

What is Searchable PDF

Searchable PDF is essentially a PDF image file. Unlike static image formats such as TIFF, JPEG and BMP, every PDA document has the ability to contain several layers of information i.e. image layer and text layer. The image layer carries information including the actual image, resolution, compression method, color depth, etc. Similarly, the text layer includes the actual ASCII text and an identification of the text's location on the page. In simple terms the Searchable PDF's text portions of the scanned document gets stored in a text layer, allowing the user to easily search for and locate any keyword within the scanned document.

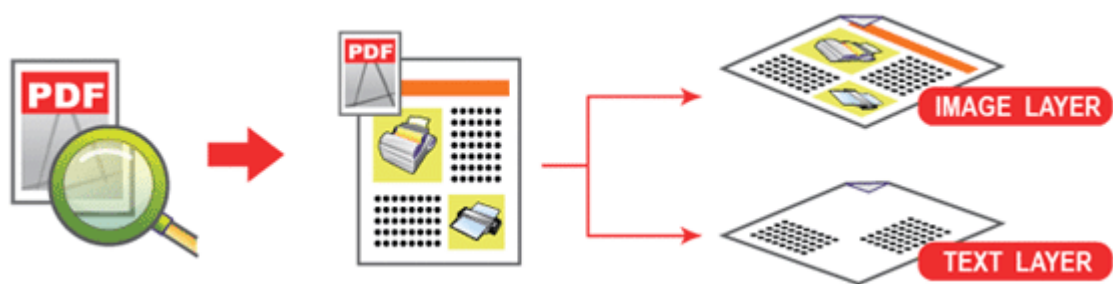


Figure 1: Searchable PDF is a combination of two layers: Image-Layer and Text-Layer.

Who needs Searchable PDFs?

Searchable PDFs are the ideal scanning solutions due to the image and text layers concept. Scanning your documents or books and converting them into the Searchable PDF format allows them to be easily searched for along with millions of other documents. No longer is it necessary to remember the subject or name of your documents nor keep track of them in a separate word file; finding information becomes easy and quick. Simply type any keyword you remember of the document you want to find to bring up all relevant documents. Manually typing each character and word of your documents to digitize them is all but a thing of the past. Last but not least, you can organize your documents in different folders and quickly browse through them when you need to find them in a hurry.



Figure 2: Use "keywords" to search your documents, books, and so on after they are digitized and converted to the Searchable PDF format.

How to Generate a Searchable PDF?

A Searchable PDF document is created in two ways:

1: If you already have a scanned image or document, (1) Open OCR software (2) Go to 'File' and open the scanned image (3) Save it as (searchable) PDF



Figure 3: OCR flow.

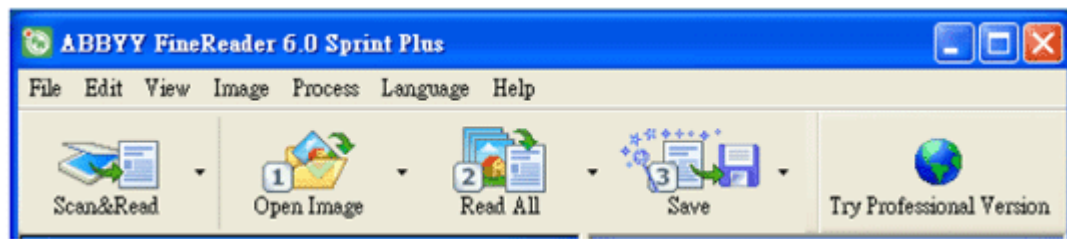


Figure 4: Example – ABBYY FineReader Sprint Plus (OCR Software).

2: If you want to convert a paper document into a Searchable PDF file, (1) Place the document on the scanner's scanning glass (2) Configure the scanner's hardware button utility and set it to save as (searchable) PDF (3) Press the button to create the searchable PDF directly

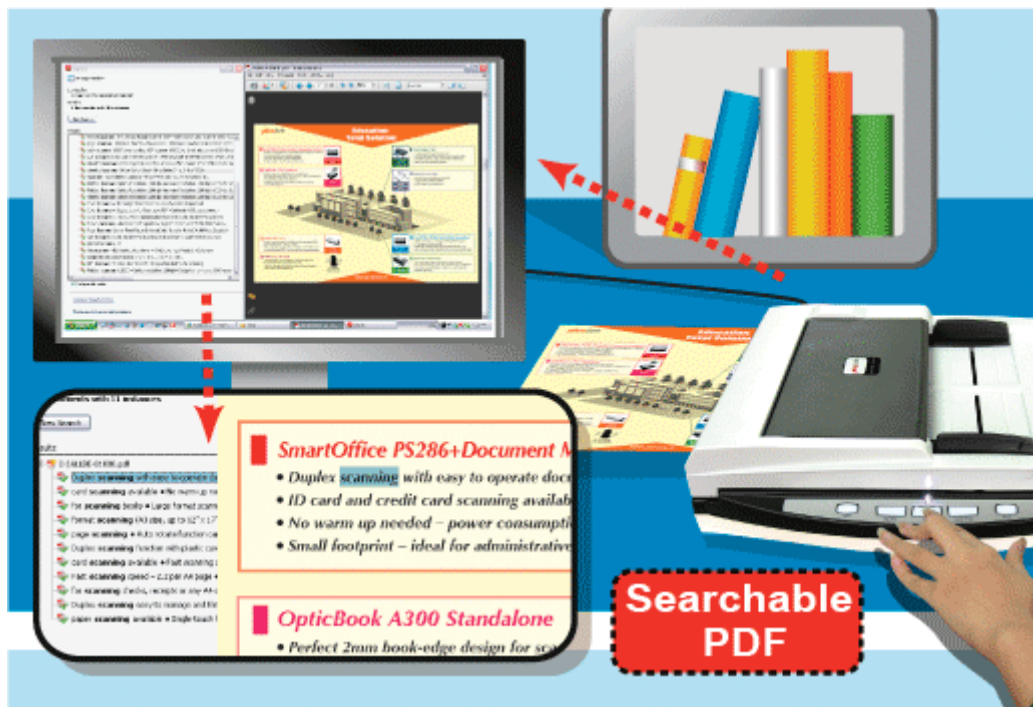


Figure 5: Plustek One-Touch button to "Searchable PDF" file.

The conversion into a Searchable PDF happens through the built-in specialized optical character recognition software commonly known as OCR. OCR converts the text portions of the image to text and stores it in the text layer. This text layer allows you to instantly search the document for keywords or phrases by using any desktop search engine.

What is OCR?

- OCR short for Optical Character Recognition is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner) into machine-editable text. The accuracy of recognition is closed to 100% if the characters on the documents are very neat and clean. Simply put an OCR program can looks at documents as images and tries to recognize the characters on it in order to convert it them to text.
- Plustek offers a number of award-winning high-quality OCR applications with their scanners, including Plustek DI Express, ABBYY FineReader. These OCR applications are used for various markets and for various situations. For example, DI Express, which is provided in the Asian market, is specifically used for recognition of traditional Chinese, simplified Chinese, Japanese Kanji, and Korean Hangul.

Benefits

Searchable PDFs provide higher security, reliability and easiness to transit and exchange documents; while preserving the look and integrity of the original

documents, including fonts, graphics, images and layouts. Files can be easily viewed, printed, and shared with anyone, across any operating system and application.



Figure 6: Scan to Searchable PDF file is the best way to share documents.

The Searchable PDF format is a way to easily manage your documents, books and images. It not only enhances your efficiency but also keeps you connected with your information and knowledge at all times.