

Formpipe.

Lasernet

Lasernet 9
OCR



Lasernet 9 – OCR [Revision 3 – May 2019]

© 2019 Formpipe Software

Lasernet is a trademark of Formpipe Software

Company website www.formpipe.com

Product website www.lasernetbyformpipe.com

1	INTRODUCTION	1
1.1	WHO SHOULD USE THIS GUIDE	2
2	TERMS OF USE	3
3	LASERNET OCR	4
3.1	INTRODUCTION	5
3.1.1	ABBYY RECOGNITION SERVER FOR LASERNET	5
3.2	GETTING STARTED - SERVER	7
3.2.1	CONFIGURING OCR ENGINE	7
3.2.2	CONNECTION	7
3.2.3	FIELDS	7
3.2.4	LINE TYPES	9
3.2.5	IDENTIFIERS	10
3.2.6	VALIDATION	10
3.2.7	FIELD NAMES AND REGIONAL PROFILES	14
3.2.8	EXAMPLE OF A WORKFLOW	17
3.2.9	INPUT MODULE	17
3.2.10	MODIFIERS – PDF TO TEXT AND JOBINFO MANIPULATION	18
3.2.11	ENGINE - OCR	19
3.2.12	OCR ENGINE – ADDITIONAL SETTINGS	20
3.2.13	MODULES - OUTPUT	21
3.3	LASERNET CLIENT	22
3.3.1	CONNECTION SETTINGS	22
3.4	LASERNET OCR	24
3.4.1	CONNECTION SETTINGS	24
3.4.2	THE USER INTERFACE	25
3.5	GETTING STARTED	28
3.5.1	CREATING A FORM	28
3.5.2	FORM PROPERTIES	28
3.5.3	FORM TOOLS	30
3.5.4	FIELD PROPERTIES	31
3.5.5	SELECTIONS	32
3.5.6	BODY PROPERTIES	36
3.5.7	PAGE PROPERTIES	37
3.5.8	VIEW PROPERTIES	38
3.6	WALK-THROUGH – HOW TO CREATE A FORM	39
3.6.1	CREATE A FORM	39
3.6.2	SETTING UP FORM PROPERTIES	39
3.6.3	SETTING UP CRITERIA	40
3.6.4	OCR FIELD	40
3.6.5	COVERED DATA FIELDS	44
3.6.6	CREATING THE BODY AREA	45
3.6.7	DATA FIELDS IN BODY	46
3.6.8	ADDITIONAL ITEM LINES	49
3.6.9	VIEW OUTPUT	51
3.6.10	UPDATE OCR FORM	52
3.6.11	RETRIEVE OCR FORM	52
3.6.12	FORMS LIST	53
3.7	DICTIONARY SERVICE	55
3.7.1	INSTALL LASERNET DICTIONARY SERVICE	55

3.7.2	CONNECT TO DICTIONARY SERVER.....	56
3.7.3	MANAGE DICTIONARY.....	58
3.7.4	CREATE OCR FORMS WITH DICTIONARY.....	63
3.8	CONVERT OCR ENGINE XML TO YOUR OWN WORKFLOW FORMAT	65

1 INTRODUCTION

1.1 Who Should Use This Guide

This guide is written for end-users who are responsible for the incoming document workflow. It is intended primarily as a reference to the various functionalities in Lasernet OCR.

2 TERMS OF USE

No part of this publication may be reproduced, transmitted, transcribed, or translated into any language in any form by any means without the prior written permission of Formpipe Software. The information in this manual is subject to change without notice. Any company names or data is fictive unless otherwise stated.

Formpipe Software shall not be liable for any loss or damage whatsoever arising from the use of this manual and the information contained therein (including errors or omissions).

Trademarks of other companies mentioned in this document appear for identification purposes only and are the property of their respective companies.

© 2019 Formpipe Software

3 LASERNET OCR

3.1 Introduction

Lasernet OCR is a client/server solution used for maintaining incoming business information and easily extracting data for import directly into your workflow system. By setting up simple text recognition you can work with data fields in forms such as invoices, order confirmations etc, from any of your business partners. The only requirement is that documents must be delivered as a PDF, TIFF or Text file to a channel available to the Lasernet Server.

Lasernet OCR is included in two of the Lasernet Input Management packages, which support the following incoming formats:

Package	Input Format	Output Format	Licensed modules in package
Input Management Basic	PDF, PDF/A with embedded text data	Text	PDF to Text Excel to XML
Input Management Standard (Lasernet OCR)	PDF, PDF/A TIFF with embedded text data or images*	Text, XML**	PDF to Text Excel to XML OCR Module
Input Management Full (Lasernet OCR)	PDF, PDF/A, TIFF with embedded text data or images*	Text, XML**	PDF to Text Excel to XML OCR Module XML Out

* For converting images to text, ABBYY Recognition Server is required for OCR scanning of TIFF files.

** An additional license for XML Input (Form Engine) is required.

Excel to XML conversion is also included in the Input Management packages. This process is not covered in this documentation, as XML is not a supported input for the Lasernet OCR application.

3.1.1 ABBYY Recognition Server for Lasernet

ABBYY Recognition Server provides functionality for optical character recognition and exchanging documents with Lasernet. The Lasernet edition supports the following basic features:

Functionality	Lasernet Edition
Document Input	TIFF, PDF, PDF/A
Export Formats	Original Image, TIFF, JPEG, PDF, PDF/A
Recognition	Normal, Typewriter, Matrix, OCR-A, OCR-B, MICR-E1 3B
Languages	More than 190 Recognition Languages (No Asian support)

Barcode processing	1D + 2D
CPU Cores	1 single
Scanning Station	No *
Indexing Station	No *
Verification Station	No *

* Supported via Lasernet Modules

ABBYY Recognition Server has an intuitive interface and can be set up and operated without prior training. Processing stages that require minimal user involvement, such as OCR and document conversion, are automated together with Lascript.

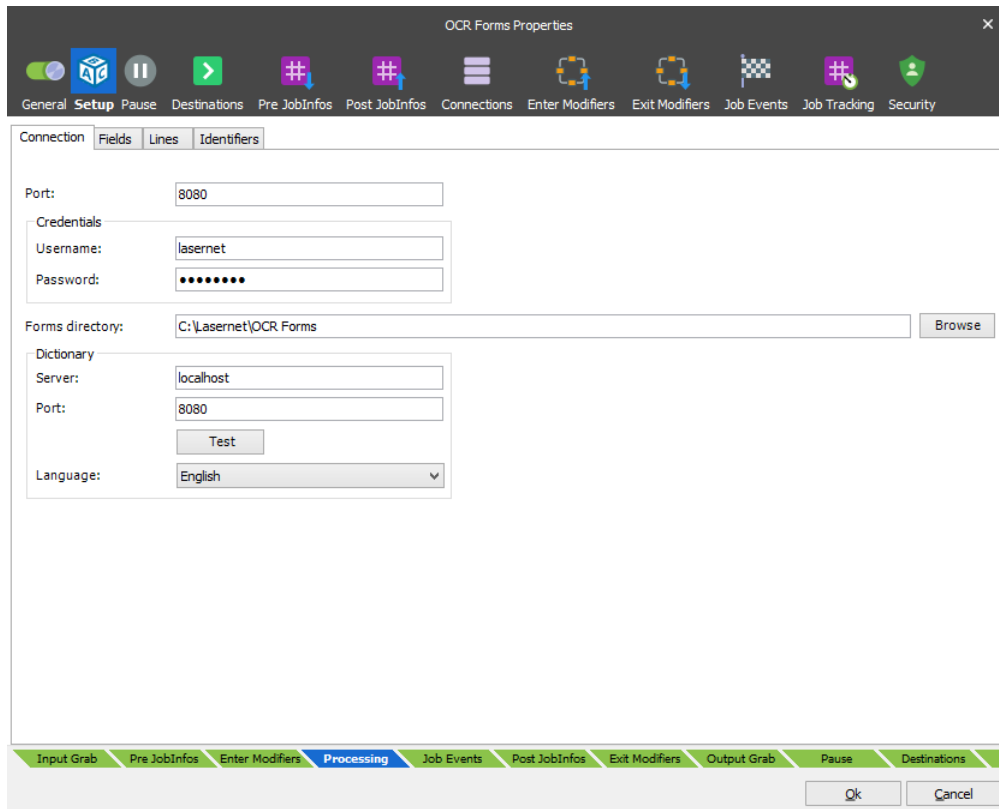
Documentation for the ABBYY Recognition Server is installed with "ABBYY Recognition Server for Lascript". Please note that the documentation may include information, not available in the Lascript edition of the ABBYY Recognition Server.

3.2 Getting Started - Server

3.2.1 Configuring OCR Engine

In the Lasernet Developer an OCR Engine must be added as a module. The connection between the client and server is established via a listening port over http; user login information must be defined. The OCR Engine can contain information on the list of documents you want to setup for OCR scanning and what fields are required for each document. All the information defined in the OCR Engine will be available in the OCR Designer, to help the end user to set up their document.

3.2.2 Connection



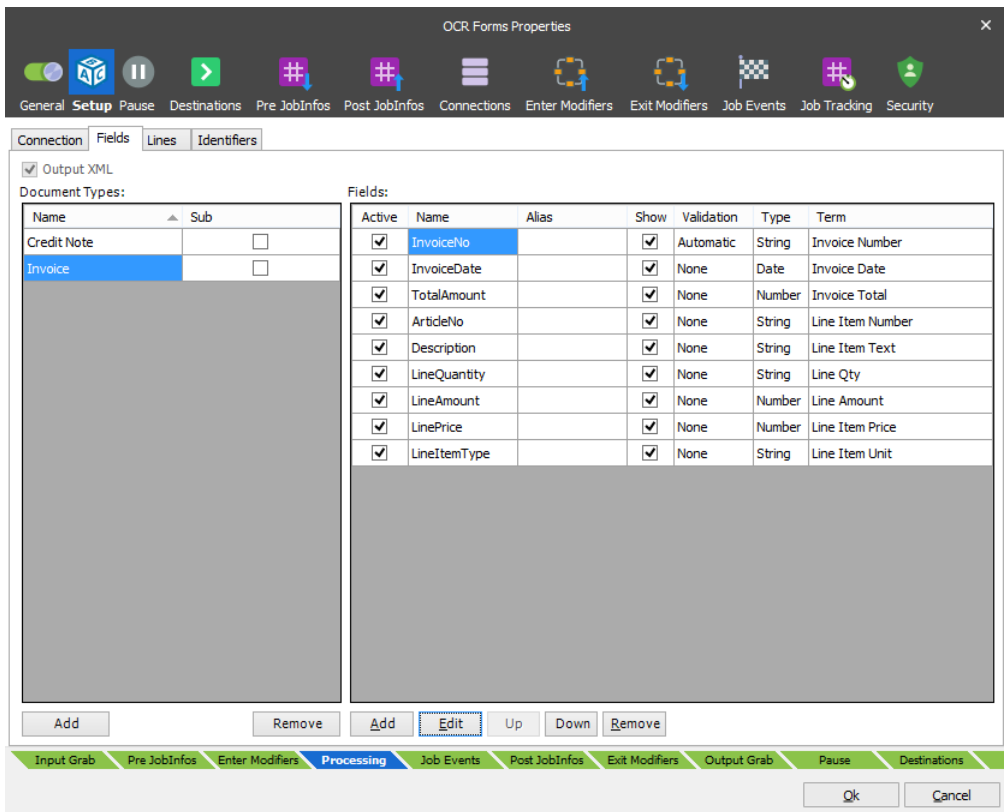
The screenshot shows the 'OCR Forms Properties' dialog box with the 'Connection' tab selected. The 'Port' field is set to 8080. Under 'Credentials', the 'Username' is 'lasernet' and the 'Password' is masked with dots. The 'Forms directory' is 'C:\Lasernet\OCR Forms'. Under 'Dictionary', the 'Server' is 'localhost', the 'Port' is 8080, and the 'Language' is 'English'. The 'Processing' step is highlighted in the bottom progress bar.

The following settings are required for setting up the connection between client and server and for storing forms.

Port	Port number for OCR Engine to listen on.
Username	User name required to login.
Password	Password required to login.
Forms directory	Directory for the OCR Engine to store the list of OCR forms. Caution: Do not use the same directory as the Server runtime data directory, defined in the server settings for your Lasernet Server, or you will corrupt your OCR forms.
Dictionary server	Name or IP address for server running the Lasernet Dictionary Service.
Dictionary port	Port number for Lasernet Dictionary Service to listen on.
Language	Default language for dictionary in OCR Editor. By default the solution is delivered with an English and Danish dictionary. Other languages must be maintained by the end-user.

3.2.3 Fields

For each Document Type, a list of Field Names and Line Types must be added. Field Names must have a data type, which is used for formatting numbers and date fields.



Output XML

Fixed output format from OCR Engine is XML.

Document Types

A Document Type has a reference to a list of Field Names and Line Types. The Document Type is linked to each OCR form when the incoming document is mapped in the Lasernet Form Editor. The listed OCR Fields must be mapped to cover the required data fields in the external workflow system.

Field Names

Any Field Name added to a document will be available in the Lasetnet OCR Editor.

The **Name** column must contain the field name to be included in TEXT/XML data created as output from OCR Engine. If an illegal tag name is entered you will be prompted with an !

The **Alias** column must contain the logical field name seen by the end user in the OCR Designer. Logical names will help the end-user to locate the required data fields in the incoming OCR documents.

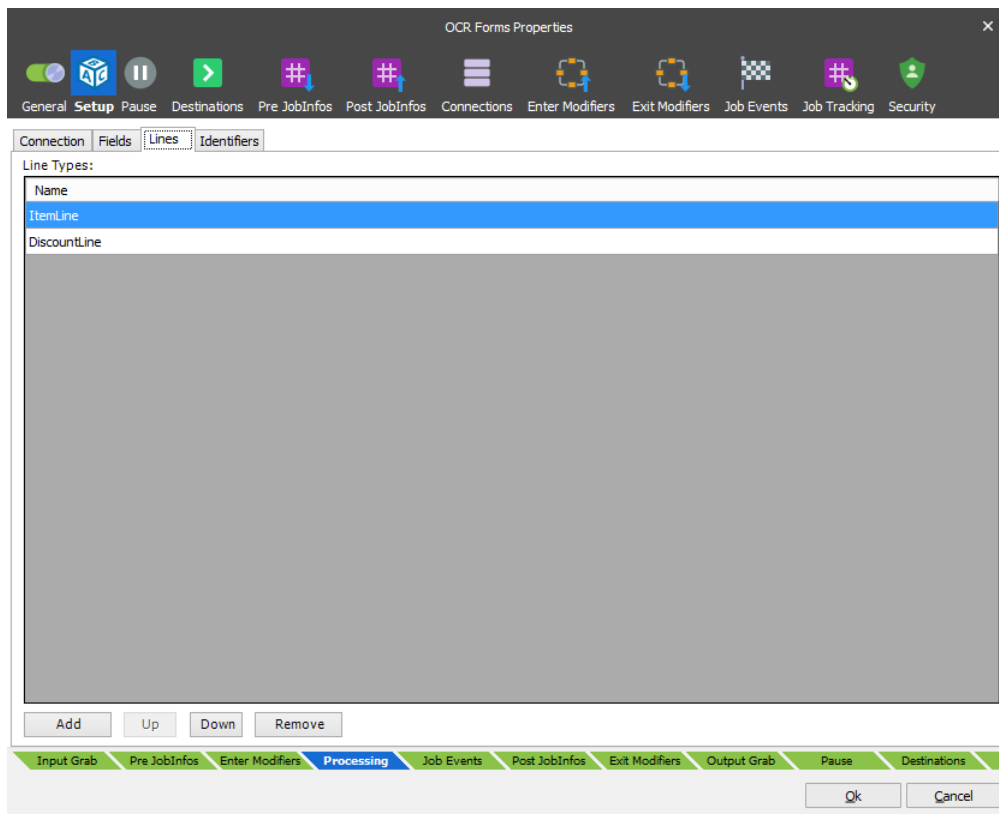
Activate **Show** to enable viewing of the field name, or alias name if entered in the Lasetnet Client. Please notice that if this field is not activated, validation cannot be performed by the end user in the Lasetnet Client.

Validation column will contain value "None" if no validation for this field is required or "Automatic" if a validation rule has been entered. Select a Field Name and click the **Edit** button to setup validation rules.

Each field can be defined as a **Type**, such as string, number or date. Use number and date formats for converting various incoming number and date formats, via Regional Profiles, into a standard format. Numbers will be converted to 99999.9999 (four decimals) and dates formatted to yyyyymmdd.

The **Term** column is used to connect a field name with the name of a term in the dictionary. In the dictionary a term consists of one or several aliases (words that appear in job data). When the dictionary understands a phrase it will automatically be connected to the defined field name.

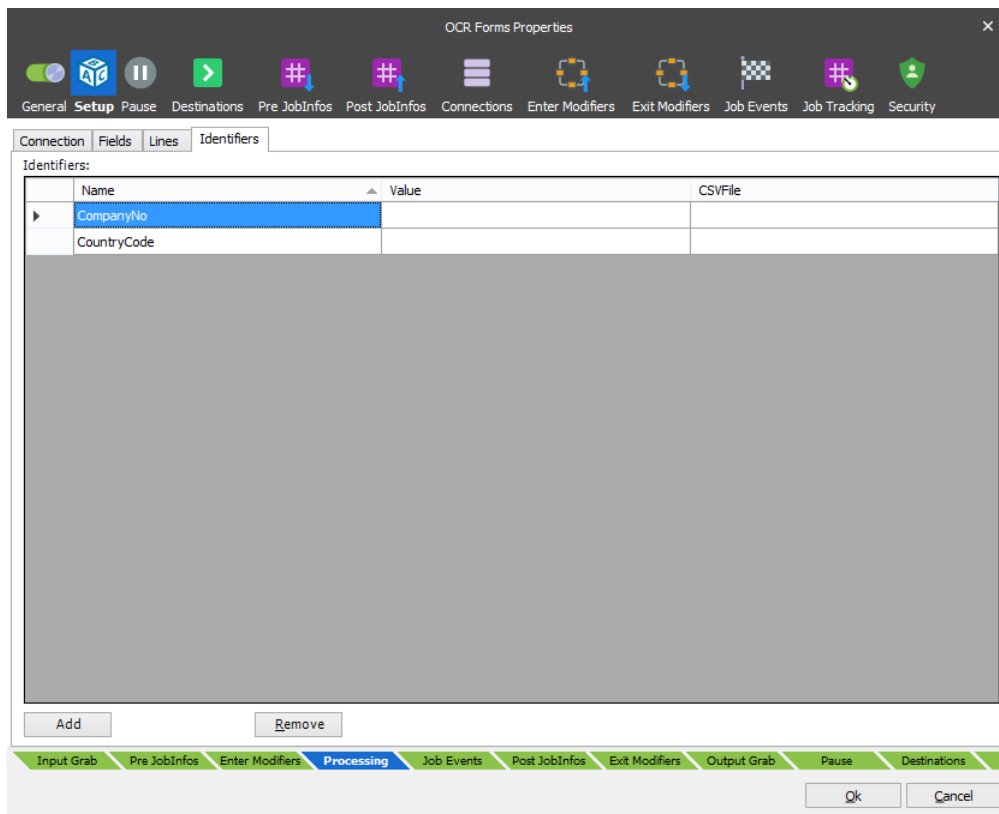
3.2.4 Line Types



Line Types

Any Line Type added to a document will be available in the Lasernet OCR Designer. It is recommended that Line Types are given logical names to help the end user to capture the required body lines in the incoming OCR documents. The Line Types will be embedded in the output job created by the OCR Engine. The list can be left empty if no lines need to be captured. When recognizing Line Types in the body area of the incoming documents, the order of Line Types can be important. The Line Types at the top of the list have highest priority when the OCR Engine is recognizing lines. Move Line Types up and down to set the priority.

3.2.5 Identifiers



Identifiers

Identifiers are used for inserting fixed field names and values in the output data. Only characters from A-Z are recommended as valid Identifier names and spaces are not allowed. Set a default value for the field if required or leave empty.

CSVFile

Path and filename to CSV file in the format:

"Header1","Header2","Header3"

"Value1","Value2","Value3"

"Value1","Value2","Value3"

The list of records, added to the CSV file, will be included in the dropdown list. The value of the first column can be selected and used as the value for the field. Other columns are for viewing purposes only.

Header names: Required

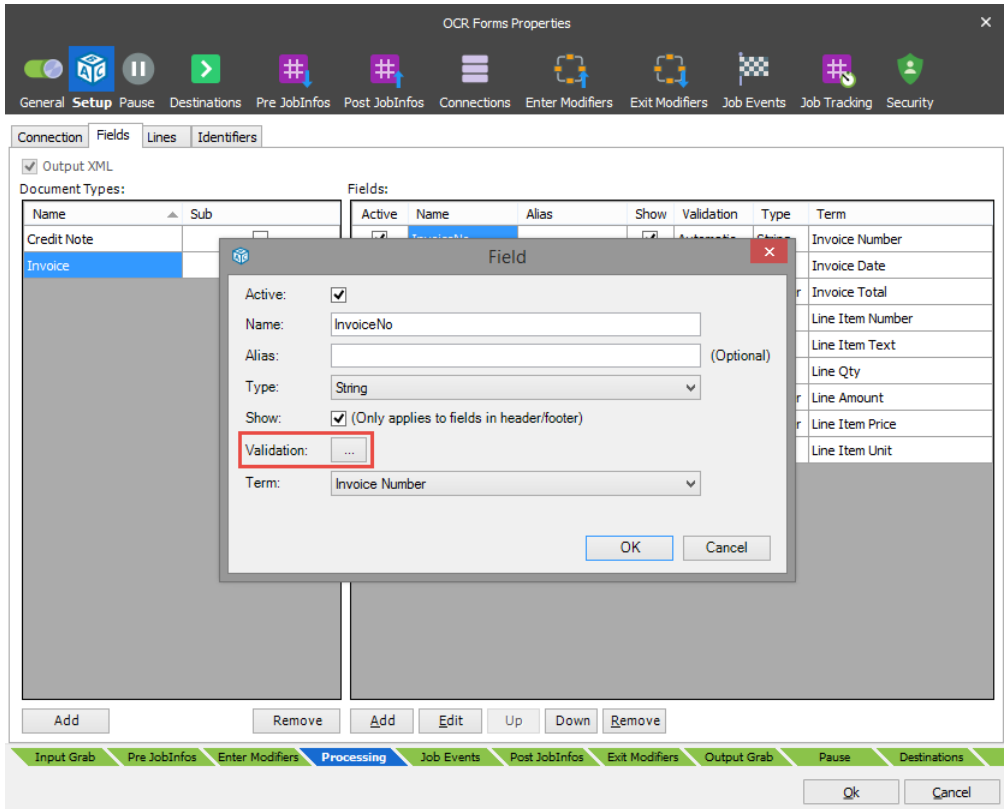
Field separator: Comma only

Quotes: Not required but recommended

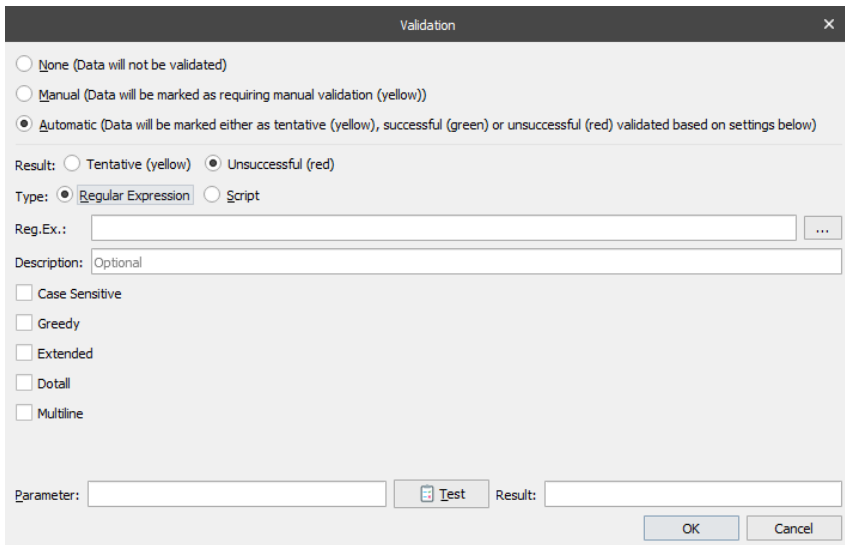
Number of columns: A maximum of 6 columns are recommended for viewing purposes.

