



This document details the workflow for converting PDFs/scanned images to Microsoft Word format using ABBYY Finereader, using TransTools suite inside Microsoft Word to help with some tedious tasks.

Table of Contents

I.	Prerequisites.....	2
1.	Configuring ABBYY Finereader.....	2
2.	Configuring Microsoft Word.....	4
2.1.	View table gridlines.....	4
2.2.	View formatting marks (non-printable symbols).....	4
2.3.	View fields.....	5
II.	Workflow.....	7
1.	Step 1 – Import document into Finereader.....	8
2.	Step 2 – Define structure and recognize each page.....	8
3.	Step 3 – Check for recognition issues in Finereader.....	21
4.	Step 4 – Save Document / Selected Pages.....	22
5.	Step 5 – Check text recognition errors in Microsoft Word.....	22
6.	Step 6 – Automatically fix / check for formatting issues in Microsoft Word.....	22
6.1.	Removing incorrect paragraph breaks and line breaks in the middle of sentences.....	22
6.2.	Document reformatting and finding key issues.....	23
6.3.	Removing redundant tags.....	36
7.	Step 7 – Manually format Word document.....	40
8.	Step 8 – Handle images containing text.....	44
9.	Step 9 – Translate.....	50
10.	Step 10 – Pre-flight check.....	51
III.	Tips: Preparing accurate quotes for PDF translation.....	53
IV.	Extra tools for Word document formatting before and after translation in CAT tools.....	59
1.	Removing excessive spaces in Word documents.....	59
2.	Finding and correcting punctuation issues, common mistakes, common text recognition issues.....	59



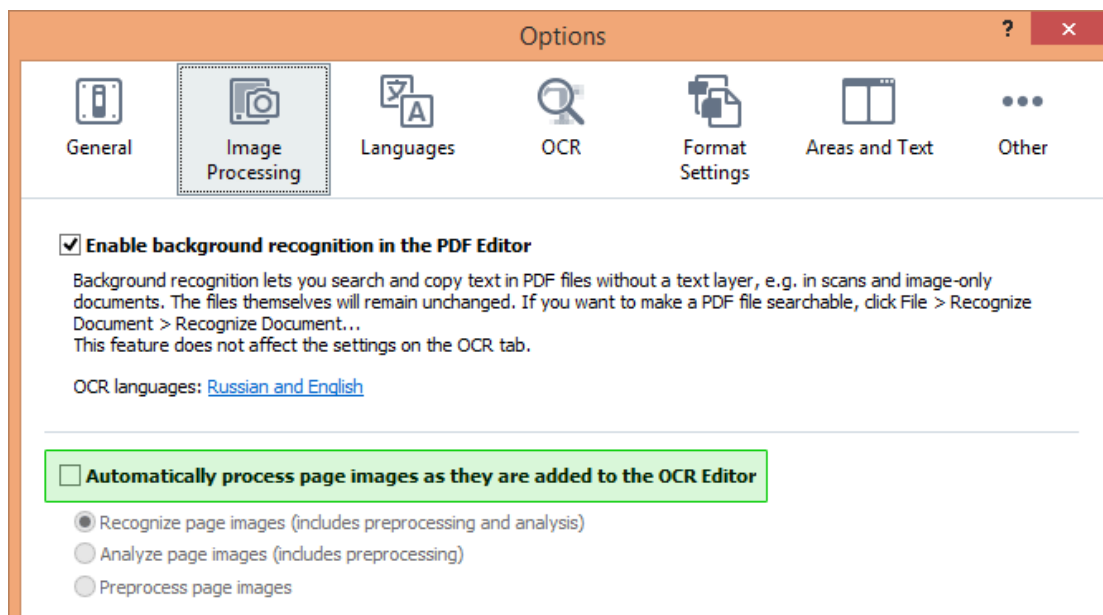
I. Prerequisites

1. Configuring ABBYY Finereader

During initial use of Finereader make sure that it has the following settings (this will need to be done only once). The instructions below are for Finereader 15:

Open **Options** dialogue by clicking **Tools -> Options...**

- Under **Image Processing** tab use the settings specified below (this is required to prevent Finereader from detecting recognition areas automatically when the document is opened):





Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

- Under **OCR** tab use the settings specified below (if *Headers and Footers* option is activated, Finereader can sometimes omit some text located in document headers or footers, so this option needs to be turned off):

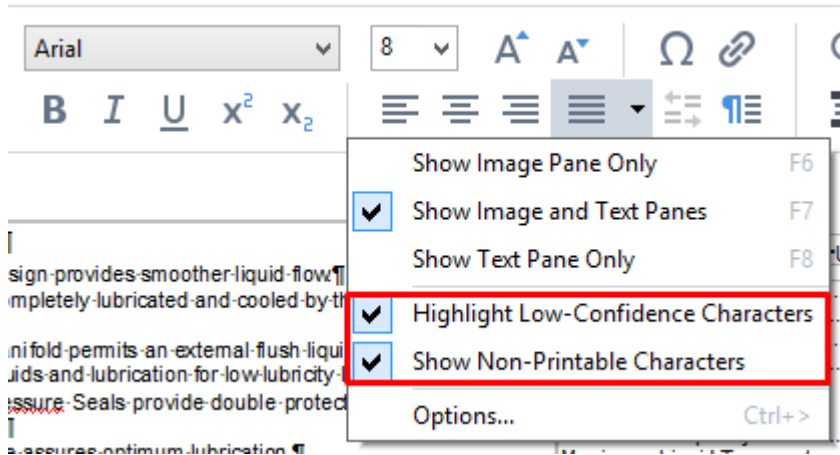
The screenshot shows the 'Options' dialog box with the 'OCR' tab selected. The 'PDF recognition mode' is set to 'Automatically choose between OCR and text from PDF'. Under 'OCR speed and accuracy', 'Thorough recognition' is selected. Under 'Document type', 'Auto' is selected. Under 'Detection of structural elements', 'Table of contents' and 'Footnotes' are checked. Under 'Barcodes', 'Convert barcodes into strings' is unchecked.

- Under **Format Settings** tab, **DOC(X)/RTF/ODT** sub-tab, use the settings specified below:

The screenshot shows the 'Options' dialog box with the 'Format Settings' tab selected and the 'DOC(X) / RTF / ODT' sub-tab active. The 'Document layout' is set to 'Editable copy'. The 'Default paper size' is 'Automatic'. 'Keep pictures' is checked and set to 'Compact size'. Under 'Keep headers, footers, and page numbers', 'Keep page breaks', 'Keep line numbers', and 'Keep text and background colors' are checked. 'Highlight low-confidence characters' is also checked.



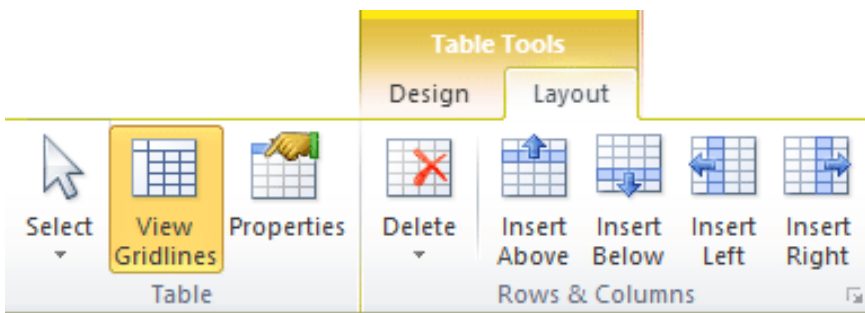
- In the toolbar that appears above the Text Pane on the right side, right-click and activate the settings specified below:



2. Configuring Microsoft Word

2.1. View table gridlines

Insert a table into a new document (*Insert -> Table*) or select any table in an existing document, put the cursor inside the table and then switch to **Table Tools – Layout** ribbon tab. **View Gridlines** option must be activated. This will allow you to detect invisible table borders in the future.



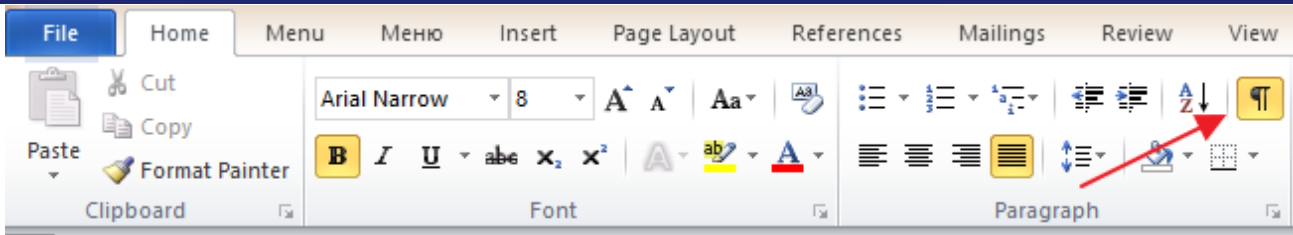
With this option activated, invisible table borders will be displayed using a thin dotted line as follows:



These thin dotted borders are not printed, so do not try to remove them. The customer will not see them, but it will be easier for you to find and work with tables.

2.2. View formatting marks (non-printable symbols)

Make sure that the option to display non-printable symbols is activated (*Home tab, Paragraph group*).



When this option is turned on, Microsoft Word displays various symbols as follows:

paragraph breaks	¶
line breaks	↵
section breaksSection Break (Next Page)
page breaksPage Break.....
column breaksColumn Break.....
tab characters	→
end-of-cell	— ⊞ —
soft hyphens	BO ¶
some other symbols (space, non-breaking space, etc.)	

It should be remembered that your customer will not be able to see these symbols on printed pages and, in most cases, on their screen, so you should not worry about this. However, when you see these symbols on your screen, it will be much easier for you to perform formatting.

2.3. View fields

Activate visibility of (automatic) fields in order to make it easy to locate such fragments as tables of contents, page numbers, total pages and other automatically updated text.

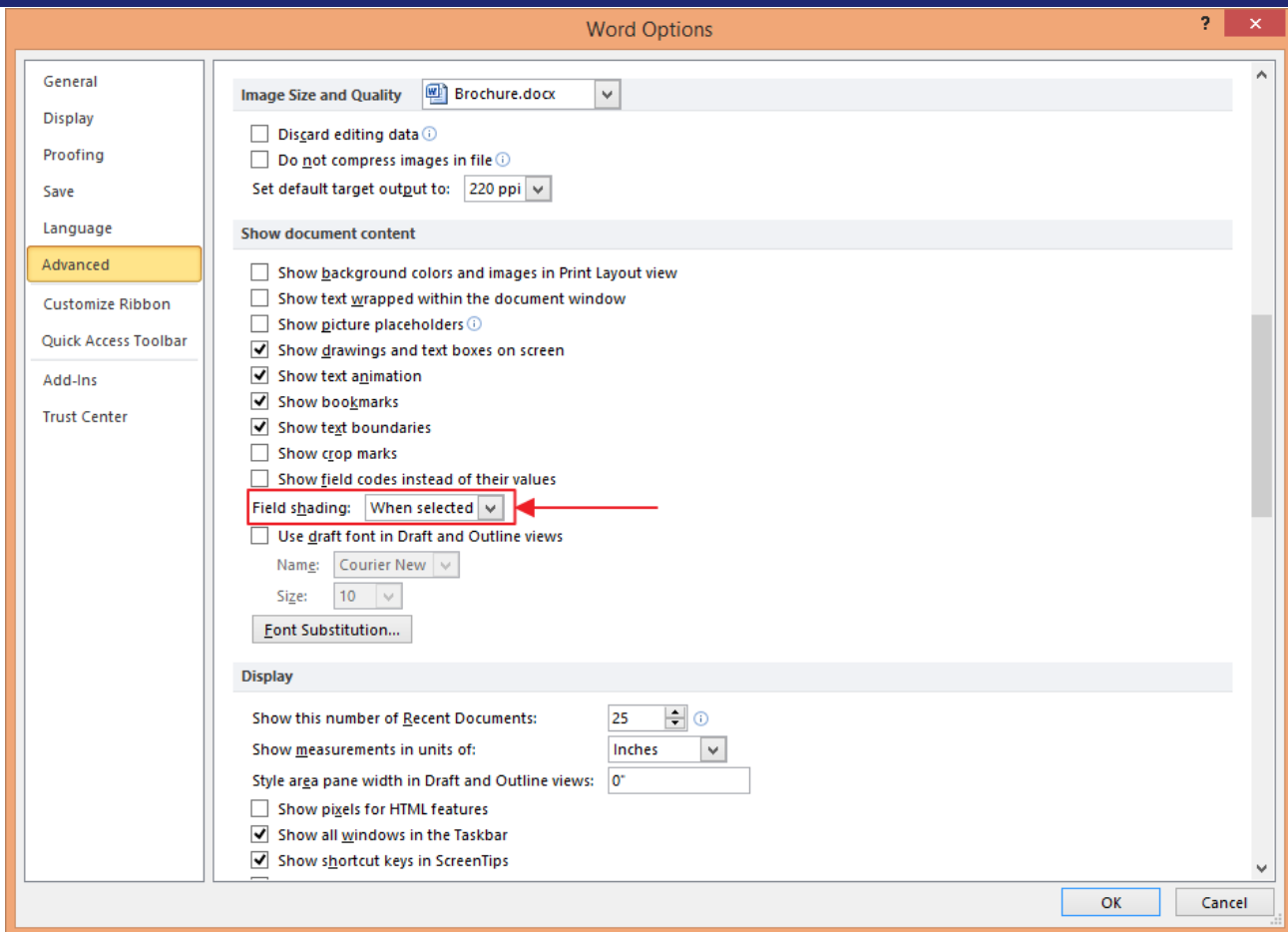
To do this, switch to *File* tab (Word 2010+), click *Options* menu, and in the Word Options dialogue under *Advanced* find the *Field Shading* setting and choose *When Selected* or *Always* as shown on the screenshot below.



Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

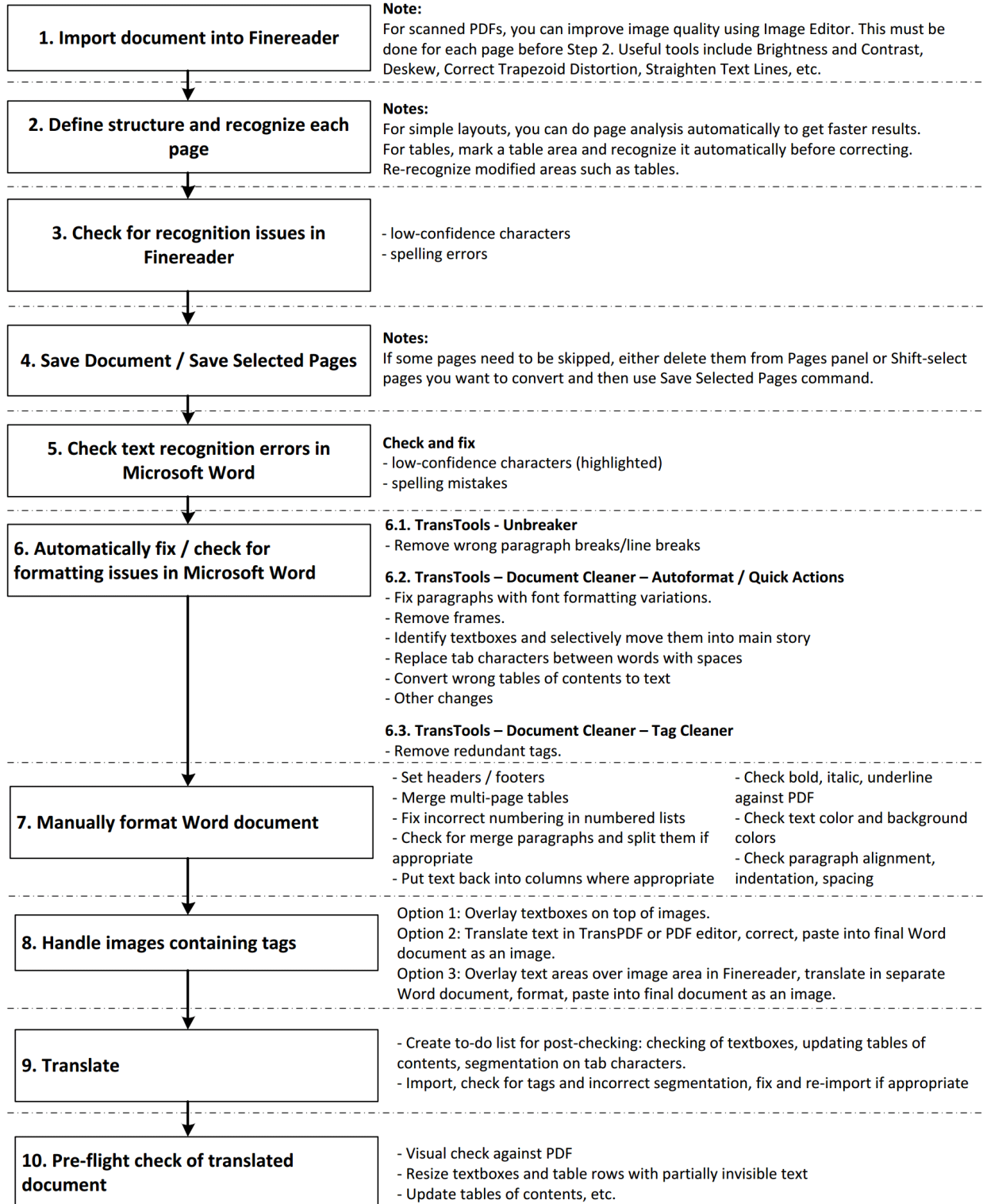


Remember that the grey shading of fields will not be displayed on printed pages, so do not try to clear this shading. To make sure that you are dealing with a field and not shaded text, you can simply right-click it – if the context menu contains *Update Field* menu item, then this is an updatable field and not shaded text.



II. Workflow

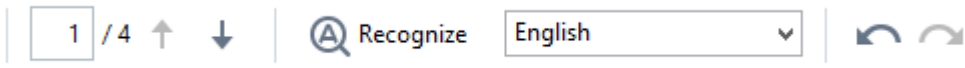
Here is the recommended workflow diagram. I would recommend you to follow it until you have experience with each step of the process, so that you can work out your own unique workflow.





1. Step 1 – Import document into Finereader

After loading a new document into Finereader, before you start the recognition process, always check that the recognition language(s) is (are) selected properly. It is possible to select multiple languages.



For scanned pages that have inadequate visual quality or if you see a lot of low-confidence characters after recognizing a page, you can improve image quality using Image Editor:



This must be done before you mark recognition areas on the page, since all areas will be cleared from the page after it has been edited. Useful tools in the Image Editor include *Brightness and Contrast*, *Deskew*, *Correct Trapezoid Distortion* and *Straighten Text Lines*. Read Finereader Reference Documentation for more information (for Finereader 15, see https://help.abbyy.com/en-us/finereader/15/user_guide/adjustimage).




2. Step 2 – Define structure and recognize each page

There are two objectives when you convert a PDF document for further translation:

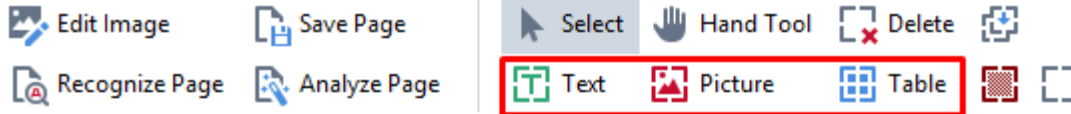
1. minimize the document layout issues as much as possible so that the document does not fall apart during text expansion or shrinking as a result of translation or editing,
2. make the document suitable for translation inside your CAT tool

Marking recognition areas manually helps you meet these objectives, preventing a number of conversion issues.

There are four primary types of recognition areas: text, picture, background picture and table. Three of them – text, picture and table – are heavily used for manual area definition:

	Text	Text area
	Picture	Picture area
	Table	Table area

To mark a specific recognition area, choose the area type on the toolbar, and then mark this area on the current page:



15 Frame Plunger Pump

Standard Model

660

Stainless Steel Model

661D

SS Flushed Model

661C

FEATURES

Superior Design

- Triplex plunger design provides smoother liquid flow.
- V-Packings are completely lubricated and cooled by the liquid being pumped.
- Special ported manifold permits an external flush liquid providing cooling for hi-temp liquids and lubrication for low lubricity liquids.
- Lubricated J-o-Pressure Seals provide double protection against external leakage.
- Oil bath crankcase assures optimum lubrication.
- Close tolerance concentricity of the ceramic plunger offers perfect alignment and maximizes seal life.

SPECIFICATIONS

	U.S. Measure	Metric Measure
Flow.....	18 GPM	(30 L/M)
Pressure Range.....	100 to 3000 PSI	(7 to 210 Bar)
Inlet Pressure Range.....	Flooded to 60 PSI	(Flooded to +4 Bar)
RPM.....	1429 RPM	(1429 RPM)
Bore.....	.787"	(20 mm)
Stroke.....	1.131"	(29 mm)
Crankcase Capacity.....	.42 oz.	(1.26 L)
Maximum Liquid Temperature.....	160°F	(71°C)
Above 130F call CAT PUMPS for inlet conditions and elastomer recommendations.		
Inlet Ports (2).....	3/4" NPTF	(3/4" NPTF)
Flushing Ports (2) (661C).....	1/8" NPTF	(1/8" NPTF)

Different area types are used for different purposes:

Text – one text column (standard appearance of text in a document) or several text columns similar to a newspaper or magazine (text from the end of one column flows into the beginning of the second column, etc.)



