

Mariner Software

Knowledgebase > Paperless for Mac OS > OCR > Why are OCR results inaccurate or simply unintelligible?

Why are OCR results inaccurate or simply unintelligible?

Jim Henson - 2024-04-25 - OCR

Why are OCR results inaccurate or simply unintelligible?

Optical Character Recognition (OCR) is a process by which a computer program (the OCR engine) makes the best guess possible as to what the text content of a visual image is. Since the OCR engine is making its best guess (and although we have done our best to instruct the OCR engine how to accurately determine text from visual content), the results of OCR in Paperless will not be perfect 100% of the time.

[Several factors contribute to the accuracy of OCR output.](#) If for some reason, one or more of these factors works against either the scanner's ability to acquire quality input or the OCR Engine's ability to discern what it has been instructed letters should look like, the results of the OCR process might vary significantly from the original. Also, it is important to understand in some detail [How OCR works](#).

Understanding how OCR makes an educated guess as to how visual input represents text characters, it should be possible to understand the limitations of the OCR process and what can be done to assist the OCR engine in producing more-accurate results.

We have made it possible to [review OCR results in Paperless by reviewing the OCR Text field for document types in Paperless](#). It is also possible to edit the content of the OCR Text field manually, in order to correct any mistakes found in OCR results.