

Comparator

- About..... 2
- Operating Procedure 3
- Data source setup 5
 - Database 5
 - Excel..... 6
 - XML 7
 - Delimited text 9
- Fields configuration settings 10
- Processing settings 11
- Review results..... 12
- Launch modes 14
- Application settings..... 15
 - Console mode settings 15
 - Common settings 16

About

This application may be useful whenever you want to compare table data from two sources, that is:

- Identify whether the rows from the determined sources match using key fields (search for pairs through keys);
- Compare contents of specific fields in the found pairs of rows;
- Identify rows which don't have a match in the other source;
- Identify rows which have more than one pair in the other source (non-unique key);
- Analyze the results and save them in common formats;
- Conduct regular compare tasks based on preconfigured settings (profiles) and send out the results.

System requirements

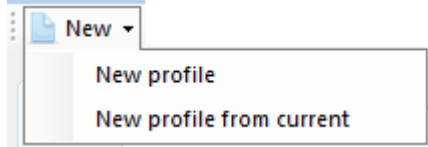
Windows XP/7/8/10/Server

.NET 4.0

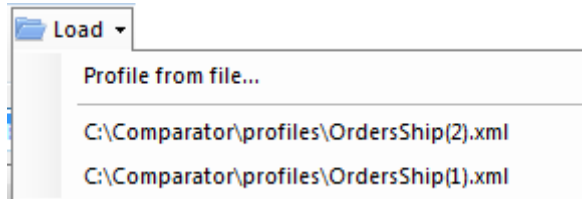
Operating Procedure

First...

Create compare settings from scratch or based on loaded profile:

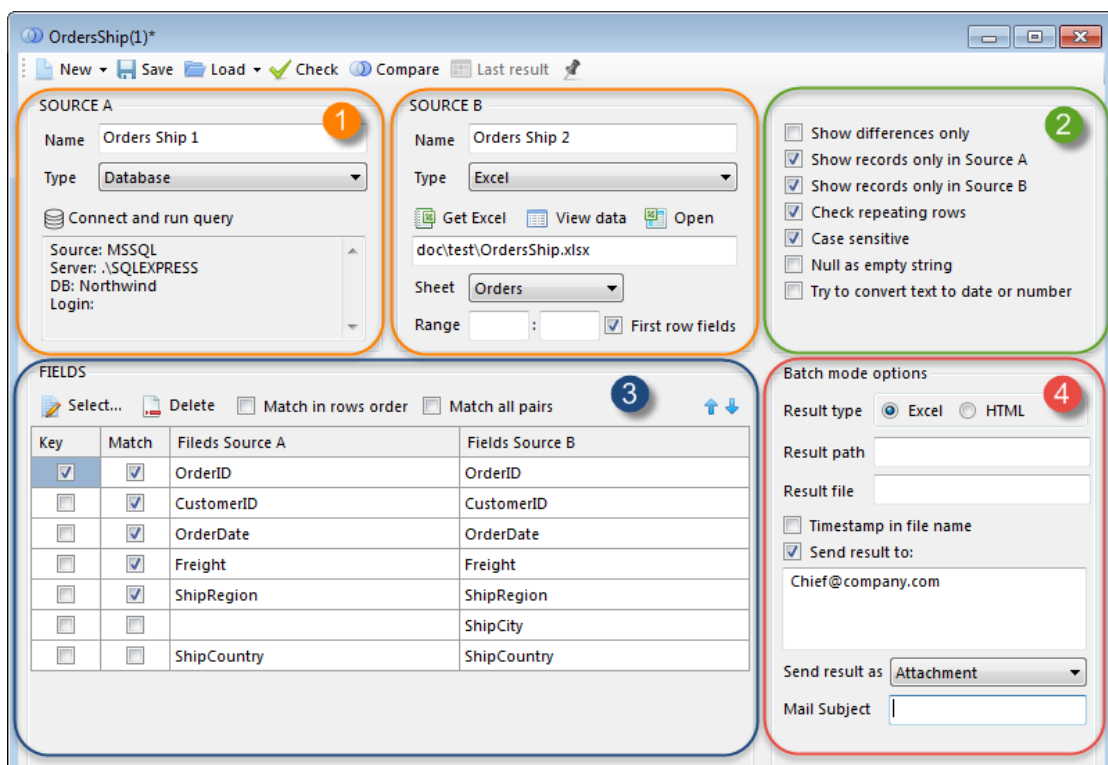


Or load an existing profile from file or recently opened list:



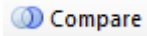
Then...

- Define the two sources;
- Configure settings for data and results processing;
- Set source fields for data compare and display, and define conditions for pair selection;
- Define options of compare startup from command line when required.



And finally...

