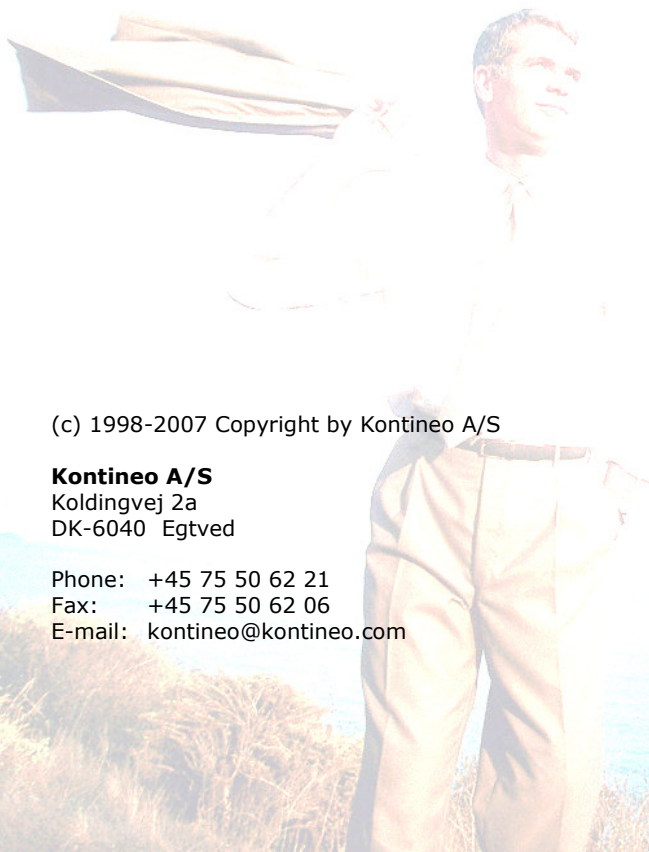


## **Duplicate Analyzer 2007**

*for SuperOffice CRM®*



(c) 1998-2007 Copyright by Kontineo A/S

**Kontineo A/S**  
Koldingvej 2a  
DK-6040 Egtved

Phone: +45 75 50 62 21  
Fax: +45 75 50 62 06  
E-mail: [kontineo@kontineo.com](mailto:kontineo@kontineo.com)



# List of contents

|   |           |
|---|-----------|
| <b>LIST OF CONTENTS .....</b>                         | <b>2</b>  |
| <b>INSTALLATION AND CONFIGURATION.....</b>            | <b>3</b>  |
| <i>System requirements .....</i>                      | <i>3</i>  |
| <i>Backup/Upgrade Recommendation.....</i>             | <i>3</i>  |
| <i>Installation.....</i>                              | <i>3</i>  |
| <i>Technical Configuration .....</i>                  | <i>3</i>  |
| <i>License Activation .....</i>                       | <i>4</i>  |
| <i>License Management and User Authorization.....</i> | <i>5</i>  |
| <i>Analysis Data Limitation .....</i>                 | <i>6</i>  |
| <i>Master Definition .....</i>                        | <i>7</i>  |
| <i>Reject Merge Configuration .....</i>               | <i>8</i>  |
| <i>Script Configuration .....</i>                     | <i>9</i>  |
| <i>Analysis Profile Configuration .....</i>           | <i>10</i> |
| <i>    Profile Line Configuration.....</i>            | <i>11</i> |
| <b>ANALYSIS .....</b>                                 | <b>13</b> |
| <i>Running Analysis .....</i>                         | <i>13</i> |
| <i>The Merge Company Screen.....</i>                  | <i>14</i> |
| <i>    The Filter Section .....</i>                   | <i>14</i> |
| <i>    The Master and Candidate Sections .....</i>    | <i>15</i> |
| <i>    The Functions Section .....</i>                | <i>16</i> |
| <b>PROCESSING ANALYSIS RESULTS.....</b>               | <b>17</b> |
| <i>Execute Bulk Merge.....</i>                        | <i>18</i> |
| <i>Person Analysis and Merge.....</i>                 | <i>18</i> |
| <b>VIEW MERGED LIST – AND UNDO MERGE.....</b>         | <b>19</b> |
| <b>ADVANCED FUNCTIONS .....</b>                       | <b>20</b> |
| <i>Compact Local Database.....</i>                    | <i>20</i> |
| <i>Remove Zipcode Prefix .....</i>                    | <i>20</i> |
| <i>Clear all Undecided .....</i>                      | <i>20</i> |
| <i>Set all Undecided.....</i>                         | <i>20</i> |
| <i>Clear Local and reload Central.....</i>            | <i>20</i> |

## Installation and configuration

### System requirements

Duplicate Analyzer 2007 requires:

- SuperOffice CRM version 6.0 up to 6.1.
- Microsoft SQL Server 2000 or Microsoft SQL server 2005, Oracle 9.
- PC requirements: Equivalent to SuperOffice 6.1

### Backup/Upgrade Recommendation

Duplicate Analyzer 2007 creates so called 'detached' tables in the SuperOffice database. Using SuperOffice tools, such as dbsetup.exe, will NOT include these tables. Therefore, to maintain the Duplicate Analyzer Log you should make sure these tables and the data therein are maintained.

### Installation

Duplicate Analyzer does not require a Server installation, but should be installed on one or more workstations.

- Run the Installation script available from [www.kontineo.com/download/DuplicateAnalyzer2007install.exe](http://www.kontineo.com/download/DuplicateAnalyzer2007install.exe)
- Follow the on-screen instructions.

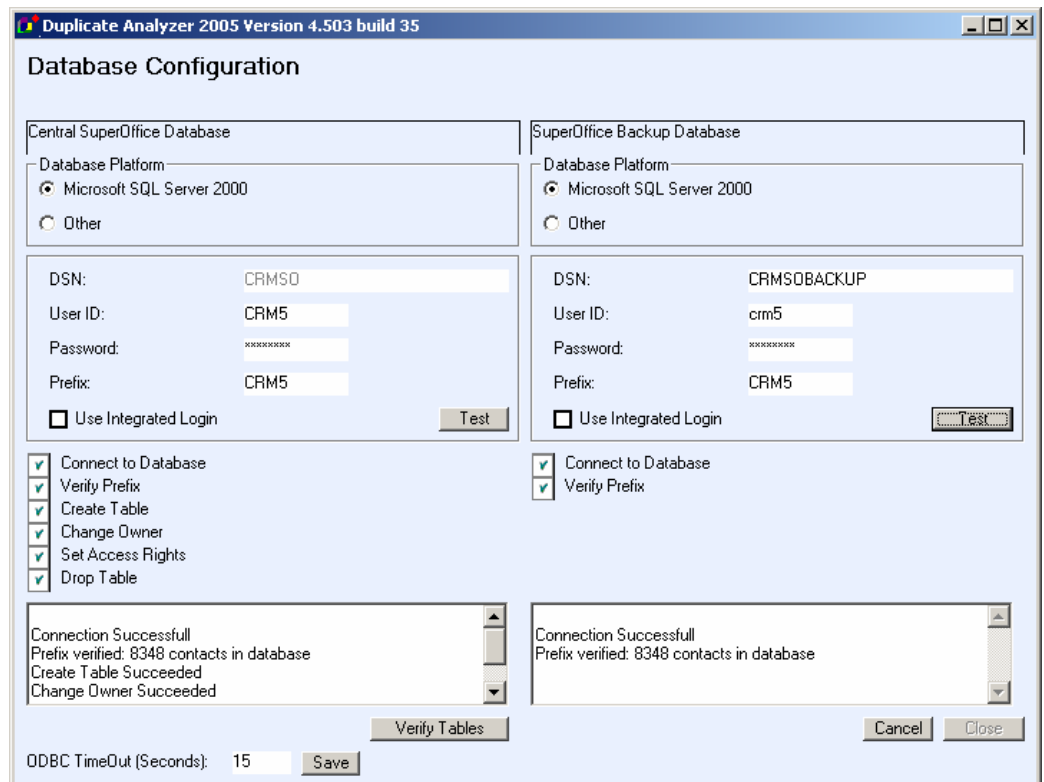
### Technical Configuration

At first run, Duplicate Analyzer 2007 will verify if the tables required to run Duplicate Analyzer 2007 exist. If not – or if it fails to connect to the SuperOffice SQL database, it will launch the Database Configuration screen.

In the left side you must provide an SQL Server login with system administrator rights on the SuperOffice SQL server database.

Usually the CRM5 user has these rights. Click the 'Test' button to test that the user has sufficient rights. When successful, click the 'Verify Tables' button. When table verification succeeds, the 'Close' button is enabled.

*Tech Note: Duplicate Analyzer require 5 tables to be created in the SuperOffice SQL Database (Please remark, this tables are so called 'detached tables' which are unknown to SuperOffice and therefore the do not in any way inflict on SuperOffice Replication. The tables will not be included if you use the SuperOffice dbsetup.exe application to rebuild/backup a database – but the tables will be included in an SQL Backup of the SuperOffice SQL Server database).*



In the right pane, you can optionally configure Duplicate Analyzer with a SuperOffice Backup Database. This is not required to run Duplicate Analyzer, but it is required to enable the 'Undo Merge' functionality in Duplicate Analyzer.

ODBC Timeout is as default set to 15 seconds. Large databases may require a longer timeout. If you get timeout errors in Duplicate Analyzer, you should increase this value.

### License Activation

At first start-up (after Database Configuration), the License Activation screen will be displayed.