Modularization

ZAHIER I. MALIK and RONALD D. KEY
The Pro-Quip Corporation
A Subsidiary of Linde AG
Tulsa, Oklahoma

A Practical and Economical Construction Technique for Process Systems in the Chemical, Petrochemical, Gas Processing, and Oil Refining Industry.

working, the precision, safety, and platform. Modules can be horizontal, vertical, single level, or multi-level depending on the plant space, equipment, and required piping configuration. The optimum split of modules to total construction efforts is determined for each individual project based upon such factors as local labor costs, transportation limitations and schedule.

As one may surmise, a true modular project has more sophisticated engineering and design requirements than a field erected project. Fortunately, the higher design costs can be offset by minimizing downstream equipment erection that is more costly than shop fabrication work. Here, the use of process applications are good candidates for modularization. The preliminary engineering factors are equipment size and overall system complexity (i.e., quantity of pipe, valves, instrumentation, tubing, modules, and other modules which interconnect with major equipment items). For a true modular project, all equipment will fit into a horizontal module which can be shipped, while providing adequate space for the installation of all piping, tubing, instrumentation, insulation, hangers, and elevated systems. A plant should be modularized if it can be engineered, designed, shop fabricated, erected, transported, installed, and commissioned at less cost and less time or the construction would afford and allow very close control of quality or when similar factors during the course of construction is improved.

The modularization technique transfers costly, weather-dependent field construction maintenance to a less costly, highly productive shop environment.

The obvious benefits of this delivery:

1. Controlled environment for fabrication resulting in higher productivity.
2. Experienced staff (produced contractors) resulting in higher productivity.
3. Improved quality control resulting in a higher quality facility.
4. Improved material control and access to material sources resulting in lower cost & higher productivity.
5. Significantly reduced on plant construction hours and time duration resulting in lower construction cost, less operating inconvenience, less risk, and improved safety performance.
6. Better schedule control.
7. Lower cost overall.

The capital equipment:

1. Engineering and design must be executed earlier in the project.
2. Potentially increased structural steel requirements.

In successful modular projects, the increased engineering, design, and structural steel costs are offset by the reduced cost of assembly, painting, piping, electrical, insulation, robotics, and other work being performed in the fabrication facility instead of the field. Transportation of the larger module is usually competitive with the combined cost of shipment of individual items. As control is more easily facilitated, complicated systems are erected in shorter calendar for modularization.

FACTORS AFFECTING MODULAR DESIGN

The following factors affect modular unit design:

1. Process Engineering
2. Equipment Selection
3. Equipment Layout
4. Piping
5. Instrumentation
6. Size of module
7. Equipment Erection
8. Installation
9. S. Rigidity

These factors are discussed below:

Process Engineering

The process design of the plant must be carried out with modular construction techniques in mind. Process design of modular units requires a thorough knowledge of modular equipment. Techniques to facilitate modularization include items such as combining different heat exchanger services in a common shell. A good example is the combination of radial flow heat exchangers in a common shell heat body, or locating all outlet condensers in a common shell. Some may have operating flexibility may be lost in the modification, however, cost saving experience does not indicate a problem is created by such a combination.

Equipment Selection

Selection of the equipment is very important in applying the modular construction technique. There are certain styles of equipment which lend themselves more to modular construction than others. A single example is vertical vs. horizontal heat exchanger and column pumps. A vertical air line pump requires less connecting and disassembly and then one over the potential piping difficulties that may be encountered with a horizontal pump.

Over the years, Pro-Quip has gained considerable knowledge as to the suitability of process unit equipment, thanks to many of our client's feedback. The feedback has given us more insight as to the selection of more suitable equipment for transportation into established facilities. By utilizing equipment items which meet client's specifications, and have proven superior reliability, the end-user can approach the project knowing maintenance costs will be reduced. Equipment is selected due to its construction, other features, compatibility with the module construction, maintenance features, and factors operating reliability.

Equipment Layout

After equipment selection, the overall layout of the equipment is developed. This is the most critical phase of the design, as it defines the number of modules, length of piping runs, assembly, and ultimately the overall plant plan.

Often, modular construction is especially beneficial for locations where plot space is a premium. In these situations, the access to

Regular text (one column)

Multi-column text flowing from one column into another like in a newspaper

Picture – regular graphical image, including an image containing many text labels

Table

a) simple table structure with columns and rows (including a table with invisible borders)

ITEM	PART NUMBER			DESCRIPTION	QTY
	660	MATL	661D	MATL	
2	30067	STL	30067	STL	1
5	92519	STZP	92538	S	8
	125824	STCP R	—	—	8
8	43496	AL	43496	AL	2
9	815281	FBR	815281	FBR	2
	—	—	814800	S	2
10	11340	NBR	11340	NBR	2
11	43495	NBR	43495	NBR	2
15	39060	STL	39060	STL	2
20	48600	TNM	48600	TNM	3
25	43494	FCM	43494	FCM	1
				Key (M8x7.5x25)	
				Screw, HHC Sems (M6x16)	
				Screw, HHC Sems (M6x16)	
				Cover, Bearing	
				Split Shim, Bearing Cover, 2 Pc.	
				Split Shim, Bearing Cover, 2 Pc.	
				O-Ring, Bearing Case - 70D	
				Seal, Oil, Crankshaft - 70D	
				Bearing, Roller	
				Rod Assy, Connecting [11/00]	
				Crankshaft - Dual End	

b) complex table structure (text does not exactly look like a table, but it resembles a grid in which text in one column is logically linked to text in another column)



15 Frame Plunger Pump

*Standard
Model*

660

*Stainless Steel
Model*

661D

*SS Flushed
Model*

661C

A table is a perfect area type. Text which is enclosed into such an area will always preserve its layout during translation, even if it expands too much. So, if it is difficult for you to determine what area type must be used, you can simply choose a table.



In order to define recognition areas on a page, you should use your imagination and **split the page into horizontal bands** that occupy the entire width of the page. Each band should represent a logical block of information.




15 Frame Plunger Pump

Standard Model 660

Stainless Steel Model 661D

SS Flushed Model 661C

FEATURES

Superior Design

- Triple plunger design provides smoother liquid flow.
- V-Packings are completely lubricated and cooled by the liquid being pumped.
- Special ported manifold permits an external flush liquid providing cooling for hi-temp liquids and lubrication for low lubricity liquids.
- Lubricated Lo-Pressure Seals provide double protection against external leakage.
- Oil bath crankcase assures optimum lubrication.
- Close tolerance concentricity of the ceramic plunger offers perfect alignment and maximizes seal life.

Quality Materials

- Precision design 304 and 316 Stainless steel valves and seals are hardened and polished for ultimate seating and extended valve life.
- Bronze or Duplex Stainless Steel manifolds are strong and corrosion resistant.
- Special concentric, high-density, polished, solid ceramic plungers provide a true wear surface and extended seal life.
- Specially formulated, CAT PUMP exclusive, V-Packings offer unmatched performance and seal life.
- Die cast aluminum crankcase provides high strength, minimum weight and precision tolerance control.
- Chrome-moly crankshaft gives unmatched strength and surface hardness.
- Overized crankshaft bearings with greater loading capacity mean longer bearing life.

Easy Maintenance

- Wet-end is easily serviced without entering crankcase, requiring less time and effort.
- Inlet and discharge valve assemblies interchangeable for easier maintenance without disturbing piping.
- Pressed packings mean no packing gland adjustment is necessary, reducing maintenance costs.

WARNING

All systems require both a primary pressure regulating device (PRV, regulator, valve) and a secondary pressure safety relief device (PSD, relief valve, safety valve). Failure to install such relief devices could result in personal injury or damage to the pump or to system components. CAT PUMPS does not assume any liability or responsibility for the operation of a customer's high pressure system.

SPECIFICATIONS

	U.S. Measure	Metric Measure
Flow 0-30 GPM	(0-113 L/M)
Pressure Range 100 to 3000 PSI	(7 to 210 Bar)
Inlet Pressure Range Flooded to 60 PSI	(Flooded to +4 Bar)
RPM 1425 RPM	(1425 RPM)
Bore781"	(20 mm)
Stroke 1.181"	(30 mm)
Crankcase Capacity 42 cc.	(1.26 L)
Maximum Liquid Temperature 160°F	(71°C)

Above 1200 psi CAT PUMPS for inlet conditions and material recommendations.

Alternate Specifications

	U.S. Measure	Metric Measure
Flow 7.5 GPM	(27 L/M)
Pressure Range 100 to 3000 PSI	(7 to 210 BAR)
Inlet Pressure Range -5 to +60 PSI	(-0.35 to +4)
RPM 1000 RPM	(1000 RPM)

ELECTRIC HORSEPOWER REQUIREMENTS

FLOW	PRESSURE			MOTOR PULLEY SIZE		
	PSI 2000	PSI 2500	PSI 3000	Using 1725 RPM Motor & 10" O.D. Pump Pulley		
U.S. GPM	140	175	210	RPM	Pulley O.D.	
10.0	38	13.7	17.2	26.6	1425	8.1
9.0	34	12.4	15.4	18.5	1285	7.3

DETERMINING THE PUMP R.P.M. Rated G.P.M. * "Desired" G.P.M.

DETERMINING THE REQUIRED H.P. Rated R.P.M. * "Desired" R.P.M.

DETERMINING MOTOR PULLEY SIZE G.P.M. x PSI * Electric Motor H.P. Required

DETERMINING MOTOR PULLEY SIZE Motor Pulley O.D. * Drive Pulley O.D. Motor R.P.M.

See complete Drive Packings (Double Pulleys, Balls, Hubs, Keys) Tech Bulletin 202. Refer to pump Service Manual for repair procedures, additional technical information and pump warranty.

"Customer confidence is our greatest asset"

(above, the horizontal bands into which the document is split are shown in different colours)

The horizontal bands must be defined in such a way that each band has one of the following layouts:

1. A band that consists of a *single text area*

Modularization

ZAHEER I. MALIK and RONALD D. KEY
The Pro-Quip Corporation
A Subsidiary of Linde AG
Tulsa, Oklahoma

A Practical and Economical Construction Technique for Process Systems in the Chemical, Petrochemical, Gas Processing, and Oil Refining Industry.



2. A band that consists of a *single picture area*



3. A band that consists of a *single table area*

ITEM	PART NUMBER			DESCRIPTION	QTY	
	660	MATL	661D	MATL		
2	30067	STL	30067	STL	Key (M8x7.5x25)	1
5	92519	STZP	92538	S	Screw, HHC Sems (M8x16)	8
	125824	STCP R	—	—	Screw, HHC Sems (M8x16)	8
8	43496	AL	43496	AL	Cover, Bearing	2
9	815281	FBR	815281	FBR	Split Shim, Bearing Cover, 2 Pcs.	2
	—	—	814800	S	Split Shim, Bearing Cover, 2 Pcs.	2
10	11340	NBR	11340	NBR	O-Ring, Bearing Case - 70D	2
11	43495	NBR	43495	NBR	Seal, Oil, Crankshaft - 70D	2
15	39060	STL	39060	STL	Bearing, Roller	2
20	48800	TNM	48800	TNM	Rod Assy, Connecting [11/00]	3
25	43494	FCM	43494	FCM	Crankshaft - Dual End	1
31	828710	—	828710	—	Protector, Oil Cap w/Foam Gasket	1
32	43211	ABS	43211	ABS	Cap, Oil Filler	1
33	14177	NBR	14177	NBR	O-Ring, Filler Cap - 70D	1
37	92241	—	92241	—	Gauge, Oil, Bubble w/Gasket - 80D	1
38	44428	NBR	44428	NBR	Gasket, Flat, Oil Gauge - 80D	1
40	92520	STZP	92542	S	Screw, HHC Sems (M8x20)	4
	126541	STCP R	—	—	Screw, HHC Sems (M8x20)	4
48	25625	STCP	25625	STCP	Plug, Drain (1/4"x19BSP)	1

4. A band that consists of several text columns located next to each other, where the text flows from one column to the next, i.e. after reading text at the bottom of one column, you continue reading from the top of the next column

1 Die cast aluminum **crankcase** means high strength, lightweight, and excellent tolerance control.

2 Oversized crankshaft **bearings** provide extended bearing life and pump performance.

3 Chrome-moly **crankshaft** provides unmatched strength and surface hardness for long life.

4 Matched oversized TNM **connecting rods** noted for superior tensile strength and bearing quality.

5 Special stainless steel **plunger rods** with high strength crossheads for longevity and corrosion resistance.

6 The stainless steel **slinger** provides back-up protection for the crankcase seal, keeping pumped liquids out of the crankcase.

7 Special concentric, high-density, polished, solid ceramic **plungers** provide a true wear surface and extended seal life.

8 **Manifolds** are a high tensile strength bronze or Duplex stainless steel for long term, continuous duty and corrosion resistance.

9 100% wet **seal** design adds to service life by allowing pumped liquids to cool and lubricate on both sides.

10 Stainless steel **valves, seats and springs** provide corrosion-resistance, ultimate seating and extended life.

11 Specially formulated, CAT PUMP exclusive, **V-Packings** offer unmatched performance and seal life.

12 **Crossheads** are 360° supported for uncompromising alignment.

13 Optional special **ported Flushed Manifold** permits external flush for lubricating and cooling seals when pumping liquids such as DI water, Hi-Temp or low lubricity. (Not shown).



ABSTRACT

The following paper outlines the aspects of using the modular construction concept. Each point is detailed from a general perspective so that it may be applied to any industry. Major points discussed include engineering, design, equipment size, type, and instrumentation.

INTRODUCTION

Over the years the notion of modularized processing units invariably has conjured up images of equipment packed into the smallest possible area with little concern for the necessary access for safety, proper operation, and maintenance of the unit. These concerns, which more often than not are justified, have been further perpetuated over the years by many skid fabricators who, in their efforts to provide low-cost processing units for installation in the limited space often available, have built them in just such fashion.

Unfortunately, the space available for the installation of new units is still limited as most existing facilities have established and fully developed plot areas, thereby creating a challenge for both the client and the engineering contractor.

Pro-Quip recognized the need for operating and maintenance accessibility combined with the challenge of limited plot space and over the years has endeavored to change the image of the modularized process units by developing modular design concepts in such manner that safety, maintenance and ease of operation are given prime consideration. In this concept modular construction is not simply putting the most in the least space, but involves a sophisticated art requiring special expertise to engineer, design and fabricate process facilities compatible with facility operator needs.

Modularized construction has many positive aspects to consider. The modules contain the equipment, piping, heat tracing, electrical instrumentation systems, specialized

5. A band that consists of several columns including text area(s), picture area(s) and/or table area(s). The difference from section 4 is the following: besides text, a column may contain image(s), table(s) or both.

FEATURES

Superior Design

- Triplex plunger design provides smoother liquid flow.
- V-Packings are completely lubricated and cooled by the liquid being pumped.
- Special ported manifold permits an external flush liquid providing cooling for hi-temp liquids and lubrication for low lubricity liquids.
- Lubricated Lo-Pressure Seals provide double protection against external leakage.
- Oil bath crankcase assures optimum lubrication.
- Close tolerance concentricity of the ceramic plunger offers perfect alignment and maximizes seal life.

Quality Materials

- Precision design 304 and 316 Stainless steel valves and seats are hardened and polished for ultimate seating and extended valve life.
- Bronze or Duplex Stainless Steel manifolds are strong and corrosion resistance.
- Special concentric, high-density, polished, solid ceramic plungers provide a true wear surface and extended seal life.
- Specially formulated, CAT PUMP exclusive, V-Packings offer unmatched performance and seal life.
- Die cast aluminum crankcase provides high strength, minimum weight and precision tolerance control.
- Chrome-moly crankshaft gives unmatched strength and surface hardness.
- Oversized crankshaft bearings with greater loading capacity mean longer bearing life.

Easy Maintenance

- Wet-end is easily serviced without entering crankcase, requiring less time and effort.
- Inlet and discharge valve assemblies interchange for easier maintenance without disturbing piping.
- Preset packings mean no packing gland adjustment is necessary, reducing maintenance costs.

WARNING

All systems require both a primary pressure regulating device (i.e., regulator, unloader) and a secondary pressure safety relief device (i.e., pop-off valve, safety valve). Failure to install such relief devices could result in personal injury or damage to the pump or to system components. CAT PUMPS does not assume any liability or responsibility for the operation of a customer's high pressure system.

SPECIFICATIONS

	U.S. Measure	Metric Measure
Flow.....	10 GPM	(38 L/M)
Pressure Range.....	100 to 3000 PSI	(7 to 210 Bar)
Inlet Pressure Range.....	Flooded to 60 PSI	(Flooded to +4 Bar)
RPM.....	1429 RPM	(1429 RPM)
Bore.....	.787"	(20 mm)
Stroke.....	1.181"	(30 mm)
Crankcase Capacity.....	42 oz.	(1.28 L)
Maximum Liquid Temperature.....	160°F	(71°C)

Above 130F call CAT PUMPS for inlet conditions and elastomer recommendations.

Inlet Ports (2).....	3/4" NPTF	(3/4" NPTF)
Flushing Ports (2) (661C).....	1/8" NPTF	(1/8" NPTF)
Discharge Ports (2).....	1/2" NPTF	(1/2" NPTF)
Pulley Mounting.....	Either Side	(Either Side)
Shaft Diameter.....	1.181"	(30 mm)
Weight.....	44 lbs.	(19.9 kg)
Dimensions.....	16.28x12.99x6.44"	(413.50x330x163.50 mm)

Alternate Specifications

	U.S. Measure	Metric Measure
Flow.....	7.0 GPM	(27 L/M)
Pressure Range.....	100 to 3000 PSI	(7 to 210 BAR)
Inlet Pressure Range.....	-5 to +60 PSI	(-0.35 to +4)
RPM.....	1000 RPM	(1000 RPM)

ELECTRIC HORSEPOWER REQUIREMENTS

FLOW	PRESSURE			MOTOR PULLEY SIZE		
	PSI 2000	PSI 2500	PSI 3000	Using 1725 RPM Motor & 10" O.D. Pump Pulley		
U.S. GPM	BAR	BAR	BAR	RPM	Pulley O.D.	
10.0	38	13.7	17.2	20.6	1429	8.1
9.0	34	12.4	15.4	18.5	1286	7.3

DETERMINING THE PUMP R.P.M.	$\frac{\text{Rated G.P.M.}}{\text{Rated R.P.M.}}$	=	"Desired" G.P.M.
DETERMINING THE REQUIRED H.P.	$\frac{\text{GPM} \times \text{PSI}}{1460}$	=	"Desired" R.P.M.
DETERMINING MOTOR PULLEY SIZE	$\frac{\text{Motor Pulley O.D.}}{\text{Pump R.P.M.}}$	=	Electric Brake H. P. Required
		=	Pump Pulley O.D. Motor R.P.M.

See complete Drive Packages (Inclds: Pulleys, Belts, Hubs, Key) Tech Bulletin 003. Refer to pump Service Manual for repair procedure, additional technical information and pump warranty.


6. A band that consists of a single table area, which contains nested tables inside its cells (this layout is rare).

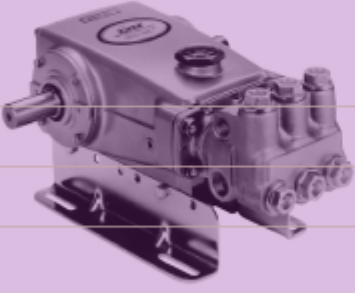


Examples of document recognition

Below you will find how several typical layout examples can be split into horizontal bands and areas:

1. Example 1



	15 Frame Plunger Pump	
	Standard Model	660
	Stainless Steel Model	661D
	SS Flushed Model	661C

FEATURES

Superior Design

- Triplex plunger design provides smoother liquid flow.
- V-Packings are completely lubricated and cooled by the liquid being pumped.
- Special ported manifold permits an external flush liquid providing cooling for H-temp liquids and lubrication for low lubricity liquids.
- Lubricated Lo-Pressure Seals provide double protection against external leakage.
- Oil bath crankcase assures optimum lubrication.
- Close tolerance concentricity of the ceramic plunger offers perfect alignment and maximizes seal life.

Quality Materials

- Precision design 304 and 316 Stainless steel valves and seats are hardened and polished for ultimate seating and extended valve life.
- Bronze or Duplex Stainless Steel manifolds are strong and corrosion resistant.
- Special concentric, high-density, polished, acid ceramic plungers provide a true wear surface and extended seal life.
- Specially formulated, CAT PUMP exclusive, V-Packings offer unmatched performance and seal life.
- Die cast aluminum crankcase provides high strength, minimum weight and precision tolerance control.
- Chrome-moly crankshaft gives unmatched strength and surface hardness.
- Oversized crankshaft bearings with greater loading capacity mean longer bearing life.

Easy Maintenance

- Wet-end is easily serviced without entering crankcase, requiring less time and effort.
- Inlet and discharge valve assemblies interchange for easier maintenance without disturbing piping.
- Pre-set packings mean no packing gland adjustment is necessary, reducing maintenance costs.