Run the compare process by pressing and



check the results in a new window:

Compare "Orders Ship 1"(A) and "Orders Ship 2"(B)							
Differences		Identicals		Only in Source A		Only in Source B	
Send to ▾ Search		Find 🔍 Filter 📏 Clear filter ✕		Difference ↔ Key 🔑 Duplicate 📄			
A	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCity	ShipCountry
B	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCity	ShipCountry
	10403	ERNSH	03/01/1997	73.79	[NULL]	Graz	Austria
	10403	ERNSH	03.01.1997 0:00:...	73.792	[NULL]	Graz	Austria
R	10405	LINOD	06/01/1997	34.82	Nueva Esparta		Venezuela
	10405	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	I. de Margarita	Venezuela
R	10405	LINOD	06/01/1997	34.82	Nueva Esparta		Venezuela
	10405	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	I. de Margarita	Venezuela
	10406	QUEEN	07/01/1997	108.04	SP		Brazil
	10406	Queen	07.01.1997 0:00:...	108.04	SP	Sao Paulo	Brazil

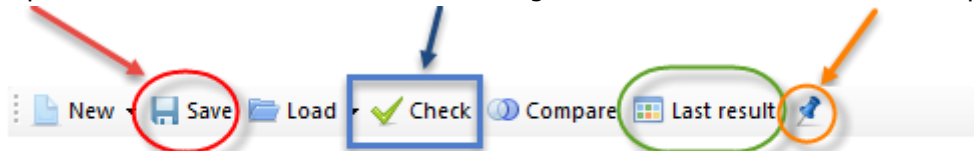
4 pairs : Pair 1

Besides:

Save profile

Check the settings

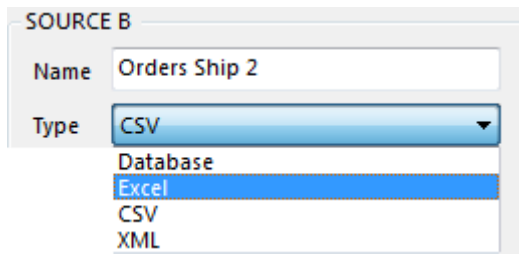
Pin the window on top



Show last compare launch results

Data source setup

Source name will be displayed in compare results, profile name and mailing subject, suggested by default.



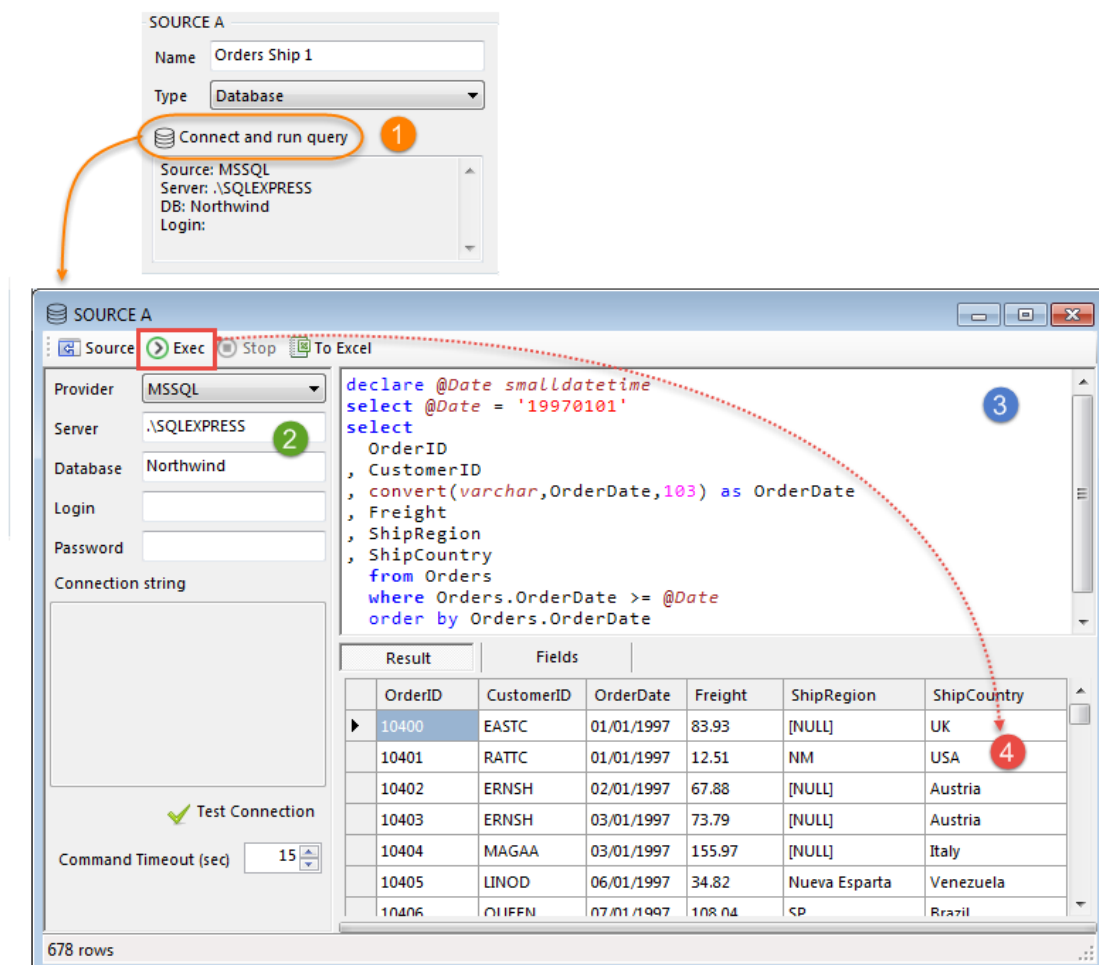
Depending on source type: database, excel, text (csv) or XML file, you will have to configure various settings.

