

Formpipe. Lasernet



Lasernet 9 – OCR [Revision 3 – May 2019]
© 2019 Formpipe Software
Lasernet is a trademark of Formpipe Software
Company website <u>www.formpipe.com</u>
Product website <u>www.lasernetbyformpipe.com</u>

<u>1</u>	INTRODU	ICTION	1
1.1	. Wно S	HOULD USE THIS GUIDE	2
2	TERMS O	FUSE	3
_			
3	LASERNE	T OCR	4
_			
3.1	INTROD	UCTION	5
	3.1.1	ABBYY RECOGNITION SERVER FOR LASERNET	-
3.2	GETTIN	g Started - Server	7
-	3.2.1	CONFIGURING OCR ENGINE	
	3.2.2	CONNECTION	
	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	LASERN	ET CLIENT	. 22
	3.3.1	CONNECTION SETTINGS	. 22
3.4	LASERN	ET OCR	.24
••••	3.4.1	CONNECTION SETTINGS	
	3.4.2	THE USER INTERFACE	
3.5	••••=	G STARTED	
5.5	3.5.1	CREATING A FORM	
	3.5.2	FORM PROPERTIES	
	3.5.2	FORM TOOLS	
	3.5.3		
	3.5.5	SELECTIONS	
	3.5.6	BODY PROPERTIES	
	3.5.7	PAGE PROPERTIES	
	3.5.8	View Properties	-
3.6		FIROUGH – HOW TO CREATE A FORM	
5.0	3.6.1	CREATE A FORM	
	3.6.2	SETTING UP FORM PROPERTIES	
	3.6.3	SETTING UP FORM PROPERTIES	
	3.6.4	OCR FIELD	-
	3.6.5	Covered Data Fields	-
	3.6.6	COVERED DATA FIELDS CREATING THE BODY AREA	
	3.6.7	DATA FIELDS IN BODY AREA	-
	3.6.8	Additional Item Lines	-
	3.6.9	VIEW OUTPUT	-
	3.6.10	UPDATE OCR FORM	-
		RETRIEVE OCR FORM	-
		Forms List	-
3.7		VARY SERVICE	
5.7	3.7.1	Install Lasernet Dictionary Service	
	5.7.1		

3.8	CONVE	RT OCR ENGINE XML TO YOUR OWN WORKFLOW FORMAT6	5
	3.7.4	CREATE OCR FORMS WITH DICTIONARY	3
	3.7.3	MANAGE DICTIONARY	8
	3.7.2	CONNECT TO DICTIONARY SERVER	6



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



3.1 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 Lasernet.

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

		OCR Forr	ns Properties				×
eo 🙀 🕕	1	# =	(]	[]) 888	# .	2
	Destinations Pre JobInfos		ns Enter Modifiers	Exit Modifiers	Job Events		Security
Connection Fields Line	es Identifiers						
Port:	8080						
Credentials							
Username:	lasernet						
Password:	•••••						
Forms directory:	C:\Lasernet\OCR Forms						Browse
Dictionary							
Server:	localhost						
Port:	8080						
	Test						
Language:	English	~					
	-						
Input Grab Pre Job	oInfos Enter Modifiers Pro	cessing Job Events	Post JobInfos 🛛 E	it Modifiers O	utput Grab	Pause	Destinations
						Qk	Cancel

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.

	use Destinations		H	faa Connactiona	L		斗 📑	none	Iob Tracking Security
nnection Fields			POSt JODI	ilos connections	Enter Modifiers	EXITING	differs Job (_vents	ob tracking security
Output XML	Lines Identifie	rs							
cument Types:			Fields:						
Name			Active	Name	Alias	Show	Validation	Туре	Term
redit Note			~	InvoiceNo		~	Automatic	String	Invoice Number
nvoice			~	InvoiceDate		~	None	Date	Invoice Date
			-	TotalAmount		~	None	Number	Invoice Total
			~	ArticleNo		~	None	String	Line Item Number
			•	Description		~	None	String	Line Item Text
			•	LineQuantity		~	None	String	Line Qty
			~	LineAmount		~	None	Number	Line Amount
			-	LinePrice		~	None	Number	Line Item Price
			-	LineItemType		~	None	String	Line Item Unit

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 Lasernet 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 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 Lasernet Client. Please notice that if this field is not activated, validation cannot be performed by the end user in the Lasernet 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 yyyymmdd.

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

					OCR Forms P	roperties					×
	0	>	#	#		:	[]	88	# .	•	
General Setup	Pause De	stinations	Pre JobInfos	Post JobInfos	Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking	Security	
Connection Fie	elds Lines	Identifier	s								
Line Types:											
Name											
ItemLine DiscountLine											
Discourraine											
Add	Up	Down	Remove								
Input Grab	Pre JobInf	os Enter	Modifiers Pro	ocessing Jo	b Events P	ost JobInfos Exi	it Modifiers 🔪 O	utput Grab	Pause	Destination	15
									Qk	Can	cel

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

				OCR Forms F	Properties					×
co 🙀 🗉		#	#		[]	[]	88	# ,	•	
General Setup Paus	e Destinations	Pre JobInfos P	ost JobInfos	Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking	Security	
	ines Identifiers	3								
Identifiers:										
Name CompanyNo			▲ Value			CS	VFile			_
CountryCode										
Add		<u>R</u> emove								
Input Grab Pre	JobInfos Enter 1	_	essing Jo	b Events P	ost JobInfos Ex	rit Modifiers 🔪 C	Output Grab	Pause	Destination	15
Inpot Grab Pre	countros cinter i	Proc		PEVEND	Ost JODINIOS EX	at mounters C	ouput orab			
								Qk	Can	icel

IdentifiersIdentifiers are used for inserting fixed field names and values in the output data. Only
characters from A-Z are recommended as valid Idenfier 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

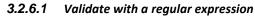
Validation is used to validate the contents of an OCR Field. Metadata not matching specified conditions will be marked either as tentative (yellow) or unsuccessful (red) in the Lasernet Client, which is used for approval of documents in the workflow.

			OCR Forms Pr	roperties						×
ඟ 🙀 🕕 🔉	#	#.			<u></u>	- <u>388</u>		# _	2	
General Setup Pause Destination		Post JobInfos	Connections	Enter Modifiers	Exit Modifie	rs Job Even	ts .		Security	
Connection Fields Lines Iden	tifiers									
✓ Output XML										
Document Types:		Fields:								
Name 🔺 Sub		Active Nam	ne	Alias		lidation Ty	pe	Term		
Credit Note			Field	4		X	1	Invoice Num		
Invoice	_		FIEIC	4		_		Invoice Date		
	Active:]					ľ	Invoice Tota		
	Name: In	voiceNo]	Ŀ	Line Item Nu		
	Alias:					(Optional)	ŀ	Line Item Te	ext	
	Type: S	tring			~	1	ŀ	Line Qty		
	Show:	- (Only applies t	o fields in hea	der/footer)		1	ſ	Line Amount	-	
							ľ	Line Item Pr		
						7	h	Line Item Ur	זונ	
	Term: Ir	voice Number			~		1			
							1			
				(ОК	Cancel	1			
Add	Remove	<u>A</u> dd	<u>E</u> dit Up	Down <u>R</u> e	emove					
Input Grab Pre JobInfos E	Enter Modifiers Pro	cessing Job	Events Po	ost JobInfos 🛛 Exi	it Modifiers	Output Grab	$\overline{}$	Pause	Destination	ns
								Qk	Car	ncel

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

Validation	×
○ None (Data will not be validated)	
Manual (Data will be marked as requiring manual validation (yellow))	
• Automatic (Data will be marked either as tentative (yellow), successful (green) or unsuccessful (red) validated based on setti	ngs below)
Result: O Tentative (yellow) Unsuccessful (red)	
Type: Regular Expression Script	
Reg.Ex.:	
Description: Optional	
Case Sensitive	
Greedy	
Extended	
Dotall	
Multiine	
Parameter: I I I I I I I I I I I I I I I I I I I	
OK	Cancel

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



Validation X
○ None (Data will not be validated)
O Manual (Data will be marked as requiring manual validation (yellow))
Automatic (Data will be marked either as tentative (yellow), successful (green) or unsuccessful (red) validated based on settings below)
Result: O Tentative (yellow) Unsuccessful (red)
Type: Regular Expression Script
Reg.Ex.: ^[0-9]{1,3}(,[0-9]{3})*([0-9]{2})?\$
Description: Optional
Case Sensitive
Greedy
Extended
Dotal
Multiline
Parameter: 999.99 P
OK Cancel

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 Lasernet Developer.

Parameter: 999.99 P

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 <u>internal</u> date or number format handled by the OCR Engine and not as the format present in the document.

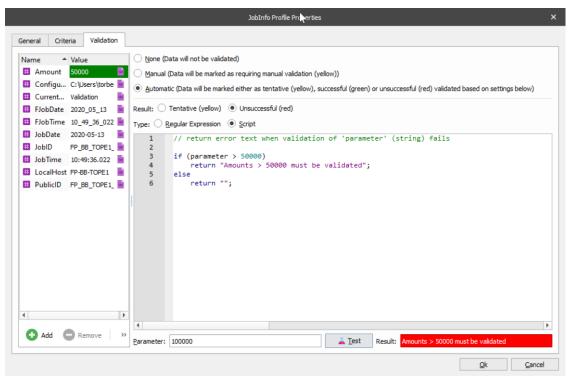
The internal format for dates is: yyyyMMdd

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

	Fields:						
Name InvoiceDate - Text	Active	Name	Alias	Show	Validation	Туре	Term
Format Date (ddMMyyyy) - Split		SupplierName		\checkmark	Automatic	String	
Required V Split String		InvoiceNo		\checkmark	None	String	Invoice Nu
incounted in a spin bound		TotalAmount			None	Number	Invoice Total
		ArticleNo			None	String	Line Item N
Date9Size: (25, 1) 19-01-2016		Description			None	String	Line Item T
Sales order SO-100195 Requisition		InvoiceDate			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



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.

Parameter:	100000	🗐 <u>T</u> est	Result:	Amounts greater than 50000 must be validated
Parameter:	10000	🔋 <u>T</u> est	t Result:	Parameter was successfully validated

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).

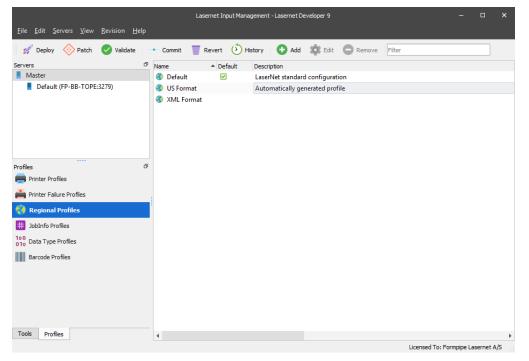
Arrays:

- 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 Lasernet Developer will be available in any OCR Forms designed in the Lasernet OCR client.



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

General Numb		×					
Name:	US Format						
Description:	Automatically generated profile						
Created date:	25-09-2018 16:32:48						
Modified date:	25-09-2018 16:32:48 25-09-2018 16:32:48						
Created by:	Unknown						
Modified by:	Automatically generated profile 25-09-2018 16:32:48 25-09-2018 16:32:48						
Form Editor:	V						
OCR Designer:	V						
Import pre-defin	ed regional profile Import						

Qk <u>C</u>ancel

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 Curro	🗙 🕔 🛗 ency Time Date			
Positive: 123,456,789.0	0 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 •			
			Ok	Cancel

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.

	US Format Properties			×
General Numbers Currency Time Date				
Show in OCR Designer: 🗸				
Format Week days Months				
Short date				
Short date example: 2018-10-11 Short date format: yyyy-MM-dd 💌				
O Long date				
Long date example: 11. oktober 2018 Long date format: d. MMMM yyyy 💌				
		Г		
			<u>O</u> k	Cancel

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 Lasernet.

File Input Input Management C:\Lasernet\Input*.pdf	 C:Lasernet\OCR Forms	 File Output A Workflow C:\Lasernet\Output
DIF to Text	0 🖽 🕈 🔒	■, ■,

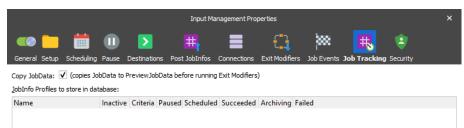
A PDF to Text modifier is used to convert PDF into a text document, which is used by the Lasernet OCR Engine to extract the contents into an XML file. The output can be stored on any output location accessible to your external workflow system.

3.2.9 Input Module

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

			Input Manageme	nt Properties				×
General Setup Schedu	Jing Pause Desti	_	obInfos Conner	tions Exit Modifier	s Job Events	H, Job Tracking	(E) Security	
Directory path:	C:\Lasernet\Input							••••
Scan mask:	*.pdf					Mask	🔵 Regular Exp	pression
Recursive:	Read files in spe	cified directory	and all subdirecto	ries.				
Regular Expression Option	ns							
Case-sensitive	Greedy Exte	nded 📃 Dota	ll 📃 Multiline					
✓ Use Polling								
Polling interval: 1 sec	ond(s)							
Multi-threaded								
Scan File Details to Jo	bInfos							
Sort								
On when file was on To ensure sorting	reated O Alphab order, polling must	· · · · ·	-					
Post Processing Deletion:	Defer deleting in	nput file until it h	nas completed pro	cessing.				
Read Headers:	Read associated	d .hdr file and ex	xtract JobInfos.					
Schedule Proce	essing Input Gra	ab 🔰 Job Eve	ents Post JobIr	nfos Exit Modifiers	Output Grab	Pause	Destinatio	ons
						Ok	Ca	ncel

Enable Job Tracking \rightarrow **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 Lasernet Client and Lasernet 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.