3.2.6 Validation

Validation is used to validate the contents of an OCR Field. Metadata not matching specified conditions will be marked either as tentative (yellow) or unsuccessful (red) in the Lasernet Client, which is used for approval of documents in the workflow.

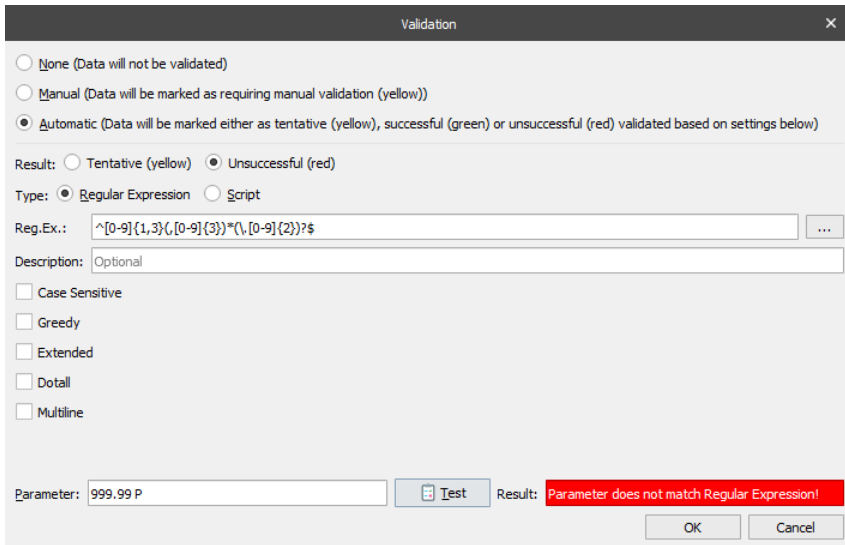


Select a **Field Name**, click **Edit** and select the **Validation** button to edit or create new validation rules.



Select Regular Expression or Script to validate the contents of an OCR Field.

3.2.6.1 Validate with a regular expression



In the example we have created a regular expression matching a US formatted number with comma as thousand delimiter and dot as comma delimiter.

Regular Expression: `^([0-9]{1,3}([0-9]{3})*(\.[0-9]{2})?)?$`

The \$ sign at the end of the regular expression is essential to match the full string and not a part of the string.

The **Parameter field** and **Test button** are useful to validate the expected result of the regular expression. This functionality is for online testing only in the Lascript Developer.



For more information regarding regular expression syntax, we recommend using the internet for reference.

3.2.6.1.1 Validating dates and numbers

All fields defined with date or number as the field type must be validated against the internal date or number format handled by the OCR Engine and not as the format present in the document.

The internal format for dates is: yyyyMMdd

The internal format for numbers is: 99999.9999 (dot as the decimal splitter and four decimal places)

Name: InvoiceDate Text: ...

Format: Date (ddMMyyyy) Split:

Required: Split String: ...

Date: ----- Size: (25, 1) 19-01-2016

Sales order SO-100195

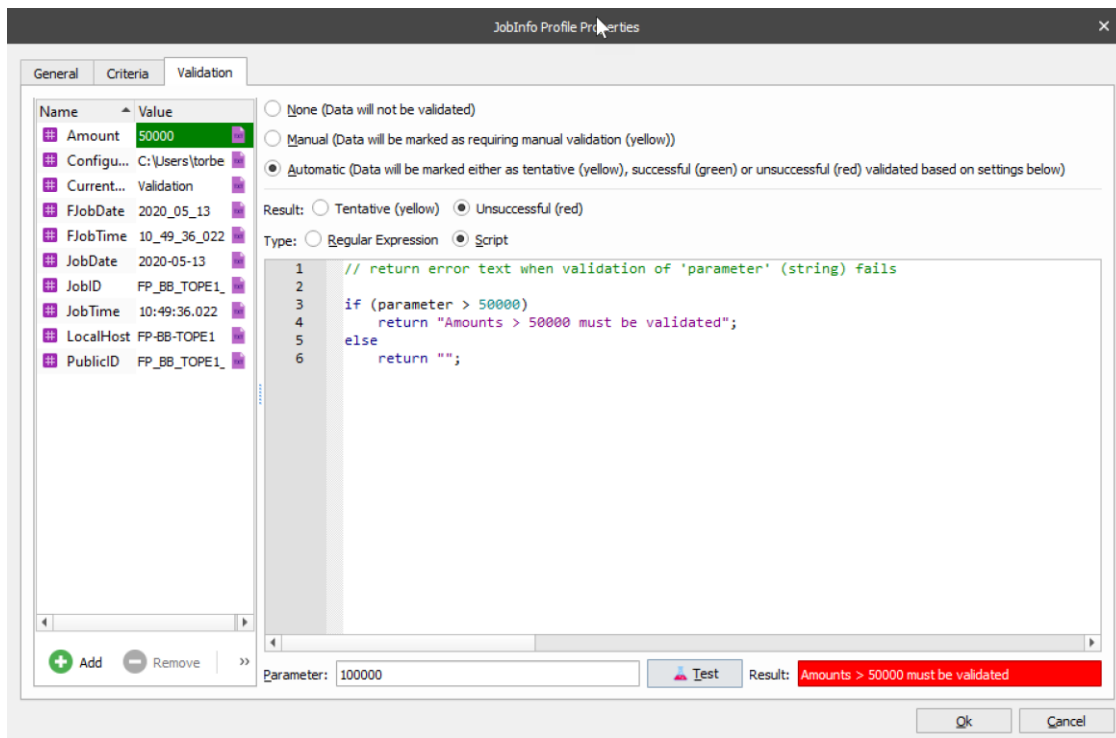
Requisition

Active	Name	Alias	Show	Validation	Type	Term
<input checked="" type="checkbox"/>	SupplierName		<input checked="" type="checkbox"/>	Automatic	String	
<input checked="" type="checkbox"/>	InvoiceNo		<input checked="" type="checkbox"/>	None	String	Invoice Nu...
<input checked="" type="checkbox"/>	TotalAmount		<input checked="" type="checkbox"/>	None	Number	Invoice Total
<input checked="" type="checkbox"/>	ArticleNo		<input type="checkbox"/>	None	String	Line Item N...
<input checked="" type="checkbox"/>	Description		<input type="checkbox"/>	None	String	Line Item T...
<input checked="" type="checkbox"/>	InvoiceDate		<input checked="" type="checkbox"/>	Automatic	Date	Invoice Date

In this example, the date 19-01-2016 is extracted and converted into the internal format 20160119.

For the date field you must define a regular expression that validates to the internal format of yyyyMMdd. For numbers you must validate to the format 99999.9999

3.2.6.2 Validate via JavaScript



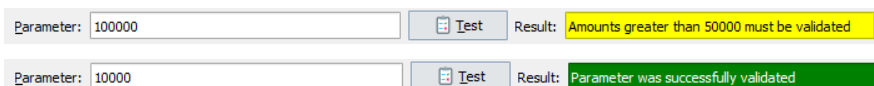
In this example we have created a JavaScript matching if a parameter is higher than a specific amount. The **parameter string** is essential and contains the contents of the JobInfo to be validated. The expression will return an empty string when validation of the parameter is successful and error text when it fails validation.

```
if (parameter > 50000)
    return "Amounts > 50.000 must be validated";
else
    return "";
```

If you want to validate against a **JobInfo** included in the script you can press **Add** to add the JobInfo **Name** and **Value** in the JobInfo grab window.

```
if (parseFloat(parameter) > parseFloat(job.getJobInfo("Amount")))
    return "Amounts > 50.000 must be validated";
else
    return "";
```

The **Parameter** field and **Test** button are useful to validate the expected result of the JavaScript. This functionality is for online testing only in the Lasernet Developer.



3.2.6.3 Supported JavaScript functions

The JavaScript tool in the OCR validation solution has support for a limited range of system functions compared to the Job Engine. The supported convert functions, job classes and arrays are:

Static functions:

- Convert.toDate
- Convert.toDateRP

- Convert.toNumber
- Convert.toNumberRP

Job class:

- All functions in the job class (see the Scripting manual section 4.2).

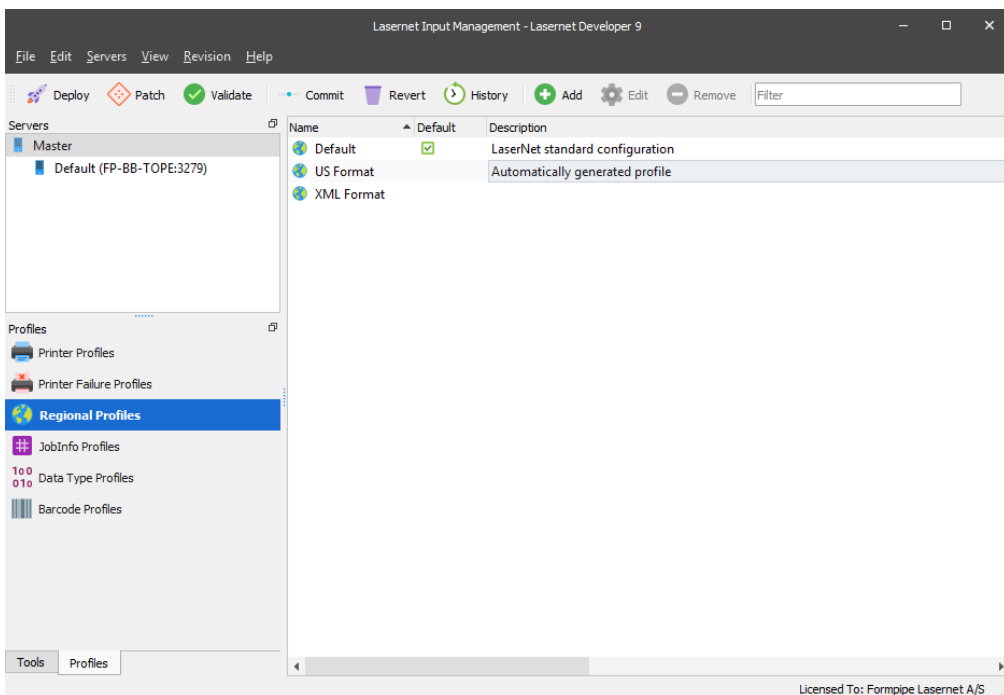
Arrays:

- Azure commands
- Database commands
- Modifier commands
- SharePoint commands

Note: The system functions are not supported for validation in Lasernet Meta, where validation runs on client side and not server side.

3.2.7 Field Names and Regional Profiles

When setting up unique formats for amounts and dates in the output format, regional profiles must be created, based on the incoming data fields. Regional Profiles which are defined in the Lasetnet Developer will be available in any OCR Forms designed in the Lasetnet OCR client.



Formats for Dates and Numbers are accessed via Regional Profiles by the Lasetnet OCR client.

US Format Properties

General Numbers Currency Time Date

Name: US Format

Description: Automatically generated profile

Created date: 25-09-2018 16:32:48

Modified date: 25-09-2018 16:32:48

Created by: Unknown

Modified by: Unknown

Form Editor:

OCR Designer:

Import pre-defined regional profile

Numbers and Date formats can be defined for any format that exists in the incoming documents and will be available for the Lasernet OCR Designer.

US Format Properties

General Numbers Currency Time Date

Show in OCR Designer:

Example

Positive: 123,456,789.00 Negative: -123,456,789.00

Decimal symbol: .

Digit grouping symbol: ,

Digit grouping: 123,456,789

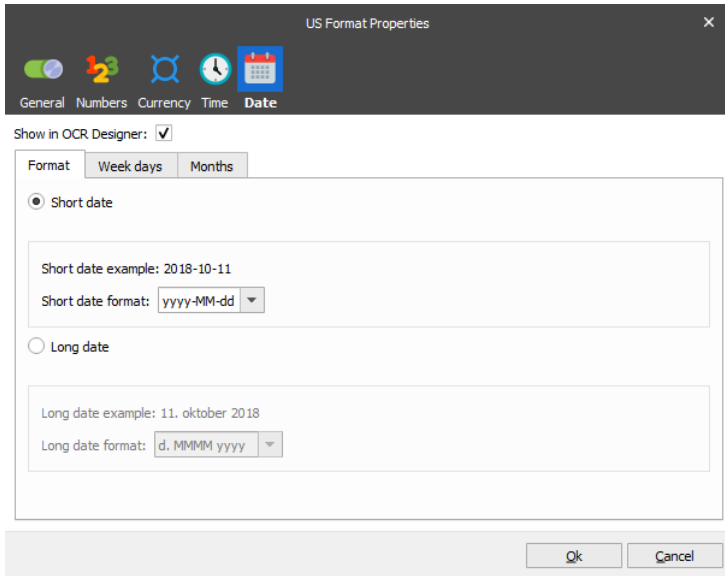
Negative sign symbol: -

Negative number format: -1.1

Display leading zeros: 0.7

Field types defined as Numbers in the OCR Engine will make use of Number formats defined via the Number settings in the Regional Profiles.

Activate “Show in OCR Designer” to show the Regional Profile containing this Number format.



Date formats make use of the Date settings. Field types defined as strings do not make use of any regional settings. Activate "Show in OCR Designer" to show the Regional Profile containing this Date format.

3.2.8 Example of a workflow

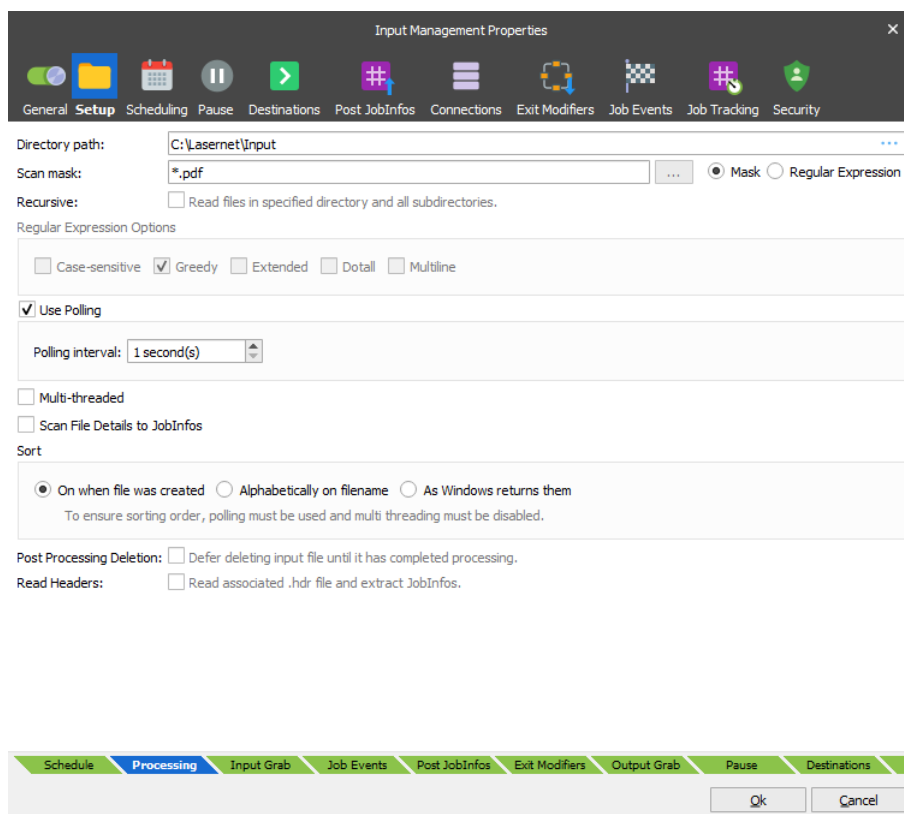
The input format for the Lasernet workflow is a PDF document. If your documents are scanned TIFF files you must install and configure the ABBYY Recognition Server to convert a scanned TIFF file into a text readable PDF document before it is delivered to a File Input module accessible for Lasetnet.



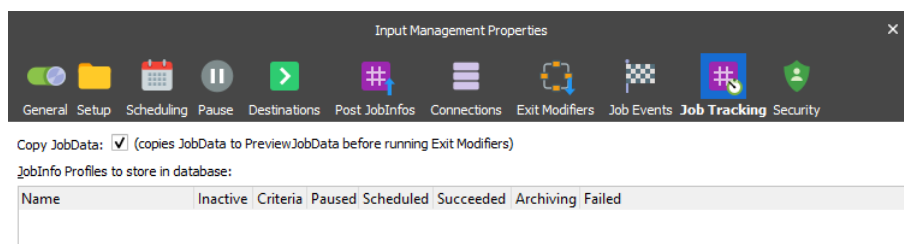
A PDF to Text modifier is used to convert PDF into a text document, which is used by the Lasetnet OCR Engine to extract the contents into an XML file. The output can be stored on any output location accessible to your external workflow system.

3.2.9 Input Module

Add a **File Input** Port and set the scan mask to read for PDF files.



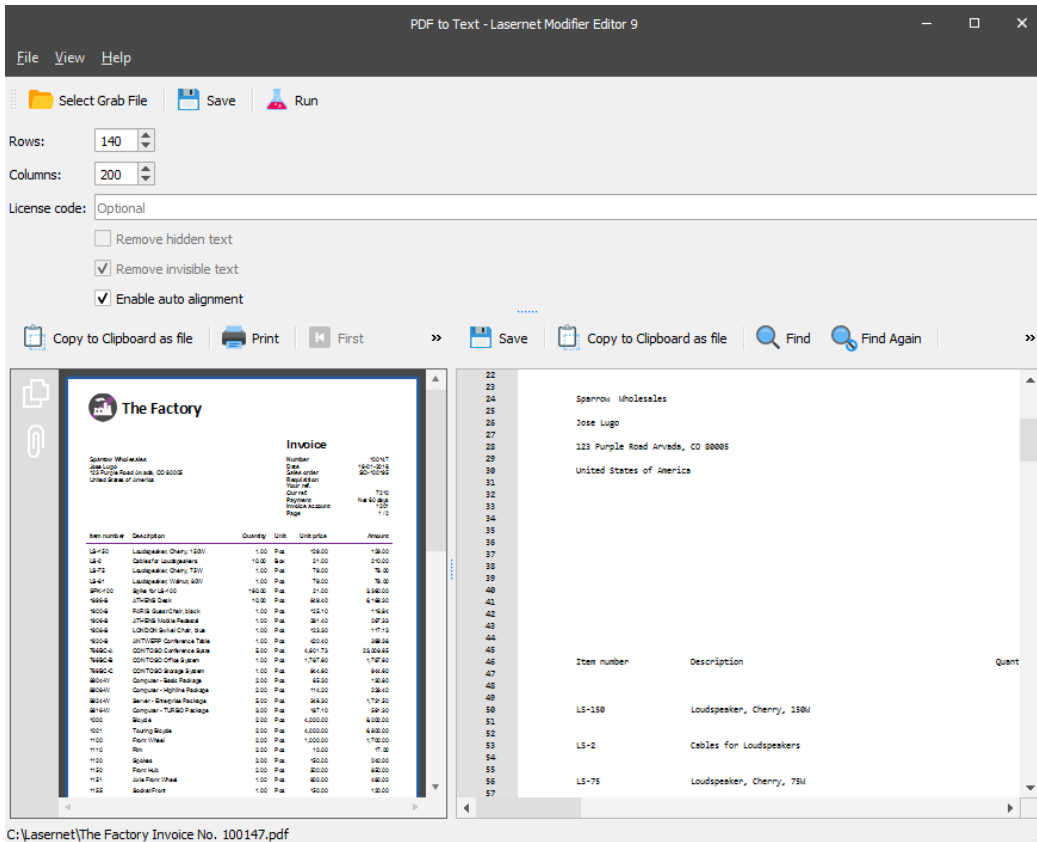
Enable Job Tracking → **Copy JobData** on the input module to create a copy of the JobData before it is converted into text.



The incoming PDF document is copied into the JobInfo named **PreviewJobData**. The Lasetnet Client and Lasetnet OCR Editor are now able to detect that the preview window shows the incoming document instead of JobData, which is a text file only.

3.2.10 Modifiers – PDF to Text and JobInfo Manipulation

Add a **PDF to Text** modifier to the File Input module to convert PDF to a text format.