Database

Any source that can be accessed via OleDb or ODBC to request data using SQL code.

To setup:

1. Open the 'connect and run query' configuration form:



SOURCE A

Name: Orders Ship 1

Type: Database

Connect and run query (1)

Source: MSSQL
Server: .\SQLEXPRESS
DB: Northwind
Login:

SOURCE A

Provider: MSSQL (2)

Server: .\SQLEXPRESS (2)

Database: Northwind

Login:

Password:

Connection string:

Test Connection (green checkmark)

Command Timeout (sec): 15

SQL Query:

```
declare @Date smalldatetime
select @Date = '19970101'
select
    OrderID
  , CustomerID
  , convert(varchar, OrderDate, 103) as OrderDate
  , Freight
  , ShipRegion
  , ShipCountry
from Orders
where Orders.OrderDate >= @Date
order by Orders.OrderDate
```

Result Table:

OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry
10400	EASTC	01/01/1997	83.93	[NULL]	UK
10401	RATTC	01/01/1997	12.51	NM	USA
10402	ERNSH	02/01/1997	67.88	[NULL]	Austria
10403	ERNSH	03/01/1997	73.79	[NULL]	Austria
10404	MAGAA	03/01/1997	155.97	[NULL]	Italy
10405	LINOD	06/01/1997	34.82	Nueva Esparta	Venezuela
10406	QUIFF	07/01/1997	108.04	CP	Brazil

678 rows

2. Setup connection settings. To make it easier, the main connection settings via OleDb for MSSQL, Sybase and Oracle are shown in separate menu. For common cases (OleDb or ODBC) connection string is required.
3. Input SQL code script.
4. Run the script. In order to have the up-to-date fields list from the source, SQL script must be run before you setup the fields to compare.

Provider	OleDb
Server	MSSQL Sybase SybaseASE15 Oracle
Database	OleDb
Login	ODBC
Password	
Connection string	Provider=SQLOLEDB;Data Source=.\SQLEXPRESS;Integrated Security=SSPI;Initial Catalog=Northwind

Notes:

- Does the script return several datasets, only the first one will be considered.
- Selected parts of the script can be run for debugging purpose.

Excel

To setup excel file data extract:

1. Open excel file or select one of the currently open files. You can indicate file path based on the application path.
2. Select sheet.

The diagram illustrates the steps to setup Excel file data extract. It shows a 'SOURCE B' dialog box, an Excel file 'OrdersShip.xlsx', and an Excel spreadsheet with data. Arrows and numbers 1 through 6 indicate the sequence of actions:

1. Select file path (From file... OrdersShip.xlsx (C:\Comparator\data)).
2. Select sheet (Orders).
3. Open file (Open button).
4. Select range (A1 : G7).
5. Select first row fields (First row fields checkbox).
6. Review data (View data button).

The Excel spreadsheet shows the following data:

OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry	ShipCity
10402	ERNSH	02.01.1997 0:00:...	67.88		Austria	Graz
10403	ERNSH	03.01.1997 0:00:...	73.792		Austria	Graz
10404	MAGAA	03.01.1997 0:00:...	155.97		Italy	Bergamo
10405	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	Venezuela	I. de Margarita
10405	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	Venezuela	I. de Margarita
10406	Queen	07.01.1997 0:00:...	108.04	SP	Brazil	Sao Paulo
10407	OTTIK	07.01.1997 0:00:...	91.48		Germany	Köln
10408	FOLIG	08.01.1997 0:00:...	11.26		France	Lille