	PDF to Text - Lasernet Modifier Editor 9 - 🗆											×
<u>F</u> ile <u>V</u> i	ew <u>H</u> el	р										
se 📄	elect Grab	File 💾 S	ave 📕 👗	Run								
Rows:	140	*										
Columns:	200	*										
License co	de: Opti	onal										
	R	lemove hidden ti	ext									
	V R	emove invisible.	text									
	V E	nable auto align	ment									
Co	py to Clip	board as file	름 Print	K	First	»	💾 Sa	ve 📋 Copy to (Clipboard as file 🔍 Find	🔍 Find Again		»
						A	22					
L)	6	The Factory					23 24 25	Sparrow Whol	esales			
Û	-						26 27	Jose Lugo				
יש	Spanner Wh	ciersies.		nvoice	100147 19-01-2016 90-100195		28 29	123 Purple Ro United States	ad Arvada, CO S0005			
	125 Purgle R Unhed State	lead ón ada, CO 60005 L of ómerica		ales order spuid ton bur mit	7210		30 31 32	United States	of America			
			P.	symene wolce account age	Ner 60 days. 1201 1/2		33 34					
	ten numbe	- Description	Cuanty Unit	Unit price	Amount		35 36					
	19450	Loudspeaker, Chang 150V Cables for Loudsreaders	1.00 Pas 1000 Bax	129.00	129.00		37					
	19-75	Loudepasker, Chany, 75W	1.00 Pas	78.00	79.00		38 39					
	19-91 99K-100	Laudepeaker, Wahut, 90V Spile for L9-100	1.00 Pas 140.00 Pas	79.00	79.00		40					
	1996-0	ATHENS Dask PORIS GuerChair, black	1000 Pas 100 Pas	669.40 125.10	6 169.50 119.64		41					
	1900-6	ATHENG Mobile Pedagal	100 Pat 100 Pat	25.10	119.04		42					
	1909-0	LOIDOI Skielchar, Sue	100 Pat	129.90	197.19		43 44					
	1920-0 79690-0	ANTWERP Conference Table CONTOSO Conference System	1.00 Pas 2.00 Pas	420.40	200.50 23,000.65		45					
	2690-9	CONTOSO Office System	1.00 Pat	1,767.60	1,767.60		46	Iten number	Description		0	uant
	7668C-C 60064V	CONITOSO Storage Siyaan Computer - Bado Package	1.00 Pat 2.00 Pat	904.90 95.90	966.60		47 43					
	ADDR-W	Computer - Highline Package	200 Pa	114.20	228.40		48 49					
	60264W	Server - Entegrise Package Computer - TURISO Package	200 Pat 200 Pat	197.10	1,791.50 Set.90		50	LS-150	Loudspeaker, Chenry, 1500			
	1000	Noyd e	200 Pat	4,000.00	6,000.00		51					
	1001	Touring Bloyde Font Wheel	200 Pas 200 Pas	4,000.00	6,600.00 1,700.00		52					
	1110	Rm	200 Pas	10.00	7.00		53 54	L5-2	Cables for Loudspeakers			
	+120 +150	Spokes Forr Hub	200 Pas 200 Pas	150.00 200.00	240.00		55					
	1121	Jole Fight Wheel	1.00 Pat	600.00	440.00		56	LS-75	Loudspeaker, Cherry, 75W			-
	**55	SodierFrom	1.00 Pat	\$0.00	120.00		57					*
C-N	t)The Fac	tory Invoice No	100147 odf			Þ	•					•

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 Lasernet, an additional license code was required to run the PDF to Text modifier. To ensure backwards compatibity 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 Lasernet OCR Editor.

S <u>F</u> ile <u>V</u> iew <u>H</u> elp	et LnTXT Extensior	n - Lasernet Moo	difier Editor 9) — 🗆	×				
Select Grab File	Bave	👗 Run							
Command	JobInfo	Parameter	Replace	LogLevel					
Сору	Extension	.Intxt		JobInfo					
O Add	(†) Move Up	(Move Dow	n D <u>R</u> e	emove					
			Input M	lanagement Pr	operties				×
🥶 📒 🚺		\geq	#		€ }) 888	ж,	•	
General Setup Sche	eduling Pause D	estinations Po	st JobInfos	Connections	Exit Modifiers	Job Events	Job Tracking	Security	
Name	Inactive	e Run On	Resu	ult In Criteria	Modifier Type		Description		Inline
B		JobData	Job	Data	PDF to Text				
👂 PDF to Text		JUDData	1001	Juca	PDF to Tex	L			

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).

	OCR Forms Properties									
ඟ 🙀 🕕	$\mathbf{>}$	# <mark>1</mark>	#			[]) 888	# .	2	
General Setup Pause	Destinations			Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking	Security	
Connection Fields Line		-								
Tields Link	es identifiers	,								
Port:	8080			7						
Credentials										
Username:	lasernet									
Password:	•••••									
Forms directory:	C:\Lasernet\(OCR Forms							Brow	Nse
Dictionary										
Server:	localhost									
Port:	8080									
	Test									
Language:	English									
	Lingilori									
										_
Input Grab Pre Job	Dintos Enter 1	Iodifiers Pr	ocessing Jo	b Events 💦 P	Post JobInfos Ex	it Modifiers O	utput Grab	Pause	Destination	
								Qk	Can	cel

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

eneral Setup Pa		#	#			[]	- 222	# <mark>.</mark>	2
	use Destinat	ions Pre JobInfos		Connections	Enter Modifiers	Exit Modifiers	Job Events	-	Security
onnection Fields	Lines Ide	ntifiers							
Output XML		indici 3							
ocument Types:			Fields:						
Name	🔺 Sub		Active	Name	Alias	Show	Validation	Type	Term
nvoice			Image: A state of the state	InvoiceNo		~	Automatic	String	Invoice Number
			 Image: A set of the set of the	InvoiceDate		~	None	Date	Invoice Date
			 Image: A set of the set of the	TotalAmount		~	None	Number	Invoice Total
			~	ArticleNo		~	None	String	Line Item Number
			•	Description		~	None	String	Line Item Text
			v	LineQuantity		~	None	String	Line Qty
			~	LineAmount		~	None	Number	Line Amount
			v	LinePrice		~	None	Number	Line Item Price
			-	LineItemType	:	~	None	String	Line Item Unit
				- 10					
Add		Remove	<u>A</u> dd	<u>E</u> dit Up	Down R	emove			
Input Grab P	re JobInfos	Enter Modifiers Pr	rocessing Jo	b Events Po	st JobInfos 🔷 Ex	it Modifiers 🔪 O	utput Grab	Pause	Destinations
Input Grab P									

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.

OCR Forms Properties										×
••• 📦		#	#		[]	[]) 888	#	2	
General Setup	Pause Destinations	Pre JobInfos	Post JobInfos	Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking	Security	
✓ Paused										
<u>C</u> riteria for pa	using Job:									
Name	JobInfo	Operator	Case ser	nsitive Mat	ch expression					
A	OCRMatch	=		0						

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.

					OCR Forms P	Properties					×
ඟ 🎲 🕕	$\mathbf{>}$		# <mark>1</mark>	#			[]) 888	# &	۲	
General Setup Pause De	stination	s Pre Jo	obInfos Pos	t JobInfos	Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking	Security	
Destination	Inactive	Virtual	Alternative	Descriptio	n Criteria						
OCR Forms					Success: A	A=OCRMatch =	0, B=OCRValida	ated = 0 [A C	OR B]		
📒 A Workflow					Success: /	A=OCRMatch =	1, B=OCRValida	ated = 1 [A A	ND B]		

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

General Setup		OCR JobInf	o Profile Pr	roperties					×
Name	Alias		Inactive	Criteria	Show	Writeable	Validation		
FileNameWithOutExt	Document				lf data				
🖸 Add 🔯 Edit	(†) Move Up	(Move Do	wn	Remov	/e				
	<u> </u>	-							
							Qk	Cance	9

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

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

OCR Forms Properties									
ඟ 🎲 🕕	$\mathbf{\Sigma}$	#	#.		[]	[]	888	H	2
General Setup Pause De	estinations	Pre JobInfo	s Post JobInfos	Connections	Enter Modifiers	Exit Modifiers	Job Events	Job Tracking S	ecurity
Copy JobData: (copies J	obData to P	review JobDa	ta before running	Exit Modifiers)					
JobInfo Profiles to store in da		review 50000		Exit Houndry					
Name		Criteria Pa	used Scheduled	Succeeded	Archiving Failed				
OCR JobInfo Profile	indetire	<pre>citication i do</pre>							
- Certific Frome				-					

3.2.13 Modules - Output

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

						A Workfi	ow Properties							×
General Setup	Scheduling Com	bining Pause		Destinations	#	Post JobInfos	Connections	Enter Modifiers	Exit Modifiers	Data Written	Job Events	K Job Tracking	(1) Security	
Directory Path:	C:\Lasernet\(Dutput												•••
File exist:	Overwrite	Rename	🔾 Skip											
Splitter symbol:	_ 👻 Exar	nple: <filenan< td=""><td>ne>_0001.<</td><td>extension> -</td><td><filename>_99</filename></td><td>999. <extension></extension></td><td>•</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></filenan<>	ne>_0001.<	extension> -	<filename>_99</filename>	999. <extension></extension>	•							
Number of digits:	4 🗘													
Embed JobInfo	s into output file													
Write JobInf	fo tag as #JobInf	o Name=Value	#											
Use <u>C</u> odepage:	UTF-8	*												
Embed JobInfos	:													
O Add 🏘	Edit († Mov	re yp 🕕 M	love <u>d</u> own	C Remove	2									
Input Grab	Pause	Schedule	Combine	e Pre Job	Infos Enter I	Modifiers Proc	essing Dat	a Written Post :	JobInfos Exit I	Modifiers Ou	tput Grab	Job Events	Destinations	
												Qk	Cano	el

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 Lasernet Client needs to be installed on the same computer as the Lasernet OCR Editor.

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

<u>F</u> ile <u>V</u> iew <u>H</u> elp							Lasemet Client 9	
🔂 Add 🤤 Remove	Si	uccessful (F6)	🦉 R	equired (F7) 🛛 😻 Fa	ailu	re (F8)		
Servers		# Job <u>I</u> nfos) <u>L</u> in	es 🧧 Pre <u>v</u> iew 🥥) 0	pen in OCR	<u>R</u> elease as is <u>Destinations</u> C	efresh
E Default (FP-BB-TOPE:8080)	0	Add	eader he	ere to group by that colu	mn			
	*	<u>E</u> dit					OCR JobInfo Profile	
	•	Remove	us	Date	<u>=</u>	Extension	Document	
	-		used	15-10-2018 11:16:07		.Intxt	Adobe Invoice No. 0709345831	
		→ Pa	used	15-10-2018 11:16:07		.Intxt	The Factory Invoice No. 100147	

3.3.1 Connection settings

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

Options					Options 🔀
Client OCR Misc.	Clier	t OC	R N	lisc.	
Instance Name: Default		Engine:		OCR Form	
Lasernet Server: FP-BB-TOPE	Port			8080	*
Port: 8080 C Secure: (HTTPS)				Creden	tials
Default gredentials:				<u>U</u> sername	: lasernet
Credentials				Password	******
Username: admin				Test	
Password: *****				Test	
Test					
Poll interval: 60 🗘 seconds.					
OK Cancel					OK Cancel

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

<u>F</u> ile <u>V</u> iew <u>H</u> elp					Lasernet	Client 9				-		×
Add C Remove	Successful (F6) 🦉 Req	uired (F7) 🛛 🔹 F	ailure (F8)			_			_		
Servers	Iobluto:	s 🥥 Lines	📄 Pre <u>v</u> iew 🧉	Open in OCR	Release as is Destinations	🗘 Refresh	-	Preview			4	×
Default (FP-BB-TOPE:8080)	Drag a colum	n header here	to group by that col	JMN				😋 Add	Save 🔘 Remove			
	Laserne	t			OCR JobInfo Profile			¢.			,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	1.
	🧳 S	itatus D	ate	Extension	Document				🛃 The Factory			
			5-10-2018 11:16:07	Intxt	Adobe Invoice No. 0709345831			O		Invoice		
	→ P	aused 1	5-10-2018 11:16:07	.Intxt	The Factory Invoice No. 100147				Sparrow Wholesation 2006 - Logo 1915 - Fuel Anada, 500 - 2005 United Sease of America	Number Gen Seine onter Regulation Vieuron	10007 10007	
Tools										Nambar Bea Sala onter Bagulahten Tisarren, Carren, Pagmas s Innio ca account Pag	1010 Ner 60 dala 1001 110	
Paused									ten number Description LE-130 Loudgeske; Chery, 1011	Ouncity Unit Drittprice	Amount 128.00	
Queues									Lefte Loodgeaner, charg, foot Left Collect for Loodgeaner Left Loodgeaner, Charg, 75V Left Loodgeaner, Value, 80V	0.00 Pa 10800 0.00 Pa 76.0 100 Pa 76.0	10400 21000 7600 7600	
									Sherto Syle Brushoo 198-5 Shitting Deal 198-6 Pullis Guerdher, bas	1000 Par 21.00 10.00 Par 21.00 100 Par 21010	100000 4-0000 11640	
OCR Forms (2)									192-9 Provide Substituter, basis 192-9 Children Substituter, basis 192-9 Children Substituter Substituter 192-9 Children Substituter Substituter	100 Pia 10210 100 Pia 20140 100 Pia 10500 100 Pia 40040	247.55	
									NEC4 CONTOSC Conference System NEC4 CONTOSC Conference System NEC4 CONTOSC Office System	200 Par 640175 100 Par 178740 100 Par 178740	15,00445 (787.40 944.40	
									2010-W Computer - Secto Package 2010-W Computer - Highline Package 2010-W Santar - Emergina Package	200 Pm 45.8 200 Pm 114.20 200 Pm 548.20	19080 20840 179180	
									50%/W Computer - TURBO Pastage 100 Bitgele 100 Taxing Bicgele	200 Par 427.40 200 Par 4,000.00 200 Par 4,000.00	201.00 6,000.00 6,000.00	
									112 Favelited 110 Ein 112 Spins	000 Pas 1,000.00 000 Pas 10.00 000 Pas 150.00	(70000 1700 24000	
Q Search									112 Fort Hub 112 Distantional 112 Sedan Pert	100 Par 50000 100 Par 60000 100 Par 12000	480.00 480.00 400.00	
Search									1120 Tia 1170 Tula 1280 End/Minael	100 Par 40.00 10.00 Par 10.00 100 Par 10000	68.00 80.00 1,000.00	
Paused									Torubr w mitsjage		8,6563	
Scheduled								l h				
Archived									👩 The Factory			
Q Failed									Somow Wholesdes Jacion 123 Fuge Road Janeira CO HEIG	Invoice Noter Data other	1000 1000	
📄 Printers									United Street of America	Bon Sakk order Reguletion No ref. Regnant Regnant Involce account	7010 Nar60304 1001 212	
		2							tern number - Description	Page Ountry Unit Unitories	212 Amount	
***	: Page 1 o	f 1 🛛 Ei	st 🔹 Previous	D Next D	Last Page 1 🗘 👻			1		and the second)
Last updated: 15-10-2018 11:19:21												

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

🔳	÷	New Form - Lasernet OCR
]	tionary
	New	Recent Forms: Description
	<u>O</u> pen	i Identifiers
	<u>S</u> ave	Header
	Save <u>a</u> s	
	Grab File	
	Abou <u>t</u>	
		S Exit
		e Item Unit

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

		Options		×
Misc.				
Use wizard:	\checkmark			- 1
				- 1
				- 1
				- 1
				- 1
				- 1
				- 1
			ок са	ancel

3.4.1 Connection settings

Open the Optios dialog in the Lasernet OCR Editor, without using the Lasernet 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.

