you save a new address it will be checked immediately for duplicates. You can exclude the address from future duplicate searches during that check. This option only refers to the immediate search, the duplicates will still be detected via the Duplicate Finder pro.

## Toolbar functions



-  Click **Next duplicate** in the company search run to switch to the next company comparison or to the individual contact search run.
-  Click **Next duplicate** in the individual contact search run to switch to the next individual contact comparison or to the contact person search run.
-  Click **Next duplicate** in the contact person search run to switch to the next contact person comparison.
-  Click **Merge duplicates** to combine the duplicate addresses.
-  Click **Dossier** to show or hide the dossier.

### 3.1 Merging duplicates

-  Click **Merge duplicates** to merge duplicate addresses.

If there are no conflicts, the duplicates will be merged automatically.

The fields in the two addresses are then merged into one address. This merge includes any links and participants.

If conflicts occur, the **Merge duplicates** window opens. In this window you can see all conflicts.

Conflicts are, for example, different spellings of names that are detected as duplicates during a phonetic search.



Field name	1st entry	2nd entry	Result
Created on	15.04.2015 10:19:30	15.04.2015 10:14:38	
Salutation	Dear Mr Smyth [Apply value >]	Dear Mr Smith [Apply value >]	Dear Mr Smyth [Change value]
Person responsible	Ralph Smyth [Apply value >]	Ralph Smith [Apply value >]	Ralph Smyth [Change value]
Name	Smyth [Apply value >]	Smith [Apply value >]	Smyth [Change value]
	[Adopt all >>]	[Adopt all >>]	

Some fields of the found addresses are unique and will automatically be transferred to the resulting address.

- Check the field values and adopt the correct values by clicking **Apply value**.
- If you like to adopt all values of a data record click **Adopt all**.
- Click **OK**.

The result will be saved in the original data record and the duplicate will be deleted.



When you merge two addresses, you also merge their links. If the two addresses are not public, then their participants and rights are also merged.

## 4 Configuring Duplicate Finder pro

Configure the duplicate search using optional filters and the search profile in the **Duplicate Finder pro** window.

### Restricting address selection

The default settings search your total address stock. You can restrict this selection via address filters.

-  Select the **Filter** button behind the **Check duplicates for the following addresses** field to set the respective settings.
-  Click on the magnifying glass symbol to check the filtered data records.

## Working with the Easy Invoice or ERP connect modules

The following information is important for you if you deploy the Easy Invoice module or the ERP connect module.


The default settings also check addresses that belong to the Easy Invoice or the ERP connect modules. These duplicates should not be merged to avoid inconsistent data.

Exclude the addresses from Easy Invoice and ERP connect with a filter, so duplicates will not be found. Set the following filter criteria in the **Check duplicates for the following addresses** field:

- Easy Invoice: Customer ID is empty
- ERP connect: the number field for the synchronization is empty

## Restrict result list

Select the **Only show duplicate for addresses from the following subset selection of addresses** option to show only specific addresses as duplicates.

 Select the **Filter** button and select the addresses to which duplicates should be identified.

You can, for example, compare imported addresses with existing addresses without comparing all addresses in CAS genesisWorld.

 Click on the magnifying glass symbol to check the filtered data records.


## Defining search profiles

Select the **Search profile** for the duplicate search.

The default settings use the **Default search profile**. The default search profile is based on the settings for the duplicate search of the Management Console and cannot be edited in CAS genesisWorld.

The administrator can adjust the settings in the Management Console at **Miscellaneous > Addresses > Duplicate check**.

 Select **Edit** to edit a search profile.

 To create a new search profile, select the **New** button

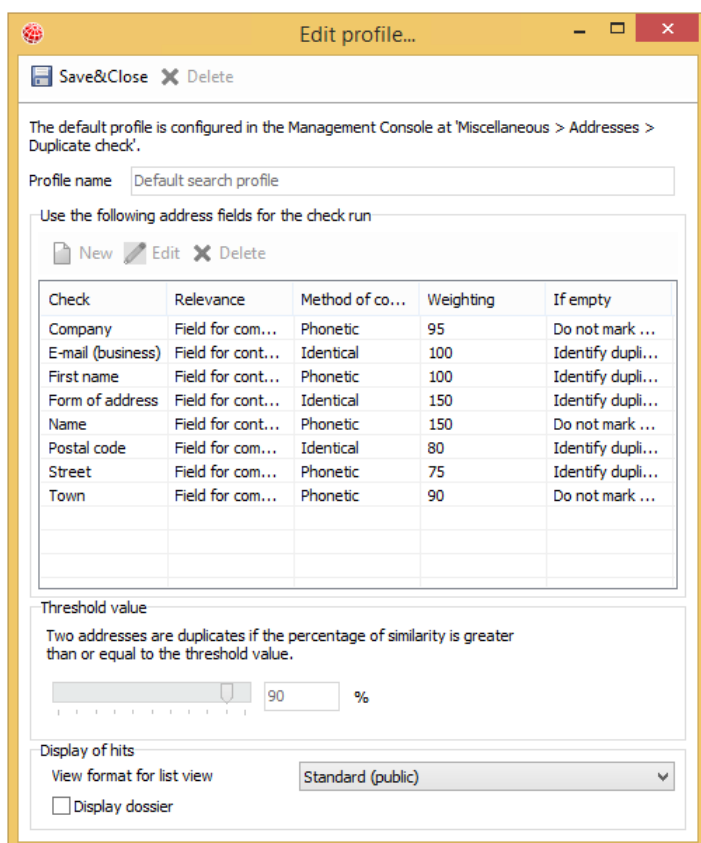
You can duplicate the selected search profile by clicking the arrow next to the **New** button and choosing **Duplicate**.