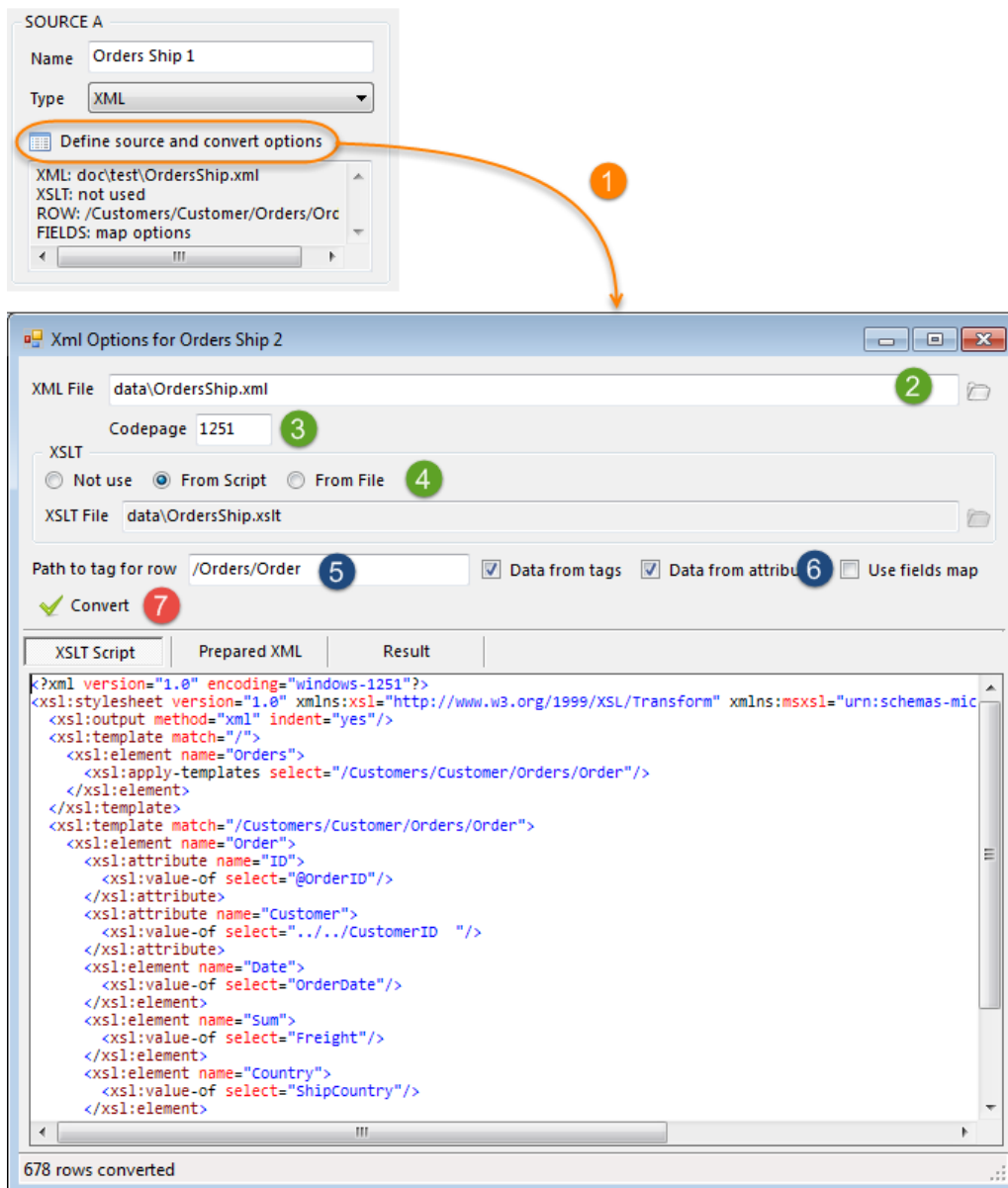
You can also:

3. Select field names to be filled from first row values of the selected range (or the entire sheet if the range is not set). Otherwise field name will be display as the column letter.
4. Open/bring on top selected excel book.
5. Set range either manually or select it directly on the sheet. If any of the range boundaries are not defined, the boundaries of the "used range" will be considered.
6. Review data that will be received in accordance with the settings configured.

XML

To compare, XML data must be first converted into table. To configure conversion process:

1. Open XML configuration settings form:



2. Select XML file (you can indicate file path based on the application path)
3. Set code page number (e.g. 1251, 866) when required or common page name (e.g (UTF8, cp866). When not set, encoding will be defined by default by 'encoding' attribute in XML file header.
4. Select an option of XSLT pre-use before conversion:
 - Not use – not used;
 - From Script – XSLT Script button enables an editor where you need to input XSL script for prior conversion of the XML script into an XML which will be then converted into table view;
 - From File – select XSL file for the same purposes (you can indicate file path based on the application path);
5. Set path (in XPath format) for tag which will be used as basis for the table data row, that will be created as a result of the conversion.