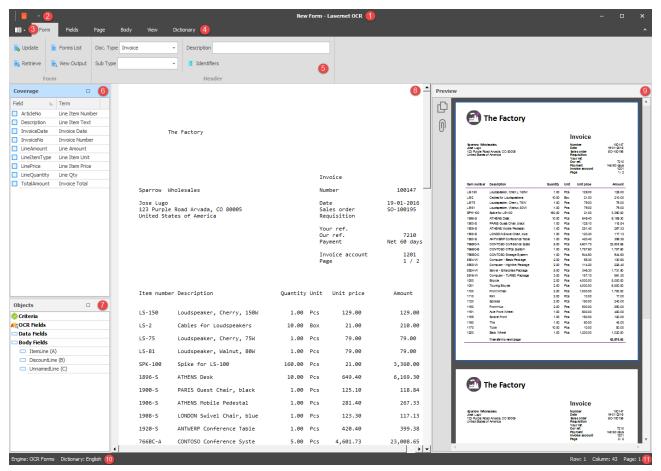
		Optior	าร	×
Client	OCR	Misc.		
OCR En <u>P</u> ort: Default	gine: gredentia			>
		Credentials Username: lasernet Password: Test		
			ОК С	ancel

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 Lasernet 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 Lasernet OCR, since the output format is already pre-defined (compared to the Lasernet 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)

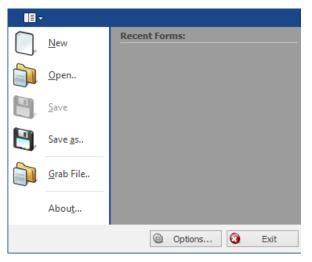
Set the Options for the Lasernet OCR Editor: edit connection and user credentials, enable/disable the wizard for creating OCR Fields.

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 recommened that you open it via the Form \rightarrow 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.

당 Update	🧐, Forms List	Doc. Type Invoice 🗸	Description	
🗟 Retrieve	🔋 View Output	Sub Type 🗸 🗸 🗸	J Identifiers	
Fo	rm 🖬		Header	5

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.

C	Coverage 🗖 म
	TotalAmount
	ArticleNo
	Description
	InvoiceDate
	InvoiceNo
	LineQuantity
	LineAmount
	SupplierName
	SupplierAddress1
	SupplierAddress2
	SupplierAddress3
	LinePrice
	LineItemType
	LineShipmentDate

Object list (7)

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

Objects	口 単
😪 Criteria	
ocr Fields	
🖓 Data Fields	
Sody Fields	
🚰 ItemLine (A)	
🚰 DiscountLine (B)	
🛐 UnnamedLine (C)	

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:

📙 Update	📕 Forms List	Doc. Type Invoice
📑 Retrieve	📘 View Output	Sub Type 🔹 📱 Identifiers
F	orm	Header

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

lame 🔺	Sub	Active	Name	Alias	Show	Validation	Туре	Term
voice			InvoiceNo		~	Automatic	String	Invoice Number
			InvoiceDate		~	None	Date	Invoice Date
			TotalAmount		~	None	Number	Invoice Total
			ArticleNo		✓	None	String	Line Item Number
			Description		~	None	String	Line Item Text
			LineQuantity		✓	None	String	Line Qty
			LineAmount		✓	None	Number	Line Amount
			LinePrice		✓	None	Number	Line Item Price
			LineItemType		✓	None	String	Line Item Unit

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

Sub TypeSub 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.

	Кеу	Value	
*	Click here to	add a new row	
\rightarrow	CompanyNo -		
	CountryCode		
- 1	4		Þ
-	4	<u>0</u> K	⊳ <u>C</u> ancel
-	4	<u>0</u> K	
	 sction Fields Lines Identifier 		
onne	· · · · · · · · · · · · · · · · · · ·		
onne	ction Fields Lines Identifier		<u>C</u> ancel
onne	ction Fields Lines Identifier		 <u>C</u> ancel

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.

CSVFile

C:\Lasernet\suppliers.csv

	Кеу	Value			L9-150 L9-2 L9-75	Loudspeaker, Cherry, 150 Cables for Loudspeakers Loudspeaker, Cherry, 75%	10.00	Box
*	Click here	e to add a r			L5-81	Loudspeaker, Walnut, 80	W 1.00	Rts
I	CompanyNo				SPK-100 1/898-6	Spike for LS-100 ATHENS Desk	160.00	
			Index	Supplier Number	Corporate Group ID	Name1	Country Coded	
			1	24	1	The Factory	DK	
			2	26	1	Microsoft	US	

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.

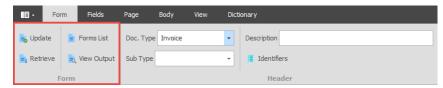
The Count value shows the total number of successfully imported records in the grid.

	Index	Supplier Number
	1841	9999999142
Þ	1842	NL60ZZZ64938
		Count: 1834

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.

)rag a column header here t	to group by that column	1						
Lasernet								
Name	Index 🗄	Description	DocType	SubType	Created	Created by	Modified	Modified by
8 8 C	=	8 <mark>8</mark> C	R C	88C	=	8 0 0	=	R B C
Invoice - The Factory			Invoice		15-10-2018 14:17:15	FP-BB-TOPE/TorbenP		

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.

Drag a column header he	re to group by	that column						
Lasernet								
Name	Index 🗉	Description	DocType	SubType	Created	Created by	Modified	Modified by
8 8 C	=	88C	в∎с	RBC	=	RBC	=	RBC
Invoice - The Factory	1		Invoice		15-10-2018 14:17:15	FP-BB-TOPE/TorbenP	15-10-2018 14:18	FP-BB-TOPE/Torben
	1							
Form name: Invoice - Th								

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.

Name Index 🖭 Inactive	Description						
	Description	DocType	SubType	Created	Created by	Modified	Modified by
):	RBC	88 C	RBC	=	8 8 0	=	8 0 0
ivoice - The Factory 1		Invoice					

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 \rightarrow 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.

Output Viewer		×
- <lasernet module="OCR" name="Invoice - The Factory"></lasernet>		~
- <header></header>		
<pre><documenttype>Invoice</documenttype></pre>		
<documentsubtype></documentsubtype>		
<description></description>		
<identifiers></identifiers>		
- <rearranges></rearranges>		
<invoiceno></invoiceno>		
<invoicedate></invoicedate>		
<totalamount></totalamount>		
<articleno></articleno>		
<description></description>		
<linequantity></linequantity>		
<lineamount></lineamount>		
<lineprice></lineprice>		
<lineitemtype></lineitemtype>		
<patterns></patterns>		
<pre><grab>DQoNCg0KDQoNCg0KDQogICAgICAgICAgICAgICBUaGUgRmFjdG9yeQ0</grab></pre>	KDQoN	Cg 👝
		~
<	>	
Save as	ОК	

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.

	Form Fi	elds	Page B	ody Vie	w Dict	ionary		
💍 Undo	2	3	-	AC		-	00	•
C Redo	Understand *	Select	Criterion	OCR Field	Data Field	Field Criterion	Link	Exception
Edit				Tool	5			

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. <i>See Link tool for creating a Data Field which is positioned relative to the OCR Field.</i> 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

Criterion

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

A Criterion is marked as a dark grey object:

Invoice

• •		Invoice - The Factory - Lasernet OCF			
B + Form Fields Page	Body View Dictionary	me Text	··· Name	- Text Invoice	Line
	For	mat String - Split	Format String	Row	Column End *
Redo Understand Select Criterio	n OCR Field Data Field Field Link Exception Rec	quired Split String	Required =	Column *	Greedy
Edit	Tools		Field		
overage 🗆 🗏			Preview		
id 🛌 Term			D		
ArticleNo Line Item Nu				The Factory	
Description Line Item Text InvoiceDate Invoice Date			0	,	
InvoiceNo Invoice Number	The Factory				Invoice
LineAmount Line Amount			10 an Jose 1 122 B	ow Wholesales Lupo Urple Road Avada, CO 80005 3 Sates of America	Number 100457 Chele onder 19-01-2016 Mesa Nillon 50-010156 Year and Our ref. 7210 Payment Net 50 days Invisce account 1201
LineItemT Line Item Unit			Unter	d Bates of America	Requisition Yournet, 7210
LinePrice Line Item Price					Pagmant Net60 days Involte account 1201 Page 1/2
LineQuan Line Qty TotalAmo Invoice Total		Envoice		number Desofption	Quantity Unit Unit price Amount
	Sparrow Wholesales	Number	100147	0 Loudspeaker, Cherry, 150W	1.00 Pos 12900 129.00
	Jose Lugo	Date	19-01-2016		10.00 Box 21.00 210.00 1.00 Pos 79.00 79.00
	123 Purple Road Arvada, CO 80005 United States of America	Sales order Requisition	50-100195 LD01	100 Boke for LB-100	1.00 Pcs 7500 75.00 160.00 Pcs 21.00 3.360.00
	oniced searces of Ancrica		1394- 1920-	S RARIS Quest Chair, black	10.00 Pos 64940 6.199.30 1.00 Pos 125.10 115.54
		Your ref. Our ref.	7210	5 LONDON Svivel Creit, blue	1.00 Pcs 201.40 207.33 1.00 Pcs 123.30 117.13
		Payment	Net 60 days	DA CONTOSO Conference Siste	1.00 Pos 420.40 399.38 5.00 Pos 4601.73 23,003.65
		Invoice account	1201 7650	DC CONTOBO Storage Byslem	1.00 Pcs 1,707.00 1,707.00 1.00 Pcs 94460 944.60
		Page	1 / 2	W Computer - Highline Peckage	2.00 Pos 6530 130.60 2.00 Pos 11420 228.40
			2224	W Computer - TURBO Package	5.00 Pcs 34630 1,721.00 3.00 Pcs 187.10 581.30
			1000	Touring Bicycle	2.00 Pcs 4000.00 8,000.00 2.00 Pcs 4000.00 6,800.00
	Item number Description	Quantity Unit Unit price	Amount 1100	Rim	2.00 Pcs 1,000.00 1,700.00 2.00 Pcs 1000 17.00
bjects 🗆 4			1120	Bookes Front Hub	2.00 Pcs 150.00 240.00 2.00 Pcs 500.00 850.00
Criteria	LS-150 Loudspeaker, Cherry, 150W	1.00 Pcs 129.00	129.00 1151	Axie Front Wheel Book e Front	1.00 Pcs 600.00 420.00 1.00 Pcs 150.00 120.00
The Factory	LS-2 Cables for Loudspeakers	10.00 Box 21.00	210.00 190	Tire Tube	1.00 Pcs 60.00 48.00 10.00 Pcs 10.00 80.00
OCR Fields	LS-75 Loudspeaker, Cherry, 75W	1.00 Pcs 79.00	79.00	Eact Wheel Transferto next page	1.00 Pes 1,000.00 1,020.00 40,678.80
Data Fields	LS-81 Loudspeaker, Walnut, 80W	1.00 Pcs 79.00	79.00		
Body Fields	SPK-100 Spike for LS-100	160.00 Pcs 21.00	3,360.00		
ItemLine (A) DiscountLine (B)	1896-5 ATHENS Desk	10.00 Pcs 649.40	6,169.30		
UnnamedLine (C)	1900-S PARIS Guest Chair, black	1.00 Pcs 125.10	118.84	The Factory	
	1906-S ATHENS Mobile Pedestal	1.00 Pcs 281.40	267.33		Invoice
	1908-S LONDON Swivel Chair, blue	1.00 Pcs 123.30	117.13 torr	ow Wholesales Lugo Unpe Road Avada, CO 2005 I Sates of America	
	1920-S ANTWERP Conference Table	1.00 Pcs 420.40	399.38	d Bates of America	Regulation Yournet 7210
	766BC-A CONTOSO Conference Syste	5.00 Pcs 4,601.73	23,008.65		Payment Net 60 days Invoice aboount 1201 Page 2/2
		4,001775			

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:

Number

		Invoice - The Factory - Lasernet OCR		
- Form Fields Page Body	View Dictionary			
	CR Field Data Field Field Link Exception	Name _ Text _ Name _ Format String _ Splt _ Format String _ Required _ Splt String _ Required _ Field	Column	Line Column End Column End
Coverage 🗆 👎			Preview	#
Field Im Term Artidelio Line Item Nu Description Line Item Text InvoiceDate Invoice Number LineAmount Line Amount LineNetric Line Item Price LineNetric Line Item Price LineQuant Line Item Price LineQuant Line Item Price LineQuant Line Item Price	The Factory Sparrow Wholesales Jose Lugo 123 Purple Road Arvada, CO 80005 United States of America	Invoice Vour ref. Vour ref. Payment Net 60 days Invoice account 1201 Page 1 / 2		
	Item number Description	Quantity Unit Unit price Amount	The Factory	
Objects III	LS-150 Loudspeaker, Cherry, 150W	1.00 Pcs 129.00 129.00	Sprew Westerla	Invoice
Criteria The Factory The Factory The View of the V	LS-2 Cables for Loudspeakers	10.00 Box 21.00 210.00	Card Sector Sect	Kanhar (2017) Data (2012) Extransfer (2012) Explain Tarrie (2017) Frynan K (2017) Frynan (2017) Frynan (2017) Frynan (2017)
ACOCR Fields	LS-75 Loudspeaker, Cherry, 75W	1.00 Pcs 79.00 79.00	browber Decipier	Cardy Dir Ditpla Anare
✓ K ² _C Number □ InvoiceNo	LS-81 Loudspeaker, Walnut, 80W	1.00 Pcs 79.00 79.00	The set of	10 Pes 1100 920 10 Pes 1100 920 10 Pes 200 420 10 Pes 200 400 20 Pes 200 1920

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		
	Invoice -	The Factory - Lasernet OCR	– 🗆 ×
- Form Fields Page Bod	y View Dictionary		*
Undo Andread Select Criterion C	CR Field Tools	Split String Required Co Field	d Une w 1Column End umn 21Greedy
Coverage 🗆 #		Prev	iew #
Field N Term Articlelo Line Item Net Description Invice Item Text InviceClate Invice Number Uneternit Uneternit Uneternit Invice Number Uneternit Uneternit Uneternit Invice Number Uneternit Invice Number Uneternit Invice Number Uneternit Invice Number Invice Number Invice Number	The Factory Sparrow Wholesales Jose Lugo 123 Purple Road Arvada, CO 80005 United States of America	Invoice Number Joid Sales order So-10015 Requisition Your ref. Our ref. Our ref. Our ref. Invoice account 1201 Page 1 / 2	
	Item number Description Quantity	/ Unit Unit price Amount	The Factory
Objects	LS-150 Loudspeaker, Cherry, 150W 1.00 LS-2 Cables for Loudspeakers 10.00	Pcs 129.00 129.00 Box 21.00 210.00	Innoise distance (Relation distance (Relation) distance (Relation) d
Invoice Invoice	LS-75 Loudspeaker, Cherry, 75W 1.00	Pcs 79.00 79.00	benunter Decision Danity Uni Uktyria Amare
V Ros Number	LS-81 Loudspeaker, Walnut, 80W 1.00	Pcs 79.00 79.00	Transformation page Ball to the back hit To Page TO Ball to the back hit CD Jans back there CD Page TO Ball to the back hit CD Jans back there CD Page Ball to the back hit Ball to the back hit Ball
InvoiceNo	CDV-100 Cnika for IC-100 160.00	Dee 21 00 3 360 00	

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 <u>not</u> 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 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 \rightarrow Output View).

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	V	LaserNet standard configuration
🌏 US Format		Automatically generated profile
🌏 XML Format		

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

String	•
Date (ddMMyy)	T
Date (ddMMyyyy)	
Date (MMddyy)	
Date (MMddyyyy)	
Date (yyMMdd)	
Date (ууууMMdd)	
String	
US Format	
Default	
XML Format	

Name

Format

Text

Split

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.

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.

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)	Name of the second field when running in split mode (works similarly to single Field Names).
		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.
		The second Field Name will contain the rest of the string.
		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.
	Format (2)	Name of the Regional Profile for the second Data Field (works similarly to single Field Names).
	Required (2)	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 the output.
	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		e for the Link tool. The tool is used for linking a Data Field with an nows which fields are connected.
	Number	100147
Exception	exception for excluding p	le for the Exception tool. The tool is used in the body area as an atterns with specific text symbology. The value of an Exception is I and is shown in the list view along with active objects in the OCR

= -				Invoice - 1	The Factor	y - Lasernet OCR					-	o x
Form Fields Page	Body View			Name		Text	··· Name		- Text		Line	,
🔿 Undo 🛛 🔏 🎽 🧃	🐼 🕺	C	9	_								
C Redo Understand Select Cri	terion OCR Field			Format String		Split	Format	t String	Row		Column End	
C Redo Understand Select Cri	Lenon OCK Held	Criterion		Required 🗸		Split String	Require	ed 🗸	Colur	mn	Greedy	
Edit	Tools							Field				
Coverage 🗆 🕂									Previe	ew		4
Field E. Term	1								-C1		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	8
ArticleNo Line Item Nu										The Factory		
Description Line Item Text									6	-	levels -	
InvoiceDate Invoice Date		The Factory							0	Sperman Windowske	Invoice	1017 10121 20121
☑ InvoiceNo Invoice Number										Carbon Residence constant	Bandar Dan Zanautor Rayandar Curret Curret Rodonation Rodonation Rodonation	
☑ LineAmount Line Amount											Popular Popular Popular	120
☑ LineItemT Line Item Unit										Innerter Decision	Danly Dir Displa	Annan
UnePrice Line Item Price										Li-St Lastipates Dary Still Li-3 Other & Lastipates Li-5 Lastipates Dary Sill	1.0 Pos 0400 19.0 Bex 0100 1.0 Pos 7400	126.00 19.00 19.00 19.00
LineQuan Line Qty					Invoi	ce				GHS Ladigate Day, SW Gar Ladigate Tata, SW SWAX Splatticks SRAX STORES	120 Post 7600 120 Post 7600 16010 Post 2100	340
 TotalAmo Invoice Total 										SEA Children SEA PURSSanChildren SEA Children	100 Pcz 7000 100 Pcz 7000 100 Pcz 1100 100 Pcz 1000 100 Pcz 1000 100 Pcz 1000 100 Pcz 1000 100 Pcz 1000 100 Pcz 1000	2000 6/000 11520 11730 11730 11730
	Sparr	ow Wholesales			Numbe	er	- 100147			984 LOCOLSAIDE NA	10 Per 2000	17.5
	Jose	Lugo			Date-		- 19-01-2016			Neo-3 Confrosofunkersalare Neo-3 Confrosofunkarsalare Neo-3 Confrosofunkarsalare	10 80 178 40	178.40
		urple Road Arvada, (0 80005			order	50-100195			Billi-H Company-Badd Package Billi-H Company-Mphile Package	100 Pes 8440 200 Pes 6520 300 Pes 7420 500 Pes 2450	96.40 12.40 28.40 179.50
	Unite	d States of America			Requi	sition				82-11 Sener-EreptusFodige 816-11 Congues TURBORubige 102 Rights 00 Turing Seals	500 Pos 9450 500 Nas 6710 500 Pos 60000 500 Pos 60000	179.50 29.30 400.00 440.00
					Your					CO TarrigBasa 10 Fen Vited	10 Pc4 0710 10 Pc4 0710 10 Pc4 00000 10 Pc4 00000 10 Pc4 1000 10 Pc4 1000 10 Pc4 5000 10 Pc4 5000	40.00
					Our r Payme		7210 Net 60 days			102 Fon Vited 110 Rei 123 Spile 123 Fon Hol	210 Pos 10000 210 Pos 1000 210 Pos 5000 210 Pos 5000	178.00 1720 140.00
					r uyine					12 Jah Frentited 15 SoderFrent 10 Th	1.00 Pea 6000 1.00 Pea 5000 1.00 Pea 6000	12.00
						ce account	1201			10 Tak 10 Tak 00 Bet (Part	1020 Pea 0000 1020 Pea 1000 1020 Pea 10200	48.00 12.00 6.00 8.00 9.00
					Page		1 / 2			- Treatmanage		9710
	Item	number Description		Quantity	Unit U	Init price	Amount			The Factory		
Objects 🗆 🕸										Theractory		
	A LS-15	0 Loudspeaker,	Cherry, 150W	1.00	Pcs	129.00	129.00			Spraw Weskelan	Invoice	18:57
• NQ Humber										Contraction of the local sector	Kuthar Data Data Data Data Variat Orrad Data Data Data Data Data Data Data Da	2023
	LS-2	Cables for Lo	oudspeakers	10.00	BOX	21.00	210.00				Bagrann Freihann ann	
Total TotalAmount	A LS-75	Loudspeaker,	Cherry, 75W	1.00	Pcs	79.00	79.00			bronte bester	Canto Uni Unicio	
V AC Date	A LS-81	Loudspeaker,	Walnut 800	1.00	Pcc	79.00	79.00			Torrador hompstellan page 122 - Back H2		947.0
InvoiceDate										CD Jin Bak (Paul 15 Sobelast 05 Dati Jay 50 Dati		48.00 600
Data Fields	A SPK-1	.00 Spike for LS-	100	160.00	Pcs	21.00	3,360.00				1.00 Pos 1000 1.00 Pos 2000 1.00 Pos 5000 1.00 Pos 6000 1.00 Pos 6000	(420) 420 (420) 1250 1250 1260 4200 40000
Body Fields	A 1896-	S ATHENS Desk		10.00	Pcs	649.40	6,169.30			10 244	10 Per 0000 10 Per 0000 10 Per 0000 10 Per 1000 10 Per 1000 10 Per 1000	5200
V ItemLine (A)										TOD Forces	1820 Can 270 1820 Can 270 280 Pen 2010	630
	A 1900-	S PARIS Guest (hair, black	1.00	Pcs	125.10	118.84			2014 0052754505 2015 0052764505 2014 105276 001505 2014 105506 001505	510 Pcs 5540 510 Pcs 1410	15.00
LineQuantity	A 1906-	S ATHENS Mobile	Pedestal	1.00	Pcs	281.40	267.33			ROS TERUTOTOM	LD Hu CODE LD Fut 500 700 LD Fut 150 700 LD Fut 1500 700 SD Fut 4500 700	2.40 3.80 19.45
											100 Pet 2000 010 Pet 200 0100 Pet 200 100 Pet 200 100 Pet 400 100 Pet 400 100 Pet 400 100 Pet 400 100 Pet 400 100 Pet 100 100	
 Description 	A 1908-	S LONDON Swive:	. cnair, blue	1.00	PCS	123.30	117.13			4100 1010-00100 101	10 Pas 120 10 Pas 120	50
ArtideNo	A 1920-	S ANTWERP Confe	rence Table	1.00	Pcs	420.40	399.38			801 Outre, Subard 820 Science NorrO Tige Dr. 989 JOR 100	2510 Pea 540 2510 Pea 100	14.00
✓ □ LineItemType	A 766BC	-A CONTOSO Confe	mence Syste	5.00	Pcc	4,601.73	23,008.65			104 children 124 Artistachicter	10 Na 8840 10 Ra 0210	122
😑 Box 🔻	4 700DC	CONTOSO CONTR	a calce bysite	5.00		-,	25,000.05	1		4	Anna ann	
Engine: OCR Forms Dictionary: English					_						ow: 1 Column: 1	06 Page: 1
angular out round - bicdonary, digisi										n	Columniti 1	oo ruger t

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.

	Item number Description
	LS-150 Loudspeaker, Cherry, 150W
	LS-2 Cables for Loudspeakers
	Form Fields Page Body View Dictionary
	Start Line * Start Text Item number Start Offset 3
	End Line * End Text Payment per End Offset -2 🔷
	Area
	1896-S ATHENS Desk
	1900-S PARIS Guest Chair, black
	Payment per 18-02-2016
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.

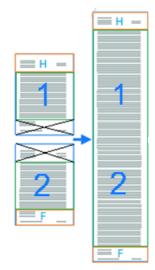
In the field **Header Size**, type in number of lines for header size; if the number of lines in header is exact the same for all pages you want to combine.

