



RAPPORT DE STAGE

Filière : « Sciences des Données, Big Data & IA »

Développement d'un module intelligent d'analyse de CV

Présenté par

Mariam Mouh

Soutenue le..... à l'Heure

Devant le jury

Pr	Professeur à l' ENSIASD de Taroudant	Président
Pr	Professeur à l' ENSIASD de Taroudant	Examineur
Pr	Professeur à l' ENSIASD de Taroudant	Examineur
Pr	Mohammed Kassri	Encadrant
Pr/Mr	Neuman Charhbili	Encadrant

Année Universitaire 2024- 2025

DEDICACES

Je dédie ce travail avec tout mon cœur à :

Mes chers parents et à toute ma famille, pour votre amour, votre patience et votre soutien sans limite. Merci d'avoir toujours cru en moi et d'avoir été là dans chaque étape de mon parcours.

À mes amis et collègues, merci pour vos encouragements et votre présence tout au long de ce parcours.

À mes enseignants de l'École Nationale Supérieure d'Intelligence Artificielle et Science des Données de Taroudant, merci pour votre accompagnement et vos précieux conseils.

Et à toute l'équipe de VALA Blue, encadrants et employés, merci pour votre accueil, votre gentillesse et votre aide pendant mon stage.

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à Monsieur Mohammed Kassri, mon encadrant académique à l'École Nationale Supérieure d'Intelligence Artificielle et Science des Données de Taroudant, pour son accompagnement, ses conseils avisés et sa disponibilité tout au long de ce projet. Son soutien m'a permis d'avancer avec confiance et rigueur dans ce travail.

Je remercie également Monsieur Neuman Charhbili, mon encadrant au sein de l'entreprise VALA Blue, pour son encadrement bienveillant, sa patience et ses conseils techniques précieux. Son aide m'a permis de mieux comprendre le monde professionnel et de mettre en pratique mes connaissances dans un cadre réel et stimulant.

Je remercie l'ensemble des enseignants, le personnel administratif et la direction de l'École Nationale Supérieure d'Intelligence Artificielle et Science des Données de Taroudant pour leur accompagnement tout au long de ma formation. Grâce à eux, j'ai pu acquérir des compétences solides, aussi bien sur le plan technique qu'humain.

Je souhaite également remercier chaleureusement toute l'équipe de VALA Blue, ainsi que ses employés, pour leur accueil, leur gentillesse et leur disponibilité. Travailler à leurs côtés a été une expérience très enrichissante, tant sur le plan professionnel que personnel.

Enfin, je remercie toutes les personnes qui m'ont soutenue de près ou de loin durant ce projet. Que ce soit par un mot d'encouragement, un conseil ou simplement une écoute, votre présence m'a été précieuse.

RESUME

Ce présent rapport présente les différentes étapes réalisées lors de mon stage d'initiation effectué au sein de l'entreprise VALA Blue. L'objectif de ce stage était de développer un outil intelligent capable de lire un CV au format PDF, d'en extraire les informations principales, puis de les analyser afin d'évaluer leur correspondance avec les critères d'un poste défini.

Le travail a porté sur la mise en place d'un outil qui permet d'extraire le texte depuis un fichier PDF, de nettoyer les données, de les organiser par catégories (compétences, formations, expériences, etc.), puis de les analyser automatiquement. Pour cela, j'ai utilisé plusieurs technologies telles que Python, pdfplumber pour l'extraction de texte, spaCy pour le traitement du langage naturel, et Streamlit pour la création de l'interface.

Ce stage m'a permis de découvrir le fonctionnement d'un environnement professionnel et de mettre en pratique mes connaissances dans le domaine de l'intelligence artificielle et du traitement automatique du langage.

TABLE DES MATIERES

DEDICACES -----	2
REMERCIEMENTS-----	3
RESUME -----	4
TABLE DES MATIERES -----	5
LISTE DES ABREVIATIONS -----	7
LISTE DES FIGURES -----	9
LISTE DES TABLEAUX -----	10
INTRODUCTION GENERALE-----	1
CHAPITRE1 : ENTREPRISE D’ACCUEIL -----	2
I. Introduction :-----	2
II. Présentation de l’entreprise : -----	2
II.1 Position géographique :-----	2
II. 2 Vision et idée :-----	3
II. 3 Fiche technique de l’entreprise : -----	4
I. Organisation et structure :-----	4
I.1 Organisation hiérarchique de l’entreprise : -----	4
I.2 Architecture de la plateforme Vala multi-serveurs :-----	5
II. Les services de l’entreprise :-----	7
III. Le service d’affectation :-----	8
IV. Conclusion :-----	9
CHAPITRE 2 : SPÉCIFICATIONS-----	10
I. Introduction :-----	10
II. Présentation du projet : -----	10
II.1 Contexte général du projet :-----	10
II. 2 Les objectifs de la plateforme :-----	10
III. Cahier de charge :-----	11
III.1 Problématique : -----	11
III.2 Les acteurs : -----	11
III.3 Les besoins fonctionnels :-----	11
III.4 Les besoins non fonctionnels :-----	12
III.5 Règles de gestion :-----	13
III.6 Nom et Logo de site : -----	14
IV. Conclusion :-----	14
CHAPITRE 3 : ANALYSE ET CONCEPTION -----	15

Tables des matières

I.	Introduction :-----	15
II.	Diagramme de cas d'utilisation :-----	15
III.	Diagramme de séquence :-----	16
IV.	Diagramme d'activités :-----	18
V.	Conclusion :-----	20
CHAPITRE 4 : ENVIRONNEMENT DE TRAVAIL ET OUTILS-----		21
I.	Introduction :-----	21
II.	Outils de travail :-----	21
II.1	Visual studio code :-----	21
II.2	Git et GitHub :-----	22
II.3	Enterprise Architect :-----	22
III.	Outils de développement :-----	23
III.1	Python :-----	23
III.2	Streamlit:-----	23
III.3	pdfplumber:-----	24
III.4	Regex :-----	24
III.5	spaCy:-----	24
III.6	pandas:-----	25
IV.	Conclusion :-----	25
CHAPITRE 5 : RÉALISATION DE PROJET -----		26
I.	Introduction :-----	26
II.	Présentation de l'interface :-----	26
II.1	En-tête de l'interface :-----	26
II.2	Le formulaire des exigences :-----	27
II.3	Le champ du dépôt des CV :-----	28
II.4	Tableau des résultats :-----	28
II.5	Exportation des résultats :-----	29
II.6	Répartition détaillée pour chaque candidat :-----	30
III.	Difficultés rencontrées lors de la réalisation :-----	31
IV.	Conclusion :-----	31
CONCLUSION GENERALE ET PERSPECTIVES -----		32
ANNEXE A-----		34

LISTE DES ABREVIATIONS

Abréviation	Signification
CV	Curriculum Vitae
PDF	Portable Document Format
S.A.R. L	Société à Responsabilité Limitée
PME	Petites et Moyennes Entreprises
STE	Société
SOE	Optimisation pour les moteurs de recherche (Search Engine Optimization)
RH	Ressources Humaines
CDN	Content Delivery Network (Réseau de distribution de contenu)
DDoS	Distributed Denial of Service (Déni de service distribué)
SSD	Solid State Drive (Disque à état solide)
MySQL	My Structured Query Language (Système de gestion de base de données)
FTP	File Transfer Protocol (Protocole de transfert de fichiers)
PHP	Hypertext Preprocessor (langage de programmation web côté serveur)
CMS	Content Management System (Système de gestion de contenu)

Liste des abréviations

SSL	Secure Sockets Layer (Protocole de sécurisation des échanges sur Internet)
WYSIWYG	« What You See Is What You Get » (Éditeur de pages visuel)
E-commerce	Commerce électronique / Vente en ligne
NLP	Natural Language Processing (Traitement du langage naturel)
UML	Unified Modeling Language (Langage de modélisation unifié)
NER	Named Entity Recognition (Reconnaissance d'entités nommées)
CSV	Comma-Separated Values (Valeurs séparées par des virgules) – format de fichier très utilisé pour stocker des données tabulaires

LISTE DES FIGURES

Figure 1.1:Logo de l'entreprise.....	2
Figure 1.2:Position géographique de l'entreprise.....	3
Figure 1.3:Organisation hiérarchique de l'entreprise.....	5
Figure 1.4:Architecture de la plateforme Vala multi-serveurs.....	6
Figure 2.1:Logo du système.....	14
Figure 3.1:Diagrammes des cas d'utilisation du système.....	16
Figure 3.2:Diagramme de séquence du système.....	17
Figure 3.3:Diagramme d'activités du système.....	19
Figure 4.1:Logo de Visual studio code.....	21
Figure 4.2:Logo de Git.....	22
Figure 4.3:Logo de Github.....	22
Figure 4.4:Logo de Enterprise Architect.....	22
Figure 4.5:Logo de Python.....	23
Figure 4.6:Logo de Streamlit.....	24
Figure 4.7:Logo de SpaCy.....	25
Figure 4.8:Logo de Pandas.....	25
Figure 5.1:En-tête de l'interface.....	26
Figure 5.2:Première section du formulaire des exigences.....	27
Figure 5.3:Deuxième section du formulaire des exigences.....	27
Figure 5.4:Le champ du dépôt des CV.....	28
Figure 5.5:Tableau des résultats.....	28
Figure 5.6:Tableau des résultats après tri.....	29
Figure 5.7:Exportation des résultats.....	29
Figure 5.8:Exemple de fichier des résultats.....	30
Figure 5.9 : Répartition détaillée pour chaque candidat.....	30
Figure 5.10 : Répartition détaillée pour chaque candidat (suite).....	31

LISTE DES TABLEAUX

<i>Tableau 1.1:Fiche technique de l'entreprise.....</i>	<i>4</i>
---	----------

INTRODUCTION GENERALE

Dans le monde professionnel, lorsqu'un poste est publié, il peut recevoir un très grand nombre de candidatures. Cette situation représente un véritable défi pour les recruteurs, qu'ils soient en équipe ou seuls, car il devient difficile de parcourir manuellement tous les CV et de vérifier si l'expérience et les compétences des candidats correspondent aux exigences du poste. C'est dans ce contexte que j'ai développé une solution intelligente permettant de réduire considérablement le temps consacré à cette phase initiale de tri. Cette solution consiste à analyser automatiquement les CV en format PDF afin d'en extraire les informations clés, puis de les comparer aux critères attendus par l'employeur.