6. Select defining method of table fields that will be created as a result of the conversion:
- Data from tags – fields will be defined by child tags of the tags selected according to XPath, set in 5.
 - Data from attributes – fields will be defined by attributes of the tags selected according to XPath, set in 5.
 - Use fields map – this option will disable all previous fields defining options, and enable path setup (in XPath format) for tags or attributes to define fields. These paths must be set based on the tag defined in 5.
- In this case pressing «Fields Map» button will enable a list in which name and path must be indicated, and also, when required, default value for occurrences when there's no data found in the indicated path.
- If conversion result has already been received, you can add fields from it into the list by selecting respective menu item.

Path to tag for row: ☒ Data from tags ☒ Data from attributes ☒ Use fields map

☒ Convert

Field Name	Path to node for field	Default value
Customer	.././CustomerID	[NULL]
OrderID	@OrderID	[NULL]
OrderDate	OrderDate	[NULL]
Freight	Freight	[NULL]
ShipCountry	ShipCountry	[NULL]
ShipRegion	ShipRegion	

Add field (Ctrl+Ins)
 Add from Result
 Delete field (Ctrl+Del)
 Delete All

7. To verify XML conversion process press «Convert».
- Then:

- Selected XMLfile will be read and pre-converted via XSLT when needed. To view the results press «Prepared XML»;
- Based on settings configured in 5. and 6., data table will be created from the prepared XML file. To view this table press «Result».

```

<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<Customers>
  <Customer>
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <Order OrderID="10643">
        <OrderDate>25/08/1997</OrderDate>
        <Freight>29.4600</Freight>
        <ShipCountry>Germany</ShipCountry>
      </Order>
      <Order OrderID="10692">
        <OrderDate>03/10/1997</OrderDate>
        <Freight>61.0200</Freight>
        <ShipCountry>Germany</ShipCountry>
      </Order>
    </Orders>
  </Customer>

```

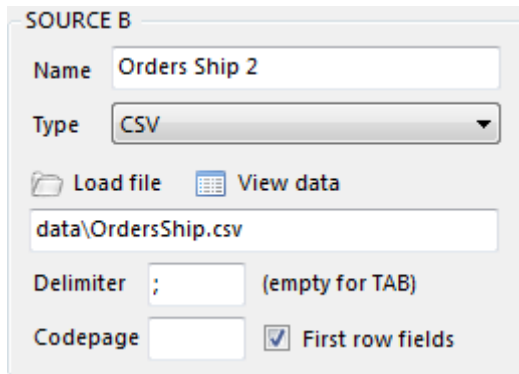
Prepared XML	Result	Fields Map				
	Customer	OrderID	OrderDate	Freight	ShipCountry	ShipRegion
▶	ALFKI	10643	25/08/1997	29.4600	Germany	
	ALFKI	10692	03/10/1997	61.0200	Germany	
	ALFKI	10702	13/10/1997	23.9400	Germany	

Delimited text

Define fields delimiter for selected file ('tab' will be used by default) and encoding: code page number (e.g. 1251, 866) or common page name (e.g. UTF8, cp866). When not set, encoding will be defined by default as current ANSI code page of the Operating System.

Field names will be defined by first row names, similarly to excel file source. To view delimited data in a separate window press "View data".

You can indicate file path based on the application path.

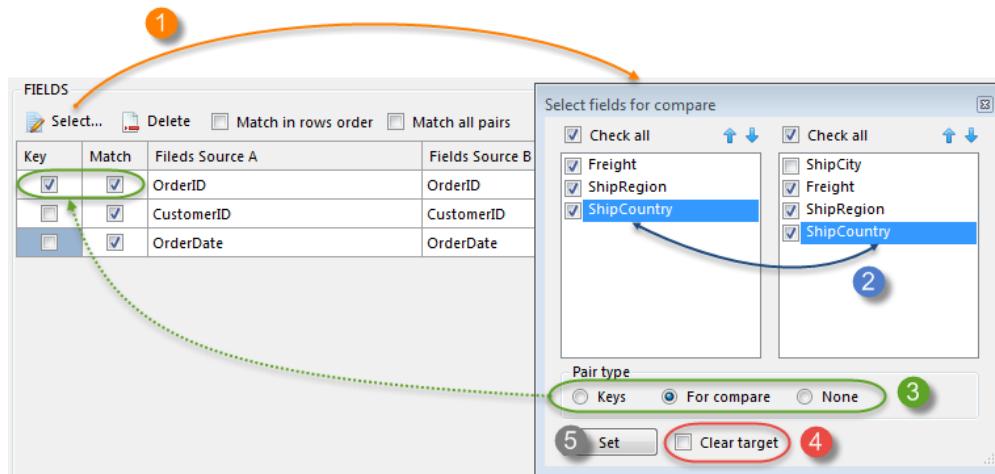


The image shows a dialog box titled "SOURCE B" with the following fields and controls:

- Name:** A text box containing "Orders Ship 2".
- Type:** A dropdown menu set to "CSV".
- Buttons:** "Load file" (with a folder icon) and "View data" (with a grid icon).
- File Path:** A text box containing "data\OrdersShip.csv".
- Delimiter:** A text box containing ";" with the note "(empty for TAB)".
- Codepage:** An empty text box.
- First row fields:** A checked checkbox.