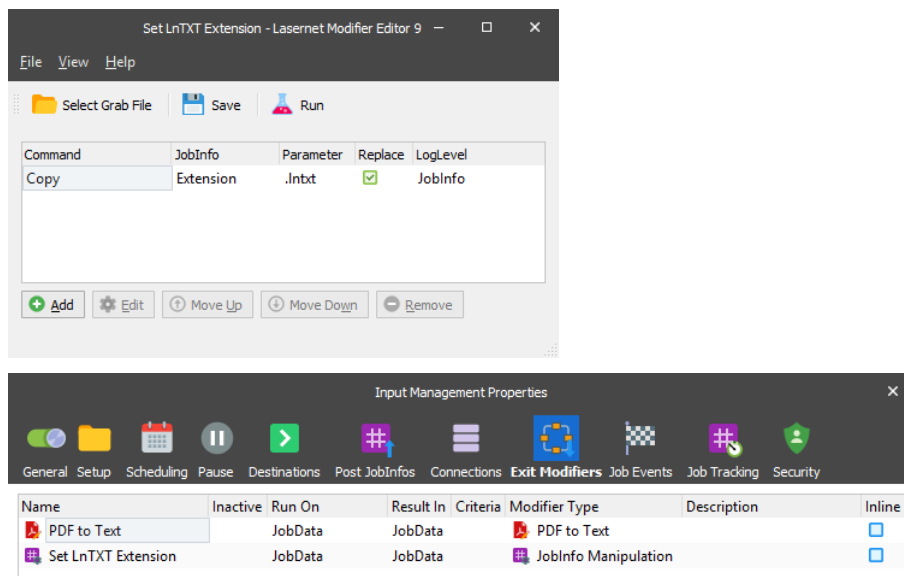


Values for **Rows** and **Columns** are used for the algorithm to extract the text strings, embedded in the PDF document, into a text grid. Rows represents the number of lines per page and columns represent the width of the text file.

In previous versions of Lascript, an additional license code was required to run the PDF to Text modifier. To ensure backwards compatibility you must enter you old license code here, otherwise leave the field empty.

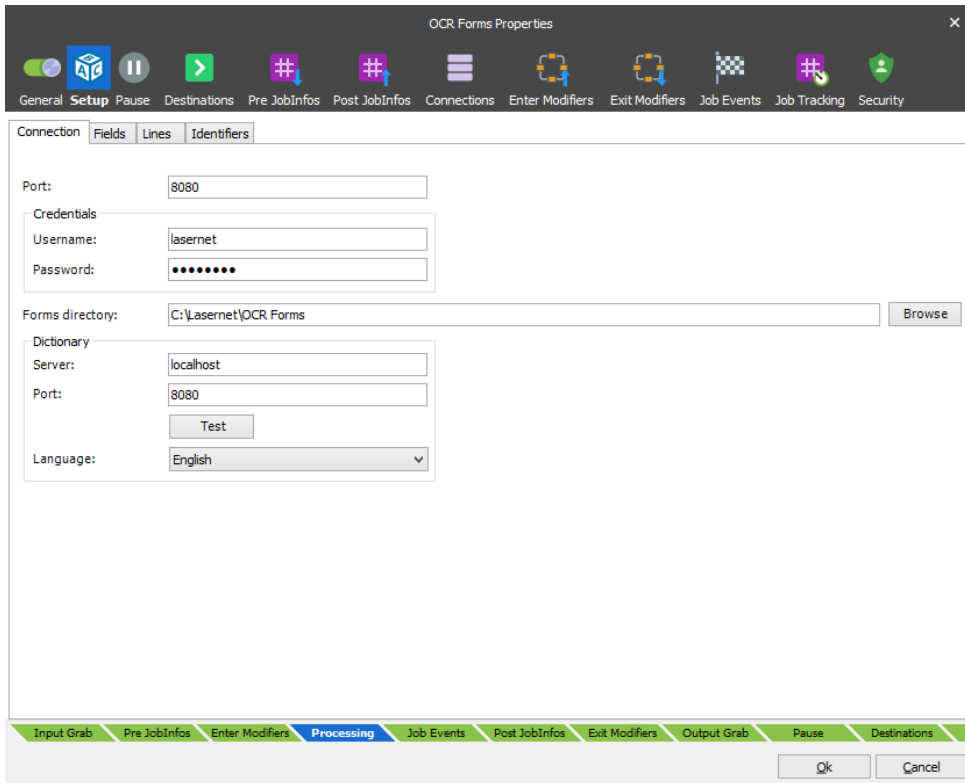
'Hidden' and 'invisible' text options allow you to remove text from PDF files which are usually not needed in the final text file, after conversion.

Create a JobInfo Manipulation and add the object as a second Exit Modifier on your File Input port. The modifier will associate the extension of text format with the Lascript OCR Editor.

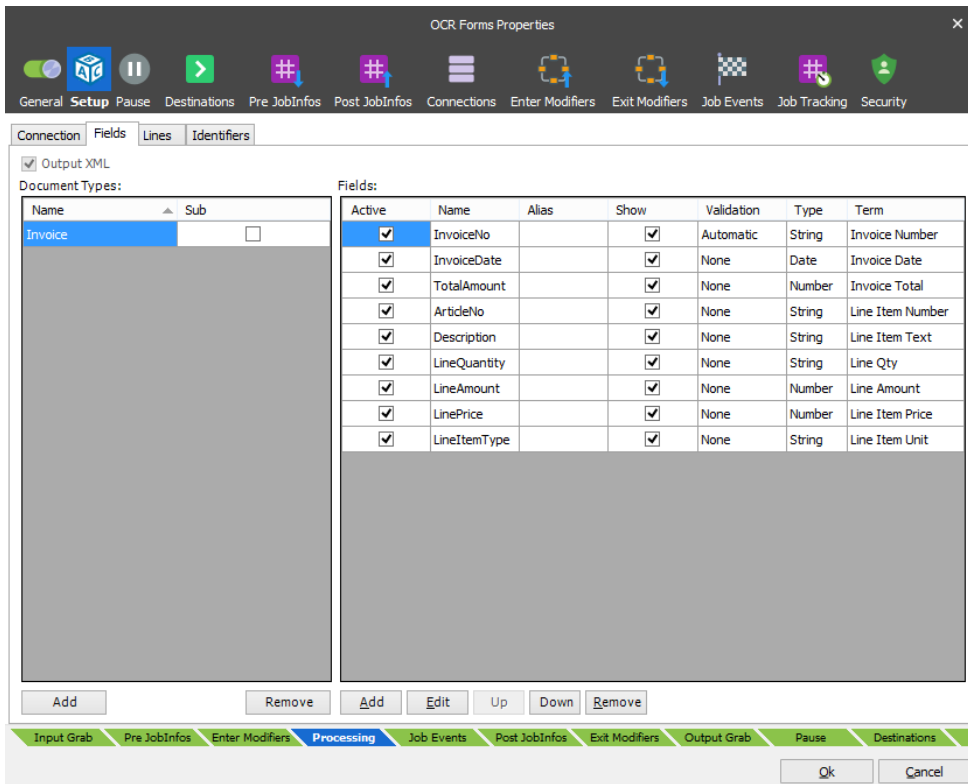


3.2.11 Engine - OCR

Add an **OCR Engine** with the settings for port/listen number, username and password, and forms directory (location where to store uploaded OCR Forms on server site).



Setup the list of document types you want to maintain in your OCR solution and a list of Field Names and Line Types.



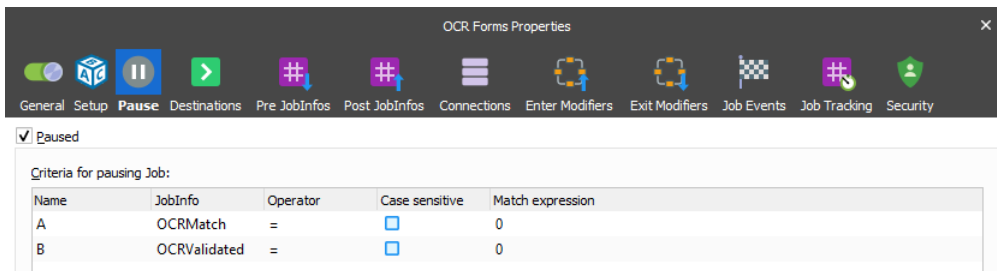
3.2.12 OCR Engine – additional settings

You are ready to set up a pause function for incoming documents, which are not recognized by the OCR Engine.

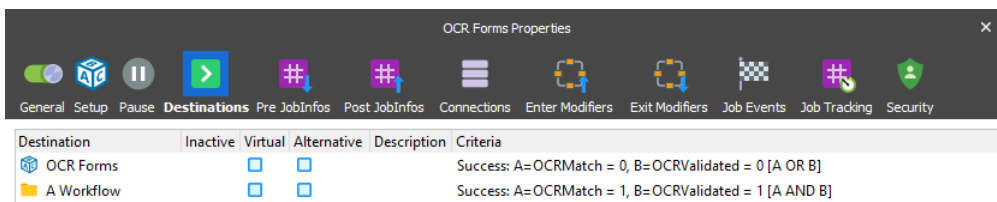
For processed documents the JobInfo: **OCRMatch** will be set to either **0 = not recognized** or **1 = recognized**.

By setting up a Job Destination and a validation criteria in the Pause tab you can either: pause the jobs that haven't been matched or validated successfully by the OCR Engine, or forward them to another module in the workflow.

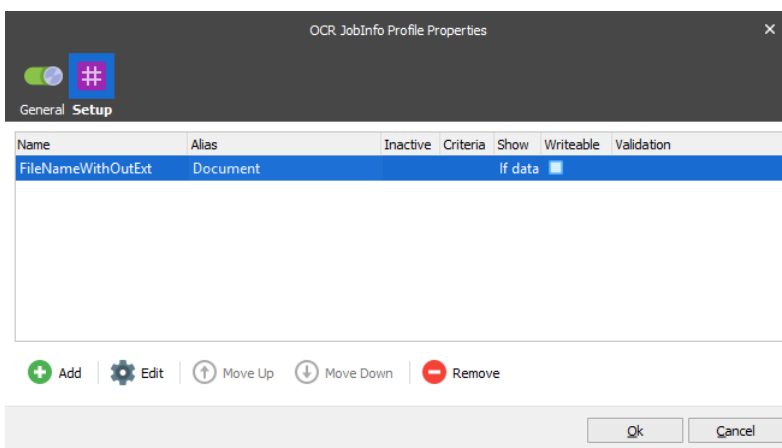
Note: OCR validation is an optional feature and if it is not configured, the JobInfo should not be added as a criterion for pausing jobs.



When the job has been processed by the OCR Engine, the JobInfos OCRMatch (Boolean) or OCRValidated (Boolean) (optional) can be used for checking if the job should be forwarded to the Form Engine or sent back to the OCR Engine.

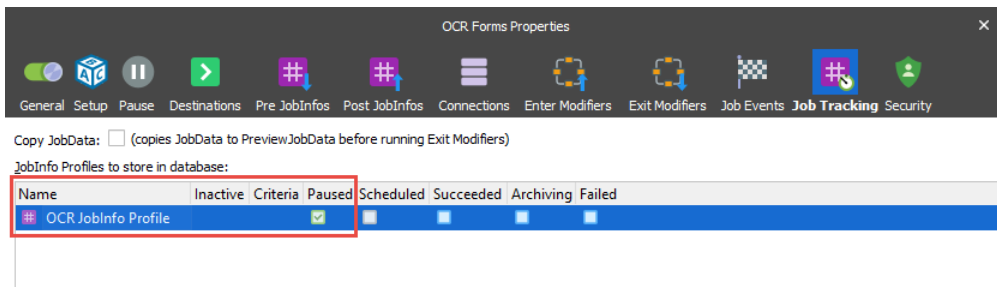


When saving jobs in pause mode you are able to save other available JobInfos as metadata for the job. This is maintained via a JobInfo Profile



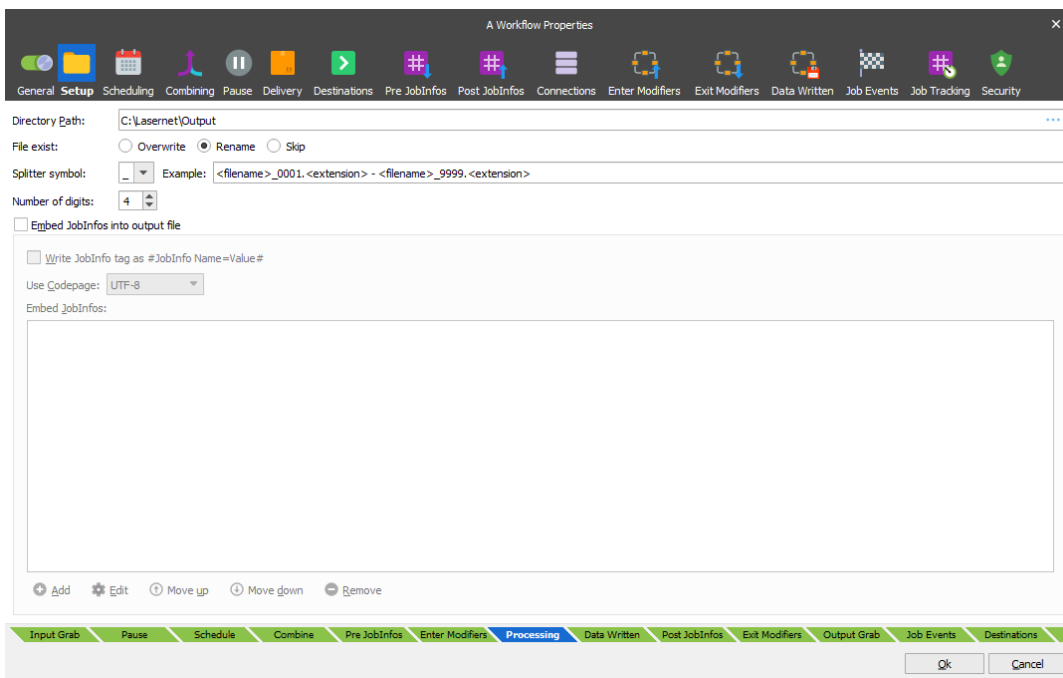
In this example, a JobInfo Profile named "OCR JobInfo Profile" is added for storing additional metadata like the document name without the file extension (FilenameWithoutExt) to a database.

When the JobInfo Profile is created you can assign the profile to the OCR Engine → Job Tracking tab and activate it for paused jobs only.



3.2.13 Modules - Output

Add a File Output module to save the data extracted by the OCR Engine on a location accessible from your external workflow system.

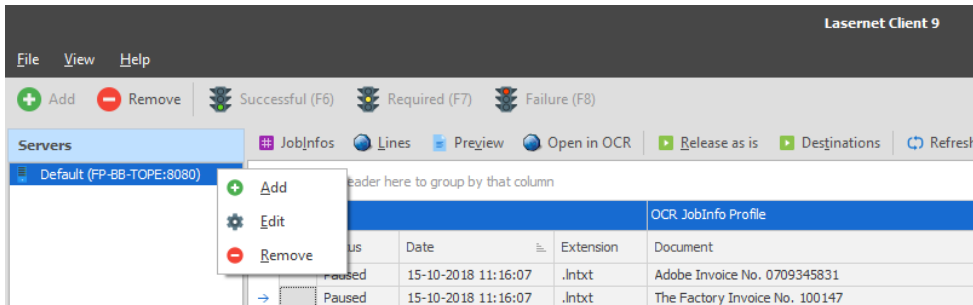


Set whether you want to overwrite, rename or skip the output if a file already exists with the same file name.

3.3 Lasernet Client

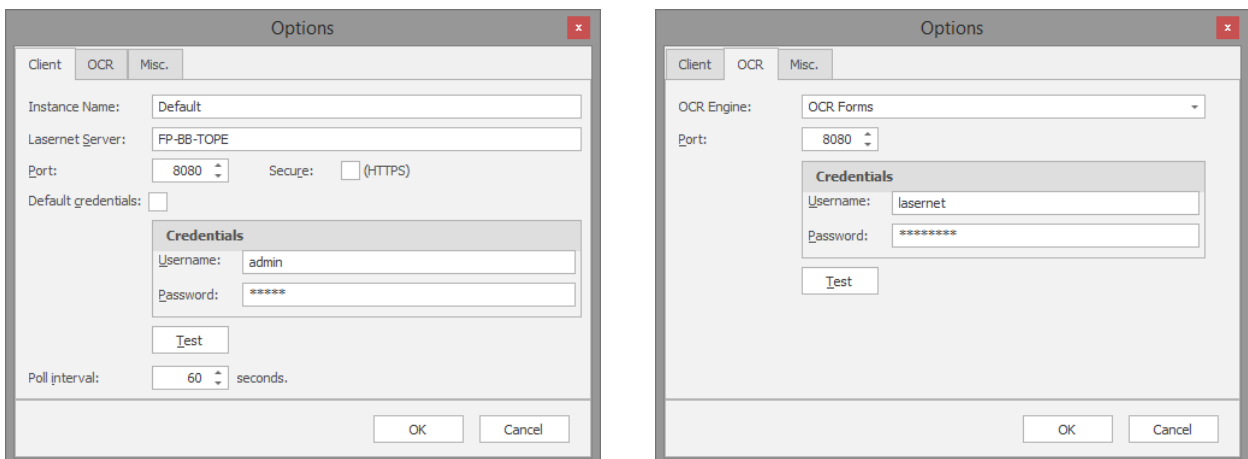
To open a paused job saved by the OCR Engine, the Lasetnet Client needs to be installed on the same computer as the Lasetnet OCR Editor.

Before you can retrieve the paused job, you must set up a valid connection to the Lasetnet server.

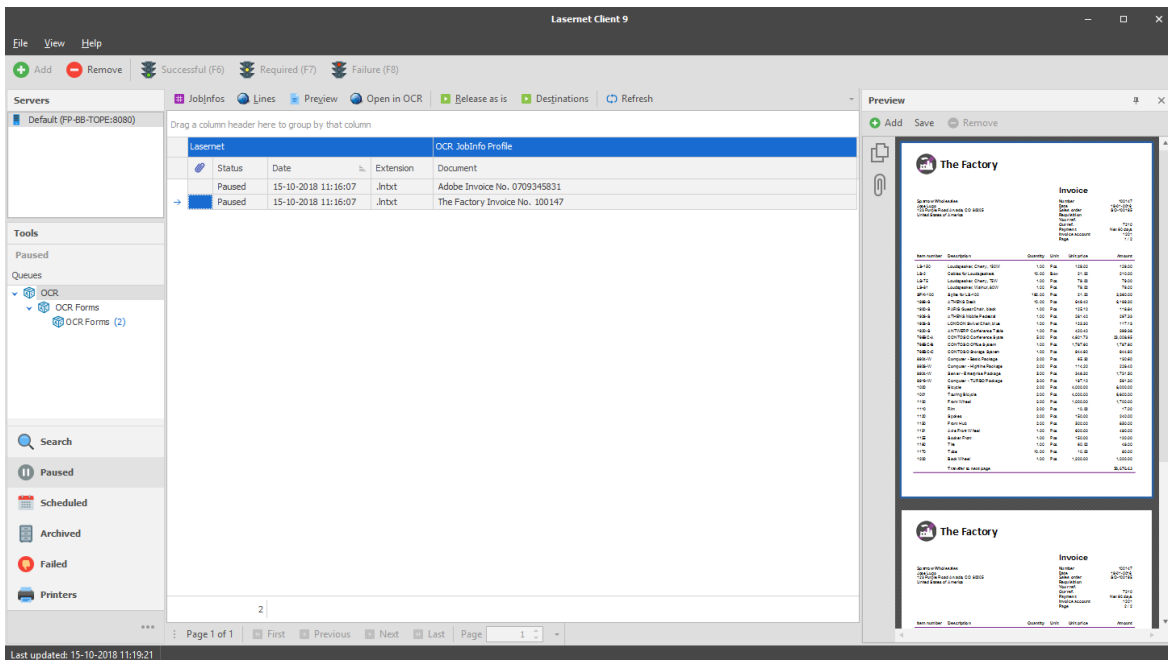


3.3.1 Connection settings

In the Server list you can maintain your server options and user credentials for the Lasetnet Client and Lasetnet OCR.



You are now able to view the paused jobs stored by the OCR Engine and edit them in the OCR Editor.

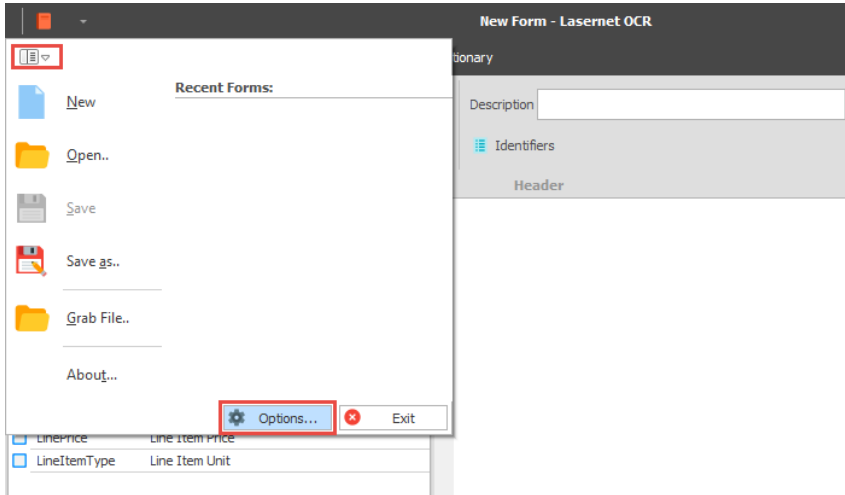


Select **Paused** documents and click **Open in OCR** to create a OCR form.

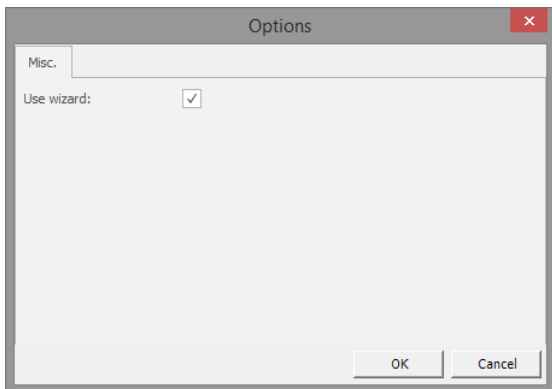
If you want to retry/release a paused job, first select the row, then right click on it and choose **Release as is**. If the job is recognized by the OCR Engine (e.g. if you have finished setting up an OCR form for it), it will be removed from the queue and forwarded to a Job Destination added to the OCR Engine in your Lasernet configuration.

3.4 Lasernet OCR

You can also start the Lasetnet OCR Editor and maintain your OCR forms without the Lasetnet Client. This requires that you also configure your connection settings in the Lasetnet OCR Editor as previously done in the Lasetnet Client.

