SAKIRA

TUTORIEL 3

Sortir un tableau propre de factures ou de PDF

© SAKIRA

© SAKIRA. Reproduction et diffusion interdites sans accord écrit.
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933

TUTORIEL 3 · DU DÉBUTANT À L'EXPERT

Sortir un tableau propre de factures ou de PDF

Recopier à la main les montants de vingt factures donne des fautes de frappe. Une IA le fait vite, mais un tableau qui a l'air propre peut contenir un chiffre faux que personne ne voit.

La règle à respecter

Recalculez avant d'utiliser : un tableau propre peut contenir un chiffre faux : l'addition des lignes ne tombe pas toujours sur le total de la facture. Un chiffre vrai pris sur une autre ligne passe le contrôle, et le taux de TVA n'est pas vérifié. Ne partagez que des factures que vous pouvez partager.

AVANT DE COMMENCER

Il vous faut

Des factures dont vous avez le droit de partager le contenu, en PDF avec du texte (pas des scans), un assistant d'IA, et pour le niveau 3 un terminal avec `pdftotext` (sur Mac : `brew install poppler`) et Claude Code ou Codex.

Résultat

Un tableau (fournisseur, numéro, date, HT, TVA, TTC) et la liste des factures dont les chiffres ne tombent pas juste.

01 Savoir ce qu'on cherche

Une facture peut contenir une erreur d'addition. Un tableau recopié fidèlement la reproduit sans la signaler. Dans notre exemple de trois factures, l'une contient une erreur : le total HT imprimé ne correspond pas à la somme de ses lignes.

02 Débutant : une phrase à coller

Joignez vos PDF ou collez leur texte, puis écrivez :

À COPIER

```
Mets ces factures dans un tableau avec les montants.
```

GARDE-FOU

Sur notre exemple, l'IA a signalé l'écart d'elle-même. Rien ne garantit qu'elle le fasse chaque fois : elle peut le voir un jour et l'oublier le lendemain. Ne vous reposez pas sur ce signal.

03 Confirmé : un assistant qui refait les calculs

Dans ChatGPT (« Créer un GPT »), dans un Projet Claude, ou collé au début d'une conversation :

© SAKIRA. Reproduction et diffusion interdites sans accord écrit. · sakira.fr
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933

À COPIER

```
Tu transformes des factures en tableau pour un gérant qui n'est pas comptable.

Si tu ne le sais pas encore, demande en une question : quelles colonnes veut-il, et
pour quel usage (comptable, suivi des dépenses, autre) ?

Pour chaque facture, recopie les montants exactement comme ils sont écrits, sans les
arrondir, sans les corriger et sans en deviner un qui serait illisible : écris «
illisible ». Une ligne par facture : fournisseur, numéro, date, total HT, TVA, total
TTC.

Ensuite, pour chaque facture, fais le calcul toi-même, ligne par ligne : quantité x
montant unitaire, somme des lignes, HT + TVA. Écris le résultat de chaque calcul. Si
ton résultat diffère de ce qui est imprimé, écris « ÉCART » avec les deux chiffres, et
ne choisis pas lequel est juste.

Finis par le total de toutes les factures, et par la liste des factures à faire
vérifier. Aucun compliment.
```

GARDE-FOU

Sur la facture qui contient l'erreur, cette consigne rend trois écarts (HT, TVA, TTC) : c'est une seule erreur, comptée trois fois. L'IA ne tranche pas, et c'est voulu : le bon chiffre, c'est le fournisseur qui le donne.

04 Expert : un programme qui calcule à la place de l'IA

L'IA ne fait que recopier les chiffres dans un tableau. Les contrôles (montant écrit dans le PDF, quantité fois montant, lignes égales au HT, HT plus TVA égal au TTC) sont faits par le programme, sans IA.

Voir le programme `tableau-factures.sh` (64 lignes)

PROGRAMME À COPIER

```
#!/usr/bin/env bash
# ./tableau-factures.sh dossier-de-pdf
# Il faut pdftotext (paquet poppler). Les PDF doivent contenir du texte, pas être des scans.
# MOTEUR="codex exec --skip-git-repo-check" pour utiliser Codex
export LC_ALL=C   # sinon, sur un ordinateur en français, awk lit 4.50 comme 4
dossier="${1:-.}"; moteur="${MOTEUR:-claude -p --model haiku}"
command -v pdftotext >/dev/null || { echo "pdftotext est absent (brew install poppler)" >&2; exit 1;
}
sortie="tableau-$(date +%Y%m%d-%H%M%S)"; mkdir -p "$sortie"
: > "$sortie/tableau.csv"; : > "$sortie/alertes.txt"

for pdf in "$dossier"/*.pdf; do
  nom="$(basename "$pdf")"
  pdftotext -layout "$pdf" "$sortie/$nom.txt"
  # Même texte, avec les montants écrits à l'anglaise : 1 250,00 devient 1250.00
  perl -pe 's/(?<=\d) (?=\d{3}\b)//g; s/(?<=\d),(?=\d)/./g' "$sortie/$nom.txt" > "$sortie/$nom.norm"
  consigne="Voici le texte d'une facture. Rends une ligne par ligne de facture
et une ligne par total, sous cette forme exacte, séparée par des points-virgules,
sans titre et sans commentaire :
TYPE;libellé;quantité;montant unitaire;montant
TYPE vaut LIGNE, HT, TVA ou TTC. Les montants s'écrivent avec un point
décimal, sans espace et sans le signe euro (1250.00). Pour un total, laisse
vides la quantité et le montant unitaire. Recopie les chiffres tels qu'ils sont
écrits, même s'ils te semblent faux.
```

