

Paris, le 7 Juin 2016
HERAIZ-BEKKIS Daphné

Manuel 1 de stage :

Sujet de stage : Identités Nommées

Objet du manuel : Cahier des Charges

Langue : Français

Logiciel Bureautique Téléchargeable => + IHM

Input :

Fichier PDF représentant pour chacun une ou plusieurs factures

Output :

Fichier Excel (= Classeur)

un fichier par facture ?

Données à récupérer :

Info sur l'expéditeur de la facture	Nom, type (SARL...),capital de l'entreprise ?,Ville RCS, Num RCS, Adresse envoyeur, adresse siege ?,Tel+Fax, Num TVA (FR)
Info sur le destinataire de la facture	Nom, type (SARL...),à l'attention de?, Adresse receveur,Tel+Fax, Num TVA (FR)
Facture	Titre ?,Numero, date
Infos factures simplifiées	Bloc de texte prestations, montant total HT, taux TVA, mont total TVA, montant total autres taxes, montant total TTC
Infos factures (groupement de toutes les prestations)	montant HT, taux TVA, montant TVA, autres taxes, montant TTC(total), moyen de paiement ?
Infos pour une prestation	Nature prestation,montant HT, taux TVA, montant TVA, autres taxes, montant TTC(total prestation)

Deroulement :

Phase Classement Manuel

Entrer fichier de facture par fichier des données attendues que l'algorithme doit retrouver

Phase Utilisateur

A voir :

PDF → PNG ?

OCR

Retour d'une OCR ?

Séparation des factures ?

Lecture et classement des données (*)

Ecriture Sortie → Fichier CSV (fichier texte avec séparation ,)

CSV lisible partout et réadaptable par excel pour une lecture classeur

Conversion nécessaire en Excel ?

Ajout : Possibilité pour l'utilisateur de préciser la présence d'une erreur ce qui permettra au système d'ajouter un entraînement supplémentaire

Cote prog:

Installation d'OpenCV pour une OCR python ?

Phase Apprentissage

A voir :

(*)

Apprentissage et entraînement sur chaque données apprises et vérification avec les données classées correspondantes entrées manuellement

Cote prog :

Scikit-learn pour le machine learning ?

Threads ?