You can choose:

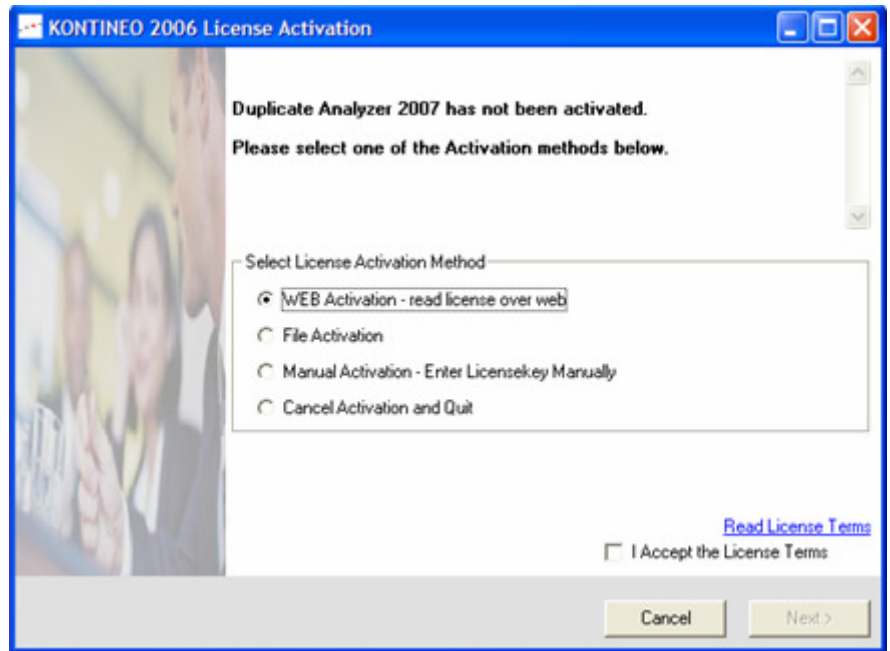
#### WEB Activation

If a License has been issued by Kontineo, the License will be loaded from Kontineo WEB Server.

If no license has been issued, a Demo License with restricted functionality will be generated and read.

WEB activation may not be possible, due to security restrictions. In that case you will get an error. You should contact Kontineo for a license file.

Select this option, set a tickmark in *Accept the License Terms* and click *Next*



#### File Activation

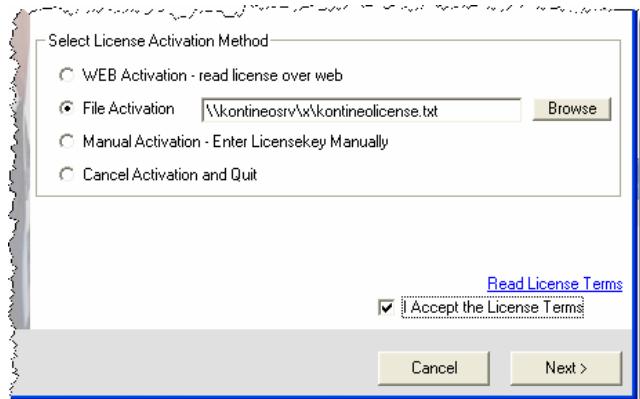
If you have had a License file sent from Kontineo, you can load the License from this file.

Select this option, set a tickmark in *Accept the License Terms* and click *Next*

#### Manual Activation

You can manually type in the License Key.

Select this option, set a tickmark in *Accept the License Terms* and click *Next*



### **License Management and User Authorization**

After Activation, the License Management screen will be displayed.

Administrator users can at any time access this screen from the Help Menu.

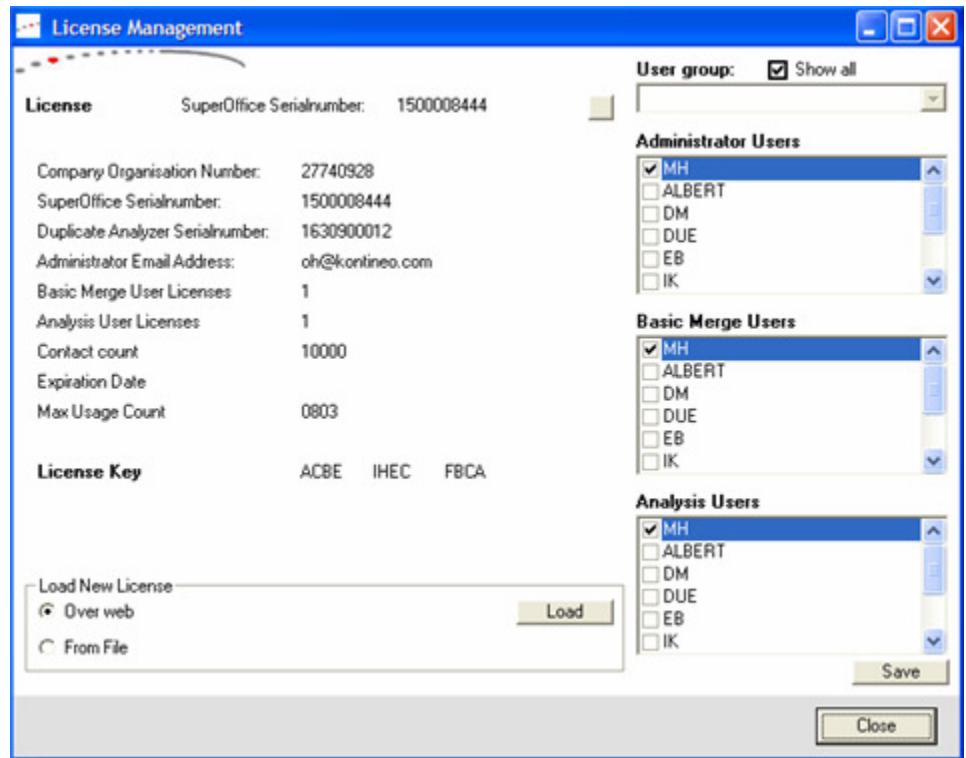
If a new License has been issued, you can load it over WEB or from file.

In the right side, you can administer user Authorization:

Administrator Users can manage Licenses.

Basic Merge Users can perform Merge of companies and persons, but cannot run Analysis. They can process analysis results though.

Analysis users can manage and run duplicate identification Analysis.



**Analysis Data Limitation**

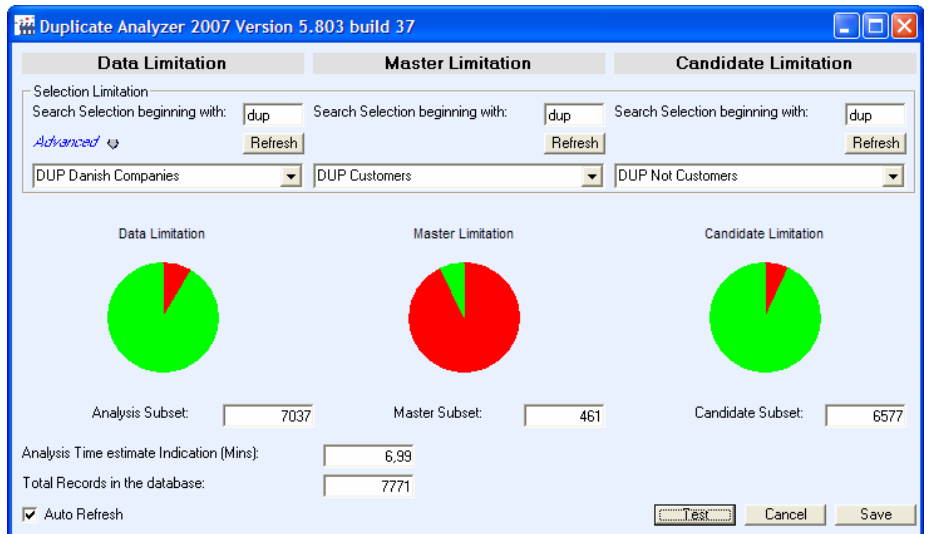
In the Navigator Bar under Analysis, you can define advanced Analysis Data Limitations. Click the 'Change...' button.

You can use Selections to define data limitations.

In the configuration picture illustrated to the right, we have made a static selection in SuperOffice of all Danish Companies which is defined as the analysis subset.

Another SuperOffice selection contains all existing customers. That can be both Danish and foreign, but when the Master Limitation is Danish, the Master Subset will only be those that are in both selections, i.e. Existing Danish Customers.

The third selection contains all contacts in SuperOffice, that are not customers.



The result is, that for 461 Danish Companies, we will search for duplicate candidates amongst 6577 danish companies that are not customers.

Data Limitation lets you handle the issues of defining masters and candidates, but it is also particularly relevant for large databases where the Master Limitation lets you run the analysis in 'bites', and you can let multiple users work on different analysis subsets.

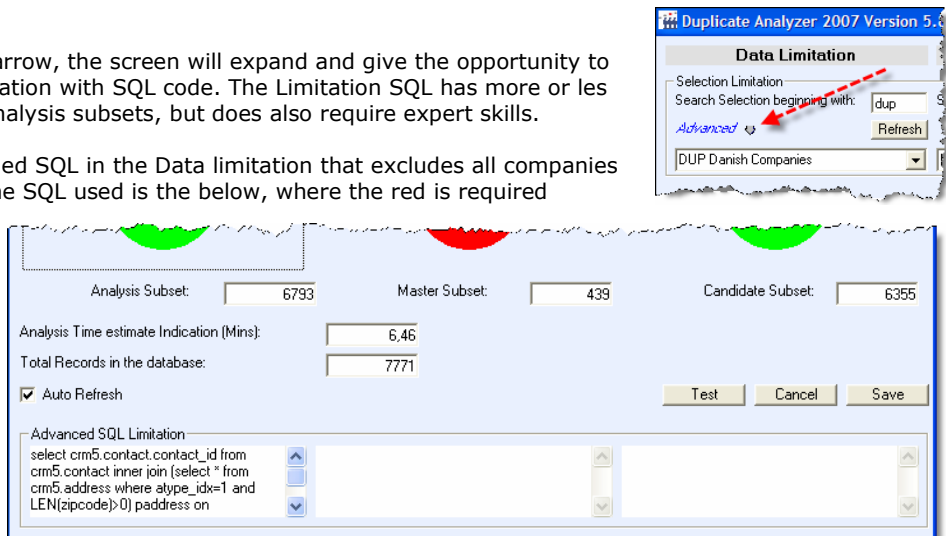
**Advanced Limitation**

If you click the blue "Advanced" down arrow, the screen will expand and give the opportunity to define data, master and candidate limitation with SQL code. The Limitation SQL has more or less unlimited capabilities of limiting your analysis subsets, but does also require expert skills.