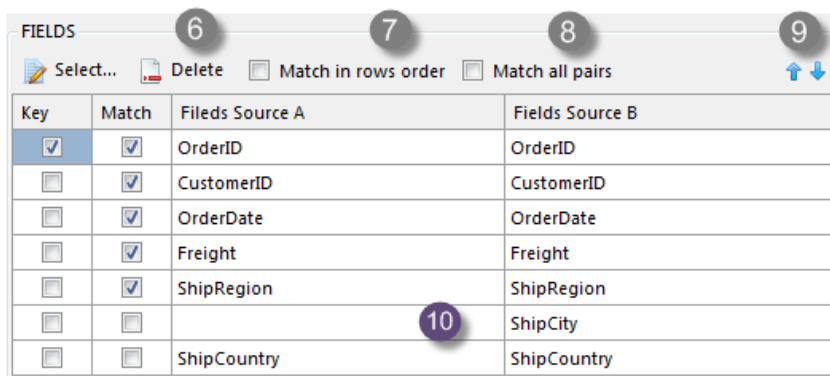
Fields configuration settings

After source setup is complete, define data fields to compare:



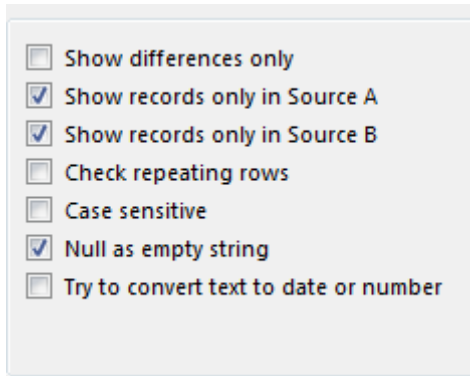
1. Open field list (source A will be shown on the left, source B on the right).
2. Check fields to select them for compare. Correspondence (pairs) will be set in order downwards. To define a specific pair change fields order using blue arrows. The quantity of selected fields in the two sources can be unequal, in this case unpaired fields won't be compared but will be displayed in the results.
3. Define type of selected fields – keys, compare or display only (None).
4. Clear previous selection when needed.
5. Add selected fields to the list.

Fields list can be edited when needed:



6. Delete selected rows.
7. Define pairs by row order, ignoring keys.
8. Match all the fields.
9. Change fields order in the results.
10. Define fields processing type:
 'Key' – key fields, when they match, it will define row pairs from the sources;
 'Match' – this data will be compared in the found row pairs.

Processing settings

A screenshot of a software dialog box titled "Processing settings". It contains seven checkboxes with corresponding labels. The checkboxes for "Show records only in Source A", "Show records only in Source B", and "Null as empty string" are checked, while the others are unchecked.

<input type="checkbox"/>	Show differences only
<input checked="" type="checkbox"/>	Show records only in Source A
<input checked="" type="checkbox"/>	Show records only in Source B
<input type="checkbox"/>	Check repeating rows
<input type="checkbox"/>	Case sensitive
<input checked="" type="checkbox"/>	Null as empty string
<input type="checkbox"/>	Try to convert text to date or number

- **Show differences only** – do not include rows that match.
- **Show records only in Source A** – include unpaired fields from source A.
- **Show records only in Source B** – include unpaired fields from source B.
- **Check repeating rows** – search for all pairs for each key fields set; if more than one pair found it will be marked as repeating. This setting may slow down the process. When not checked, the search will cease after finding first pair.
- **Case sensitive** – use case for compare process.
- **Null as empty string** – null value will be processed as empty char.
- **Try to convert text to date or number** – will try to convert text fields into date or number format. E.g. '20151231' and '31.12.2015' will be matched as equal values.