Ce projet a été réalisé dans le cadre de mon stage d'initiation au sein de l'entreprise VALA Blue, une société qui propose divers services numériques liés à la présence web, tels que l'hébergement, la messagerie professionnelle et la création de sites. L'objectif principal de ce stage était de concevoir une application simple, efficace et accessible, capable d'assister les recruteurs dans la sélection initiale des candidats en automatisant l'analyse des CV.

Ce rapport est structuré en cinq chapitres principaux :

Le premier chapitre présente l'entreprise VALA Blue, sa mission, son organigramme et sa fiche technique.

Le deuxième chapitre explore les fonctionnalités développées, le cahier des charges, ainsi que l'idée générale du projet.

Le troisième chapitre détaille le processus de conception de la plateforme, en exposant les étapes, les choix techniques et les solutions apportées.

Le quatrième chapitre dresse une liste complète des outils et technologies utilisés durant le développement.

Enfin, le cinquième chapitre décrit le produit final et démontre comment il répond aux besoins identifiés tout en facilitant le travail de présélection des recruteurs.

CHAPITRE1 : ENTREPRISE D'ACCUEIL

I. Introduction :

Ce chapitre présente l'entreprise d'accueil en mettant en avant son identité, son organisation et son environnement. J'aborderai successivement ses informations générales, sa structure interne, ses produits et marchés, ainsi que sa position géographique et sa nature juridique.

II. Présentation de l'entreprise :



Figure 0.1: Logo de l'entreprise

Creative Internet Solutions est une société située à Agadir, créée le 30 septembre 2020. Elle est enregistrée sous la forme juridique S.A.R.L et se spécialise dans l'hébergement et la création de solutions digitales. Opérant dans le secteur de la production digitale, l'entreprise offre des services innovants adaptés aux besoins numériques actuels. Creative Internet Solutions fonctionne également sous la franchise Valla Blue, qui dispose d'autres départements notamment au Royaume-Uni, où elle est enregistrée en tant qu'entité légale.

II.1 Position géographique :

Valla blue est stratégiquement localisée à Immeuble Safwa, Boulevard hassan 1er , 2e étage N°6 Dakhla , dans la ville dynamique d'Agadir, portant le code postal 80000.

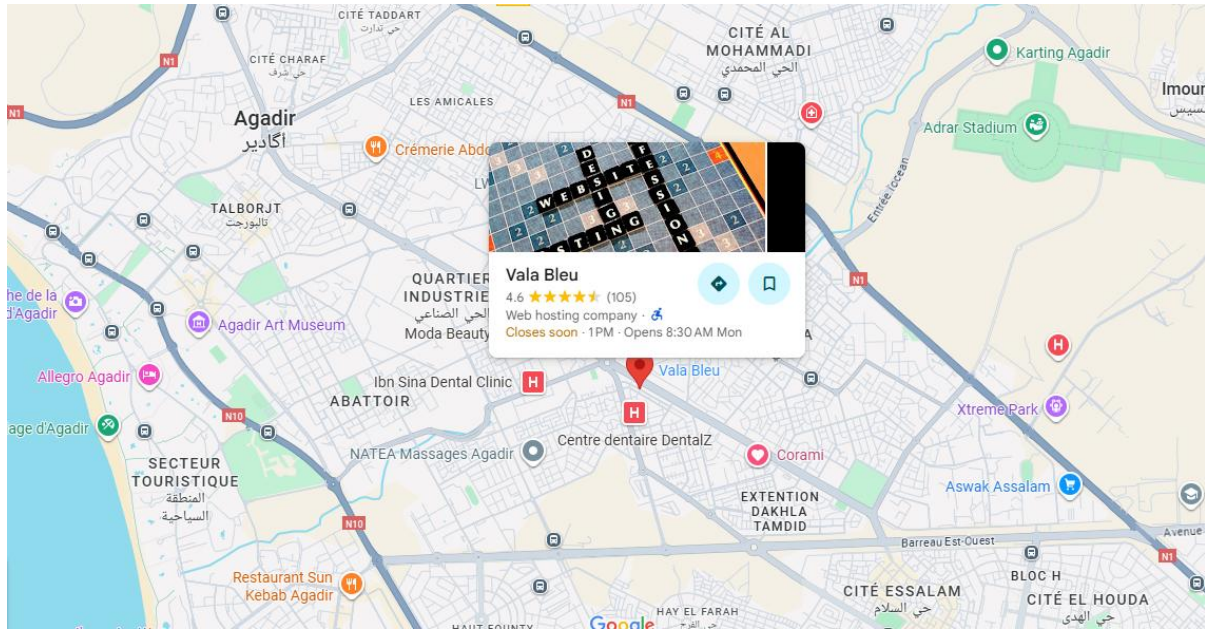


Figure 0.2: Position géographique de l'entreprise

II. 2 Vision et idée :

VALA BLEU s'est imposée comme un acteur majeur en Europe, en Amérique et en Afrique du Nord dans le domaine de l'hébergement web et de l'enregistrement de noms de domaine. Sa vision est de faciliter la présence en ligne des PME et des professionnels en offrant des solutions à la fois économiques et techniques, grâce à une infrastructure de qualité et un processus automatisé de gestion des serveurs. Partant d'un seul serveur, l'entreprise dispose aujourd'hui d'un réseau rapide, sécurisé et étendu, garantissant fiabilité et performance. Investissant continuellement dans la recherche et le développement, VALA BLEU propose des offres compétitives et innovantes qui accompagnent ses clients dans leur succès sur Internet.

II. 3 Fiche technique de l'entreprise :

Tableau 1.1:Fiche technique de l'entreprise

Date de Création de la STE	30/09/2020
NOM de la STE	Creative Internet Solutions
Forme Juridique	S.A.R.L
Les Activités de la STE	Hébergement et Création De Solutions Digitales
Secteur D'Activité	Production Digital
Adresse	Immeuble Safwa, Boulevard hassan 1er , 2e étage N°6 Dakhla – Agadir 80000 Maroc
Téléphone	00 212 528 21 3045 / 00 212 528 22 5522
Effectif De L'Entreprise	10
Directeur Général	Mr. Kayouh Abdelbast

I. Organisation et structure :

I.1 Organisation hiérarchique de l'entreprise :

L'entreprise est dirigée par Mr. Abdelbast Kayouh, Organisateur Général, qui supervise l'ensemble des services. La structure comprend plusieurs pôles : le pôle Web et Projets, dirigé par Mr. Khalid Jbour, assisté de Mr. Wafafa Fouks et du service comptabilité ; la direction et administration, assurée par Mme Asma El Boudi et Mr. Bachir El Omari (comptabilité et ressources humaines) ; le pôle commercial et relation client, avec Mme Fatiha Disky (responsable commercial) et Mme Hanane Bayen (service clientèle) ; ainsi que le pôle SOE, dirigé par Mr. Neumann Charhbili, assisté de M. Kamna Chaoui et du service de gestion relation client.

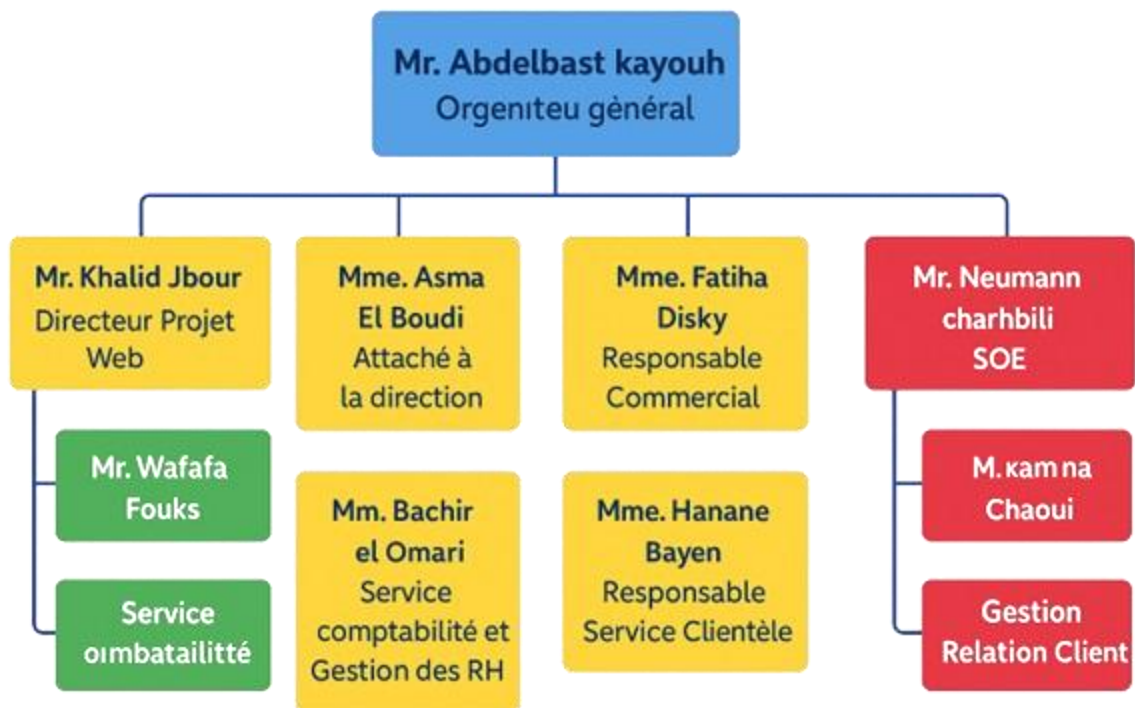


Figure 0.3: Organisation hiérarchique de l'entreprise

I.2 Architecture de la plateforme Vala multi-serveurs :