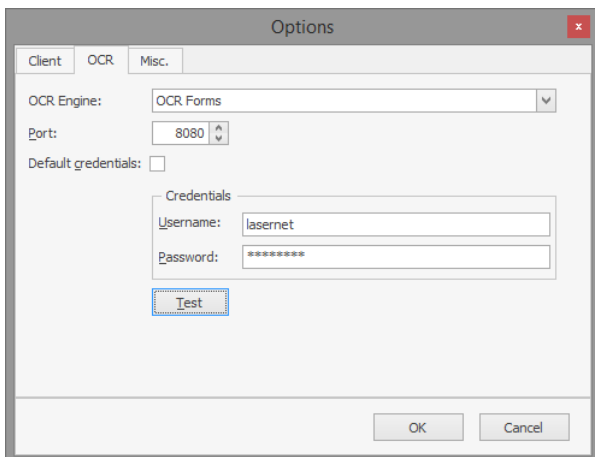


If you have logged in to the Lasetnet OCR from the Lasetnet Client, please note that the Options dialog is limited to a Misc tab only. The connections settings are inherited from and maintained by the Lasetnet Client and will overwrite the settings defined in the Lasetnet OCR Editor.



3.4.1 Connection settings

Open the Optios dialog in the Lasetnet OCR Editor, without using the Lasetnet Client, and you will be able manage the connection and user credentials. Enter the connection and credentials settings as configured in the OCR Engine on the server side.



OCR Engine

Logical module name defined for the OCR Engine. Multiple OCR Engines are supported on the same server.

Server

Server name on which the Lasernet Server is running.

Port

Port number to connect to. Check that the firewall settings on your server allow the Lasetnet OCR to connect on this port.

Username

Username defined in the OCR Engine.

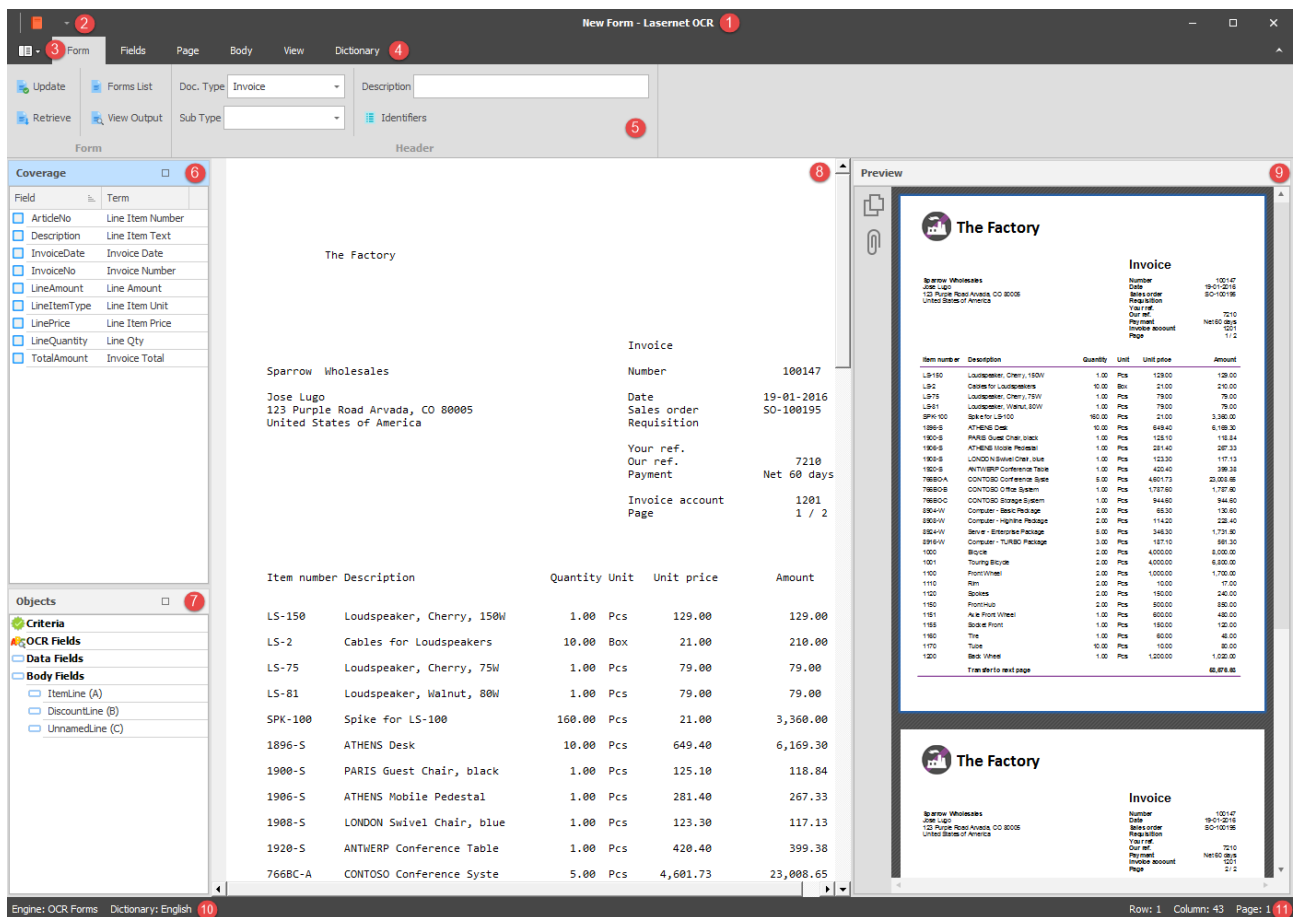
Password

The password for the username defined in the OCR Engine.

Click **Test** to try the connection between the client and server. If the connection succeeds, click 'OK' to accept the settings.

3.4.2 The user interface

The user interface consists of a window showing text data extracted from PDF or TIFF documents. Only the input view is needed for Lasetnet OCR, since the output format is already pre-defined (compared to the Lasetnet Form Editor where both the input and output format must be defined manually).



Form name (1)

The name of the form. If no name has been given it will be named New Form.

Dictionary (2)

Enter Dictionary mode. Add and manage terms and aliases to understand forms.

Options (3)

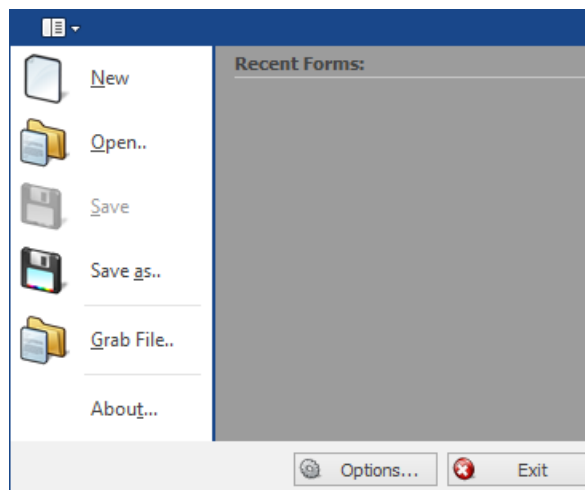
Set the Options for the Lasetnet OCR Editor: edit connection and user credentials, enable/disable the wizard for creating OCR Fields.

Menu bar (4)

You can access all the application functions and options through the various menu headings.



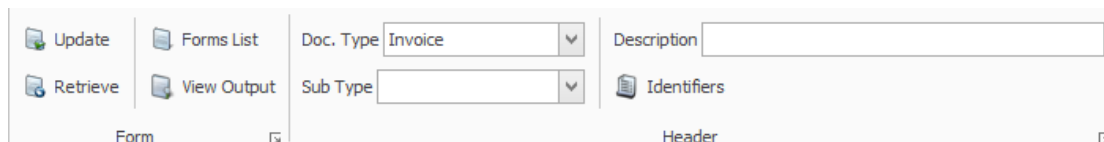
From the File menu you are able to Open, Save and create New OCR Forms. Grab files can be loaded into an existing form and the Options for the application can also be set (see chapter “Setting up the connection”).



Recent forms can be opened from the Recent Forms list. If you want to load a form that has been saved on the server, we recommend that you open it via the Form → Retrieve functionality. This is because the path is not shown in the Recent Forms list and you may end up using an old, temporary or locally stored version of the file instead.

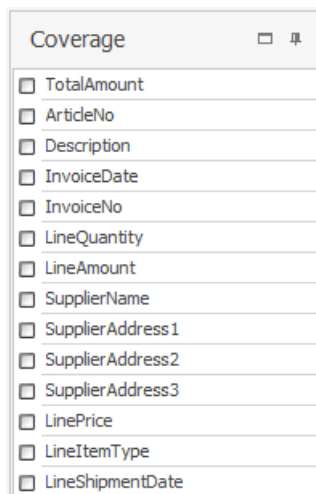
Menu settings (5)

Settings for form, field, body and view options.



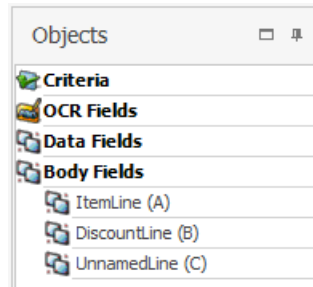
Coverage (6)

A list containing available Data Fields for the document type defined in the OCR Form. Activated Data Fields indicate that the field is already included in the OCR Form.



Object list (7)

A list containing objects added to the form. The list includes Criteria, OCR Fields, Data Fields and Body Fields.



Design View (8)

Document (JobData) represented in text format.

Preview (9)

A preview of the incoming PDF document (contents of the JobInfo PreviewJobData if it exists).

OCR Engine and Dictionary (10)

The name of the OCR Engine the application is connected to and the default dictionary language.

Row and column (11)

Positioning for selected object.

Row: 5 Column: 5

3.5 Getting started

3.5.1 Creating a Form

When you have added an OCR Engine to your project and a workflow to convert incoming PDF or TIFF files to a Text file, you are ready to start working with Lasernet OCR.

OCR Forms cannot be designed without having some current (or sample) business data to work from. This data needs to be stored in a text file on a share that is accessible to the Lasernet OCR client, or saved in paused mode in the internal Lasernet database.

You can create a new OCR form by following these easy steps:

Scenario 1 – Lasernet has converted and saved a text copy on a share.

Select Lasernet OCR 9 from your Lasernet Output Management program group. When the application is open, go to File – Grab file and browse for a valid text file.

Scenario 2 – Lasernet has converted and saved a text copy in the internal database.

Select Lasernet Client 9 from your Lasernet Output Management program group. When the application is open, click the Paused tab, and double click on any of the documents listed for the OCR Engine. The Lasernet OCR will open the document.

When the text/grab file has been loaded you are ready to start creating an OCR Form.

3.5.2 Form Properties

In the Form properties you are able to manage the properties of your form. The following settings are available:

Form Properties dialog box showing 'Doc. Type' set to 'Invoice' and 'Sub Type' set to an empty dropdown. The 'Description' field is empty. The 'Identifiers' button is visible.

Doc. Type

Type of document. List retrieved from the OCR Engine. Field names belonging to this document type will be available in the form.

Fields tab in the Form Properties dialog box. The 'Output XML' checkbox is checked. The 'Document Types' list shows 'Invoice' selected. The 'Fields' table lists various fields with their properties.

Active	Name	Alias	Show	Validation	Type	Term
<input checked="" type="checkbox"/>	InvoiceNo		<input checked="" type="checkbox"/>	Automatic	String	Invoice Number
<input checked="" type="checkbox"/>	InvoiceDate		<input checked="" type="checkbox"/>	None	Date	Invoice Date
<input checked="" type="checkbox"/>	TotalAmount		<input checked="" type="checkbox"/>	None	Number	Invoice Total
<input checked="" type="checkbox"/>	ArticleNo		<input checked="" type="checkbox"/>	None	String	Line Item Number
<input checked="" type="checkbox"/>	Description		<input checked="" type="checkbox"/>	None	String	Line Item Text
<input checked="" type="checkbox"/>	LineQuantity		<input checked="" type="checkbox"/>	None	String	Line Qty
<input checked="" type="checkbox"/>	LineAmount		<input checked="" type="checkbox"/>	None	Number	Line Amount
<input checked="" type="checkbox"/>	LinePrice		<input checked="" type="checkbox"/>	None	Number	Line Item Price
<input checked="" type="checkbox"/>	LineItemType		<input checked="" type="checkbox"/>	None	String	Line Item Unit

The list will include all documents added as document types in the OCR Engine that are not activated as Sub types.

Sub Type

Sub type for document. List retrieved from the OCR Engine. Field names belonging to this sub type will be available in the form.

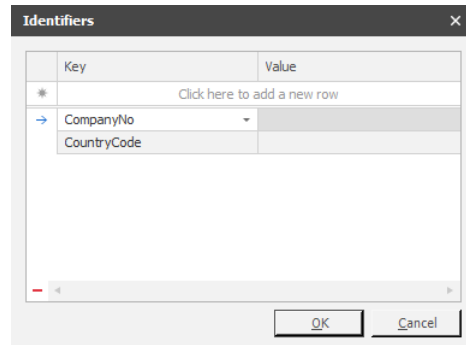
The list will include all documents added as sub types in the OCR.

Description:

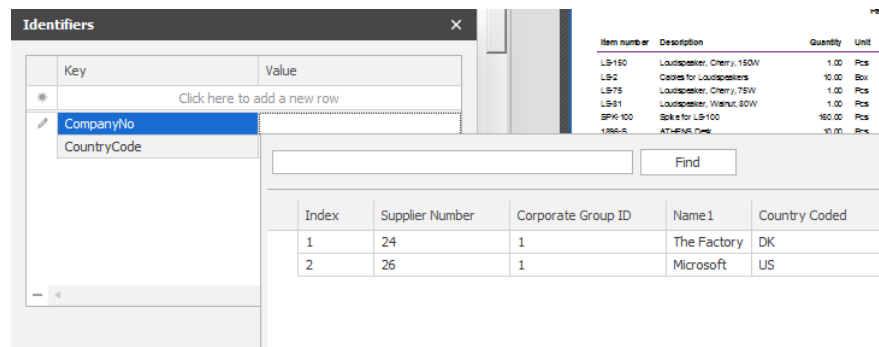
Description of the form. The text will be added as a value in the final output.

Identifiers:

Identifiers are fixed field names and values that can be added to output data. Only characters from A-Z are recommended as valid key names.

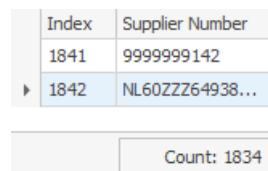


If a CSV file is assigned to the Identifier key, a grid will pop up where you can select the value presented in the column next to the Index number.



If the CSV file contains a parsing error, due to an incorrect CSV format, a dialog will pop-up with a warning and the line number saying: "Error parsing CSV file. Records with incorrect formatting will not appear in the list, but are still allocated an Index number."

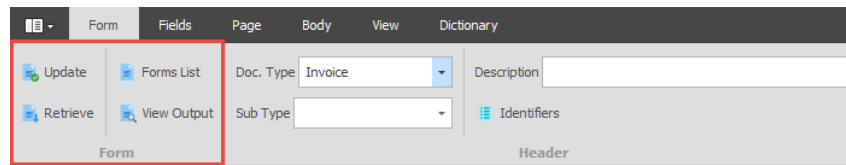
The Count value shows the total number of successfully imported records in the grid.



If there is mismatch between the highest index number and the total count number, please report the error to an administrator of the Lasernet Configuration Server, since this is often caused by a corrupted CSV file.

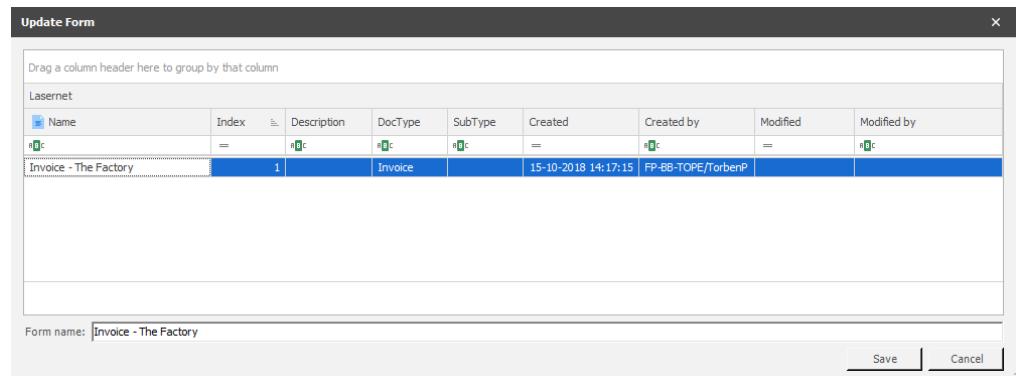
3.5.3 Form Tools

Administrative tools are available in the ribbon bar. You can update, retrieve, administrate and view OCR Forms.



Update

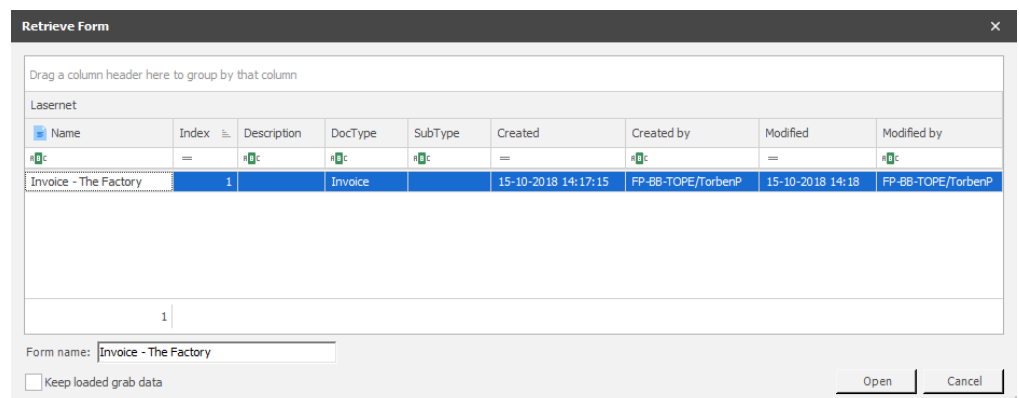
Click Update to save the OCR Form in the OCR Engine (production mode). Before an update you must ensure that the OCR Form contains unique criteria. The OCR Engine will parse through the OCR Forms in the list index order. The first form with a criteria match will be processed. The order of the index is managed from the Form Lists tool.



You can either save the OCR Form as a new document or overwrite an existing one.

Retrieve

Retrieves an already existing live OCR Form from the OCR Engine.



Click on an OCR Form in the form list and you will be able to open and edit its settings. A form that is already open must be saved before retrieving a new form.

Activate **Keep loaded grab data** to use loaded grab data instead of last used grab data stored within the form at last update.

Forms Lists

Lists the active OCR Forms already updated on the OCR Engine and running in live mode.

Forms List									
Drag a column header here to group by that column									
Lasernet									
Name	Index	Inactive	Description	DocType	SubType	Created	Created by	Modified	Modified by
Invoice - The Factory	1	<input type="checkbox"/>	Invoice	Invoice		15-10-2018 14:17:15	FP-BB-TOPE/TorbenP	15-10-2018 14:18	FP-BB-TOPE/TorbenP

Forms can be moved up and down to change the recognition order or deleted from the form list in the OCR Engine. Forms are deleted immediately and removed from the live server. To save a copy you must retrieve the form and use the File → Save as function.

View Output

View a preview of the final output, in XML or Text mode according to the setting on the OCR Engine.

```

- <LaserNet name="Invoice - The Factory" module="OCR">
  - <Header>
    <DocumentType>Invoice</DocumentType>
    <DocumentSubType></DocumentSubType>
    <Description></Description>
  </Header>
  <Identifiers/>
  - <Rearranges>
    <InvoiceNo></InvoiceNo>
    <InvoiceDate></InvoiceDate>
    <TotalAmount></TotalAmount>
    <ArticleNo></ArticleNo>
    <Description></Description>
    <LineQuantity></LineQuantity>
    <LineAmount></LineAmount>
    <LinePrice></LinePrice>
    <LineItemType></LineItemType>
  </Rearranges>
  <Patterns/>
  <Grab>DQoNCg0KDQoNCg0KDQogICAgICAgICAgICBUaGUGRmFjdG9yeQ0KDQoNCg</Grab>
</LaserNet>

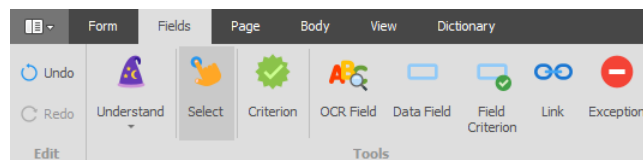
```

Shown above is the Output Viewer with an example of an OCR Form with XML as the output format.

Click "Save as..." if you want to save a local copy of the XML or Text file output. The output file can be opened in the Lasernet Form Editor or any other Text or XML editor.

3.5.4 Field Properties

The OCR toolbar provides easy access to various tools for defining Criteria, OCR Fields, Data Fields, Link Fields and Exceptions into the form design.



Understand

To use the Understand button a dictionary must be available. Click the arrow key to select a language and click Understand to automatically create OCR Criteria, OCR Fields and Data Fields for words/labels in the form that matches words found in the dictionary. By default the solution is delivered with an English and Danish dictionary. Other languages must be maintained by end-user.

Select

Select an existing object by clicking it.

Criterion

Manually select a criterion to tell the OCR Engine to scan for specific data strings and match the form. The full page will be scanned to detect the text string in any position. We recommended that you create several criteria for each OCR Form to ensure more accurate recognition e.g., document type, company name etc.

OCR Field

Used for detection of unique text strings in the document. An object will be created for the first text string in the document matching the OCR Field string (used in combination with Data Field and Link tool).

Data Field

When marking a text string in the OCR Form, the positioning for the Data Field is fixed (by default) to a defined row and column. See *Link tool for creating a Data Field which is positioned relative to the OCR Field*. Several Data Fields can be linked to one OCR Field.

Link

By using the Link tool you can draw a dotted line between an OCR Field and a Data Field, to define the relative positioning of the two objects. The Data Field will then be grabbed according to its relative position from the OCR Field.

Exception

Used in the body area as an exception for excluding patterns with specific text symbology. Exceptions uses 'CONTAINS' as an operator when matching strings. Exceptions in the object list can be edited by pressing F2.

3.5.5 Selections

Each tool has a selection value showing the settings of the object.

Criterion

A Criterion is marked as a dark grey object:



The screenshot displays the OCR software interface. The 'Criteria' tool is active, and the 'Objects' list on the left shows 'The Factory' and 'Invoice' selected. The main workspace shows the OCR scan of an invoice with 'The Factory' and 'Invoice' marked as dark grey objects. A preview window on the right shows the original invoice document.