SPECIFICATIONS	U.S. Measure	Metric Measure
Flow	10 GPM	(38 L/M)
Pressure Range	100 to 3000 PSI	(7 to 210 Bar)
Inlet Pressure Range	Flooded to 60 PSI	(Flooded to +4 Bar)
RPM	1429 RPM	(1429 RPM)
Bore	7/8"	(20 mm)
Stroke	1.181"	(30 mm)
Crankcase Capacity	42 oz	(1.26 L)
Maximum Liquid Temperature	150°F	(71°C)
<i>Note: For use of CAT PUMPS in wet conditions and corrosive environments.</i>		
Inlet Ports (2)	1/2" NPTF	(1/2" NPTF)
Flushing Ports (2) (661C)	1/8" NPTF	(1/8" NPTF)
Discharge Ports (2)	1/2" NPTF	(1/2" NPTF)
Pulley Mounting	Either Side	(Either Side)
Shaft Diameter	1.181"	(30 mm)
Weight	44 lbs.	(19.9 kg)
Dimensions	16.28x12.99x6.44"	(413.50x330x163.50 mm)
Alternate Specifications		
Flow	7.0 GPM	(27.1 L/M)
Pressure Range	100 to 3000 PSI	(7 to 210 BAR)
Inlet Pressure Range	-5 to +60 PSI	(-0.35 to +4)
RPM	1000 RPM	(1000 RPM)

ELECTRIC HORSEPOWER REQUIREMENTS						
FLOW	PRESSURE			MOTOR PULLEY SIZE		
	PSI 2000	PSI 2500	PSI 3000	Using 1725 RPM Motor & 10" O.D. Pump Pulley		
U.S. GPM	BAR	BAR	BAR	RPM	Pulley O.D.	
10.0	38	13.7	17.2	20.6	1429	8.1
9.0	34	12.4	15.4	18.5	1286	7.3

DETERMINING THE PUMP R.P.M.	Rated G.P.M.	=	"Desired" G.P.M.
	Rated R.P.M.	=	"Desired" R.P.M.
DETERMINING THE REQUIRED H.P.	G.P.M. x PSI	=	Electric Brake H.P. Required
DETERMINING MOTOR PULLEY SIZE	Motor Pulley O.D.	=	Pump Pulley O.D. Motor R.P.M.

See complete Drive Packages (Pulley, Belt, Hub, Key) from Station 003.
Refer to pump Service Manual for setup procedure, additional technical information and pump features.

"Customer confidence is our greatest asset"

This example uses the following types of bands: 2 (single picture area), 3 (complex table with one picture cell and merged cells), 5 (several columns with text and table areas), 1 (single text area).

In the first table area the pump's photo is placed inside a single cell, and *Treat selected cells as picture* is specified in the Area Properties of the cell.



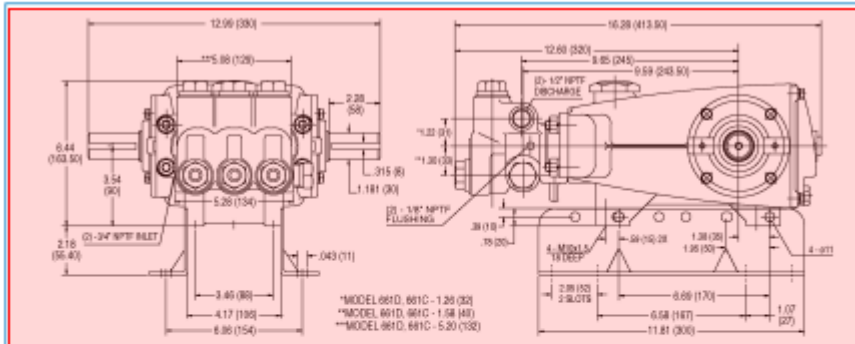
2. Example 2

PARTS LIST						
ITEM	660	PART NUMBER	661D	MATL	DESCRIPTION	QTY
7	32007	STL	32007	STL	Key (M6x7.5x25)	7
8	32519	STZP	32519	S	Screw, HHC Serie (M5x16)	8
9	12824	STZP R	---	---	Screw, HHC Serie (M6x10)	8
10	43496	AL	43496	AL	Cover, Bearing	2
11	815281	FBR	815281	FBR	Solt Shim, Bearing Cover, 2 Pcs.	2
12	---	---	814800	S	Solt Shim, Bearing Cover, 2 Pcs.	2
13	11340	NBR	11340	NBR	O-Ring, Bearing Case - 70D	2
14	43495	NBR	43495	NBR	Seal, Oil, Crankshaft - 70D	2
15	39050	STL	39050	STL	Bearing, Roller	2
16	48900	TNM	48900	TNM	Rod Assy, Connecting (11100)	3
17	43494	PCM	43494	PCM	Crankshaft - Dual End	1
18	83970	---	83970	---	Protector, Oil Cap w/Foam Gasket	7
19	43211	ABS	43211	ABS	Cap, Oil Filler	1
20	14177	NBR	14177	NBR	O-Ring, Filler Cap - 70D	1
21	80241	---	80241	---	Seal, Oil, Bubble w/Gasket - 80D	1
22	44428	NBR	44428	NBR	Gasket, Flat, Oil Sump - 80D	1
23	92520	STZP	92542	S	Screw, HHC Serie (M5x20)	4
24	128543	STZP R	---	---	Screw, HHC Serie (M5x20)	4
25	25625	STCP	25625	STCP	Plug, Drain (1.1/4" x 1.875")	1
26	23170	NBR	23170	NBR	O-Ring, Drain Plug - 70D	1
27	43491	AL	43491	AL	Cover, Rear	1
28	44834	NBR	44834	NBR	O-Ring, Rear Cover - 70D	1
29	115471	AL	115471	AL	Crankcase (Incls: 54)	1
30	27488	S	27488	S	Pin, Guide	2
31	44864	PCP	44864	PCP	Pin, Oil	1
32	43507	CM	43507	CM	Pin, Crosshead	3
33	48228	SHS	---	---	Rod, Plunger, Long	3
34	---	---	48892	SSHB	Rod, Plunger, Short (1302)	3
35	126962	STZP R	126962	STZP R	Washer, Seal	3
36	43500	NBR	43500	NBR	Seal, Oil, Crankcase - 80D	3
37	45976	S	45976	S	Spacer, Bearing	3
38	45975	S	45975	S	Washer, Keyhole	3
39	45966	BS	45966	BS	Spacer, Plunger	3
40	43300	CC	43300	CC	Plunger (M20x77)	3
41	---	---	36661	SS	Stud, Plunger Retainer (M6x70)	3
42	43235	PTFE	43235	PTFE	Back-up-Ring, Plunger Retainer	3
43	17399	NBR	17399	NBR	O-Ring, Plunger Retainer - 80D	3
44	14160	PPM	14160	PPM	O-Ring, Plunger Retainer - 70D	3
45	48204	EPCM	48204	EPCM	O-Ring, Plunger Retainer - 70D	3
46	48891	CU	44041	SS	Gasket, Plunger Retainer	3
47	---	---	44031	SS	Retainer, Plunger (M6)	3
48	164360	S	---	---	Retainer w/Std, Plunger (M5)	3
49	855001	PCP	855001	PCP	Retainer, Seal, 3-cc (0705)	3
50	43508	---	43508	---	Wick, Long Tab	3
51	44871	NBR	44871	NBR	Seal, LPS w/SS-Seal	3
52	44381	PPM	44381	PPM	Seal, LPS w/SS-Seal	3
53	707874	SPTB*	707874	SPTB*	Seal, LPS w/SS-Seal	3
54	48291	BS	48294	SS	Case, Seal (0705)	3
55	14762	NBR	14762	NBR	O-Ring, Seal Case - 70D	3
56	17737	PPM	17737	PPM	O-Ring, Seal Case - 70D	3
57	48205	EPCM	48205	EPCM	O-Ring, Seal Case	3
58	43816	BS	48162	PTFE	Adaptor, Female (0104)	3
59	44807	STG	44807	STG	V-Packing	6
60	44386	PPM	44386	PPM	V-Packing	6
61	43817	BS	44186	SS	Adaptor, Male	3
62	20326	BBZP	44382	SS	Plug, Inlet (3/4" NPT)	1
63	43248	PTFE	43248	PTFE	Back-up-Ring, Seat	6
64	43249	NBR	43249	NBR	O-Ring, Seat - 80D	6
65	44383	PPM	44383	PPM	O-Ring, Seat - 70D	6
66	48209	EPCM	48209	EPCM	O-Ring, Seat - 70D	6
67	43722	S	44031	SS	Seat	6
68	43751	S	48791	SS	Valve	6
69	43751	S	44036	SS	Spring	6
70	44864	PVDF	44864	PVDF	Retainer, Spring	6
71	17617	NBR	17617	NBR	O-Ring, Valve Plug - 90D	6
72	17621	PPM	17621	PPM	O-Ring, Valve Plug - 90D	6
73	48207	EPCM	48207	EPCM	O-Ring, Valve Plug - 70D	6
74	43850	BS	46235	SS	Plug, Valve	6
75	48978	NAB	48788	SSD	Head, Manifold	1
76	---	---	48338	SSD	Head, Manifold Flashed (7.0" Flushing Port) 88TC Only	1
77	126948	STZP R	44036	S	Stud (M10x45)	4
78	126231	STZP R	16847	S	Lockwasher (M10)	4
79	126922	STZP R	31268	S	Nut, Hex (M10)	4
80	30764	NY	30764	NY	Protector, Shaft w/Two Screws	7
81	---	---	307648	NY	Protector, Shaft w/Two SS Screws	7
82	30264	STZP	30264	STZP	Mount, Drive	7
83	30673	STZP	30673	STZP	Mount, Rod Assy	7
84	860550	STZP	860550	STZP	Mount, Rod (Required on Diesel/Gas)	7
85	30209	F	30209	F	Key (7.0" AB [See Drive Packages, Tech Bulletin 003])	7

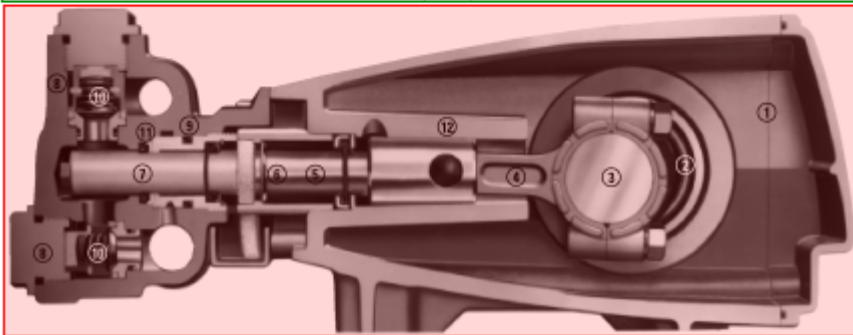
This example uses the following types of bands: 1 (single text area), 3 (simple table area).



3. Example 3



Models 660, 661D, 661C



- | | | |
|---|---|---|
| <p>1 Die cast aluminum crankcase means high strength, lightweight, and excellent tolerance control.</p> <p>2 Oversized crankshaft bearings provide extended bearing life and pump performance.</p> <p>3 Chrome-moly crankshaft provides unmatched strength and surface hardness for long life.</p> <p>4 Matched oversized TNM connecting rods noted for superior tensile strength and bearing quality.</p> <p>5 Special stainless steel plunger rods with high strength crossheads for longevity and corrosion resistance.</p> | <p>6 The stainless steel slinger provides back-up protection for the crankcase seal, keeping pumped liquids out of the crankcase.</p> <p>7 Special concentric, high-density, polished, solid ceramic plungers provide a true wear surface and extended seal life.</p> <p>8 Manifolds are a high tensile strength bronze or Duplex stainless steel for long term, continuous duty and corrosion resistance.</p> <p>9 100% wet seal design adds to service life by allowing pumped liquids to cool and lubricate on both sides.</p> | <p>10 Stainless steel valves, seats and springs provide corrosion-resistance, ultimate seating and extended life.</p> <p>11 Specially formulated, CAT PUMP exclusive, V-Packings offer unmatched performance and seal life.</p> <p>12 Crossheads are 360° supported for uncompromising alignment.</p> <p>13 Optional special ported Flushed Manifold permits external flush for lubricating and cooling seals when pumping liquids such as DI water, Hi-Temp or low lubricity. (Not shown).</p> |
|---|---|---|

<p>World Headquarters CAT PUMPS 1681 - 94th Lane N.E. Minneapolis, MN 55449-4324 Phone (763) 780-5440 — FAX (763) 780-2958 e-mail: techsupport@catpumps.com www.catpumps.com</p> <p>International Inquiries FAX (763) 785-4329 e-mail: info@catpumps.com</p>	 <p>The Pumps with Nine Lives</p>	<p>CAT PUMPS (U.K.) LTD. 1 Pinfel Business Park, Sandy Lane, Chalfont Cookham Bucks, High Wycombe, Bucks HP8 2BP, England Phone: +44 (1295) 622331 — Fax: +44 (1295) 626655 e-mail: sales@catpumps.co.uk www.catpumps.co.uk</p> <p>N.V. CAT PUMPS INTERNATIONAL S.A. Hevelinkens 14, B-2640 Kottich, Belgium Phone: 32 (3) 463 71 68 — Fax: 32 (3) 463 71 61 e-mail: cat@catpumps.be www.catpumps.be</p> <p>CAT PUMPS DEUTSCHLAND GmbH Buchenweg 2, D-65510 Idstein, Germany Phone: 49 6126 6003 0 — Fax: 49 6126 6003 33 e-mail: catpumps@t-online.de www.catpumps.de</p>
---	--	--

Ph 5621 13 Rev D 04/92 661C

This example uses the following types of bands: 2 (single picture area), 1 (single text area), 2 (single picture area), 4 (multi-column text), 3 (table with one image cell) and 1 (single text area).



4. Example 4

		TOTAL E&P RUSSIE			
		KHARYAGA PIII DEVELOPMENT PROJECT			
		TECHNICAL QUERY			
Contract No: 4600001201					
Technical query No: RU-KH3-87-SLT-QUE-000052				Discipline:	Work Unit:
				HVA	292
Technical query title: HVAC system and derogation					
ENG. CONTRACTOR		Name	Job Position	Signature	Date
	Originator		Engineer		9/12/2013
	Checked by		Lead Procurement Engineer		9/12/2013
	Approved by		Engineering Manage		9/12/2013
Response requested date:					
Question:					
According to para 6.1 of GS EP HVA 208 All outdoor ductwork shall be made of Stainless steel A/Sl 316L (Ca < 0.03% / Mo > 2.5%) continuously welded. Outdoor ducts shall be 3 mm thick up to a width of 2.500 mm and 4 mm thick above 2500 mm. As we were informed during the weekly management meeting on 4.09.13, Company had already issued DEROGATION in this respect but hadn't approved it yet.					
Please advise us how to proceed: a) follow the GS requirements; b) follow the DEROGATION; or c) put our design on HOLD until DEROGATION is approved					
Company's Response					
b) ENGINEER to follow the DEROGATION (not formally approved yet):					
1. Rectangular Outdoor and Indoor Air Ducts (with length of straight parts of ductwork not more 2000 mm for duct width from 901 mm and not more 2500 mm for duct width up to 901 mm) for system with static pressure up to 800 Pa shall be fabricated of Stainless Steel sheet thickness being a function of duct width:					
• Up to 300 mm: 0.6 mm					
• From 301 to 500 mm: 0.7 mm					
• From 501 to 900 mm: 0.8 mm					
• From 901 to 1.200 mm: 1.0 mm					
• From 1.201 to 1.400 mm: 1.2 mm					
• From 1.401 to 3.000 mm: 1.5 mm					
2. Transverse joints of Outdoor and Indoor Air Ducts shall be made with flanges or continuously welded depending on pollutant and safety requirement.					
Additional comments					
COMPANY		Name	Job Position	Signature	Date
	response by		Senior HVAC Engineer		9/27/2013
	checked by		Package 2&3 Extension Project Engineer		27.09.13.
	approved by		Package 2/3 Extension Manager		30/09/2013

This example uses the following types of bands: 3 (table area). In this table some cells are recognized as images using *Treat selected cells as picture* option.



5. Example 5

Modularization

ZAHEER I. MALIK and RONALD D. KEY
*The Pro-Quip Corporation
A Subsidiary of Linde AG
Tulsa, Oklahoma*

A Practical and Economical Construction Technique for Process Systems in the Chemical, Petrochemical, Gas Processing, and Oil Refining Industry.




Photo: Modular Construction employed in an Ethylene Recovery Unit built by Pro-Quip for Mobil Chemical in Houston, Texas

ABSTRACT

The following paper outlines the aspects of using the modular construction concept. Each point is detailed from a general perspective so that it may be applied to any industry. Major points discussed include engineering, design, equipment size, type, and instrumentation.

INTRODUCTION

Over the years the notion of modularized processing units inevitably has required by images of equipment packed into the smallest possible area with little concern for the necessary access for safety, proper operation, and maintenance of the unit. These concerns, which more often than not are justified, have been further heightened over the years by many skilled fabricators who, in their efforts to provide low-cost processing units for installation in the limited space often available, have built them in just such fashion.

Unfortunately, the space available for the installation of new units is still limited as most existing facilities have established and fully developed plot area, thereby creating a challenge for both the client and the engineering contractor.

Pro-Quip recognized the need for operating and maintenance accessibility combined with the challenge of limited plot space and over the years has endeavored to change the image of the modularized process units by developing modular design concepts in such manner that safety, maintenance and ease of operation are given prime consideration. In this concept modular construction is not simply packing the most in the least space, but involves a sophisticated art requiring special expertise in engineer, design and fabricate process facilities compatible with facility operator needs.

Modularized construction has many positive aspects to consider. The modules contain the equipment, piping, heat tracing, electrical instrumentation systems, specialized

Copyright Linde BOC Process Plants LLC
All Rights Reserved

This example uses the following types of bands: 1 (single text area), 2 (single picture area), 1 (single text area), 4 (multi-column text), 1 (single text area).

After the document is divided into horizontal bands and the areas are defined in all pages that require conversion, select these pages in the Pages Window, right-click and choose *Recognize Pages*.

Tips:

- To save time in pages which have a more standard layout, you can let Finereader determine areas automatically rather than mark them manually. To do this, select such page(s) in the Pages Window on the left side, right-click the page(s) and choose *Recognize Pages*. If recognition areas are defined incorrectly or it will be necessary to make a lot of changes, simply select all the areas using the Arrow tool, delete them with Delete key, then define the areas manually and recognize them.



- The text that repeats on every page (headers and footers) should be marked with areas only once, skipping the repetitions on other pages. However, remember that a document may contain more than one unique header and footer, i.e. there could be minor differences between headers and footers on different pages.
- If the document contains pictures with text written on them (text labels), one of the best options is to recognize the entire picture as an image and then insert textboxes on top of it inside Microsoft Word (see Step 8 below for more details on this and other image handling options).

3. Step 3 – Check for recognition issues in Finereader

During this step, check the following in the Text window on the right side:

- All areas have been recognized correctly.
- All low-confidence characters are correct:

technique· transfers· cosy,

By default, low-confidence characters will be highlighted in turquoise colour if you previously activated visibility of low-confidence characters in Finereader options as recommended in the Prerequisites section.

- Potential spelling errors (with red underlines) are correct:

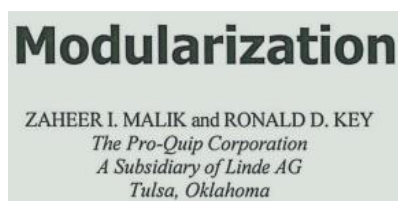
pre-insulated piping

You can also run Verify command to spell-check the recognized text:



- Text is divided into paragraphs correctly:

PDF page



Recognized text (note that 4 paragraphs were converted into one)

Modularization¶

ZAHEER I. MALIK and RONALD D. KEY The Pro-Quip Corporation A Subsidiary of Linde AG Tulsa, Oklahoma¶



This issue mostly occurs with text that does not span the entire width, for example with centred text. To speed up checking, pay attention to paragraph break and line break symbols at the end of lines. When you find a paragraph which was merged by Finereader by mistake, insert missing paragraph breaks or line breaks if appropriate.

4. Step 4 – Save Document / Selected Pages

At this step, use File > Save As > Microsoft Word Document command from the main menu (if all pages will be converted) or Save As > Microsoft Word Document from the right-click menu appearing after you select specific pages in the Pages panel on the left side (if only specific pages selected in the Pages window will be converted).

Use Ctrl key when selecting pages in the Pages window to select specific pages. If there are many pages and some pages are not required, it is easier to delete them from Pages window using *Delete...* command in the right-click menu.

5. Step 5 – Check text recognition errors in Microsoft Word

Open the generated Microsoft Word document side by side with the PDF and check low-confidence characters (which will be highlighted in turquoise colour) and potential spelling errors.