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 <u>first</u> incoming page
- the body lines for <u>all</u> incoming pages
- the footer for <u>last</u> 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 Facto	ry - Lasernet OCF	ł				-		×
📑 - Form Fields Page		Body View	Dictionary										
Header Size 0 🗘 🚺													
Footer Size 0 2													
Offsets o													
Coverage 🛛 🕂	A	1170	Tube		10.00 Pcs	10.00	80.00		_ р	review			4
Field 🛌 Term	A	1200	Back Wheel		1.00 Pcs	1,200.00	1,020.00			л. Г		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	-
ArticleNo Line Item Nu									[The Factory			
Description Line Item Text			Transfer to next page				58,676.63						
InvoiceDate Invoice Date										0	Invoice	10117 901-216 201-216	
InvoiceNo Invoice Number										Carling And	Kantar Dan Estabuter Raphtiter Tairet Garnet Raphan Indensition		
☑ LineAmount Line Amount											Pagnani Pagnani Pagn	720 No 0 4,4 121 172	
UneItemT Line Item Unit										Brounter Georgie	Carely Life Debuta	America	
UnePrice Line Item Price										Lindo Lautipaire, Dany, Schl Lind Collect & Lautipaires		12.00 210.00 70.00 70.00	
LineQuan Line Qty		Page 2								LATS Lastipates (Dary Til) LATE Lastipates (Dary Til) SPICE Spite truth (Dary SPICE Spite truth (Dary	100 Pea 7600 100 Pea 7600	N.00 7600	
☑ TotalAmo Invoice Total										SE4 ChDisber 924 RRssacht(ber	1930 Per 4640 130 Res 1990	61630 1584	
		Jose Lug	The Factory Wholesales 10 16 Road Arvada, CO 80005		Inve Numt Date	er	100147 19-01-2016 50-100195			1 - Service and the service of the		5000 1100 1100 1000 1000 1000 1000 1000	
Objects 🗆 🕸			itates of America			isition	50-100195				Invoice		
					Our Pays	ent ice account	7210 Net 60 days 1201 2 / 2			The second secon			
Body Fields										10 Herdinamidad Sala 12 Herdinamidad Sala 120 Herdinamidad Sala	120 Per 0000 120 Per 0000	00.00	
V D ItemLine (A)		Item num	ber Description		uantity Unit	Unit price	Amount			7000 Teles Intel 8005 2755-77-44 05 8004 055-77-44 05 8005 908-77-44 05	510 Pea 2510 510 Pea 3540	15.50	
LineAmount										205 95016100308 205 125016 1001208 207 245506 1001208	230 Pos 1910 120 Pos 2040 130 Dos 2441	16.00 16.40 16.40	
LineQuantity			Transfer from previous	page			58,676.63			207 SELON-GIVED 200 Primp Regr 200 Ox 210 200 SeptCastro	100 Pes 200 1000 Pask 114 1000 Pes 200	16.45	
LinePrice	A	1250	Back Hub	_	1.00 Pcs	1,100.00	935.00			425 Geptication 484 Une 404 505 Gender 429 Disk:05 (20 Des) 494 Endos (20 Des)	1020 Pea 640 1030 Pea 640	6.00 6.00	
 Description 										E259 DOLLAR DE CONTRA E210 DOLLAR DE CONTRA E211 DOLLAR VA (Stat) E220 Dollar V (Stat) Tage Str	410 Pos 1150 510 Pos 5270 2510 Pos 540	16.00	
ArticleNo	A	1251	Axle Back Wheel		1.00 Pcs	600.00	480.00			EDD See a New OTes Dr. 986 (T-D)Stee 904 NUTSSan Cret, Inc.	100 Pea 400 100 Pea 4640 100 Pea 4640	02.00 04.00 14.00	
LineItemType Box T	A		Socket Back		1.00 Pcs	10.00	8.00				A41114101		
Box V	4							•	•			71 0	

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 Fac	tory - Lasernet OCF	i.			-		;
📑 - Form Fields Page	Во	dy View	Dictionary									
Header Size 37 📜 🚹												
Footer Size 10 2												
Offsets D												
Coverage 0 #	A	1100	Front Wheel	2.00	Pcs	1,000.00	1,700.00		w			
Field is Term	A	1110	Rim	2.00	Pcs	10.00	17.00					
	-								The Factory			
-	A	1120	Spokes	2.00	Pcs	150.00	240.00		meraculy			
Description Line Item Text InvoiceDate Invoice Date	A	1150	Front Hub	2.00	Pcs	500.00	850.00	0	San San Westerne	Invoice	1217	
InvoiceDate InvoiceDate	A	1151	Axle Front Wheel	1,00	0	600.00	480,00	Ŭ	TOPOTA Part on the CO MICS Unit See of Sector	Number Das Sassorder Restort en Valvet Our ret Paytest	2023	
LineAmount Line Amount		1151	Axie Front wheel	1.00	PCS	000.00	400.00			Corrie Reprint Incoloring corri	N= 0.34	
UneItemT Line Item Unit	A	1155	Socket Front	1.00	Pcs	150.00	120.00		bourier Decisie	Page Canto Uni Dituña	1/2	
UnePrice Line Item Price	A	1160	Tire	1.00	Pre	60.00	48.00		19-50 Ladjeler Der; 5011	10 Res 1000 100 Res 1000	19.00	
UneQuan Line Oty			1210			00.00	40100		LI-S Lospeler, Day, SV LI-D Lospeler, West, SW	1.0 Pcs 7600 1.0 Pcs 7600	%.00 %.00	
TotalAmo Invoice Total	A	1170	Tube	10.00	Pcs	10.00	80.00		SNAS Statuting SSG Indiana SSG RESSACTION	1010 Pos 2100 1010 Pos 4640 100 Pos 10210	5540.00 6140.00 116.44	
	A	1200	Back Wheel	1.00	Pcs	1,200.00	1,020.00		554 LOOD Station	10 Na 2140 10 Pot 1990	17.19	
									124 Infilition Continue Tale NGC-3 CONTOSOCIENTERSEE	10 Pos 06075	525.6	
	A Per	• 2 1250	Back Hub	1.00	Pcs	1,100.00	935.00		NEOS CONTOSCOMA Spain NEOS CONTOSCE ago Spain ECHI Contos Sed Talana	10 Res (12/40 10 Res Resi0 50 Res 4220	0127.40 946.60 125.60	
	A	1251	Axle Back Wheel	1.00	Pcs	600,00	480,00		MG-11 Conguer-Highle Pacings MG-11 Serier-Direg Collecting	210 Pos 7420 210 Pos 2420	25640	
									Billi Groper 1,820fedige CD Billio Of TerryBillio	50 Per 5710 50 Per 60000 50 Per 60000	59.50 400.00 440.00	
	A	1255	Socket Back	1.00	Pcs	10.00	8.00		TO FORTHA	210 Pea 10200	(782.00 (7.00	
	A	1300	Chain Assy	2.00	Pcs	800.00	1,360.00		12 Spin 12 Familia 12 Jak Familian	2.0 Pes 1000 3.0 Pes 2000 1.0 Pes 2000	240.00 420.00 420.00	
									TE Bedarfrom	120 Pea 9000 120 Pea 9000	*20.00	
	A	1310	Chain	2.00	Pcs	150.00	255.00		110 Tale 100 Red (Pre)	10.00 Pea 1000 1.00 Pea 100000	12200	
	A	1500	Lamp	2.00	Pcs	150.00	240.00		Такатонкорр		2,750	
									8			

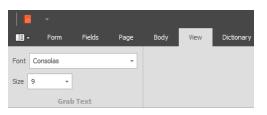
3.5.8 View Properties

Set up the preferred font type for grab text.

Font character name. Default is Consolas.

Font Size

Font character size. Default is 9.



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 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.

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

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

Start Lasernet 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 Lasernet 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 Lasernet Server. From the Lasernet Client we are now able to open the invoice saved as a temporary text file.

E Form	Fields	Page	Body View	Dictionary	Invoice - The	e Factory - Lasernet OC	R			- 0
Update	Forms List	Doc. Type	Invoice	- Description						
Retrieve	View Output	Sub Type		- Identifiers						
Form				Header						
overage								₽r	eview	
ield 🗽 Ter	rm							d	5	
ArtideNo Lin	e Item Nu								The Factory	
Description Lin	e Item Text							ſ		Invoice
	voice Date	_		The Factory				0	Same Waters Balance Taken Records CO KOS	Nation 1211 Data 201-215 States 201-215
	voice Number e Amount	_								Hindoxic (2007) Dat (2007) Dat (2007) Date (2007) Date (2007) Tatrict Date (2007) Date (20
	e Amount e Item Unit	_							Annual Annual	Danly US Displan Amark
	e Item Onic								Li-Si Ladgele: Der; SH Li-Si Ladgele: Der; SH Li-Si Ladgele: Der; SH	
LineQuan Lin						Invoice			U-S Lastpele: Der; SV Lost Lastpele: Nex, SV	100 Per 9500 1000 100 Per 9400 1000 100 Per 9400 8000 100 Per 9400 8000 100 Per 940 8000 100 Per 940 9400 100 Per 940 9400 100 Per 940 100 P
TotalAmo Inv	voice Total								SN-X Spin trides 1984 - 2740/2546 1924 - RifeStarchet tes	40.00 Per 0100 5540.00 1020 Per 4840 4740.00 100 Per 10210 11644
			Sparrow	Wholesales		Number	100147	_	054 ST4031054-bass 055 L000154404 Na 924 ST109 Colores Tale	10 Per 0210 1984 10 Per 810 2073 10 Per 9100 10735 10 Per 0200 1775
				o le Road Arvada, CO 80005 tates of America		Date Sales order Requisition	19-01-2016 SO-100195		NB04 CONTOSC enternation NB04 CONTOSC enternation NB04 Control Contos and Sam B041 Congram East Policy B041 Congram East Policy B041 Congram Faits Policy B1441 Congram TABOPalage B1441 Congram TABOPalage	100 Page 400/75 B0/844 100 Page 1078/0 1078/0 100 Page 1080 9440 100 Page 4500 12840
						Your ref. Our ref. Payment	7210 Net 60 days		00 Tarrysky 10 Den Wee 10 Re 10 Re	10 Per (0000 48000 10 Per (0000 17800 10 Per 100 1780 10 Per 500 1400 10 Per 500 4800 10 Per 500 4800
						Invoice account Page	1201 1 / 2		12 Soberforr 14 Te 17 Ta 12 East the Transfer trians page	120 Por 5000 1200 120 Por 5000 800 120 Por 1000 800 120 Por 1000 800 120 Por 1000 900 120 Por 1000 900
		_	Item num	ber Description	Quantity Ur	nit Unit price	Amount		The Factory	
bjects		4	LS-150	Loudspeaker, Cherry, 150W	1.00 Pc	s 129.00	129.00		Spritter Westerlag	Invoice Natur 1217
Criteria OCR Fields			LS-2		10.00 Bo	x 21.00	210.00		Colores Tear In the Colores Units Same of Linears	Kanthar 12.1/7 5 million 6 million 7 million 7 annel 7 annel 7 annel 7 annel 7 million 7 million
Data Fields				Cables for Loudspeakers						Barnel, 720 Papaan Ne Dina Inga 20
Body Fields		_	LS-75	Loudspeaker, Cherry, 75W	1.00 Pc	s 79.00	79.00		Innumber Deciption	Ganty Un Unpha Anazz
ItemLine (A)			LS-81	Loudspeaker, Walnut, 80W	1.00 Pc	s 79.00	79.00		Terate for protocological CE Back HZ CD and Each The I	047540 100 Pez (12000 20200 100 Pez (2000 20200 100 Pez 1000 400
DiscountLine (SPK-100	Spike for LS-100	160.00 Pc	s 21.00	3,360.00		15 Soarbox 03 Outriay 51 Out	1.00 Pca 1000 800 1.00 Pca 8000 (185.00 1.00 Pca 9000 185.00
UnnamedLine	(C)	_							02 Lang 10 East 10 Hardner dat East 12 Hardner dat East	10 Part 8000 44800 10 Ray M 1000 800 10 Ray M 1000 1000 10 Ray M 1000 800
			1896-S	ATHENS Desk	10.00 Pc		6,169.30		TO Refer the late T2 Refer the late T00 For Lat	10 Pot 0000 1000 10 Pot 1000 (800 10 Car 270 630 10 Car 270 630
			1900-S	PARIS Guest Chair, black	1.00 Pc	s 125.10	118.84		205 0755/744 02 205 055/744 02 205 02/744 02	20 Pes 02-0 -020 20 Pes 2580 -1920 20 Pes 1940 -850
			1906-S	ATHENS Mobile Pedestal	1.00 Pc	s 281.40	267.33		2014 125 the following of the following	100 Pea 2040 2040 100 Pea 2040 2040 100 Pea 110 2040
			1908-5	LONDON Swivel Chair, blue	1.00 Pc	s 123.30	117.13		EES Dig 216 EES Septicative EES Unit 101120 Cavillar	50.0 Pack 110 162.0 15.0 Pack 200 44.0 15.0 Pack 440 44.0 15.0 Pack 440 44.0
			1920-5	ANTWERP Conference Table	1.00 Pc		399.38		0 and bet Pland 0 and bet Pland 0 and and the pland	Toto Pair Base Base Base 10 Pair 100 100 100 10 Pair 100 100 100 10 Pair 100 100 100 10 Pair 100 100 1000 10 Pair 100 1000 1000
			766BC-A	CONTOSO Conference Syste	5.00 Pc	s 4,601,73	23,008.65		1024 Rokasanchi ba	10 Au 0210 1924

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 Lasernet 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.

🔳 🕒				Invoice - The Factory - Lasernet OCR
E∎ - Fo	rm Fields	Page Body Vie	ew Dictionary	
📙 Update	📄 Forms List	Doc. Type Invoice	Description The Factory	
📑 Retrieve	📘 View Output	Sub Type	- Identifiers	
F	orm		Header	

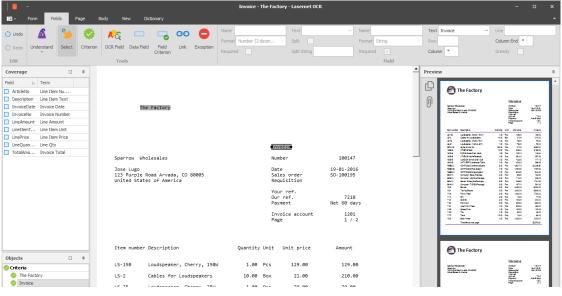
From the Doc. Type list select the proper document type. In this example we have selected **Invoice** as document type and that the contents belong to a company called **The Factory**. A list of required OCR fields are defined in the OCR Engine for the invoice and they will now appear in the Coverage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

Coverage		Щ.
Field 🖦	Term	
ArticleNo	Line Item Nu	
Description	Line Item Text	
InvoiceDate	Invoice Date	
InvoiceNo	Invoice Number	
LineAmount	Line Amount	
LineItemT	Line Item Unit	
LinePrice	Line Item Price	
LineQuan	Line Qty	
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 or select OCR Field 👻 🖂 Mark or select Data Field 👻 InvoiceNo 👻 Select Field Format 💌 🖉 Cance			_		_			
	<u>í</u>		•	InvoiceNo	•	Select Field Format	▼ Ø	Cancel

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 Line of	or Field		×
Header/Fo	oter		
Line types:	ItemLine (Code:A)		Ÿ
		ОК	Cancel

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.

	4	Rg 👘	Number	+	Mark or select Data Field		InvoiceNo	•	Select Field Format	-	0	Cancel
_	_					_						

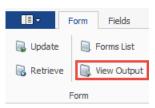
To mark the actual invoice number, click the Data Field and draw a box around the invoice number 100147.