© SAKIRA. Reproduction et diffusion interdites sans accord écrit. · sakira.fr
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933

```
$(cat "$sortie/$nom.txt")"
  $moteur "$consigne" < /dev/null 2>>"$sortie/erreurs.log" |
    tr -d '`' | grep -E "^(LIGNE|HT|TVA|TTC);" | sed "s|^|$nom;|" >> "$sortie/tableau.csv"
done

echo "== Tableau =="; cat "$sortie/tableau.csv"; echo

# Contrôle 0 : chaque PDF a donné des lignes, et aucune ligne n'est en double
for pdf in "$dossier"/*.pdf; do
  nom="$(basename "$pdf")"
  grep -q "^$nom;" "$sortie/tableau.csv" || echo "AUCUNE LIGNE LUE : $nom (relancer)" >>
"$sortie/alertes.txt"
done
sort "$sortie/tableau.csv" | uniq -d | sed 's/^/LIGNE EN DOUBLE : /' >> "$sortie/alertes.txt"

# Contrôle 1 : chaque montant du tableau est écrit dans le PDF
while IFS=';' read -r f type lib q pu m; do
  for v in "$q" "$pu" "$m"; do
    [ -n "$v" ] || continue
    a="${v#-}"
    grep -qE "(^|[^0-9.])-?${a//./\\.}([^0-9]|$)" "$sortie/$f.norm" ||
      echo "ABSENT DU PDF : $f, $lib $type, valeur $v" >> "$sortie/alertes.txt"
  done
done < "$sortie/tableau.csv"

# Contrôle 2 : quantité x montant = montant ; lignes = HT ; HT + TVA = TTC
awk -F';' '
  { f=$1; t=$2 }
  t=="LIGNE" { s[f]+=$6
      if ($4*$5 - $6 > 0.005 || $6 - $4*$5 > 0.005)
        printf "CALCUL FAUX : %s, %s : %s x %s ne fait pas %s\n", f,$3,$4,$5,$6 }
  t=="HT" { ht[f]=$6 } t=="TVA" { tva[f]+=$6 } t=="TTC" { ttc[f]=$6 }
  END { for (f in ht) {
      d=s[f]-ht[f]; if (d>0.005||d<-0.005)
        printf "TOTAL FAUX : %s : les lignes font %.2f, la facture dit HT %.2f\n", f,s[f],ht[f]
      d=ht[f]+tva[f]-ttc[f]; if (d>0.005||d<-0.005)
        printf "TTC FAUX : %s : %.2f + %.2f ne fait pas %.2f\n", f,ht[f],tva[f],ttc[f]
      n++ } printf "Factures lues : %d\n", n }' "$sortie/tableau.csv" >> "$sortie/alertes.txt"

echo "== Alertes =="; cat "$sortie/alertes.txt"
echo "Détail dans le dossier $sortie"
```

Lancez `chmod +x tableau-factures.sh`, puis `./tableau-factures.sh mon-dossier`. Avec Codex : `MOTEUR="codex exec --skip-git-repo-check" ./tableau-factures.sh mon-dossier`. Le tableau est dans le fichier `tableau.csv` du dossier créé.

© SAKIRA. Reproduction et diffusion interdites sans accord écrit. · sakira.fr
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933



Figure 1. Le tableau lu dans les trois PDF d'exemple, puis l'alerte « TOTAL FAUX » sur la troisième facture.

> **GARDE-FOU**
> L'IA saute parfois une facture ou en répète une. Si le programme affiche « AUCUNE LIGNE LUE » ou « LIGNE EN DOUBLE », relancez : le tableau n'est pas fiable.

> **CE QUI N'A PAS ÉTÉ MONTRÉ**
> Le programme contrôle l'arithmétique et la présence des chiffres dans le PDF, pas leur sens : il ne vérifie pas que le taux de TVA est le bon, et un chiffre vrai pris sur une autre ligne passe.

Qui fait quoi

L'IA fait	Vous faites
Recopie les chiffres dans un tableau	Choisissez les colonnes et l'usage
(assistant) Refait les calculs et marque les écarts	Appelez le fournisseur pour savoir quel chiffre est juste
(programme) Contrôle sans IA ce qu'elle a recopié	Relancez quand une facture manque ou se répète
Ne sait pas quel chiffre est juste	Décidez de payer, de corriger ou de demander un avoir

© SAKIRA. Reproduction et diffusion interdites sans accord écrit. · sakira.fr
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933

Ce que ce guide ne fait pas

Il ne lit pas les PDF scannés (il faut d'abord une lecture optique du texte), et il ne vérifie pas le taux de TVA applicable. Avec une facture à plusieurs taux ou à plusieurs pages, ou un gros dossier, relisez le tableau par sondage. Pour la comptabilité, faites valider par votre expert-comptable. Les commandes du niveau 3 sont écrites pour un Mac.

© SAKIRA

© SAKIRA. Reproduction et diffusion interdites sans accord écrit. · sakira.fr
SAKIRA SAS · 6 rue d'Armaille, 75017 Paris · SIREN 108 493 933