



Database Indexer 1.1

User Guide

Contents

Installation	2
Translation	2
Getting Started.....	3
Add a New Database	4
Add a data source.....	5
Field Display Names Option	6
Set Indexing Options	7
Custom Indexing Options.....	8
Default Indexing Options.....	9
Schedule the Index updates.....	10
Remove an index.....	10
Incremental Index Updates	11
Custom Queries	12
Document Extraction	13
Enumerable Fields.....	14
Get Help.....	15
Settings	15
Feedback	15
Appendix	16
Evaluation using the demo Azure SQL database:.....	16
Evaluation using the demo Azure SQL database containing documents	18
Evaluation of Custom Queries using MySQL:.....	19
Evaluation of Custom Query using Azure MSSQL.....	19

Installation

Database Indexer is an add-on product for use with dtSearch Desktop or dtSearch Network.

Before installing Database Indexer you must install a copy of dtSearch Desktop or dtSearch Network* on the machine where you will create and maintain the indexes.

dtSearch Network licenses include use of an extra copy on a single additional machine for indexing and/or for shared network index access.

Install Database Indexer on the same machine where your copy of dtSearch Network is installed, it will automatically detect where dtSearch Network is installed and will warn if a copy cannot be found.

The default install folder is ...**Program Files (x86)\DTSUK\Database Indexer**

IMPORTANT: Before installing any **beta** build, from Windows Control Panel, uninstall previous versions from the machine. Ensure that the install folder is completely empty before installing the beta version.

*Needs dtSearch version 7.94.8620 or later

Translation

The default language of the user interface is English. To translate the user interface, make a copy of the **Database Indexer - template.lang** file which is in the **\I18N** subfolder.

Instructions are in the file, translate the appropriate text using a text editor (e.g. Notepad), save it in utf-8 format (we suggest you rename it using IETF language tags (e.g. **fr-CA.lang** for Canadian French) and place the file in the same folder as Database Indexer.exe.

If a file with filename extension **.lang** is found when Database Indexer is run, it will read the file to translate the user interface.

Place the untranslated copy of the template file in the same folder as the **Database Indexer.exe** file and run Database Indexer, you will see all the English text is replaced by pseudo-language with a width 30% wider than English to take into account the typical expansion that occurs in translating from English to many other languages. Text Labels in Database Indexer will automatically expand to accommodate the length of the text. If you experience any issues translating the user interface please contact support@dtsearch.co.uk for assistance.

Getting Started

The steps to create an index are

- [Add a New Database](#)
- [Add a Data Source](#)
- [Set the index options](#)
- [Set a schedule](#) for the index updating if needed.

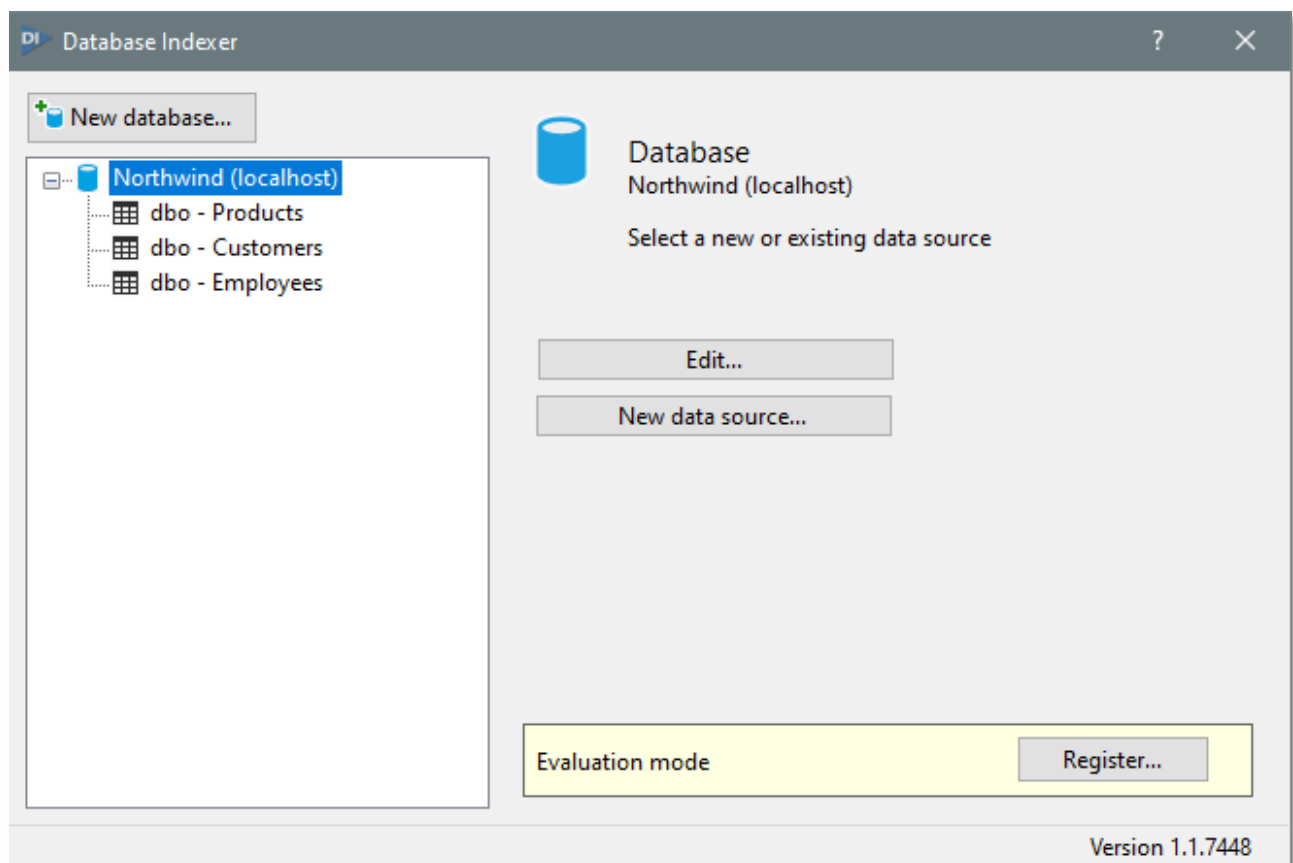
Click the **Edit...** button to change any of the settings later.

