

Report 4: Multi Page Invoice Fixes

Aliyan Faisal · the problems found in Report 3, fixed and scored again

These fixes were made using the same 40 multi-page invoices they are scored on. The new score shows that the problems listed in Report 3 are gone. It does not show how the extractor reads other multi-page invoices; that figure comes from invoices that were not used for any fix.

1. Summary

Report 3 scored the rules extractor on 40 multi-page invoices and listed what went wrong. For this report I fixed those problems in the extractor and scored the same 40 invoices in the same way, so the numbers can be compared directly. The *Evaluation Harness Guide* explains how the scoring works.

Rules extractor (the real one)	Report 3 (before)	Report 4 (after fixes)
Field accuracy	72.2%	99.98%
Planted mistakes found	13 of 20	20 of 20
False alarms	50	0
Correct invoices with no findings	8 of 20	20 of 20
Swiss-franc invoices (10), field accuracy	3.6%	100.0%
Other invoices (30), field accuracy	94.6%	100.0%

Result. Field accuracy on the multi-page invoices rose from 72.2% to 99.98%, all 20 planted mistakes are still found, and false alarms fell from 50 to 0. Every one of the 2,521 line-item cells is now read correctly. The only fields still wrong are two tax rates, which the extractor leaves blank instead of guessing, because the printed tax on those invoices carries a planted mistake and no longer matches a rate.

The two older sets were scored again to check that nothing got worse (section 4). They did not.

2. What was fixed

Each row is a problem from Report 3 with the count measured then and now.

Problem found in Report 3	What changed in the extractor	Before	After
Swiss number format not read (1'234.56)	An apostrophe is now accepted as a thousands separator.	All 10 Swiss invoices at 3.6%	100.0%
Amounts followed by a currency code or "kr" not recognised (246'197.86 CHF, 194 974,16 kr)	An ISO currency code or kr beside an amount no longer stops it being read as an amount. The code also counts as the invoice currency, and "kr" counts as the Swedish krona.	Totals, subtotal and tax missing on most invoices	Totals read on all 40
A description that wraps onto a second line is cut short	A row directly under a line, in the description column and with no numbers, is joined to that line. Part-number notes are left out.	230 wrong descriptions	0 wrong
An item code in front of the description	A short code printed in its own column before the description is no longer part of the description.	166 wrong descriptions (coded layout 39%)	0 wrong (coded layout 100%)
Section totals and carried-forward rows read as invoice totals	A row that closes a section ("Parts total"), or equals the sum of the last lines, is not used as a total. Carried-forward and brought-forward rows are skipped.	15 wrong totals, 12 spurious shipping amounts	0 wrong totals, 0 spurious shipping amounts
"Less deposit paid" read as a discount	A deposit or payment already received is a separate kind of row, and "Balance due" after it is not used as the total.	6 spurious discounts	0

Problem found in Report 3	What changed in the extractor	Before	After
Currency decided from "\$"	A bare "\$" is settled by what else the invoice says: an ABN means Australian dollars, an HST or GST/HST number or a business number ending RT0001 means Canadian dollars.	14 Australian and Canadian invoices read as US dollars	0 wrong
Supplier and customer swapped when the customer block is printed first	A name far to the right of the "Invoice to" label is no longer taken as the customer, and the delivery and customer blocks are skipped when looking for the supplier.	9 invoices	0 wrong
Customer missing when it follows a label on one line ("Billed to: Name")	The name after the colon is read.	8 invoices	0 wrong
Dates with a month name in US order not read ("June 14, 2026")	Month-first dates are now read.	Dates missing on all 7 Canadian invoices	0 missing
Tax rate and subtotal not found where the totals rows carry no labels	A tax row is also found by its size against the subtotal, so a tax amount with a planted mistake is still read as tax.	Tax rate missing on 17, subtotal on 10	Tax rate blank on 2, subtotal wrong on 0

Found while fixing, and also fixed: the invoice number written as "Ref. FV 2026/659" or in a headline (7 invoices wrong, now 0), tax IDs under the labels "Momsreg.nr" and "BTW-nr." (10 wrong, now 0), and credit lines that lost their minus sign (16 line amounts, now 0 wrong). Two smaller changes: "Amount payable" is now recognised as the label of a total, and the tax rates used to recognise an unlabeled tax row now include 2.6%, 3.8%, 8.1% (Switzerland), 12% (Sweden), 13% and 15% (Canada) and a few others. The checking rules themselves were not changed.

3. Results on the multi-page invoices

Fields read right, before and after. Green is 95% or better, amber 90 to 95%, red below 90%:

Field	Report 3	Report 4
Total	9/40	40/40
Tax rate	13/40	38/40
Tax amount	14/40	40/40
Customer name	19/40	40/40
Currency	20/40	40/40
Subtotal	20/40	40/40
Shipping	23/40	40/40
Line description	1510/2521	2521/2521
Supplier name	29/40	40/40
Tax ID	30/40	40/40
Line unit price	1896/2521	2521/2521
Line amount	1896/2521	2521/2521
Line quantity	1915/2521	2521/2521
Invoice number	33/40	40/40
Issue date	33/40	40/40
Due date	33/40	40/40
Discount	33/40	40/40
Page structure	Report 3	Report 4
coded	38.9%	100.0%

Page structure	Report 3	Report 4
stacked	52.6%	100.0%
inline	79.6%	99.9%
wide	82.7%	100.0%
classic	95.9%	100.0%

Planted mistake type	Planted	Found in Report 3	Found in Report 4
grand total	5	4 of 5	5 of 5
line total	9	7 of 9	9 of 9
missing tax ID	1	0 of 1	1 of 1
subtotal	4	2 of 4	4 of 4
tax	1	0 of 1	1 of 1

4. The older sets, scored again

The same extractor was run on the two single-page sets to check that the fixes did not make them worse. Report 2 figures are shown for comparison.

Set	Field accuracy (Report 2 / now)	Mistakes found (Report 2 / now)	False alarms (Report 2 / now)
Corpus A: 30 independent invoices	98.1% / 99.1%	31 of 32 / 31 of 32	6 / 4
Corpus B: 40 generator invoices	99.9% / 99.9%	20 of 20 / 20 of 20	0 / 0

Corpus A improved a little as a side effect and corpus B stayed at the top. One attempted fix was dropped along the way: a guess at the tax rate from an amount that carried a planted mistake gave 21% where the invoice used 20%, and caused two false alarms on corpus B. The extractor now leaves the rate blank in that situation.

Conclusion

The problems listed in Report 3 are fixed: field accuracy on the multi-page invoices went from 72.2% to 99.98%, with all planted mistakes found and no false alarms, and the two older sets did not get worse.

Read the 40-invoice score with care. The invoices are synthetic and few, and I fixed the extractor against these same invoices, so the score is higher than it would be on invoices the extractor has not seen. Some of the fixes are general conventions that real invoices use, such as the Swiss number format, currency codes beside amounts, deposits and section totals. Others handle layouts that appear in these invoices, such as a customer block printed first or a wrapped description spaced a certain way, and may not carry over to other layouts. The figure to rely on is a score on invoices that were not used for any fix. Field accuracy is the harness's weighted score (money fields count more than names) everywhere in this report.

The invoices, their answer keys and the raw results for this report can be downloaded from the Harness Reports page of this site.