

Mariner Software

Knowledgebase > Paperless for Mac OS > OCR > How come I can't select text after performing recognize text (OCR) on a document?

How come I can't select text after performing recognize text (OCR) on a document?

Jim Henson - 2024-04-20 - OCR

How come I can't select Text after performing recognize text (OCR) on a document?

The OCR function does not appear to be working as after applying Recognize Text, text is not selectable.

Whenever Optical Character Recognition (OCR) is performed on a document in a Paperless library, the raw OCR output is written to a special field: OCR Text.

The OCR Text field is included by default in the Fields list under Library Configuration > Data Types, but it is not assigned by default to a document type.

In order to show raw OCR results for a document:

Create a new document type (or modify an existing one) to include the OCR Text field.

Make sure that the document type for a document you would like to see raw OCR output for is set to a document type that the OCR Text field has been assigned to.

Furthermore, that the OCR'ed text is not "mapped", through layout analysis, and saved onto the PDF as a multi-layer PDF (Image over Text). We hope to provide this functionality in the future.

OCR text is still used for searches and in Spotlight and can be edited, corrected and selectable. Overall the OCR is very functional and helpful - but the info isn't (currently) written back to the PDF as image over text multi-layer PDF.

Please note that there are several FAQ's that deal with improving the quality of the OCR as it varies greatly depending on several factors, like dpi, size, font, paper, scanner, image correction.