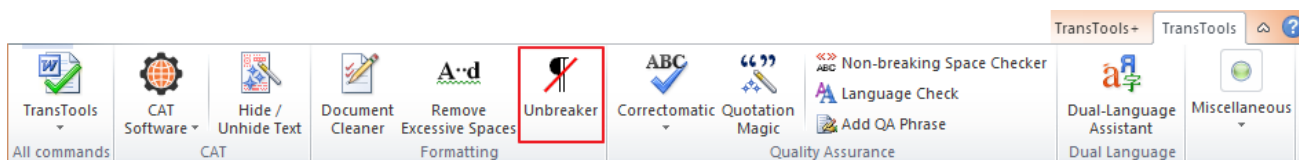
6. Step 6 – Automatically fix / check for formatting issues in Microsoft Word

For this step, you will need to use Document Cleaner and Unbreaker tools available as part of TransTools suite. If TransTools is not yet installed on your computer, install it using the installation program downloaded from <http://translatortools.net/downloads>

6.1. Removing incorrect paragraph breaks and line breaks in the middle of sentences.

Often, Word documents converted with Finereader will contain paragraph breaks and line breaks in the middle of sentences. Unless such breaks are removed before the document is imported into your CAT tool, the text will be segmented incorrectly and you will have to spend time manually merging such segments, placing the resulting tags into correct places in the translation, and removing them from the final document. Incorrect breaks can be removed using Unbreaker tool.

To run Unbreaker, click *Unbreaker* button on TransTools ribbon tab.





15-Fran

Standard Model

Stainless S

SS Flushed

- FEATURES**
- Superior Design**
 - Triplex plunger design provides smoother liquid flow.
 - V-Packings are completely lubricated and cooled by the liquid being pumped.
 - Special ported manifold permits an external flush liquid providing cooling for hi-temp liquids.
 - Lubricated Lo-Pressure Seals provide double protection against external leakage.
 - Oil-bath crankcase assures optimum lubrication.
 - Close tolerance concentricity of the ceramic plunger offers perfect alignment and maximizes
- Quality Materials**
 - Precision design 304 and 316 Stainless steel valve and seats are hardened and polished for
 - Bronze or Duplex Stainless Steel manifolds are strong and corrosion resistance.
 - Special concentric, high-density, polished, solid ceramic plungers provide a true wear surface.
 - Specially formulated, CAT PUMP exclusive, V-Packings offer unmatched performance and sea
 - Die cast aluminum crankcase provide high strength, minimum weight and precision toleranc
 - Chrome-moly crankshaft give unmatched strength and surface hardness.
 - Oversized crankshaft bearings with greater loading capacity mean longer bearing life.
- Easy Maintenance**
 - Wat and is easily serviced without entering crankcase requiring less time and effort.

Unbreaker

Remove incorrect paragraph and/or line breaks

Scope
 Entire document Selection

Main options
 Incorrect paragraph breaks Incorrect line breaks

Manual correction | Automatic correction/highlighting | Extra options...

Search (click to search in active document)

Incorrect / uncertain breaks found: Show All

<input checked="" type="checkbox"/>	Standard Model	Paragr.
<input checked="" type="checkbox"/>	Pulley O.D.	Paragr.
<input checked="" type="checkbox"/>	...ats and springs provide corrosi...	Paragr.
<input checked="" type="checkbox"/>	...lushed Manifold permits externa...	Paragr.
<input type="checkbox"/>	..., MN 55449-4324 Phone (763) 780...	Paragr.
<input type="checkbox"/>	... (763) 780-2958 e-mail: techsup...	Paragr.
<input type="checkbox"/>	...Church Crookham FLEET, Hampshir...	Paragr.

Preview: Total: 17, uncertain: 3

Context: Standard Model

Action
Highlight certain corrected breaks in: Yellow

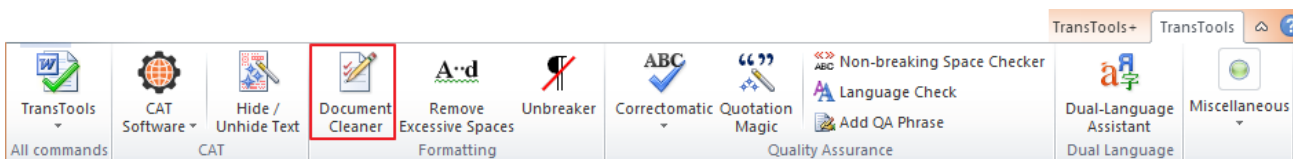
Highlight uncertain corrected breaks in: Red

Correct Document

? 3. Finally, click to remove checked breaks Close

6.2. Document reformatting and finding key issues.

To perform initial processing of the Word document, use the *Autoformat* command in Document Cleaner tool (<http://www.translatorstools.net/docs/transtools/doccleaner>).



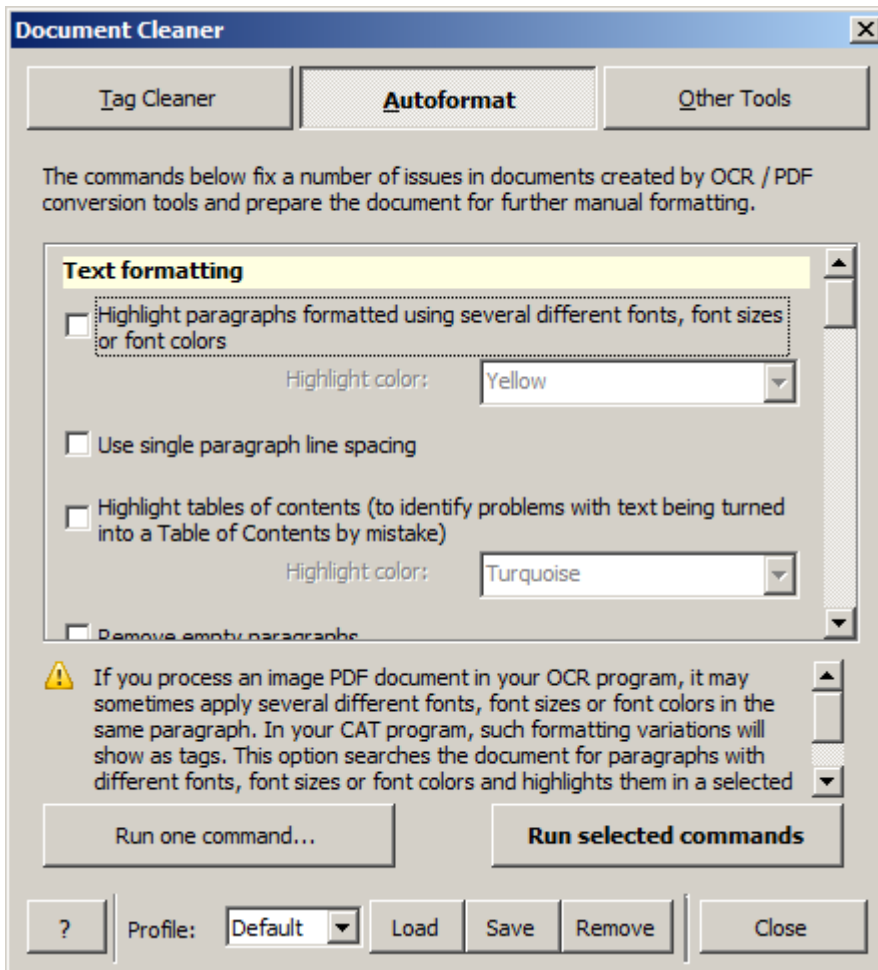


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

Once you open *Document Cleaner* dialogue, switch to Autoformat tab:



Below, you need to check the boxes for the actions that you want to apply to the current document, and also configure additional settings of these actions (see below). All Autoformat actions are explained below:

	<p>Highlight paragraphs formatted using several different fonts, font sizes or font colours</p> <p>This action applies colour highlighting to the paragraphs that contain text of varying font size, font colour or font name, e.g. when one word is 9 pt. while the rest of the paragraph is 10 pt. This issue is common during recognition or conversion of PDFs and scanned documents. When such paragraphs are imported into a CAT tool, they will contain tags due to changes in formatting.</p> <p><u>It is recommended to activate this action.</u></p>
--	---



<input checked="" type="checkbox"/> Use single paragraph line spacing	<p>Use single paragraph line spacing (for all paragraphs)</p> <p>This action applies single line spacing to all paragraphs.</p> <p><u>It is recommended to activate this option to speed up further formatting of the document.</u></p>
<input checked="" type="checkbox"/> Highlight tables of contents (to identify problems with text being turned into a Table of Contents by mistake) Highlight color: <input type="text" value="Turquoise"/>	<p>Apply colour highlighting to all automatic tables of contents (to detect fragments of text which was converted into a table of contents by mistake).</p> <p>Sometimes Finereader can erroneously detect a certain fragment of text as a table of contents (TOC). Most of the time a TOC will not be imported into a CAT tool, so you will not even see such text inside your CAT tool and will discover the non-translated text too late. Also, if document fields are updated (which can happen in Word automatically, or upon your instruction), the textual contents of such TOC will be lost.</p> <p>This action applies colour highlighting to all automatic TOC's. Normally, the very first TOC in a document is a "true" one and all the others are "false", but there are exceptions. Check such text fragments manually and if you find a "false" TOC, convert it to text using Document Cleaner (see below).</p> <p><u>It is recommended to activate this action.</u></p>
<input checked="" type="checkbox"/> Remove empty paragraphs	<p>Remove empty paragraphs</p> <p>This action removes empty paragraphs in the entire document. Later you will be able to insert empty paragraphs if needed or adjust the paragraph spacing through paragraph options.</p> <p><u>It is recommended to activate this action.</u></p>



<p><input checked="" type="checkbox"/> Remove text shading</p>	<p>Remove text shading (paragraph shading and text shading)</p> <p>This shading can be applied to text if <i>Highlight Low-confidence Characters</i> option is activated in Finereader Save Options, or if the source document was a scan containing highlighter marks. This shading causes extra tags inside CAT tools.</p> <p>This action clears colour shading of paragraphs or individual characters.</p> <p><u>It is recommended to activate this action.</u></p>
<p><input checked="" type="checkbox"/> Set default text spacing</p>	<p>Set default text spacing (scale, spacing, position and kerning)</p> <p>Non-standard text spacing settings are often applied to text when you convert scanned documents. This causes extra tags in CAT tools.</p> <p>This action applies standard text spacing settings that are typical of most documents.</p> <p><u>It is recommended to activate this action.</u></p>
<p><input checked="" type="checkbox"/> Clearly show where all tab characters are</p> <p>Highlight color: <input type="text" value="Bright green"/></p>	<p>Apply colour highlighting to tab characters between words</p> <p>If the original PDF file contained text justified on both sides, different words have different amount of whitespace between them. This causes Finereader to insert a tab character between words rather than a regular space. Inside a CAT tool, this causes incorrect segmentation (depending on CAT tool segmentation settings) or extra tags.</p> <p>This action applies colour highlighting to all tab characters between words.</p> <p><u>It is recommended to activate this action.</u></p>



<input checked="" type="checkbox"/> Remove soft (manual) hyphens	<p>Remove soft (manual) hyphens in the middle of words</p> <p>This action removes invisible hyphens in the middle of words. Normally, Finereader inserts such symbols if hyphens (used to force a break to a new line between word syllables) are visible in the source document.</p> <p><u>It is recommended to activate this action.</u></p>
<input type="checkbox"/> Left-align all tables	<p>Left-align all tables</p> <p>Normally Finereader aligns tables in the centre. When this action is activated, all tables in the document will be left-aligned.</p> <p><i>This action can be activated if needed.</i></p>
<input type="checkbox"/> Use "AutoFit to Window" option for all tables	<p>Use "Autofit to Window" option for all tables</p> <p>As a rule, tables in documents created with Finereader have different widths. You can use this action if all tables in your document must occupy the entire width of the document's working area.</p> <p><i>This action can be activated if needed.</i></p>
<input type="checkbox"/> Use consistent left/right cell margins for all tables Left/right cell margins: <input type="text" value="0,1"/> pt	<p>Use identical left and right margins for cells in all tables</p> <p>Documents created with Finereader can have very wide or very narrow table cell margins (i.e. the amount of space between text and the cell border). This action sets identical left and right margins for all table cells in the document.</p> <p><i>This action can be activated if needed.</i></p>



<p><input checked="" type="checkbox"/> Apply variable row height to fixed-height rows</p> <p>Options: <input type="text" value="Set variable height for all rows"/></p>	<p>Apply variable height for table rows if the row has fixed height</p> <p>As a rule, Finereader formats tables in such a way that table rows always keep a certain minimum height even if the text inside them shrinks. Besides, some PDF conversion programs can set fixed row height, so if the text does not fit inside such rows after translation, it will be visible only partially.</p> <p>This action applies the selected setting to all rows in all tables. The sub-option called “Set variable height for all rows” is best suited for initial processing of converted document – when it is selected and the action is run, all table rows start to adjust vertically to the amount of text in the cells.</p> <p><u>It is recommended to activate this action. Use “Set variable height for all rows” sub-option.</u></p>
<p><input type="checkbox"/> Apply transparent background to all tables</p>	<p>Apply transparent background to all tables</p> <p>If the source document uses different cell background colours or various tables have different background colours due to the quality of the scanned document, it makes sense to apply transparent background colour to all tables and later adjust it manually if required.</p> <p><i>This action can be activated if needed.</i></p>
<p><input type="checkbox"/> Use consistent cell alignment for all tables</p> <p>Alignment: <input type="text" value="Top Left"/></p>	<p>Use identical cell alignment in all tables</p> <p>Finereader often applies various text alignment settings to various cells (and often they are incorrect).</p> <p>To speed up document formatting it is sometimes advisable to apply identical text alignment to all cells in the document before changing it for a small number of cells later.</p> <p><i>This action can be activated if needed.</i></p>



<p><input type="checkbox"/> Apply consistent borders to all tables</p> <p>Borders: <input type="text" value="All borders"/> Border width: <input type="text" value="½ pt"/></p>	<p>Apply identical border settings for all tables</p> <p>If this action is selected all tables will be reformatted to have identical borders.</p> <p><i>This action can be activated if needed.</i></p>
<p><input checked="" type="checkbox"/> Use consistent page margins for all sections in the document</p> <p>Page margins: <input type="text" value="Narrow"/></p>	<p>Use identical page margins for all sections in the document</p> <p>Often, when you convert a PDF document, different pages in the converted Word document may have different page margin settings (left, right, top and bottom margins). When this action is activated all the pages in the document will have the same page margin settings.</p> <p><u>It is recommended to activate this action. In most cases you will want to select <i>Narrow</i> setting in order to apply narrow margins.</u></p>
<p><input checked="" type="checkbox"/> Minimize the number of section breaks in the document</p>	<p>Minimize the number of section breaks in the document</p> <p>In many cases, a document obtained through PDF conversion will contain several sections which can have different headers/footers, page settings, etc. This causes a lot of issues. This action removes excessive section breaks if pages have the same headers/footers, etc.</p> <p><u>It is recommended to activate this action.</u></p>
<p><input checked="" type="checkbox"/> Format all text in one text column</p>	<p>Format all text using a single text column</p> <p>Finereader prefers to lay out text that has several visible columns by using special column sections (called continuous sections) with several text columns. This causes a large number of text layout issues, so it is recommended to reformat the text into one visual column (default text appearance in a document) and change the layout manually later.</p> <p><u>It is recommended to activate this action.</u></p>



Use a single paper size for all pages

Page size:

Use a single paper size for all pages

In PDF documents different pages may have different size (A4, A3, non-standard sizes). When Finereader converts a page into Microsoft Word format, it tries to use the same page size as in the original PDF. This can cause issues during formatting.

This action applies the same paper size (e.g., A4 or A3) to all sections in the document.

It is recommended to activate this action. Use A4 setting in most cases.

Convert page breaks and/or column breaks to regular paragraph breaks

Page breaks Column breaks

Convert page breaks and/or column breaks to regular paragraph breaks

Converted documents often contain multiple page breaks or column breaks which make formatting more difficult.

This action replaces page breaks and column breaks with regular paragraph breaks.

It is recommended to activate this action. It is better to activate both sub-options – Page breaks and Column breaks.

Do not put text / tables / images inside frames

Remove frames around text, tables and images

Very often Finereader places certain fragments of text (especially tables and images) inside frames. The downside of frames is that the contents of the frame cannot wrap to the next page if they do not fit on one page. Also, frames make it more difficult to format the document.

This action removes frames, leaving the contents of the frames in the original place.

It is recommended to activate this action.



<p><input checked="" type="checkbox"/> Format invisible (1-point size) paragraphs using default font size</p>	<p>Format empty invisible (1-point size) paragraphs using default font size</p> <p>Occasionally Finereader and similar programs insert very small empty paragraphs (font size – 1 pt) into a Word document in order to create vertical spacing between objects. As a rule, such empty paragraphs are invisible and can make further document formatting more complicated.</p> <p>When this action is activated, standard font size (usually 8 pt) will be applied to such empty paragraphs.</p> <p><u>It is recommended to activate this action.</u></p>
<p><input checked="" type="checkbox"/> Make all images inline (to avoid floating images and simplify formatting)</p>	<p>Make all images inline</p> <p>Oftentimes, documents obtained with the help of Finereader and similar programs contain images with text flowing around them. In most cases such images are difficult to work with. This action makes all document images inline, i.e. the text does not flow around them.</p> <p><u>It is recommended to activate this action.</u></p>
<p><input checked="" type="checkbox"/> Clearly identify all textboxes</p> <p>Text Highlight Color Violet</p>	<p>Clearly identify textboxes</p> <p>Textboxes are a real nightmare for documents which were converted from PDFs. In real life textboxes are used when specific text must float above images and other text, but Finereader often uses textboxes in order to enforce more simple text layouts. From translation perspective textboxes are a big complication because they tend to shift their position away from the original location or expand as a result of translation, sometimes with unpredictable results.</p> <p>This action marks textboxes using colour highlighting or textbox background colour.</p> <p><u>It is recommended to activate this action. Use the settings shown on the left: select <i>Text Highlight Color</i> in the first list (to apply colour highlighting to the text inside the textbox) and then choose the highlight colour that is not used by the other actions.</u></p>

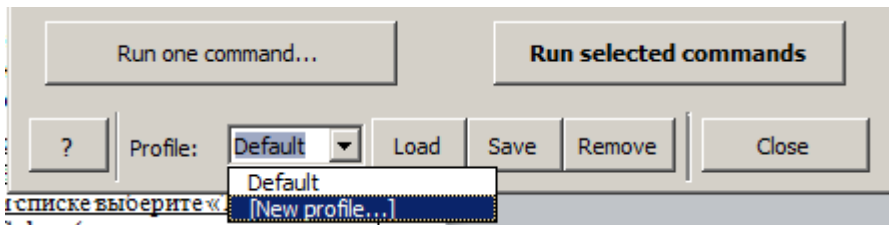


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

In order to process your first document you can set the checkboxes and settings as specified in the above table. Once the settings have been adjusted, you may want to save them in a profile in order to avoid this in the future. To create a new profile select *[New profile...]* in the list at the bottom of Document Cleaner dialogue, enter the profile name and then click *Save* button. The current settings of all Document Cleaner commands will be saved to the selected profile. In the future, when you want to run Autoformat command, simply select this profile from the list and click *Load* to apply the settings from the profile.



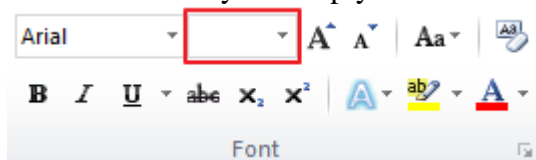
When all the appropriate actions have been activated and configured, click **Run selected commands** button. All selected actions will be executed for the current document.

Now go through the entire document and check it:

- Check the paragraphs that are fully highlighted in colour. Such paragraphs contain text formatted in different font, font size or font colour. [Note: if you move text out of a textbox using the appropriate Quick Action from inside Document Cleaner, the moved text will also be highlighted, even though it may contain the same font, font size and colour].

Figure 3.4: General Layout of Combined Pigging and BV Station 3.6 Block Valve Stations

When the paragraph contains text formatted with several different fonts or fonts sizes, you will be able to tell this by the empty font or font size field on the Home tab (see screenshot below):



As for font colour differences, it is often difficult to notice them with a naked eye, so it is simpler to select the entire paragraph and select Automatic font colour for it.

In some rare cases these differences are valid, but most of the time this is a Finereader's mistake. In the specific example shown above, we have encountered a common issue: Finereader merged two different paragraphs that had different font size, producing a single paragraph. Therefore, you must insert a paragraph break before "3.6" to separate these paragraphs.

In other cases you should assign an identical font, font size and font colour to the whole paragraph.

