



-Tutorial-

Introduction

This document contains a short tutorial of how to use the standard edition of DATPROF Match. DATPROF Match can help you finding duplicate records in various data sources. This can be of great value when cleaning and deduplicating various types of files such as customer records, inventories or mailing lists.

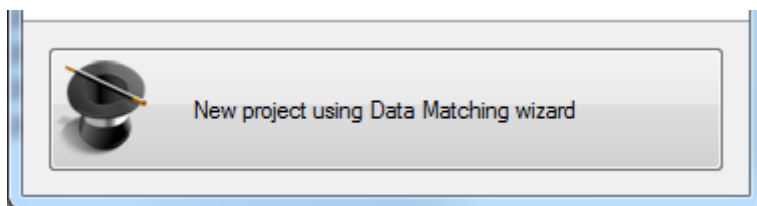
For news and downloads concerning DATPROF Match or other DATPROF products, check our site: www.datprof.com.

Installation and prerequisites

Download and install DATPROF Match on a Windows XP, Vista or 7 machine. Start the program and enter the license key. This key will be mailed to you after registering on the DATPROF site.

Step 1 – Start the wizard

Click the 'New project using Data Matching wizard' button to start the wizard.



Click 'Next'.

Step 2 – Select data source

Choose the type of data source to deduplicate. The standard edition of DATPROF Match supports Excel, Access, CSV and text files¹. Choose one of these and click 'Next', then indicate where the file is located. On our website is a demo file downloadable which is used in this tutorial. Click 'Next' to proceed.

Step 3 – Loading a data source

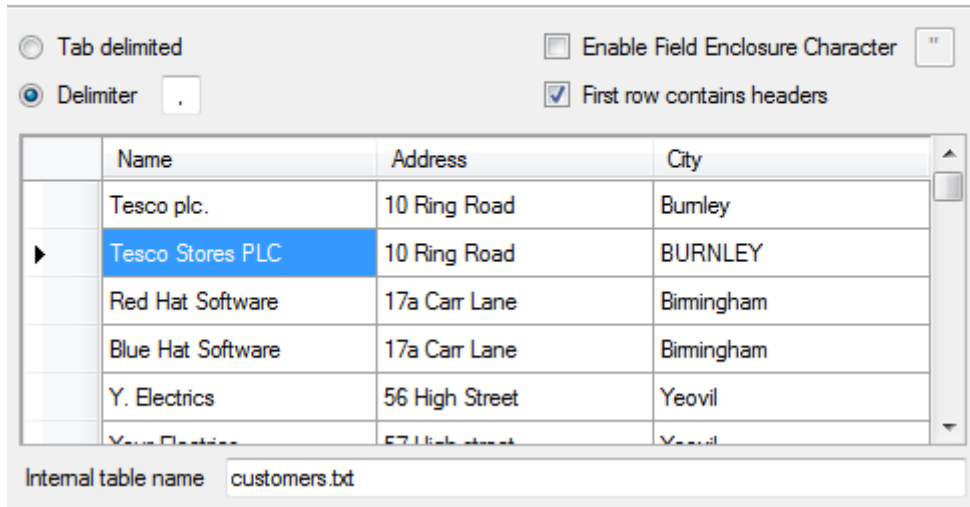
Depending on which data source type you choose, the wizard will slightly differ:

- Excel files require you to select the correct sheet. If the first row contains the column headers you can indicate this using the checkbox 'First row is header';
- Access files require you to select a table;
- CSV files require you to indicate which character marks the end of a field. Usually this is configured automatically. If the first row contains the column headers you can indicate this using the checkbox 'First row is header'.

¹ Besides basic file types DATPROF Match Expert also supports connections to Oracle and Microsoft SQL Server databases. This version is also available from www.datprof.com.

Examine text file

Alter text file properties until the grid displays the text data correctly.



The above screenshot shows the correct settings for the demo file *customers.txt*.

Click 'Next'.

Step 4 – Configuration

In this screen you can indicate which fields will be used in the matching procedure. In a customer file you typically use customer name and address fields. In an inventory file product name and description are common candidates.

When matching, you can choose one of two types of data matching. You can use the 'Fuzzy' setting when you expect values to be similar to one another, but not identical. Use 'Exact' when you want values to be identical.

It is also possible to combine a combination of 'Fuzzy' and 'Exact' settings in your matching procedure. For example, in a customer file you could use 'Fuzzy' on customer name and surname fields while using 'Exact' on address data such as place of residence.

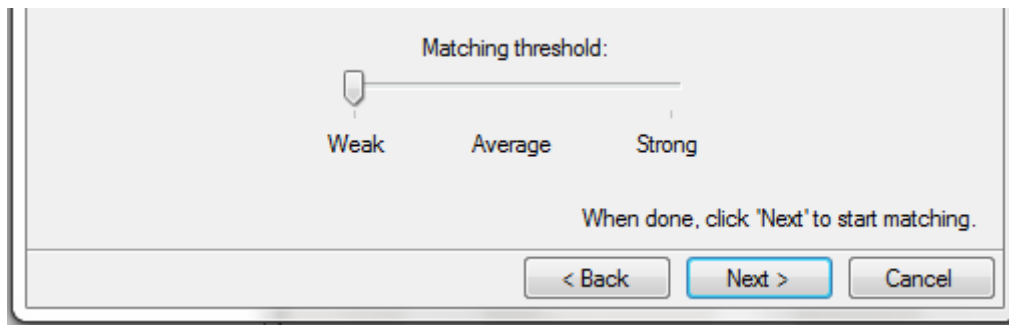
	Column name	Fuzzy match	Exact match	Column contents
	Name	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Tesco plc. ; Tesco Stores PLC ; Red
▶	Address	<input type="checkbox"/>	<input checked="" type="checkbox"/>	10 Ring Road ; 10 Ring Road ; 17a C
	City	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Bumley ; BURNLEY ; Birmingham ; B

When ready, click "Next".

