|  |  |  |
| --- | --- | --- |
|  |  |  |



AkeenJobs

« Plateforme de recherche d’emplois »

par

Rakotoasimbola Mihary Lalaina

Mémoire présenté  
MASTER de SCIENCES, TECHNOLOGIES, SANTE, mention INFORMATIQUE  
Spécialité MOBIQUITE, BASES DE DONNEES ET INTEGRATION DE SYSTÈMES (MBDS)

Septembre 2023

Jury :

M. Harilalanjatovo Vahatriniaina, encadreur professionnel

© Rakotoasimbola Lalaina, 2023

Table des matières

[Liste des tableaux iii](#_Toc146043462)

[Liste des figures iv](#_Toc146043463)

[Glossaire v](#_Toc146043464)

[Avant-propos vi](#_Toc146043465)

[L’IT University vi](#_Toc146043466)

[Remerciements vii](#_Toc146043467)

[Introduction viii](#_Toc146043468)

[1 Présentation du stage 1](#_Toc146043469)

[1.1 Présentation de l’entreprise 1](#_Toc146043470)

[WYLog Madagascar 1](#_Toc146043471)

[1.2 Présentation du sujet et objectifs du projet (problème traité) 2](#_Toc146043472)

[2 État de l’art sur le sujet traité 3](#_Toc146043473)

[2.1 Critères de comparaison 3](#_Toc146043474)

[2.1.1 La lecture de CV 3](#_Toc146043475)

[2.1.2 Le modèle de catégorisation 3](#_Toc146043476)

[2.1.3 Le système bilingue 3](#_Toc146043477)

[2.2 Étude de chaque solution au vu des critères 4](#_Toc146043478)

[2.2.1 Lecture de CV 4](#_Toc146043479)

[2.2.2 Modèle de catégorisation 6](#_Toc146043480)

[2.2.3 Système bilingue 7](#_Toc146043481)

[2.3 Tableau comparatif des solutions au vu des critères 8](#_Toc146043482)

[2.3.1 Lecture de CV 8](#_Toc146043483)

[2.3.2 Modèle de catégorisation 9](#_Toc146043484)

[3 Étude de l’existant et solution envisagée 10](#_Toc146043485)

[3.1 Étude de l’existant 10](#_Toc146043486)

[3.2 Critique de l’existant 10](#_Toc146043487)

[3.3 Solutions envisagées 10](#_Toc146043488)

[3.4 Objectifs principaux et livrables 11](#_Toc146043489)

[4 Démarche projet 12](#_Toc146043490)

[4.1 Principes de la démarche projet 12](#_Toc146043491)

[4.1.1 Activités d’ingénierie logicielle 12](#_Toc146043492)

[4.1.2 Méthode de gestion de projet utilisée 12](#_Toc146043493)

[4.1.3 Rôles et responsabilités 13](#_Toc146043494)

[4.1.4 Outils 13](#_Toc146043495)

[4.1.5 Gestion de la configuration 15](#_Toc146043496)

[4.2 Contraintes et risques sur le projet 15](#_Toc146043497)

[4.3 Démarche projet mise en œuvre 16](#_Toc146043498)

[4.4 Budget du projet 17](#_Toc146043499)

[5 Exigences réalisées dans le projet (vision externe/utilisateur) 18](#_Toc146043500)

[5.1 Exigences fonctionnelles – Cas d’utilisation 18](#_Toc146043501)

[5.1.1 Cas d’utilisation 1 : Recherche des emplois et offres 18](#_Toc146043502)

[5.1.2 Cas d’utilisation 2 : Gestion du modèle de données 19](#_Toc146043503)

[5.1.3 Cas d’utilisation 3 : Tableau de bord 20](#_Toc146043504)

[5.2 Exigences non fonctionnelles transverses 21](#_Toc146043505)

[1.1.1. Utilisabilité 21](#_Toc146043506)

[1.1.2. Performances 21](#_Toc146043507)

[1.1.3. Robustesse 22](#_Toc146043508)

[1.1.4. Sécurité 22](#_Toc146043509)

[1.1.5. Supportabilité (maintenabilité, évolutivité) 22](#_Toc146043510)

[5.3 Interfaces détaillées 22](#_Toc146043511)

[5.3.1 IHM 22](#_Toc146043512)

[5.3.2 Interfaces avec d’autres systèmes 27](#_Toc146043513)

[6 Architecture(s) système 28](#_Toc146043514)

[6.1 Architecture technique 28](#_Toc146043515)

[6.2 Architecture logicielle 28](#_Toc146043516)

[7 Conception du système logiciel réalisée dans le projet (vision interne/développeur) 29](#_Toc146043517)

[7.1 Plate-forme technique 29](#_Toc146043518)

[7.1.1 Architecture matérielle 29](#_Toc146043519)

[7.1.2 Framework logiciel 29](#_Toc146043520)

[7.2 Conception du logiciel développé 29](#_Toc146043521)

[7.2.1 Conception du code source 29](#_Toc146043522)

[7.2.2 Le code source – vue statique 31](#_Toc146043523)

[7.2.3 Modélisation de données 32](#_Toc146043524)

[7.2.4 Réalisation des cas d’utilisation 32](#_Toc146043525)

[7.2.5 Les composants et leur déploiement 33](#_Toc146043526)

[8 Tests du système logiciel 34](#_Toc146043527)

[9 Conclusion générale 35](#_Toc146043528)

[10 Références et Bibliographie 36](#_Toc146043529)

[11 Annexes 37](#_Toc146043530)

# Liste des tableaux

[Tableau 1 : Comparaison entre les méthodes de lecture de CV 8](#_Toc146043708)