Review results

Source A:

Result	Fields				
OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry
10401	RATTC	01/01/1997	12.51	NM	USA
10402	ERNSH	02/01/1997	67.88	[NULL]	Austria
10403	ERNSH	03/01/1997	73.79	[NULL]	Austria
10404	MAGAA	03/01/1997	155.97	[NULL]	Italy
10405	LINOD	06/01/1997	34.82	Nueva Esparta	Venezuela
10406	QUEEN	07/01/1997	108.04	SP	Brazil

Source B:

A	B	C	D	E	F	G
OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry	ShipCity
10402	ERNSH	02.01.1997	67.88		Austria	Graz
10403	ERNSH	03.01.1997	73.792		Austria	Graz
10404	MAGAA	03.01.1997	155.97		Italy	Bergamo
10405	LINOD	06.01.1997	34.82	Nueva Sparta	Venezuela	I. de Margarita
10405	LINOD	06.01.1997	34.82	Nueva Sparta	Venezuela	I. de Margarita
10406	Queen	07.01.1997	108.04	SP	Brazil	Sao Paulo

Compare "Orders Ship 1"(A) and "Orders Ship 2"(B)

1 Differences

Identicals

Only in Source A

Only in Source B

Send to

Search

4

Find

Filter

Clear filter

5 Difference

Key

Duplicate

A	B	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCity	ShipCountry
		10403	ERNSH	03/01/1997	73.79	[NULL]		Austria
		10403	ERNSH	03.01.1997 0:00:...	73.792	[NULL]	Graz	Austria
R		10405	LINOD	06/01/1997	34.82	Nueva Esparta		Venezuela
		10405	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	I. de Margarita	Venezuela
R		10405	LINOD	06/01/1997	34.82	Nueva Esparta		Venezuela
		Repeated row #5	LINOD	06.01.1997 0:00:...	34.82	Nueva Sparta	I. de Margarita	Venezuela
		10406	QUEEN	07/01/1997	108.04	SP		Brazil
		10406	Queen	07.01.1997 0:00:...	108.04	SP	Sao Paulo	Brazil

4 pairs : Pair 1

Send to 7

Excel (F9) HTML HTML File...

A	B	C	D	E	F	G	H
1	Compare "Orders Ship 1"(A) and "Orders Ship 2"(B)						
2	Rows in A: 678						
3	Rows in B: 832						
4	Differences: 4 row pairs						
5	Identicals: 676 row pairs						
6	Only in A: 0 rows						
7	Only in B: 152 rows						
8							
9	Differences						
10	A	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry
11	B	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCity
12		10403	ERNSH	03.01.1997	73.79		Austria
13		10403	ERNSH	03.01.1997 0:00	73.792		Graz
14	R117	10405	LINOD	06.01.1997	34.82	Nueva Esparta	Venezuela
15		10405	LINOD	06.01.1997 0:00	34.82	Nueva Sparta	I. de Margarita
16	R117	10405	LINOD	06.01.1997	34.82	Nueva Esparta	Venezuela
17		10405	LINOD	06.01.1997 0:00	34.82	Nueva Sparta	I. de Margarita
18		10406	QUEEN	07.01.1997	108.04	SP	Brazil
19		10406	Queen	07.01.1997 0:00	108.04	SP	Sao Paulo
20	Top						
21							
22	Identicals						
23	A	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCountry
24	B	OrderID	CustomerID	OrderDate	Freight	ShipRegion	ShipCity
25	RQ	10400	EASTC	01.01.1997	83.93		UK
26		10400	EASTC	01.01.1997 0:00	83.93		London

result

- Results are shown in a separate window where various results set will be available, depending on compare settings. You can switch between the sets:
 - Differences – row pairs with differences;
 - Identicals – row pairs with no differences;
 - Only in Source A – rows in Source A for which a pair couldn't be found in Source B;
 - Only in Source B – rows in Source B for which a pair couldn't be found in Source A.

2. Differences are marked red; key fields are marked blue; matched fields are marked green.
3. Headers and rows in the result sets "Differences" and "Identicals" are showed in pairs: first row from source A, second row from source B (the latter is marked with colour). If one row was included in more than one pair (6), it will be marked as 'repeating' with letter 'R', and have a hint showing its number in the source (starting from 0).
4. Use 'Search' field to search the entire result table or a specific column, or use it to filter data in the current column.
5. Use 'Difference' button to jump to the next difference, 'Key' button – to the next key field, 'Duplicate' button – to the next 'repeating' row.
There are hints on all buttons for the hot keys definition.
7. The results can be opened in Excel or HTML, a file 'results.xls' or 'results.htm' will be created in folder defined in application settings. the results can be also saved as HTML.

Launch modes

1. Windows mode (using common application settings)

comparator [-profile:"[path]filename"]

[-profile:"[path]filename"] – profile will be loaded.

When profile not defined, application will be launched with new empty profile or template profile (as configured in application settings).

2. Console mode (using console mode application settings)

comparator -batch -profile:"[path]filename" [-path:"path"] [-log:"[path]filename"] [-type:excel/html] [-sendto:"address[;address...]" [-file:"filename"] [-open]

Required settings:

-batch – defines console mode – console mode of the specified profile will be used (see further);

-profile:"pathfilename" – file containing the profile;

Optional settings (ignored if -batch is not defined), have higher priority than respective settings from the application profile and settings:

-path:"path" – path where the results will be saved (do not use '/' in the end);

-log:"[path]filename" – log file;

-type:excel/html – result file type (excel or html);

-sendto:"address[;address...]" – mailing list; when specified but the list is empty, the results will not be sent;

-file:"filename" – result file name;

-open – when specified, the result file will be opened.