In the picture to the right, we have added SQL in the Data limitation that excludes all companies not having a postal address zipcode. The SQL used is the below, where the red is required syntax:

```
select crm5.contact.contact_id from
crm5.contact inner join (select * from
crm5.address where atype_idx=1 and
LEN(zipcode)>0) paddress on
crm5.contact.contact_id=paddress.owner_id
```

This is still combined with the selections, where we can see, that the analysis subset has decreased, and so has both the master and candidate subsets.



For company data limitation, Advanced SQL Limitation must always start with the below:

```
select crm5.contact.contact_id from crm5.contact
```

*Note: Limitation specification from Duplicate Analyzer versions prior to SR2 will not work.*

## Master Definition

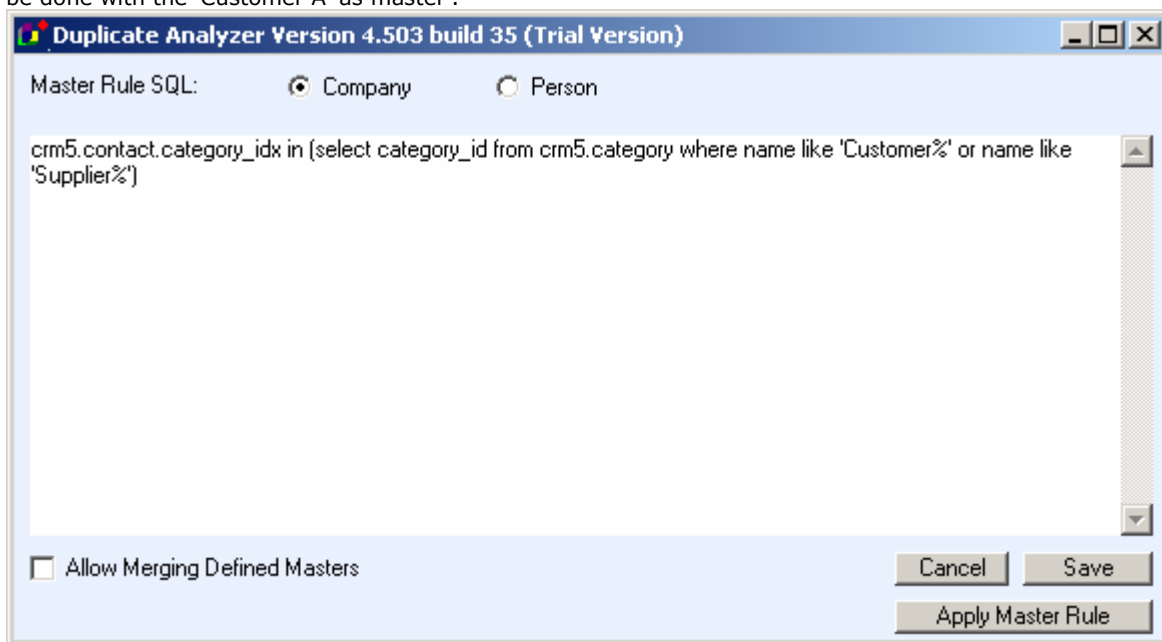
When analysing and merging companies or persons, you may want to set up rules for which ones to remain as masters, and which ones should be merged to the master and deleted.

When merging companies, all data that is of a 'one to many' type (i.e. activities, interests, etc.) are moved from the candidate to the master – the candidate is then deleted. Such 'one to many' information will therefore remain in the database.

*Tech Note: Duplicate Analyzer 2007 does not merge data in custom tables that are not part of the standard SuperOffice CRM 5 database structure. See Script Configuration for merging data in custom tables.*

'One to one' type information, such as the category, the Our Contact, data on the 'More' page and various other information is more problematic – since one company can only have one category for instance. When merging two companies with different categories, the category of the Master will remain.

Therefore, merging e.g. a company with category 'Customer A' and a company with 'Imported – category undecided' should be done with the 'Customer A' as master'.



The Master rule SQL is defined similar to data limitation. Above as set, that a company with 'Customer' or 'Supplier' in the Category, should always be masters.

If you try to merge a 'Defined Master' as a candidate, you will be told that it is defined master. If the 'Allow Merging Defined Masters' tick box is ticked, you will be allowed to manually merge when you confirm.

Automatic Merge will never merge 'Defined Masters' as candidates.

The 'Apply Master Rule' function button applies the rules to analysis result sets. Results are not applied during analysis, but should be applied after. Applying a master rule will compare identified analysis result sets. If it find one with a candidate being a defined master, it will 'toggle' (i.e. set the candidate to master and the master to candidate), unless the master is also a defined master.

Using Bulk Merge will skip merge of defined masters, possibly leaving a set of duplicates where both are defined masters. These should be checked manually, or another master rule should be applied.

*Tech Note: The base SQL for Master RULE SQL is: select contact\_id from contact where contact\_id>0*

**Reject Merge Configuration**

The typical usage of Reject Merge is to import account numbers from a Finance System or other system integrated with SuperOffice.

Thereby, you ensure, that companies that exist in the account system are not merged.

Any identified duplicates in the Finance System will in this manner be identified – and you will be able to disable these accounts in the Finance System before merging the contact in SuperOffice.

Under 'Configuration', select 'Reject Merge Configuration...'

Tick 'Enable Reject Merge'

Select the field in SuperOffice the values you will import from another system relate to. (Link Field)

Click the 'Build..' button to define how to connect to your finance system (or other system).

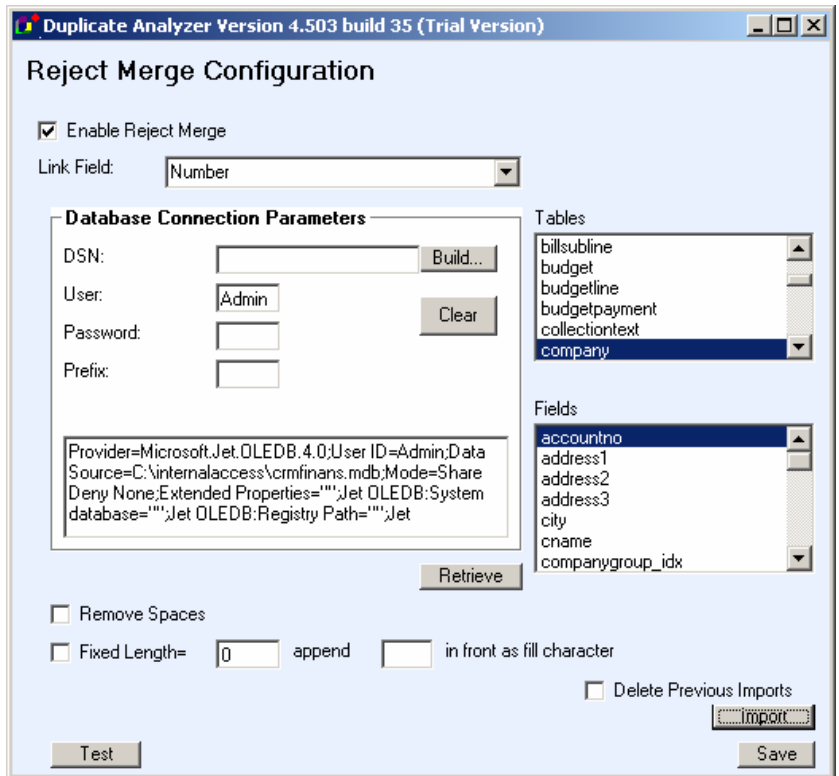
Click 'Retrieve' will list all tables in the database you have connected to.

Clicking a table will list all the fields in that table.

Click the field that hold the account number or other number linked to the selected SuperOffice Field.

Click 'Import' will import all values in the selected field (As string Values)

Click the 'Test' button will verify that your imported values correctly links to contacts in SuperOffice:



If none of the imported values has a related record in SuperOffice, some manipulation may be required.

Tick 'Remove Spaces' will, during import remove spaces in the imported values.

Tick 'Fixed length' = '10' append 'x' will, during the import, append x's in front of values that have a length less than 10, e.g. 12345678 becomes XX12345678.

### Script Configuration

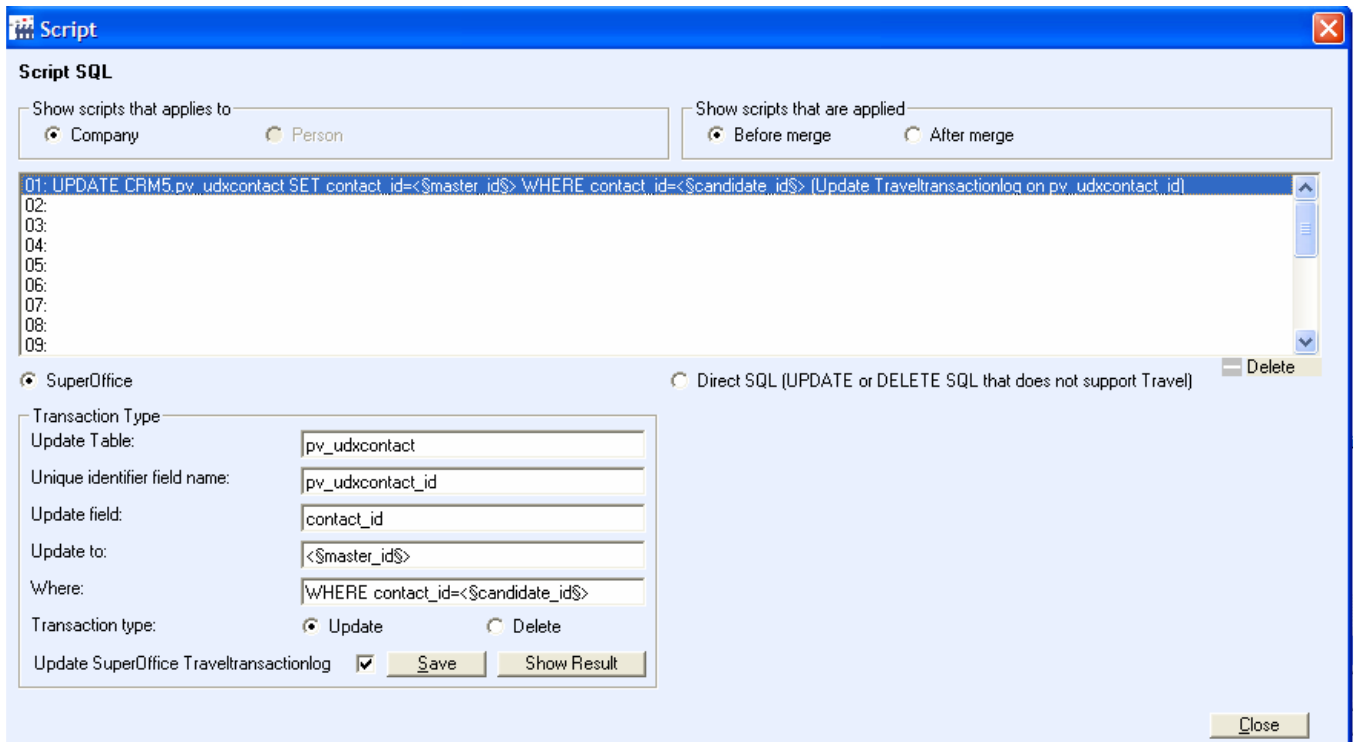
Scripts allow you to also Merge data in custom tables, in tables that are not part of SuperOffice, or even in tables in other systems.

