Chapitre 1 : Présentation du cadre de projet

Introduction

Dans ce chapitre nous commençons par présenter l’organisme d’accueil « Bee coders » au sein duquel j’ai effectué ce projet de fin d’études par la suite je vais introduire les solutions existantes et leurs limites et finir par présenter la solution proposée.

1. **Présentation de l’organisme d’accueil**

Beecoders, fondée en 2020, est une entreprise de services numériques ayant achevée plus que 500 projets et plus que 1400 heures de support dans le secteur de consulting.

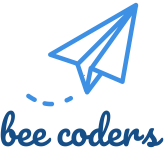
 Beecoders est spécialisée dans le développement web et mobile des solutions digitales répondant aux besoins du client (Android et IOS), le design web, le consulting IT pour booster les performances des employés dans le secteur d’IT et les formations à distance à travers la plateforme « Bridge ».

Figure 1 : Logo Bee coders

1. **Etude de l’existant**

Dans le sillage de la pandémie mondiale, le recrutement à distance est une pratique   
inévitable. L'objectif des entreprises est de pouvoir maintenir ses activités face aux bouleversements du monde du travail. Pour accompagner cette évolution du recrutement et optimiser sa recherche d'emploi, les demandeurs d'emploi ont également l'obligation de savoir manipuler les canaux digitaux. Pour y voir plus clair, nous allons exposer l'évolution du e-recrutement en Tunisie, ainsi que les sites de recrutement les plus connus.

De nombreuses entreprises prennent l'habitude d'embaucher des employeurs en ligne pendant la vague de coronavirus. Les entreprises placent des offres d'emploi sur de nombreuses plateformes pour cibler les meilleurs talents selon leurs propres conditions. De plus, la dématérialisation du recrutement facilite certaines activités qui peuvent être réalisées en freelance ou à domicile. La crise sanitaire a également facilité le passage au e-recrutement ou recrutement en ligne.

Ceux qui envisagent de recruter publient leurs offres d'emploi sur plusieurs portails de recrutement en ligne et même dans des journaux locaux pour découvrir des candidats dignes de ce nom. Le processus de recrutement peut se dérouler en plusieurs étapes : tri des lettres de motivation et CV, tests, entretiens en visioconférence pour présélectionner les candidats et exprimer les attentes des recruteurs, formation aux nouveaux outils à distance.

En conséquence, le recrutement en ligne peut mieux cibler les candidats avec des offres d`emploi largement diffusées, réduisant ainsi la charge de travail des recruteurs. Les offres d'emploi sont également accessibles 24 heures sur 24, 7 jours sur 7 et peuvent être mises à jour.

En Tunisie par exemple, on retrouve un certain nombre de plateformes qui exploitent le potentiel d’internet dans la recherche du candidat idéal. Citons entre autres emploi.nat, keejob, Tanitjob, Tunisie travail, Linkedin. Ces plateformes de recrutement en ligne offre des centaines de nouvelles offres chaque jour.

Citons par exemple :