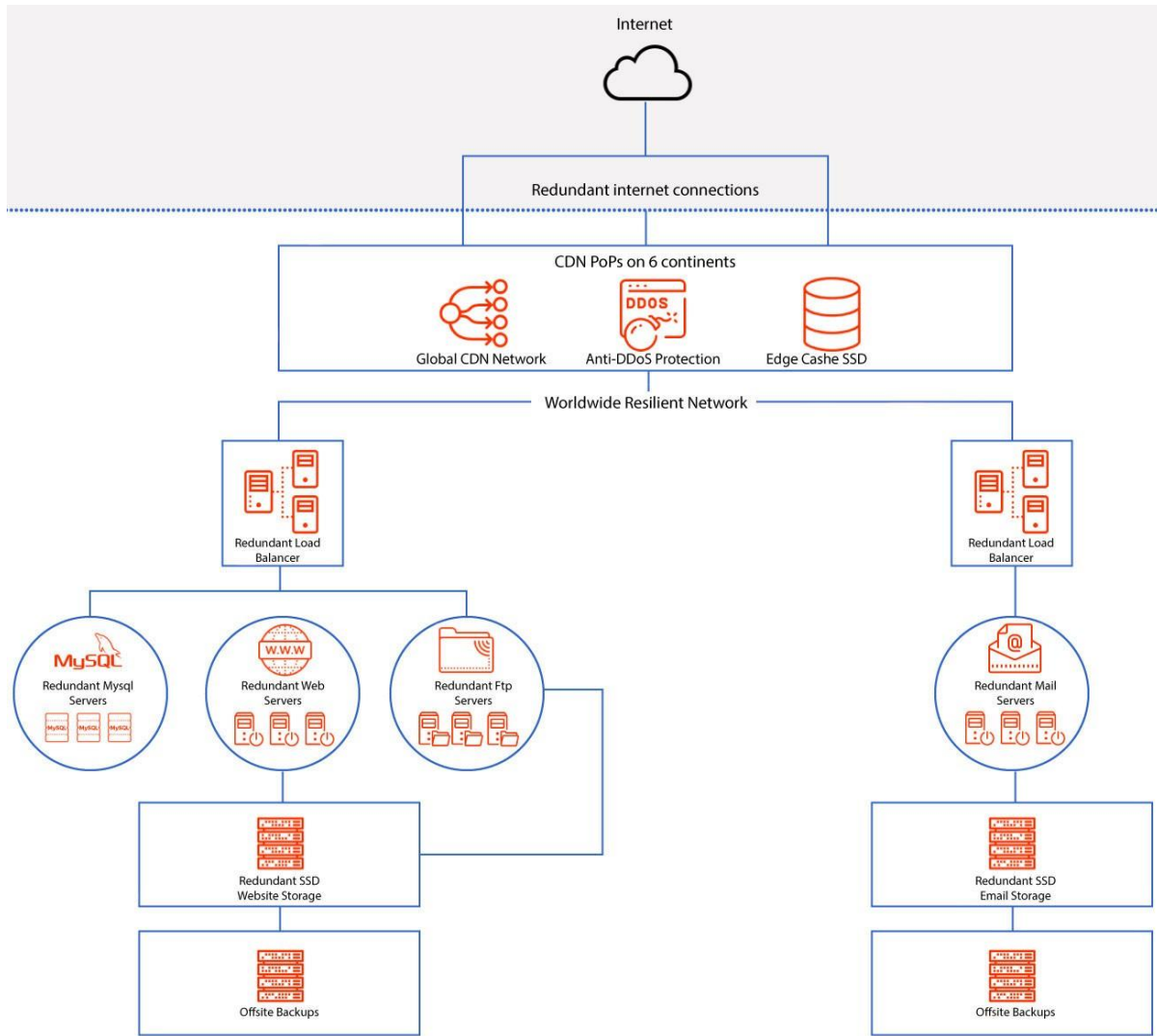


Figure 0.4: Architecture de la plateforme Vala multi-serveurs

L'architecture de la plateforme Vala est construite sur un modèle hautement résilient avec des connexions internet redondantes alimentant un réseau CDN global présent sur six continents, incluant protection anti-DDoS et cache SSD. Le système distribue le trafic via deux centres de données géographiquement séparés, chacun équipé d'équilibreurs de charge redondants et de clusters de serveurs spécialisés (MySQL, web, FTP, messagerie). L'infrastructure comprend des systèmes de stockage SSD redondants pour les sites web et emails, complétés par des sauvegardes hors site, garantissant une continuité de service optimale.

II. Les services de l'entreprise :

- **Hébergement Cloud pour Applications Web**

L'entreprise propose une gamme complète de solutions d'hébergement cloud adaptées à différents environnements de développement. Pour les applications PHP, l'hébergement garantit performance et flexibilité, permettant une gestion optimale des ressources. Les sites basés sur WordPress bénéficient d'une infrastructure sécurisée et rapide, assurant une expérience utilisateur fluide et fiable.

- **Hébergement pour Revendeurs et Plateformes E-commerce**

Creative Internet Solutions met à disposition des services d'hébergement permettant aux revendeurs de créer et gérer leurs propres offres. Par ailleurs, des solutions spécialisées sont proposées pour les boutiques en ligne, notamment via les plateformes Prestashop et Magento, assurant stabilité, sécurité et performances adaptées aux exigences du commerce électronique.

- **Hébergement et Messagerie Professionnelle**

L'entreprise offre également des services de messagerie professionnelle sécurisée, incluant une messagerie e-mail performante ainsi que des solutions collaboratives basées sur Microsoft, favorisant une communication efficace au sein des entreprises.

- **Création et Gestion de Sites Web E-commerce**

Spécialisée dans la création de sites e-commerce, l'entreprise développe des solutions complètes intégrant un design professionnel, une ergonomie intuitive, et un éditeur de pages WYSIWYG. Les fonctionnalités comprennent la gestion des produits, promotions, panier d'achat, confirmation des commandes en temps réel, et optimisation SEO, garantissant une présence en ligne performante et facile à administrer.

- **Solution CMS Sans Programmation : Site Moujoud**

Pour répondre aux besoins des particuliers et professionnels sans compétences techniques, Creative Internet Solutions propose la solution Site Moujoud. Ce CMS intuitif,

basé sur un système de glisser-déposer, permet de créer facilement des sites vitrines ou e-commerce, optimisés pour le référencement et compatibles avec tous les supports, facilitant ainsi le développement et la mise en ligne rapide.

- **Stratégie et Optimisation SEO**

Creative Internet Solutions déploie une stratégie SEO complète incluant la recherche de mots-clés, un audit technique et éditorial, ainsi que la correction des erreurs essentielles. L'optimisation du contenu vise à améliorer l'expérience utilisateur, tandis qu'un suivi régulier avec rapports mensuels garantit l'adaptation aux évolutions des moteurs de recherche. L'acquisition de backlinks renforce l'autorité du site, un facteur clé pour un référencement performant.

- **Désinfection et Sécurisation des Sites Piratés**

L'entreprise identifie d'abord la source de l'infection puis sauvegarde les données avant toute intervention. Elle supprime les logiciels malveillants, met à jour les systèmes et plugins, et renforce la sécurité avec des mots de passe forts et restrictions d'accès. Des tests réguliers garantissent la protection durable du site.

- **Certificat SSL**

Valla propose des certificats SSL offrant un cryptage symétrique de 256 bits pour sécuriser les échanges en ligne. L'authentification certifie la légitimité du site, renforçant la confiance des visiteurs. Ces certificats, accessibles à moindre coût, respectent les normes de sécurité les plus strictes. Le SSL favorise des relations durables avec les clients, assure la protection contre les cyberattaques, et améliore le référencement naturel du site sur les moteurs de recherche.

III. Le service d'affectation :

En tant que département d'une grande entreprise, Creative Internet Solutions fait face à un nombre important de candidatures pour ses différents postes. La gestion et la vérification manuelle des CV représentent un défi majeur en termes de temps et d'efficacité. Pour répondre à cette problématique, un projet innovant a été développé : un

service d'extraction et d'analyse automatique des CV. Ce système permet de classer rapidement et précisément les candidats les plus qualifiés en fonction des critères prédéfinis, optimisant ainsi le processus de recrutement et améliorant la qualité des sélections.