Item number	Description	Quantity	Unit	Unit price	Amount
LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00
LS-75	Loudspeaker, Cherry, 75W	1.00	Pcs	79.00	79.00
LS-81	Loudspeaker, Walnut, 80W	1.00	Pcs	79.00	79.00
SPK-100	Spike for LS-100	160.00	Pcs	21.00	3,360.00
1896-S	ATHENS Desk	10.00	Pcs	649.40	6,493.60
1900-S	PARIS Guest Chair, black	1.00	Pcs	125.10	125.10
1906-S	ATHENS Mobile Pedestal	1.00	Pcs	267.40	267.40
1908-S	LONDON Swivel Chair, blue	1.00	Pcs	123.30	123.30
1920-S	ANTWERP Conference Table	1.00	Pcs	420.40	420.40
7668C-A	CONTOSO Conference System	5.00	Pcs	4,601.73	23,008.65

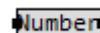
When selecting one of the objects, the selected field will show the value of the criterion in the following field:

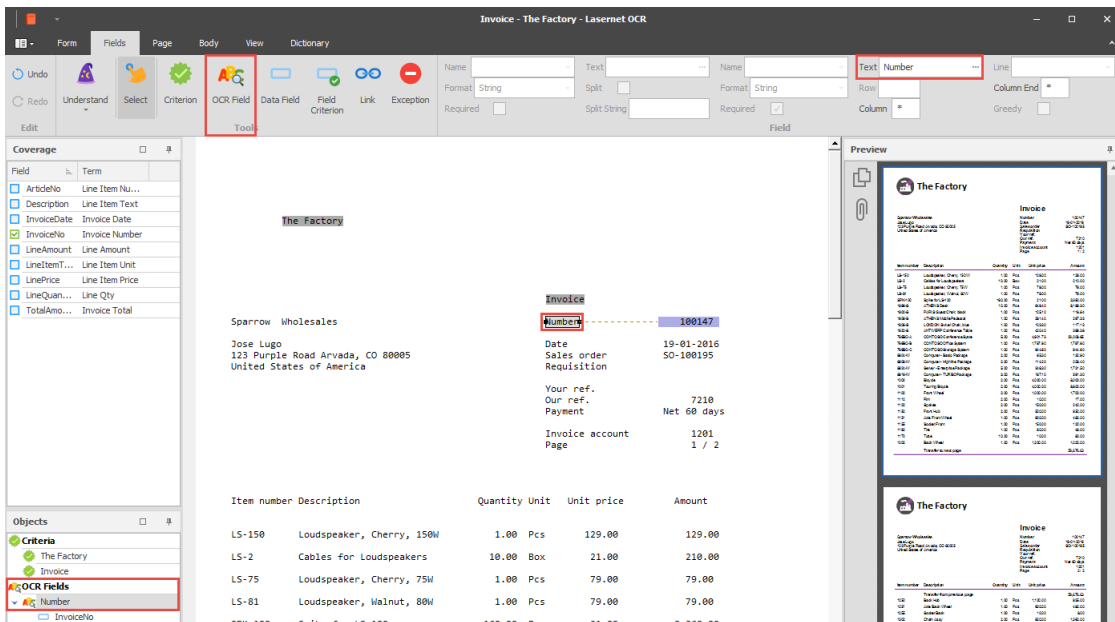
Text

Value of a criterion.

OCR Field

An OCR Field is marked as a light grey object:





When selecting the object, the selection field will show the value of the OCR Field in the following field:

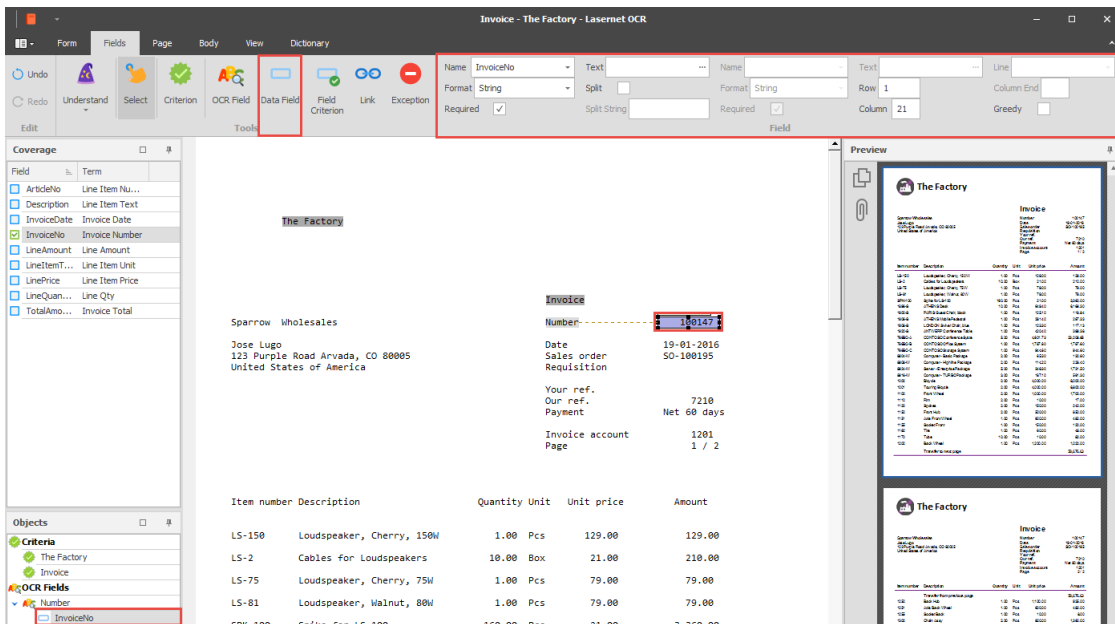
Text

Value of an OCR Field. The value is included in the output format, as it is only used for easily finding linked Data Fields.

Data Field

A Data Field is marked as a blue object:

100147



Name and Format are defined in the properties of a Data Field. If a Data Field is linked to an OCR Field, the value of the Row and Column is in a relative position to the OCR Field. If the Data Field is not linked to an OCR Field the Row and Column values are in an absolute position to the start of the form data.

Objects List View

Data Fields linked to an OCR Field are listed as a child to a specific OCR Field.

Data Fields not linked to an OCR Field are added to the list of general Data Fields.

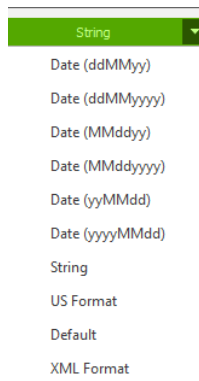
Name Name of the field. The list of available field names are obtained from the document type, managed in the OCR Engine.

The Field Name and its value will be included in the output format created by the OCR Engine (also viewable from the Form → Output View).

Format Name of the Regional Profile. The list of available profiles is retrieved from the Lasernet Server on startup. A field can either be a string, number or date (see Field Names list). Numbers and Amounts must match what is specified in the current profile (via the Regional Profiles), for successfully converting them into the final output format.

Name	Default	Description
Default	<input checked="" type="checkbox"/>	LaserNet standard configuration
US Format	<input type="checkbox"/>	Automatically generated profile
XML Format	<input type="checkbox"/>	

(An example of Regional Profiles created in the Lasetnet Developer and accessible from the Lasetnet OCR)



A range of predefined parameters for Dates and String are available.

The output result for a string field will be exactly as the data is presented in the document.

When using a Regional Profile for number and date formats, the separators specified in the Regional Profile must match the separators in the value of the field, otherwise the field will be ignored.

Required (1) Activated by default. This flags differently in the header/footer as opposed to the body. Header/footer: If this setting is not active, the Field Name will not be represented in the output format.

Body: A valid value is required for the row to be included in output.

Text Used as an exact match of a Data Field. If a text string has been entered into this field only Data Fields containing this match will be captured. Wildcards are not supported.

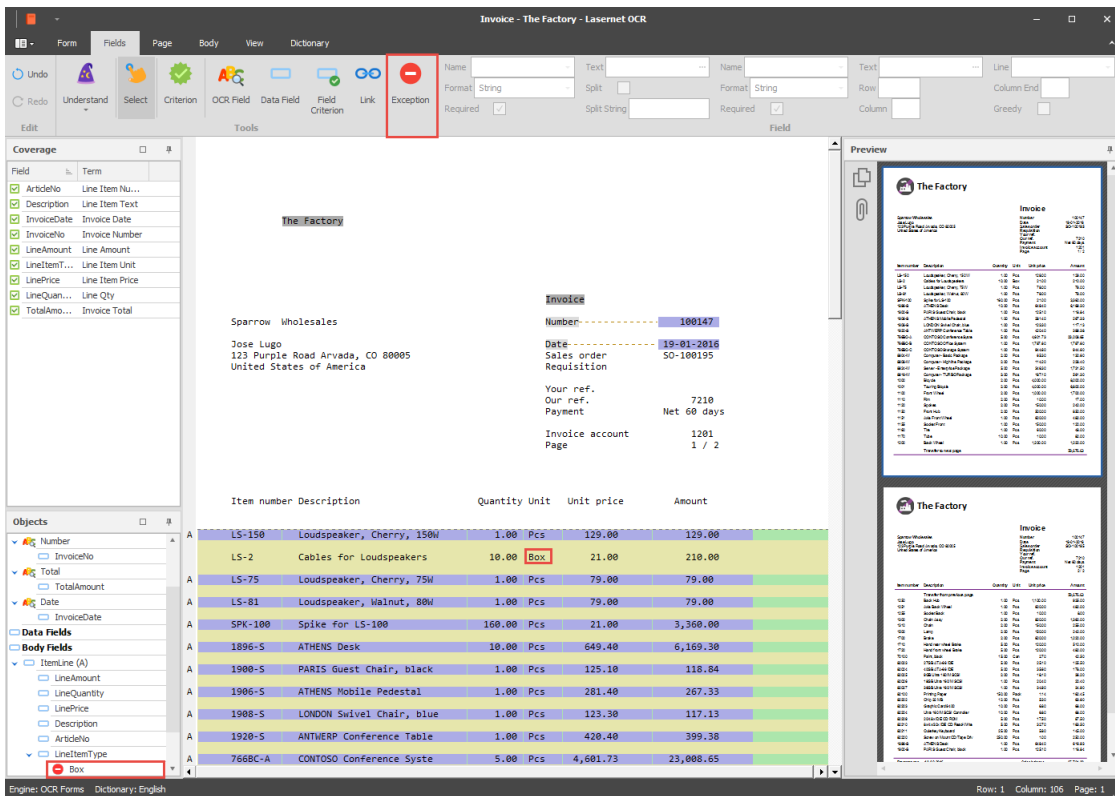
Split If a Data Field contains two different kinds of data information, like a string and a number, or a number and a date and there is no logical way of creating two Data Fields due to overlapping, the Split Field must be activated. This can typically happen for two columns listed in the body area after a PDF or TIFF document is converted into Text.

Name (2)	<p>Name of the second field when running in split mode (works similarly to single Field Names).</p> <p>The rule for splitting strings into two Field Names is that first Field Name contains the value of a date or a number, if represented in the string.</p> <p>The second Field Name will contain the rest of the string.</p> <p>If a number or date is not detected, both Field Names will be set as text strings. The splitter finds the first space character and adds the text string before it into the first Field Name and any character after the first space into the second Field name.</p>
Format (2)	Name of the Regional Profile for the second Data Field (works similarly to single Field Names).
Required (2)	<p>Activated by default. This flags differently in the header/footer as opposed to the body. Header/footer: If this setting is not active the Field Name will not be represented in the output format.</p> <p>Body: A valid value is required for the row to be included in the output.</p>
Row / Column	Position for a Data Field represented as rows and columns in a text document. Position of a Data Field is relative if linked to an OCR Field. If not linked the position is fixed.

Link No selections are available for the Link tool. The tool is used for linking a Data Field with an OCR Field; a dotted line shows which fields are connected.

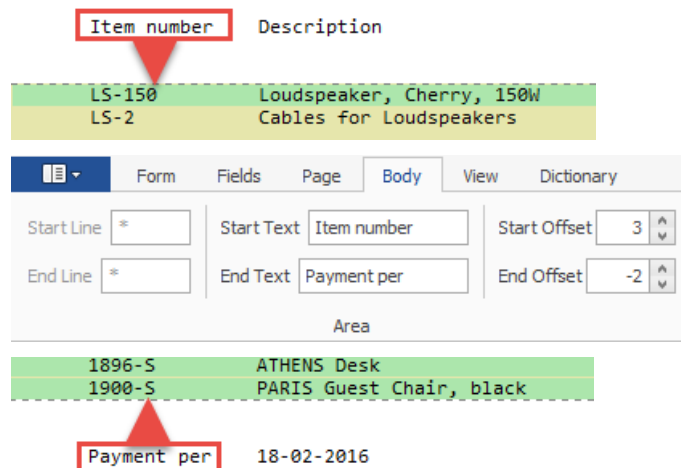
Number ----- 100147

Exception No selections are available for the Exception tool. The tool is used in the body area as an exception for excluding patterns with specific text symbology. The value of an Exception is connected to a Data Field and is shown in the list view along with active objects in the OCR Form.



3.5.6 Body Properties

The OCR toolbar enables you to set up the body area so that a dynamic number of article lines can be detected and used in the final output format.



Start Line Set fixed number of lines from where the body area starts within the input data file. The start line number must be a positive numeric value. This value is only required if a fixed start text cannot be found.

End Line Set fixed number of lines where the body area ends in the input data file. The end number must be a positive numeric value higher than the start line number or an asterisk (*) which means that the body stops at End of Page (as shown in the example above). This value is only required if a fixed end text cannot be found.

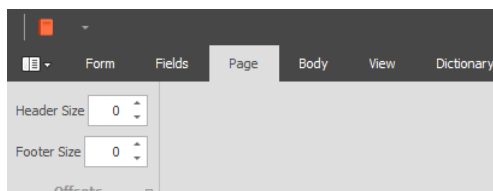
Start / End Text The body area can have its own Start and Stop text criteria which gives you flexibility beyond just defining its location in an input data file, by ensuring the data also satisfies the criteria set.

Start / End Offset

Used together with start criteria in the body area. The header (start) or footer (end) area can either be extended with a positive number of additional lines or decreased with a negative number of lines.

3.5.7 Page Properties

Each page will typically contain a header, body and footer. Very often the numbers of printed text lines, for the header and footer size, are defined with exactly the same size. In the **Page tab** you can define the size of the Header and Footer of the OCR Form.



The settings must be handled with care, since important data can be lost if the defined size does not fit the forms conditions. The benefit of removing headers and footers from middle pages in a document is easier maintenance of item lines.

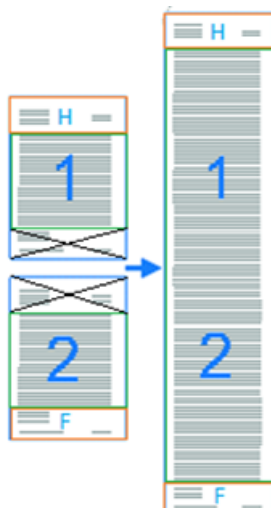
In the field **Header Size**, type in number of lines for header size; if the number of lines in header is exact the same for all pages you want to combine.

In the field **Footer Size**, type in number of lines for footer size; if the number of lines in footer is exact the same for all pages you want to combine.

The effect will be that:

- the header for the first incoming page
- the body lines for all incoming pages
- the footer for last incoming page

will be combined/converted into one grab file/job, in the illustration.



By default, the headers and footers for all pages will be visible in the data. This is not critical since the item lines are recognized as expected and the headers and footers in between are ignored.

Invoice - The Factory - Lasetnet OCR

Form Fields Page Body View Dictionary

Header Size: 29
Footer Size: 9

ArticleNo	Description	Quantity	Unit	Unit price	Amount
1170	Tube	10.00	Pcs	10.00	80.00
1200	Back wheel	1.00	Pcs	1,020.00	1,020.00
Transfer to next page					58,676.63
Page 2					
The Factory					
Sparrow Wholesales					
Jose Lugo					
123 Purple Road Arvada, CO 80005					
United States of America					
Invoice					
Number 100147					
Date 19-01-2016					
Sales order 50-100195					
Requisition					
Your ref. 7210					
Our ref. Net 60 days					
Payment					
Invoice account 1201					
Page 2 / 2					
Transfer from previous page					
58,676.63					
1250	Back Hub	1.00	Pcs	1,100.00	935.00
1251	Axle Back Wheel	1.00	Pcs	600.00	480.00
1255	Socket Back	1.00	Pcs	10.00	8.00

Engine: OCR Forms Dictionary: English Row: 88 Column: 71 Pages: 1

In this example the **Header Size** is 29 lines and **Footer Size** is 9 lines. The pages are combined and only one header (first page) and one footer (last page) appears in data.

Invoice - The Factory - Lasetnet OCR

Form Fields Page Body View Dictionary

Header Size: 37
Footer Size: 10

ArticleNo	Description	Quantity	Unit	Unit price	Amount
1100	Front Wheel	2.00	Pcs	1,000.00	1,700.00
1110	Rim	2.00	Pcs	10.00	17.00
1120	Spokes	2.00	Pcs	150.00	240.00
1150	Front Hub	2.00	Pcs	500.00	850.00
1151	Axle Front Wheel	1.00	Pcs	600.00	480.00
1155	Socket Front	1.00	Pcs	150.00	120.00
1160	Tire	1.00	Pcs	60.00	48.00
1170	Tube	10.00	Pcs	10.00	80.00
1200	Back Wheel	1.00	Pcs	1,020.00	1,020.00
1250	Back Hub	1.00	Pcs	1,100.00	935.00
1251	Axle Back Wheel	1.00	Pcs	600.00	480.00
1255	Socket Back	1.00	Pcs	10.00	8.00
1300	Chain Assy	2.00	Pcs	800.00	1,360.00
1310	Chain	2.00	Pcs	150.00	255.00
1500	Lamp	2.00	Pcs	150.00	240.00

Engine: OCR Forms Dictionary: English Row: 88 Column: 71 Pages: 1

3.5.8 View Properties

Set up the preferred font type for grab text.

Font Font character name. Default is Consolas.

Size Font character size. Default is 9.

Form Fields Page Body View Dictionary

Font: Consolas
Size: 9

Grab Text

None of the settings will have effect on the final output. This is for viewing only.

3.6 Walk-through – How to create a Form

3.6.1 Create a Form

When you have added an OCR Engine to your project and a workflow to convert incoming PDF or TIFF files to a Text file, you are ready to start working with Lasernetet OCR.

OCR Forms cannot be designed without having some current (or sample) business data to work from. This data needs to be stored in a text file on a share that is accessible to the Lasetnet OCR client, or saved in paused mode in the internal Lasetnet database.

You can create a new OCR form by following these easy steps:

Scenario 1 – Lasetnet has converted and saved a text copy on a share.

Start Lasetnet OCR 9. When the application is open go to File → Grab file and browse for a valid text file.

Scenario 2 – Lasetnet has converted and saved a text copy in the internal database.

Start Lasetnet Client 9. When the application is open, click the Paused tab and double click on any of the documents listed for the OCR Engine. The Lasetnet OCR will open the document.

In our walk-through example we will set up an incoming invoice from a supplier named “The Factory”. A PDF/A file is received by email and converted into text format by the Lasetnet Server. From the Lasetnet Client we are now able to open the invoice saved as a temporary text file.

The screenshot displays the Lasetnet OCR application interface. The main window is titled "Invoice - The Factory - Lasetnet OCR". The interface is divided into several sections:

- Form Fields:** Includes "Update", "Retrieve", "Forms List", "View Output", "Doc. Type" (set to "Invoice"), "Description", "Sub Type", and "Identifiers".
- Coverage Table:** A table with columns "Field" and "Term". It lists various fields like "ArticleNo", "Description", "InvoiceDate", "InvoiceNo", "LineAmount", "LineItem...", "LinePrice", "LineQuan...", and "TotalAmo..." with their corresponding terms.
- Objects:** A list of objects including "Criteria", "OCR Fields", "Data Fields", "Body Fields", "ItemLine (A)", "DiscountLine (B)", and "UnnamedLine (C)".
- Main Invoice Content:**

The Factory
Sparrow Wholesales
Jose Lugo
123 Purple Road Arvada, CO 80005
United States of America

Invoice
Number 100147
Date 19-01-2016
Sales order SO-100195
Requisition
Your ref. 7210
Payment Net 60 days
Invoice account 1201
Page 1 / 2

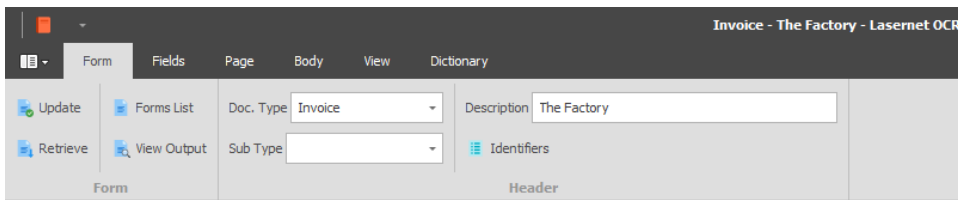
Item number	Description	Quantity	Unit	Unit price	Amount
LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00
LS-75	Loudspeaker, Cherry, 75W	1.00	Pcs	79.00	79.00
LS-81	Loudspeaker, Walnut, 80W	1.00	Pcs	79.00	79.00
SPK-100	Spike for LS-100	160.00	Pcs	21.00	3,360.00
1096-S	ATHENS Desk	10.00	Pcs	649.40	6,494.00
1900-S	PARIS Guest Chair, black	1.00	Pcs	125.10	125.10
1906-S	ATHENS Mobile Pedestal	1.00	Pcs	267.33	267.33
1908-S	LONDON Swivel Chair, blue	1.00	Pcs	123.30	123.30
1920-S	ANTHERP Conference Table	1.00	Pcs	420.40	420.40
766BC-A	CONTOSO Conference System	5.00	Pcs	4,601.73	23,008.65
- Preview Window:** Shows a visual representation of the original PDF invoice from "The Factory".

The incoming PDF file will often have logos and graphical elements embedded. The key data (text) will be inserted in to rows and columns in a text file. The preview window will show a copy of the original PDF document.

When the text/grab file is loaded in Lasetnet OCR you are ready to create an OCR Form.

3.6.2 Setting up Form Properties

Set up the form properties. Click the Form tab.