* **La plateforme emploi.nat.tn :**C’est le site d’offres d’emploi de l’Etat ou du réseau de l’ANETI. Il propose des offres et demandes d’emploi, des avis de concours, des annonces de presse, des appels d’offres et résultats, des programmes d’encouragement à l’emploi, de présentation de l’ANETI, projet BAD, coopération internationale etc. On peut également y trouver un espace candidat, un espace promoteur et un espace entreprise ainsi qu’un espace pour les investisseurs et le travail indépendant. Les membres disposent d’un login et mot de passe pour accéder au site.
* [**Keejob**](https://www.tanitjobs.com/blog/355/les-sitesd-emploi-en-tunisie/)**:** Ce site d’offres d’emploi propose également un espace candidat et espace entreprise avec inscription. Les candidats peuvent mettre leur CV sur le site et paramétrer leur compte. Les entreprises peuvent aussi gérer leurs annonces et recevoir des candidatures
* [**Tunisie-travail.net**](https://www.tanitjobs.com/blog/355/les-sitesd-emploi-en-tunisie/)**:** Cette portail d’offres d’emploi diffuse des concours de la fonction publique, des offres d’emploi actualisées en Tunisie, en freelance et à l’étranger. Il permet aux recruteurs de publier gratuitement leurs offres d’emploi.
* [**Tanitjobs**](https://www.tanitjobs.com/)**:** Un site d’offres d’emploi spécialisé. Les candidats peuvent déposer leur CV sur le site. Les recruteurs peuvent aussi s’inscrire sur le site pour dénicher rapidement le profil en adéquation à leurs besoins. Le site propose une expérience utilisateur satisfaisante en permettant à ce dernier de trier parmi beaucoup de critères l’offre adaptée à ses besoins. Les offres se déclinent aussi en diverses catégories. Le site est mis à jour d'une façon régulière.
* [**Linkedin**](https://www.tanitjobs.com/)**:** Un site de réseautage professionnel. c'est tout à la fois un réseau social et un site d’emploi. Vous l'utilisez pour rester en contact avec vos relations professionnelles et rechercher de nouvelles opportunités de carrière. Mais pour les recruteurs, LinkedIn est un moteur de recherche plus qu'autre chose.

1. **Problématique**

La grande majorité des sociétés qui recrutent décident de contacter les anciens employeurs indiqués dans le CV, surtout s’il s’agit d’un poste à responsabilité. Cela leur permet de vérifier l’authenticité des renseignements sur les profils et le comportement du candidat. Il y a également des postes qui imposent le test psychotechnique ou le test graphologique. L’autre problématique de ce processus,

C’est que les chercheurs d’emplois se démotivent à faire de recherche sur le net dès qu’ils ont envoyé quelques demandes sans suite. Ils se souviennent de l’entreprise qui ne répond pas aux demandes.

Ce type de recrutement a aussi le désavantage d’être moins efficace sur la fidélisation des candidats embauchés notamment s’ils sont des jeunes de la génération Y et Z. Du fait de leur maîtrise des nouvelles technologies et de leur créativité, ils ne restent pas longtemps dans une entreprise.

En outre, l’inexistence de statistiques sur ces sites d’embauche se traduit par le manque d’informations nécessaires créant ainsi un besoin à étudier. En effet les données publiées sont stockées et non analysées, à travers des statistiques primordiales les professionnels seront capables de se différencier et se positionner dans le marché ce qui n’est pas possible sans analyse.

1. **Critique de l’existant**

La dispersion des données sur plusieurs sites et le manque d’analyse au sein de ses sites nous a permis de dégager les défaillances suivantes :

* Perte de temps : Le fait que les offres d’emploi sont éparpillées, il est difficile de les accéder toutes à la fois. Par conséquence on a une nécessité de faire le tour des offres disponibles sur le marché ce qui entraine une énorme perte de temps pour les étudiants et les professionnels.
* Manque de visibilité globale sur le marché : Vu que certaines entreprises choisissent de publier leurs offres sur un site bien précis, les chercheurs risquent par conséquent un accès limité d’offres.

* Inexistence d’analyses :Les utilisateurs ne sont pas capables de se positionner sur le marché à cause de l’absence d’analyse et statistiques des offres sur ses sites

1. **Solution proposée**

Après une étude de l’existant, j’ai proposé une collecte de données nécessaires qui sont dispersées sur différents sites. Les publications de chaque site ont été récoltées en appliquant sur ces derniers la méthode WEB SCRAPING.

Par suite un traitement a été effectué sur ces données dans le but d'offrir une information qui est de meilleure qualité pour une utilisation ultérieure.

Ces données ont abouti à une aptitude de repérer en temps réel les KPI (indicateurs de performance) des propositions d’embauche afin de suivre les éléments clefs.

Cette distinction facilite la recherche optimale pour le profil du demandeur.

En effet ces KPI construisent un Tableau de bord qui est un outil de pilotage aidant à la prise de décision, dans le but de mettre des analyses complexes à la portée des professionnels et des étudiants car ceux-ci n'ont pas toujours la capacité d'interpréter la data brute qu’est les offres d’emploi.