**WARNING: Incorrect Scripting can have fatal results.**

Scripting requires expert SQL skills.

You can define scripts to run:

- **before merging a company**  
If you include data in your script that is deleted or updated during the merge, you must use the Before Merge type.
- **after a company has been merged.**  
If your script depend on data that must have been updated during the merge, you must use the After merge type.



You can apply two types of Scripts:

- **SuperOffice**  
If you use Remote Travel or Satellites, your Scripts must support update of the SuperOffice Travel Transactionlog. You should apply only SuperOffice dictionary tables scripts.
- **Direct SQL**  
You can apply direct SQL scripting. Direct SQL scripting does not support SuperOffice replication. The most common use is to fire a script after a merge, that updates data in another system. You may also want to use Direct SQL scripting to update data in SuperOffice, add data to a datawarehouse and many other purposes.

The above pictured example moves records in a custom SuperOffice table from the Candidate to the Master.

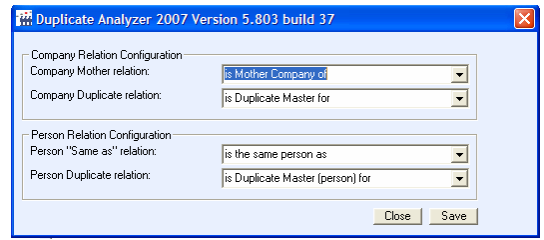
#### Direct SQL scripting Example (SQL Server)

This example sets a merged candidate with the same customer number to inactive in a Finance System, if one is found:  
UPDATE XALSRV.XAL.debtors set active=0 where customerno=(select number2 from CRM5.contact where contact\_id=<\$candidate\_id\$>)

### Relations Configuration

Relations Configuration allows you to add relations between companies or persons identified as potential duplicates.

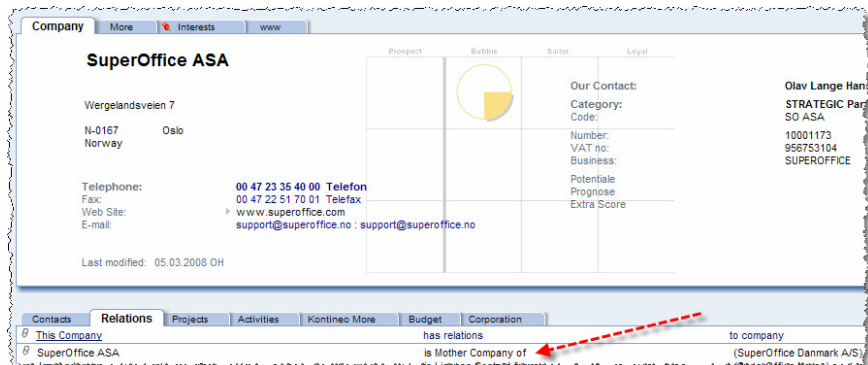
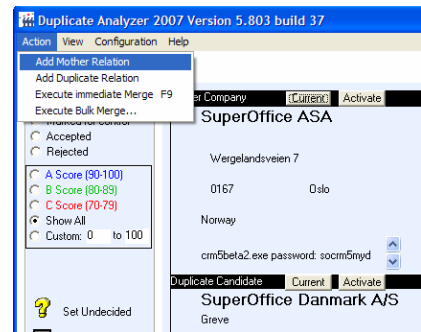
Configured Company relations should be defined as Company to Company Relations in SuperOffice Relations configuration, and like wise Person to Person for configured Person Relations.



In the User Interface in the Actions Menu, you have:

- Add Mother Relation
- Add Duplicate Relation.

In the illustrated example a Mother Relation is added between the suspected master and candidate, which then appears in SuperOffice.



### Userdefined Fields Configuration

In Duplicate Analyzer 2007 SR2, you can configure up to seven User defined fields on company level. (Fields from the Company More Page).

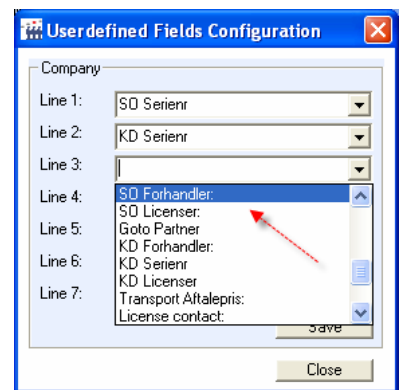
These fields will:

- Be visible in the Master/Candidate window
- You can edit data in the userdefined fields directly in Duplicate Analyzer by double clicking
- You can include the selected userdefined fields in Analysis profile comparison

To activate the required userdefined fields, you have to go into Analysis and go to:

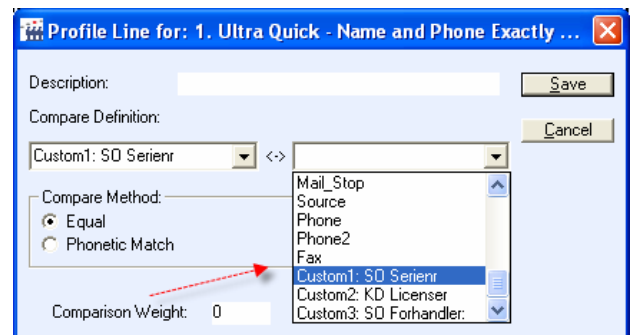
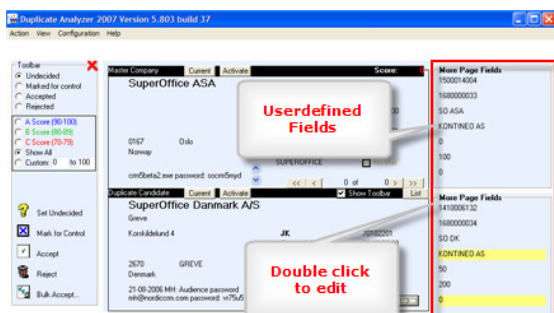
Configuration → Userdefined Fields Configuration

Here you can select the values that your analysis requires.

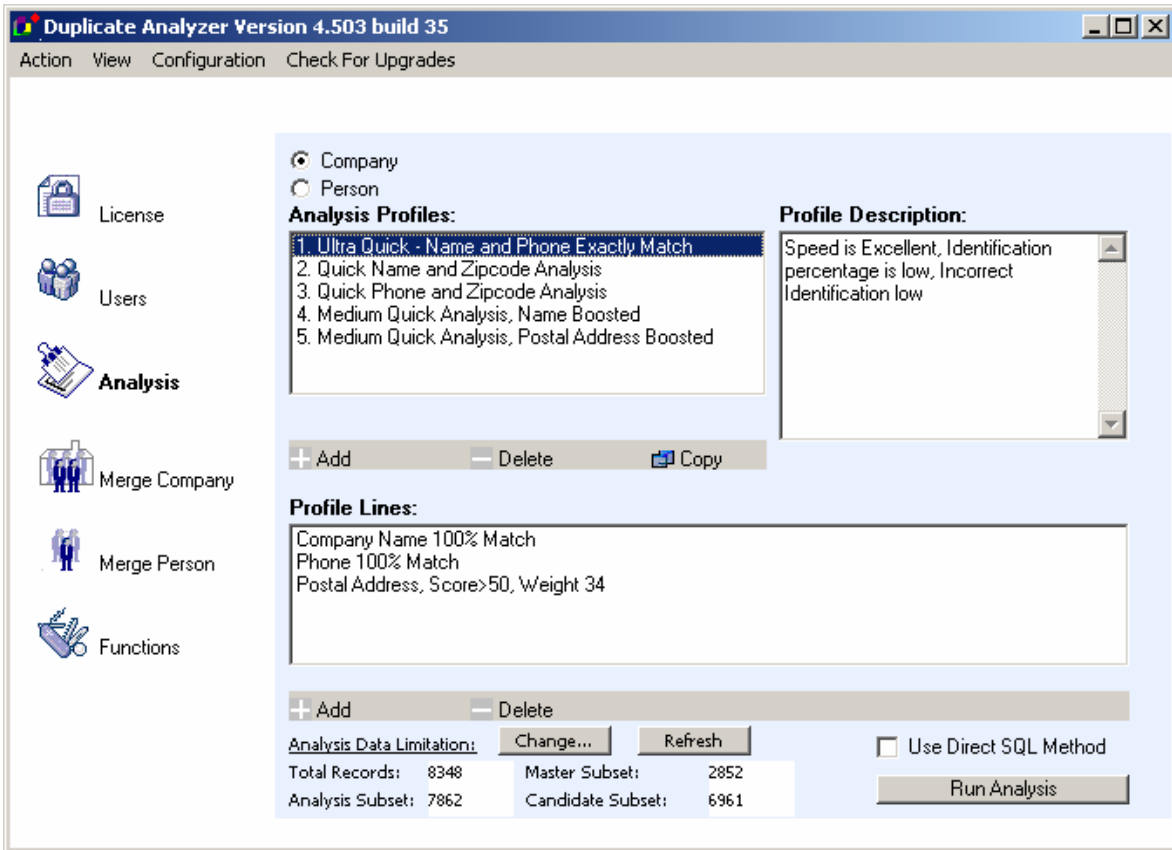


To add the Userdefined fields into the analysis, you have to add a Profile Line into the Analysis Profiles.