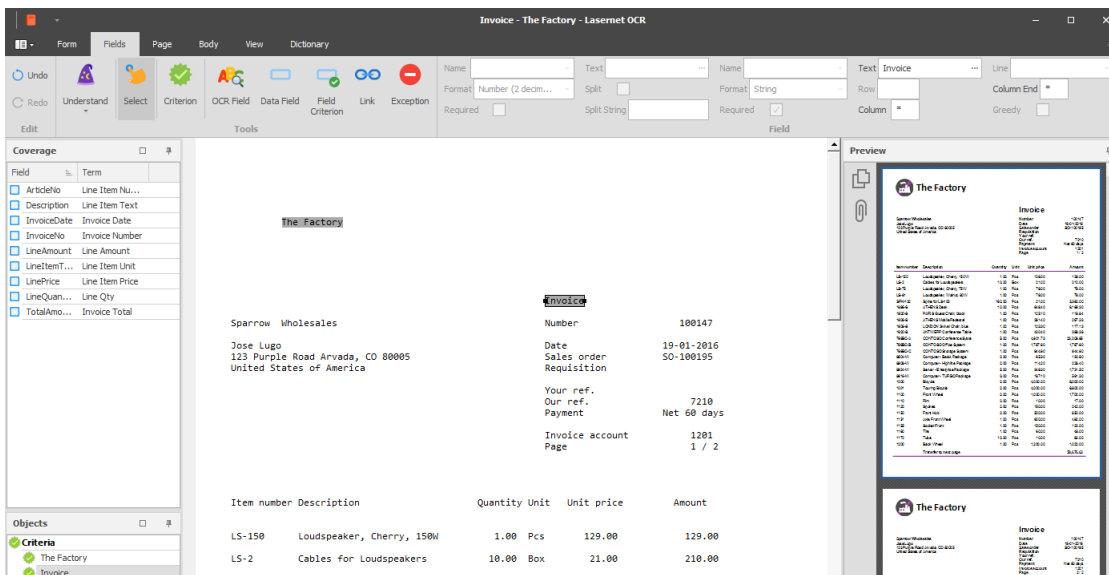
From the Doc. Type list select the proper document type. In this example we have selected **Invoice** as document type and that the contents belong to a company called **The Factory**. A list of required OCR fields are defined in the OCR Engine for the invoice and they will now appear in the Coverage window as a kind of to do list.

3.6.3 Setting up Criteria

To successfully match an invoice from “The Factory”, we must set up a couple of unique criteria. In this example we are setting up two criteria. One matching the document type “Invoice” another matching the supplier name “The Factory”.

Click the Criterion tool and mark the supplier name “The Factory”, by drawing a box in the text string. Then draw a second box around the form name “Invoice”.

A criterion match can only match a single line of text.



Two dark grey objects shows how to match the OCR Form:

The Factory **Invoice**

Documents that are processed through the OCR Engine and match the two strings will be identified as an **Invoice** from **The Factory**. Add additional criteria to make the matching process more unique to provide greater differentiation between other processed documents. Matching only the word ‘Invoice’ could end up matching every invoice from every supplier, not just “The Factory”.

3.6.4 OCR Field

The next step is to set up the OCR and Data Fields. Select the **Field** band.

Firstly you need to pick up the invoice number.

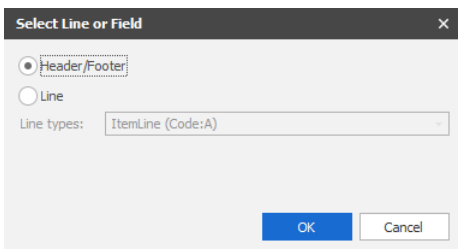
Select the **InvoiceNo** in the **Coverage** list:

Field	Term
<input type="checkbox"/> ArticleNo	Line Item Nu...
<input type="checkbox"/> Description	Line Item Text
<input type="checkbox"/> InvoiceDate	Invoice Date
<input type="checkbox"/> InvoiceNo	Invoice Number
<input type="checkbox"/> LineAmount	Line Amount
<input type="checkbox"/> LineItemT...	Line Item Unit
<input type="checkbox"/> LinePrice	Line Item Price
<input type="checkbox"/> LineQuan...	Line Qty
<input type="checkbox"/> TotalAmo...	Invoice Total

A wizard will guide you through the required selections for marking/selecting the OCR Field, Data Field, Field Name and Field Format.



Mark the first **OCR Field**, containing the **Number**, by drawing a box around the text string or you can double click on the word.



Select whether the **Number** exists in the **Header/Footer** or is a part of an item line. In this example the label **Number** is positioned in the header. Click **OK** to confirm.



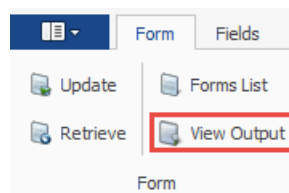
To mark the actual invoice number, click the **Data Field** and draw a box around the invoice number **100147**.



When the Field Name is selected the wizard will automatically close. The selections in the wizard vary depending on the field types, since a field format is required for date types and amounts.

Two new objects are created. The light grey one is the OCR Field, and the blue one is the Data Field. A dotted line between the two objects shows that they are linked together and that the position of the Data Field is relative to the position of the text string "Invoice No.".

The first field is mapped for the The Factory. If you want to view the result of the the extracted data, select the **Form** tab and click the **View Output** button.



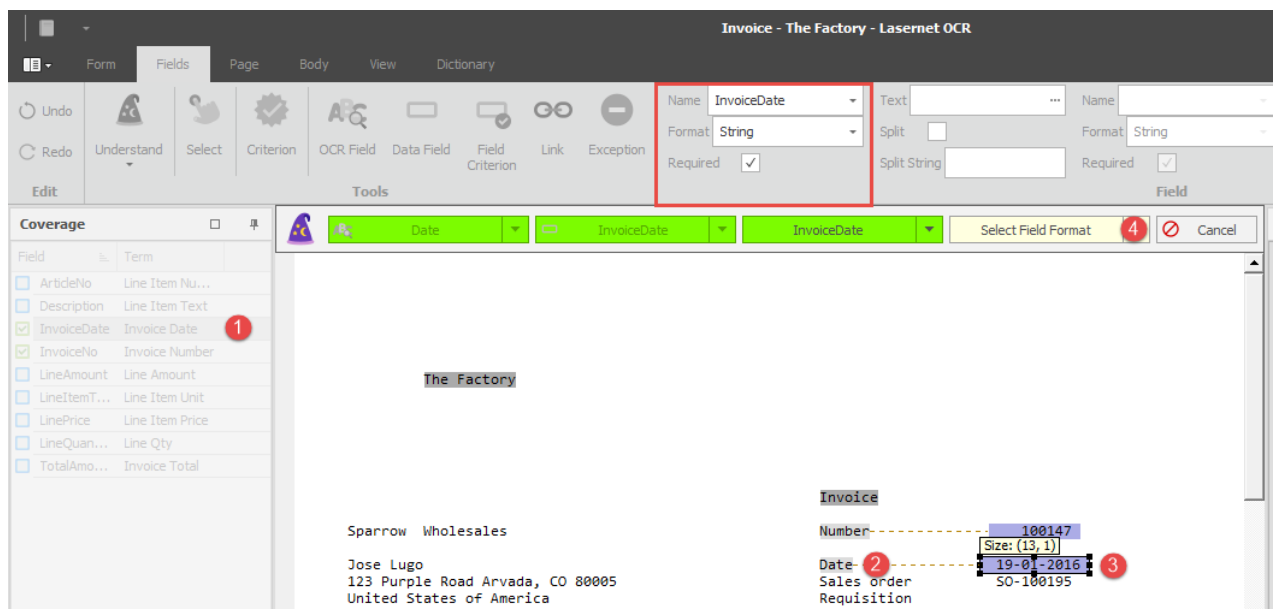
This is the preview result of the pre-defined XML format created by the Lasernet OCR Editor.

```

- <LaserNet name="Invoice - The Factory" module="OCR">
- <Header>
  <DocumentType>Invoice</DocumentType>
  <DocumentSubType></DocumentSubType>
  <Description>The Factory</Description>
</Header>
<Identifiers/>
- <Rearranges>
  <InvoiceNo>100147</InvoiceNo>
  <InvoiceDate></InvoiceDate>
  <TotalAmount></TotalAmount>
  <ArticleNo></ArticleNo>
  <Description></Description>
  <LineQuantity></LineQuantity>
  <LineAmount></LineAmount>
  <LinePrice></LinePrice>
  <LineItemType></LineItemType>
</Rearranges>
<Patterns/>
<Grab>DQoNCg0KDQoNCg0KDQogICAgICAgICAgICAgICBUaGUGRmFjdG9yeQ0KDQoNCg0KDK</Grab>
</LaserNet>

```

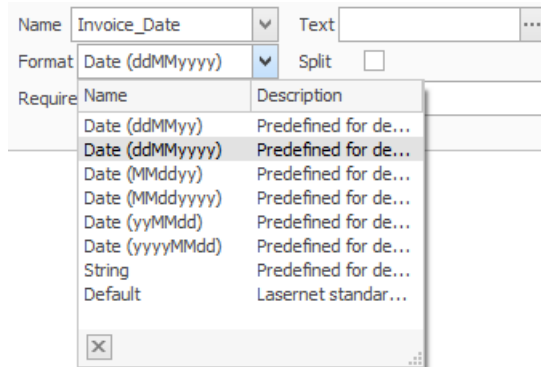
The next step is to go through the other fields from the coverage list.



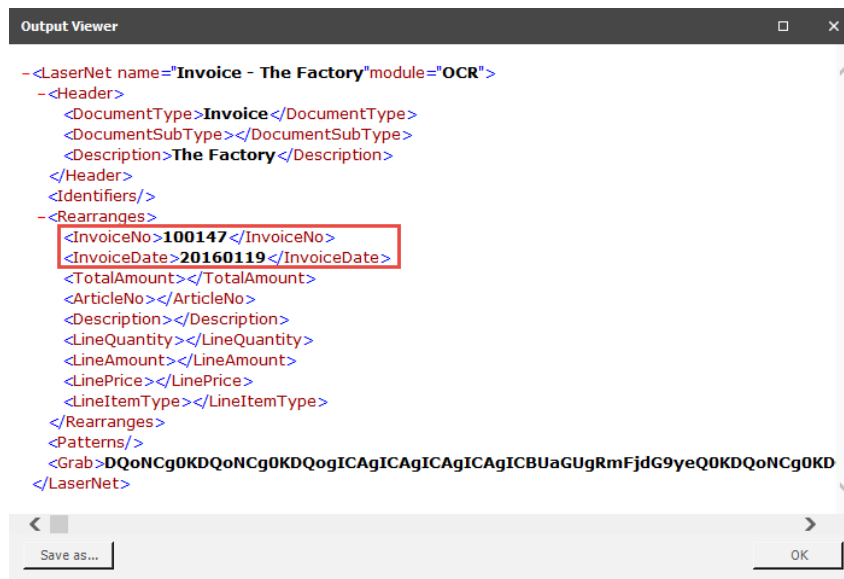
Now we will map the invoice date by starting the wizard from the Coverage window. Click on the field **InvoiceDate** in the Coverage windows and follow the same steps as in the previous example. Select text string "Date" and select InvoiceDate as Field Name. Mark "24-01-11" as the value for the Data Field.

Two new objects are created; an OCR Field recognizing the Date string and a Data Field to capture the date value.

Set the proper data format for the Date field. In this example the format must be set to **ddMMyyyy** that will match **19-01-2016**.



Two fields are now mapped. Go to the Form tab and click the View Output button.



The date value is now added to the XML structure and converted to a standard format **yyyyMMdd** which is readable by most workflow systems.

Finally we will set up OCR and Data Fields to match the total amount of the invoice. The amount will be captured in a position relative to the text string **Total**.

Select the **TotalAmount** in the Coverage window and go to the wizard to create the **OCR Field** and **Data Field**. Set **Field Format** to a **Number** with **2 decimals**.

Total		TotalAmount	TotalAmount	Select Field Format		Cancel
1710	Hand rear wheel Brake	5.00	Pcs	120.00	510.00	
1720	Hand front wheel Brake	5.00	Pcs	120.00	480.00	
70100	Paint, black	15.00	Can	2.70	40.50	
80023	27GB ATA-66 IDE	5.00	Pcs	25.10	125.50	
80024	40GB ATA-66 IDE	5.00	Pcs	35.60	178.00	
80025	9GB Ultra 160/M SCSI	2.00	Pcs	18.10	36.20	
80026	18GB Ultra 160/M SCSI	1.00	Pcs	20.40	20.40	
80027	36GB Ultra 160/M SCSI	1.00	Pcs	34.90	34.90	
80100	Printing Paper	150.00	Pack	1.14	162.45	
80202	Chip 32 MB	12.00	Pcs	5.30	63.60	
80203	Graphic Card 9400	10.00	Pcs	6.80	68.00	
80204	Ultra 160/M SCSI Controller	10.00	Pcs	6.80	68.00	
80209	20/48x IDE CD ROM	5.00	Pcs	17.50	87.50	
80210	8x/4x/32x IDE CD Read-Write	5.00	Pcs	32.70	163.50	
80211	Quietkey Keyboard	25.00	Pcs	5.80	145.00	
80220	Screw on Mount CD/Tape Driv	250.00	Pcs	1.00	250.00	
1896-S	ATHENS Desk	1.00	Pcs	649.40	616.93	
1900-S	PARIS Guest Chair, black	1.00	Pcs	125.10	118.84	
Payment per 18-02-2016				Sales balance	65,731.00	
				Round-off	0,00	
				Total	65,731.00	

We have now added the total amount to the list of covered OCR Fields. Go to the **Form tab** and click the **View Output** button.

```

Output Viewer
- <LaserNet name="Invoice - The Factory" module="OCR">
- <Header>
  <DocumentType>Invoice</DocumentType>
  <DocumentSubType></DocumentSubType>
  <Description>The Factory</Description>
</Header>
<Identifiers/>
- <Rearranges>
  <InvoiceNo>100147</InvoiceNo>
  <InvoiceDate>20160119</InvoiceDate>
  <TotalAmount>65731.0000</TotalAmount>
  <ArticleNo></ArticleNo>
  <Description></Description>
  <LineQuantity></LineQuantity>
  <LineAmount></LineAmount>
  <LinePrice></LinePrice>
  <LineItemType></LineItemType>
</Rearranges>
<Patterns/>
<Grab>DQoNCg0KDQoNCg0KDQogICAgICAgICAgICAgICBUaGUGRmFjdG9yeQ0KDQoNCg0KDC
</LaserNet>

```

The value for the total amount is added to the XML structure as a standard number format with four decimals. You are now ready to add the rest of the required fields to the OCR Form. Body lines are covered in the next chapter.

3.6.5 Covered Data Fields

In the Coverage window a check mark will show the Data Fields that are covered in the form. This is a guideline to help you to cover all the required fields. The OCR Engine is still able to process OCR Forms even if all of the fields are not covered.

Field	Term
<input checked="" type="checkbox"/> InvoiceNo	Invoice Number
<input checked="" type="checkbox"/> InvoiceDate	Invoice Date
<input checked="" type="checkbox"/> TotalAmo...	Invoice Total
<input type="checkbox"/> ArticleNo	Line Item Nu...
<input type="checkbox"/> Description	Line Item Text
<input type="checkbox"/> LineQuan...	Line Qty
<input type="checkbox"/> LineAmount	Line Amount
<input type="checkbox"/> LinePrice	Line Item Price
<input type="checkbox"/> LineItemT...	Line Item Unit

3.6.6 Creating the Body Area

The next step is to define the **Body Area** of the form to recognize the item lines and capture the **Data Fields**. OCR Fields are not required in the body area as more complex algorithms are used to intelligently match the Data Fields instead.

Click the **Body** tab and set **Start Line** to **36**, to define the start of the item lines.

The screenshot shows the 'Body' configuration tab in the software. The 'Start Line' is set to 29, indicated by a red arrow. The 'Coverage' list on the left includes 'Invoice_Total', 'Invoice_Number', and 'Invoice_Date'. The main area displays an OCR'd invoice for 'The Factory' with a table of items at the bottom. The table has columns for Item number, Description, Quantity, Unit, Unit price, and Amount.

Item number	Description	Quantity	Unit	Unit price	Amount
LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00

The end of the body can either be set to a fixed **End Line** value or an **End Text** that marks the end of the body or the start of a footer.

In this example we will add an End Text string to define a body with a variable length. Type **Payment per** to locate the end of the body/footer start.

Additional settings to use an Offset relative to a **Start Text** or **End Text** value are also available. The values can be set to either a positive or negative value. An offset will include fewer or more lines in the body area relative to where the Start or End Text is located.

For more info about Start/End Offsets, see the chapter “Body Properties”.

3.6.7 Data Fields in Body

Next step is to cover the Data Fields that appears in the body.

Select the **Data Field** tool or the Data Field name for **Line_Amount** from the Coverage window. We recommended that you start capturing columns that contain numbers with decimals. The algorithms filtering the Data Fields and finding item lines are optimized for this purpose.

Because you do not need to create an OCR Field you can select **Mark or select Data Field** in the wizard and afterwards mark the column with the line amount.

Draw a box around the value in the column and ensure that it is wide enough to contain the maximum number of characters possible for that column, otherwise data will get cut off or included in the wrong Data Field for other documents of same type.

Invoice - The Factory - Lasernet OCR

Form Fields Page Body View Dictionary

Undo Redo Understand Select Criterion OCR Field Data Field Field Criterion Link Exception

Name: [] Text: [] Name: []
 Format: Number (0 decim...) Split: [] Format: String
 Required: [x] Split String: [] Required: [x] Field

Field	Term	Item number	Description	Quantity	Unit	Unit price	Amount
<input checked="" type="checkbox"/>	InvoiceNo	Invoice Number	A LS-150	Loudspeaker, Cherry, 150W	1.00 Pcs	129.00	129.00
<input checked="" type="checkbox"/>	InvoiceDate	Invoice Date	A LS-2	Cables for Loudspeakers	10.00 Box	21.00	210.00
<input checked="" type="checkbox"/>	TotalAmo...	Invoice Total	A LS-75	Loudspeaker, Cherry, 75W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	ArticleNo	Line Item Nu...	A LS-81	Loudspeaker, Walnut, 80W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	Description	Line Item Text	A SPK-100	Spike for LS-100	160.00 Pcs	21.00	3,360.00
<input checked="" type="checkbox"/>	LineQuan...	Line Qty	A 1896-S	ATHENS Desk	10.00 Pcs	649.40	6,169.30
<input checked="" type="checkbox"/>	LineAmount	Line Amount					
<input checked="" type="checkbox"/>	LinePrice	Line Item Price					
<input checked="" type="checkbox"/>	LineItemT...	Line Item Unit					

Now we have six blue objects which represent the Data Fields in the item lines.

Name ArticleNo
Format String
Required Yes
Size 15 characters

Invoice - The Factory - Lasernet OCR

Form Fields Page Body View Dictionary

Undo Redo Understand Select Criterion OCR Field Data Field Field Criterion Link Exception

Name: ArticleNo Text: [] Name: []
 Format: String Split: [] Format: String
 Required: [x] Split String: [] Required: [x] Field

Field	Term	Item number	Description	Quantity	Unit	Unit price	Amount
<input checked="" type="checkbox"/>	InvoiceNo	Invoice Nu...	A LS-150	Loudspeaker, Cherry, 150W	1.00 Pcs	129.00	129.00
<input checked="" type="checkbox"/>	InvoiceDate	Invoice Date	A LS-2	Cables for Loudspeakers	10.00 Box	21.00	210.00
<input checked="" type="checkbox"/>	TotalAmount	Invoice Total	A LS-75	Loudspeaker, Cherry, 75W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	ArticleNo	Line Item N...	A LS-81	Loudspeaker, Walnut, 80W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	Description	Line Item Text	A SPK-100	Spike for LS-100	160.00 Pcs	21.00	3,360.00
<input checked="" type="checkbox"/>	LineQuantity	Line Qty	A 1896-S	ATHENS Desk	10.00 Pcs	649.40	6,169.30
<input checked="" type="checkbox"/>	LineAmount	Line Amount					
<input checked="" type="checkbox"/>	LinePrice	Line Item Price					
<input checked="" type="checkbox"/>	LineItemType	Line Item Unit					

Name Description
Format String
Required Yes
Size 33 characters

Invoice - The Factory - Lasernet OCR

Form Fields Page Body View Dictionary

Undo Redo Understand Select Criterion OCR Field Data Field Field Criterion Link Exception

Name: Description Text: [] Name: []
 Format: String Split: [] Format: String
 Required: [x] Split String: [] Required: [x] Field

Field	Term	Item number	Description	Quantity	Unit	Unit price	Amount
<input checked="" type="checkbox"/>	InvoiceNo	Invoice Num...	A LS-150	Loudspeaker, Cherry, 150W	1.00 Pcs	129.00	129.00
<input checked="" type="checkbox"/>	InvoiceDate	Invoice Date	A LS-2	Cables for Loudspeakers	10.00 Box	21.00	210.00
<input checked="" type="checkbox"/>	TotalAmount	Invoice Total	A LS-75	Loudspeaker, Cherry, 75W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	ArticleNo	Line Item Nu...	A LS-81	Loudspeaker, Walnut, 80W	1.00 Pcs	79.00	79.00
<input checked="" type="checkbox"/>	Description	Line Item Text	A SPK-100	Spike for LS-100	160.00 Pcs	21.00	3,360.00
<input checked="" type="checkbox"/>	LineQuantity	Line Qty	A 1896-S	ATHENS Desk	10.00 Pcs	649.40	6,169.30
<input checked="" type="checkbox"/>	LineAmount	Line Amount					
<input checked="" type="checkbox"/>	LinePrice	Line Item Price					
<input checked="" type="checkbox"/>	LineItemType	Line Item Unit					

Name LineQuantity
Format String
Required Yes
Size 10 characters

Field	Term	Item number	Description	Quantity	Unit	Unit price	Amount
InvoiceNo	Invoice Num...			1.00	Pcs	129.00	129.00
InvoiceDate	Invoice Date						
TotalAmount	Invoice Total						
ArticleNo	Line Item Nu...						
Description	Line Item Text						
LineQuantity	Line Qty	LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
LineAmount	Line Amount						
LinePrice	Line Item Price						
LineItemType	Line Item Unit						

If you are sure about number formats you can always set the format type to string. The value will then keep the same format in output data.