[Tableau 2 : Comparaison des modèles de catégorisation 9](#_Toc146043709)

[Tableau 3 : Critique de l'existant 10](#_Toc146043710)

[Tableau 4 : Outils utilisés 13](#_Toc146043711)

[Tableau 5 : Contraintes et risques 16](#_Toc146043712)

[Tableau 6 : Budget du projet 17](#_Toc146043713)

# Liste des figures

[Figure 1 : Diagramme de GANTT 17](#_Toc146043714)

[Figure 2 : Cas d'utilisation 1 - Recherche d'emplois et d'offres 19](#_Toc146043715)

[Figure 3 : Cas d'utilisation 2 - Gestion du modèle de catégorisation 20](#_Toc146043716)

[Figure 4 : Cas d'utilisation 3 - Tableau de bord 21](#_Toc146043717)

[Figure 5 : Import de CV 22](#_Toc146043718)

[Figure 6 : Liste des compétences et professions 23](#_Toc146043719)

[Figure 7 : Choix des critères 23](#_Toc146043720)

[Figure 8 : Liste des emplois recherchés 24](#_Toc146043721)

[Figure 9 : Offres d'emplois 25](#_Toc146043722)

[Figure 10 : Gestion des compétences 26](#_Toc146043723)

[Figure 11 : Gestion des professions 26](#_Toc146043724)

[Figure 12 : Tableau de bord 27](#_Toc146043725)

[Figure 13: Architecture technique 28](#_Toc146043726)

[Figure 14 : Architecture logicielle 28](#_Toc146043727)

[Figure 15 : Architecture matérielle 29](#_Toc146043728)

[Figure 16 : Vue statique ExpressJs 31](#_Toc146043729)

[Figure 17 : Vue statique React 31](#_Toc146043730)

[Figure 18 : Modèle de données 32](#_Toc146043731)

[Figure 19 : Réalisation CU 1 - Recherche d'emplois et d'offres 33](#_Toc146043732)

[Figure 20 : Interface authentification 37](#_Toc146043733)

# Glossaire

* Upload : importer un fichier dans le système.
* Framework : En programmation informatique, c’est un ensemble de composants à la base d’une application ou d’une plateforme logicielle. Composants constituant son socle. Son but est d’uniformiser et de simplifier le travail des artisans logiciels.
* GANTT : Diagramme permettant d’afficher efficacement l’avancement d’un projet.
* API : Une API est un ensemble de protocoles qui permet à un service de communiquer avec d’autres services et qui facilite l’intégration de nouveaux composants dans une architecture existante.
* JS : Langage de programmation permettant de rendre un site interactif.
* MVC : C’est un design pattern qui permet d’organiser l’arborescence de notre code source afin de bien séparer sa logique.
* REST : Ensemble de contraintes architecturales
* API RESTful : C’est une API qui respecte les contraintes du style d’architecture REST.
* SQL : C’est un langage informatique utilisé pour interagir avec les SGBDR.
* IDE : Un environnement de développement intégré (IDE) est une application logicielle qui aide les programmeurs à développer efficacement le code logiciel.

# Avant-propos

Le présent mémoire présente les résultats du travail effectué lors de mon stage de fin d’études de Licence en Informatique de l’IT University; stage effectué au sein de l’entreprise Wylog Madagascar durant trois mois et demi environ, de Juin à Septembre 2023.

Afin de poser clairement le contexte de ce mémoire, je vais présenter succincte d’une part de l’IT University et d’autre part de mon institution/entreprise d’accueil.

## L’IT University

Fondée en 2011, l’IT University (ou ITU) est une université privée, spécialisée en informatique, formant les jeunes bacheliers scientifiques :

* en trois ans, pour l‘obtention d’une Licence, option Développement d’applications, Réseaux et Bases de Données ou Web et Design
* en cinq ans, pour l’obtention d’un Master MBDS en coopération avec l’Université de Nice Sophia Antipolis - France

Étant une formation professionnalisante, l’ITU a tissé des liens forts avec ses partenaires industriels, dont l’opérateur convergent TELMA, le GOTICOM et la plupart des entreprises et institutions du secteur des TIC[[1]](#footnote-1). Ces partenaires participent effectivement à la formation par la fourniture de connexion Internet à haut débit, l’envoi de conférenciers ou encore par l’accueil des étudiants en stage.  
D’autre part, le corps enseignant de l’ITU est constitué intégralement de spécialistes de très haut niveau et obligatoirement actifs professionnellement dans leurs domaines respectifs.  
Enfin, l’ITU fait partie du programme Microsoft Imagine en plus d’être un Oracle Gold Partner.

## Remerciements

En premier lieu, je tiens à remercier l’IT University et l’Université Côte d’Azur, les enseignants MBDS, pour les enseignements reçus et les connaissances acquises pendant ces années d’apprentissage.

Puis, je tiens à montrer ma reconnaissance envers mon encadreur pédagogique représentant l’établissement et ses enseignant, ainsi que les membres du jury, pour leur dévouement dans l’examen et la correction de ce présent mémoire.

En particulier, je voudrais remercier **M.** **Harilalanjatovo Vahatriniaina**, mon encadreur professionnel, pour m’avoir consacré temps, et pour m’avoir fournie l’aide nécessaire pour l’accomplissement de ce stage. Mes sincères remerciements, aussi à l’entreprise Wylog Madagascar pour m’avoir accueilli durant ces trois mois de stage. De même pour les employés de l‘entreprise et particulièrement ceux qui ont partagé le même projet que moi pendant cette période.

Enfin, ma gratitude envers ma famille de m’avoir soutenu, émotionnellement ainsi qu’économiquement pour l’aboutissement de mes études jusqu’au point où j’en suis, et aussi mes amis pour leur soutien incomparable durant ces années d’études.