- Find textboxes (the text inside them will be highlighted with the colour you specified in Autoformat settings):



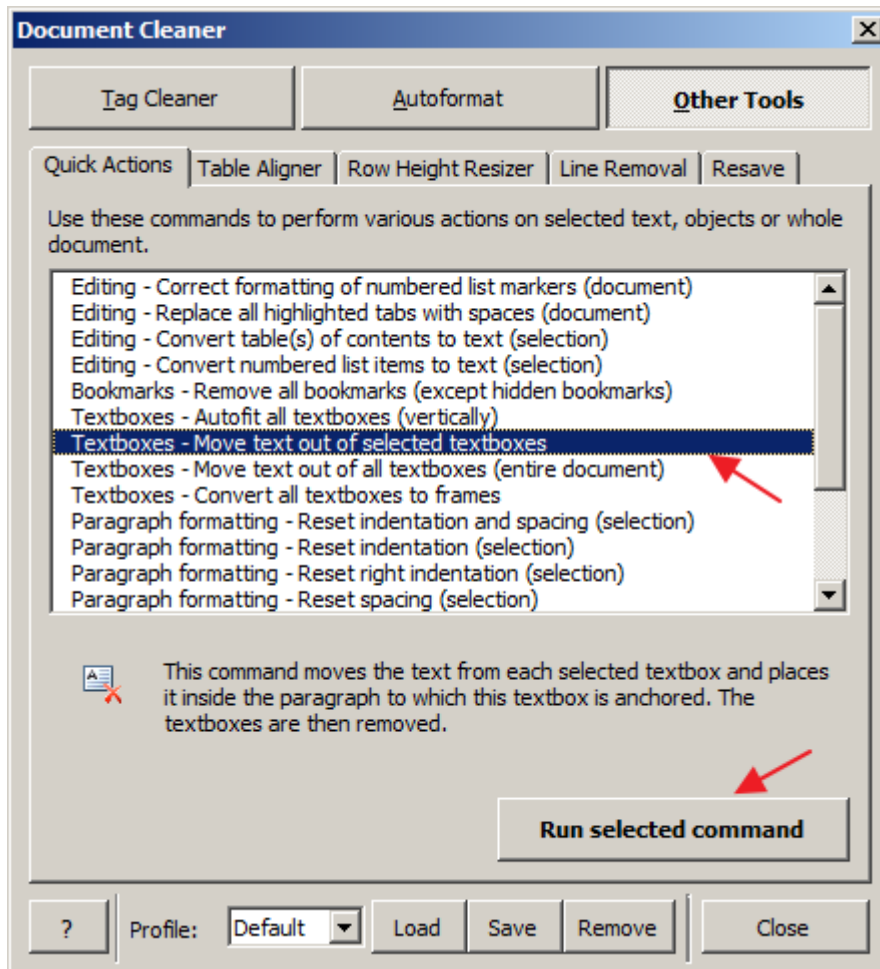
▪ Easy Maintenance

- Wet-end is easily serviced without entering crankcase, requiring less time and effort.
- Inlet and discharge valve assemblies interchange for easier maintenance without
- Preset packings mean no packing gland adjustment is necessary, reducing maint

WARNING

All systems require both a primary pressure regulating device (i.e., regulator, unloader) and a secondary pressure safety relief device (i.e., pop-off valve, safety valve). Failure to install such relief devices could result in personal injury or damage to the pump or to system components. CA PUMPS does not assume any liability or responsibility for the operation of a customer's high pressure system.

In most cases the contents of the textbox should be relocated into the body of the document and the textbox should then be removed. To do this, put the cursor inside the textbox and use a special command included in Document Cleaner:



When you do this, the text will be moved from inside the textbox and placed next to the paragraph to which the textbox was attached (“anchored”), and then the textbox will be removed. The moved text will be highlighted in yellow. The text may be moved to the wrong location, so check with the PDF and move it into the appropriate location if necessary.



Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatortools.net

- If the document contains tab characters which are highlighted in colour (see example below), check whether a regular space must be used instead of a tab character.

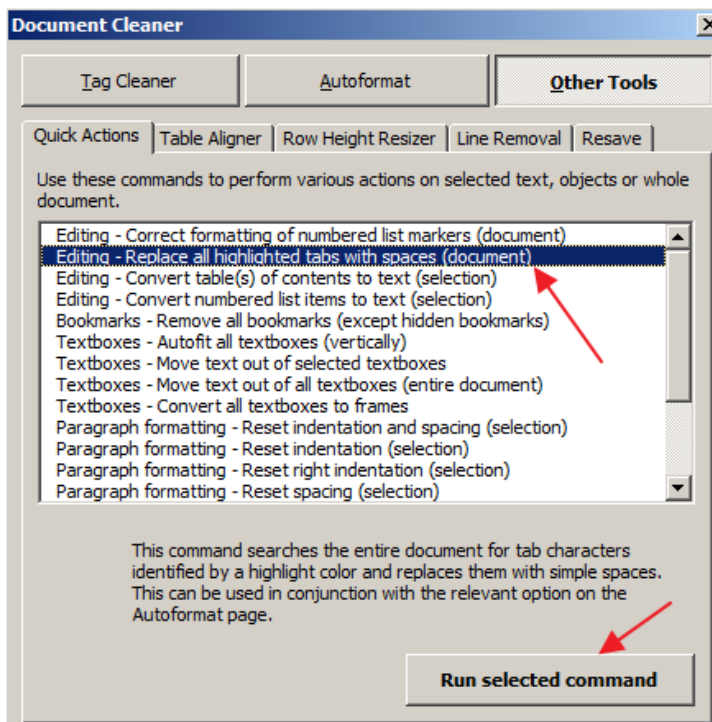
Administration building → including → first → aid / medical → facilities; ¶

Control building, including → station control → room → control → and telecom equipment ¶
room, engineering room, training room, electrical room document room, meeting room, lavatory and
sanitary installation; ¶

In this example regular spaces must be used between words instead of tab characters.

Manually replace the tab characters with spaces.

If you want to replace **all** highlighted tab characters with spaces in the entire document, you can speed up the operation using a special command included in Document Cleaner.



(before you use the command, double-check that all the tabs that are highlighted in colour must indeed be replaced with spaces)

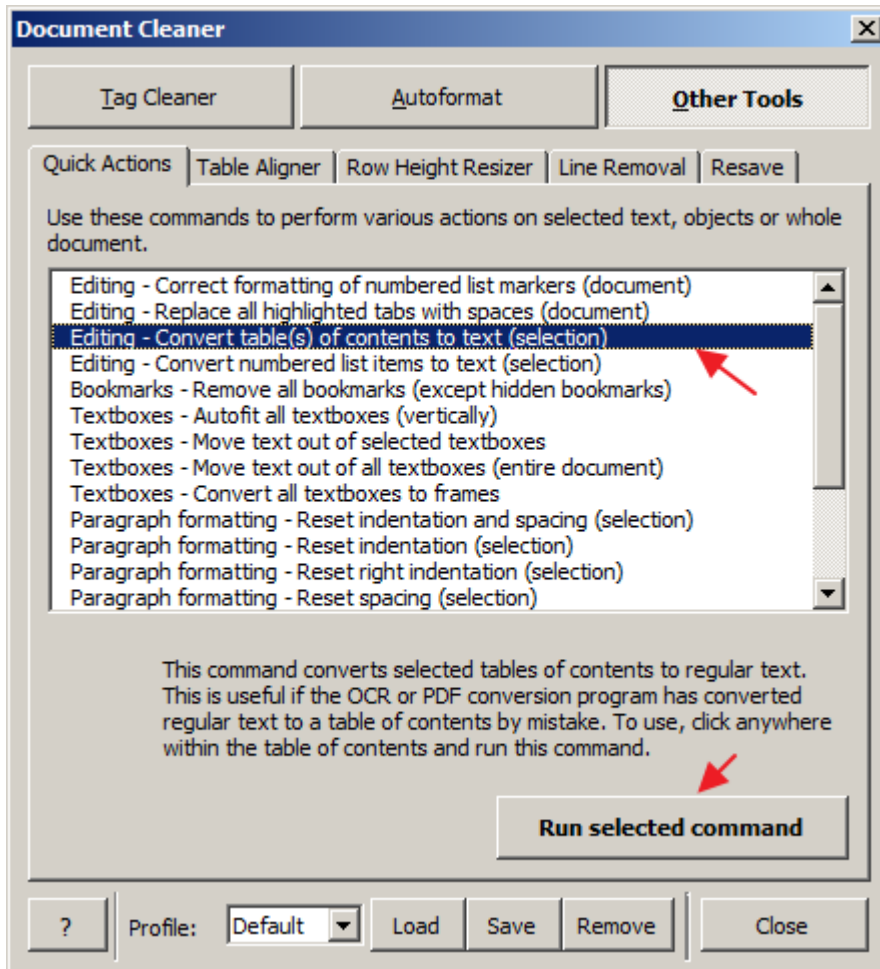


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

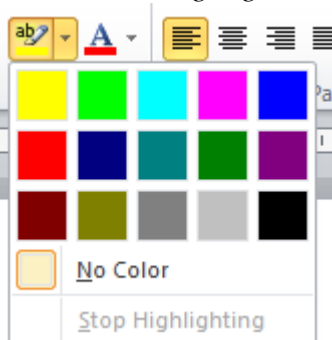
www.translatortools.net

- Check whether the document contains automatic tables of contents which look like normal text (they will have colour highlighting in accordance with the Autoformat settings that you used). If you found it, you can quickly convert it to regular text using a special command included in Document Cleaner. To do this, place the cursor in any place within such field and run the following command:



This completes the initial formatting step.

Clear colour highlighting in the entire document (select the entire text using Ctrl+A key combination and use the *Text Highlight Color* tool:

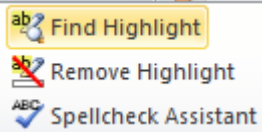


TransTools – Useful tools for every translator

www.translatortools.net

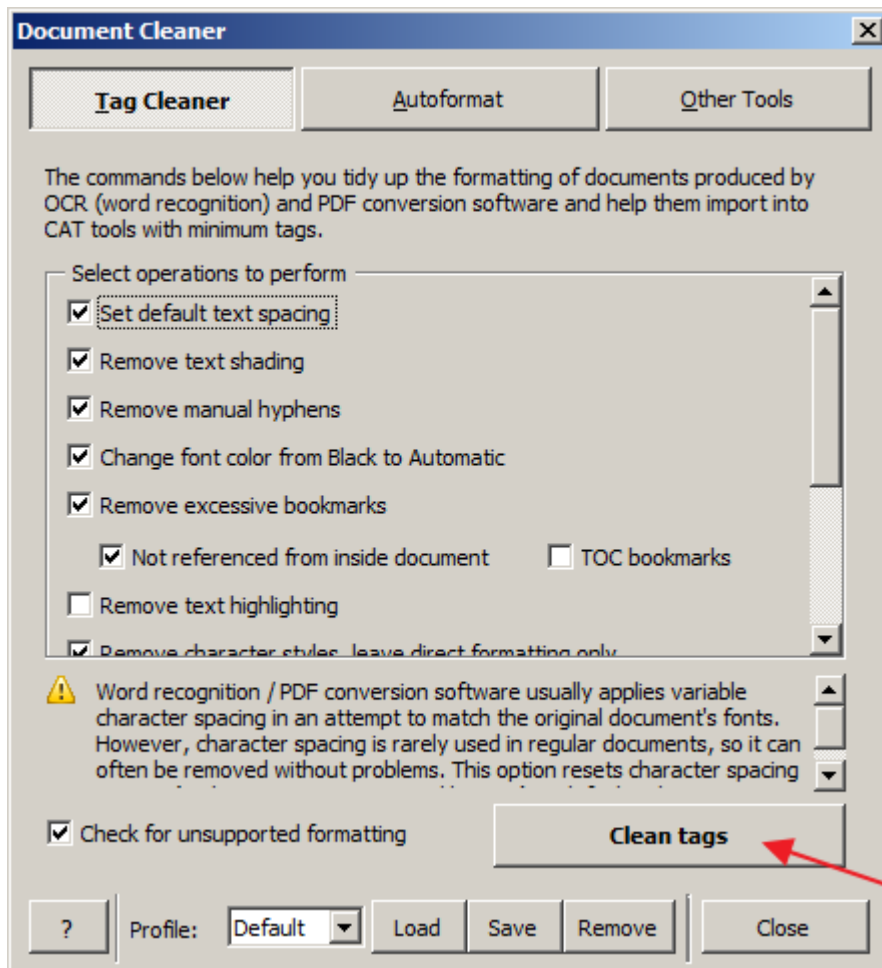


You can also clear colour highlighting using Clear Highlight tool from TransTools ribbon:



6.3. Removing redundant tags.

To remove redundant tags caused by invisible formatting differences, run Document Cleaner tool and click **Clean Tags** button under *Tag Cleaner* tab. You may leave default settings.



Here is the detailed information about Tag Cleaner options and the recommended settings:

<input checked="" type="checkbox"/> Set default text spacing	<p>Set default text spacing (scale, spacing, position and kerning)</p> <p>Non-standard text spacing settings are often applied to text when you convert scanned documents. This causes extra tags in CAT tools.</p> <p>This option applies standard text spacing settings that are typical of most documents.</p>
--	--



	<p><u>It is recommended to activate this option.</u></p>
<p><input checked="" type="checkbox"/> Remove text shading</p>	<p>Remove text shading (paragraph shading and text shading)</p> <p>Shading can be applied to text if <i>Highlight Low-confidence Characters</i> option is activated in Finereader Save Options, or if the source document was a scan containing highlighter marks. Such shading causes extra tags inside CAT tools.</p> <p>This option clears colour shading of paragraphs or individual characters.</p> <p><u>It is recommended to activate this option.</u></p>
<p><input checked="" type="checkbox"/> Remove manual hyphens</p>	<p>Remove manual (soft) hyphens</p> <p>This option removes invisible hyphens in the middle of words. Normally, Finereader inserts such symbols if hyphens (used to force a break to a new line between word syllables) are visible in the source document.</p> <p><u>It is recommended to activate this option.</u></p>
<p><input checked="" type="checkbox"/> Change font color from Black to Automatic</p>	<p>Change font colour from Black to Automatic</p> <p>Automatic and Black are two different colours, so every change between such colours is converted to tags by CAT tools. These two colours may occur together in documents converted by PDF converters or assembled from various sources.</p> <p>Use this option to change the 'black' colour to 'automatic'.</p> <p><u>It is recommended to activate this option.</u></p>
<p><input checked="" type="checkbox"/> Remove excessive bookmarks</p> <p><input checked="" type="checkbox"/> Not referenced from inside document <input type="checkbox"/> TOC bookmarks</p>	<p>Remove excessive bookmarks</p> <p>Finereader inserts bookmarks in converted Word documents to enable Table of Contents functionality and for some other reasons. Bookmarks cause tags inside your CAT tool so they should be removed unless they are necessary.</p> <p>This option removes certain bookmarks depending on the following sub-options:</p> <p><i>Not referenced from inside document</i> option</p>



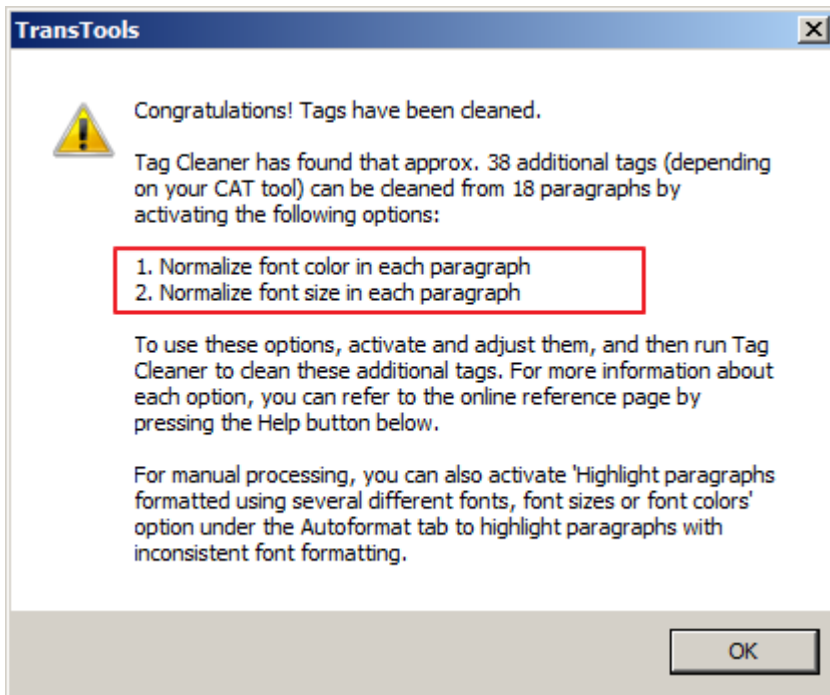
	<p>removes bookmarks which are not referenced from any hyperlinks or fields (including TOC hyperlinks).</p> <p>TOC bookmarks option removes bookmarks which are created by Microsoft Word when a TOC is generated in order to enable jumping from the TOC to specific sections of the document. In most cases, TOC bookmarks can be removed before translation and re-created after translation by updating the TOC.</p> <p><u>For documents converted with Finereader, it is recommended to activate ‘Remove excessive bookmarks’ + ‘Not referenced from inside document’ options.</u></p>
<input type="checkbox"/> Remove text highlighting	<p>Remove text highlighting</p> <p>When text highlighting is applied to parts of a paragraph (as opposed to whole paragraphs), it is usually imported into CAT tools as tags. This option removes all text highlighting from text.</p> <p><u>This option can be activated if necessary.</u></p>
<input checked="" type="checkbox"/> Remove character styles, leave direct formatting only	<p>Remove character styles, leave direct formatting only</p> <p>Most documents produced by Finereader will contain character styles on top of paragraph styles, causing tags in CAT tools. If you use this option, the command will strip character styles, leaving equivalent direct formatting untouched.</p> <p><u>It is recommended to activate this option.</u></p>
<input type="checkbox"/> Accept tracked changes and switch off change tracking	<p>Accept tracked changes and switch off change tracking</p> <p>Documents produced by Finereader will not have tracked changes, however tracked revisions in general may cause excessive tags in some CAT tools. It is recommended to accept them before translating the document. When you use this option, the command will accept all tracked changes and turn off change tracking.</p> <p><u>This option can be activated if necessary.</u></p>
<input type="checkbox"/> Normalize font color in each paragraph	<p>Normalize font colour in each paragraph</p> <p>When processing PDFs, especially scanned</p>



	<p>PDFs, a single paragraph may contain several different text colours which are similar. This causes tags in CAT tools. When you use this option, Tag Cleaner will apply the first text colour to other parts of the paragraph if these colours are very similar.</p> <p><u>This option can be activated if necessary, e.g. if you see a warning about remaining tags after the first run of Tag Cleaner.</u></p>
<p><input type="checkbox"/> Normalize font size in each paragraph</p> <p>Fix paragraph font size differences of 1 pt or smaller</p>	<p>Normalize font size in each paragraph</p> <p>When processing PDFs, especially scanned PDFs, a single paragraph may contain several different font sizes. This causes tags to appear when the document is imported into a CAT tool. When you use this option, the command will apply the first font size to the rest of the paragraph (subject to the sub-options).</p> <p><u>This option can be activated if necessary, e.g. if you see a warning about remaining tags after the first run of Tag Cleaner.</u></p>
<p><input type="checkbox"/> Normalize font in each paragraph</p>	<p>Normalize font in each paragraph</p> <p>When processing PDFs, a single paragraph may contain several different fonts (e.g., Calibri alongside Arial). This causes tags to appear when the document is imported into a CAT tool. When you use this option, Tag Cleaner will apply the first font face to the rest of the paragraph (except for symbol fonts).</p> <p><u>This option can be activated if necessary, e.g. if you see a warning about remaining tags after the first run of Tag Cleaner.</u></p>
<p><input type="checkbox"/> Replace special symbols with regular characters where appropriate</p>	<p>Replace special symbols with regular characters where appropriate</p> <p>Some documents may contain special symbols from Symbol, Wingdings and some other symbol fonts. CAT tools display these symbols as tags. When you use this option, certain symbols which are also present in standard fonts are replaced with regular text, reducing the number of tags.</p> <p><u>This option can be activated if necessary.</u></p>



When *Tag Cleaner* command completes, you may see a warning that the command found paragraphs whose text is formatted using different font, font size or colour (such differences will also cause extra tags inside your CAT tool):



This means that you should run Autoformat command again (see Step 6.2), activating *Highlight paragraphs formatted using several different fonts, font sizes or font colors* option only. Then reformat such paragraphs.

You can also follow the advice given in the warning and activate additional Tag Cleaner options (*Normalize font color*, *Normalize font size* and *Normalize font*) and then re-run Tag Cleaner in order to “normalize” the formatting of such paragraphs automatically.

7. Step 7 – Manually format Word document

During this step, do the following:

- Put text into headers/footers, inserting section breaks and unlinking headers/footers from previous sections if necessary.
- Merge multi-page tables.
- Correct numbering in automatically numbered lists (if any automatic numbering is inconsistent with the PDF).
- Check document for merged paragraphs and split them if appropriate (see notes below).
- Put text back into columns where appropriate.
- Check bold, italic, underline formatting of paragraphs and individual phrases/words.
- Check text colours and background colours.



- Adjust spacing between paragraphs.
- Adjust paragraph and list indentation, tab positions, alignment (left/centre/right/justified).

Note 1: Checking document for merged paragraphs.

Occasionally, Finereader combines several paragraphs into a single paragraph.

Example:

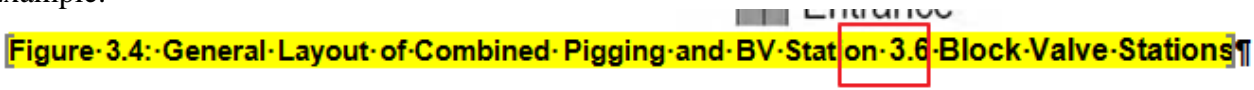


Figure 3.4: General Layout of Combined Pigging and BV Station 3.6 Block Valve Stations

Usually such merged paragraphs will be highlighted in yellow when you run Autoformat option called “Highlight paragraphs formatted using several different fonts, font sizes or font colours” (see above).