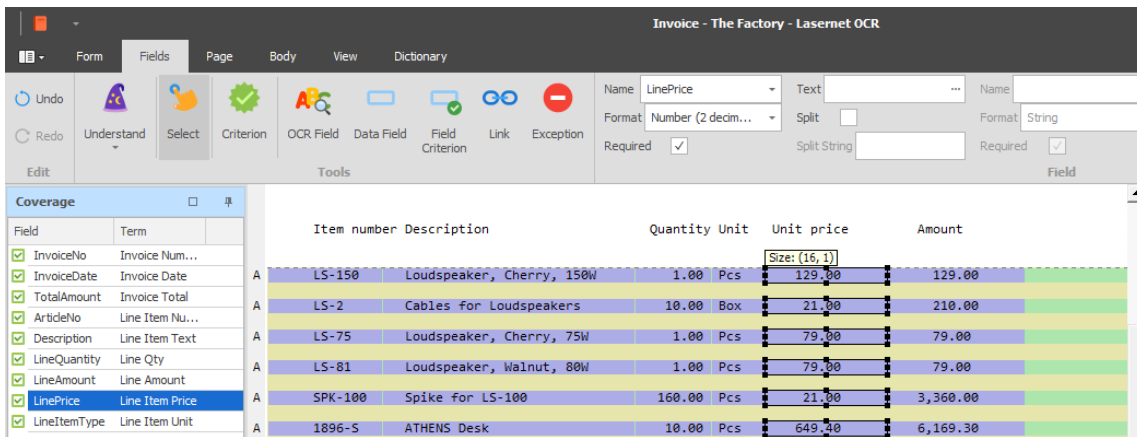
Name LineItemType
Format String
Required No
Size 7

Field	Term	Item number	Description	Quantity	Unit	Unit price	Amount
InvoiceNo	Invoice Num...			1.00	Pcs	129.00	129.00
InvoiceDate	Invoice Date						
TotalAmount	Invoice Total						
ArticleNo	Line Item Nu...						
Description	Line Item Text						
LineQuantity	Line Qty	LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
LineAmount	Line Amount						
LinePrice	Line Item Price						
LineItemType	Line Item Unit						

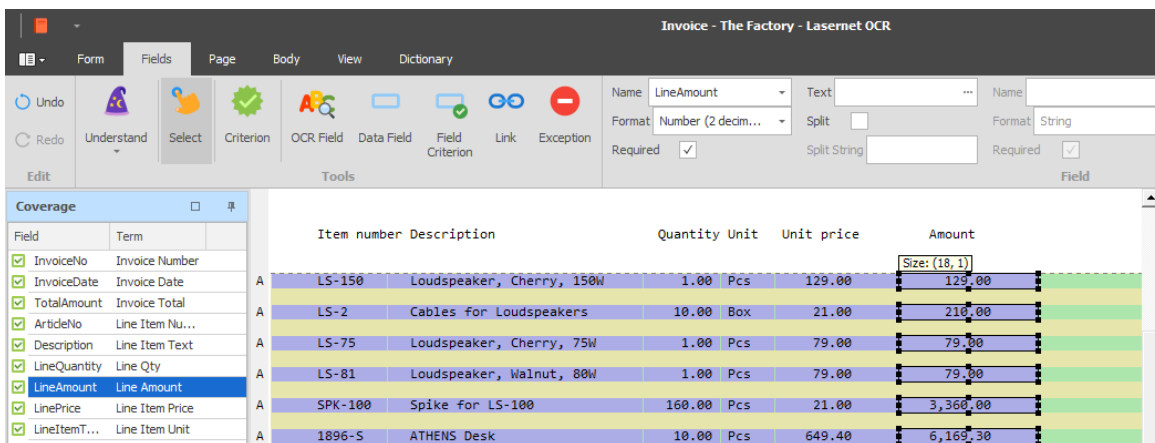
If the Required property is set to inactive then item lines will still be recognized and extracted.

If the Required property is set to active, only item lines where this column includes data will be extracted. Other Data Fields in the same column will also be ignored.

Name LinePrice
Format Number (2 decimals)
Required Yes
Size 16 characters



Name LineAmount
Format Number (2 decimals)
Required Yes
Size 18 characters

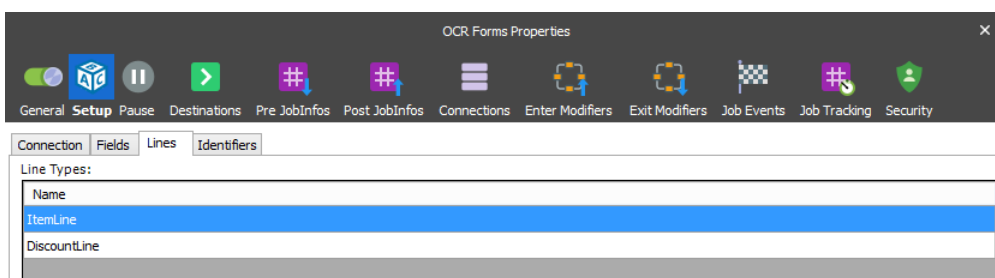


All the required fields are now mapped for the invoice from “The Factory” and the form is ready to be deployed.

3.6.8 Additional Item Lines

Each item line must be attached to a Line Type. The default value is **UnnamedLine** and is always available in OCR forms. If you only extract one type of item lines you can use the default name.

Additional line types can be defined in the OCR Engine and selected afterwards in the OCR forms. The line type is used as a reference for the workflow system to know what kind of item line it is. In the XML structure the item lines will be grouped into the selected line types.



Two line types are defined for this form type and a default system line type named UnnamedLine is also present.

Item number	Description	Quantity	Unit	Unit price	Amount
A LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
A LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00
A LS-75	Loudspeaker, Cherry, 75W	1.00	Pcs	79.00	79.00
A LS-81	Loudspeaker, Walnut, 80W	1.00	Pcs	79.00	79.00
A SX-100	Spike for LS-100	160.00	Pcs	21.00	3,360.00
A 1896-S	ATHENS Desk	10.00	Pcs	649.40	6,169.30

You must be able to find matches in the item lines to differentiate between an item line with or without a discount. In the user interface you can see the difference between Line Types. In front of each line you will find a letter (starting from A, B, C...). The letter will refer to which line type the pattern belongs to.

3.6.9 View output

Go to Form tab and click the View Output button.

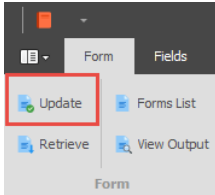


```
-<LaserNet name="Invoice - The Factory" module="OCR">
  -<Header>
    <DocumentType>Invoice</DocumentType>
    <DocumentSubType></DocumentSubType>
    <Description>The Factory</Description>
  </Header>
  <Identifiers/>
  -<Rearranges>
    <InvoiceNo>100147</InvoiceNo>
    <InvoiceDate>20160119</InvoiceDate>
    <TotalAmount>65731.0000</TotalAmount>
  </Rearranges>
  -<Patterns>
    -<ItemLine>
      <LineAmount>129.0000</LineAmount>
      <LineQuantity>1.0000</LineQuantity>
      <LinePrice>129.0000</LinePrice>
      <Description>Loudspeaker, Cherry, 150W</Description>
      <ArticleNo>LS-150</ArticleNo>
      <LineItemType>Pcs</LineItemType>
    </ItemLine>
    -<ItemLine>
      <LineAmount>210.0000</LineAmount>
      <LineQuantity>10.0000</LineQuantity>
      <LinePrice>21.0000</LinePrice>
      <Description>Cables for Loudspeakers</Description>
      <ArticleNo>LS-2</ArticleNo>
      <LineItemType>Box</LineItemType>
    </ItemLine>
    -<ItemLine>
      <LineAmount>79.0000</LineAmount>
      <LineQuantity>1.0000</LineQuantity>
      <LinePrice>79.0000</LinePrice>
      <Description>Loudspeaker, Cherry, 75W</Description>
      <ArticleNo>LS-75</ArticleNo>
      <LineItemType>Pcs</LineItemType>
    </ItemLine>
  </Patterns>
</LaserNet>
```

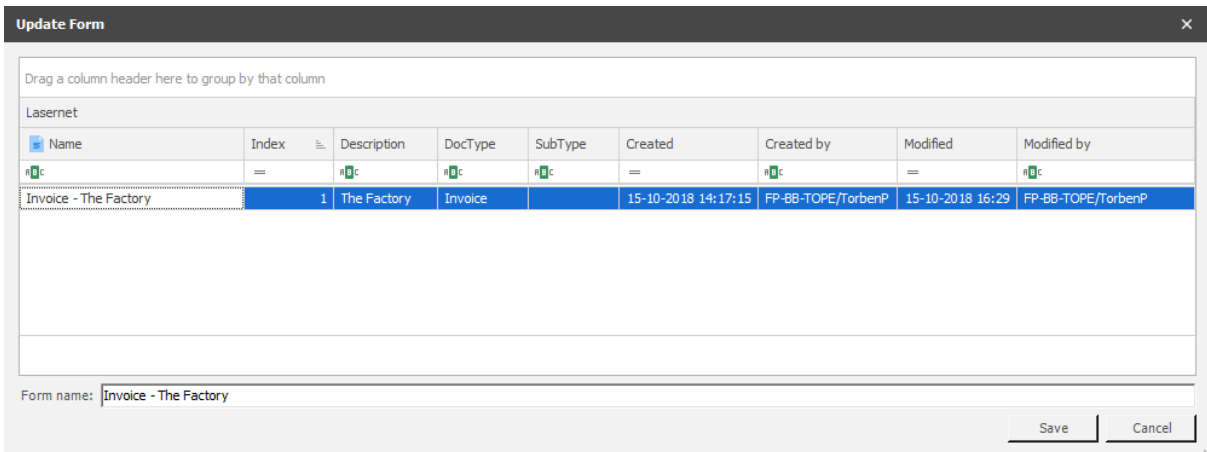
This is an example of the XML format created by the Lasernet OCR Engine. All the mapped field names and values are included in the job.

3.6.10 Update OCR Form

When all the required fields are added to the OCR Form, you are ready to deploy the OCR Form to the Lasernet OCR server.



Select the **Form** tab and click **Update**.



Update Form

Drag a column header here to group by that column

Lasernet

Name	Index	Description	DocType	SubType	Created	Created by	Modified	Modified by
#c	=	#c	#c	#c	=	#c	=	#c
Invoice - The Factory	1	The Factory	Invoice		15-10-2018 14:17:15	FP-BB-TOPE/TorbenP	15-10-2018 16:29	FP-BB-TOPE/TorbenP

Form name: Invoice - The Factory

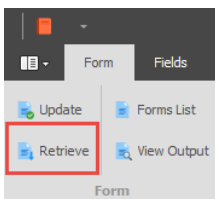
Save Cancel

Type in a logical name for your OCR Form, usually the name of the invoice supplier - then press **Save**.

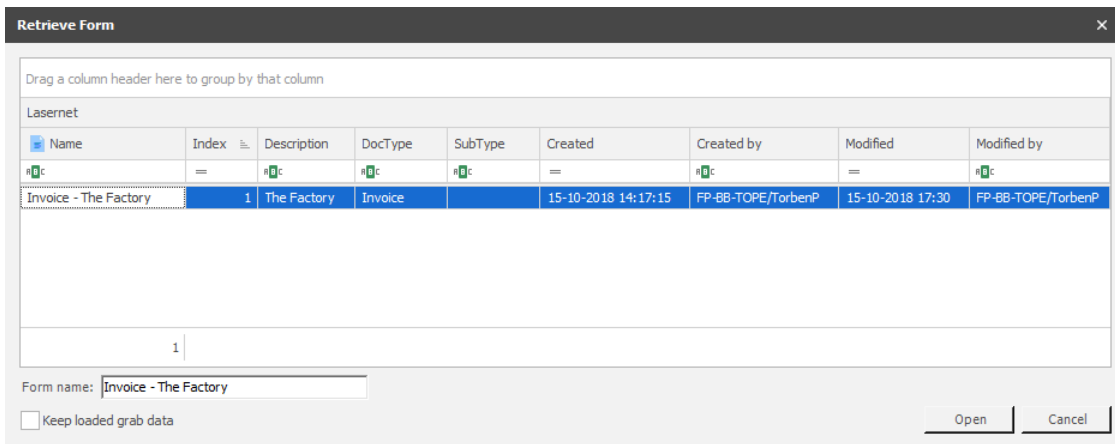
The OCR Form will be uploaded to the Lasernet server and activated. Documents received by the Lasernet server that match this document type and supplier name will be converted to an XML format that can be used by other modules in Lasernet or an external workflow system.

3.6.11 Retrieve OCR Form

You can retrieve and edit an existing OCR Form from the Lasernet OCR server.

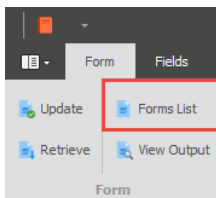


Select the **Form** tab and click **Retrieve**. Changes can now be made to the OCR Form and afterwards uploaded with the new changes to the Lasernet OCR server.

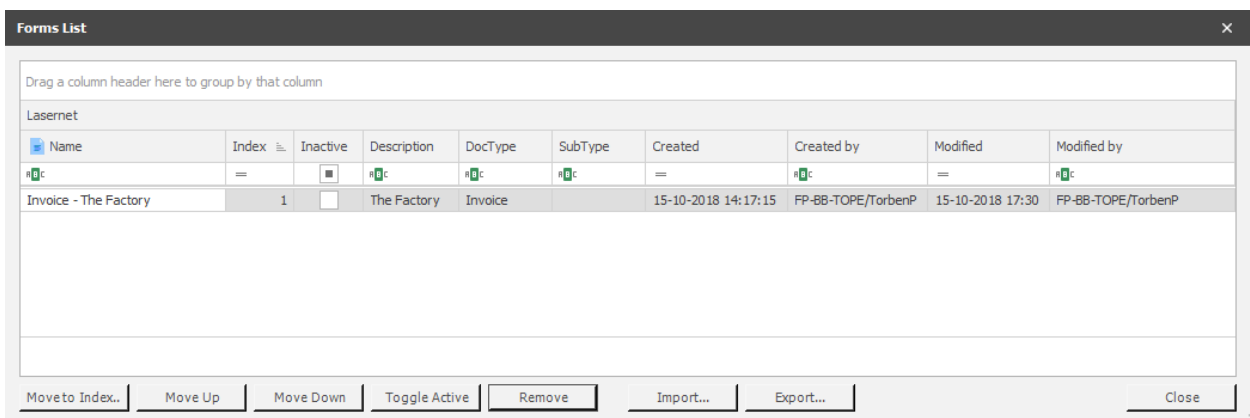


3.6.12 Forms List

Select the **Forms List** button for the following actions: view **Created/Modified By**, **Move to Index**, **Remove** discontinued OCR Forms and **Import/Export** OCR Forms between Lasernet projects.

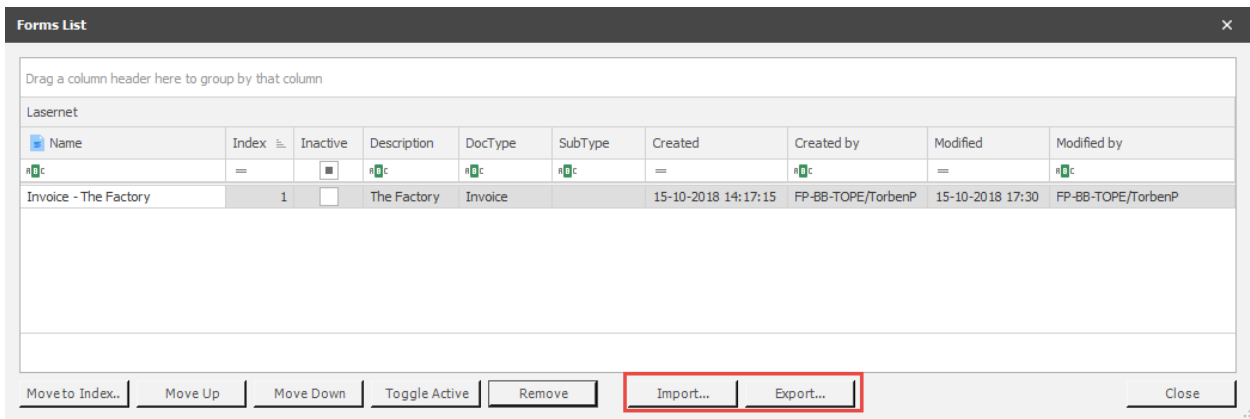


We recommend that you move regularly used OCR Forms to the lowest index number for the best performance. Documents are analyzed in index order and the first match for a form criterion will extract the OCR Fields based on the defined rules in the recognized form.



3.6.12.1 Mass input/export support

Select the Forms List dialog to **Import** or **Export** a single or multiple OCR Forms. Mark the OCR Forms in the grid and click Export. Select a temporary folder on disk to export to.



OCR Forms can be imported to other Lasernet projects. Open the Lasernet OCR application for another project and select the **Import** tool.

Browse to the folder where the OCR Forms are exported to. All the OCR Forms located in the selected folder will be imported. You have to manually remove the files from the folder, before the import action, to exclude specific OCR Forms from being imported. If the name of an imported OCR Form already exists, a new copy of the OCR Form is created. The new copy will be appended with an “_2” suffix, as shown in the next screen shot.

During import, the OCR Forms will be immediately deployed to the Lasetnet server. It is not necessary to manually run the Update functionality afterwards.

3.7 Dictionary Service

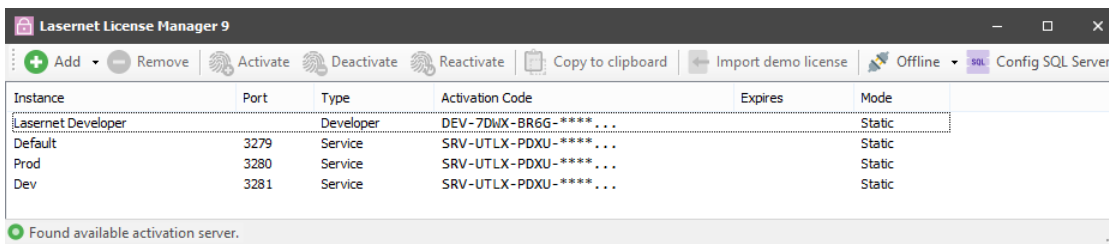
In the previous chapters we covered how to create OCR Fields with the built-in OCR tools and wizard to create Form Criterion, OCR Fields and Data Fields.

The solution also provides a Dictionary tool to automatically create objects with known phrases in the OCR Form. The solution requires a central database with a list of terms and regular expressions to recognize labels and extract data from documents.

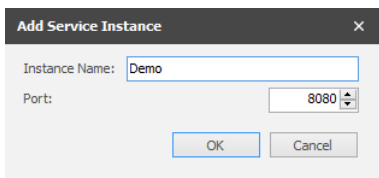
3.7.1 Install Lasernet Dictionary Service

A Dictionary that contains languages, words, terms and recognition rules is required.

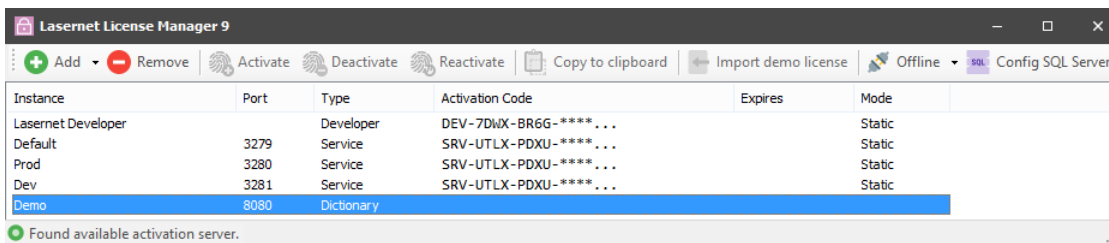
To install the **Lasernet Dictionary Service** you must start the **Lasernet License Manager**.



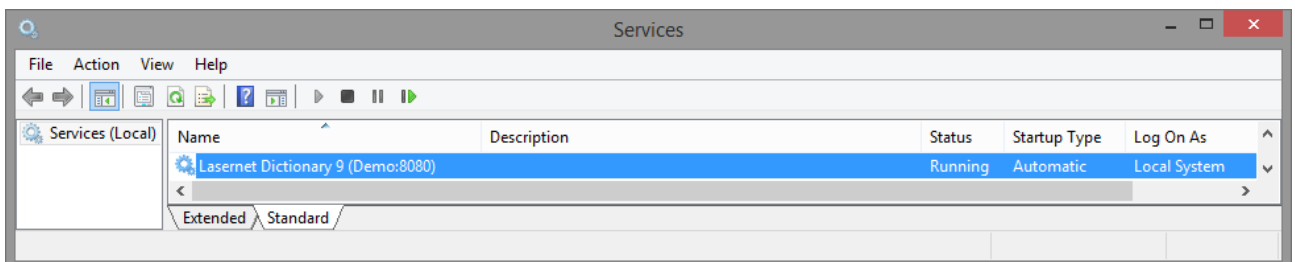
Select **Add** → **Dictionary Service** enter an **Instance Name** and a listen **Port**.



A new **Dictionary** will be created with the selected **Instance Name**. This instance does not require an additional license.



A new Windows service named **Lasernet Dictionary 8 (Instance:Port)** will now be running.