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.

Output Viewer		×
- <lasernet module="OCR" name="Invoice - The Factory"></lasernet>		^
- <header></header>		
<pre><documenttype>Invoice</documenttype></pre>		
<documentsubtype></documentsubtype>		
<pre><description>The Factory</description></pre>		
<identifiers></identifiers>		
- <rearranges></rearranges>		
<invoiceno>100147</invoiceno>		
<invoicedate></invoicedate>		
<totalamount></totalamount>		
<articleno></articleno>		
<description></description>		
<linequantity></linequantity>		
<lineamount></lineamount>		
<lineprice></lineprice>		
<lineitemtype></lineitemtype>		
<patterns></patterns>		
<grab>DQoNCg0KDQoNCg0KDQogICAgICAgICAgICBUaGUgRmFjdG9yeQ0KDQo</grab>	NCgOKI	00
		\sim
< .	>	
Save as	ОК	

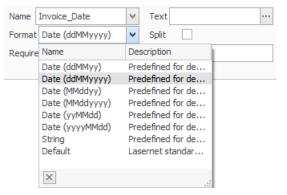
The next step is to go through the other fields from the coverage list.

	*									Invoice -	The Factory	- Lasernet O	CR		
•	Form Fie	lds P													
🖒 Undo	8	3		AG			00	0		InvoiceDate		Text	··· Nam		Ţ
C Redo	Understand	Select	Criterion		Data Field	Field Criterion	Link	Exception	Format Require		*	Split Split String		nat String	*
Edit				Tool	5									Field	1
Coverage	•		*	d.	Date	-		InvoiceDa	te	•	InvoiceDate	•	Select Field Format	4 Ø	Cancel
 Invoicel Invoicel LineAma LineIter LinePrice LineQua 		a Text Date Number Junt I Unit			The	Factory					Invoic				<u> </u>
				Jose 123	row Whol Lugo Purple Ro ed States	ad Arvad		80005			Number Date- Sales Requis	2 order	100147 Size: (13, 1) 19-01-2016 S0-100195	3	

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**.

/Bg	Fotal 🔻 📼 TotalAmount	•	TotalAm	iount 💌	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		s	ales balance	65,731.00	
				ound-off Size: otal	(20,1) 0,00 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 module="OCR" name="Invoice - The Factory"></lasernet>		~
- <header></header>		
<documenttype>Invoice</documenttype>		
<pre><documentsubtype></documentsubtype></pre>		
<pre><description>The Factory</description></pre>		
<identifiers></identifiers>		
- <rearranges></rearranges>		
<invoiceno>100147</invoiceno>		
<invoicedate>20160119</invoicedate>		
<totalamount>65731.0000</totalamount>		
<articleno></articleno>		
<description></description>		
<linequantity></linequantity>		
<lineamount></lineamount>		
<lineprice></lineprice>		
<lineitemtype></lineitemtype>		
<patterns></patterns>		
< <u>Grab</u> >DQoNCg0KDQoNCg0KDQogICAgICAgICAgICAgICBUaGUgRmFjdG9yeQ0KDQo	NCgOk	(DC
		\sim
<	>	,
Save as	ОК	

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.

Co	verage]	ф.
Fie	ld	Term		
4	InvoiceNo	Invoice Number		
4	InvoiceDate	Invoice Date		
4	TotalAmo	Invoice Total		
	ArticleNo	Line Item Nu		
	Description	Line Item Text		
	LineQuan	Line Qty		
	LineAmount	Line Amount		
	LinePrice	Line Item Price		
	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.

📾 🗎 📮		Invoice - The Factory - Lasernet OCR
Form Fields Page Body	View Dictionary	
Start Line 29 Text End Line * End Text	Start Offset 0 0	
Area		
Coverage 🗖 💷		
Field Term Invoice_Total Invoice Total		
Invoice_Number Invoice Number	The Factory	
Invoice_Date Invoice Date		
Line_Item_Number Line Item Number		
Line_Qty Line Qty		
Line_Item_Price Line Item Price		Invoice
Line_Amount Line Amount	Sparrow Wholesales	Number
Line_Item_Text Line Item Text	Jose Lugo	Date19-01-2016
Line_Item_Unit Line Item Unit	123 Purple Road Arvada, CO 80005 United States of America	Sales order SO-100195 Requisition Your ref. 7210 Dur ref. 7210 Payment Net 60 days Invoice account 1201
		Page 1 / 2
	Item number Description	Quantity Unit Unit price Amount
Objects 🗆 म	LS-150 Loudspeaker, Cherry, 150W LS-2 Cables for Loudspeakers	1.00 Pcs 129.00 129.00 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.

= -	Body View	Dictionary	Invoice - 1	The Factory	/ - Lasernet OCF	2			-	
tart Line 36 Start Text Ind Line End Text Payment per	Start Offs End Offse	et 门								
Area	1200	Lamp	2.00	PCS	150.00	240.00				
orelage	1700	Brake	2.00	Pcs	600.00	1,020.00	Previe	W		
InvoiceNo Invoice Number	1710	Hand rear wheel Brake	5.00	Pcs	120.00	510.00	C	The Factory		
InvoiceNo Invoice Number	1720	Hand front wheel Brake	5.00	Pcs	120.00	480.00	6			
TotalAmo Invoice Total							U	Sperier Walkacke Statunge Calificate Read to also 00 all 013	Invoice	10117 901-216 801-0193
ArticleNo Line Item Nu	70100	Paint, black		Can	2.70	40.50		CHERT PLANE	Ruedan Dan Istnander Raginten Valmet Der net Poplant Poplant Pop	100 Na Dias 121 112
Description Line Item Text	80023	27GB ATA-66 IDE	5.00	Pcs	25.10	125.50				
LineQuan Line Qty LineAmount Line Amount	80024	40GB ATA-66 IDE	5.00	Pcs	35.60	178.00		Innunter Deutydan 18-50 Lastgelei Deut 5011 18-1 Order friuntselen	Carely Uds Ubbyta 1.0 Pcs 1900 15.0 Eax 2100 1.0 Pcs 7800	128.00 919.00
LinePrice Line Item Price	80025	9GB Ultra 160/M SCSI	2.00	Pcs	18.10	36.20		LI-S Ladgele, Day, SV LI-B Ladgele, Wes, SV		N.00 76.00
LineItemT Line Item Unit	80026	18GB Ultra 160/M SCSI	1.00	Pcs	20,40	20,40		SRAX Spin trisks 1984 - ADROSee 1924 - RikksanChir box 1924 - ADROSerChir box	1000 Pos 2100 100 Pos 4040 100 Pos 10210 100 Pos 2010	1940.00 4440.00 196.64 547.35
	80027	36GB Ultra 160/M SCSI		Pcs	34,90	34.90		954 27-0121-stateframe 954 LOBO Shall Set ba 954 STITUT Contract Tate 956 STITUT Contract Tate 956 CONTOSCOME Spain	1.00 Pea 20140 1.00 Pea 0.000 1.00 Pea 0.000 2.00 Pea 0.0075 1.00 Pea 0.0075	17.19 20.20 19.20.45
								NBCC COPOSIDerge Span	1.00 Pea 81420	0187.60 846.60 122.60
	80100	Printing Paper	150.00	Pack	1.14	162.45		BSHI Gergen-Hythe Tariga BSHI Serar-EmphaRodya BSHI Gergen-TulkOheaga 102 Bigda	5.00 Pes 7420 5.00 Pes 5450 5.00 Pes 6710 5.00 Pes 600.00	08.40 (79.50 59.50
	80202	Chip 32 MB	12.00	Pcs	5.30	63.60		102 Bigde 00 TairigBigae 102 FontVised		400.00 440.00 170.00
	80203	Graphic Card 9400	10.00	Pcs	6.80	68.00		Biology Biology Congress Canada Congression Congressio	2.00 Pea 100.00 2.00 Pea 1000 2.00 Pea 5000 2.00 Pea 2000	14.00 14.00
	80204	Ultra 160/M SCSI Controller	10.00	Pcs	6.80	68.00		10 bis Fordinal 15 Sodarfron 10 Te 10 Te	1.00 Pes 4000 1.00 Pes 4000 1.00 Pes 4000 1.00 Pes 4000	48.00 1200 4.00 8.00
	80209	20/48× IDE CD ROM	5.00	Pcs	17.50	87.50		00 Sea their Treatmanage	(2 Na (2200	010100
	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				
jects 🗆 🕸	80220	Screw on Mount CD/Tape Driv	250.00	Pcs	1.00	250.00		The Factory		
iriteria	1896-5	ATHENS Desk	1.00		649,40	616.93		Sprite Website	Invoice	12107
The Factory								Constant Sept. In pile 00 4005	Kunthar Ben Schlander Gernel Gernel Freihanseum Traje	232
Invoice	1900-S	PARIS Guest Chair, black	1.00	Pcs	125.10	118.84			Pagnant Intelessors Pagn	101
OCR Fields								benanter Decription Transfer forspreise page	Carely Ult Ditpla	Anant BJTLO
R Number	Payment pe	er 18-02-2016		Sale	s balance	65,731.00		CE Basing CD Jobson Umal CE Adderbos	1.20 Pea (12.00 1.30 Pea 6000 1.30 Pea 1000	48.00 48.00
InvoiceNo						(20, 1) 0,00		50 Datas 50 Datas	220 Na 2000 230 Res 5000	100 (100 100 100 100 100 100
Date				Tota		65,731.00		000 Oversawy 900 Over 000 Lang 100 Bote 100 Rotes 100 Retrieve Bate 100 Retrieve Bate 100 Retrieve Bate	10 Per 0000 10 Per 0000 10 Per 0000 10 Per 0000 10 Per 1000 10 Per 1000	1220.00
Total					-			8/00 1705/T/40/DC		48.00 41.00 195.00
 TotalAmount 								201 0052714602 201 0020m-00100 202 1050a 90100	510 Pcs 5560 510 Pcs 1610 110 Pcs 2010	1%.00 #.00 20.00

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.

	ARG Mark or select	: OCR Field 🔻 🗖 LineAmount	Line	Amount 🔹	Select Field Format	O Cancel
					Number (0 decimals)	_
	Item number	r Description	Quantity Uni	t Unit price	Number (1 decimal)	
					Number (2 decimals) 3	
A	LS-150	Loudspeaker, Cherry, 150W	1.00 Pcs	129.00	Number (3 decimals)	
A	LS-2	Cables for Loudspeakers	10.00 Box	21.00	Number (4 decimals)	
A	LS-75	Loudspeaker, Cherry, 75W	1.00 Pcs	; 79.00	String	
A	LS-81	Loudspeaker, Walnut, 80W	1.00 Pcs	; 79.00	US Format	
A	SPK-100	Spike for LS-100	160.00 Pcs	21.00	Default	
A	1896-S	ATHENS Desk	10.00 Pcs	649.40	XML Format	
A	1900-S	PARIS Guest Chair, black	1.00 Pcs	125.10	110.04	
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 2	

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.

🔳 -	-									Invoi	e - The Fa	actory - Lasernet OCR		
∎-	Form Fie	lds	Page	В	ody Vie	ew Dict	ionary							
💍 Undo	2	5	-	F	AC		-	00	•	Name Format Numbe	(0 decim	- Text	··· Name	Ptring
\bigcirc Redo	Understand *	Select	Criterio	n	OCR Field	Data Field	Field Criterion	Link	Exception	Required V	(o decim	Split String	Required	
Edit					Tools	5								Field
Coverage			4											
Field	Term				Item	number D	escripti	on		Quant	ity Unit	Unit price	Amount	
Invoice			-	A 📷	LS-15	50 L	oudspeak	er, Ch	nerry, 150	W 1.	00 Pcs	129.00	129.00	
TotalAm ArticleN	no Invoice	Total		4	LS-2	С	ables fo	or Loud	Ispeakers	10.	00 Box	21.00	210.00	
Descript				4	LS-75	5 L	oudspeal	cer, Ch	nerry, 75W	1.	00 Pcs	79.00	79.00	
LineQua					LS-81	1 L	oudspeal	er. Wa	alnut, 80W	1.	00 Pcs	79.00	79.00	
🗹 LineAmo	ount Line Am	ount			6.D.V. 4								2 260 00	
LinePrice	e Line Iter	n Price		A	SPK-1	100 5	pike for	LS-10	90	160.	00 Pcs	21.00	3,360.00	
LineIter	mT Line Iter	n Unit		4	1896-	-S A	THENS De	sk		10.	00 Pcs	649.40	6,169.30	

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 - T	he Factory - Lasernet OCR		
🔳 - Form	Fields	Page	Body View Dictionary				
🔿 Undo	8 3	-	💦 🗖 🗖 🖓	Name ArticleNo	- Text	··· Name	
				Format String	- Split	Format String	
C Redo Unde	rstand Select	Criterion	OCR Field Data Field Field Link Criterion	Exception Required 🗸	Split String	Required 🗸	
Edit			Tools			Field	
Coverage		д					-
Field	Term		Item number Description	Quantity	Unit Unit price	Amount	
InvoiceNo	Invoice Nu		Size: (15, 1)				
 InvoiceDate 	Invoice Date	А	LS-150 Loudspeaker, Che	erry, 150W 1.00	Pcs 129.00	129.00	
 TotalAmount 	Invoice Total		LS-2 Cables for Louds	peakers 10.00	Box 21.00	210.00	
ArticleNo	Line Item N	<u>^</u>		peakers 10.00	21.00	210.00	
 Description 	Line Item Text	Α	LS-75 Loudspeaker, Che	erry, 75W 1.00	Pcs 79.00	79.00	
 LineQuantity 	Line Qty	А	LS-81 Loudspeaker, Wal	nut, 80W 1.00	Pcs 79.00	79,00	
 LineAmount 	Line Amount			100 1.00	/5.00	/5.00	
 LinePrice 	Line Item Price	Α	SPK-100 Spike for LS-100	160.00	Pcs 21.00	3,360.00	
✓ LineItemType	Line Item Unit	А	1896-S ATHENS Desk	10.00	Pcs 649.40	6,169.30	