Step 5 – Set matching threshold (optional)

If you used 'Fuzzy' on one or more fields you can indicate a matching threshold on this screen. When using 'Weak', values can be quite different and still be a match while using 'Strong' will require values to be almost exactly alike. 'Average' is the default matching threshold and a good place to start. If you haven't used 'Fuzzy' in the previous step, you only need to click 'Next'.

Handleiding DATPROF Match

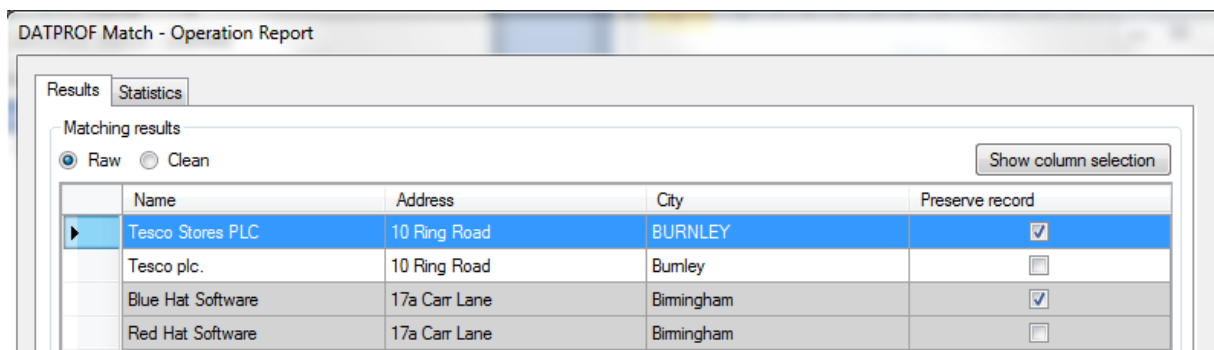


Click 'Next' to start the matching process.

Step 6 – Reviewing results

When the process has finished, click 'View results' to review the matching results. If no results were found, this button will remain disabled.

If this happens and you had expected some results, you can go back using the 'Back button' and change settings.



In the 'Operation report' you can see which duplicate records DATPROF Match found using current settings. You can also review duplicate record distribution along with several other statistics on the 'Statistics' tab.

This concludes the DATPROF Match tutorial. If you have any question or remarks, please contact us at our site: www.datprof.com.

DATPROF Match features:

In the full versions of DATPROF Match Standard and DATPROF Match Expert it is also possible to export your matching results to a variety of files:

- You can export these to a HTML report;
- You can export these to a CSV file (can be read by Excel and most current databases);
- You can export a deduplicated version of your file to CSV. This exported file contains the original records minus the found duplicate records.

DISCLAIMER

ALL WARRANTIES, EXPRESS OR IMPLIED, ARE DISCLAIMED, INCLUDING WITHOUT LIMITATION, ANY AND ALL WARRANTIES CONCERNING THE ACCURACY OF THE CONTENTS, ITS FITNESS OR APPROPRIATENESS FOR A PARTICULAR PURPOSE OR USE, ITS MERCHANTABILITY AND ITS NONINFRINGEMENT OF ANY THIRD PARTY'S INTELLECTUAL PROPERTY RIGHTS. DATPROF EXPRESSLY DISCLAIMS ANY AND ALL RESPONSIBILITIES FOR THE ACCURACY OF THE CONTENTS AND MAKES NO REPRESENTATIONS OR WARRANTIES REGARDING THE CONTENT'S COMPLIANCE WITH ANY APPLICABLE STATUTE, RULE OR REGULATION, OR THE SAFETY OR HEALTH EFFECTS OF THE CONTENTS OR ANY PRODUCT OR SERVICE REFERRED TO IN THE DOCUMENT OR PRODUCED OR RENDERED TO COMPLY WITH THE CONTENTS. DATPROF SHALL NOT BE LIABLE FOR ANY AND ALL DAMAGES, DIRECT OR INDIRECT, ARISING FROM OR RELATING TO ANY USE OF THE CONTENTS CONTAINED HEREIN, INCLUDING WITHOUT LIMITATION ANY AND ALL INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES (INCLUDING DAMAGES FOR LOSS OF BUSINESS, LOSS OF PROFITS, LITIGATION, OR THE LIKE), WHETHER BASED UPON BREACH OF CONTRACT, BREACH OF WARRANTY, TORT (INCLUDING NEGLIGENCE), PRODUCT LIABILITY OR OTHERWISE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THE FOREGOING NEGATION OF DAMAGES IS A FUNDAMENTAL ELEMENT OF THE USE OF THE CONTENTS HEREOF, AND THESE CONTENTS WOULD NOT BE PUBLISHED BY DATPROF WITHOUT SUCH LIMITATIONS.

Copyright ©2011 DATPROF. All rights reserved.