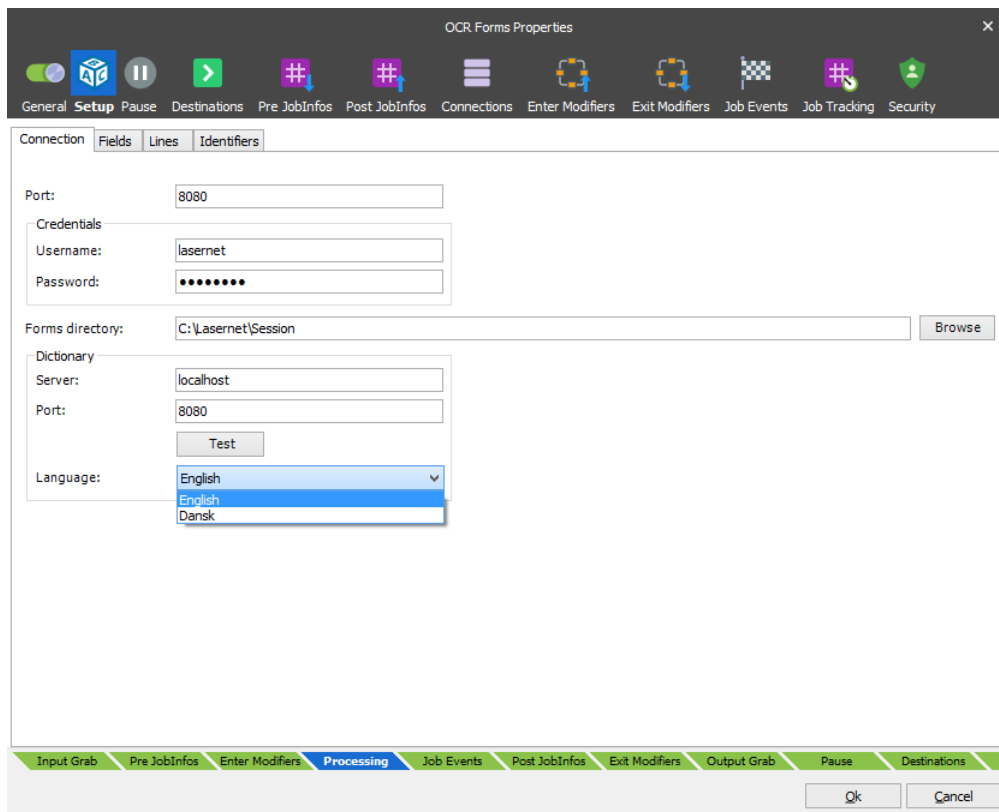


The dictionary service connects to a SQL Compact database used as a backend to store the dictionary rules. The database is located on the Lasernet Server in the following location:

C:\ProgramData\Formpipe Software\Lasernet 9\[Demo]\Databases\Dictionary.sdf

3.7.2 Connect to Dictionary Server

In the OCR Engine you must configure how to connect to the Dictionary Server.



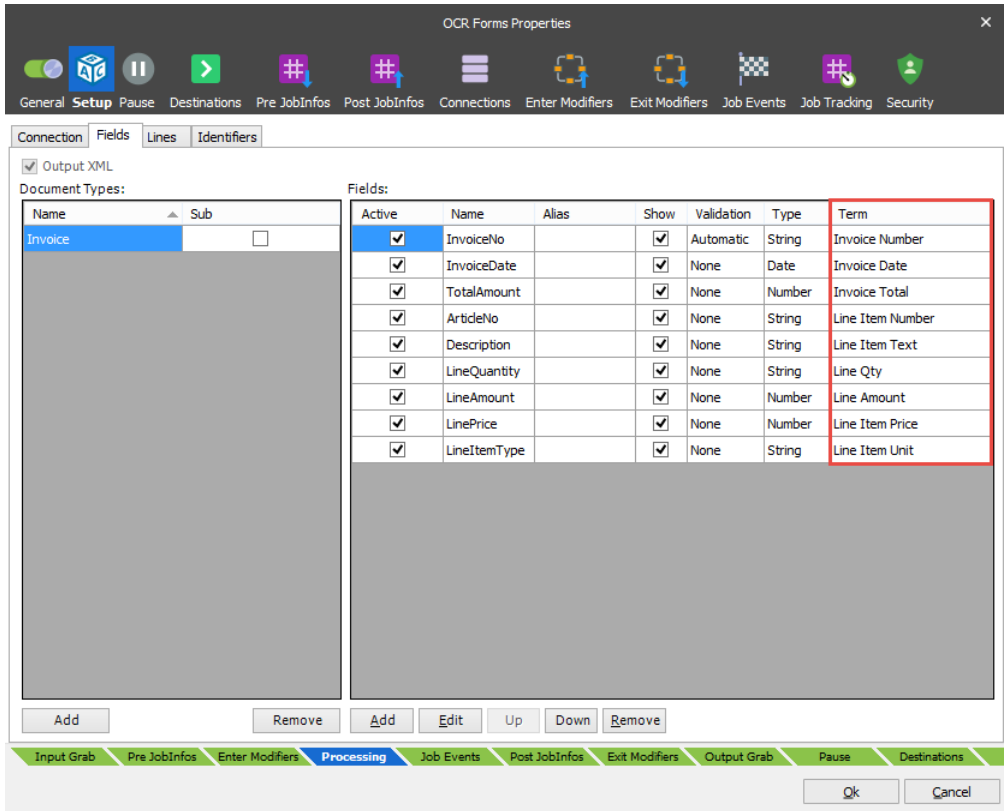
Dictionary server Name or IP address for server running the Lasernet Dictionary Service.

Dictionary port Port number for Lasernet Dictionary Service to listen on.

Language Default language for the dictionary in the OCR Designer. By default, the solution is delivered with an English and Danish dictionary. Other languages must be maintained by end-user.

Words or phrases stored in the dictionary are grouped in to Terms.

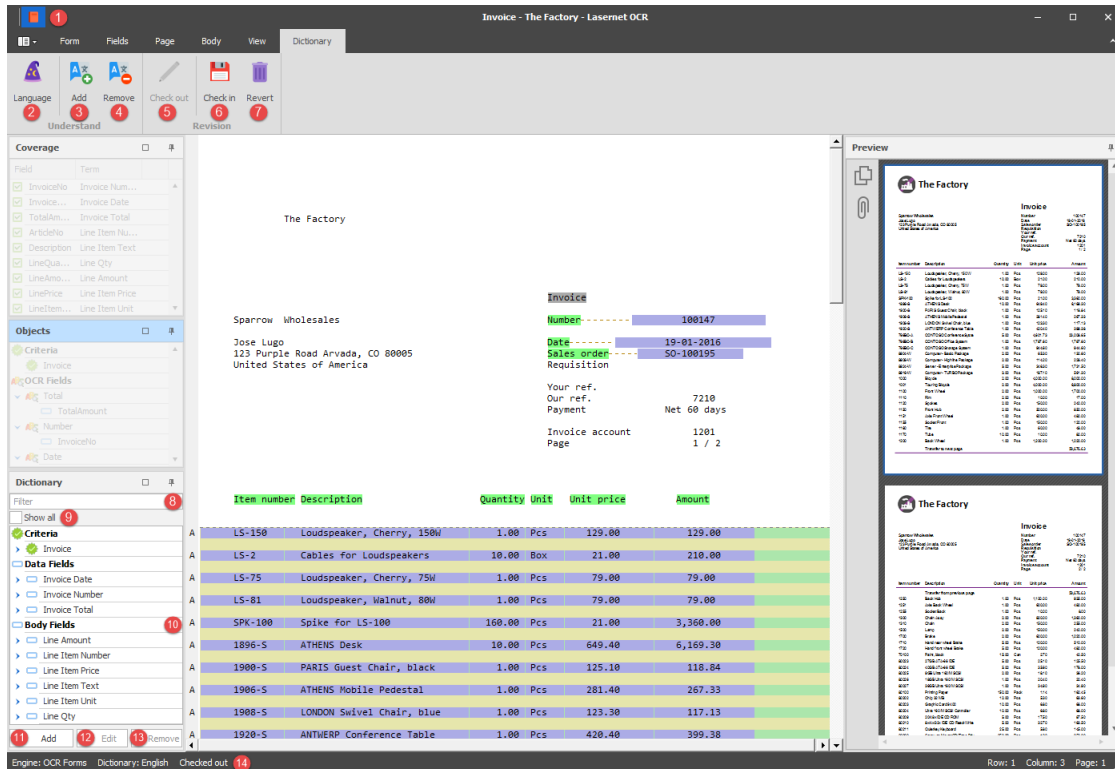
The Term column is used to connect an OCR Field with the name of a term in the dictionary. A Term consists of one or several Aliases (words that appear in job data). When the dictionary understands a phrase, it will be automatically assigned to the OCR Field. The association between a Term and an OCR Field is managed in the Fields tab.



By default the solution has an English and Danish dictionary, with labels and phrases used in invoices and credit notes. All the default terms are available in English only, regardless of the language. New Terms must be created in the Lasernet OCR and checked into the database before they can be associated to an OCR Field in the OCR Engine module.

3.7.3 Manage Dictionary

In Lasernet OCR you can manage the **Dictionary** by clicking the Dictionary icon (1) in the top of the application. In this mode you can only manage the Dictionary. All settings and functions used to manage OCR Forms, are disabled.



Select which **Language** you want to manage words for in the Dictionary. Before you can add or manage entries to the selected language you must **Check out** (5) a private copy of the current version. Click **Check in** (6) to make your changes available to other users of the Dictionary. The selected Language and Check out status are available in the Status Bar (14).

3.7.3.1 Tool bar

Language

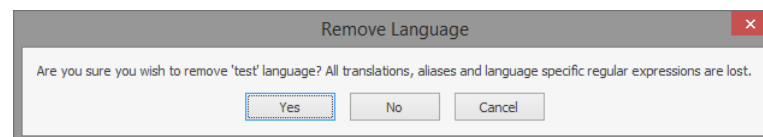
Select which **Language** (2) you want to manage words for in the Dictionary.

Add

Enabled in Check out mode. Click **Add** (3) to add a new Language to the Dictionary.

Remove

Enabled in Check out mode. Click **Remove** (4) to remove a Language and all contents in the Dictionary for the selected Language. A warning will appear, before all translations, aliases and language specific regular expressions are lost.



Check out

Click **Check out** (5) to prepare a new revision of the Dictionary. A private copy of the Dictionary will be created. Changes made to the Dictionary in Check out mode will not be available for other users after a Check in.

Check in

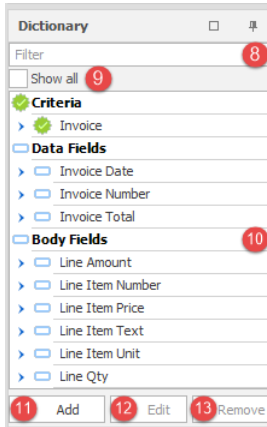
Enabled in Check out mode. Click **Check in** (6) to update the Dictionary and deploy the modifications to other users of the dictionary.

Revert

Enabled in Check out mode. Click **Revert (7)** to discard all changes made to the Dictionary since last Check in.

3.7.3.2 Dictionary

You can View, Add, Edit and Remove Terms from the Dictionary in the tree view in the Dictionary window.



Filter

Type a string to **Filter (8)** the contents of the Dictionary by.

Show all

Enable **Show all (9)** to view all existing entries in the Dictionary for the selected language. Disable to only show the entries in the Dictionary that exist in the OCR Form that you are currently working on.

Dictionary

The list of Criteria, Data Fields and Body Fields available in the **Dictionary (10)** for the selected Language.

Add

Click **Add (11)** to manually add a new term to the Dictionary. Type the name in **English** for the term and a **Description** to describe the usage of the term.

A screenshot of the 'Add Term' dialog box. It has a title bar with 'Add Term' and a close button. There are two text input fields: 'English:' and 'Description:'. Below these is a 'Type:' section with three radio buttons: 'Criterion' (selected), 'Column', and 'Field'. At the bottom are 'OK' and 'Cancel' buttons.

If you have selected different language than English, an additional field for the selected language will be visible. In the field you can type in a translation for the language you are currently working with.

In dialogs, where the Term is present, the translated Term will be shown. The Term is bound to its English name and this will be the primary key in the Dictionary.

Edit

Select a Criterion, a Data Field or a Body Field in the Dictionary. Click **Edit (10)** to change the contents of an existing Term.

A screenshot of the 'Edit Term' dialog box. It has a title bar with 'Edit Term'. There are four main sections: 'Name:' with a text box containing 'Invoice'; 'Description:' with a text box containing 'Document Type'; 'Type:' with three radio buttons: 'Criterion' (selected), 'Column', and 'Field'; and 'Aliases (Case insensitive):' with a text box containing 'Invoice' and 'Credit Note' on separate lines.

Edit Term for a Criterion

Edit Term

Name: Invoice Date

Description:

Type: Criterion Column Field

Data Type: String Number Date

Areas: Left Below Right

Aliases (Case insensitive):
 Invoice Date
 Credit Note Date
 Date

Data recognition (All languages):
 (\d{2,4}[-./]\d{2}[-./]\d{2,4})|(\d{1,2}\.\s+[A-Za-z]{3,})\.\s+\c
 ^\d\d(\x20|-)[A-Z][a-z](\x20|-)\d\d\d\d\$

Add Edit Move Up Move Down Remove

Data recognition (For English):

Add Edit Move Up Move Down Remove

Dependencies (Boolean OR):
 Invoice

OK

Edit Term for a Data Field

Remove

Select a Criterion, Data Field or Body Field in the Dictionary and click **Remove (11)** to remove all aliases and language specific regular expressions for the selected Term. A warning will appear, before all translations, aliases and language specific regular expressions are lost.

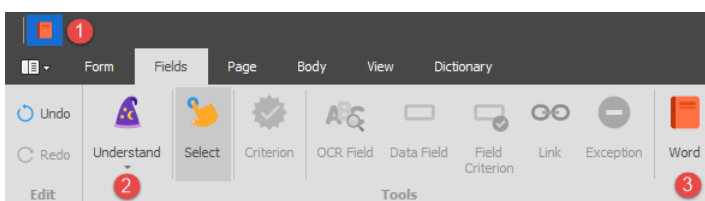
Remove Term X

Are you sure you wish to remove term 'Invoice Number'?

Yes No Cancel

3.7.3.3 Understand tool

In Dictionary (1) mode you will find tools to manage the Dictionary.



Click the Understand (2) tool and any Alias, connected to a Term that matches a string in the loaded document, will be highlighted in green boxes.

The screenshot shows the Lighthouse OCR interface. The main window displays an invoice from 'The Factory' to 'Sparrow Wholesales'. Several terms are highlighted in green boxes: 'Invoice', 'Number', 'Date', 'Sales order', and 'Requisition'. The 'Preview' pane on the right shows a thumbnail of the invoice. Below the document, a table lists the invoice items with their descriptions, quantities, units, unit prices, and amounts.

Item number	Description	Quantity	Unit	Unit price	Amount
A LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
A LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00
A LS-75	Loudspeaker, Cherry, 75W	1.00	Pcs	79.00	79.00
A LS-81	Loudspeaker, Walnut, 80W	1.00	Pcs	79.00	79.00
A SPK-100	Spike for LS-100	160.00	Pcs	21.00	3,360.00
A 1896-S	ATHENS Desk	10.00	Pcs	649.40	6,169.30
A 1900-S	PARIS Guest Chair, black	1.00	Pcs	125.18	118.84
A 1906-S	ATHENS Mobile Pedestal	1.00	Pcs	281.40	267.33
A 1908-S	LONDON Swivel Chair, blue	1.00	Pcs	123.30	117.13
A 1920-S	ANTWERP Conference Table	1.00	Pcs	420.40	399.38

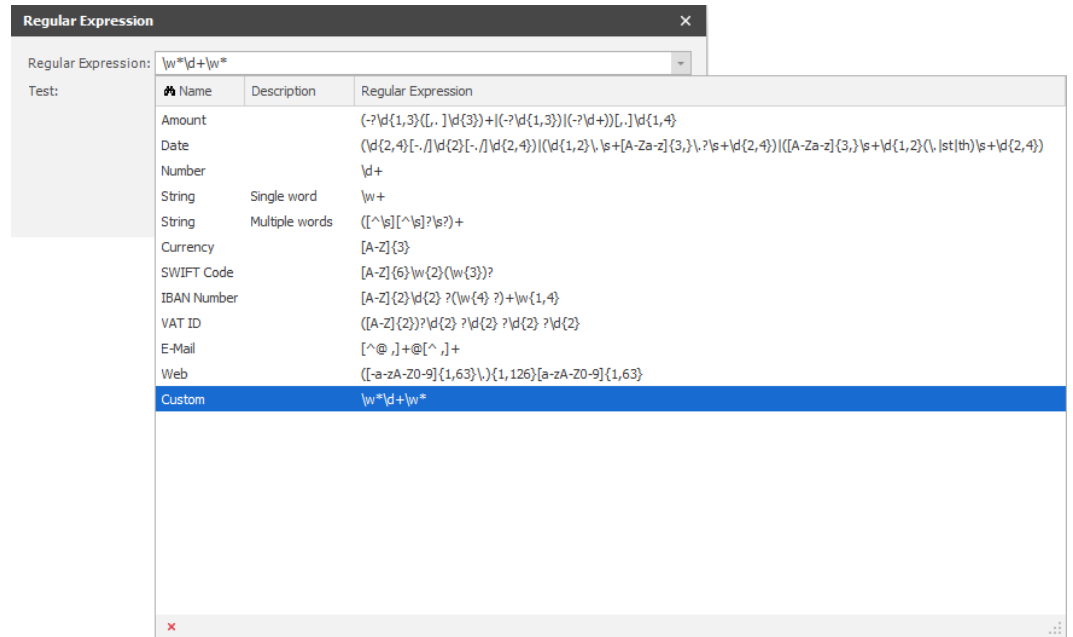
Aliases connected to Terms must be present in the loaded document as well as a Data Recognition (regular expression) in a specified area relative to an Alias. A Data recognition can be set for all languages or a specific language.

Click on any of the found Terms (green boxes) to view the location of area(s) where data recognition is running for the selected object. Drag'n'drop one of the anchor points to increase or decrease the size of the area for data recognition.

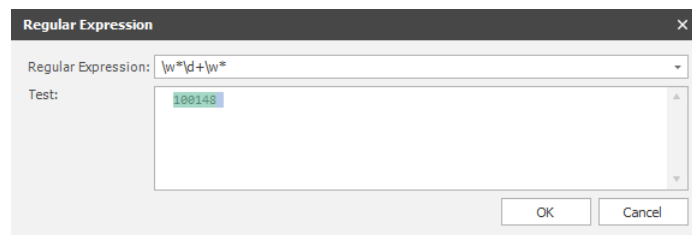
This close-up shows a green box around the term 'Invoice' in the document. A red hatched area below it represents the data recognition region. The text 'Invoice' is highlighted, and the data recognition area is positioned to capture the 'Invoice' field in the document's header.

Double click on the green item to edit the properties of the Term.

- Name** **Name** of Term that is stored in the Dictionary.
- Description** User-defined **Description** of Term.
- Type** **Criterion** defines that the Term is used to create a Criterion object that can recognize an OCR Form.
Column defines that the Term is used to automatically create OCR and Data Fields in the body of the OCR Form.
Field defines that the Term is used to automatically create OCR and Data Fields in the header and footer of the OCR Form.
- Data type** Selected **Data type** will automatically be assigned to the OCR Field.
- Aliases** One of the **Aliases** in the list must appear in an OCR Form to bind a Term to an OCR Field. The text strings are case insensitive and the first found string will bind the Term to an OCR Field if Data recognition is true.
- Data recognition** Regular expressions are used for **Data recognition** either for **all languages** or a **specific language**. The field has a list of pre-defined regular expressions to help you set up data recognition for common field types like amounts, dates, numbers, single words, multiple words, currency, SWIFT code, IBAN Number, VAT ID, E-mail and Web. You can customize your own regular expression(s) to handle other types of data recognition.



You can always test various combination of text strings to validate a regular expression. Strings marked with a green background show that the test string has been successfully recognised.



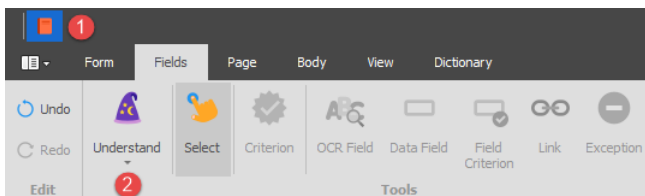
Dependencies

For the most accurate recognition of a Term you can set up **Dependencies** to specific document type(s). This is often useful if you have an OCR solution running for several Document types and you want to ensure that the Term is only found on a specific type of document.

If you want to add additional words to the Dictionary you can use the Word (3) tool and mark words located in the Document.

3.7.4 Create OCR Forms with Dictionary

Turn off the Dictionary (1) mode and tools used to manage the Dictionary will be disabled.



You can always switch to Dictionary (1) mode and add new Terms and Aliases to the Dictionary. Existing objects added to the OCR Form won't be lost.

Click the Understand (2) tool and any Alias, connected to a Term that matches a string in the loaded document, will be created as Criteria, OCR Fields and Data Fields.

The screenshot displays the 'Invoice - The Factory - Lasetnet OCR' application. The main window is titled 'Invoice - The Factory - Lasetnet OCR'. The interface includes a top menu bar with 'Form', 'Fields', 'Page', 'Body', 'View', and 'Dictionary'. Below the menu is a toolbar with various icons for editing and viewing. The central area shows the extracted invoice data, including the company name 'The Factory', address 'Sparrow Wholesales', and a table of items. The 'Coverage' pane on the left lists fields like 'InvoiceNo', 'InvoiceDate', and 'TotalAmount' with their corresponding OCR fields. The 'Objects' pane shows criteria and data fields. The 'Preview' pane on the right shows a visual representation of the invoice, including a table of items and a total amount.

Item number	Description	Quantity	Unit	Unit price	Amount
A LS-150	Loudspeaker, Cherry, 150W	1.00	Pcs	129.00	129.00
A LS-2	Cables for Loudspeakers	10.00	Box	21.00	210.00
A LS-75	Loudspeaker, Cherry, 75W	1.00	Pcs	79.00	79.00
A LS-81	Loudspeaker, Walnut, 80W	1.00	Pcs	79.00	79.00
A SPK-100	Spike for LS-100	160.00	Pcs	21.00	3,360.00
A 1896-S	ATHENS Desk	10.00	Pcs	649.40	6,494.00
A 1908-S	PARIS Guest Chair, black	1.00	Pcs	125.10	125.10
A 1906-S	ATHENS Mobile Pedestal	1.00	Pcs	267.33	267.33
A 1908-S	LONDON Swivel Chair, blue	1.00	Pcs	123.30	123.30
A 1928-S	ANTWERP Conference Table	1.00	Pcs	420.40	420.40

The Coverage window will show which fields have been automatically covered using the rules in the Dictionary.

The Objects window will show the property values for each object more in detail. If there is data in the document that hasn't been found, you can add the OCR Fields and Data Fields manually to finish the data extraction.

It is important to ensure that the objects which are automatically found are connected to the expected OCR and Data Fields and have the expected formatting properties. If the data is not extracted as expected, you must either manually correct the OCR Form objects or switch to Dictionary mode to optimize the contents of the Dictionary.

We recommend that you work with a Check In version of the Dictionary to ensure you are using the latest revision.

3.8 Convert OCR Engine XML to your own workflow format

XML formats created by the OCR Engine are produced to a standard format and are not specific to any workflow system.

It is possible to convert the XML format, created by the Lasernet OCR Engine, to one that is supported by your workflow system. The configuration and design can be done in the Lasetnet Form Editor or Lasetnet XML Transformer and processed by the Lasetnet Form Engine. Please ensure that you have the required XML Input and Output licenses for Lasetnet.

For more information on setting up XML forms in the Form Editor please read the Lasetnet 9 – Form Editor manual