To see the result, you have to go select 'Merge Company' where it will be listed.



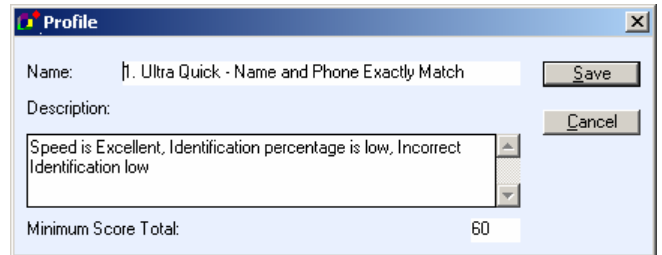
## Analysis Profile Configuration



Duplicate Analyzer comes with 5 preconfigured Company analysis profiles, and one Person Analysis Profile.

- A Profile has only three properties:
- Name: The line displayed in the list
  - Description
  - Minimum Score Total

The minimum Score total is a value (max 99 – Duplicate Analyzer never gives score 100) determining at what level Duplicate Analyzer should include an analyzed Master/Candidate Set in the result set. A minimum score total less than 60 is not recommended. Take a database with 10.000 contacts. If you set a minimum total score to 0, Duplicate would analyse the first contact to 9.999 others – and return them all as possible candidates. So basically, it would return something like 100 million possible duplicate sets! And if you have 100.000 contacts, hmmm... – 10 billion...!!! – So – don't set the minimum total score too low.



### Profile Line Configuration

Each Profile can have up to 9 profile lines. The properties of a Profile line are:

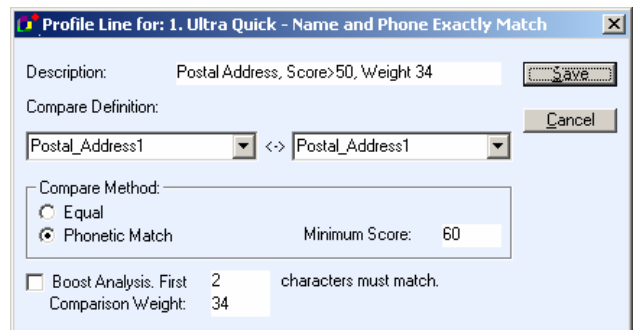
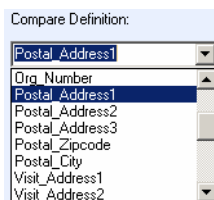
**Description:** The line displayed in the list

**Compare Definition Fields:** This profile line is set up to compare Postal address 1 to Postal address 1 during analysis, but you can also configure 'cross field' analysis, e.g. visit address to postal address, etc.

**Compare Method:** Compare Method is for improving analysis speed.

Setting a profile line to **equal** will speed the analysis process significantly up, as it will only compare other companies in the database having a 100% match in the field. This will, however, also mean that no candidates will be found where the field differs just a little bit. A good field to set to Equal is zipcode.

**Phonetic Match** will compare all and be slow – but this combined with 'Boost' on 2 characters will compare e.g. all addresses starting with the same two characters as the address being looked at.



**Comparison Weight:** Each Profile line has a Comparison Weight. When giving a Master/Candidate set a score, Duplicate Analyzer takes the score of the individual profile line and multiply it with the weight of the line and divides it by the sum of the comparison weight of all profile lines. Below are two different configurations, and an example of a comparison score:

| Profile Line        | Line Score  | Line Weight | Line Weighted Score    |           |
|---------------------|-------------|-------------|------------------------|-----------|
| Name <-> Name       | 70          | 40          | $70 \times 40 / 100 =$ | 28        |
| Address <-> Address | 60          | 40          | $60 \times 40 / 100 =$ | 24        |
| Phone <-> Phone     | 99          | 20          | $99 \times 20 / 100 =$ | 20        |
|                     | Sum Weight: | 100         | <b>Total Score</b>     | <b>72</b> |

| Profile Line        | Line Score  | Line Weight | Line Weighted Score    |           |
|---------------------|-------------|-------------|------------------------|-----------|
| Name <-> Name       | 70          | 30          | $70 \times 30 / 100 =$ | 21        |
| Address <-> Address | 60          | 30          | $60 \times 30 / 100 =$ | 18        |
| Phone <-> Phone     | 99          | 40          | $99 \times 40 / 100 =$ | 40        |
|                     | Sum Weight: | 100         | <b>Total Score</b>     | <b>79</b> |

We are looking at two companies, where the name and address are looking similar, but with a relatively low score.

In the top profile, the weight is low on Phone (20) where in the bottom more weight is put on Phone (40). With the two different configurations only on Comparison Weight, we get a total score difference from 72 to 79.

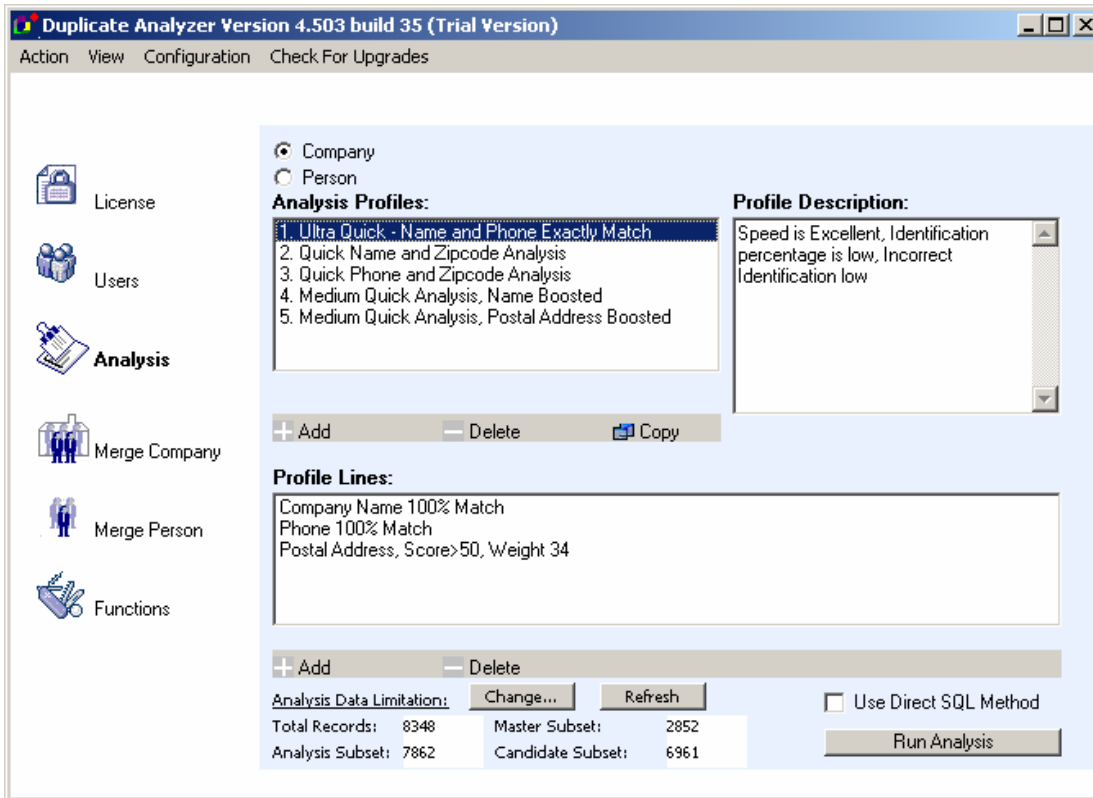
This sample is included to clarify the meaning of the Comparison Weight – and how running the same analysis configurations, only with changed Comparison Weights can be a good method to improve duplicate identification.

**Minimum Score:** Is for the line score. If the score is smaller than a minimum score of e.g. 60 when comparing e.g. address, then it will not count in the total score (e.g. above, had the Line score for address been 59, the Line Weighted score for address would have been 0. A score of less than 60 is usually not relevant, but it does not have the any 'fatal impact' as setting the Profile Minimum Score to a lower value. Both 50 and 40 can help 'catch the last duplicates'. For Phone a 'not near' match is not very relevant, so increasing the Minimum Score to 85 can be recommended.

## Analysis

### Running Analysis

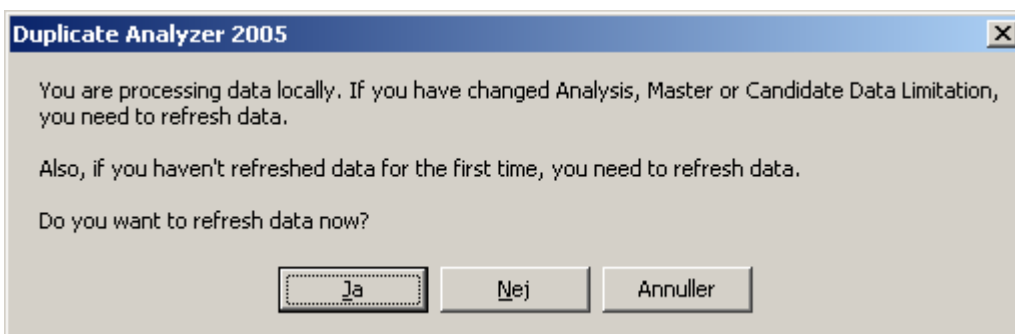
Go to 'Analysis' in the Navigator Bar



Click the first Analysis Profile and Click 'Run Analysis'.