Evaluation mode

If you are evaluating, you might want to try our demonstration databases before testing with your own data. You will find details in the [Appendix](#).

Evaluation mode is limited to 10,000 rows per index and scheduled indexing is not functional.

To remove the limits, click **Register...** and enter your serial number; you can find your serial number later by clicking on the help (?) icon in the title bar.

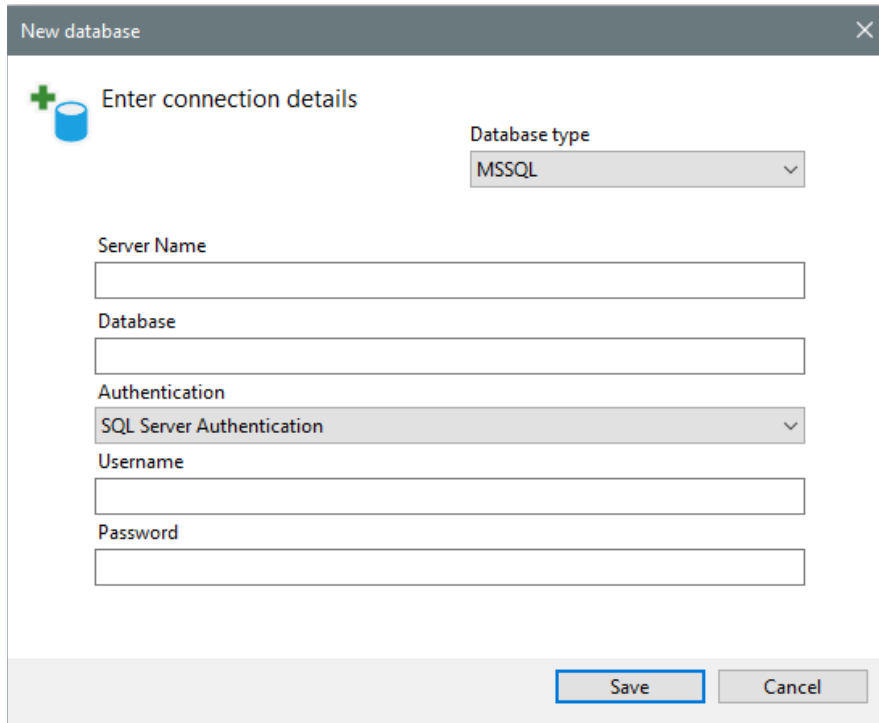


Add a New Database

Select the **database type** from the drop-down list, enter the requested parameters of your database then **Save** your settings.

You may prefer to use our cloud hosted demonstration databases while evaluating. See [Appendix](#).

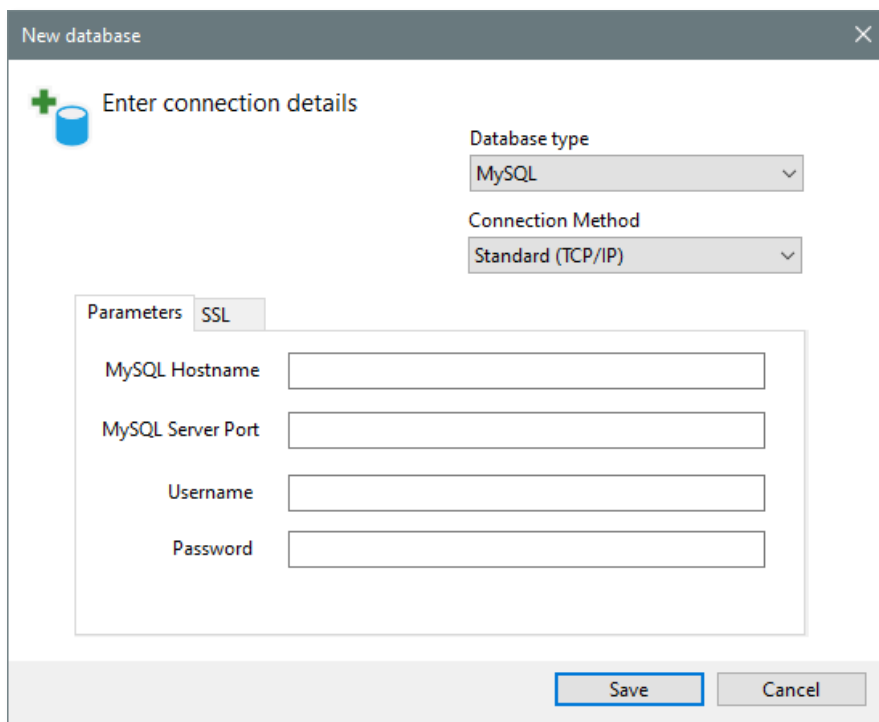
The images below show the options for the MS SQL and mySQL database types.



The screenshot shows a 'New database' dialog box with a title bar and a close button. It features a green plus icon and a blue database cylinder icon. The main heading is 'Enter connection details'. On the right, there is a 'Database type' dropdown menu set to 'MSSQL'. Below this are several input fields: 'Server Name', 'Database', 'Authentication' (a dropdown menu set to 'SQL Server Authentication'), 'Username', and 'Password'. At the bottom right, there are 'Save' and 'Cancel' buttons.

MS SQL database type

Select **Authentication** from **SQL Server Authentication** or **Windows Authentication**



The screenshot shows a 'New database' dialog box with a title bar and a close button. It features a green plus icon and a blue database cylinder icon. The main heading is 'Enter connection details'. On the right, there are two dropdown menus: 'Database type' set to 'MySQL' and 'Connection Method' set to 'Standard (TCP/IP)'. Below these is a 'Parameters' section with a sub-tab 'SSL'. Inside this section are four input fields: 'MySQL Hostname', 'MySQL Server Port', 'Username', and 'Password'. At the bottom right, there are 'Save' and 'Cancel' buttons.