After launching in console mode, the following will happen: profile data check, search and connect to the sources, compare and create result file, mail and/or open the result file.

The process will be written in the log file:

```
[08.02.2016 20:37:07.776] >>> Start process for "testprofile.xml"
[08.02.2016 20:37:08.192] Profile loaded
[08.02.2016 20:37:09.977] Compared
[08.02.2016 20:37:10.002] Save result in "C:\Comparator\results\testprofile__20160208_203709.xls"
```

Application settings

Console mode settings

Batch mode options

Result type ☒ Excel ☐ HTML

Result path \\tmp\compare_results\

Result file orders_ship.xls

☒ Timestamp in file name

☒ Send result to:

Chief@company.com;
Manager@company.com

Send result as Attachment

Mail Subject Compare orders

Attachment
GZIP
Text
Link

- **Result type** – result file type (excel или html).
- **Result path** – result file path (application settings will be used by default).
- **Result file** – result file name (profile name plus file extension depending on the result type will be used by default).
- **Timestamp in file name** – when specified, date and time of the compare launch will be added to the file name.
- **Send result to** – when set:
 - Specify mailing list (separated by semicolon)
 - Choose sending option of the result file in **Send result as**
 - Attachment;
 - ZIP – file attached as Gzip archive (filename.gz) – optimal for large files;
 - Text – sent as HTML text in message body;
 - Link – link to the result file, no attachment;
 - Optional: you can define message subject in **Mail Subject** ("Compare 'Source A Name' and 'Source B Name'" by default)

When not specified or the list is empty, the results will not be sent.

For other mailing options, such as connect SMTP server see common settings.

Common settings

For global settings use **Comparator.xml** file which should be stored in application folder.

If it doesn't exist, the file will be created at first launch.

Example:

```
<?xml version="1.0" encoding="windows-1251"?>
<Config xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <optionsList>
    <Options Host="BATCH">
      <ProfileFolder>C:\Comparator\profiles</ProfileFolder>
      <ResultFolder>C:\Comparator\results</ResultFolder>
      <LogFile>C:\Comparator\Comparator.log</LogFile>
      <HtmlStylesFile>C:\Comparator\styles.css</HtmlStylesFile>
      <SendOptions>
        <Server>smtp.server.ru</Server>
        <Port />
        <Ssl>>false</Ssl>
        <User />
        <Pwd />
        <PwdEncr />
        <From>Comparator@server.ru</From>
        <FromAlias>Comparator</FromAlias>
      </SendOptions>
    </Options>
    <Options Host="HOST1">
      <RecentFiles>
        <file>C:\Comparator\profiles\Источник1_Источник2.xml</file>
        <file>C:\Comparator\profiles\test.xml</file>
      </RecentFiles>
      <ProfileFolder>C:\Comparator\profiles</ProfileFolder>
      <ResultFolder>C:\Comparator\results</ResultFolder>
      <HtmlStylesFile>C:\Comparator\styles.css</HtmlStylesFile>
      <PatternFile>C:\Comparator\pattern.xml</PatternFile>
    </Options>
    <Options Host="HOST2">
      .....
    </Options>
  </optionsList>
</Config>
```

Console mode settings are in the tag <Options> with attribute Host="BATCH".

Settings for normal (windows) mode are defined for each host in the tag <Options> with attribute Host="Host_name".

On first launch the host will set default settings, which can be changed by editing the file manually.

Following tags can be used to store settings:

- <ProfileFolder> - folder where the profiles are stored ('profile' in the application folder by default);
- <ResultFolder> - folder where the result files will be stored ('result' in the application folder by default);
- <LogFile> - log file name (comparator.log in the application folder by default);
- <HtmlStylesFile> - css file for results style design (style.css in the application folder by default);
- <PatternFile> - default pattern profile;


```
<SendOptions> - mailing options:
  <Server> - SMTP server;
  <Port> - Port (25 by default);
  <Ssl> - use SSL (true/false);
  <User> - login for SMTP server;
  <Pwd> - password for SMTP server, specified here and then encrypted on first launch,
    stored in <PwdEncr> and erased from <Pwd>;
  <PwdEncr> - encrypted password for SMTP server;
  <From> - sender for mailing list;
  <FromAlias> - alias of the sender;
</SendOptions>

<RecentFiles> - recent profiles:
  <file> path/name of the latest opened profile;
  <file> path/name of the previous opened profile and so on;
</RecentFiles>
```