

EXTRACTING TEXT FROM UNNECESSARY TABLES IN PDF

NTRODUCTION

Extracting text from unnecessary tables in Adobe Acrobat is important for several reasons related to accessibility and usability. Unnecessary tables can confuse users, especially those using screen readers, so removing them helps streamline navigation and it simplifies the layout by removing non-essential tables making the document easier to read. Screen readers often struggle with complex or unnecessary tables, leading to misinterpretation of the content and removing them ensures clearer communition of information and allows the information to be presented in a more logical and accessible format.

TABLE OF CONTENTS

Click a title below to go to the destination page.	
Introduction	1
Setting Up the Document Properties	2
Document Title	2
Setting the Document Title	2
Setting PDF Metadata to Use the Document Title	4
Setting the Language in PDF	5
Extracting Text from Unnecessary Tables	6
Extract text from unnecessary tables	6
Full Accessibility Check on Document	8
Full Accessibility Check - Newer Versions of Adobe	8
Full Accessibility Check Older Versions of Adobe Acrobat	0

There are three parts to setting the basic document properites:

- 1. Setting the Document Title
- 2. Setting PDF Metadata to Use the Document Title
- 3. Setting the Language in PDF

Performing these steps first when opening any PDF document will help resolve the most common errors before you start the remediation process.

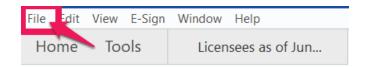
DOCUMENT TITLE

SETTING THE DOCUMENT TITLE

First check if the document has a document title. If it doesn't add the title. If the source document (such as Word, Excel, or PowerPoint) includes a title, the PDF will typically carry it over. Otherwise, you will need to add a title to the PDF. Follow the steps below to determine if the document title has been set properly.

Older Versions of Adobe Acrobat

1. Click on "File" at the top of the menu bar.



2. Select "Properties" to open the Document Properties window.

File Edit View E-Sign Window Help	
Dpen	Ctrl+O
Protect Using Password	
Request E-signatures	
1 Share File	
Revert	
<u>C</u> lose File	Ctrl+W
Prop <u>e</u> rties	Ctrl+D
Print	Ctrl+P

Newer Versions of Adobe Acrobat

1. Click on "Menu" at the top left of the window.



2. Select "Document Properties" to open the Document Properties window.



3. View the Description tab to determine if the document has a title and/or if the title is an appropriate description of the document. If the title is accurate, proceed to next section. If the title is not there or is not accurate, proceed to step 4.

Note: Steps 3 and 4 are the same in both older and newer versions of Adobe Acrobat

Document Properties	×
Description Security Fonts Initial View Custom Advanced	
Description	1
File: Adding Table of Content Links PDF.pdf	
Title: Adding Table of Content Links in PDF	ł
Author:	
Subject:	

4. If there is no title, or if the title does not provide an accurate description of the document click in the Title field and select any text in it. Type in the new accurate document title.

SETTING PDF METADATA TO USE THE DOCUMENT TITLE

When adding a title to the document, the metadata must be updated to reflect the new document title instead of the file name.

Note: These steps are the same in both older and newer versions of Adobe Acrobat.

- 1. Make sure you are still in the Document Properties. See last section for guidance.
- 2. Click on the "Initial View" tab.

cument Properties			
escription Security I	Fonts Initial View	stom Advanced	
Layout and Magnifica	tion		
Navigation tab:	Page Only	l.	
Page layout	Default	*	
Magnification	Default	~	

3. If the "Show" dropdown menu is set to "File Name" change this to "Document Title"

	Fonts Initial View Cus	ustom Advanced	
Layout and Magnifica	tion		
Navigation tab:	Page Only	*	
Page layout	Single Page	v	
Magnification	Default	~	
Open to page:	1 of 11		
Window Options			
Resize window	to initial page		
Center window	on screen		
Open in Full Sci	een mode		
Show: Documen	t Title 🔍		
File Name			
Iser Interface Option	Inde		
Hide menu bar			
Hide menu bar			
	ontrois		
Hide tool bars	ontrols		
Hide tool bars	ontrols		
Hide tool bars	ontrols		
Hide tool bars	ontrols		
Hide tool bars	entrols		

SETTING THE LANGUAGE IN PDF

Screen reader technology will read the document to the user in the language set for the document. A document can only have one language setting. The steps to set the correct language in a PDF are:

Note: These steps are the same in both older and newer versions of Adobe Acrobat.

1. Click on the "Advanced" tab in the Document Properties window.



4. Choose the "Language dropdown" and change it to the correct language. In this case "English."

Description Security Fonts Initial	View Custom Advanced		
PDF Settings			
Base URL:			
Search Index:		Browse	Clear
Trapped: Unknown	0		
Accessibility			
PDF/UA Compliance:	None		
Print Dialog Presets			
Page Scaling:	Default 🔍		
DuplexMode:	Simplex		
Paper Source by Page Size:	1		
Print Page Range:			
Number of Copies:	Default 🗠		
Reading Options			
Binding: Left Edge	~		
Language: English	×		
	Arabic support Hebrew support		
Help Estonian Finnish		ОК	Cancel
French French (Moro German	(CC0)		

5. Click the "OK button to save any changes to the PDF title or language.

EXTRACTING TEXT FROM UNNECESSARY TABLES

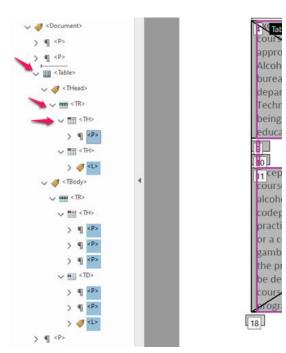
Sometimes paragraphs or lists will be nested inside tables due to formatting.