MySQL database type

Select **Connection method** from **Standard (TCP/IP)** or **Local Socket/Pipe** or **Standard TCP/IP Over SSH**

Add a data source

After you have added a new database, select it from the list and press the **New Data Source** button.

Select the **schema** from the drop-down list.

Choose to select a **Table** or create a [Custom Query](#).

If any of the columns contain a document or file path to a document, you might want to set up [Document Extraction](#) so that it is displayed faster from search results.

If the data source contains columns that can be enumerated, you might want to set up [Enumerable Fields](#) to provide faceted search navigation. (New in version 1.1)

Dialog box titled "Edit Data Source" showing configuration options for indexing data. The "Schema" is set to "dbo". The "Table" option is selected, with "Products" chosen from the dropdown. The "Custom Query" option is unselected, with "Not Set" in the text box and a "Setup..." button. The "Columns" section has "All" and "None" buttons, and a list of columns with checkboxes: ProductID, ProductName, SupplierID, CategoryID, and QuantityPerUnit, all of which are checked. The "Document Extraction" section has a "Don't Extract" text box, a close button (X), and a "Setup..." button. The "Enumerable Fields" section has a "3 Enumerable Fields" text box, a close button (X), and a "Setup..." button. At the bottom, there are "Next >" and "Cancel" buttons.

When finished click the **Next >** button to optionally set up user friendly Field Display Names and setup the index options.

Field Display Names Option

You can optionally enter user friendly “display names” for each field as shown in the example below. (New in version 1.1)

Click **Next >** when finished.

Column	Display Name
BirthDate	D.O.B.
HireDate	Hire Date
Address	Address
City	City
Region	State
PostalCode	Zip Code
Country	Country
HomePhone	Phone (Home)

These user-friendly Display Names will appear in dtSearch Desktop Search Dialog in the **fields...** drop-down list and in the Fields... list in the Universal Search Client.

Set Indexing Options

You can create a **New Index** or **Use an Existing dtSearch index**. The index name can consist of several words and can contain spaces.

Choose an index **Location** on the same machine where you have read/write permissions.

CAUTION: If you choose **Use Existing Index** and select **Recreate Index** any data not previously indexed by the Database Indexer will be lost. (The software will also warn you if you select this option)

Incremental Update requires a column in the table that has a Last Modified Date.

Database Indexer needs read-only permission to the database. Incremental updates are recommended where possible to reduce database activity.

The screenshot shows the 'Edit Data Source' dialog box with the 'Set Indexing Options' tab selected. The dialog is divided into several sections:

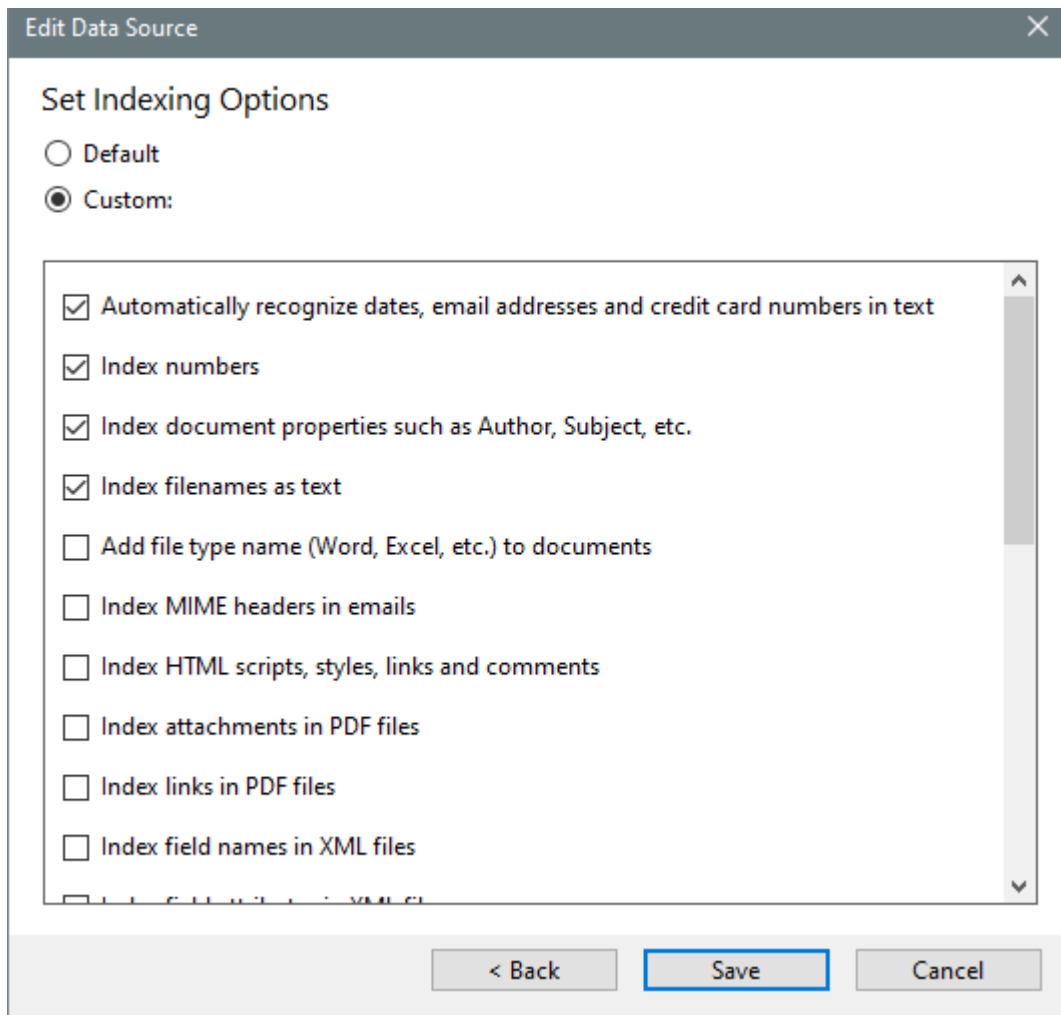
- Indexing Options:** Two radio buttons are present. The 'New Index' option is unselected, while the 'Use Existing Index' option is selected.
- New Index Section:** Contains a 'Name' text box with the value 'dbo - Products' and a 'Location' text box with the value 'D:\Documents\Database Indexes'. A 'Browse...' button is located to the right of the location text box.
- Use Existing Index Section:** Contains a 'dtSearch® Index File (*.ix)' text box with the value 'D:\Documents\Database Indexes\dbo - Products\his'. A 'Browse...' button is located to the right of the text box.
- Indexer options Section:** Contains two radio buttons. The 'Recreate Index (All existing content will be removed)' option is selected. The 'Incremental Update' option is unselected. Below the 'Incremental Update' option is a text box containing 'Not Setup' and a 'Setup...' button.
- Navigation:** At the bottom of the dialog are three buttons: '< Back', 'Next >', and 'Cancel'. The 'Next >' button is highlighted with a blue border.