# Introduction

La population active obtient des avantages, financiers, sociaux ou autres, grâce à leur emploi. Or, les études de la banque mondiale affirment que le taux de chômage est de 5.8% dans le monde, dont environ 8% en France et 2% à Madagascar. Ainsi, afin de réduire ce nombre, l’entreprise propose une solution parmi tant d’autres, pour faciliter la recherche d’emploi, aussi pour les chômeurs que pour les employés : AkeenJobs. Ce projet a donc pour objectif de simplifier la recherche de travail, sans se déplacer ou faire beaucoup de procédures.

Étant passionné de développement et ayant pour projet de devenir un développeur professionnel et bien au-delà dans l’avenir, j’ai choisi Wylog Madagascar pour effectuer mon stage de fin d’études durant lequel mes objectifs ont été d’appliquer les connaissances acquises durant mes études universitaires à l’Université de Nice Sophia Antipolis et d’acquérir plus d’expériences dans le monde professionnel.

Durant ce stage, mes principales attributions consistent à :

* Établir une application web, pour les besoins nécessaires du projet
* Construire un modèle de données identifiant certaines catégories de mots
* Participer à la conception, au développement et au déploiement du projet
* Rédiger les documents nécessaires des outils que je développe

Ces objectifs et missions orientent à poser la question suivante : comment ces outils vont-ils contribuer à améliorer la recherche d’emploi susdite?

Afin de répondre à cette question, ce document détaillera dans les prochains chapitres, en premier lieu l’état de l’art du projet, puis l’analyse de l’existant qui est le centre Pole Emploi qui existe en France, suivi des démarches et approches dans la réalisation du projet. Ensuite, détailler les exigences réalisées dans le projet et enfin finir par la conception du projet.

# Présentation du stage

## Présentation de l’entreprise

## WYLog Madagascar

WYLog est une société française dont le siège est basé à Paris spécialisée dans les prestations informatiques depuis 2003. Forte de plus de 18 années d’expérience, la société WYLog a acquis une grande capacité à allier expertise technique et compréhension métier.

La société WYLog dispose aujourd’hui de six bureaux, répartis dans cinq pays (France, Madagascar, Italie, Philippines et États-Unis) composée d’une équipe de 120 professionnels passionnés. Les équipes de WYLog sont organisées autour de plusieurs pôles technologiques (Développements Web & Mobile ; Développements AR/ VR ; Blockchain ; Département IoT/ Live Streaming). WYLog travaille aussi bien en langue française qu’en langue anglaise.

Des centaines de projets informatiques ont été menés à bien avec des entreprises aux profils variés (startup, comptes corporate, organisme public) basées aussi bien en France qu’à l’étranger.

WYLog travaille dans plusieurs secteurs d’activités, développant des applications pour ceux-ci :

* Industrie
* Media & Communication
* Fintech

## Présentation du sujet et objectifs du projet (problème traité)

Le travail est un facteur important dans la vie d’un adulte actif. Cela lui permet essentiellement de gagner de quoi vivre et de quoi planifier son futur. Or, la recherche d’emploi requiert beaucoup d’étapes et de difficultés. Alors, la société a proposé une solution pour faciliter cette dernière, une application Web. En envoyant un simple CV, le système va aller chercher les compétences et emplois du possesseur du CV. À partir de ces données, le système va rechercher les emplois et offres disponibles via la base de données de Pole-Emploi. Les administrateurs auront aussi besoin d’un espace pour suivre les actions des utilisateurs.

Vu les besoins précisés ci-dessus, on a étudié la faisabilité qui nous a amené aux objectifs réalisables du projet :

* Obtenir des offres d’emplois
  + Choix de compétences et emplois pour la recherche
  + Définir des critères de recherches, tels la région, la nature du travail voulues par l’utilisateur
  + Faire la recherche d’emplois liés aux données précisées ci-dessous
  + Faire ressortir les offres disponibles d’un emploi choisi, avec les critères voulus
* Récupérer les compétences et emplois à partir d’un CV
  + Import et lecture de CV
  + Reconnaissance des éléments du CV en construisant un modèle de catégorisation des mots.
* Espace dédiée pour les administrateurs
  + Ajout de nouveaux données d’apprentissage du modèle, avec mis-à-jour du modèle
  + Statistiques des critères, pour savoir les critères les plus demandés par les utilisateurs

La société, étant une société bilingue, a demandé que l’application soit compréhensible par autant d’utilisateurs possibles. Afin de répondre à cette demande, la solution proposée est :

* Application Web bilingue : Français et anglais.

# État de l’art sur le sujet traité

## Critères de comparaison

Afin de répondre aux objectifs citées précédemment, on a étudié multiples solutions, dont on va les séparer selon les critères de comparaison.

### La lecture de CV

Nous allons porter les études sur les méthodes pour faire la lecture du CV. Cette étape consiste à donner le texte contenu dans le fichier, afin de continuer avec l’application du modèle.

### Le modèle de catégorisation

La plus importante étape de l’application se porte sur la catégorisation des mots. Elle permet aux systèmes de ressortir les compétences et les professions de l’utilisateur présent dans son CV. Un chercheur suédois, Nick Bostrom a dit : « L'IA est comme une feuille blanche, elle a le potentiel d'être tout ce que nous pouvons imaginer ».  On va donc utiliser l’intelligence artificielle afin de faire cette catégorisation.  
À partir de données solides, une étude des algorithmes est nécessaire pour obtenir lequel on pourra utiliser le mieux dans le système. De ce fait, on va utiliser des algorithmes existants pour les étudier. Nous allons les comparer en fonction de leur taux de succès et leur taux d’erreur.   
Pour la comparaison, on va diviser les données en deux types de données :

* les données d’apprentissage : 75% des données initiales
* les données de test : 25% des données initiales

### Le système bilingue

Pour une meilleure compréhension du contenu du système, deux langues ont été demandées, l’anglais et le français. L’étude qu’on va effectuer sur ce domaine se portera sur la flexibilité de la méthode.