To identify such text more easily, you can also select the entire text within the document and set paragraph’s right indentation to 0.

Pay particular attention to bulleted and numbered lists, since Finereader tends to combine several list items into one.

Note 2: Finding all text fragments which must be moved into headers/footers, and moving them.

Move the text that corresponds to original document’s headers and footers (repeated at top/bottom of pages) into the header and footer of the appropriate section.

Once the text has been moved into the headers or footers, find numbers that denote the current page number and total number of pages and replace them with updatable fields using *Quick Parts – Field* under *Insert* tab. Use *Page* field to insert a page number field and *NumPages* field to insert the total number of pages. Use *SectionPages* field to insert the number of pages in the current section.

Note 3: Merging multi-page tables.

Sometimes, a table in the original PDF may span several pages, and such table will be split into several tables in the generated Word document. Merge these tables by removing the paragraph breaks that separate them.

If the top rows of the table were repeated on each page, you will need to remove such rows in all tables except the first one, and then activate the option called “Repeat as header row at the top of each page” in Table Properties for the first row(s) of the table.

If the column borders within the combined table are not aligned (see example below), you will need to adjust them:



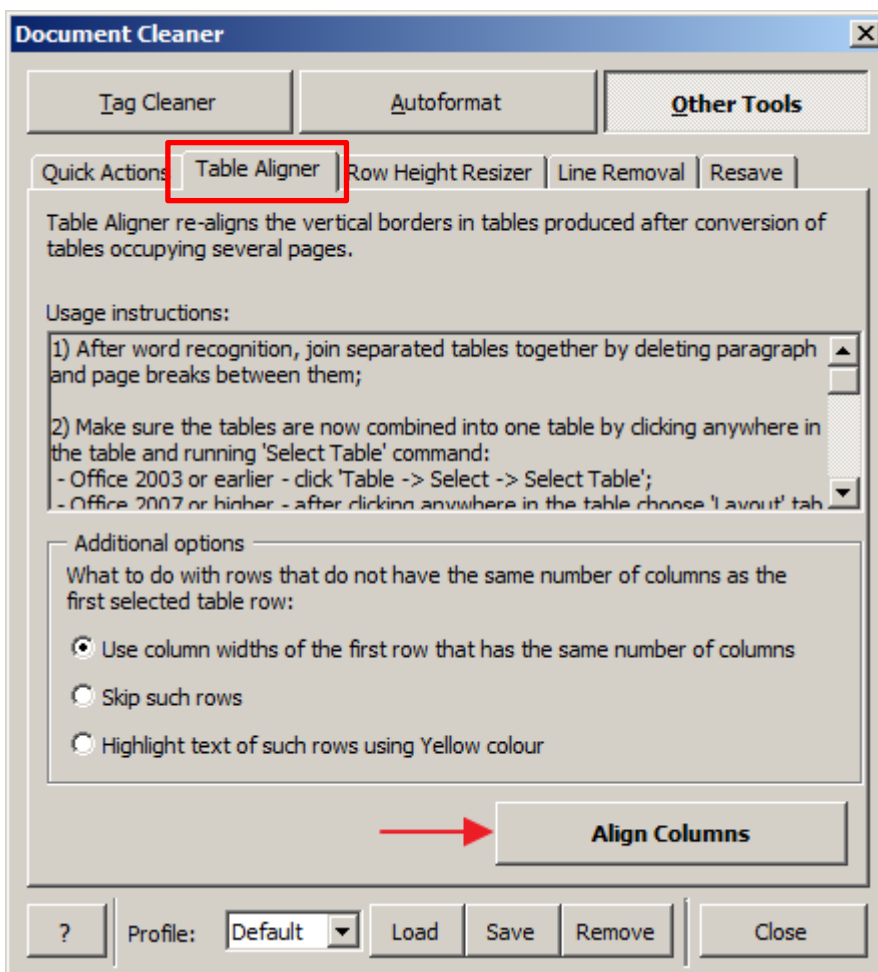
Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

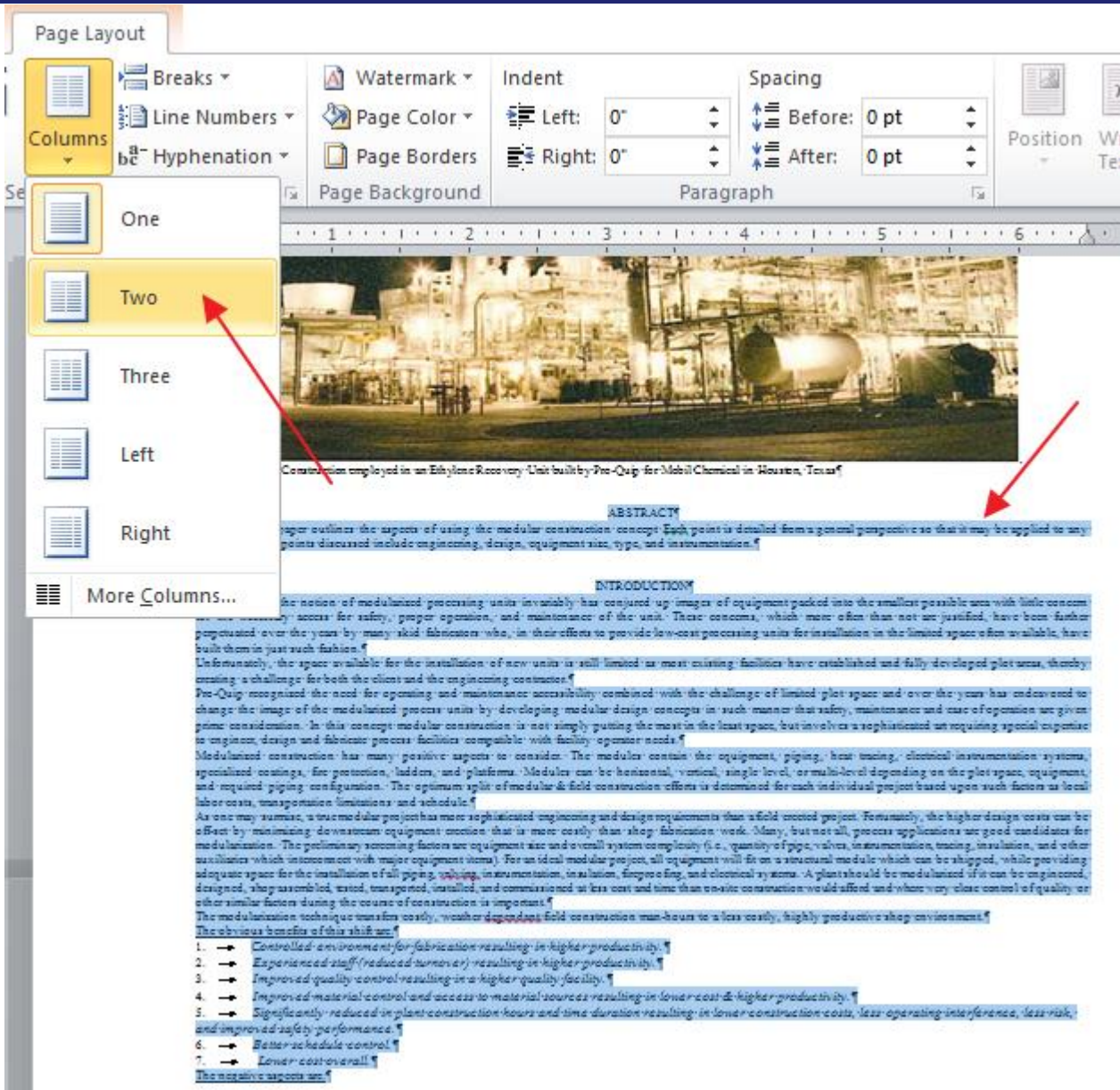
V=25-м³; ШxВxД – 2,4м x 2,9м x 9,9м; Ø3000мм; P _{раб.} =0,07-МПа; T _{раб.} =60°C; P _{расч.} =0,6-МПа; T _{расч.} =100°C	ООО «Саулит-Инжиниринг»	шт.	1	7500	Тип НГС-25
H=2,5 кгс/см²; Q=100-м³/час; Nэ.дв.=37 кВт.	-	шт.	2	-	Марка винтовой насос.

You can achieve this using the special *Table Aligner* tool included in Document Cleaner. Select the entire combined table and run Table Aligner:



Note 4: Putting text back into columns.

If specific text must be laid out in several columns flowing into each other, select such text and then choose the appropriate number of columns from *Columns* list under *Page Layout* tab.





This will convert the text into multi-column text and place continuous section breaks before and after the text:



Photo: Modular Construction employed in an Ethylene Recovery Unit built by Pre-Quip for Mobil Chemical in Houston, Texas

Section Break (Continuous)

ABSTRACT

The following paper outlines the aspects of using the modular construction concept. Each point is detailed from a general perspective so that it may be applied to any industry. Major points discussed include engineering, design, equipment size, type, and instrumentation.

INTRODUCTION

Over the years the notion of modularized processing units inevitably has conjured up images of equipment packed into the smallest possible area with little concern for the necessary access for safety, proper operation, and maintenance of the unit. These concerns, which more often than not are justified, have been further perpetuated over the years by many skid fabricators who, in their efforts to provide low-cost processing units for installation in the limited space often available, have built them in just such fashion.

Unfortunately, the space available for the installation of new units is still limited as most existing facilities have established and fully developed plot areas, thereby creating a challenge for both the client and the engineering contractor.

Pre-Quip recognized the need for operating and maintenance accessibility combined with the challenge of limited plot space and over the years has endeavored to change the image of the modularized process units by developing modular design concepts in such manner that safety, maintenance and ease of operation are given prime consideration. In this concept modular construction is not simply putting the most in the least space, but involves a sophisticated art requiring special expertise to engineer, design and fabricate process facilities compatible with facility operator needs.

Modularized construction has many positive aspects to consider. The modules contain the equipment, piping, heat tracing, electrical instrumentation systems, specialized coatings, fire protection, ladders, and platforms. Modules can be horizontal, vertical, single level, or multi-level depending on the plot space, equipment, and required piping configuration. The optimum split of modular & field construction efforts is determined for each individual project based upon such factors as local labor costs, transportation limitations and schedule.

As one may surmise, a true modular project has more sophisticated engineering and design requirements than a field erected project. Fortunately, the higher design costs can be offset by minimizing downstream equipment erection that is more costly than the fabrication work. Many, but not all, process applications are good candidates for modularization. The primary screening factors are equipment size and overall system complexity (i.e., quantity of pipe, valves, instrumentation, tracing, insulation, and other auxiliaries which interconnect with major equipment items). For an ideal modular project, all equipment will fit in a structural module which can be shipped, while providing adequate space for the installation of full piping, valving, instrumentation, insulation, fireproofing, and electrical systems. A plant could be modularized if it can be engineered, designed, shop assembled, tested, transported, installed, and commissioned as fast as a field erected construction would afford and where very close control of quality or other similar factors during the course of construction is important.

The modularization technique transfers costly, weather dependent field construction man-hours to a less costly, highly productive shop environment. The obvious benefits of this shift are:

1. → Controlled environment for fabrication resulting in higher productivity.
2. → Experienced staff (reduced turnover) resulting in higher productivity.
3. → Improved quality control resulting in a higher quality facility.
4. → Improved material control and access to material sources resulting in lower cost & higher productivity.
5. → Significantly reduced in plant construction hours and time duration resulting in lower construction costs, less operating interference, less risk, and improved safety performance.
6. → Better schedule control.
7. → Lower cost overall.

The negative aspects are:

1. → Engineering and design must be executed earlier in the project.
2. → Potentially increased structural steel requirements.

In successful modular projects, the incremental engineering, design, and structural steel costs are offset by the reduced cost of assembly, painting, piping, electrical, insulation, refractory, and other work being performed in the fabrication facility instead of the field. Transportation of the large module is usually competitive with the combined costs of shipment of individual items. As controls more easily facilitated, complicated systems are excellent candidates for modularization.

Note 5: Removing colour highlights

Clear colour highlighting if it is present by using *Remove Highlight* command (TransTools tab, Miscellaneous group) or using *Text Highlight Color* tool (Home tab, Font group).

8. Step 8 – Handle images containing text

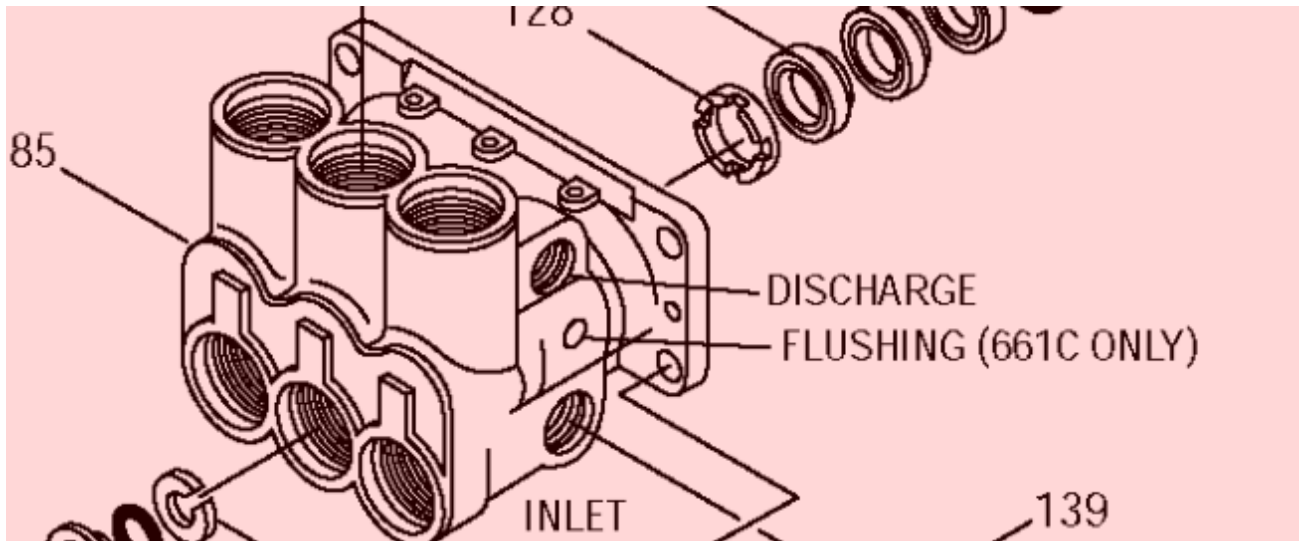
There are several different ways to translate images that contain text on them:

Option 1. Overlaying textboxes above image in Microsoft Word

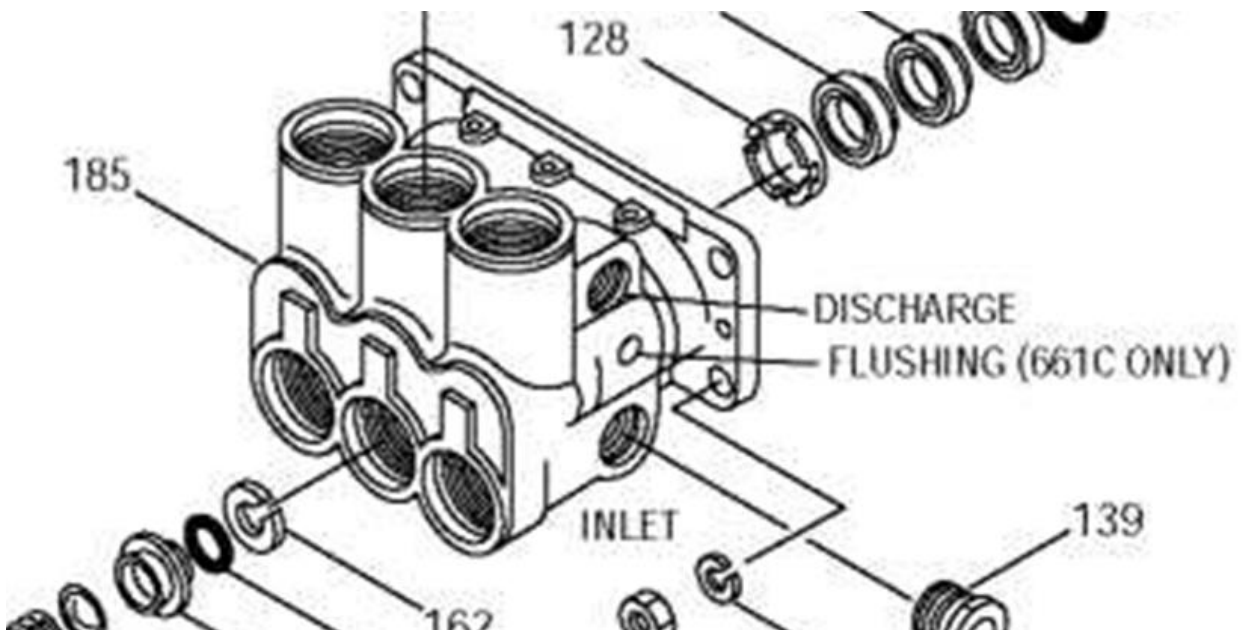
Create a single picture area around the image containing text, without creating text areas for the text on the image. This will create an image in Microsoft Word. The text written on the image will not be editable.



In ABBYY Finereader:



In Microsoft Word:



Then, overlay textboxes on top of each image, obscuring the text that is written on the image:

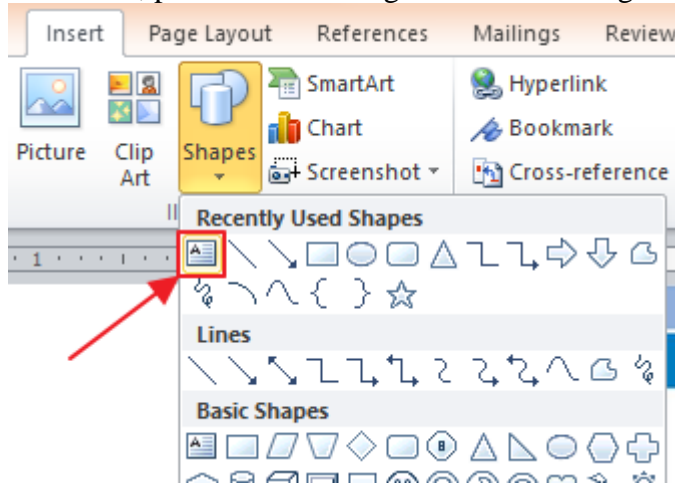


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

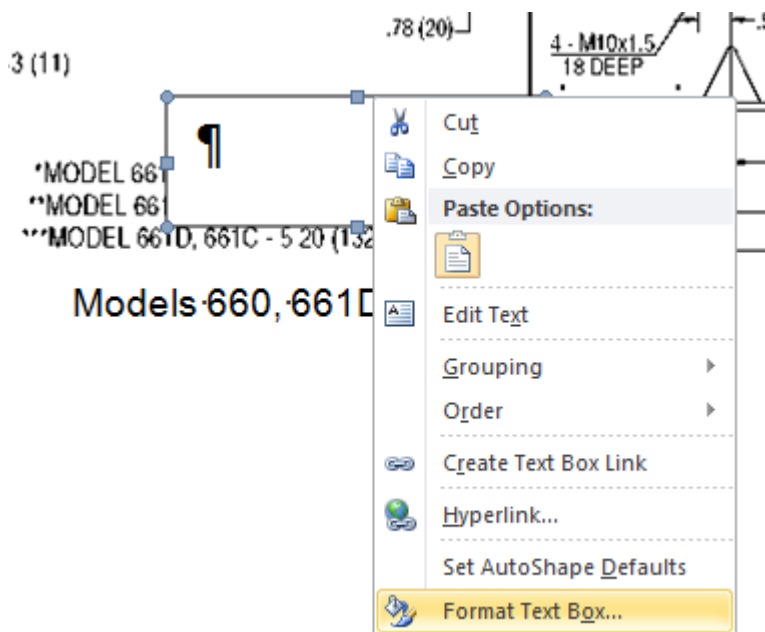
TransTools – Useful tools for every translator

www.translatorstools.net

To do this, place the cursor right before the image and insert a textbox:



Specify an invisible border and narrow margins for the textbox:



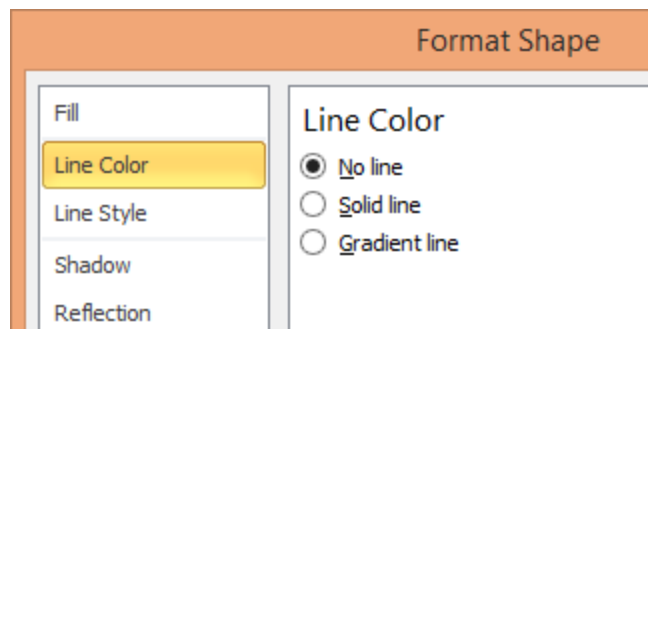


Conversion of PDFs/scanned images to Microsoft Word format using ABBY Finereader

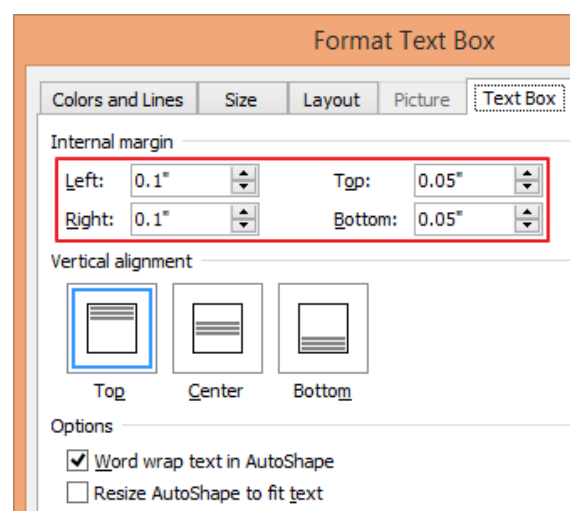
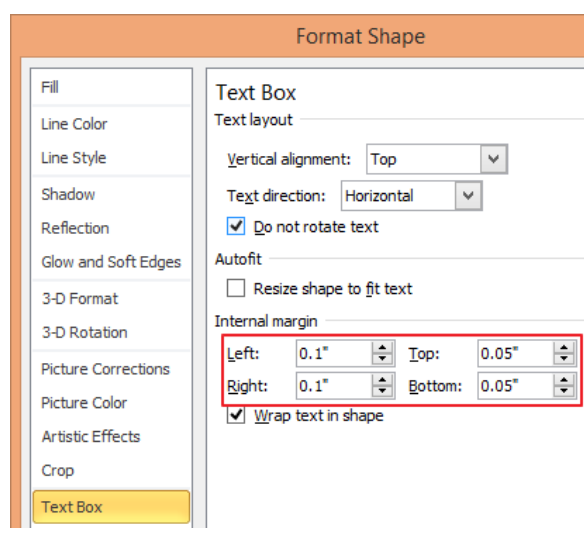
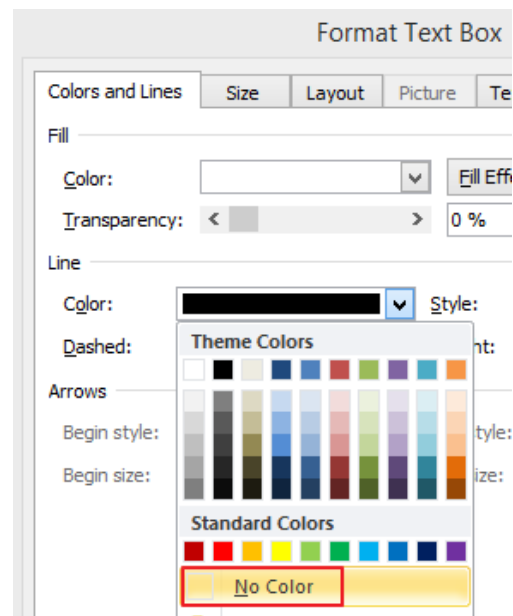
TransTools – Useful tools for every translator

www.translatortools.net

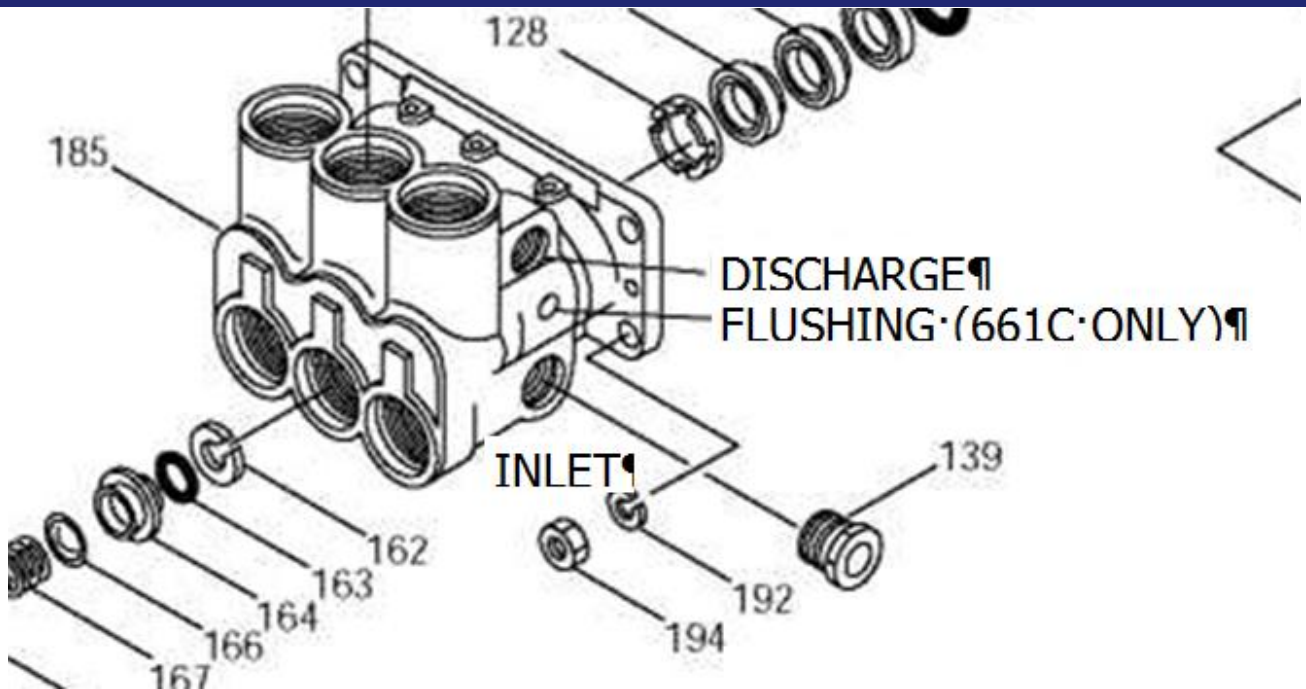
Word 2007, Word 2010, Word 2013, Word 2016, Office 365, DOCX format with compatibility mode switched off



Word 2003, Word 2000, other Word versions with compatibility mode switched on



Once the first textbox has been inserted and its formatting (including font properties) has been specified, make copies of this textbox to create new text labels for this and other images.



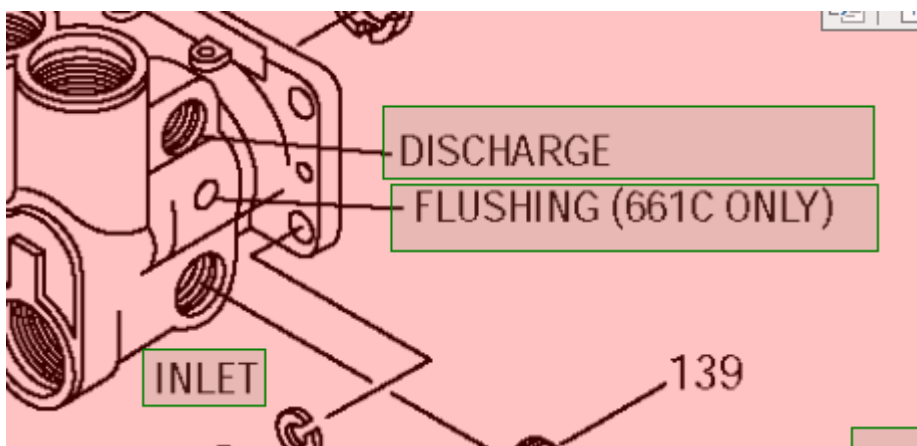
Pros: the text will be imported into the CAT tool and translated together with the document, and only minimum formatting will be required after translation.

Cons: Extra DTP before translation begins, some post-formatting after the document is translated to fit the text.

Option 2. Overlay text areas above picture area in Finereader

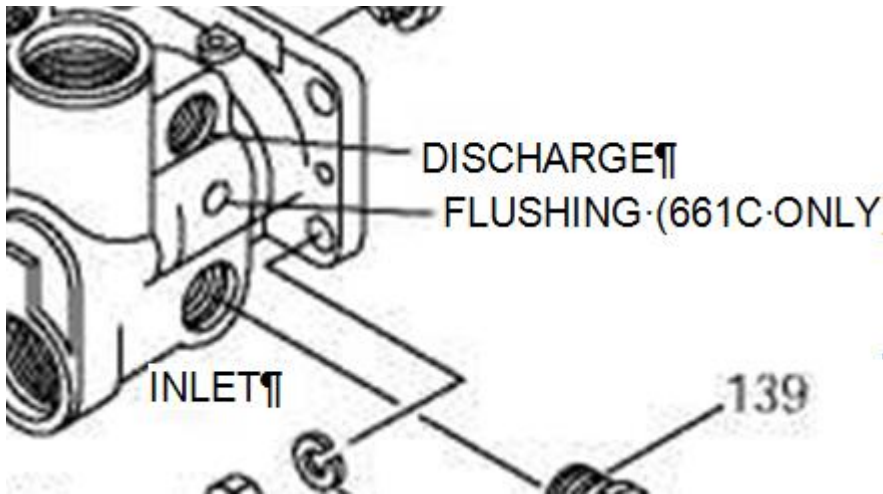
In Finereader, place text areas above picture area. This text will be converted to textboxes floating above the image.

In ABBYY Finereader:

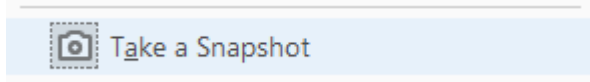




In Microsoft Word:



However, in this case Finereader will remove the text from the image in the generated Word document, so there will be no text written on the image itself. At the same time, the text will be placed inside a frame, and the frame will be fixed to the top of the page and it will be difficult to work with this image if it is placed in the translated document. Therefore, translate the individual text labels, resizing the frames around them if necessary and making sure that they are positioned correctly. Once all such images are translated, save the Word document as a PDF file, open the PDF file in Adobe Reader, zoom in so that the image occupies the entire screen, click Edit > Take a Snapshot, and then select the area around the image to copy the image to the clipboard:



After that paste the image in the translated Word document.

Note: if you have to use Autoformat command on a Word document that contains images prepared as described above, make sure that you untick “Do not put text / tables / images inside frames” option, as this option will cause the text labels to be re-formatted.

Pros: the text will be imported into the CAT tool and translated together with the document, and only minimum formatting will be required after translation.

Cons:

- 1) The translated Word document will contain non-editable text over the image.
- 2) This approach cannot be used if the text overlaps other image contents, since Finereader removes the image contents under the text areas.
- 3) Extra care must be taken to position text above image, using the source PDF as reference, since no text will remain on the image.
- 4) Extra DTP before translation begins, some post-formatting after the document is translated to fit the text.



Option 3. Translate text in TransPDF or PDF editor

If the text on the image can be selected within your PDF viewer, you can use Infix Editor or TransPDF web service to translate the pages that contain such images. Then open the “translated” PDF in Adobe Reader and copy large-scale screenshot to Microsoft Word using *Take A Snapshot* command (*Edit – Take a snapshot* menu). This approach does not work if the text is “embedded” into the image, i.e. it cannot be selected in your PDF viewer.

Pros: Ability to translate rotated text, compact text, etc.

Cons:

- 1) The translated Word document will contain non-editable text over the image.
- 2) A single text label in a PDF document may be split into several textboxes, so it may be cumbersome to edit and format the text.
- 3) Not all images in an editable PDF document may contain editable text.

Option 4. Translate images using image editor

Translate images using Paint.net (see <http://www.translatortools.net/blog/articles-and-workflows/164-article-translating-images-paint-net>) or another image editing program. This approach takes longer, but it allows you to translate very compact text or text written against complex backgrounds (e.g. gradient background).

Pros: better control of the text, ability to edit text that has complex backgrounds. Some image editors like Photoshop have numerous options to control text formatting and layout.

Cons:

- 1) If typos or errors are found in the translation, the image will need to be modified and re-inserted into the document.
- 2) To translate the text, you must put it in a separate document or inside a dual-column table below the image in the source document.
- 3) More time required than the textbox workflows described above.
- 4) Several tools are required to perform this operation.

9. Step 9 – Translate

- Before importing the document into the CAT tool, write down all the things that may need attention after the translated document is exported from the CAT tool, so that you can use this list to expedite checking of the translated document and also import the document into the CAT tool correctly:
 - Presence of textboxes: they will need to be checked for partially invisible text and resized/reformatted if necessary.
 - Presence of table(s) of contents: they will need to be updated in the final document.



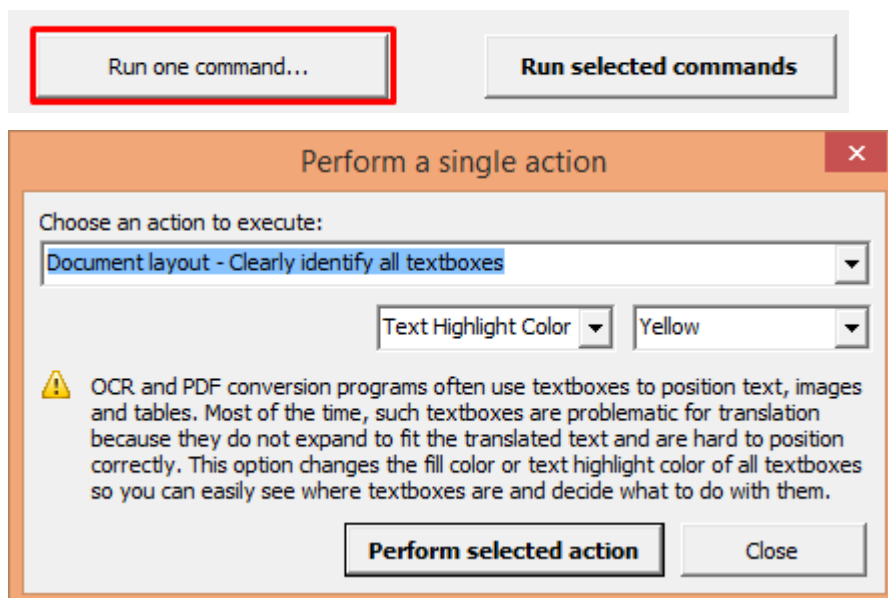
- Presence of tab characters which are used to align text into columns: you may need to alter the segmentation settings of your CAT tool during importing so that the CAT tool breaks the text into separate segments at these tab characters.
- Import the document into your CAT tool to check for excessive tags and incorrect segmentation. If required, correct these issues in the source document and re-import it.

10.Step 10 – Pre-flight check

At this step, do the following:

1. Visually check the document against the PDF.
2. Update tables of contents and perform other actions you normally do in final Word documents before delivery.
3. Check whether all text fits inside textboxes (they will usually be images).

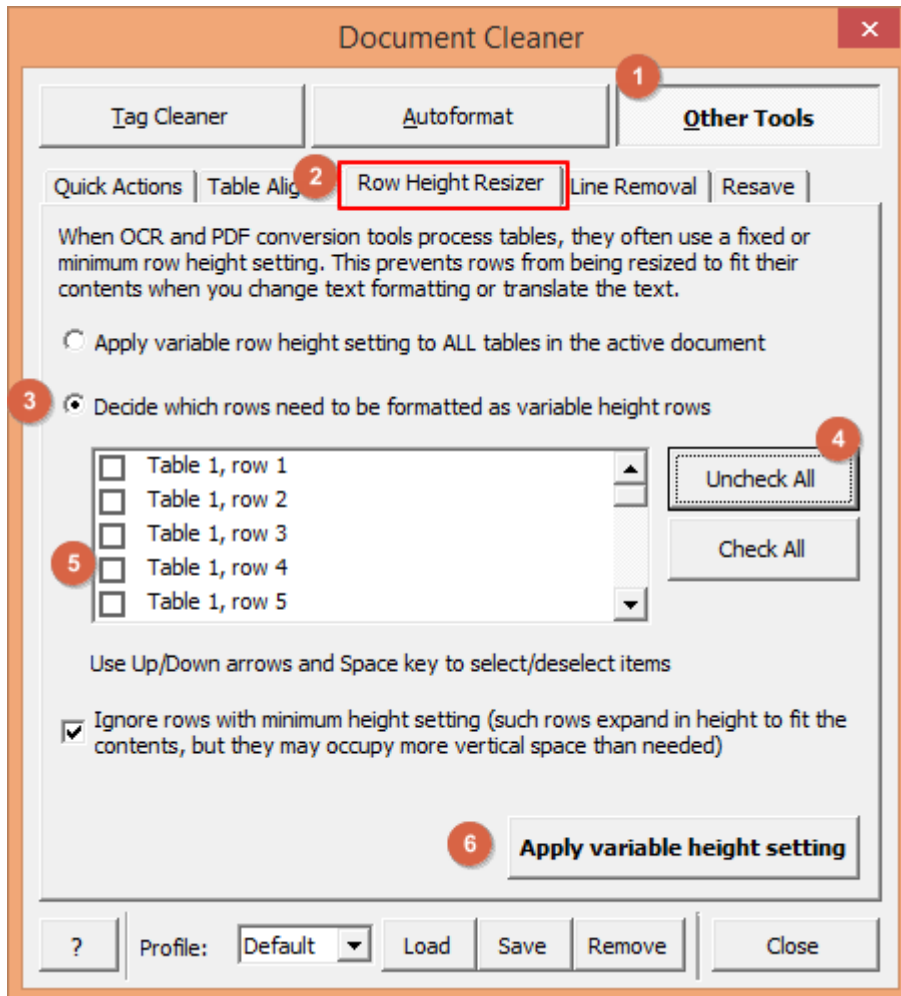
You can use Document Cleaner’s Autoformat command as explained above to highlight the contents of all textboxes in order to make them more prominent, but make sure that you run only “Clearly identify textboxes” command as follows:



4. Check the document for table rows that have fixed height, preventing the text from being fully visible inside such rows.



To find such rows easily, use the following tool available in Document Cleaner:



Here is what you need to do:

- Select Other Tools tab.
- Select Row Height Resizer tab below.
- Select “Decide which rows need to be formatted as variable height rows” option. This will search the document for all fixed-height rows.
- Click Uncheck All button.
- Go through the list, selecting each entry. You will see the corresponding row selected inside the document window. If you think this row needs to be converted to a variable-height row, tick the checkbox.
- When you are finished reviewing, click Apply Variable Height Setting button to re-format the ticked rows.



III. Tips: Preparing accurate quotes for PDF translation

Your quote will include two things: the word count and the additional effort for conversion and preparation before and after translation. Here is a suggested workflow for creating quotes for PDF translation:

1) Begin by checking the type of PDF document.

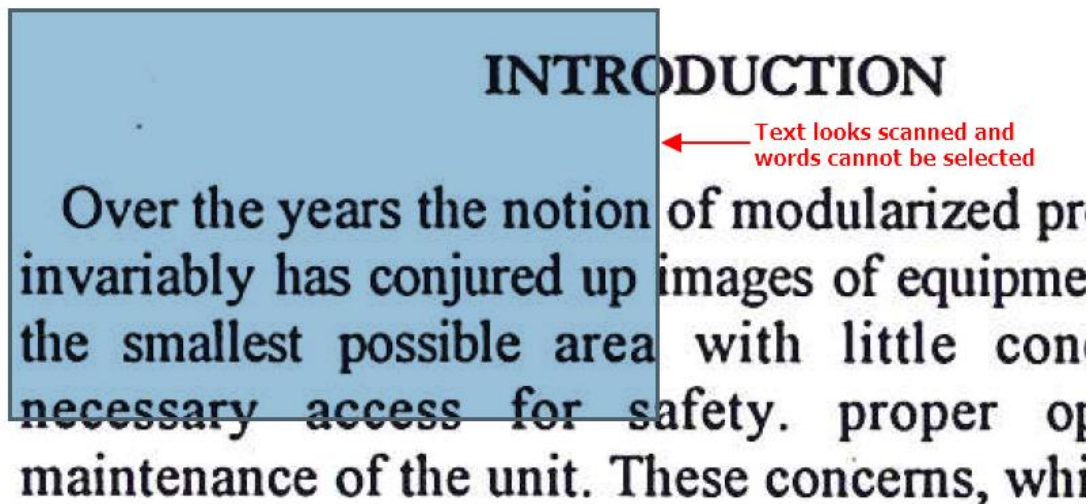
First, you need to determine the type of the PDF document. There are three types of PDF documents:

- **Text PDFs.** These files contain text that you can select and copy to another application. How can you recognize a text PDF? First, you can select individual words. And second, when you zoom in on the text, you can see smooth outlines of characters.

1.4 Scope of Document

The scope of this

- **Image PDFs.** These files are created by photocopying paper documents. How can you recognize an image PDF? First, when you zoom in on the text, you will see jagged text outlines, uneven lines of text and visual distortions. And second, you cannot select individual words.





