

TechnoFeature™

Practice management and technology articles written by experts.

Review: ABBYY FineReader 9.0

By John Heckman

TechnoScore: 4.5

1 = Lowest Possible Score; 5 = Highest Possible Score

www.ABBYY.com

INTRODUCTION

ABBYY has just released version 9.0 of its FineReader OCR software, which claims a number of new or improved features — including the ability to recognize headers, footers, paragraph numbering, and footnotes as well as the ability to reproduce them correctly as Word functions (as opposed to fixed text boxes or frames). It can also OCR digital photographs, and of course it supports Office 2007, dual-core processors, etc. It also ships with a “Screenshot Reader” which enables you to OCR screenshots, Windows Explorer lists, etc.

How does all this stack up? At a time when law firms have dramatically increased their scanning capabilities and moved toward a “paperless” office, it is critical to define terms.

The vast majority of documents will be scanned as text-searchable PDFs. Only when a firm needs to edit a document will the issue of converting it to Word or WordPerfect using a product like ABBYY FineReader 9.0 arise. However, since more and more firms are by default sending documents out in PDF format, the question of conversion from PDF to Word when you receive a PDF document is also critical.

In evaluating a program such as ABBYY FineReader, it is important to keep in mind that some functions are designed for volume scanning, but that many features touted by various companies are fairly labor intensive and apply to only a relatively small number of documents.

VOLUME SCANNING

For volume scanning, ABBYY FineReader enables you to designate a “watch” directory and will perform OCR on any document dropped into that directory. This feature is potentially a great productivity enhancer since it puts the most time-consuming process — OCR — into a background function. However, whatever PC or server is “watching” the directory needs to be robust.

ABBYY FineReader scanned a clean five page laser printout of a standard letter with zero mis-recognized characters.

The Corporate Edition of ABBYY FineReader will also watch your email inbox for scanned documents emailed to you. Since this is a very common setup for multifunction machines (albeit a rather inefficient one), this feature could also prove very valuable. However, for this review, I was unable to test it.

RECOGNITION QUALITY

I have used ABBYY FineReader personally for a number of years, and have always found its OCR recognition to be better than OmniPage, its main competitor. Where OmniPage and other software traditionally produced Word documents that make extensive use of “frames” (making them extremely

(Continued on next page)

difficult to edit), ABBYY FineReader does a better job.

However, I do not have a copy of the most recent version of OmniPage, so I was not able to conduct a direct comparison of the two products, or in particular compare some of the new features (header/footer recognition) with similar features claimed by OmniPage.

Round 1: Standard Letter Laser Printout

ABBYY FineReader scanned a clean five page laser printout of a standard letter with zero mis-recognized characters — obviously, your mileage may vary, but this marks the first time I can remember 100% accuracy. It also recognized the signature as a graphic and left it positioned correctly in the document.

ABBYY FineReader did not do quite as well with formatting. The line breaks on the inside address of a standard letter were not correct and rather than recognizing that a page break did not necessarily indicate a real break, it placed a hard return at the end of every page.

Round 2: Old, Typed Document on Erasable Bond

So then I dug out a 20-year old typed document on erasable bond (if you recall, the contrast is not nearly as good as that of a laser printer). The document scanned flawlessly with one exception — in five instances an “h” was recognized as an “n.” This was easily picked up by a spell check however. In looking back at the original, I saw that the top of the “h” key was somewhat broken and very faint on the original — looking more like an “n.” Garbage in, garbage out.

Round 3: Complex Invoice

For my final test I took a highly formatted invoice with logos and more or less odd fonts frequently used by law firms (Copperplate Gothic). Scanning “as is” was a mess. However, if before conversion to Word I designated the logos as pictures and the body of the invoice as a table, the resulting Word document, while still requiring a lot of work, was usable.

ABBYY FineReader seems to have a problem with paper size recognition. When I scanned a more complex letter size document (such as the above invoice), ABBYY FineReader thought it was an 11x17 piece of paper! And when I ran the recognition on a letter-size PDF, it thought it was legal size. While this was easy enough to correct, it means that paper size recognition is always a manual process. So while you can set the program to default to letter-size, if you want to scan legal size paper, you have to manually change the setting and then manually change it back to letter size. This could become onerous depending on exactly what you are scanning. ABBYY says that it is aware of this issue and is working hard to improve it.

In my test documents, the header/footer/footnote recognition worked extremely well.

You should also be aware that when converting even a text-searchable PDF to Word, ABBYY FineReader does not convert any hyperlinks. This seems to be a common failing among conversion programs.

By default, the installation includes recognition for 179 languages, including some you may never have heard of and will never need — Crow, Faroese, or Sami (Lappish). On the other hand firms with specialty practices, or with a heavy immigration practice, may appreciate the many foreign languages, in particular all the variations on Eastern European languages. Otherwise, you may want to perform a custom install and eliminate them.