Finally press the **Next >** button and choose [Default](#) or [Custom](#) indexing options

Custom Indexing Options

Select Default or Custom indexing options.

If you want most of your indexes to have the same settings, select **Default** and edit the [Default Indexing Options](#) from the Help (?) dialog later, otherwise select **Custom** then check the options you need and **Save** your settings.



Generally the more indexing options that are selected the larger the index size will be.

If the documents being indexed contain many numbers, and if these numbers do not have to be searchable, this setting can reduce index size considerably. Note: 'numbers' are defined as a word starting with a digit, a word such as PO325379 is NOT considered a number.

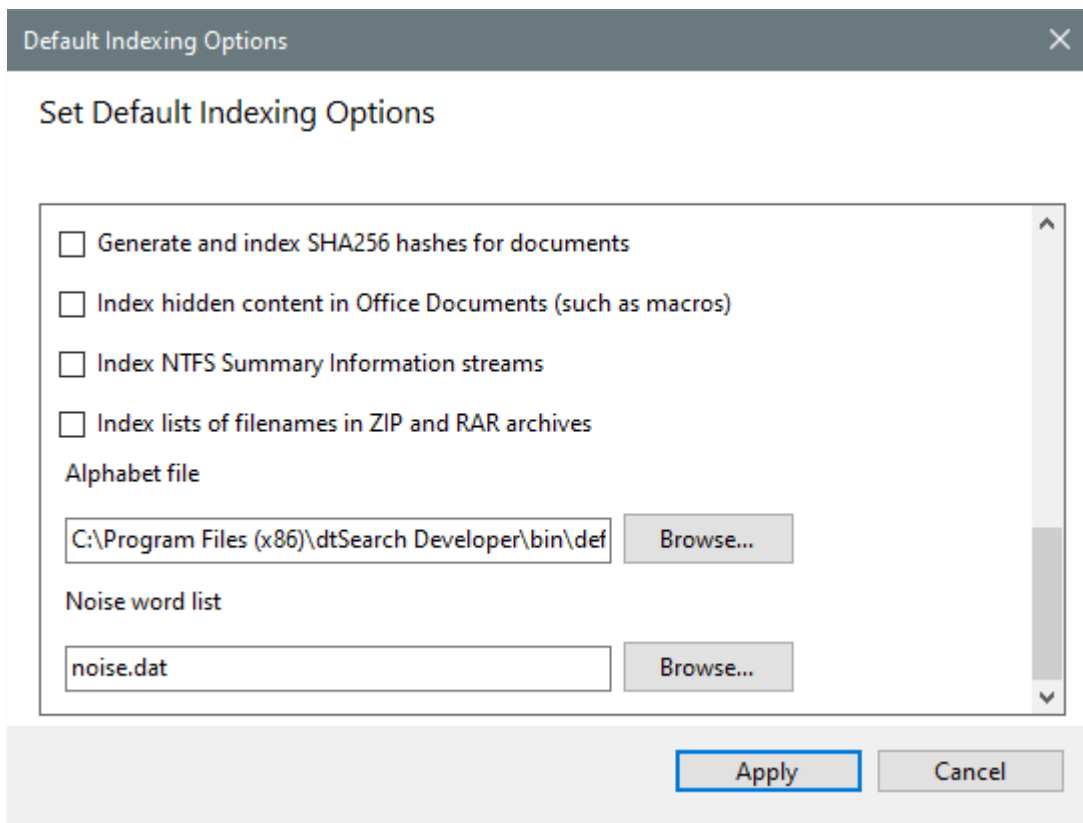
See Index Size & performance: <https://support.dtsearch.com/dts0142.htm>

Default Indexing Options

To change the default indexing options, click on the help (?) icon in the title bar and click on the Default Indexing Options button.

The settings you choose will be applied to all index tasks that are set to use the default options.

From here you can also choose your Alphabet file and noise word list from your installed copy of dtSearch Desktop\Network.

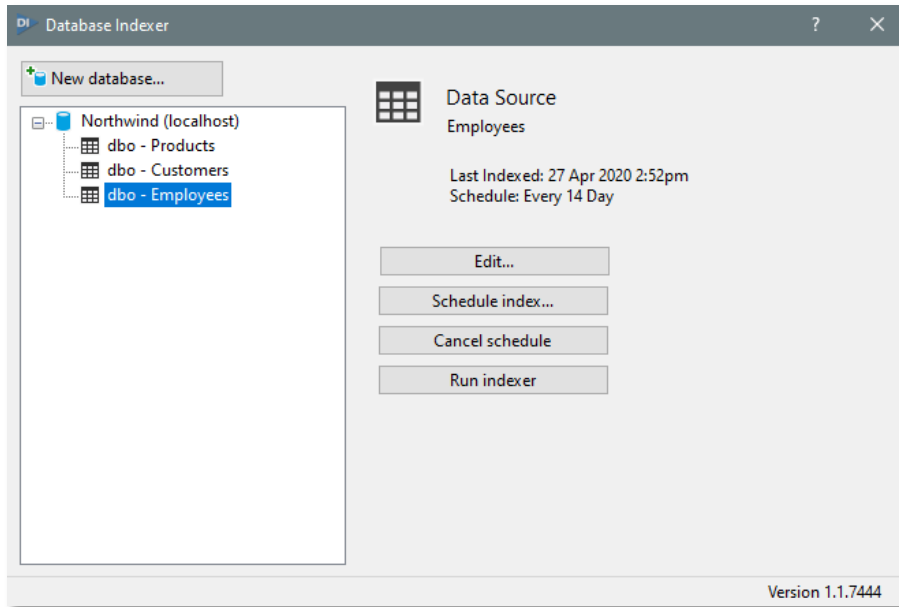


Note:

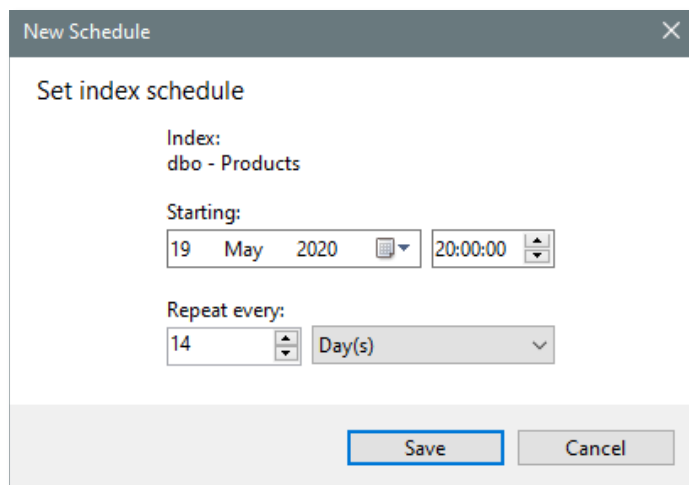
Noise word list. When you create an index, the index will store its own copy of the noise word list. Changes you make to the noise word list will be reflected in future indexes you create but will not affect existing indexes. If you make changes to the noise word file and want them applied to an existing index, you should rebuild the index. The noise.dat file supplied with dtSearch contains English words. If you are searching non-English text, either edit the file to remove all the words and save it as none.dat or enter words appropriate to the language of the text you will be searching and save it with an appropriate name.

Alphabet file. The alphabet file determines how dtSearch interprets certain characters in your documents. If you modify the default.abc file, for example to make a character such as + searchable, we recommend using **Save as...** to save with a different filename. You will need to rebuild your index if you edit the alphabet file.

Schedule the Index updates



Select the index to schedule in the left-hand pane, then click on **Schedule index...**

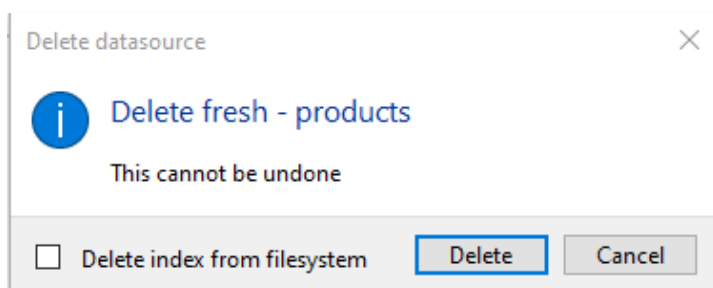


Set the **Starting** date and time, then set the repetition period in days or months.

To rebuild or update the index daily, you need to ensure that it can be rebuilt or updated within 24 hours; for large databases rebuilding might not be possible within 24 hours, it is recommended to use [incremental updates](#) on a scheduled basis.

Remove an index

To remove an index, right-click on an item in the list and choose **Delete**. From the dialog that appears, select the **Delete index from filesystem** checkbox and press the **Delete** button.

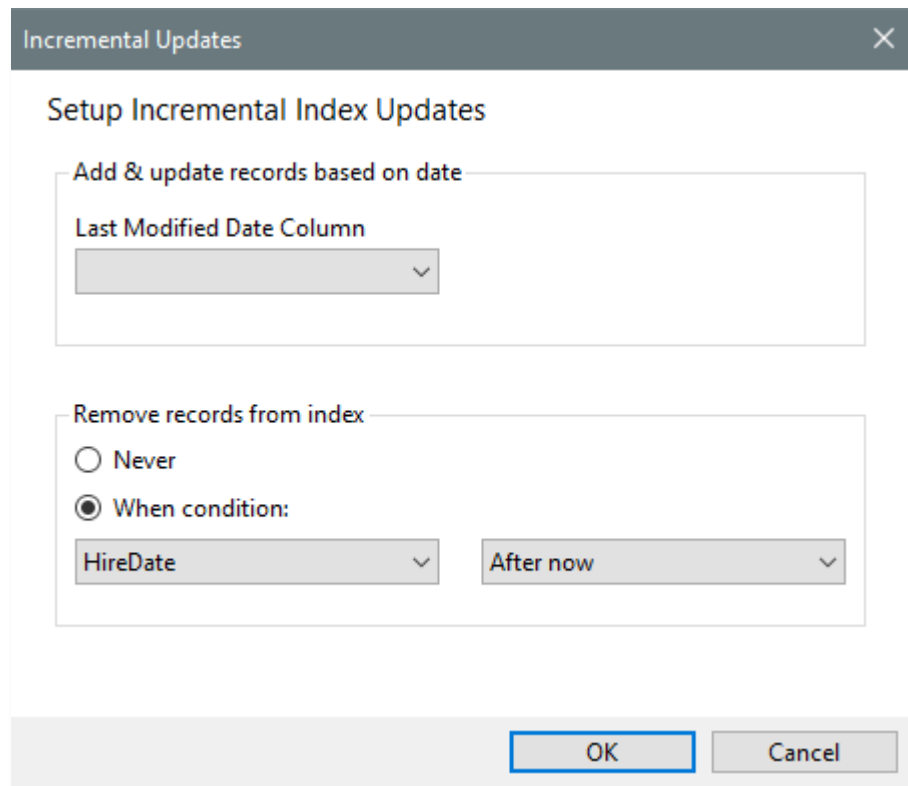


Incremental Index Updates

To set incremental index updates, select the index, then click on **Edit... Next > Next >** then select the **Incremental Update** radio button and click **Setup...**

Incremental updates re-index only rows in the table that have been modified.

