**Difficultés rencontrées lors de la récupération de données sur Hellowork.fr :**

Le site Hellowork propose des offres d'emploi, la plupart provenant d'annonces sur d'autres sites web, sources d'employeurs. Il assure ainsi la liaison entre les annonceurs et les chercheurs d'emploi ou de stages.

Bien que le site soit dynamique, il offre la possibilité d'accéder à une page de recherche en modifiant l'URL d'entrée. Cela permet de contrôler la pagination en rendant l'URL mobile et en incluant les termes de recherche, le tout dans cette dernière. Cependant, lors de la première connexion, le site Hellowork demande d'accepter ou de refuser les cookies, ce qui renvoie null lors de l'utilisation de packages de scraping statique tels que Beautiful Soup. Pour contourner cette difficulté, nous avons utilisé Selenium, qui permet de reproduire les mouvements d'un navigateur.

Dans un premier temps, nous avons récupéré les informations sur la page de recherche. Cependant, toutes les informations recherchées se trouvent au sein de chaque annonce, accessible via une URL. La solution a consisté à récupérer sur chaque page les URLs des annonces, puis à extraire toutes les informations souhaitées de chaque annonce.

Les URLs des annonces sont affichées avec une extension HTML, permettant cette fois-ci l'utilisation du package Beautiful Soup. Cependant, nous avons rencontré quelques petits défis, notamment sur le type d'emploi et le lieu, qui ont tous deux la même sélection "soup.find('span')" avec des spans portant le même nom. Nous avons dû, au sein de leur div respective, récupérer tous les spans, puis les diviser en traitement de texte sur Python. Un traitement similaire a également été nécessaire pour retrouver la date.

Enfin, nous avons dû gérer les informations manquantes, telles que notre parseur HTML ne trouvant pas d'objet à l'endroit indiqué. En effet, il est possible qu'un élément tel que le type d'emploi ou, le plus souvent, le salaire soit absent de l'annonce. Notre solution a simplement été de vérifier avec des conditions si le parseur ne trouvait pas d'objet, il retournait None, et de même pour les autres informations.

Voici, dans les grandes lignes, non pas des difficultés, mais des défis que nous avons rencontrés en effectuant le scraping du site Hellowork.