IV. Conclusion :

En conclusion, ce chapitre a présenté l'entreprise d'accueil, son organisation, son évolution et ses services. Le service d'affectation, avec son projet d'analyse automatique des CV, illustre une réponse aux besoins de l'entreprise. Ces informations sont importantes pour bien comprendre le contexte de mon projet et guider la suite de mon travail.

CHAPITRE 2 : SPÉCIFICATIONS

I. Introduction :

Dans ce chapitre, je présenterai mon projet et son contexte. J'exposerai les objectifs à atteindre, puis les problèmes et contraintes identifiés. Je préciserai également les besoins fonctionnels et non fonctionnels, afin de proposer des solutions adaptées aux exigences du cahier des charges.

II. Présentation du projet :

II.1 Contexte général du projet :

Le projet consiste en une application web innovante destinée à aider les recruteurs dans la sélection des meilleurs candidats pour un poste. Cette solution vise à optimiser et à simplifier le processus de recrutement en offrant un gain de temps et une meilleure précision dans le choix des profils adaptés. À partir des critères et exigences définis par le recruteur au début du processus, le système analyse automatiquement les CV de chaque candidat, en comparant leurs compétences et informations aux attentes spécifiées. Ensuite, l'application classe les candidats du plus qualifié au moins qualifié, tout en fournissant les informations essentielles extraites, telles que les coordonnées et tout autre détail supplémentaire demandé par le recruteur.

II. 2 Les objectifs de la plateforme :

L'application vise à rendre le recrutement plus efficace, précis et fluide grâce à une interface simple et intuitive. Elle a pour but de simplifier le processus de recrutement en automatisant une étape clé de la sélection des candidats, permettant ainsi au recruteur de gagner du temps et de se concentrer sur les étapes suivantes du recrutement. Elle offre un système intelligent qui analyse et compare les compétences des candidats aux critères définis, facilitant la gestion simultanée de plusieurs recrutements. Le recruteur peut ainsi suivre l'avancement des candidatures, obtenir un classement clair des profils du plus qualifié au moins qualifié, et accéder rapidement aux informations essentielles comme les contacts ou détails supplémentaires demandés.

III. Cahier de charge :

III.1 Problématique :

Dans le monde professionnel actuel, la gestion d'un grand nombre de candidatures représente un défi majeur pour les recruteurs. La nécessité d'examiner minutieusement chaque CV pour identifier les profils les plus adaptés est une tâche répétitive, chronophage et épuisante. Ce processus manuel engendre une perte de temps considérable et mobilise une énergie importante, ce qui peut réduire l'efficacité globale du recrutement. De plus, l'excès d'informations à analyser augmente le risque d'erreurs ou de sélection inappropriée, impactant la qualité des décisions prises. La question centrale est donc de savoir comment une application web intelligente peut alléger cette charge en automatisant l'analyse des candidatures, tout en garantissant une sélection précise et rapide, afin d'optimiser le temps des recruteurs et améliorer le processus de recrutement dans son ensemble.

III.2 Les acteurs :

Acteur principal de notre plateforme : Recruteur

- **Recruteur** : L'unique utilisateur et bénéficiaire direct de la plateforme. Le recruteur utilise la plateforme pour définir les besoins en recrutement, préciser les critères souhaités, puis déposer les fichiers PDF contenant les candidatures. La plateforme se charge ensuite de trier automatiquement les candidats en fonction des critères établis, permettant ainsi un gain de temps considérable et une meilleure précision dans la sélection des profils adaptés.

III.3 Les besoins fonctionnels :

- **Création d'une offre de recrutement** : Le recruteur peut définir les besoins en recrutement via un formulaire, en précisant les critères tels que le poste, les compétences requises et la date limite.
- **Téléversement de candidatures** : Le recruteur peut importer plusieurs fichiers PDF contenant les CV des candidats.

- **Tri Automatique des candidats** : La plateforme analyse les fichiers PDF importés et classe les candidats selon les critères définis par le recruteur.
- **Accès à des informations organisées sur chaque candidat** : Le recruteur peut consulter un profil détaillé de chaque candidat, incluant ses compétences, son expérience, sa formation et ses documents associés, dans un format clair et structuré.
- **Téléchargement des résultats** : Le recruteur peut télécharger la liste triée des candidats.
- **Triage des candidats** : Le recruteur peut trier les candidats selon n'importe quelle information.
- **Recherche des candidats** : Le recruteur peut effectuer des recherches parmi les candidats.

III.4 Les besoins non fonctionnels :

→ Sécurité :

- Aucune donnée personnelle ou fichier PDF téléchargé n'est conservé après l'analyse ; toutes les informations sensibles sont uniquement utilisées dans le cadre du traitement et ne sont ni sauvegardées, ni partagées.
- Cryptage temporaire des données lors du transfert pour éviter toute interception.

→ Performance :

- Temps de traitement et de tri rapide, même pour un grand nombre de candidatures.
- Optimisation de l'algorithme d'analyse pour garantir un classement fluide et précis.
- Gestion efficace des ressources pour éviter tout ralentissement pendant l'utilisation.

→ Interface Utilisateur :

- Conception claire et intuitive, adaptée à un recruteur non technique.
- Présentation hiérarchisée des informations extraites de chaque candidat pour un accès rapide.

- Interface responsive pour un usage sur ordinateur, tablette et smartphone.

→ **Facilité d'Utilisation :**

- Explications claires et guides intégrés pour faciliter la prise en main.
- Feedback visuel (barres de progression, messages de confirmation) lors des étapes de téléversement et de tri.
- Navigation simplifiée avec menus logiques et actions en un minimum de clics.

→ **Réactivité :**

- Temps de chargement minimal des pages et résultats.
- Mise à jour instantanée des listes de candidats après analyse.
- Support de technologies modernes pour un affichage fluide sans rechargement complet des pages

III.5 Règles de gestion :

Les règles de gestion sont l'ensemble des principes et directives qui encadrent le fonctionnement et le comportement d'un système. Elles assurent sa cohérence, sa sécurité et son bon déroulement. Dans ce rapport, je décris les règles de gestion essentielles qui servent de fondement et guident le fonctionnement de mon système.

- Les fichiers PDF téléchargés par le recruteur sont analysés uniquement pour extraire les informations pertinentes et ne sont pas sauvegardés.
- La plateforme se limite à analyser et à trier les candidatures selon des critères prédéfinis, et sert uniquement d'outil d'aide à la décision pour les recruteurs.
- Les critères de sélection doivent être définis avant le dépôt des candidatures pour garantir un tri cohérent et précis.

III.6 Nom et Logo de site :



Figure 2.1: Logo du système

J'ai choisi le nom *RecruitMate* pour ma plateforme afin de refléter son rôle dans le processus de recrutement. Ce nom se compose de deux parties : *Recruit*, qui signifie « embaucher », et *Mate*, qui signifie « ami » ou « assistant ». L'idée est de montrer que la plateforme agit comme un véritable assistant de confiance, facilitant et accompagnant chaque étape du recrutement.

J'ai décidé d'utiliser l'anglais pour toute la plateforme, car cette langue donne une image plus moderne et internationale, renforçant ainsi son attractivité. De plus, les technologies de traitement du langage naturel (NLP) que j'utilise fonctionnent de manière plus performante et stable avec l'anglais, offrant une expérience plus fluide et précise aux utilisateurs.

IV. Conclusion :

Dans cette section, j'ai résumé les grandes lignes du projet en présentant le cahier des charges, les défis, les objectifs, ainsi que le contexte général. Ce chapitre aborde également la problématique, les acteurs, les besoins fonctionnels et non fonctionnels, les règles de gestion, ainsi que le nom et le logo de la plateforme.

CHAPITRE 3 : ANALYSE ET CONCEPTION

I. Introduction :

Dans ce chapitre, je présente la modélisation et la conception de mon projet à travers l'utilisation de diagrammes UML. Cette démarche m'a permis de structurer un système clair, cohérent et parfaitement adapté aux besoins des utilisateurs. Je commence par l'élaboration du diagramme de cas d'utilisation, avant d'analyser le diagramme de séquence.

II. Diagramme de cas d'utilisation :

Le rôle des diagrammes de cas d'utilisation est de collecter, d'analyser, d'organiser les exigences et de lister les principales fonctionnalités d'un système. C'est donc la première étape UML pour la conception du système. Le diagramme de cas d'utilisation comporte trois éléments principaux.

- **Acteur** : Représente un utilisateur externe ou une entité qui interagit avec le système.
- **Cas d'utilisation** : Représente une action ou une fonctionnalité spécifique que le système offre à ses utilisateurs, décrivant une interaction entre un acteur et le système.
- **Relation** : Montre les liens entre les acteurs et les cas d'utilisation, ou entre différents cas d'utilisation.

