

75 rue de richelieu Paris 75002 -- > 0658374768

Tu as une étude de cas à préparer pour l'entretien, si tu as besoin de plus de temps, n'hésite pas.

Comme tu l'as sûrement vu dans l'intitulé, un des projets auquel tu participeras sera la récolte automatisée de données depuis des documents pdf. L'objectif n'est pas d'avoir en sortie la donnée exacte mais d'obtenir un document pdf simplifié. Séparer le document et restructurer les informations par relation sémantique, de sorte à faciliter la saisie d'analystes. Puis avoir dans une bibliothèque les bribes de pdf associées aux mots clés définis. Ainsi, classification de paragraphes et extraction de données depuis le texte et tableaux sont les objectifs de ce projet. Pour l'entretien, pourrais-tu préparer quelle démarche tu mettrais en place pour un tel projet, quels packages tu utiliserais etc. ? (Pas besoin de démarrer à coder)

L'objectif est de mettre en avant tes compétences et ta compréhension de la problématique. Tu prépares ce que tu veux pour le jour de l'entretien (extrait de code, présentation etc...)

Voici ci-joint un exemple de documents à traiter: <https://drive.google.com/file/d/1zJ-D34NMGFst2FtF1m3tuZShxbKAC0m/view?usp=sharing>. La partie qui nous intéresse est en général la déclaration de performance extra financière (le volet social). L'objectif est donc de séparer les informations de cette partie là, pour n'avoir que les informations pertinentes, à savoir les effectifs de la société, la pyramide des âges, les accidents de travail, le nombre de femmes, les politiques associées à la parité, aux personnes en situation de handicap etc...

Nous souhaitons avoir deux types d'informations en sortie: Des indicateurs clés (Effectifs, nombre de jeunes etc...) et des sections classifiées par thèmes (Insertion des jeunes, parité femmes-hommes etc...)

Si tu souhaites avoir d'autres documents exemples, cherche des documents d'enregistrement universel sur le web.

Je reste à ta disposition si tu souhaites avoir des informations complémentaires sur l'étude de cas. Et encore une fois, n'hésite pas à me prévenir si tu as besoin de plus de temps de préparation.

Bien cordialement,