- **Searchable image PDFs.** These files are created by processing image PDFs with optical character recognition (OCR) software that automatically recognizes the text and inserts the recognized text behind the page images. This way, it is possible to search for text in such PDFs and even select the text, but the text may contain recognition errors. How can you recognize such PDFs? They look like image PDFs but it is possible to select individual words.

10. Fælles indkøb

← Text is low-quality when zoomed but words can be selected

Parterne er enige om, at det kan være hensigt
Parters vegne, hvor der bl.a. er mulighed for s
med nærværende samarbejdsaftales omfang.

To make it easier to determine if the PDF is an image PDF or searchable image PDF, you can use a free application called PDF Image Extractor (<http://www.translatorstools.net/products/miscellaneous/pdfimageextract>). This program extracts all the images from the PDF into a separate folder. You can then view the extracted images. If you find that every image is a photocopied page, the PDF is clearly an image PDF or a searchable image PDF.

2) Open the PDF in Finereader's OCR Editor

Convert Documents



Open in OCR Editor
Advanced conversion and verification

3) Recognize all pages automatically

Select all pages, right-click and choose Recognize Pages command. Finereader will automatically determine the structure of each page and recognize the text.

4) Check for PDFs with very simple structure

If the layout of the PDF is very simple – a number of paragraphs following each other, with a few simple tables and images, – then you can usually convert it to a Word document straight away and translate it.

At the quoting stage, it is recommended to perform pseudo-translation or machine translation to understand whether you will need to do any additional work when the job is approved by your client, such as manual recognition inside Finereader or special formatting of the Word document.

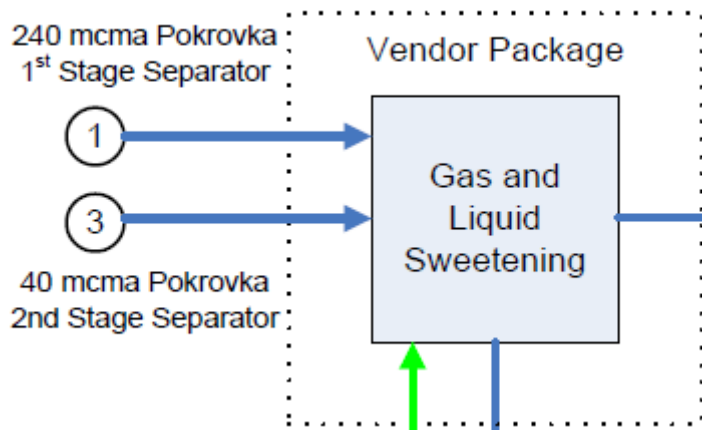


5) Analyze additional effort in Finereader for all other PDFs

- You will need to fix recognition mistakes if this is a poorly scanned PDF. So count the rough number of such mistakes in the entire document.

Årets förändringar av eget kapital	Årets förändringar av eget kapital
Belopp vid årets ingång	Belopp vid årets ingång
Resultatdisposition	Resultatdisposition
Fond för utvecklingsutgifter	Fond för utvecklingsutgifter
Erhållet ovillkorat aktieägartillskott	Erhållet ovillkorat aktieägartillskott

- You will need to translate text over images, so count the number of text labels in all images.



- Estimate the amount of effort required to take care of various layout and formatting aspects, both during conversion inside Finereader and during preparation inside Word. The table below is a sample estimate that you can use as a template for estimation. Under each aspect, you can see a description that explains why this aspect is important:

Aspect	Count	Estimated time
<p>(Scan) Poor recognition quality – rough number of mistakes</p> <p>If the PDF contains scanned pages with low scan quality, Finereader may produce text recognition mistakes which will need to be fixed before saving the document. You need to perform recognition of all pages and check how many recognition mistakes (roughly) will need to be corrected inside Finereader.</p>	Approx. 300	1.5 hours
<p>(Images) Text labels over images</p> <p>If the PDF contains images with text on them, you will need to prepare the text in images for translation using the approaches explained in Step 8 above.</p>	15	1 hour



Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatortools.net

Aspect	Count	Estimated time
<p>(Structure) Changes in page parameters (size, orientation)</p> <p>If the size or orientation of different pages in the PDF are not the same, you will need to create different sections and adjust their page properties.</p>	2	15 minutes
<p>(Structure) Changes in header/footer compared to previous page</p> <p>In case different pages of the PDF have different headers and footers, you will need to create different sections, configure headers/footers, and modify headers/footers so that they are identical to PDF.</p>	5	30 minutes
<p>(Structure) Table of contents</p> <p>If the PDF contains a table of contents, Finereader will generate a table of contents, but it may be impossible to update such table of contents after translation. You need to check which headings are used to assemble the table of contents, and possibly adjust their styles, and then modify the table of contents appropriately.</p>	–	15 minutes
<p>(Formatting) Checking and fixing list numbering</p> <p>If the PDF contains numbered lists, the numbered lists in the Word document may be incorrect, e.g. some paragraphs may be formatted as normal paragraphs and not as list paragraphs. List numbering in complex multi-level lists may also be incorrect.</p>	–	5 minutes
<p>(Formatting) Checking bold, italic, underline, text colors applied to individual words and phrases.</p> <p>If the PDF contains bold, italic, underline, or specific colours applied to individual words and phrases (and not to the entire paragraph), you will need to check that the Word document contains the same formatting. This is especially relevant for conversions made with Finereader 12.</p>	–	10 minutes
<p>(Formatting) Paragraph formatting inconsistencies</p> <p>If some paragraphs have distinct alignment, indentation, before/after spacing and line spacing compared to other paragraphs, you will need to make sure that the Word document is adjusted accordingly.</p>	–	20 minutes
<p>(Formatting) Background colors</p> <p>The PDF may have background colors applied to paragraphs or tables and table cells, so you will need to apply the same formatting in Word.</p>	–	10 minutes



Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

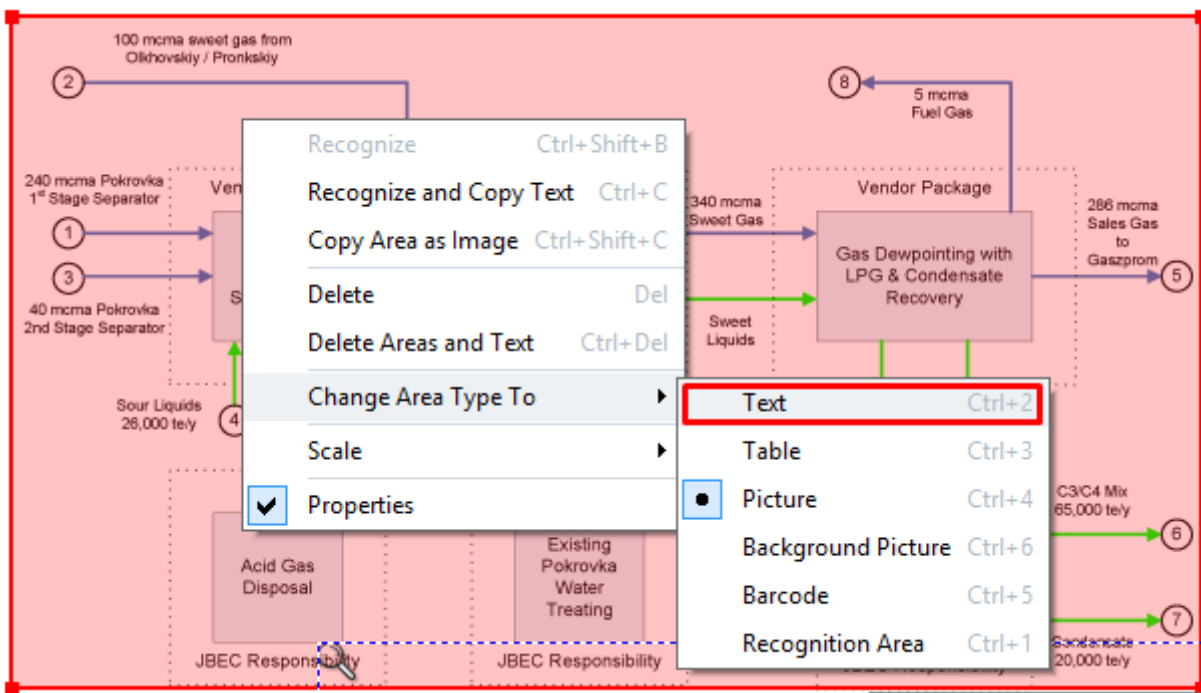
TransTools – Useful tools for every translator

www.translatorstools.net

Aspect	Count	Estimated time
(General) Manual conversion in Finereader You will need to take some time to mark recognition areas in Finereader. This will usually depend on the number of pages and complexity of page layout.	–	30 minutes
(General) Automated checks in Word You will need to take some time to process the Word document with Document Cleaner and Unbreaker tools, check all issues found by Document Cleaner's Autoformat command and make appropriate adjustments.	–	10 minutes
Total		4:55

6) Produce word count

Once you have estimated the time for all the additional effort, find the images that contain text on them and change such image areas to text areas:



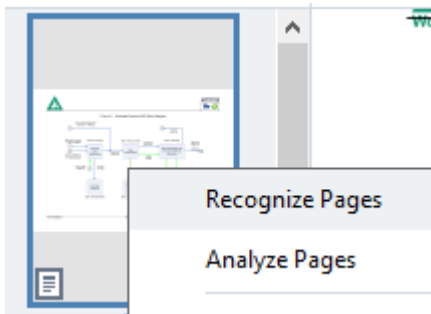


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

Then select all pages and perform Recognize Pages command so that all the text from those images gets into the Word document:



Finally save the Word document and prepare the word count in your CAT tool.

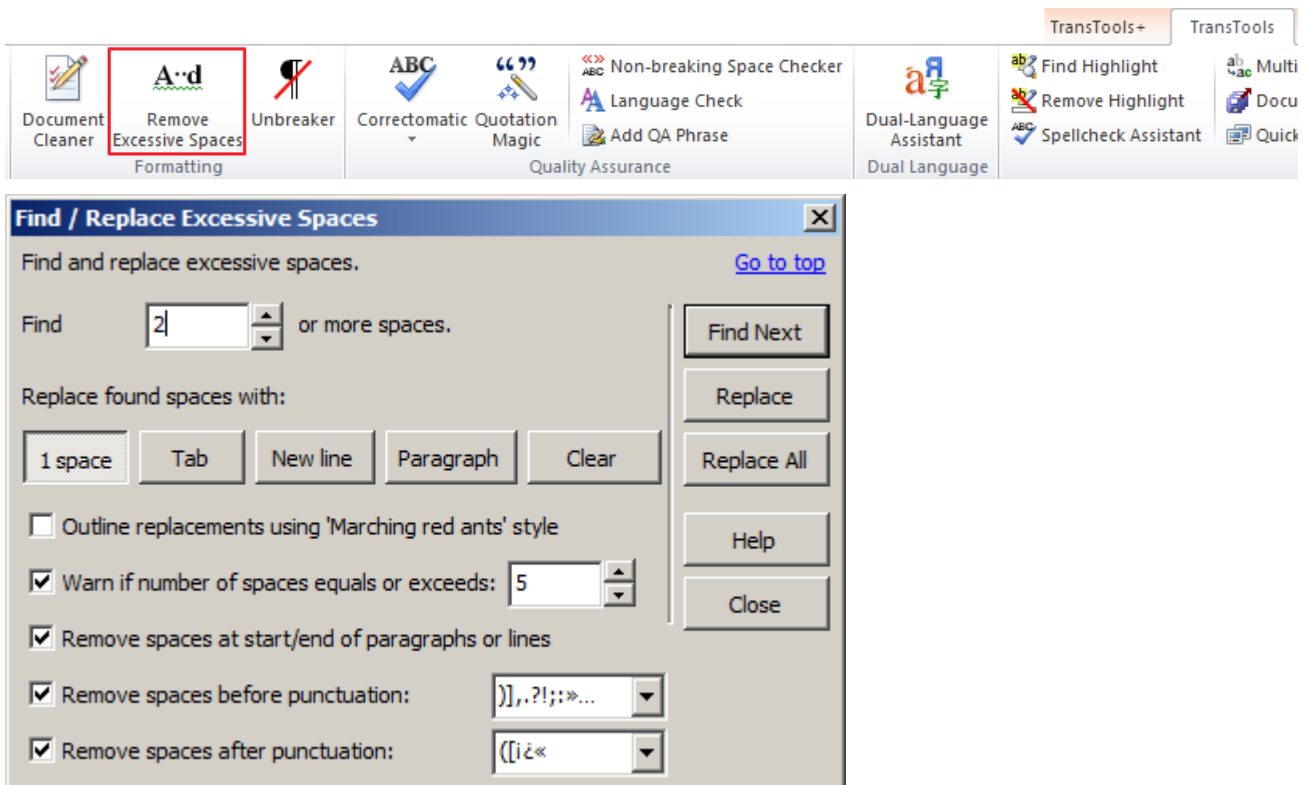


IV. Extra tools for Word document formatting before and after translation in CAT tools

1. Removing excessive spaces in Word documents

Removal of excessive spaces from a Word document before CAT tool import improves TM leverage. Also, sometimes excessive spaces are used to align text into visual columns, i.e. they are used instead of tab characters.

To find and replace excessive spaces you can use *Remove Excessive Spaces* tool in TransTools:



This tool works as a Find & Replace dialogue – when you click *Find Next* button, you locate the next sequence of spaces, and you can replace it with a single space, tab character, line break, paragraph break or nothing (no text) by clicking *Replace* button. *Replace All* button performs replacement automatically until it finds a sequence of 5 (default) or more spaces (this setting is configurable).

Remove spaces at start/end of paragraphs or lines option allows you to remove spaces at the beginning and end of paragraphs or lines.

The last two options make it possible to remove spaces before and after punctuation signs.

2. Finding and correcting punctuation issues, common mistakes, common text recognition issues

Documents converted from PDF files may contain the following:

- Common recognition mistakes, e.g. specific letter combinations are often recognized incorrectly.



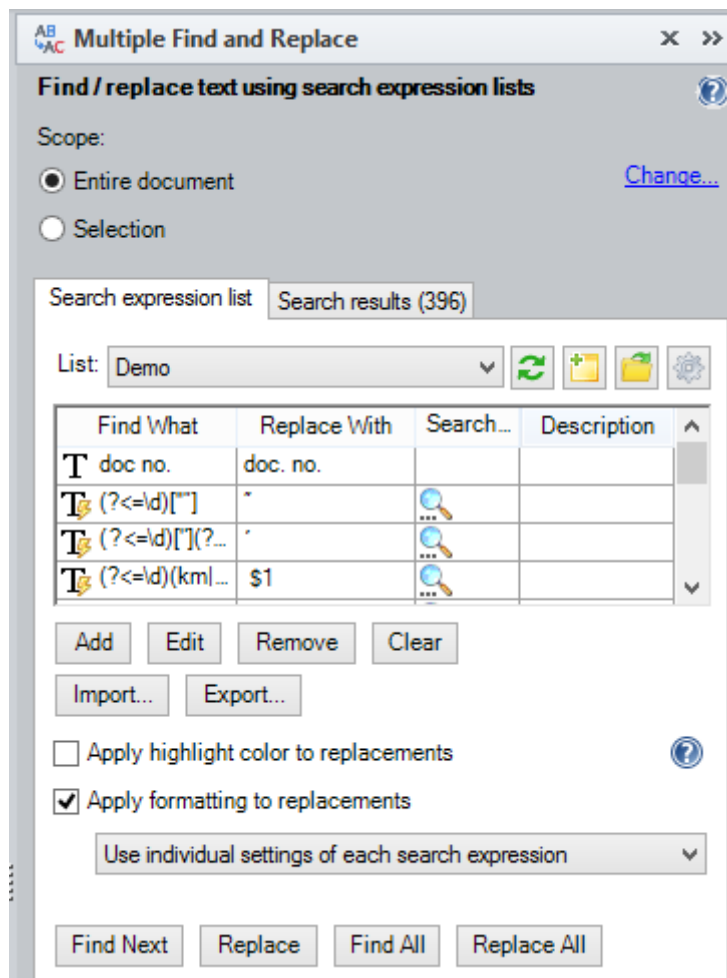
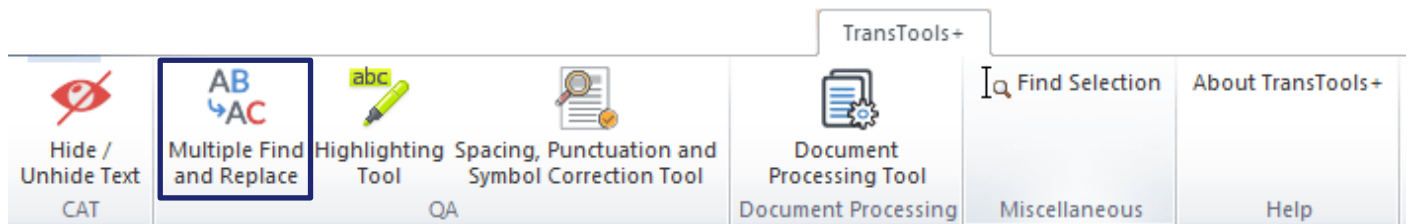
Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

- Common spelling errors.
- Common typos.
- Punctuation errors.

You can find and correct all these issues using a special tool called **Multiple Find & Replace**. It is a powerful substitute for Word's Find / Replace dialogue and Navigation pane which allows you to search for multiple words and phrases simultaneously and even use powerful regular expressions to locate and replace text that cannot be found using traditional means.



Here is what you can do using Multiple Find & Replace tool:

- Automatically replace all occurrences of words / phrases defined in your list by using *Replace All* button.

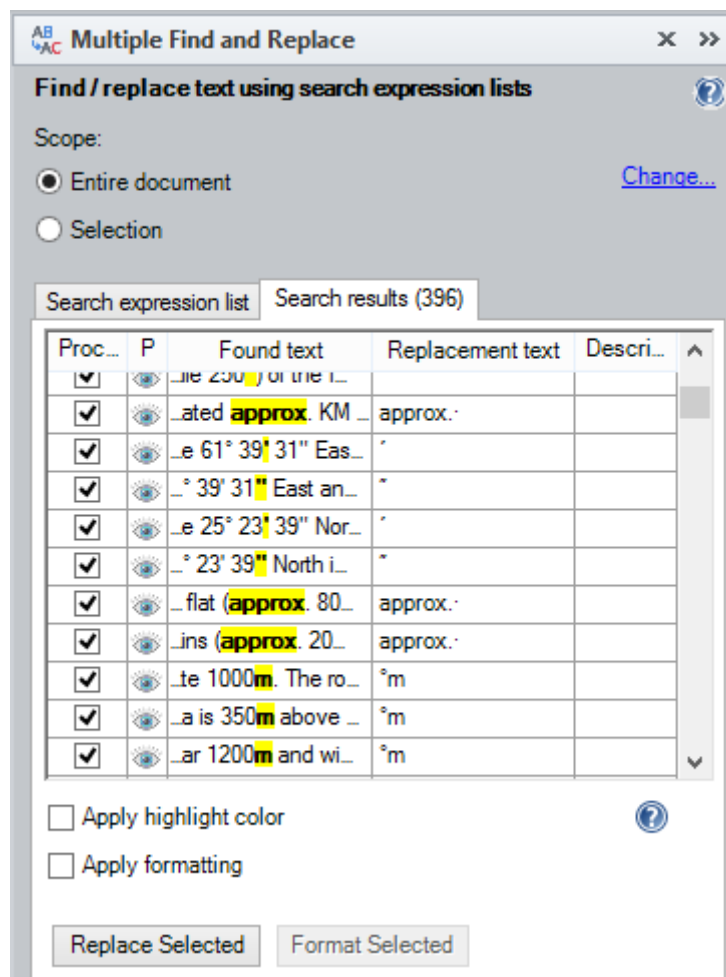


Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

- Use Microsoft Word's Find & Replace syntax for simple search and replacement, or use the powerful .NET regular expressions to perform complex searches and replacements.
- Confirm each replacement before it is made by using *Find / Replace* buttons.
- Display a list of all occurrences of words / phrases defined in your list using *Find All* button, and then see each occurrence in document context and preview how the text will appear after the replacement is made.
- Decide which found occurrences of words / phrases need to be processed, and use *Replace Selected* button to replace selected occurrences or *Format Selected* button to apply font formatting or highlight color to these occurrences.
- Search for text that has specific font formatting (including Bold, Italic, Underline style, Font Color, Font Size, Subscript and Superscript, Strikethrough and Double Strikethrough, All Caps and Small Caps), and apply specific font formatting during replacement or formatting operation.
- Modify the replacement text of specific occurrences of words / phrases based on context, if the initial replacement text must be tweaked.



List of search result which can be used to preview replacements or choose which replacements to make

Multiple Find & Replace tool is part of TransTools+, a plug-in for Microsoft Word. For more information, go to <http://www.translatorstools.net/products/transtoolsplus/word-multiplefindreplace>.



Conversion of PDFs/scanned images to Microsoft Word format using ABBYY Finereader

TransTools – Useful tools for every translator

www.translatorstools.net

The latest version can be downloaded from <http://www.translatorstools.net/downloads>