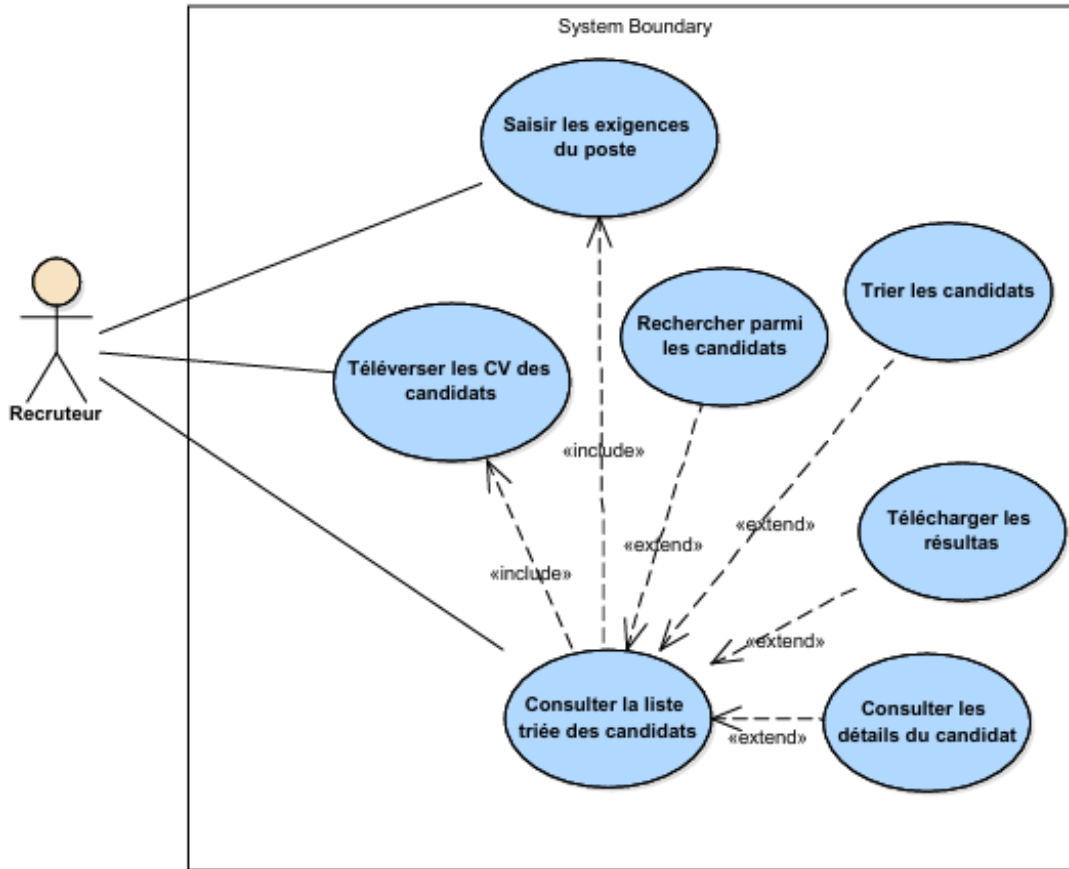


Figure 3.1:Diagrammes des cas d'utilisation du système

Le recruteur peut spécifier les critères de recrutement et télécharger les CV des candidats afin que le système puisse les analyser automatiquement. Grâce à cette analyse, le système trie les candidats en fonction des exigences définies, mais il peut également trier selon n'importe quelle information du candidat et effectuer des recherches parmi eux. Par la suite, le recruteur a la possibilité de consulter en détail le profil de chaque candidat sélectionné. Enfin, il peut enregistrer les résultats obtenus pour un suivi ou une utilisation future.

III. Diagramme de séquence :

Les diagrammes de séquence constituent la représentation graphique des interactions entre les acteurs et le système, organisées chronologiquement dans le cadre de l'UML (Unified Modeling Language).

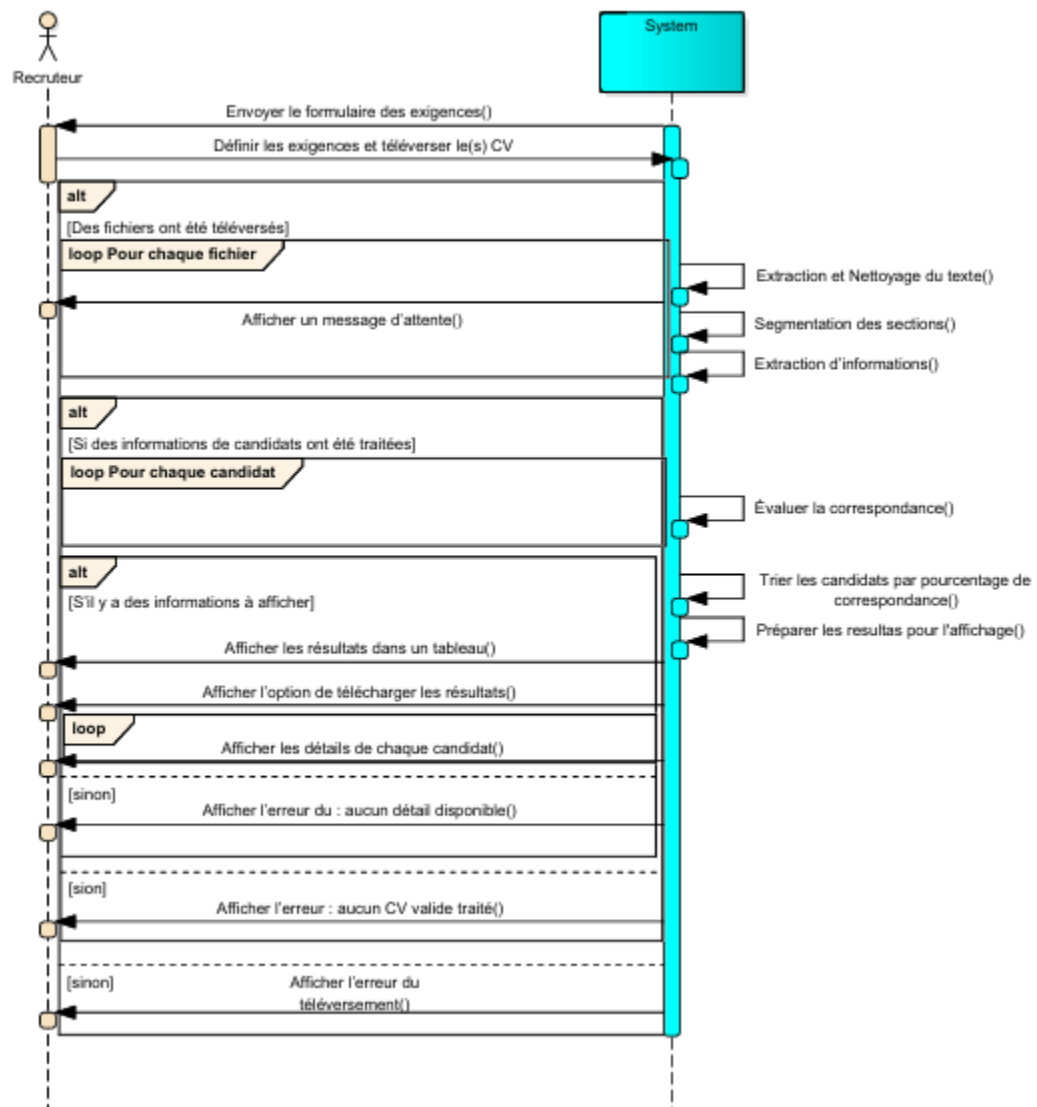


Figure 3.2: Diagramme de séquence du système

Le processus commence par le recruteur qui envoie le formulaire des exigences et téléverse les CV. Le système vérifie la présence de fichiers, puis pour chaque fichier affiche un message d'attente et effectue l'extraction et le nettoyage du texte, la segmentation des sections (pour faciliter l'extraction d'informations) et l'extraction des informations. Si des données de candidats sont disponibles, le système évalue la correspondance avec les exigences, trie les candidats par pourcentage de correspondance et prépare les résultats pour l'affichage. Les résultats sont présentés dans un tableau avec la possibilité de télécharger l'ensemble du tableau, et il est possible d'afficher les détails de chaque candidat de manière interactive. En cas d'erreur, le système indique simplement qu'aucun CV valide n'a été traité ou qu'aucun fichier n'a été téléversé.

IV. Diagramme d'activités :

Un diagramme d'activités est un type de schéma UML qui illustre le flux des actions, décisions et événements au sein d'un processus ou d'un système. Il permet de visualiser l'ordre d'exécution des tâches, les points de décision et les activités parallèles, facilitant ainsi la compréhension du fonctionnement global.