## 4.1 Editing search profiles

Search profiles are saved publicly. Every user with the **Merge duplicates** right can see, edit or delete all search profiles and create new ones. The only exception is the default search profile that cannot be edited or deleted.

You can save any number of search profiles including all your desired search criteria.

 Mark a search profile and click **Edit**.



The default profile is configured in the Management Console at 'Miscellaneous > Addresses > Duplicate check'.

Profile name:

Use the following address fields for the check run

Check	Relevance	Method of co...	Weighting	If empty
Company	Field for com...	Phonetic	95	Do not mark ...
E-mail (business)	Field for cont...	Identical	100	Identify dupli...
First name	Field for cont...	Phonetic	100	Identify dupli...
Form of address	Field for cont...	Identical	150	Identify dupli...
Name	Field for cont...	Phonetic	150	Do not mark ...
Postal code	Field for com...	Identical	80	Identify dupli...
Street	Field for com...	Phonetic	75	Identify dupli...
Town	Field for com...	Phonetic	90	Do not mark ...

Threshold value

Two addresses are duplicates if the percentage of similarity is greater than or equal to the threshold value.



%

Display of hits

View format for list view:

Display dossier

### Editing functions in the profile window

-  Click **New** in order to add a field.
- Double-click a field to edit the settings of the field.
-  Alternatively, you can select the field by clicking the **Edit** button.
- Mark a field and click **Delete** in order to remove the field.
- Two addresses are duplicates if the percentage of similarity is greater than or equal to the threshold value.

The percentage match of two compared addresses is the proportional value of the total of all field weightings that can be reached as total weighting.

The total weighting is calculated by the weighting of each individual field. Fields that are identified as duplicates are included into the total weighting with the indicated weighting. Fields with different contents are weighted with the value 0, see "Searching and merging of duplicates" on page 6.

- In the **View format for list view** field, you can define how the identified duplicates are to be displayed.
- Activate the **Display dossier** option to display the dossier directly. You can show or hide the dossier in the results window using the **Dossier** button.

## Adding or editing fields

In order to edit a field or add a new one, you can configure the following settings.

- In the **Check** field you select the field that is to be checked.
- The **Availability** field shows for which addresses the check field is available.  
You cannot edit the **Availability** field in CAS genesisWorld directly. The administration sets the settings in the Management Console under **Miscellaneous > Addresses > Field synchronization**.
- You adjust the mode in which the field content is to be compared in the **Compare method** field. You can choose between the **Identical** and the **Phonetic** compare method whereas **Identical** is the default for new fields.

In the **Identical** mode, the fields of two addresses have to be a 100% match in order to be included in the calculation with the defined weighting. The search is not case sensitive. The spelling 'Smith', for example, is identical to the spelling 'SMITH' or 'smith'.

In the **Phonetic** mode the pronunciation of two fields, established by phonetic rules, has to be a 100% match in order to be included in the calculation with the defined weighting. 'Connor' is, for example, identical to 'Konner'.

- In the **Weighting** field you define how a field should be weighted.

Each field is weighted with the indicated value. This value is included in the total weighting if the fields of two addresses are matching.

If field values of one field of the addresses differ from each other, the value 0 counts for the total weighting for this field.

Only if the total of the field weighting exceeds the given threshold value for duplication, the two addresses are detected as duplicates.

- In the **If empty** field, you define how an empty field should be rated in comparison to a filled-in field.

If you select **Identify duplicate**, an empty field will be treated and weighted like an identical field during a comparison.

If you select **Do not mark as duplicate**, the empty field will be treated as being different from the field containing a value.

Example: Select the **Identify duplicate** option for the **E-mail (business)** field. The contact persons are still identified as duplicates if all fields match but the **E-Mail (business)** field is empty.