EXTRACT TEXT FROM UNNECESSARY TABLES

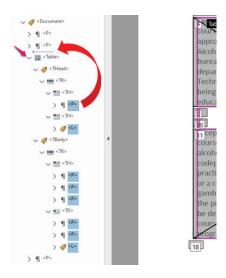
1. Open the Tags panel. In newer versions of Adobe Acrobat this tool is found on the right side of the screen. In older versions of Adobe Acrobat this tool is found on the left side of the screen.



2. Navigate to the table tag. Open the dropdowns for the "Table", "Table Row", and "TH/TD" to find the Paragraph or List tags.



3. Highlight all "P" and "L" tags and drag them above the Table tag. Strip the content out of a table that does not need to be in a table.



4. Right click on the now empty Table tag and select "Delete Tag".



5. Retag other text as necessary. This will update the page view so that the table on the page will no longer be visible.



FULL ACCESSIBILITY CHECK ON DOCUMENT

FULL ACCESSIBILITY CHECK - NEWER VERSIONS OF ADOBE

The first step is to conduct a comprehensive (full) Accessibility Check to identify any errors that need to be addressed in the Content Tool in the All Tools Menu on the left. This menu is displayed by default on the left side of the page.

1. Locate the All Tools menu on the left and click the Prepare for accessibility button.



2. Click on Check for accessibility



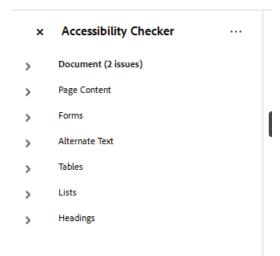
3. Verify that the Category drop-down is set to "Forms, Tables and Lists".

Category:	Forms, Tables and Lists
	Document
All form	Page content
	Forms Tables and Lists
All form	Alternate Text and Headings
TR must	be a child of Table, THead, TBody, or TFoot

4. In the Accessibility Checker Options dialog box, under Checking Options, verify that it shows (32 of 32 in all categories). Then click on Start Checking.

Checking Options (32 of 32 in all categories)				
Category:	Forms, Tables and Lists			
	Document			
✓ All form	Page content			
All form	Forms, Tables and Lists Alternate Text and Headings			
🗹 TR must	be a child of Table, THead, TBody, or TFoot			
TH and T	D must be children of TR			
🗹 Tables m	ust have headers			
🗹 Tables m	\boxdot Tables must contain the same number of columns in each row and rows in each column			
🗹 Tables m	☑ Tables must have a summary			
🗹 Ll must b	be a child of L			
Lbl and L	Body must be children of LI			
Select All Clear All				
Show this dialog when the Checker starts				
Help	Start Checking Cancel			

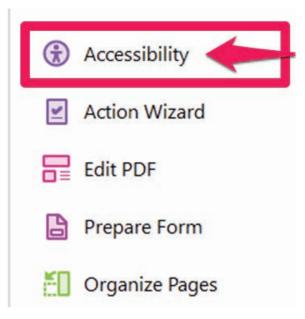
5. The Accessibility Checker panel will appear on the right side of the document. Expand the arrows to see the accessibility issues.



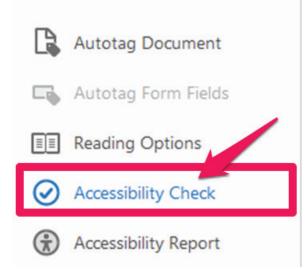
FULL ACCESSIBILITY CHECK OLDER VERSIONS OF ADOBE ACROBAT

The first step is to conduct a comprehensive (full) Accessibility Check to identify any errors that need to be addressed in the Content Tool on the left tool menu.

1. Use the Accessibility Tool on the right side of the document tool menu.



2. Select the Accessibility Check tool in the Accessibility tool to see if there are any errors.



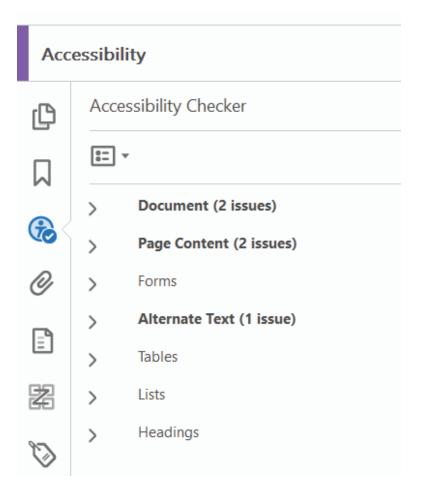
1. Verify that the Category drop-down is set to "Forms, Tables and Lists".

Category:	Forms, Tables and Lists
	Document
All form	Page content
	Forms Tables and Lists
All form	Alternate Text and Headings
TR must	be a child of Table, THead, TBody, or TFoot

2. In the Accessibility Checker Options dialog box, under Checking Options, verify that it shows (32 of 32 in all categories). Then click on Start Checking.

Checking Options (32 of 32 in all categories)			
Category:	Forms, Tables and Lists 🗸 🗸		
	Document		
	Page content		
All form f	Forms, Tables and Lists Alternate Text and Headings		
☑ TR must k	e a child of Table, THead, TBody, or TFoot		
TH and TI	D must be children of TR		
🗹 Tables mu	ust have headers		
🗹 Tables mu	$\hfill \square$ Tables must contain the same number of columns in each row and rows in each column		
☑ Tables must have a summary			
☑ LI must be a child of L			
Lbl and L	Body must be children of LI		
Select All Clear All			
Show this dialog when the Checker starts			
Help	Start Checking Cancel		

3. The Accessibility Checker panel will appear on the left side of the document. Expand the arrows to see the accessibility issues.





STATE OF NEVADA

OFFICE OF THE CHIEF INFORMATION OFFICER