Name	Description
Format	String
Required	Yes
Size	33 characters

- 📕											Invoice -	The Facto	ory - Lasernet OCR		
🔳 🕶 🛛 Form	m Fiel	ds F	age	В	ody View	Dictio	onary								
💍 Undo		9			AG			00	0	Name	Description		- Text	··· Name	
					•		e			Format	String		- Split	Format	String
C' Redo Ur	nderstand	Select	Criter	rion	OCR Field D	ata Field	Field Criterion	Link	Exception	Require	ed 🗸		Split String	Required	d 🗸
Edit					Tools										Field
Coverage			д												
Field	Term				Item n	umber De	scriptio	on			Quantity	Unit	Unit price	Amount	
✓ InvoiceNo	Invoice	Num				Size: (33, 1)								
✓ InvoiceDate	. Invoice	Date		A	LS-150		udspeak	er, Ch	erry, 150k	1	1.00	Pcs	129.00	129.00	
TotalAmoun	t Invoice	Total		A	LS-2	- (a	bles for		speakers		10.00	Box	21.00	210.00	
ArticleNo	Line Ite	m Nu		^	LJ-2	<u> </u>	0103 10				10.00	DOX	21.00	210.00	
Description	Line Ite	m Text		A	LS-75	Lo	udspeak	er, Ch	erry, 75W		1.00	Pcs	79.00	79.00	
LineQuantit	y Line Qt	у		A	LS-81	Lo	udspeak	er. Wa	lnut, 80W	-	1.00	Pcs	79.00	79.00	
LineAmount	Line An	nount													
LinePrice	Line Ite	m Price		A	SPK-10	0 🚺 Sp	ike for	LS-10	0	- 1	160.00	Pcs	21.00	3,360.00	
LineItemTyp	be Line Ite	m Unit		A	1896-5	E AT	HENS De:	sk		-	10.00	Pcs	649.40	6,169.30	

Name	LineQuantity
Format	String
Required	Yes
Size	10 characters

🗖 👻					Invoice - The Fac	tory - Lasernet OCR	
Form	Fields	Page	Body View Dictiona	ary			
🔿 Undo	<u>&</u>	-	AQ 🗆	- 00 🕒	Name LineQuantity	Text Split	
C Redo Und	erstand Select	Criterior		Field Link Exception iterion	Required 🗸	Split String	Required 🗸
Edit			Tools				Field
Coverage		д					<u> </u>
Field	Term		Item number Desc	ription	Quantity Unit	Unit price	Amount
InvoiceNo	Invoice Num				Size: (10, 1)		
 InvoiceDate 	Invoice Date	Α	LS-150 Loud	Ispeaker, Cherry, 150	1.00 Pcs	129.00	129.00
 TotalAmount 	Invoice Total	A	LS-2 Cabl	es for Loudspeakers	10.00 Box	21.00	210.00
ArticleNo	Line Item Nu						
Description	Line Item Text	Α	LS-75 Loud	Ispeaker, Cherry, 75W	1.00 Pcs	79.00	79.00
LineQuantity	Line Qty	A	LS-81 Loud	Ispeaker, Walnut, 80W	1.00 Pcs	79.00	79.00
LineAmount	Line Amount						
LinePrice	Line Item Price	A	SPK-100 Spik	e for LS-100	160.00 Pcs	21.00	3,360.00
LineItemType	Line Item Unit	A	1896-S ATHE	NS Desk	10.00 Pcs	649.40	6,169.30

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 Format Required Size	Linelter String No 7	mType Body View Dictionary	Invoice - The Factory - Lasernet OCR	
Undo Redo Understand	Select Criterio	A 🖧 🗖 🕞 🚥 🖨	Name LineItemType ▼ Text Format String ▼ Split Required ✓ Split String	Name Format String Required
Edit		Tools	but our starting	Field
Coverage	н П			
Field Term	e Num	Item number Description	Quantity Unit Unit price	Amount
_		A LS-150 Loudspeaker, Cherry, 150	Size: (7, 1) 1.00 Pcs 129.00	129.00
_	e Total			
ArticleNo Line I	tem Nu	A LS-2 Cables for Loudspeakers	10.00 Box 21.00	210.00
Description Line I	tem Text	A LS-75 Loudspeaker, Cherry, 75W	1.00 Pcs 79.00	79.00
LineQuantity Line (A LS-81 Loudspeaker, Walnut, 80W	1.00 Pcs 79.00	79.00
-	mount	A SPK-100 Spike for LS-100	160.00 Pcs 21.00	3,360.00
	tom Unit	A 1896-S ATHENS Desk	10.00 Pcs 649.40	6,169.30

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.

NameLinePriceFormatNumber (2 decimals)RequiredYesSize16 chraracters

🔳 🔹									1	(nvoice -	The Fac	tory - Lasernet OCR		
Form	Fiel	ds P	age	8	ody View	Dictionary								
🔿 Undo	2	3		2	AG 🗖		00	0	Name Li	nePrice		- Text	··· Name	
						v			Format	lumber (2	decim	- Split	Format Strin	g
C Redo Unde	erstand *	Select	Crite	erion	OCR Field Data Fie	eld Field Criterion	Link	Exception	Required	\checkmark		Split String	Required	
Edit					Tools								I	Field
Coverage			щ											_
Field	Term				Item number	Descripti	on		Q	uantity	Unit	Unit price	Amount	
InvoiceNo	Invoice	Num										Size: (16, 1)		
 InvoiceDate 	Invoice	Date		A	LS-150	Loudspeak	er, Ch	erry, 150		1.00	Pcs	129.00	129.00	
Z TotalAmount	Invoice	Total		A	LS-2	Cables fo	r Loud	cneakers		10.00	Вох	21.00	210.00	
ArticleNo	Line Ite	m Nu			25 2	cubics to	Loud	speakers		10.00	DOX	21.00	210.00	
Description	Line Ite	m Text		A	LS-75	Loudspeak	er, Ch	erry, 75W		1.00	Pcs	79.00	79.00	
LineQuantity	Line Qt	у		A	LS-81	Loudspeak	er. Wa	lnut, 80W		1.00	Pcs	79.00	79.00	
✓ LineAmount	Line Am	ount												
LinePrice	Line Ite	m Price		A	SPK-100	Spike for	LS-10	0		160.00	Pcs	21.00	3,360.00	
✓ LineItemType	Line Ite	m Unit		A	1896-S	ATHENS De	sk			10.00	Pcs	649.40	6,169.30	

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

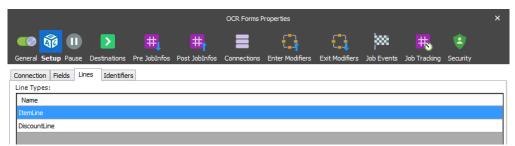
	-										I	voice - '	The Fact	ory - Lasernet	OCR			
•	Form Fie	lds	Page	Bod	y Vie	w Dict	ionary											
🔿 Undo	2	9		F I	AG			00		N	ame Lin	eAmount		- Text		Name		~
							S				ormat Nu	umber (2 d	decim	- Split		Format	String	
C' Redo	Understand *	Select	Criteri	on (OCR Field	Data Field	Field Criterion	Link	Excepti		equired	\checkmark		Split String		Required	1	
Edit					Tools												Field	
Coverage	2		д															
Field	Term				Item	number D	escripti	on			Qu	antity	Unit	Unit price	Amoun	t		
	No Invoice	Number													Size: (18, 1)	1		
✓ Invoice	Date Invoice	Date		A	LS-15	0 L	oudspeak	er, Ch	nerry, 1	150W		1.00	Pcs	129.00	129			
TotalAn	mount Invoice	Total		A	LS-2	0	ables fo	n Loud	loneaker	~ e		10.00	Вох	21.00	210	00		
ArticleN	lo Line Ite	m Nu	(<u> </u>	1.5 2		ubics it	n Louis	зреиксі			10.00	DUX	21.00				
Descrip	tion Line Ite	m Text	1	A	LS-75	i Lu	oudspeak	ter, Ch	herry, 7	75W		1.00	Pcs	79.00	79.	ð0		
🗹 LineQua	antity Line Qt	Y		A	LS-81	L	oudspeak	er, Wa	alnut, 8	BØW		1.00	Pcs	79.00	79.	ao -		
🗹 LineAm	ount Line Am	ount																
LinePrice	ce Line Ite	m Price	1	A	SPK-1	.00 Sj	pike for	LS-10	90		1	.60.00	Pcs	21.00	3,360	.00		
☑ LineIter	mT Line Ite	m Unit		A	1896-	S A	THENS De	sk				10.00	Pcs	649.40	6,169	.30		

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 UnamedLine is also present.

🗖 🤘								Ir	voice - Th	ne Factory	- Lasemet	DCR						-	o x	
III - F	Form Fiel	lds P	age	Body Vi	ew Dict	ionary														^
🔿 Undo	A	9	-	AG			30 C	Name Art	icleNo	-	Text	··· Narr	ne	~	Text		Line	ItemLine	-	
	_					v		Format St	ring	-	Split	Forr	mat String		Row 1		Colu	Name		
C' Redo	Understand *	Select	Criterior	n OCR Field	Data Field	Field Criterion	Link Exception	Required	\checkmark		Split String	Req	uired 🔽		Column	1	Gre	ItemLine		
Edit				Tool	s								Field					DiscountLin		
Coverage														- p	review			UnnamedLin	he	
Field	Term			Item	number D	escription		Qu	antity (Jnit Ur	nit price	Amount								
✓ InvoiceN	lo Invoice N	lumber		Size: (15, 1)											9	🛃 The Factory				
InvoiceD	ate Invoice D	Date	A		50 L	oudspeaker	, Cherry, 150	1	1.00 F	Pcs	129.00	129.00			N	-				
 TotalAmo 			A	LS-2	E c	ables for	Loudspeakers		10.00 E	Box	21.00	210.00			U.	Spencer Michiele 1945-00 1956-05 Foat In als CO 2005 1956-05 Foat In als CO 2005		×		
ArticleNo				LS-7	_	oudeneakon	, Cherry, 75W		1.00 F	200	79.00	79.00						Corriel Corriel Pagnans Nockang core Pagn	**************************************	
Description	ion Line Item n Line Qty		A													brounder Decision	Carrier		Ti Anner	
	unt Line Amo		A	LS-8	1 L	oudspeaker	, Walnut, 80W		1.00 0	cs	79.00	79.00				Group Garges		No. 5800	1000	
LinePrice			A	SPK-	100 S	pike for L	5-100	1	60.00 F	cs	21.00	3,360.00				Land Landgeler, Owy, 577 Landgeler, Water, 277 27670 Scherber, 270	1.0	Pea 7600 Pea 7600	Not Not	
UneItem	T Line Item	Unit		1896		THENS Desk			10.00 F		649,40	6,169.30				1044 CHARGES	1020	Pos écéco	418.20	

You must able find matches in the item lines to differentiate between an item line with our 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.

3.6.9 View output

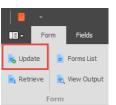
Go to Form tab and click the View Output button.

Output Viewer	_ × _
- <lasernet module="OCR" name="Invoice - The Factory"></lasernet>	~
- <header></header>	
<pre><documenttype>Invoice</documenttype></pre>	
<pre><documentsubtype></documentsubtype></pre>	
<pre><description>The Factory</description></pre>	
<identifiers></identifiers>	
- <rearranges></rearranges>	
<invoiceno>100147</invoiceno>	
<invoicedate>20160119</invoicedate>	
<totalamount>65731.0000</totalamount>	
- <patterns></patterns>	
- <itemline></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>	~
<	>
Save as	ок

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.

Drag a column header here to group by that column Lasernet	
asernet	
Name Index L Description DocType SubType Created Created by Modified Modified by	
0: - · 0: · 0: - · 0: - · 0:	
invoice - The Factory 1 The Factory Invoice 15-10-2018 14:17:15 FP-88-TOPE/TorbenP 15-10-2018 16:29 FP-88-TOPE/Torb	henP

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.

Retrieve Form X									
Drag a column header here to group by that column									
Lasernet									
📄 Name	Index 🗉	Description	DocType	SubType	Created	Created by	Modified	Modified by	
R B C	=	RBC	RBC	RBC	=	8 8 C	=	88C	
Invoice - The Factory	1	The Factory	Invoice		15-10-2018 14:17:15	FP-BB-TOPE/TorbenP	15-10-2018 17:30	FP-BB-TOPE/TorbenP	
1									
			_						
Form name: Invoice - The	Factory								
Keep loaded grab data							0	pen Cancel	

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.

pe Created Created by Modified Modified by = #@c = #@c
- 10: - 10:
15-10-2018 14:17:15 FP-BB-TOPE/TorbenP 15-10-2018 17:30 FP-BB-TOPE/TorbenP
15-10-2018 14:17:15 FP-BB-TOPE/TorbenP 15-10-2018 17:30 FP-BB-TOPE/Torb

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.

×
/TorbenP
Close
E

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. <u>All</u> 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 <u>immediately</u> deployed to the Lasernet 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.

🔂 Add 🔹 🚍 Remove	🚲 Activate	Deactivate	🔊 Reactivate 🛛 📋 Copy to clip	board 🛛 🖶 Import demo licens	e 📝 Offline	- sou Config SQL Serve
Instance	Port	Туре	Activation Code	Expires	Mode	
Lasernet Developer		Developer	DEV-7DWX-BR6G-****		Static	
Default	3279	Service	SRV-UTLX-PDXU-****		Static	
Prod	3280	Service	SRV-UTLX-PDXU-****		Static	
Dev	3281	Service	SRV-UTLX-PDXU-****		Static	

Select Add -> Dictionary Service enter an Instance Name and a listen Port.