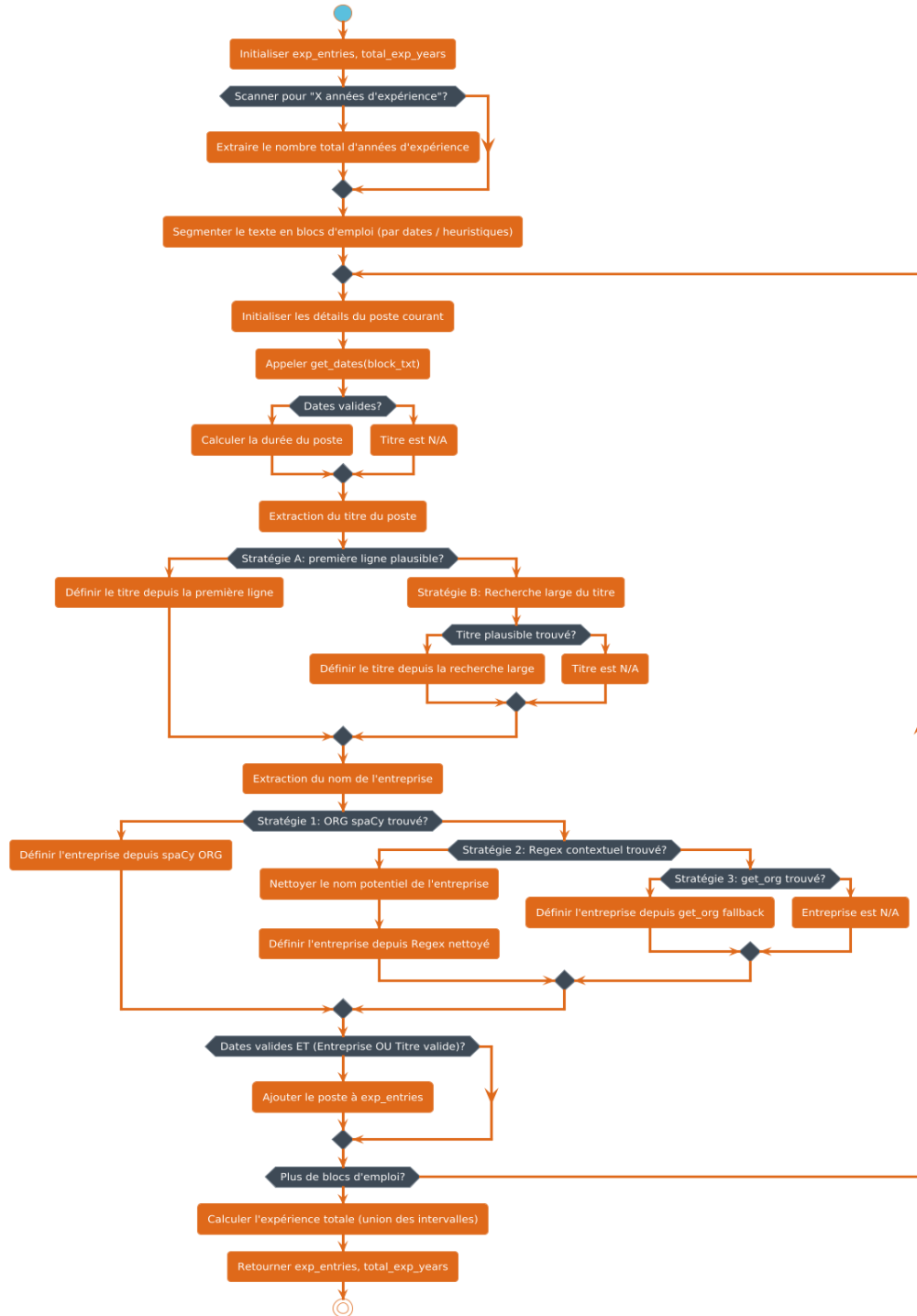


Figure 3.3: Diagramme d'activités du système

Ce diagramme décrit comment le système analyse l'expérience d'un CV. La fonction recherche d'abord l'expérience totale, puis segmente le texte en blocs de poste via les dates ou des heuristiques. Pour chaque poste, elle extrait dates et durée, identifie le titre (priorisant la première ligne), et trouve l'entreprise via une cascade de trois stratégies (détection spaCy directe, regex contextuelle avec nettoyage, ou fonction générique

get_org). Après validation de chaque poste, l'expérience totale est calculée par cumul des intervalles de temps, et le tout est retourné.

V. Conclusion :

En résumé, la modélisation et la conception à l'aide des diagrammes UML ont constitué une étape essentielle dans la réalisation de mon projet. Grâce à l'analyse des cas d'utilisation et des séquences, j'ai pu structurer un système organisé et fonctionnel, répondant précisément aux besoins des utilisateurs

CHAPITRE 4 : ENVIRONNEMENT DE TRAVAIL ET OUTILS

I. Introduction :

Dans ce chapitre, je présente les différents outils que j'ai utilisés pour créer mon projet. Chacun d'eux a joué un rôle essentiel dans les différentes étapes de développement, depuis la conception jusqu'à la réalisation finale.

II. Outils de travail :

II.1 Visual studio code :

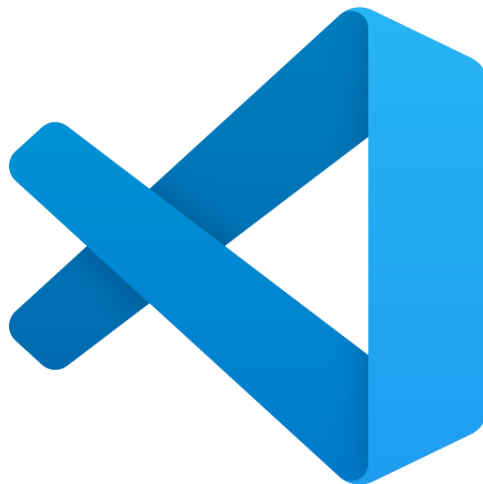


Figure 0.1: Logo de Visual studio code

Visual Studio Code est un éditeur de code source gratuit et multiplateforme développé par Microsoft. Je l'ai utilisé comme environnement principal car il m'a offert la flexibilité et la simplicité nécessaires : terminal et intégration Git intégrés, autocomplétion, outils de débogage, et un riche écosystème d'extensions (linting, formatage, support des frameworks) qui ont accéléré l'écriture, la navigation et la correction du code.

II.2 Git et GitHub :

Git est un système de gestion de versions qui enregistre et suit l'historique des modifications d'un projet. GitHub est une plateforme en ligne qui héberge des dépôts Git et facilite la collaboration.

Dans mon projet, je les ai utilisés pour suivre l'évolution du code, organiser les étapes de développement et documenter mon travail, assurant ainsi un historique clair et sécurisé.



Figure 4.3: Logo de Github



Figure 4.2: Logo de Git

II.3 Enterprise Architect :

Enterprise Architect est un outil de modélisation UML complet qui permet de créer et gérer différents types de diagrammes pour la conception et l'analyse de systèmes logiciels.

Je l'ai utilisé pour réaliser les diagrammes nécessaires, tels que les diagrammes de cas d'utilisation et de séquence, indispensables à la phase de conception de mon projet.



Figure 4.4: Logo de Enterprise Architect

III. Outils de développement :

III.1 Python :

Python est un langage de programmation interprété, polyvalent et facile à apprendre, largement utilisé dans le développement web, l'analyse de données, l'intelligence artificielle et l'automatisation. Sa syntaxe claire et sa vaste bibliothèque standard en font un outil puissant pour divers types de projets.

Je l'ai utilisé comme langage principal pour développer l'ensemble de mon projet, car il m'a permis d'intégrer facilement les différentes bibliothèques nécessaires, de traiter les données efficacement et de créer une application fonctionnelle de manière rapide et structurée.



Figure 0.5: Logo de Python

III.2 Streamlit:

Streamlit est un framework Python qui permet de créer facilement des interfaces web interactives pour visualiser et interagir avec des données ou des applications.

Je l'ai utilisé pour développer l'interface utilisateur interactive de mon projet, permettant d'afficher les résultats, de naviguer entre les fonctionnalités et de rendre l'application plus intuitive pour l'utilisateur final.



Figure 4.6: Logo de Streamlit

III.3 pdfplumber:

pdfplumber est une bibliothèque Python conçue pour extraire le texte et d'autres informations contenues dans des fichiers PDF.

Je l'ai utilisé pour lire et extraire automatiquement le contenu textuel des CV au format PDF, afin de pouvoir analyser leurs informations dans mon système.

III.4 Regex :

Regex (expressions régulières) est une technique de programmation permettant de rechercher, identifier et manipuler des motifs précis dans du texte.

Je l'ai utilisée pour extraire automatiquement certaines informations spécifiques des CV, comme les adresses e-mail, numéros de téléphone, ou compétences clés, afin de pouvoir analyser et structurer ces données efficacement dans mon système.

III.5 spaCy:

spaCy est une bibliothèque avancée de traitement du langage naturel (NLP) offrant des fonctionnalités telles que la reconnaissance d'entités nommées (NER) et la correspondance efficace de phrases ou de mots-clés.

Je l'ai utilisé pour détecter automatiquement des informations importantes dans les CV, comme les noms, les organisations, les compétences ou les certifications, ce qui a permis d'automatiser l'analyse et la catégorisation des candidats.



Figure 0.7: Logo de SpaCy

III.6 pandas:

Pandas est une bibliothèque Python puissante pour la manipulation et l'analyse de données, notamment grâce à ses structures de données comme les DataFrames.

Je l'ai utilisée pour organiser les informations extraites des CV sous forme de tableaux, faciliter leur tri et permettre l'exportation des résultats au format CSV.



Figure 0.8: Logo de Pandas

IV. Conclusion :

Dans ce chapitre, j'ai présenté tous les outils que j'ai utilisés pour gérer mon site web. J'ai aussi expliqué comment ces outils m'ont aidé à créer et faire évoluer mon application, en montrant leur importance dans le développement de mon projet.

CHAPITRE 5 : RÉALISATION DE PROJET

I. Introduction :

Cette partie du projet a pour objectif de présenter le produit final, qui est un système automatique d'analyse de CV. Je partagerai quelques captures d'écran afin de donner un aperçu du projet et de faciliter sa compréhension. Cette phase de réalisation est essentielle, car elle me permet de concrétiser ma vision et de créer un système fonctionnel et performant.