*Tech Note: 'Use Direct SQL Method' will process data directly on the SQL Server. This is slower, and lays a hard restraint on the SQL Server Performance – slowing down all other user's tasks. It is generally not recommended, but can be useful for small test runs.*

You will be prompted to Refresh Data. Click Yes to refresh if necessary (or if you fell uncertain – it will take a bit longer, depending on you database size)

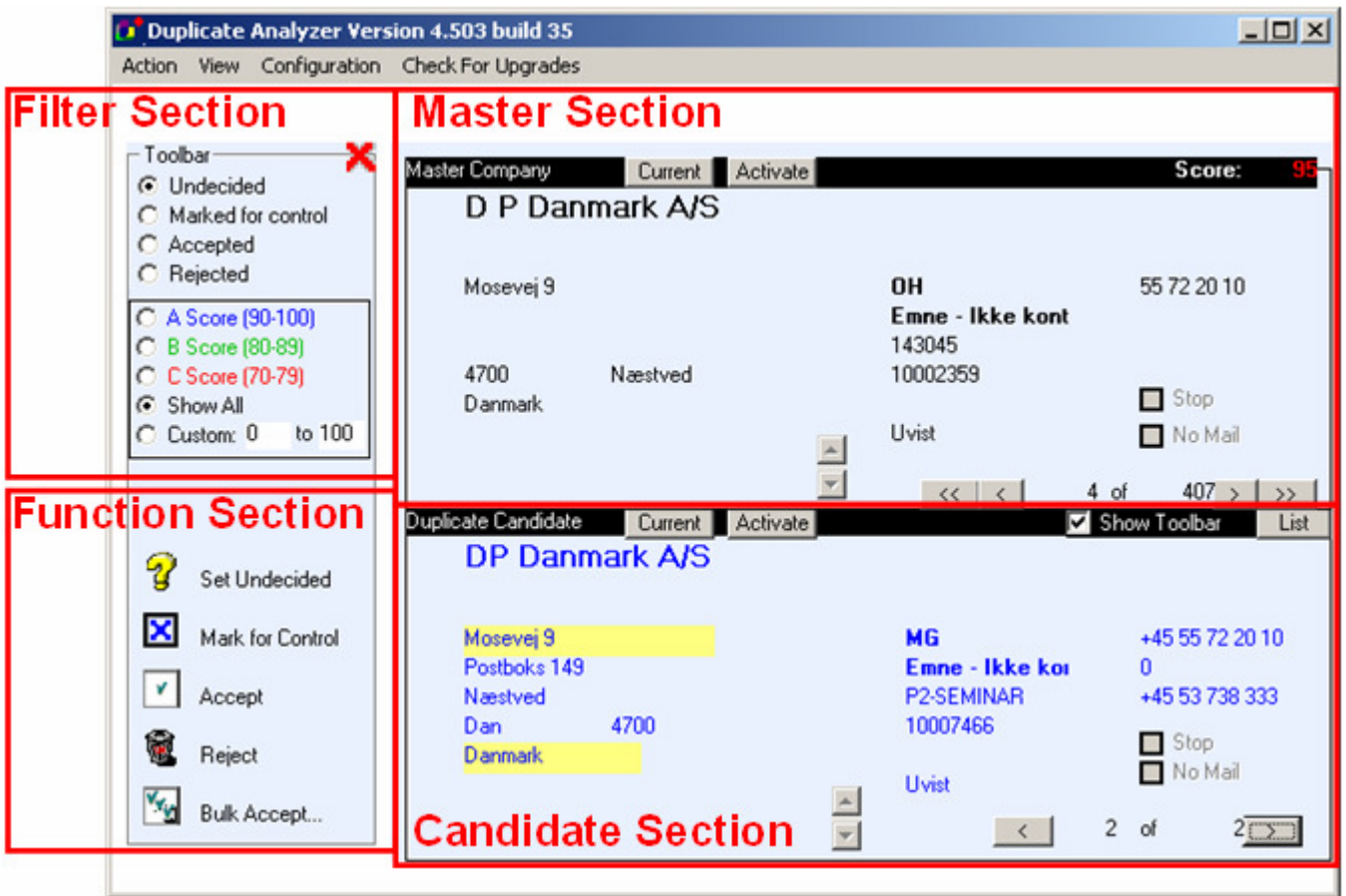


Click Yes to refresh data

The analysis will start – when completed, it will jump to the 'Merge Company' Screen' See Processing Analysis Results.

NB: You should always apply Master Rules after having run an analysis – unless you have unmerged accepted results where you have manually toggled companies (or persons).

**The Merge Company Screen**

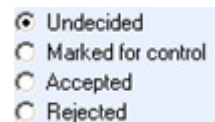


The 'Merge Company' screen has four sections:

**The Filter Section**

A 'Master/Candidate set' can have 4 statuses:

- Undecided (Has been identified automatically – it has not been processed yet)
- Marked for Control (The user was not sure if a Master/Candidate set should be accepted or rejected)
- Accepted (The Master/Candidate set has been accepted to be merged – (Has not been merged yet)
- Rejected (The Master/Candidate set has been rejected – it will not be merged (and will not be listed again))



A Master/Candidate Set is given a score, describing how well they match (upper right corner in red)

*Tech Note: The score is given according to the Analysis definition. Running an analysis with high weight on Name may give another score than an analysis with high weight on Address. Running multiple analysis will change the score of the same set, if higher when a second analysis finds the same Master/Candidate set, but is the score of the second analysis lower than that of the first, then the higher score will remain. Running multiple analyses before processing results may incur many incorrect Master/Candidate sets with high score – so we recommend to process data after each analysis.*



- A: High Score – Very likely duplicates (Can often be bulk accepted)
- B: Relatively High Score (Can often be bulk accepted – but should be checked)
- C: Medium Score, likely to be duplicates, but manually processing usually required
- Lower than C: Not likely to be duplicates, but some may be.



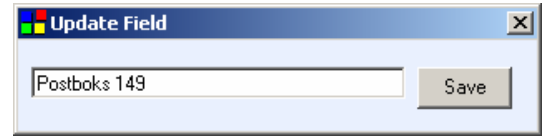
Usually you will be looking at undecided – and go down in score during processing.

You may also turn it around, running a bulk accept of all, and afterwards look at accepted with a low score, working from low score up to a higher score, where you find a satisfactory likeliness of all to be correctly identified duplicates.

**The Master and Candidate Sections**

The master section shows the data of the suggested master in a Master/Candidate set.

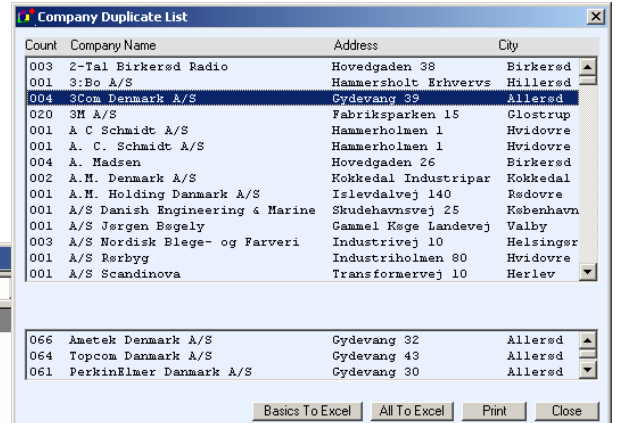
In the master section you can double-click a data field to edit data. It will default take the value from the same field of the Candidate, but you can manually change it. If you e.g. want the 'Postboks 149' added to the address of the master you double click the address field in the master section, and click save.



In the Right bottom part of the master section, you can browse the Master/Candidate Sets. It displays here that we are looking at Master no 4 of 407 (Undecided – Score is at 'Show All')



Just below is the 'List' Button. The list button will display a list of all the 407 Masters (Undecided – Score is at 'Show All') – and clicking a master in the top part of the screen will display all the identified Candidates in the bottom list.



One Click on the List button will call that Master/Candidate set to the Main Screen (and close the list).

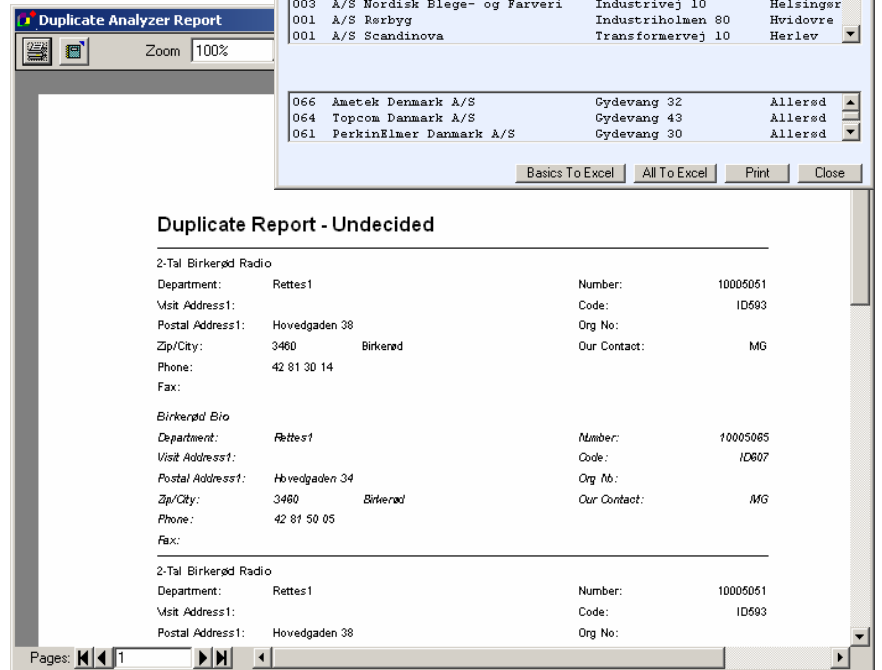
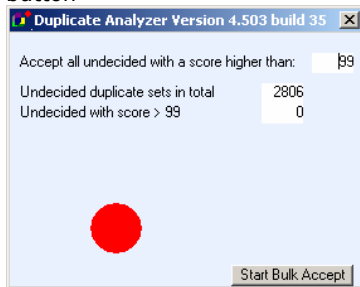
From the List Screen – you can also print a list of all the Master/Candidate Sets.