HEADERS, FOOTERS, FOOTNOTES, AND PARAGRAPH NUMBERING

ABBYY claims two major new advances for this version of the software: the ability to recognize headers, footers, and footnotes as such and the ability to correctly recognize automatic paragraph

(Continued on next page)

numbering. In the past, OCR software has generally recognized headers and footers only as part of the body copy and/or in frames, so you have to go in and remove them manually. Similarly, paragraph numbers have been hard coded; if you delete a paragraph, the numbering does not change.

In my test documents, the header/footer/footnote recognition worked extremely well. Not only does this mean that they were positioned correctly on the page, but you can edit them in Word — no longer do you have to delete them all manually on each page and create a useable header/footer. It even recognized a first-page header as different from headers starting on page 2. This is indeed a major advance.

With ABBYY FineReader you can now conduct OCR on digital photographs of documents — shades of 1950's spy movies.

As I continued testing, I added complexity: both headers and footers and then headers, footers, and footnotes, including more than one footnote on a page. ABBYY FineReader continued to perform well, although there was one particularly complex document where it fared less well. The automatic paragraph numbering feature also fared less well — or more exactly, did not work at all. The best I could do was to select the entire document, change the paragraph numbering (from numbers to bullets, for example), and then change it back. That seemed to work, but I would regard this feature as rather iffy at this point.

SCREENSHOT READER

ABBYY Screenshot Reader enables you to capture text on the screen to text, a table, an Excel spreadsheet, etc. It would be a great product if it worked better. While the text recognition is very good (though it has problems with icons designating the type of file in a Windows Explorer window),

other features such as capturing text to a table are problematic.

This feature would be perfect for a list (in my sample a client/matter list in Worldox, clearly formatted as a table) that has a “nickname” and a corresponding “description” that you want to capture to reproduce, print, and edit. Unfortunately, the resulting tables are unreliable — all too frequently multiple rows are grouped together in a single row. Drop down lists cannot be captured at all (although other screenshot capture programs such as SnagIt will do this).

In addition, if you select “table to Excel,” every capture creates a new Excel file. Since drop-down lists often have to be captured in several passes, it would be nice to have a “paste” option here to put everything into the same file.

DIGITAL PHOTOGRAPHS

With ABBYY FineReader you can now conduct OCR on digital photographs of documents — shades of 1950's spy movies. I took a photo of an open book with a 7 Megapixel camera. The book was open, so the pages were curved and not flat (always a problem with copies if the book was not pressed down on the glass and each line is curved). ABBYY FineReader correctly recognized the photo and the OCR quality was acceptable, if not great. There were no major errors and the “curve” in the pages was corrected for. So instead of trundling the book off to be copied (assuming anybody still does research with actual books in an actual library), just take a picture of a few pages and you are good to go.

So naturally, I wanted to try my cell phone (Treo 750). Try as I might — several different passes, different pages, etc. — the cell phone simply could not produce a sufficiently good quality image to OCR. So fantasies of surreptitiously capturing secret text and recognizing it with my ABBYY spy decoder software remained just fantasies.

CONCLUSION

The name of the game is character recognition, and at this ABBYY FineReader is superb. Some

(Continued on next page)

of its new features, in particular header/footer and footnote recognition also work well in general, albeit with some kinks in very complex documents. However, since most of your documents will remain in PDF format, this issue will affect only a very small minority of documents — those you want to convert to Word and which have particularly complex header/footer/footnote structures.

Copyright 2008 John Heckman. All rights reserved.

ABBYY'S RESPONSE

We invited ABBYY to respond to this TechnoFeature. Wendy Wang, Sr. Manager, Product Marketing and Marketing Program at ABBYY responded as follows:

"I would like to share my thoughts about John's comparison about Screenshot Reader and SnagIt.

"ABBYY Screenshot Reader is a free bonus utility vs. SnagIt's \$40 retail value. Our strong point is that our product has OCR capability, so it allows users to capture text from image-only sources while SnagIt cannot."

ABOUT THE AUTHOR

John Heckman has been assisting law firms with technology issues for over 25 years. Heckman Consulting is a software integration firm specializing in Amicus Attorney, PCLaw, Worldox document management, Time Matters, HotDocs and other legal-specific software. It services the greater New York area and clients range from solo practitioners to AmLaw 100 firms. Check out John's [blog](#) or the [Heckman Consulting Web site](#) for newsletters and other useful information.

Contact John:
E: heckman@heckmanco.com

About TechnoFeature

Published on Tuesdays, *TechnoFeature* is a weekly newsletter containing in-depth articles written by leading legal technology and practice management experts, many of whom have become "household names" in the legal profession. Most of these articles are TechnoLawyer exclusives, but we also scour regional legal publications for superb articles that you probably missed the first time around.