II. Présentation de l'interface :

II.1 En-tête de l'interface :

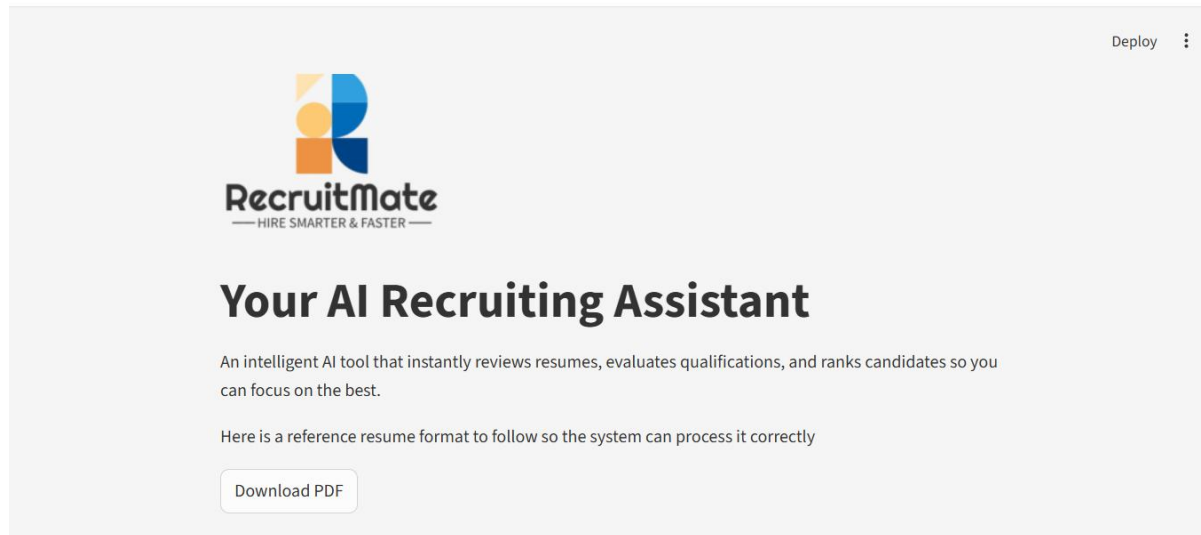
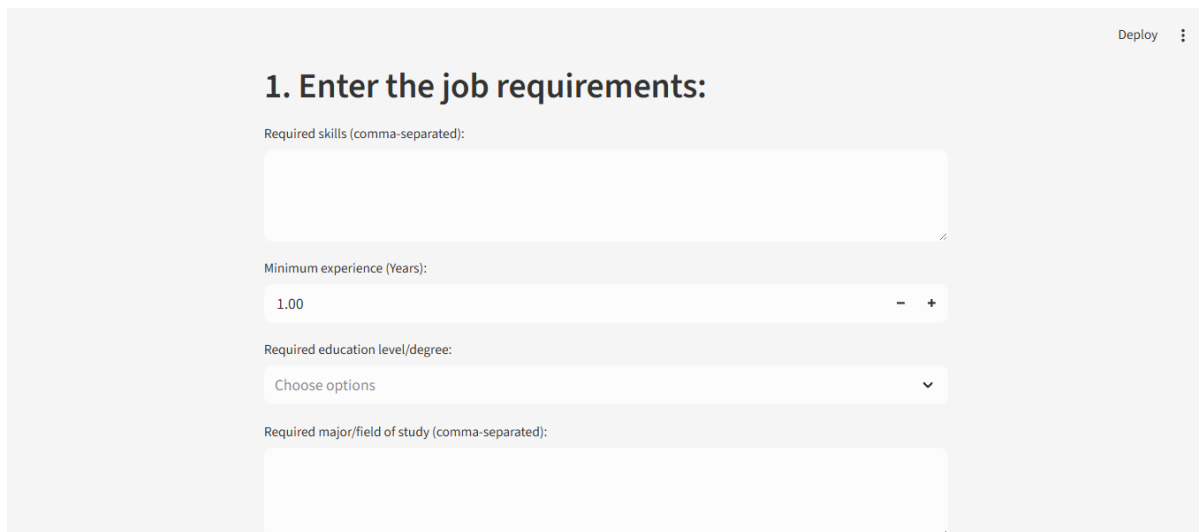


Figure 0.1: En-tête de l'interface

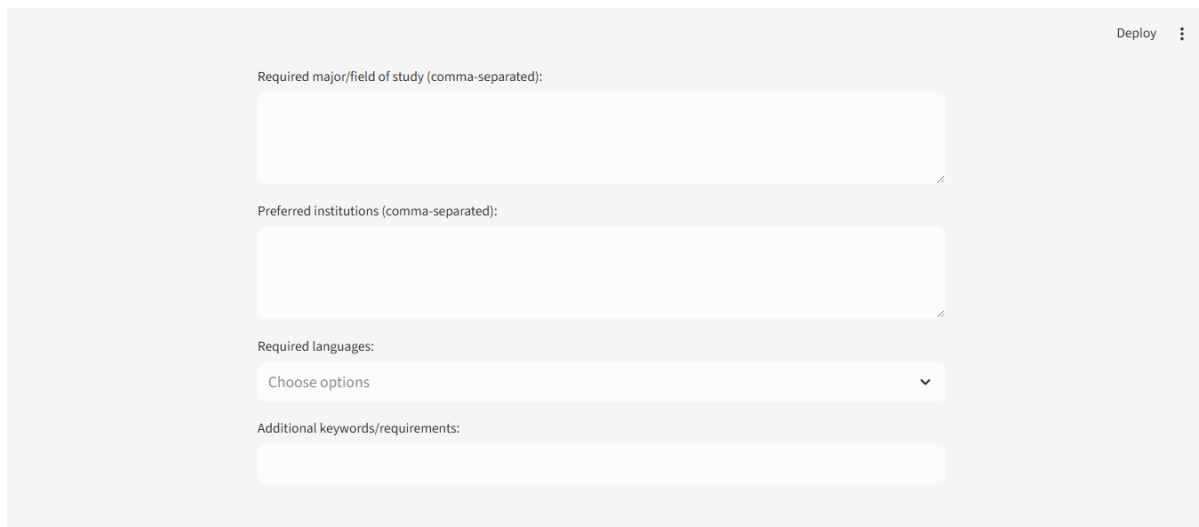
En-tête de l'interface affiche un message de bienvenue à l'utilisateur et propose un bouton pour télécharger un exemple de CV au format de référence, afin que le système puisse le traiter correctement.

II.2 Le formulaire des exigences :



The screenshot shows the first section of a web form titled "1. Enter the job requirements:". In the top right corner, there is a "Deploy" button and a three-dot menu icon. The form contains four input fields: a text area for "Required skills (comma-separated)", a numeric input for "Minimum experience (Years)" with a value of "1.00" and minus/plus buttons, a dropdown menu for "Required education level/degree" with the text "Choose options", and another text area for "Required major/field of study (comma-separated)".

Figure 0.2: Première section du formulaire des exigences



The screenshot shows the second section of the web form. It contains four input fields: a text area for "Required major/field of study (comma-separated)", a text area for "Preferred institutions (comma-separated)", a dropdown menu for "Required languages" with the text "Choose options", and a text area for "Additional keywords/requirements". The "Deploy" button and three-dot menu icon are still visible in the top right corner.

Figure 0.3: Deuxième section du formulaire des exigences

Le formulaire des exigences : contient les champs de saisie permettant de préciser les critères recherchés par le recruteur, tels que le niveau de diplôme, la spécialité, les compétences, les années d'expérience, etc.

II.3 Le champ du dépôt des CV :

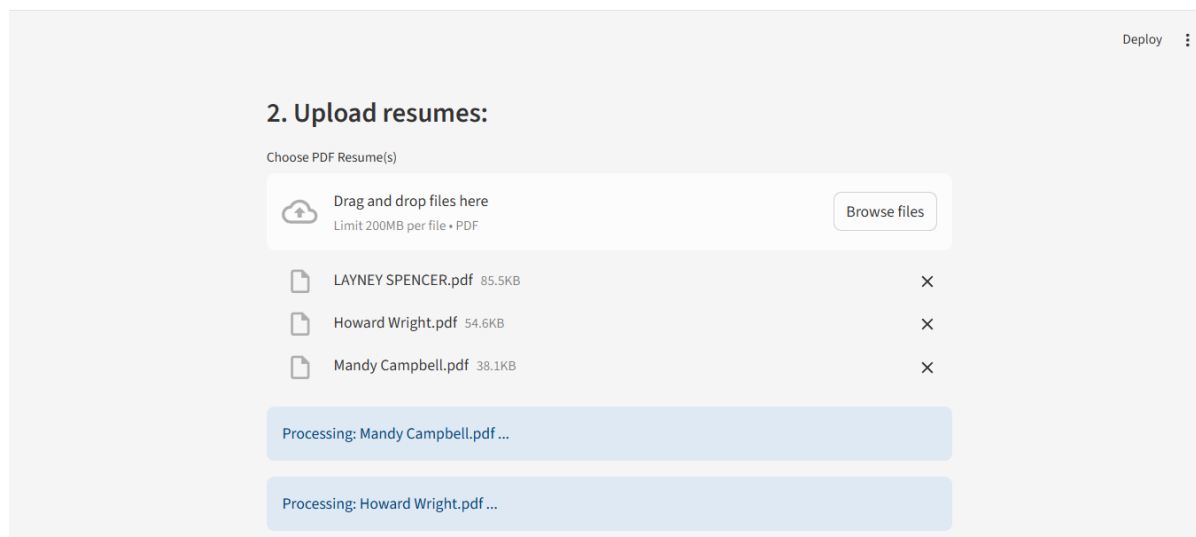


Figure 0.4: Le champ du dépôt des CV

Le champ de dépôt des CV permet à l'utilisateur de téléverser les CV en vue de leur analyse. Pour chacun d'eux, le système effectue le traitement et affiche un message de succès en cas de bonne exécution. L'utilisateur peut également supprimer des CV déjà déposés ou en ajouter de nouveaux, même après le traitement.