In the right bottom part of the Candidate Section there is also Browse Buttons.



The numbers shows, that the master displayed in the Master Section has two possible candidates – we are looking at the second.

As to the Analysis results, it returned 407 possible masters. However, each master can have multiple candidates; in this case there are actually as many as 2806 possible Master/Candidate sets. (Displayed when clicking the 'Bulk Accept button'



The candidate section use Colour Codes to make it easier to manually process data.

The Text Colour relates to the Score. Blue=A, Green=B, Red=C, Black=Lower Score. The Yellow Markings shows that these fields are 'near exact matched' to the same fields of the master, i.e. the more yellow, the more fields are exactly the same.

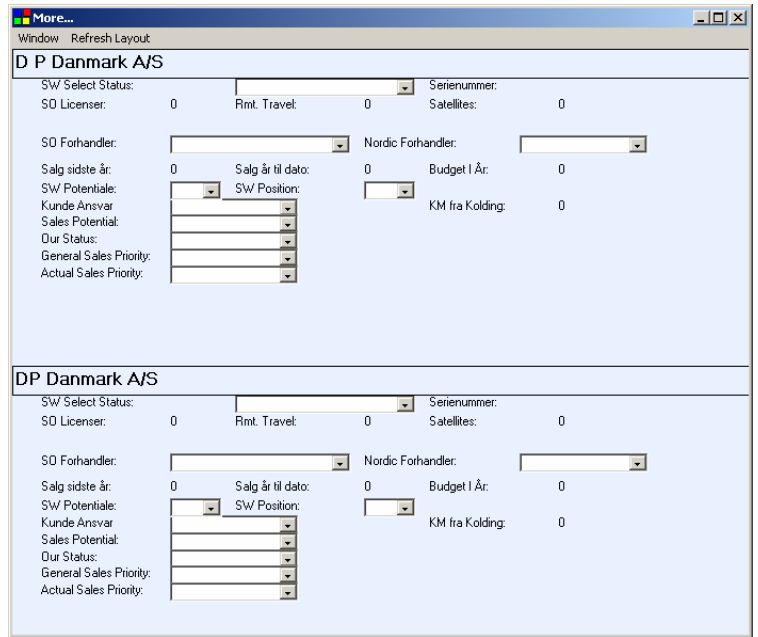
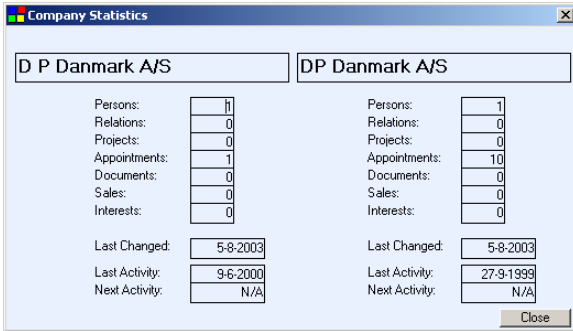


The 'Activate' buttons, which are in both the Master and Candidate sections, activates the Master or Candidate in SuperOffice.

The 'Current' buttons replace the master or the candidate with the current contact in SuperOffice (This is mainly to merge two companies ad hoc when found in SuperOffice).

Click CTRL + M to view More Pages of the Master/Candidate Set (or use 'More' in the 'View' menu).

Click CTRL + S to view statistics about the Master/Candidate set. (or use 'Statistics' in the 'View' menu).



**The Functions Section**

The four buttons encapsulated in red are the buttons used when processing analysis results manually. Depending on the Status filter, one would not be used, e.g. if you are processing Undecided, the Set Undecided is irrelevant since the displayed Master/Candidate sets are Undecided – but if you are looking at Accepted results, you may set some back to undecided.

When the filter is at Undecided, you make a decision to each of the Master/Candidate Sets. Clicking 'Mark for Control', 'Accept' or 'Reject' will change status of the Master/Candidate set, and the next Master/Candidate Set automatically displays.

See further explanation under 'The filter section' above.

The Toggle button is available only when the Status filter is at 'Accepted' (There is a long complicated explanation for this). It will move the Candidate to be master and vice versa.

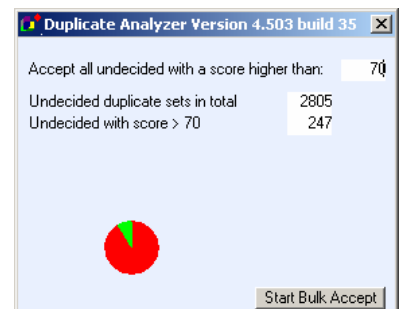


The Bulk Accept button brings up the Bulk Accept Screen:

In this screen you can automatically accept all Master/Candidate sets with a score higher than a value you set. To the right we have set it to Accept all Master/Candidate sets with a Score higher than 70.

The chart and figures shows, that 247 of 2805 Master/Candidate sets will be accepted.

You may use a method where you Bulk Accept many – then sets the Status filter to Accepted, and manually goes through accepted and rejects from there, taking scores from down and up. This method is opposed to viewing Status Undecided and manually accepting, taking scores from top and down.



Start Bulk Accept sets all to Accepted.

*Tech Note: When a Master/Candidate set is accepted, possible candidates to the accepted candidate will become candidates of the Master of the accepted candidate. This is quite complicated, and may result in some rather surprising results, e.g. a company which was given a high score to a candidate on address, may end up as a candidate with a completely different address, if it's original master was given a high score on name, and so forth...). This issue is also the reason for the 'Toggle' button only to be available when viewing Accepted results. The issue will not be dealt with further in the manual.*

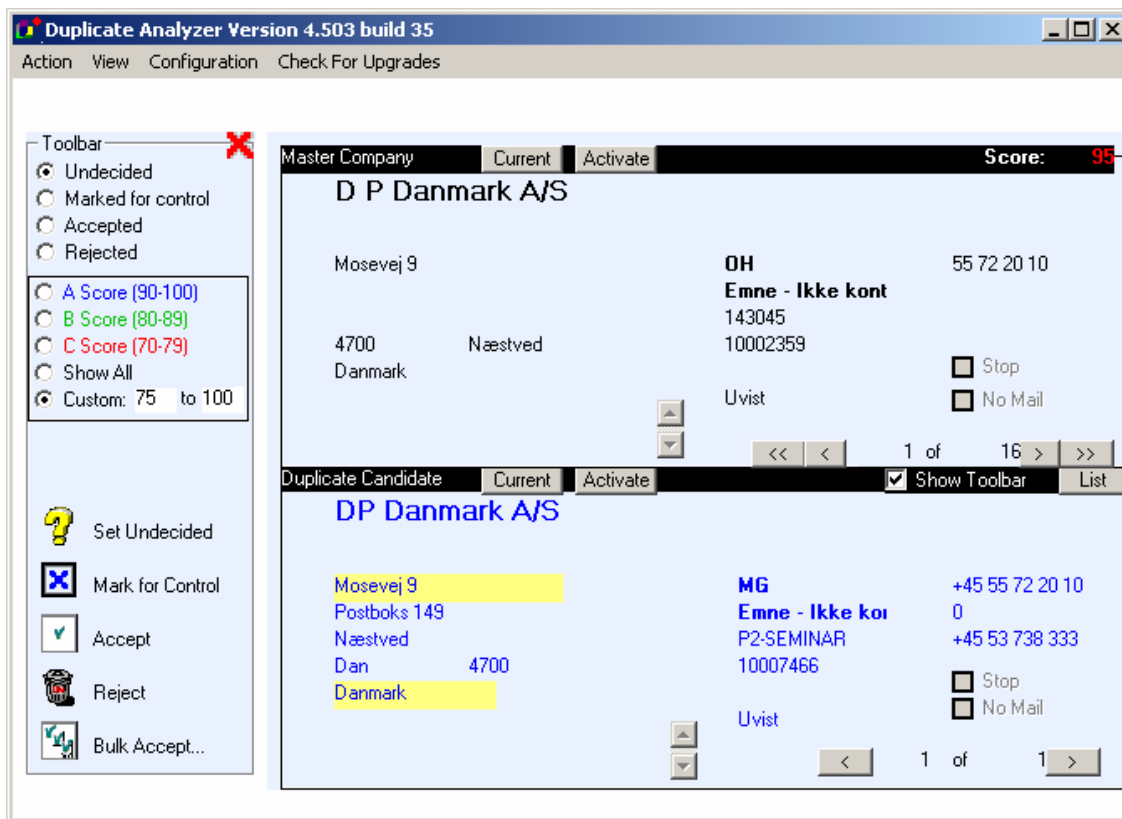
## Processing Analysis Results

The recommended approach to Duplicate Analysis and Cleansing is:

1. Run Analysis 1
2. Apply Master Rule
3. Check Undecided A Scores – are they all certain
  - a. If yes, then use Bulk Accept
  - b. If No, then process and accept manually one by one
4. Then B Scores, then C Scores, then Custom (e.g. Score 60 – 69)
5. Having Processed all Undecided
  - a. Clear All Undecided (See Functions)
  - b. Run the next analysis

An alternative method is:

1. Run Analysis 1
2. Apply Master Rule
3. Bulk Accept all with Score >69
4. Check Accepted C Scores – are they all certain
  - c. If yes, then Execute Bulk Merge
  - d. If No, then process and reject manually one by one
5. Then B Scores, then A Scores. When you reach a level where all are correct – Execute Bulk Merge
6. Having Merged all accepted
  - a. Clear All Undecided (See Functions)
  - b. Run next analysis



The Manual Process is quite efficient.