## Étude de chaque solution au vu des critères

### Lecture de CV

Les fichiers PDF sont l'un des formats de documents les plus couramment utilisés dans le monde, largement utilisés pour stocker et partager des informations sous une forme lisible par l'homme tout en préservant la mise en page et la structure du document. La nécessité de manipuler et d'extraire des données à partir de fichiers PDF est devenue une exigence fondamentale dans divers domaines, de la gestion documentaire à l'analyse de données en passant par l'automatisation des tâches liées aux documents.

#### PyPDF2 :

PyPDF2 est une bibliothèque PDF purement Python, gratuite et open source, capable de diviser, fusionner, recadrer et transformer les pages de fichiers PDF. Cette bibliothèque se présente comme une solution robuste pour accomplir une gamme de tâches liées aux fichiers PDF. PyPDF2 offre des fonctionnalités avancées de manipulation de PDF, notamment l'extraction de texte, la fusion et la division de fichiers, l'ajout d'annotations et bien plus encore. Dans cette étude, on va se concentrer sur l’extraction de texte.

Une étude a affirmé que : PyPDF2 est un outil très rapide pour traiter un fichier PDF, mais il ne fonctionne qu’avec du texte brut. En effet, la bibliothèque assure l’extraction de texte, mais seulement le titre et les paragraphes. Or, un CV a un format variant et pas de paragraphes. Ce serait ambigu pour la bibliothèque de lire un fichier d’un CV. Pour confirmer cette affirmation, voici un exemple de format d’un CV :



Un CV peut aussi être une image de document manuscrite, dont l’utilisateur vient de scanner et transformer en pdf. Alors les caractères de ce fichier seraient impossibles pour la bibliothèque en question de procéder.

Par conséquent, PyPDF2 est une solution viable pour la lecture d’un CV. Elle est rapide, mais a des lacunes, qui sont importants dans la réalisation de notre projet.

* OCR :

OCR, Optical Character Recognition, en français, reconnaissance optique des caractères, est une technologie puissante qui a révolutionné la manière dont nous traitons et interagissons avec des documents imprimés ou manuscrits. Elle permet de convertir du texte imprimé ou écrit à la main en données électroniques, rendant ainsi ces données accessibles, modifiables et exploitables par des ordinateurs.

Le processus d'OCR repose sur des algorithmes sophistiqués qui analysent des images de texte et identifient les caractères individuels, les mots et les structures de page. Les étapes principales du processus OCR comprennent :

1. **Acquisition d'image** : La première étape consiste à capturer une image du texte à l'aide d'un scanner, d'une caméra numérique ou d'un autre dispositif d'acquisition. Cette image peut être une page imprimée, une facture manuscrite, une étiquette, etc.
2. **Prétraitement de l'image** : L'image capturée subit souvent un prétraitement pour corriger des problèmes tels que la distorsion, l'éclairage inégal et le bruit, ce qui permet d'obtenir une image de meilleure qualité.
3. **Segmentation** : L'image est divisée en zones de texte, de manière à identifier les caractères individuels, les mots et les lignes.
4. **Reconnaissance des caractères** : Cette étape est au cœur de l'OCR. Les caractères identifiés sont comparés à une base de données de formes de caractères, et une correspondance est établie pour déterminer les caractères.
5. **Post-traitement** : Les résultats de la reconnaissance sont améliorés par des techniques de post-traitement qui corrigent les erreurs et reconstruisent la mise en page du texte.
6. **Production de texte exploitable** : Le résultat final est un texte exploitable, souvent dans un format éditable tel que du texte brut ou un document au format PDF.

### Modèle de catégorisation

* Réseaux de neurones (Deep Learning) :

Les réseaux de neurones profonds, également appelés réseaux de neurones profonds ou réseaux de neurones profonds, sont une classe de modèles d'apprentissage automatique basés sur des réseaux de neurones artificiels comportant plusieurs couches cachées (profondes). Ces réseaux sont conçus pour apprendre et représenter de manière hiérarchique des caractéristiques ou des informations à partir de données brutes ou d'entrées, ce qui les rend particulièrement puissants pour la résolution de problèmes complexes. Selon une étude de Yann Lecun, un informaticien français, « Les réseaux de neurones profonds sont des boîtes à outils extrêmement polyvalentes pour le traitement des données, avec un potentiel pratiquement illimité ». Alors les réseaux de neurones profonds peuvent être utilisés pour la classification des mots.

Avec Python, on a une bibliothèque capable d’implémenter cette algorithme, nommée TensorFlow. On va alors créer un modèle de catégorisation avec les données que l’on a.

* Apprentissage automatique non supervisé :

L'apprentissage automatique non supervisé est une branche de l'apprentissage automatique où l'objectif principal est de découvrir des structures intrinsèques ou des modèles cachés dans les données sans la nécessité d'étiquettes ou de supervisation externe. Contrairement à l'apprentissage supervisé, où un modèle est formé à partir de données étiquetées, l'apprentissage non supervisé explore les données de manière autonome pour en extraire des informations utiles.

Cette méthode se passe des données d’entraînement, et ne possède donc pas de taux de succès et d’erreur. L'un des principaux inconvénients de l'apprentissage non supervisé est qu'il n'utilise pas d'étiquettes ou de supervision externe pour guider le processus d'apprentissage. Cela signifie que le modèle doit découvrir des structures ou des modèles dans les données par lui-même, ce qui peut être difficile, en particulier dans des domaines complexes. Donc, on a décidé de laisser cette méthode de côté pour le projet.