II.4 Tableau des résultats :

	Filename	Name	Match %	Score	Total Exp (Years)	Skills Met	Major Met	Preferred institutions Met	languages Met
0	Mandy Campbell.pdf	Mandy Campbell	100.00%	38	2.500000	3/3	Yes	Yes	0/0
1	Howard Wright.pdf	Howard Wright	84.21%	32	3.500000	1/3	Yes	Yes	0/0
2	LAYNEY SPENCER.pdf	Layne Spencer	76.32%	29	2.300000	0/3	Yes	Yes	0/0

Figure 0.5: Tableau des résultats

Le tableau des résultats présente les candidats triés selon leur pourcentage de correspondance avec les exigences.

[illegible]

Figure 0.6: Tableau des résultats après tri

L'utilisateur peut également choisir un autre attribut de tri ainsi qu'effectuer une recherche dans le tableau.

II.5 Exportation des résultats :

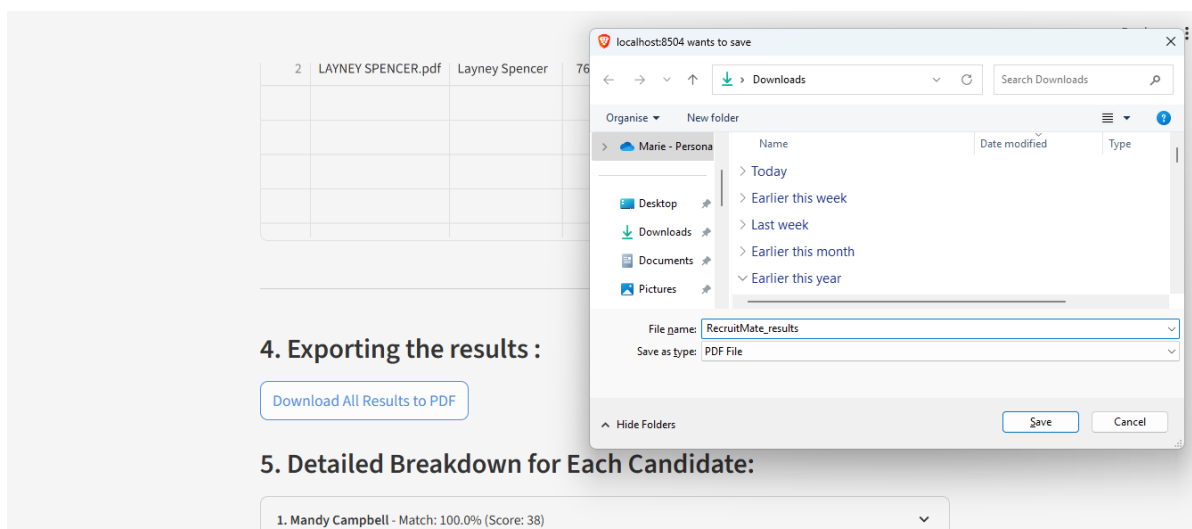


Figure 0.7: Exportation des résultats

Lorsqu’il souhaite télécharger les résultats, l’utilisateur dispose d’un bouton qui génère un PDF contenant l’ensemble des candidats et leurs informations.

Candidate 1: Mandy Campbell

Filename: Mandy Campbell.pdf
Name: Mandy Campbell
Match Percent: 100.0
Match Score: 38
Total experience (Years): 2.5
Email: email@email.com
Phone: 386-868-3442
Matched Skills: hiit, marketing, nutrition
Education Met: Yes
Major Met: Yes
Preferred institutions: Yes
Languages Met (All): Yes
Matched Required languages:
Missing Required languages:
Skills Extracted: active listening, cycling, fitness assessments, group cycling, group fitness, hiit, marketing, nutrition, program statistics, staffing, strength and conditioning, word
Education Extracted: Bachelor from Concordia University Concordia (Jul 2021 — Present)
Diploma from National CPR Foundation Miami (Jan 2020 — Jul 2021)
High School Diploma from High School Diploma (Jan 2018 — Dec 2018)
Experience Details: Certified Bootcamp Instructor at Dominist Fitness (Jan 2019 — Jul 2021)
Bootcamp Instructor at Curves Gym Ponta (Jul 2021 — Jul 2021)
Languages Extracted:
Certificates Extracted:
LinkedIn URL:
GitHub URL:
Other URLs:

Candidate 2: Howard Wright

Filename: Howard Wright.pdf
Name: Howard Wright
Match Percent: 84.21

Figure 0.8: Exemple de fichier des résultats

II.6 Répartition détaillée pour chaque candidat :

Detailed Breakdown for Each Candidate:

1. Mandy Campbell - Match: 100.0% (Score: 29)

Contact Info

Email: email@email.com

Phone: 386-868-3442

Match Details

Education Level Met: Yes

Major/Field Met: Yes

Preferred Institution Met: Yes

Extracted Information

Skills: active listening, cycling, fitness assessments, group cycling, group fitness, hiit, marketing,

Figure 0.9 : Répartition détaillée pour chaque candidat

Lorsqu'un utilisateur souhaite obtenir plus de détails sur un candidat, il peut déployer un volet qui affiche toutes les informations nécessaires le concernant.

<p>Skills: active listening, cycling, fitness assessments, group cycling, group fitness, hiit, marketing, nutrition, program statistics, staffing, strength and conditioning, word</p> <p>Education:</p> <ul style="list-style-type: none">• Bachelor in 's Degree in Exercise Science from Concordia University Concordia (Jul 2021 — Present)• Diploma in Advanced First Aid from National CPR Foundation Miami (Jan 2020 — Jul 2021)• High School Diploma in Barlow High School from High School Diploma (Jan 2018 — Dec 2018) <p>Experience:</p> <p>Total Years: 2.5</p> <ul style="list-style-type: none">• Certified Bootcamp Instructor at Dominist Fitness (Jan 2019 — Jul 2021)• Bootcamp Instructor at Curves Gym Ponta (Jul 2021 — Jul 2021)	
---	--

Figure 0.10 : Répartition détaillée pour chaque candidat (suite)

III. Difficultés rencontrées lors de la réalisation :

Au cours du développement, plusieurs difficultés importantes ont été rencontrées et résolues de manière itérative :

- **Problèmes majeurs de découpage des blocs :** La diversité des formats de CV a rendu impossible l'utilisation d'une seule stratégie de parsing. J'ai donc créé des fonctions capables de détecter le format du CV et d'appliquer la logique de traitement adaptée, ce qui a permis de résoudre les problèmes de mélange de données et de perte d'entrées.
- **Extraction des informations clés :** J'ai rencontré des difficultés à extraire des informations significatives plutôt que des mots aléatoires dans les CV. J'ai amélioré l'extraction des diplômes, spécialisations, entreprises et titres de poste en corrigeant les regex, en nettoyant les données et en ajustant les priorités pour sélectionner les informations réellement pertinentes.

IV. Conclusion :

En conclusion, cette partie m'a permis de présenter le système automatique d'analyse de CV et de montrer son fonctionnement à travers quelques captures d'écran. La réalisation concrète de ce projet a été essentielle pour mettre en pratique les choix techniques et conceptuels définis précédemment.

CONCLUSION GENERALE ET PERSPECTIVES

Le développement de RecruitMate m'a permis de concevoir une solution intelligente capable de traiter automatiquement un grand nombre de CV, d'extraire les informations clés et de les comparer aux critères attendus par les recruteurs. Cette application répond efficacement au défi majeur que représente le tri manuel des candidatures, en réduisant considérablement le temps consacré à cette phase initiale de sélection.

Au cours du projet, j'ai rencontré plusieurs difficultés, notamment l'extraction d'informations significatives à partir de documents semi-structurés et la gestion de formats variés de CV. Grâce à un processus itératif de conception, de tests et d'ajustements, j'ai pu améliorer la précision de l'extraction, la robustesse du système et l'ergonomie de l'interface utilisateur.

Ce rapport est structuré en cinq chapitres principaux : j'ai présenté l'entreprise VALA Blue, ses missions et son organisation ; exploré les fonctionnalités développées et le cahier des charges ; détaillé le processus de conception de la plateforme, en exposant les étapes et les solutions que j'ai apportées ; dressé la liste des outils et technologies utilisés ; et enfin décrit le produit final en montrant comment il répond aux besoins identifiés tout en facilitant le travail de présélection des recruteurs.

Pour l'avenir, mon objectif est d'optimiser davantage le processus d'extraction des informations et de rendre le système adaptable à d'autres langues, comme le français. De plus, plusieurs améliorations sont envisageables : intégrer des techniques de correspondance sémantique pour mieux détecter les compétences et expériences, enrichir l'interface avec des visualisations interactives, permettre la comparaison côte-à-côte des candidats, étendre la compatibilité à différents formats de CV comme DOCX, et créer des tableaux de bord et des analyses. Ces évolutions visent à rendre RecruitMate encore plus intelligent, flexible et efficace pour assister les recruteurs dans leur prise de décision.

BIBLIOGRAPHIE

- [1] <https://docs.streamlit.io/>
- [2] <https://spacy.io/api/doc>
- [3] <https://docs.python.org/3/library/re.html>
- [4] <https://pypi.org/project/pdfplumber/>
- [5] <https://zenodo.org/records/16918803>
- [6] https://www.w3schools.com/python/python_regex.asp

ANNEXE A

Lien vers le dépôt GitHub du projet : [<https://github.com/mariamouh/RecruitMate.git>]