Une représentation simple et générale est de la sorte idéale pour ces professionnels. Tel qu’énoncé dans cette partie je propose cette solution qui assure la pertinence des données, l’efficience et la rapidité de l’affaire de recherche d’emploi.

1. **Méthodologie de travail**

La méthodologie Agile est une méthode de gestion de projet basée sur un cycle de développement centré sur le client. Le client est impliqué dans la mise en œuvre du début à la fin du projet. Grâce aux méthodes agiles, les demandeurs bénéficient d'une meilleure visibilité sur la gestion des tâches qu'avec les méthodes traditionnelles. L'influence du client dans le processus permet à l'équipe d'obtenir des commentaires réguliers afin que les changements nécessaires puissent être appliqués directement.

Le principe de base est de fournir une version minimale du logiciel puis par un processus itératif d'intégrer des fonctionnalités supplémentaires dans cette base. Le processus itératif regroupe une série d'instructions, se répétant autant de fois que nécessaire selon le besoin.

* 1. **Pourquoi SCRUM**

Scrum est une méthodologie de développement agile utilisée dans le développement de logiciels basée sur un processus itératif et incrémental.

L'objectif principal de Scrum est de satisfaire le besoin du client à travers un environnement de transparence dans la communication, la responsabilité collective et le progrès continu. Le développement part d'une idée générale de ce qu'il faut construire, en élaborant une liste de caractéristiques ordonnées par priorité (product backlog) que le propriétaire du produit souhaite obtenir.

* + 1. **Les acteurs du SCRUM**

**Product owner**

Le *product owner* (PO) est le représentant des clients et utilisateurs dans le cadre du projet. Il est en charge de la tenue du backlog produit, notamment de l'ordre dans lequel vont être traitées les éléments de ce backlog (user stories, fonctionnalités à développer). En d'autres termes, il est « responsable » du succès du produit.

**Scrum master**

Le scrum master est l’animateur ou le coach de l’équipe technique, chargé principalement de définir les rôles et les objectifs en assurant que le principe scrum se déroule convenablement.

**L’équipe de développement**

L'équipe de développement est responsable de la conversion des exigences définies par le propriétaire du produit en fonctions utilisables. Il est multidisciplinaire et possède toutes les compétences nécessaires pour exécuter le projet sans avoir besoin de faire appel à des services externes.

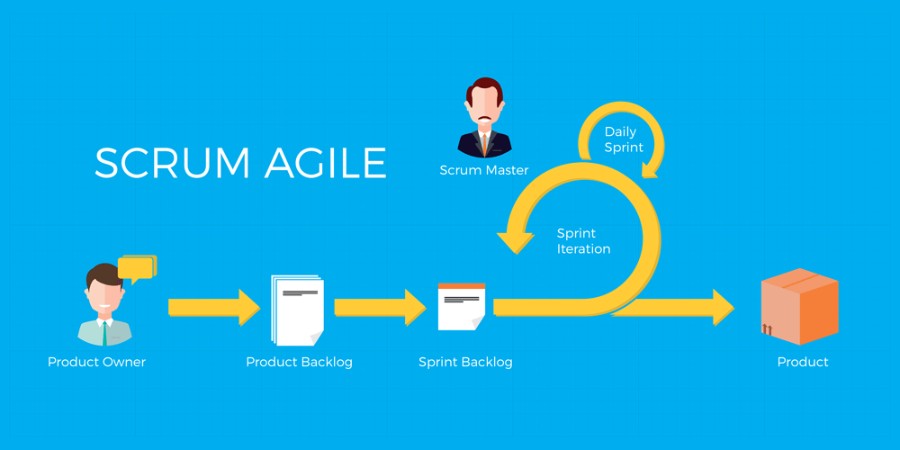


Figure 2 : Cycle de vie projet Scrum

* 1. **Méthodologie de conception**

L’UML est une abréviation de Unified Modeling Language, est un langage de modélisation standardisé composé d'un ensemble intégré de diagrammes, développé pour aider les développeurs de systèmes et de logiciels à spécifier, visualiser, construire et documenter la conception des systèmes logiciels, ainsi que pour la modélisation commerciale et d'autres systèmes non logiciels.