* NLP :

NLP (Natural Language Processing) ou TLN en français, Traitement du Langage Naturel, est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur l'interaction entre les ordinateurs et le langage humain. L'objectif principal du NLP est de permettre aux machines de comprendre, d'interpréter et de générer un langage naturel de manière à effectuer diverses tâches liées au langage.

SpaCy est une bibliothèque de traitement automatique du langage naturel (NLP). C'est une bibliothèque Python open-source conçue spécifiquement pour les tâches de NLP, et elle fournit des modèles pré-entraînés et des outils pour diverses tâches de NLP telles que la tokenisation, l'étiquetage de parties du discours, la reconnaissance d'entités nommées, l'analyse de dépendances, et plus encore.

SpaCy est largement utilisé par les développeurs, les chercheurs et les data scientists pour le traitement et l'analyse de données textuelles. Il offre un traitement efficace et précis du texte, ce qui en fait un choix populaire pour diverses applications de NLP, notamment la classification de texte, l'extraction d'informations et l'analyse de texte.

Les modèles pré-entraînés de SpaCy vont nous permettre de catégoriser les mots, ne fournissant les types de mots, comme les noms propres, les dates, les noms d’endroits etc… et laisse les autres types de mots comme « miscellaneous » ou divers. Alors, à partir de ces modèles pré-entraînés, on va ajouter des types de mots « compétences » et « professions », pour que le nouveau puisse identifier les mots divers en tant que ces nouveaux types de mots.

### Système bilingue

* React-i18next : bibliothèque d’internationalisation i18 :

i18next est une bibliothèque JavaScript populaire pour la gestion de la traduction et de l'internationalisation dans les applications web, y compris les applications React. Elle offre des fonctionnalités puissantes pour faciliter la prise en charge de plusieurs langues et de multiples clés de traduction.

i18next simplifie la gestion de l'internationalisation dans les applications web en fournissant des outils pour gérer les traductions, les langues et les éléments complexes tels que les variables et les dates. Il permet aux développeurs de créer des applications multilingues plus facilement et de fournir une expérience utilisateur localisée.

* Utilisation de fichiers de traduction :

La gestion de la traduction peut se faire avec des fichiers de traductions. On peut charger ces fichiers dans les composants que l’on veut appliquer la traduction. Pour des applications simples, l'utilisation de fichiers de traduction peut suffire. Cependant, on veut produire une application scalable dans le temps, alors cette méthode ne semble pas appropriée aux besoins du projet.

Par conséquent, pour la gestion de la traduction, il est plus convenable d’utiliser une bibliothèque spécialisée dans ce domaine. D’où le choix de la bibliothèque d’internationalisation, i18next.

## Tableau comparatif des solutions au vu des critères

### Lecture de CV

|  |  |  |
| --- | --- | --- |
| **Solutions** | **PyPDF2** | **OCR** |
| Avantages | - extraction de texte rapide | - flexible selon le format du CV  - obtention de texte exploitable |
| Inconvénients | - Meilleur seulement pour texte brute  - Pas flexible selon le format du CV | - Précision limitée.  - Traitement lent |

Tableau 1 : Comparaison entre les méthodes de lecture de CV

À partir de ce tableau de comparaison, on a décidé d’utiliser l’OCR pour faire la lecture du CV. Certes le traitement est lent et la précision limitée, mais sa flexibilité le fait passer devant les autres.

Python fournit déjà un outil pour travailler avec OCR : Tesseract.

### Modèle de catégorisation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Taux de succès | Taux d’échecs | |
| Réseaux de neurones profonds | 65% | 35% | |
| NLP | 62% | Erreur 1 | Erreur 2 |
| 30% | 8% |

Tableau 2 : Comparaison des modèles de catégorisation

Ce tableau montre les taux de réussite et les taux d’échecs des méthodes que l’on a détaillées précédemment.

Pour le modèle de NLP, on constate que les échecs de prédictions de tests se divisent en deux types :

* Les erreurs de classifications (Erreur 1) : une compétence classée profession ou vice-versa
* Les erreurs de non-catégorisations (Erreur 2): une compétence ou profession pas catégorisée.

En comparaison, les réseaux de neurones profonds vont classer, soit en compétence, soit en profession tous les mots que l’on va appliquer au modèle.

On a donc choisi la méthode NLP, SpaCy de Python, pour une meilleure flexibilité et une amélioration du modèle actuelle.

# Étude de l’existant et solution envisagée

## Étude de l’existant

Il existe beaucoup de plateformes qui permettent aux personnes de faire des recherches de travails. On a quelques exemples que l’on va étudier:

* LinkedIn
* CarreerBuilder
* Pole-Emploi

## Critique de l’existant

|  |  |  |
| --- | --- | --- |
| Plateforme | Points positifs | Points négatifs |
| LinkedIn | Réseau social où l’utilisateur peut partager son profil, voir des publications des entreprises et discuter avec | Lecture de CV indisponible pour la version gratuite pour la proposition des offres |
| CarreerBuilder | Téléchargement de CV  Offres disponibles par rapport aux CV et/ou aux compétences  Recherche d’emplois et d’offres rapide avec des mots donnés | Obligation de créer un compte pour télécharger un CV  Lecture de CV indisponible pour la version gratuite pour la proposition des offres |
| Pole-Emploi | Recherche d’emplois et d’offres rapide avec des mots donnés | Lecture de CV indisponible |

Tableau 3 : Critique de l'existant

## Solutions envisagées

AkeenJobs est donc une plateforme qui va en prendre quelques points positifs de ces plateformes existantes, ainsi que résoudre aux problèmes de ces dernières.

