

KitForma.

PRACTICAL FIELD GUIDE

OCR You Can Check

Ground truth, error decisions and a searchable-copy release record

01

02

03

SAMPLE EDITION

ENGLISH / 1.0 / 2026

How to use this guide

People digitizing ordinary English or Turkish scanned text and checking what recognition preserved.

A practical original workbook for separating visible appearance, recognized values and reading order, using four fictional practice sheets and reproducible review fixtures.

- Define a task-specific recognition acceptance contract.
- Check 25 known targets without confusing the result with whole-document accuracy.
- Identify independent value and reading-order failures.
- Record a reproducible release or rejection decision.

Scope and limitations

- The supplied OCR candidate is deliberately simulated; it is not a measured result from any OCR provider.
- Source SVG artwork must be captured or rasterized for a genuine image-recognition trial.
- This workflow does not certify accessibility, archival conformance, handwriting accuracy or universal language support.

Published by KitForma. Edition 1.0. Sources checked 1 October 2026. People and businesses in worked examples are fictional. This is an independent publication, not endorsed by the platforms mentioned.

Copyright 2026 KitForma. Usage, support and refund terms are included in LICENSE.txt in the purchased package.

Contents

How to use this guide	2
Contents	3
Give recognition a job it can fail	4
Sources and tools	6

Give recognition a job it can fail

The fictional Cedar Workshop has four paper sheets describing an equipment checkout exercise. Its coordinator wants volunteers to search for a cabinet identifier, copy a collection instruction, and find the Turkish labels that appear beside English ones. Nobody needs a legally certified archive or an automated accounting import. Defining those three uses makes the quality question concrete. A file can look excellent yet fail the cabinet search. It can also contain accurate words while presenting the two columns in an order that changes the instruction. Your task is to decide whether this particular searchable copy supports those three uses.

The included four SVG sheets are original source artwork, not scans obtained from a real organization. Open them in a current browser, check that the characters display, and print them if you want to create your own camera or scanner inputs. A clean vector PDF exported directly from that artwork already contains authored text; processing it again does not demonstrate recognition of a scan. For an actual OCR trial, use an image of the printed page or deliberately rasterize a copy, recording how you created that input. Keep the artwork, captured image, and OCR output separate so the evidence cannot become circular.

Begin a run record with an input filename, a date, the intended reader, the OCR application and version when visible, and the language settings you selected. Record observations rather than an imagined configuration. If an online service does not expose its engine version, write unknown. Do not upload private originals merely to follow this exercise: every included name and reference is fictional. A local workflow can be preferable for restricted material, but calling a tool local does not itself prove its data handling. Check the application you actually use.

Recognition produces a candidate that needs examination. W3C's PDF7 technique specifically includes checking converted text and its reading order; adding OCR alone is not a complete accessibility assessment. [\[S1\]](#) Our acceptance rule will therefore have separate gates for exact critical identifiers, ordinary sampled text, page coverage, and reading order. A high score in one gate cannot erase a failure in another. Write those gates before opening the deliberately flawed candidate so the result cannot quietly move the goalposts.

Sources and tools

[S1] **W3C PDF7: performing OCR and checking converted text**

<https://www.w3.org/WAI/WCAG22/Techniques/pdf/PDF7>

Checked: 2026-10-01

[S2] **Adobe: scan documents to PDF**

<https://helpx.adobe.com/acrobat/desktop/create-documents/scan-documents-to-pdfs/scan.html>

Checked: 2026-10-01

[S3] **Unicode Standard Annex 15: normalization forms**

<https://www.unicode.org/reports/tr15/>

Checked: 2026-10-01

Inside the complete guide

Build a reference without copying the mistake

Choose the unit before calculating a percentage

Run one controlled recognition experiment

Classify an error so the next action is useful

Read columns as instructions, not bags of words

Handle Unicode without erasing meaningful differences

Improve a candidate with an experiment log

Make the release record smaller than the problem

Capstone and answer key