L'UML représente un ensemble de meilleures pratiques d'ingénierie qui ont fait leurs preuves dans la modélisation de systèmes vastes et complexes. L'UML est une partie très importante du processus de développement de logiciels. Il utilise principalement des notations graphiques pour exprimer la conception de projets logiciels.

* 1. **Conclusion**

Dans ce chapitre, j’ai présenté l’organisme d’accueil, critiqué l’existant et ses limites et présenté la solution que j’ai envisagée.

Pour conclure je vise la conception de cette solution qui pourra remédier aux éléments existants et aux limites étudiés activant ainsi l’accès à un modèle d’analyse important pour les professionnels.

Dans le chapitre suivant, je vais étudier les besoins et la conception de mon système.

Chapitre 2 : Etat de l’art

Introduction

Dans ce chapitre on va présenter le web scraping en citant son histoire,

ces outils et conclure par identifier la solution proposée dans ce projet

1. **Représentation de l’informatique décisionnelle (BI)**

Le terme business intelligence (BI) ou business intelligence fait référence à une application, une infrastructure, un outil ou une pratique qui donne accès à l'information et permet de l'analyser pour améliorer et optimiser la prise de décision et la performance d'une entreprise.​

Ainsi, L’informatique décisionnelle rassemble une variété d'outils, d'applications et de méthodologies pour collecter des données à partir de systèmes internes et de sources externes, les préparer pour l'analyse, développer des données et lancer des requêtes sur ces ensembles de données. Utilisez ensuite ces outils pour créer des rapports, des tableaux de bord et des visualisations de données afin de mettre les analyses à la disposition des décideurs.

La capacité d'analyser et d'agir sur les données est de plus en plus importante pour les entreprises. Le rythme du changement exige que les entreprises soient capables de réagir rapidement aux demandes changeantes des clients et des conditions environnementales.

1. **Web scraping**

A l'ère du big data, la collecte de données (textes, images, fichiers hébergés) sur Internet devrait être au centre de toutes les réflexions pour les années à venir. Le web scraping ou le screen scraping est le processus d’extraction des données des sites web. Ce sont des scripts, des programmes informatiques, qui sont chargés d'extraire des informations. D’abord cette technique permet de réutiliser le contenue déjà utilisé dans un site web dans le but de l'afficher sur un autre en multipliant par conséquence sans effort le nombre de page disposant d’un même contenu.

En addition, l’augmentation exponentielle du marché de e-commerce entraine la nécessité d’être bien informé sur les concurrents et dans ce sens le web scraping peut être utilisé comme un outil de surveillance des concurrents (veille concurrentielle).

* 1. **L’histoire du Web scraping**

Le Web scraping ou Web harvesting est une technique d'indexation. La plupart l'associent à l'extraction de données Web, le moyen le plus efficace et le plus simple de copier de gros volumes d'informations en ligne. Cependant, le web scraping est né dans un but complètement différent et qu'il a fallu près de deux décennies pour qu'il se transforme en web scraping que nous connaissons maintenant.

* + 1. **La création de Wanderer**

En 1993, le premier concept de Crawler était né .World Wide Web Wanderer développé par Matthew Gray a l’institut de technologies de Massachussets était le premier robot d'exploration Web basé sur Perl dont le seul but était de mesurer la taille du Web. De plus Wanderer a été utilisé pour générer un index appelé Wandex qui lui a donné le potentiel de devenir le premier moteur de recherche universel du World Wide Web.

* + 1. **La création de JumpStation**

Dans la même année 1993, JumpStation était né et est devenu le premier moteur de recherche Web basé sur un robot. Grâce à lui, des millions de pages Web ont été indexées de sorte qu’Internet s'est transformé en une plate-forme open source de données sous diverses formes.