De ce fait, on a proposé les points suivants pour le projet :

* Téléchargement de CV libre sans créer un compte
* Recherche d’emplois et d’offres selon les compétences/professions données, mais aussi selon celles présentes dans le CV
* Redirection directe vers le lien des offres.
* Espace Backoffice

## Objectifs principaux et livrables

En vue de ces solutions envisagées, les objectifs principaux de ce projet sont de minimiser l’effort de l’utilisateur pour rechercher un travail et de faire monter un espace backoffice pour les administrateurs afin qu’ils puissent faire un suivi de la plateforme. De ce fait, il y a un (1) livrable :

* Une application web, sur laquelle l’utilisateur peut faire sa recherche et l’administrateur peut se connecter pour accéder au backoffice.

# Démarche projet

## Principes de la démarche projet

### Activités d’ingénierie logicielle

Selon Wikipédia, l’ingénierie logicielle est une science de génie industriel qui étudie les méthodes de travail et les bonnes pratiques des ingénieurs développant des logiciels. Il s’agit de procédures systématiques qui permettent d’arriver à ce que le logiciel corresponde aux attentes du client, soit fiable, performant et respecte les délais, les coûts et la qualité de construction. Ci-après les activités d’ingénierie logicielle applicables à ce projet :

* **Spécification des besoins :** consiste à récolter des informations détaillées sur les besoins et demandes sur le système à mettre en place. Durant ce projet, cette phase a été effectué avec le chef de projet Guillaume Bellanger.
* **Analyse fonctionnelle** : après la définition des demandes, cette partie amène à une étude de ce qu’on peut faire par rapport aux demandes.
* **Conception :** cette phase consiste à déterminer les grandes lignes à adopter pour le développement en soi du système, telles que la schématisation les architectures, les tables, les fonctions à utiliser.
* **Développement** : c’est la partie où on code l’outil par rapport à la conception.
* **Test et validation :** Une suite de vérifications doit être ensuite réalisée afin de déceler un maximum de bugs, de pannes ou de résultats incorrects. Quant aux validations, le chef de projet vérifie que la plateforme mise en place répond bien aux attentes et besoins définis.

### Méthode de gestion de projet utilisée

La méthode en cycle V, classique mais incontournable, a été choisie afin de mener à bien le développement de ce projet. Le cahier de charges a été défini et détaillé avant le développement, pour que l’on puisse nous concentrer sur un objectif sans vouloir changer ce qui a été prévu d’être implémenté. Deux séances hebdomadaires de tests ont été effectuées pour vérifier les fonctionnalités et évaluer les potentielles contraintes.

### Rôles et responsabilités

Les acteurs concernés par ce projet sont les suivants :

* Le PDG de Wylog : son rôle est de présenter le projet aux prospects
* Le chef de projet, Guillaume BELLANGER : acteur dans la spécification des besoins, analyse de l’existant et les tests
* L’équipe de développement, y inclus moi, Mihary Lalaina RAKOTOASIMBOLA : chargé de l’analyse, de la conception, le développement et le déploiement du projet.

### Outils

Pour la mise en œuvre de la démarche du projet, nous avons utilisé les outils ci-après :

|  |  |
| --- | --- |
| Domaines | Technologies |
| Planning | Gantt |
| Langages | Python, Javascript, Html, CSS, SQL |
| Framework | ExpressJs, React, Django |
| Base de données | PostgreSql |
| Service tiers | Pole-Emploi |
| Gestion de versions | Gitlab |
| IDE | Visual Studio Code |

Tableau 4 : Outils utilisés

1. Planning: Gantt

Pour le planning du projet, on a utilisé Gantt, dans le but d’avoir une meilleure visibilité des tâches et les périodes d’accomplissement de ces tâches.

1. Les langages et les frameworks
   1. Python : Django

La partie Machine Learning a été développé avec Python du fait que c’est un langage très facile à appréhender et qu’elle propose beaucoup de librairies pour la Machine Learning.

Django est un Framework Web de Python léger qui est très facile à prendre en main et donc une maintenance facile.

* 1. ExpressJs

ExpressJs est un framework Javascript, flexible et minimal, mais qui a la possibilité de produire des applications robustes. D’où, l’équipe de développement a choisi ExpressJs pour développer la partie backend de l’application.

* 1. React

Pour développer l’interface utilisateur, on a choisi React. Les raisons pour lesquelles cette bibliothèque de Javascript a été choisie sont majoritairement sa légèreté, et aussi sa simplicité dans l’installation.

1. La base de données : PostgreSql

PostgreSql est un système de gestion de base de données (SGBD) relationnelle et objet. L’application se connecte à une base PostgreSql, de un, ce système est gratuit, mais aussi pour sa réputation en performance, robustesse et sa fiabilité.

1. Pole-Emploi

Pole-Emploi est une plateforme, qui se tourne dans le monde du travail. Il fournit un API, utilisable par notre application, tels les emplois existants ou les offres publiés.

1. IDE : Visual Studio Code (VS Code)

VS Code est un IDE très flexible. Il nous a servi à développer avec ExpressJs, React et en Python.

Visual Studio Code prend en charge le JavaScript en tant que langage de programmation principal et offre des fonctionnalités d'achèvement automatique, de mise en évidence de la syntaxe et de débogage, ce qui facilite grandement le développement en JavaScript. Il dispose aussi d’un écosystème d’extensions très riche. Avec son intégration naturelle à Nodejs, il est le premier choix pour développer l’application avec ExpressJs et React.