Incremental index updates require a column in the table that has a Last Modified Date.



Remove Records from Index

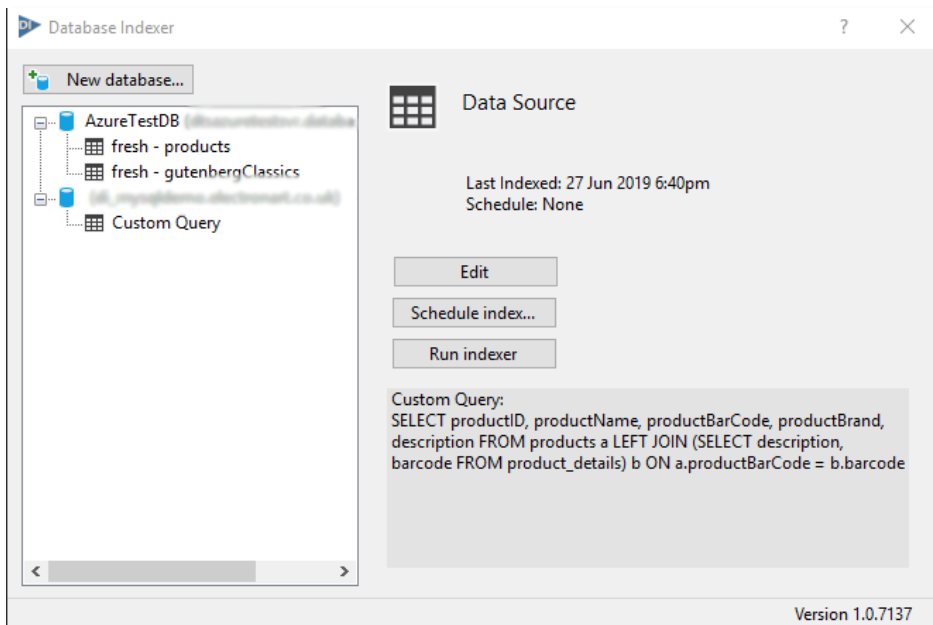
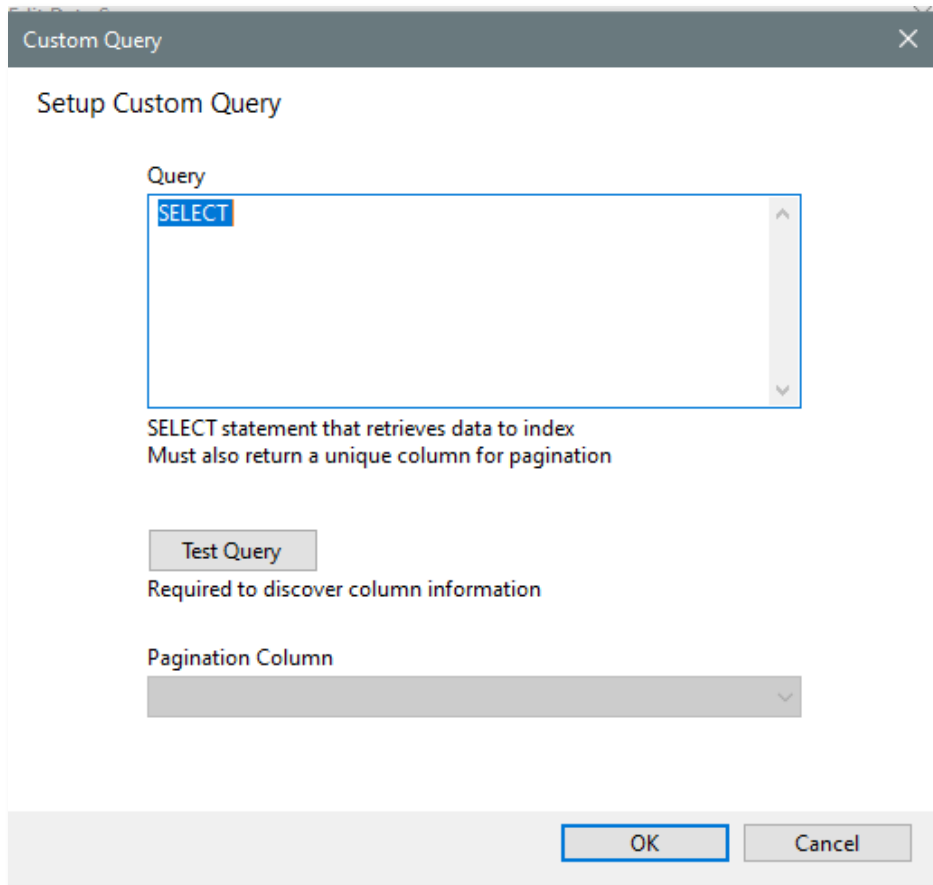
You can choose to remove records from the index when a certain condition is met. The options available are if the selected column contains a row with a true value: **is null**, **is not null**, **After now**, **Before now**, **= 0**, or **=1**.

Database Indexer never alters or deletes any data in the database itself and only requires read permission.

If you delete a row from the database it will not remove the data from the index. To clear data from the index if records have been deleted requires the index to be rebuilt, see [Indexing Options](#).

Custom Queries

Use the Custom Query option when data to be indexed spans multiple tables and/or to control what data is indexed when the table contains data that should be excluded from the index.