- Look at the Master/Candidate – take a decision
  - o Is it a duplicate – click 'Accept'
  - o Are you in doubt – click 'Mark for control'
  - o Is it two different companies – click 'Reject'

As soon as you click one of the three – the next Master/Candidate set will be displayed – so you just continue until none is left.

Then set the filter to a lower score. **You do not need to go through all – there's a good chance, that the next analysis will increase the score of those now having a low score.** So rather than spending a lot of time rejecting 90% of the low scores, run another analysis – make you own profiles.

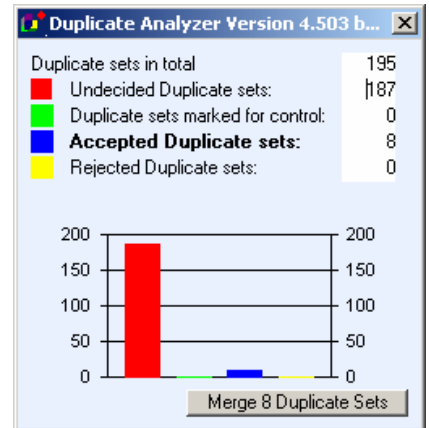
### Execute Bulk Merge

In the Action Menu – select Bulk Merge

The Screen to the right will show. The graph and figures shows the current status of the Analysis Process, i.e. how many are left undecided, how many are marked for control, rejected and accepted.

Click the button to start the physical Merge of the accepted Master/Candidate sets.

Please notice, if any candidates are either defined in the Reject Merge Table, or included in the 'Master Definition', they will be skipped and left to either manually merge or take other action on.



### Person Analysis and Merge

The person part of Duplicate Analyzer 2007 is quite similar to that of the Company Part, and will only be dealt with quite superficially.

Under Analysis, you can tick Person to display Person Profiles.

A Person Duplicate Analysis only compares persons on the same company only, and is therefore quite fast.

Therefore, Person Analysis should be run after merging contacts, since duplicate persons are most likely to occur from companies having been merged.

The information available to compare is usually very limited, basically only First- and lastname are likely to be entered for all persons, though email, phone and other fields can be compared.

The pre-configured profile is configured with 70% weight on First- and Lastname comparison.

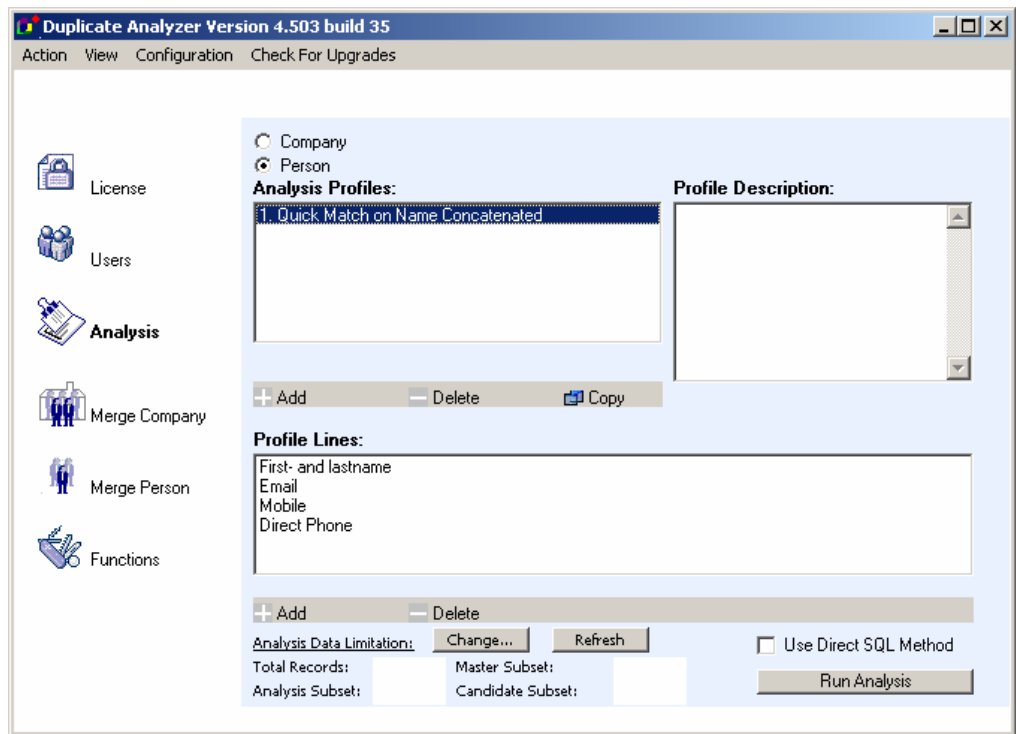
The functionality for Persons is less than that for Companies.

E.g. More page data and Statistics is not available, neither is the functionality to change data on the master – and the list and print options, as well as 'undo merge' and 'reject merge' functions are not implemented for persons.

The 'Master Definition' functionality is available and very useful for persons.

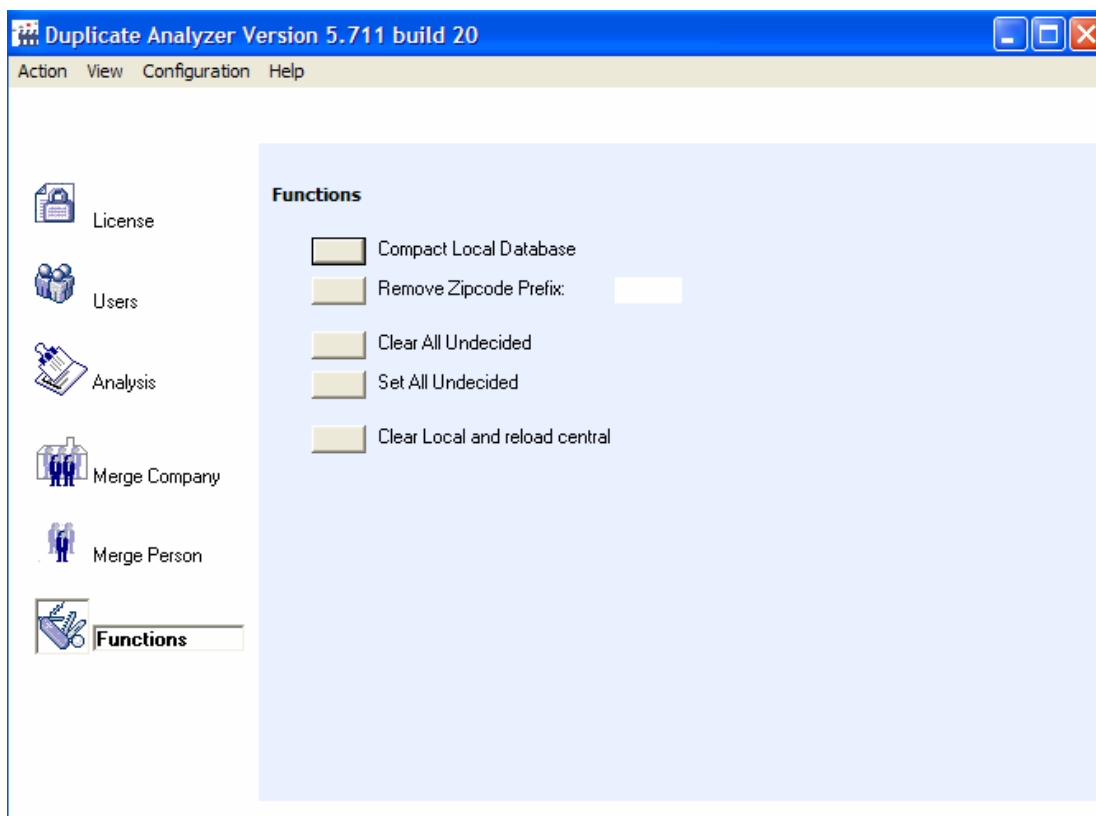
The picture to the right displays a person Master SQL rule, which will toggle all Master/Candidate sets where the candidate is active, and the Master is retired, to the Master being the Active.

Another rule could be to all persons with an email address should be masters. You can apply multiple rules – but should do it priority – with the most important last – in this case not retired is more important than email.





## Advanced Functions



Duplicate Analyzer 2007 has some functions that can be very helpful.

### **Compact Local Database**

Duplicate Analyzer use a Microsoft Access® database for local data processing. This database stores temporary data and can grow to a very large size. Its limit is 2GB. The database will automatically be truncated at every Refresh data, but the Compact Local database can be used if necessary.

### **Remove Zipcode Prefix**

Some SuperOffice Databases contain the zipcode prefix actually stored in the zipcode field, usually as a left over from previous version of SuperOffice or from import or synchronization. In current versions of SuperOffice, the zipcode prefix is retrieved from the country table when displaying zip-code.

If this is the case, using Zipcode Equal profile lines, will not provide good results. This function can remove a part of the zip code if e.g. zipcodes like 'DK-6000' will end up with just '6000' if you type 'DK-' into the field.

### **Clear all Undecided**

During process of an analysis result, you may come to a stage, where only one of the remaining Master/Candidate sets are likely duplicates, so before running a new analysis, you should clear the ones left undecided.

### **Set all Undecided**

This function is useful if you for some reason want to start over. The typical situation is that you have run a Bulk Accept, and want to undo it. However – be careful when using this function – it will clear all your work (Except those you all ready merged). For a full startover, you should first set all undecided, then clear all undecided.

### **Clear Local and reload Central**

This function will clear all history in the Access Database stored locally, and then load all results from the central SuperOffice database. All local undecided duplicate sets will be lost.

In a Duplicate Analyzer multi-user environment, this will ensure that all Merged, Rejected and Accepted duplicate candidate sets are aligned between users.

Also (and especially) if using Duplicate Analyzer on multiple different SuperOffice databases, you should use this function to align you data with the central database, and avoid carrying over incorrect data from another database.