VS Code a été choisi pour le développement Python car il offre une puissance énorme grâce à ses plugins qui peuvent le transformer pratiquement en IDE grâce à ses complétions de codes, refactorisations et toutes les autres fonctionnalités d’un IDE.

En addition, VS Code est connu pour sa légèreté et sa réactivité. Il intègre aussi Git de manière transparente, ce qui facilite la gestion des contrôles de versions directement depuis l’IDE.

1. Gestion des versions : Gitlab.wylog

Pour la gestion des versions et le partage des codes entre l’équipe, on a utilisé Gitlab de Wylog.

1. Autres

**Dbdiagram.io :** utilisé pour le Modèle Conceptuel de Données

**Draw.io :** outil de prototypage utilisé pour les maquettes

### Gestion de la configuration

Les livrables de ce projet seront configurés comme suit :

* L’application web sera déployé sur le serveur de la société
* Les codes sources seront versionnées sur le Gitlab de l’entreprise

## Contraintes et risques sur le projet

Les contraintes et les risques seront décrit dans le tableau ci-dessous :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Libellé du risque** | **Priorité** | **Facteur contribuant** | **Solutions proposées** | **Status** |
| 1 | Ne pas pouvoir construire le modèle de classification | Haute | - Manque d’expériences en machine learning  - Difficulté à obtenir des données pertinentes pour l’entrainement | - Suivre une formation en Data Science & Machine Learning  - Utiliser des données d’une source assez fiable (esco) | Réalisé |
| 2 | Système en une seule langue | Modérée | Trop de travaux pour avoir chaque mot traduit en deux langues | Appel d’une personne additionnelle pour accélérer la traduction | Réalisé |

Tableau 5 : Contraintes et risques

## Démarche projet mise en œuvre

Pour la mise en œuvre du projet, nous avons défini 4 grandes étapes.

* Phase de conception :

Cette partie consiste à étudier les demandes aux niveaux des utilisateurs finaux. Après avoir synthétiser ces besoins, l’équipe passe après à la conception des scénarios, les spécifications techniques, les tables à utiliser, les dessins d’écran.

* Phase de développement :

Cette étape a pour but de développer l’application. L’équipe procède donc à la réalisation des analyses demandées, la création des interfaces graphiques, l’implémentation des spécifications mentionnées dans la conception.

* Phase de test et de validation :

Chaque semaine, un test est effectué et une correction est ajoutée si besoin.

Après la phase de développement, un test final des cas d’utilisation possibles est effectué avant leur validation.

Après les corrections nécessaires, les métiers procèdent aux tests finaux sur bases de cahiers de tests. Une fois les corrections validées on peut procéder au déploiement.

Le diagramme de Gantt ci-dessous va résumer la planification du projet.

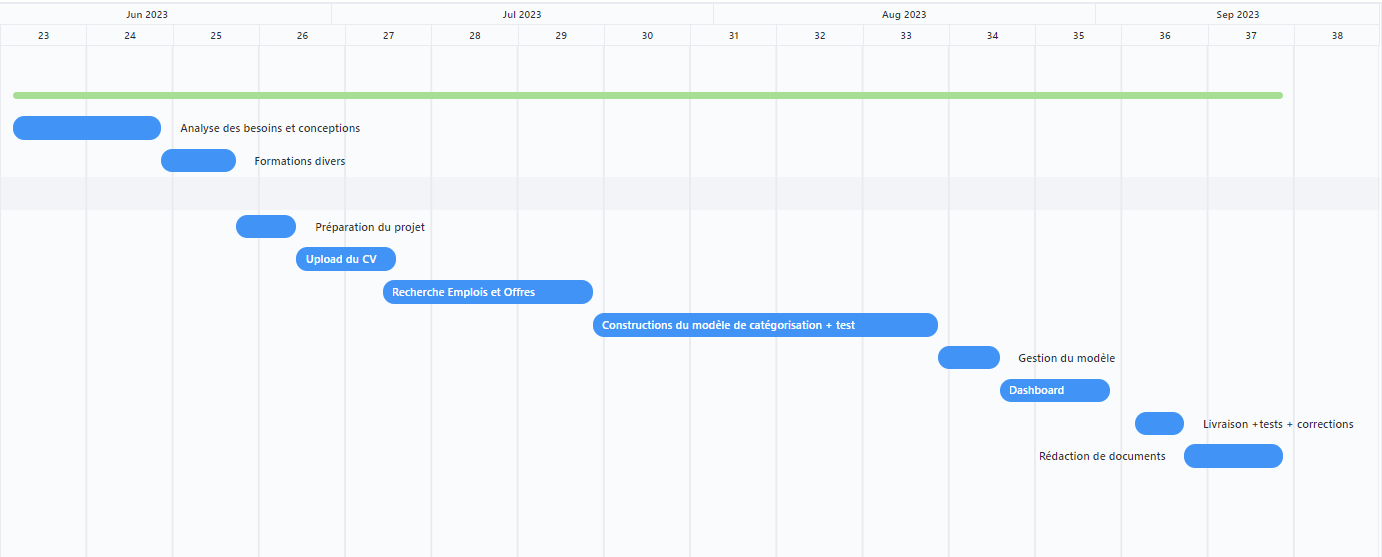


Figure 1 : Diagramme de GANTT

## Budget du projet

Ci-après se présentent brièvement et approximativement les exigences budgétaires du projet :   
Ici, on prend 1 € = 5 000 MGA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Désignation** | **Prix unitaire (MGA)** | **Quantité** | **Montant total (MGA)** |
| Ordinateur | 1 200 000,00 | 1 | 1 200 000,00 |
| Salaire développeur | 1 000 000,00 | 2 | 2 000 000,00 |
| Connexion internet | 250 000,00 | 4 | 1 000 000,00 |
| Déplacement | 120 000,00 | 4 | 480 000,00 |
| Total |  |  | 4 680 000,00 |
| Total en euro |  |  | 936,00 € |

Tableau 6 : Budget du projet

# Exigences réalisées dans le projet (vision externe/utilisateur)

## Exigences fonctionnelles – Cas d’utilisation

### Cas d’utilisation 1 : Recherche des emplois et offres

**Post-condition :**

* Les statistiques des critères et des CV importés augmentent.