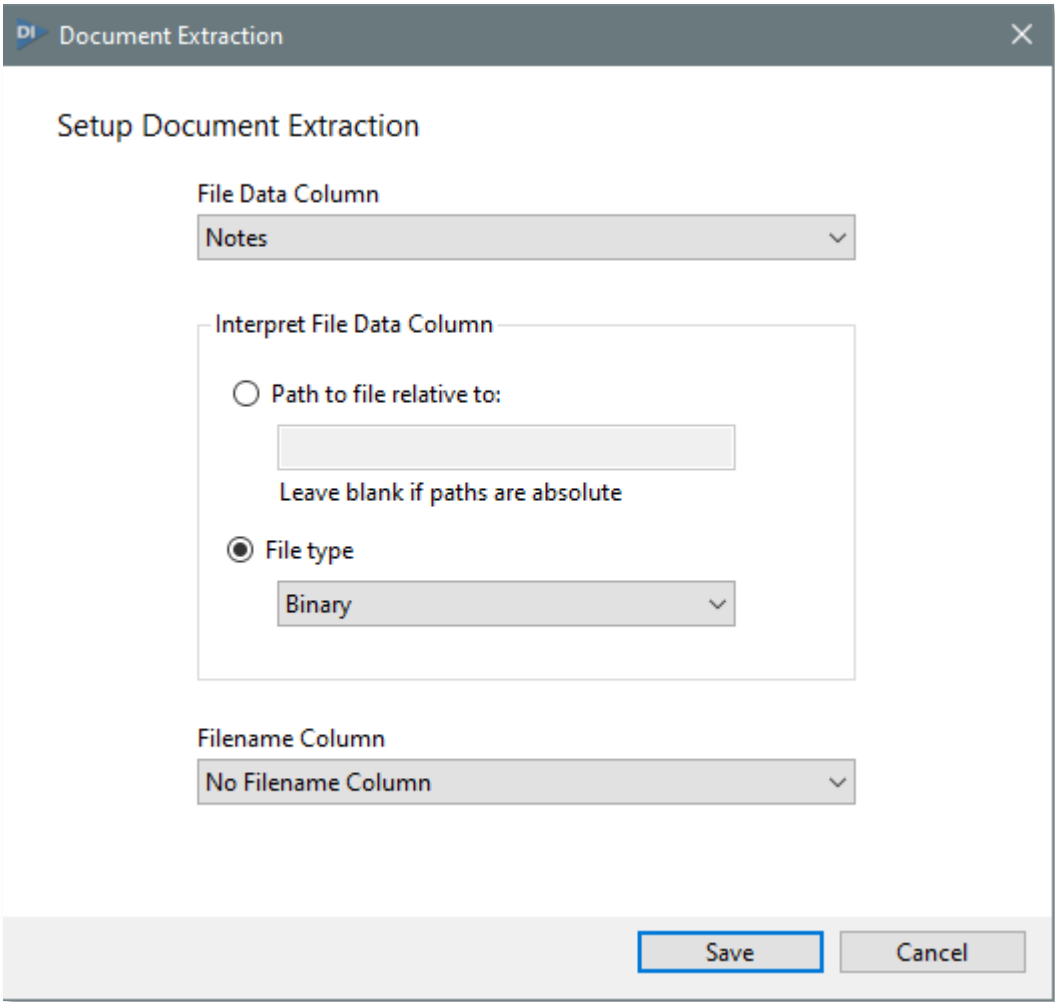


Example of setting up a [Custom Query](#)

Document Extraction

Database Indexer treats all database records as field data by default. Extracted documents are treated as additional file data.

When Document Extraction is selected, both types of data are cached in the index as files, not text. For example, if a PDF file is stored in the database as binary data, the PDF file is cached in the index and can be viewed as a PDF file.



If your database contains embedded files or a column that contains a link to a file, you can choose to extract the file and store it in the index. This option will make files open faster after a search.

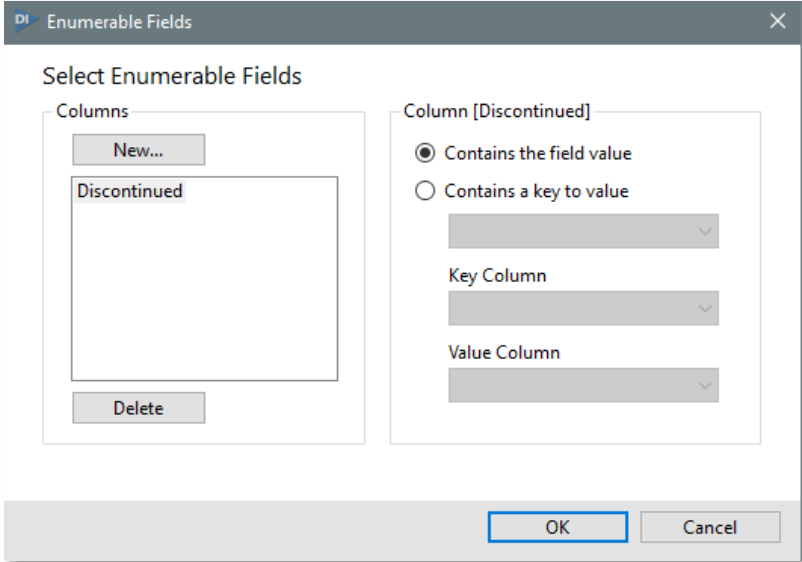
Enumerable Fields

Enumerable fields are used for categorizing search results into facets. [Universal Search Clients](#) can display faceted results or developers can use the [Universal Search API](#) to display faceted search results in their own application.

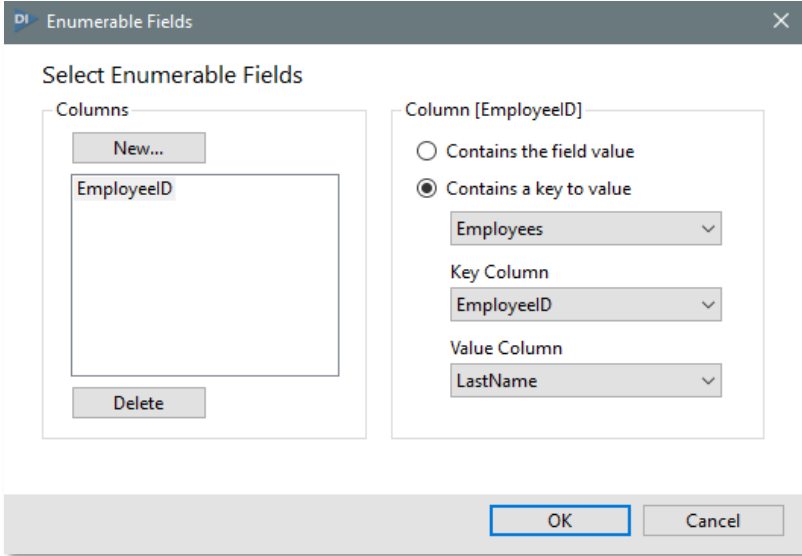
From the New Data source dialog click the **Setup...** button alongside the Enumerable fields option (or select an existing data source and click the **Edit...** button).