Add Service Ins	tance		×
Instance Name:	Demo		
Port:			8080 🛨
		OK	Cancel

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

🔒 Lasernet License Manager 9						-		×
🔂 Add 👻 😑 Remove 🛛 🔊	Activate 🖇	🔏 Deactivate 🛛 🐊	Reactivate 🛛 🚔 Copy to clipboard 🗍 👉 In	nport demo license	💉 Offline	- sal Con	fig SQL S	Server
Instance	Port	Туре	Activation Code	Expires	Mode			
Lasernet Developer		Developer	DEV-7DWX-BR6G-****		Static			
Default	3279	Service	SRV-UTLX-PDXU-****		Static			
Prod	3280	Service	SRV-UTLX-PDXU-****		Static			
Dev	3281	Service	SRV-UTLX-PDXU-****		Static			
Demo	8080	Dictionary						
Found available activation server								

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

0,		Services				×			
File Action View	w Help								
🤹 Services (Local)	Name	Description	Status	Startup Type	Log On As	^			
	🝓 Lasernet Dictionary 9 (Demo:8080)		Running	Automatic	Local System	~			
	<					>			
	Extended Standard								

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

			OCR Forms F	Properties					×
General Setup Pause	Destinations Pre JobI		Connections	Enter Modifiers	Exit Modifiers	Job Events	R Job Tracking	security	
Connection Fields L	ines Identifiers								
Port:	8080								
Credentials									
Username:	lasernet								
Password:	•••••								
Forms directory:	C:\Lasernet\Session							Brow	vse
Dictionary									
Server:	localhost								
Port:	8080								
	Test								
Language:	English		~						
	English Dansk								
	Dunak								
Input Grab Pre I	lobInfos Enter Modifiers	Processing Ja	ob Events 💦 P	ost JobInfos Ex	it Modifiers 📃 O	utput Grab	Pause	Destinations	~
							Qk	Cano	el

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.

LanguageDefault 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.

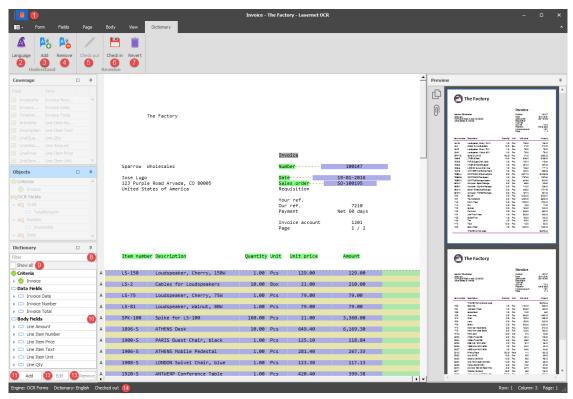
The Term column is used to connect an OCR Field with the name of a term in the dictionary. A Term consists of one or several Aliases (words that appear in job data). When the dictionary understands a phrase, it will be automatically assigned to the OCR Field. The association between a Term and an OCR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

IO 🙀 (D 돈	# ,	# <mark>†</mark>		t	t_1	, P2	8	🔣 😫	
neral Setup Pau	ise Destinations	Pre JobInfos	Post JobInfos	Connections	Enter Modifiers	Exit Modi	fiers Job Ev	ents Job	Tracking Security	
nnection Fields	Lines Identifie	rs								
Output XML										
cument Types:			Fields:						_	
Name	🔺 Sub		Active	Name	Alias	Show	Validation	Туре	Term	
nvoice			 ✓ 	InvoiceNo		✓	Automatic	String	Invoice Number	
				InvoiceDate		✓	None	Date	Invoice Date	
			✓	TotalAmount		~	None	Number	Invoice Total	
			Image:	ArticleNo		~	None	String	Line Item Number	
				Description		✓	None	String	Line Item Text	
			 Image: A start of the start of	LineQuantity		~	None	String	Line Qty	
			 Image: A start of the start of	LineAmount		~	None	Number	Line Amount	
			✓	LinePrice		-	None	Number	Line Item Price	
			✓	LineItemType		-	None	String	Line Item Unit	
				Edit Up		emove				

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.
	Remove Language
	Are you sure you wish to remove 'test' language? All translations, aliases and language specific regular expressions are lost. Yes No Cancel
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.

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

Revert

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

Dictio	narv		д
Filter			
			0
Sho	wall 🧐		
Crit	eria		
> 🗇	Invoice		
🗆 Dat	a Fields		
	Invoice Date		
	Invoice Number		
	Invoice Total		
Bod	ly Fields		10
> 🗆	Line Amount		-
	Line Item Number		
	Line Item Price		
	Line Item Text		
	Line Item Unit		
	Line Qty		
1	Add 12 Edit	13Rer	nove

Filter

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

Show allEnable 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 Langauge.

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.

Add Term					×
English:					
Description:					
_	0.01		0.01		0=11
Type:	Criteri	on	Colu	mn	Field
		(DK	C	ancel

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.

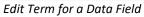
Edit

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

Name:	Invoice		
Description:	Document Type		
Type:	Criterion Column Field		
Aliases (Case insensitive):	Invoice Credit Note	Ă	

Edit Term for a Criterion

Edit Term					
Name:	Invoice Date				
Description:					
Type:	O Criterion	Column	Field		
Data Type:	String		Oate		
Areas:	Left	✓ Below	✓ Right		
Aliases (Case insensitive):	Invoice Dat Credit Note Date				×
Data recognition (All languages):			.d{2,4}) (1, a-z](\x20 -)\d	,2}\.\s+[A-Za-z] i\d\d\d\$	{3,}\.?\s+\(
	Add	Edit	Move Up	Move <u>D</u> own	<u>R</u> emove
Data recognition (For English):					
	Add	Edit	Move Up	Move <u>D</u> own	<u>R</u> emove
Dependencies (Boolean OR):	V Invoice				
				L	ОК



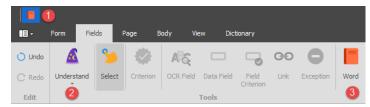
Remove

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.

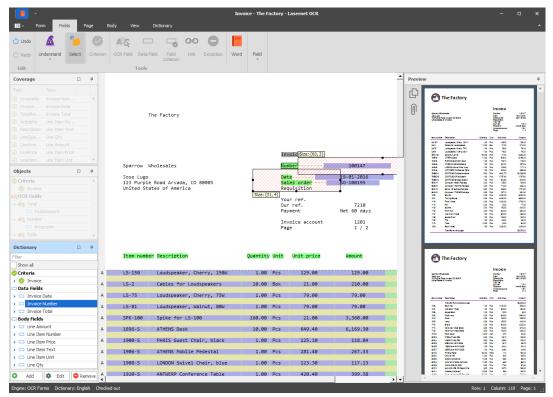
Remove Term		×			
Are you sure you wish to remove term 'Invoice Number'?					
<u>Y</u> es	No	<u>C</u> ancel			

3.7.3.3 Understand tool

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

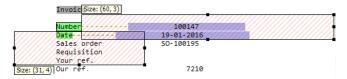


Click the Understand (2) tool and any Alias, connected to a Term that matches a string in the loaded document, will be highlighted in green boxes.



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 dicrease the size of the area for data recognition.



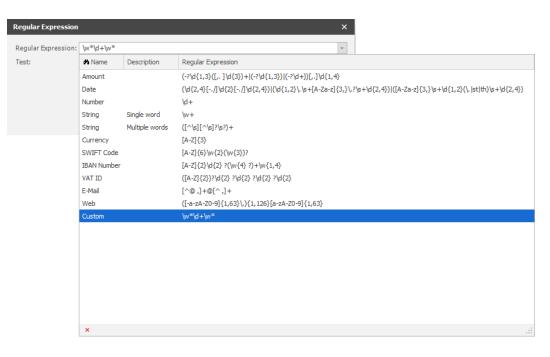
Double click on the green item to edit the properties of the Term.

Edit Term			
Name:	Invoice Number		
Description:			
Type:	Criterion OColumn I Field		
Data Type:	String Oumber O Date		
Areas:	Left 🗸 Below 🗸 Right		
Aliases (Case insensitive):	Invoice No. Invoice Number Credit Note No. Credit Note Number Number		
Data recognition (All languages):	\w*\d+\w*		
Data recognition (For English):	Add Edit Move Up Move Down Remove		
Dependencies (Boolean OR):	Add Edit Move Down Remove		
	OK		
Name	Name of Term that is stored in the Dictionary.		
Description	User-defined Description of Term.		
Гуре	Criterion defines that the Term is used to create a Criterion object that can recognize a OCR Form. Column defines that the Term is used to automatically create OCR and Data Fields in th 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 Fie		

 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

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.

Regular Expression		×
Regular Expression:	/w*/d+/w*	Ŧ
Test:	100148	*
		Ŧ
	OK Cancel	

DependenciesFor 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.

	Invoice - The Factory - Lasernet OCR	
🗄 🔹 Form 🛛 Fields Page	Body View Dictionary	
🔿 Undo 🛛 🔏		Line
C Redo Understand Select Crite	Format String - Split Format String - Row	Column End
•	Criterion Required Split String Required Column	Greedy
Edit	Tools Field	
Coverage 🗆 👎	Preview	ţ.
Field Term	The Factory	<u> </u>
InvoiceNo Invoice Number InvoiceDate Invoice Date	- Increasing	
TotalAmo Invoice Total	The Factory	Invoice
ArticleNo Line Item Nu	The Factory and the factory	Rudan (2017 Data (2012) Databate 2012/03 Rudate 2012/03 Taint 10 Part 10 Part 10 Part 10 Part 12 Part
Description Line Item Text		Tarrie, 190 Royan, Kalia, Sala, 190 Royan, Kalia, 191 Royan, 191
✓ LineQuan Line Qty	Konster Sanjin	Danly Dis Displan Arean
✓ LineAmount Line Amount	14-30 Lanager 695-500 14-30 Lanager 695-500 14-30 Lanager 695-50	1.0 Res 1940 19400 1940 Ber 2100 21940 10 Res 1940 1940 10 Res 1940 1940
LinePrice Line Item Price	Invoice Do Laberovin Pro Laberovin	
LineItemT Line Item Unit	924 Rélistador de 924	10 Pea 1070 1944
	Sparrow wholesales would be added and a second seco	100 Pete 1050 177.5 100 Pete 1050 107.5 100 Pete 1057.5 200.56 100 Pete 1057.5 200.66 100 Pete 1057.6 1057.6
	Date 10-01-2016	200 Per c20:75 83/842 10 Per c19/80 0/9/80 10 Per 8:00 8:00
		10 Res 2420 2420 10 Res 2220 1220 10 Res 1420 1220 10 Res 1420 2240 10 Res 1420 175 20 10 Res 1470 20:00
	United States of America Requisition	100 Para (100 Max (100 Max 200 Para 88400 (200 Max 200 Para 84000 440000 200 Para 600000 4400000
	Your ref. 7210 San Annual Sa Annual San Annual San Annu	10 Na 60000 44000 10 Pes 10800 17880 10 Pes 1000 17080
	Our ref. 7210 The second secon	10 Per 10000 178.0 10 Per 100 7.0 10 Per 100 8.0 10 Per 200 18.0 10 Per 200 18.0
	The second s	1.0 Pcs \$000 12.00 1.0 Pcs \$000 \$2.0
	Page 1 / 2	10.00 Pos 1000 80.00 100 Pos 102000 1020.00 Bg70.63
Objects		
Criteria	Item number Description Quantity Unit Unit price Amount	
Invoice		
ACCR Fields	A LS-150 Loudspeaker, Cherry, 150W 1.00 Pcs 129.00 129.00 services	Invoice Sater 1857
✓ Kig Total	A LS-2 Cables for Loudspeakers 10.00 Box 21.00 210.00	Karbar 1217 Data 2017 Restard 2017 Restard Control Con
TotalAmount		Variation Arration Ar
✓ KS Number	A LS-75 Loudspeaker, Cherry, 75W 1.00 Pcs 79.00 79.00 months because b	Carely Life URpha Amare
InvoiceNo	A LS-81 Loudspeaker, Walnut, 80W 1.00 Pcs 79.00 79.00	10 Pox 11200 8200
🗸 🍂 Date	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	100 Tex 6000 (60.00 100 Tex 1000 600 100 Tex 6000 (200.00
 InvoiceDate 	A SPK-100 Spike for L5-100 160.00 Pcs 21.00 3,360.00	10 Per 6000 1000 10 Per 6000 1000
Data Fields	A 1896-S ATHENS Desk 10.00 Pcs 649.40 6,169.30	2.00 Pes 4000 128.00 2.00 Pes 0000 210.00 2.00 Pes 1000 48.00
Body Fields	A SPK-100 Spike for LS-100 160.00 Pcs 21.00 3,360.00 Spike for LS-100 160.00 Pcs 21.00 3,360.00 Spike for LS-100	10 Add
V ItemLine (A)		100 Pen 14/0 8.00 100 Pen 2560 2540 100 Pen 2480 8.00
LineAmount	A 1905-5 ATHENS MODILE PEGESTAL 1.00 PCS 281.40 257.33	10 Na 200 200 500 Not 110 120 120 Na 200 200
LineQuantity	A 1908-S LONDON Swivel Chair, blue 1.00 Pcs 123.30 117.13	100 Fea 640 640 100 Fea 640 640 50 Fea 110 620
LinePrice Description	A 1920-5 ANTWERP Conference Table 1.00 Pcs 420.40 399.38	10 Per 170 270 10 Per 1270 14.00 100 Per 140 1400 V
ArticleNe *	A 1928-S ANIWERP Conference Table 1.00 Pcs 420.40 399.38	
Engine: OCR Forms Dictionary: English		: 16 Column: 102 Page: 1 .:

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 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 Lasernet Form Editor or Lasernet XML Transformer and processed by the Lasernet Form Engine. Please ensure that you have the required XML Input and Output licenses for Lasernet.

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