

General comparison between Regex and Hybrid methods (all 319,680 notes).

Metric	Regex (Baseline)	Hybrid (Proposed)
Total notes analyzed	319,680	319,680
Agencies analyzed	21	21
Analysis period	01/01/2025– 24/10/2025	01/01/2025– 24/10/2025
Notes with at least one valid identifier (Transfer or NC)	22,846	36,898
Systemic patterns mapped (transfer + NC)	67	42
Non-systemic patterns (TED)	29	64
Invalid patterns	22	11
Total identifiers extracted	52,577	111,563
Unique regular expressions	11	320

Comparison between Regex and Hybrid methods in the 48,257 notes that mention “TED”.

Metric	Regex (Baseline)	Hybrid (Proposed)
Notes that mention “TED”	48,257	48,257
Transfer numbers identified	22,487	22,226
Valid credit notes	6,585	2,153
Invalid credit notes	1,968	2,276
Notes with a valid systemic identifier (transfer or NC)	22,846	12,718
Notes that mention TED but have no valid identifiers	25,411	35,539
Notes with a TED number (non-systemic)	21,537	24,462
Notes exclusively with a TED number (depend solely on this identifier)	1,647	4,831
Notes that mention TED and are explicitly “sem vínculo”	22,440	22,399
Remaining notes without explanation (no transfer, valid NC, or “sem vínculo”)	2,971	13,183