Click the **New...** button to add columns to the list to be used for faceted search navigation.

Select a column from the list and then select **Contains the field value** or **Contains a key to a Table** as appropriate.



Example 1:
“Discontinued” column contains a Boolean (true/false) value so **Contains the field value** is selected.



Example 2:
“Employees” table contains an EmployeeID column that is a Key to a LastName.

Finally to remove any unwanted items in the list select an item and click the Delete button.
See also: <https://support.dtsearch.com/faq/dts0222.htm>
https://en.wikipedia.org/wiki/Faceted_search

Get Help

If you need assistance with evaluating or setting up your own databases, please contact support@dtsearch.co.uk.

Settings

If document extraction is configured the document will be cached in the index.

The default indexing options can be changed by clicking the Help icon (?) in the title bar and clicking the **Default Indexing Options** button.

Feedback

Please give us feedback on this user guide so that we can provide content that's useful and helpful. Thank you!

Appendix

Cloud hosted demonstration databases. Please note these are normally available weekdays during working hours, subject to breaks for maintenance. Contact support@dtsearch.co.uk for connection details and passwords.

1. Azure SQL
2. mySQL hosted on AWS
3. Custom Query

Evaluation using the demo Azure SQL database:

Connect to the Azure SQL database as follows:

- 1) Click the **New Database...** button (press Enter)
- 2) Select Database Type: MSSQL
- 3) Enter Server name: *
- 4) Enter Database: Azure DB
- 5) Select Authentication: SQL Server (or Windows)
- 6) Enter Username: *
- 7) Enter Password: *
- 8) Click the Save button (Enter)

Add demo data source (a) as follows:

- 1) Select Azure DB in the list and click **Add New Data Source** button
- 2) Select Schema: fresh
- 3) Select Table: products (18,560 records of supermarket items)
- 4) Click All (13 fields text, Boolean, numeric) productID, productName, productCreationDate, productBarcode, productLastModified, productDeleted, suggestedExpiry, suggestedLocation, ImpDate, hasAmazonLink, hasWaitroseLink, hasImageInFS, productBrand.
- 5) Leave bottom item as "Don't Extract"
- 6) Click Next (Enter)

Choose the Index Location:

- 1) Select New Index
- 2) Name: fresh – products
- 3) Location: Browse to your dtSearch index folder e.g. c : \ . . . \Documents\dtsearch\
4) Click Save (Enter)
- 5) Click the Run Index button

18,560 records should be indexed in approx. 10 seconds.

Now you can [Search the fresh – products database](#) limiting your search to particular fields...

Search the fresh – products database:

Run dtSearch Desktop and select the `fresh – products` index (If it does not appear in the list of indexes, from the Index menu>Index Manager>Recognise index...).

In the Search dialog note that clicking on the **fields...** button should show 14 fields

(13 + Filename)

Type a search query: `xfirstword`

It should return 5000 results by default.

From the **Search Dialog>More Search Options** tab - edit the limit to 20,000 so that it will show all the results from the index.

Try some field searches, for example:

Click on the **fields...** button, choose `productBarCode`, press OK and enter 3263670041279

i.e. `(productBarCode contains (3263670041279))`

Evaluation using the demo Azure SQL database containing documents

1. Select the AzureDB database and click the **New data source...** button
2. Select fresh
3. Select gutenberClassics
4. Select All Columns (id, name, author, year, filecontent)
5. Click Setup...
6. Select File Data Column: Extract filecontent
7. Interpret as File content Binary
8. Filename Column: name
9. Click save
10. Select data to Index
11. Select Schema fresh
12. Select Table gutenberClassics
13. Select All Columns
14. Document Extraction Setup...
15. Next...

Set Index Destination

New index

Name fresh – gutenberClassics

Location – c:\...\Documents\Database Indexes

Save

Run Index 5 records approx. 1 second.

Search

In dtSearch Desktop search on **abandoned**, this should return 3 results from A Tale of two Cities, Jane Eyre and Dracula with the word **abandoned** highlighted.

The document properties are displayed at the end of the document.

Evaluation of Custom Queries using MySQL:

Contact us for the Hostname, Port, Username, and Password.

1. Ensure you are running Database Indexer 1.0.7109 or later
2. Create a new database connection with the following settings:
 - Database type: *MySQL*
 - Hostname: *
 - Port: *
 - Username: *
 - Password: *
3. Using the newly created database, create a new data source:
 - Set Schema to *fresh*
 - Select Custom Query and Set the following:
 - Add the following query:

```
SELECT productID, productName, productBarCode, productBrand, description FROM products a LEFT JOIN (SELECT description, barcode FROM product_details) b ON a.productBarCode = b.barcode
```
 - Click *Test Query*. This will enable the Pagination Column Combo box.
 - Select *productid* as the pagination column.
 - Press OK. Columns will automatically be detected and displayed as options based on the result set of the custom query.
 - Select all columns, leave document extraction set to **don't extract**.

The indexer can be run as normal.

The created index will contain combined data from two tables (the **products table**, and **product_details table**) Where *productid*, *name*, *barcode*, *brand* come from the products table and *description* from the product details table.

Evaluation of Custom Query using Azure MSSQL

There is no demo pair of tables that can be used for a **join** statement; however, it is possible to demonstrate that custom queries are working in MSSQL as well:

1. Create a new data source for the Azure database with schema set to *fresh*.
2. Set the custom query to `SELECT * FROM fresh.products`
3. Select all the columns, save and run the index.

Note: a subtle difference between MySQL and MSSQL is that MySQL does not need the schema in the select statement i.e. (SELECT * FROM products) but MSSQL needs (SELECT * FROM fresh.products)