

# PRACTICAL GUIDE - HOW TO COUNT THE WORDS IN PDF FILES QUICKLY FOR QUOTATION?

Recently I received a lot of PDF files from the client and had to quote the total price in a quick time. How can we count the words in a PDF file effectively? This is a very practical question that comes up at regular intervals. Here I am introducing some practical guides based on summing up the experience of some professionals in translation and localization industry.

## 1. Distinct types of PDF files

A PDF file is used for representing documents in a manner independent of application software, personal computer hardware, and operating system. Each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, graphics, and other information needed to display it. Generally there are 2 distinct types of PDF files. The first one has editable content; it can be copied, pasted, deleted, changed, and converted. The other one has non- editable content. It looks like a scanned copy of images which contained tabs, pictures, text in it.

## 2. Converting into DOC

If the PDF is editable, you just simply copying and pasting its text into MS Word to count the word directly. Or open the document in Adobe Acrobat, from the menu bar select FILE > SAVE AS. In the SAVE AS TYPE drop-

### Our Services



Document Translation Services



Interpretation Services



Desktop Publishing Services



Video Translation Solutions

down list select RICH TEXT FORMAT (RFT), Click the SAVE button. Then open the new RFT document in Microsoft Word, from the menu bar select TOOLS > WORD COUNT. This way works well most of the time.

If the PDF file with non-editable content, there are two cases: 1. you got something like a marriage or birth certificate with a mix of handwriting text, symbols, stamps, emblems, etc., There is no short cut to count its words, but a newer version of Acrobat has an OCR (Optical Character Recognition) option already built in. Look under Tools for OCR or Recognize Text or under View>Tools>Recognize text and follow the prompts. An OCR solution is necessary if the PDF is “image” 2. The PDF file output from Quark Xpress, FrameMaker or InDesign. It is quite clean, there are ways to convert it into an editable document and proceed as above.

### 3. Useful tools

Adobe Acrobat is capable of doing a great OCR job to convert the images into the text. The end result should be a Word document where the text is really text and the graphics are boxes with graphics, and where the formatting of the original document gets preserved as much as possible.

ABBYY FineReader is just as good as Adobe Acrobat. But this is not a free tool. It costs in around £100. Some users thought that the price of FineReader is well worth of it.

Translator’s Abacus is an online tool. The user could drag and drop various file types (including PDF), and it pops up a browser with a printable report of the word count for each document. It worked fine for some professionals. (It is specifically created for word counts and is only 435 KB... that is, not a “big application”).

## Our Services



Document  
Translation Services



Interpretation  
Services



Desktop  
Publishing Services



Video  
Translation Solutions

AnyCount is a special word count tools only cost \$30 and it counts PDF files as well as many other formats in batch mode (several files at once) and generates a report. It also counts in lines, pages and characters. You can download it for 30 days trial for free.

Count Everything gives you the possibility to count everything by only clicking a button. It also gained a lot of loyal fans.

## Our Services



Document  
Translation Services



Interpretation  
Services



Desktop  
Publishing Services



Video  
Translation Solutions