* + 1. **La création de BeautifulSoup**

En 2004, l’HTML parser BeautifulSoup a été créer, c’est une bibliothèque d'algorithmes couramment utilisés écrits en langage de programmation Python. BeautifulSoup a aidé à saisir le sens de la structure du site et à analyser le contenu dans les conteneurs HTML ce qui facilite au programmeur d’économiser des heures de travail. Internet est devenu cette immense source d'informations à laquelle toute personne disposant d'un ordinateur et d'une connexion Internet avait accès, les internautes avaient commencé à en tirer parti en extrayant les informations dont ils disposaient.

En effet, la quantité de données téléchargées de nos jours ne peut pas être manuellement copiée ou collée ce qui exige que d'autres moyens d'obtention d'information devaient être développés**.**

* + 1. **L’apparition des Web Scrapers visuels**

Estimé en 2006, le web scraping tel que nous le connaissons était né. Le logiciel de Web scraping visuel Web Integration Platform version 6.0 qui a été lancé par Stefan Andresen, a permis aux utilisateurs de mettre en évidence les informations nécessaires d'une page Web et de structurer ces données dans un fichier utilisable ou une base de données permettant aux non-programmeurs de rejoindre et extraire facilement des données du Web.

* 1. **Les outils du Web scraping**

Il existe plusieurs types d'outils qui peuvent être utilisés pour effectuer du web scraping: Des outils open source qui réduisent le coût du web scraping et facilitent son exécution. Les outils les plus couramment utilisés sont Scrapy, Selenium, BeautifulSoup et Puppeteer. En ce qui suit une comparaison entre les trois outils les plus utilisés pour le Web scraping :

Beautiful Soup est une bibliothèque Python de grattage Web pour extraire des données de fichiers HTML et XML, qui prend également en charge Python 3 et Python 2.4.

Scrapy est un framework Python pour le scraping Web à grande échelle. Il vous donne tous les outils dont vous avez besoin pour extraire efficacement les données de votre site Web, les traiter comme vous le souhaitez et les stocker dans la structure et le format de votre choix.

Selenium est un outil open source pour automatiser les navigateurs Web. Il fournit une interface unique qui vous permet d'écrire des scripts de test dans des langages de programmation tels que Ruby, Java, NodeJS, PHP, Perl, Python et C#.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Extension** | **Performance** | **Ecosystème** |
| **Scrapy** | Utiliser pour les projets larges et complexes | Les bibliothèques existantes ne peuvent pas battre les performances de Scrapy grâce à sa fonctionnalité intégrée, à savoir l'utilisation d'appels système asynchrones. | Ayant un bon écosystème, nous pouvons utiliser des proxys et des VPN pour automatiser la tâche.  C'est l'une des raisons du choix de la bibliothèque pour les projets complexes. |
| **Beautiful Soup** | Utiliser pour les petits projets et à basse complexité | Beautiful Soup est assez lent pour effectuer une certaine tâche, mais nous pouvons surmonter ce problème à l'aide du concept de multithreading | Cette bibliothèque a beaucoup de dépendances dans l'écosystème.  C'est l'un des inconvénients de cette bibliothèque pour un projet complexe. |
| **Selenium** | Utiliser pour une taille de donnés limitée | Il peut gérer jusqu'à une certaine plage mais n'est pas équivalent à Scrapy. | Caractérisé par un bon écosystème pour le développement, mais les proxys ne peuvent pas être utilisés très facilement. |

1. **Visualisation des données**
   1. **Définition**

Face à des quantités de données importante, il est important d’utiliser les outils les mieux adaptés pour les rendre compréhensibles. En effet le rôle de la Data visualisation est d’[éclairer les informations récoltées](https://ia-data-analytics.fr/solution-bi/datavisualisation/datavisualisation-role-clef-data-scientist/) en les traduisant par des représentations visuelles claires et accessibles à tous. Contrairement à des données brutes ou des tableurs complexes remplis de données, la Data Visualisation va permettre de créer des visuels qui permettent de comprendre les informations essentielles rapidement.

* 1. **Création de tableau de bord**

Dashboard est la traduction anglaise du terme "tableau de bord". Il s'agit donc d'une représentation visuelle d'informations importantes sur le même écran. Il peut mesurer la performance d'une entreprise et évaluer son organisation pour atteindre un ou plusieurs objectifs.