**Scénario :**

* Sur le site, l’utilisateur importe un CV et saisit un nom en tant que propriétaire du CV.
* Après l’import, le système va récupérer les compétences et professions présentes dans le CV.
* L’utilisateur choisit les compétences et professions avec lesquelles il voudrait continuer. Il peut aussi en ajouter à la liste.
* Il choisit les critères de recherche, et puis clique sur le bouton « Suivant »
* Il reçoit une liste d’emplois disponibles, et clique sur un emploi.
* Il est redirigé vers toutes les offres publiées.

**Cas alternatif :**

* L’utilisateur n’importe aucun CV ou il ne saisit pas de propriétaire, un message d’erreur va se présenter.

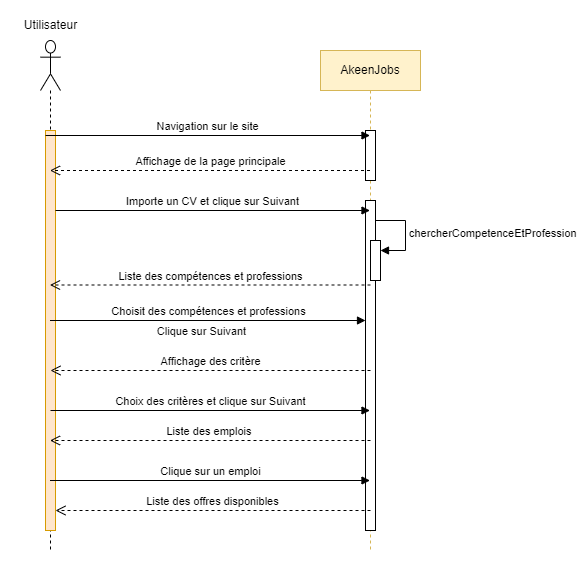


Figure 2 : Cas d'utilisation 1 - Recherche d'emplois et d'offres

### Cas d’utilisation 2 : Gestion du modèle de données

**Précondition :** être connecté en tant qu’administrateur

**Post-conditions :**

* Le modèle de catégorisation est ré entrainé

**Scénario 1: Gestion des compétences**

* L’administrateur arrive sur une liste des compétences ajoutées.
* Il peut modifier/supprimer ces compétences.
* Il sélectionne les compétences qu’il veut entrainer, puis clique sur entraîner.
* Il reçoit un message disant que l’entraînement est un succès.

**Cas alternatif :**

* L’administrateur ne sélectionne aucune compétence à entraîner, il reçoit un message d’erreur.

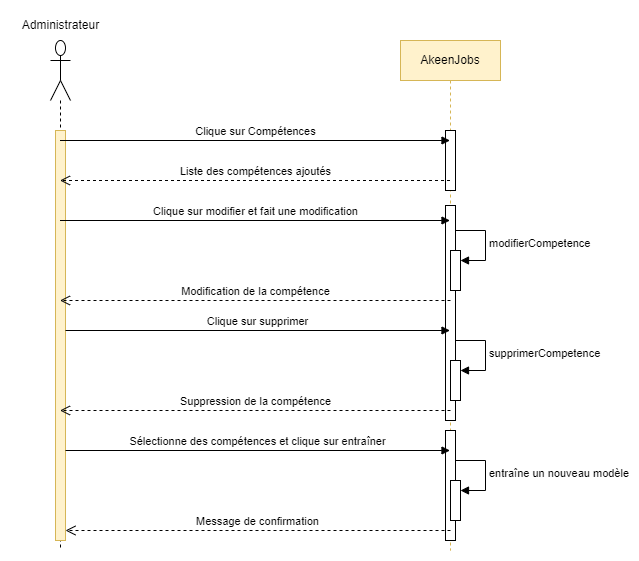


Figure 3 : Cas d'utilisation 2 - Gestion du modèle de catégorisation

**Scénario 2 : Gestion des professions**

L’utilisateur clique sur « Emplois ».

Ce scénario est similaire au précédent mais avec les professions au lieu des compétences.

### Cas d’utilisation 3 : Tableau de bord

**Précondition :** être authentifié en tant qu’administrateur

Le tableau de bord est un outil nécessaire pour l’administrateur pour le suivi de la plateforme. En effet, grâce à cet outil, il peut voir le nombre de CV importés, et la pertinence des critères en voyant les critères les plus recherchés par les utilisateurs.

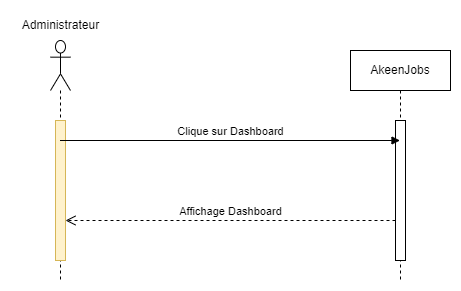


Figure 4 : Cas d'utilisation 3 - Tableau de bord

## Exigences non fonctionnelles transverses

### Utilisabilité

L’ergonomie est un point clé dans le développement d’une application web. Compte tenu qu’on doit livrer une application à caractère professionnelle, l’ergonomie doit rester simple et compréhensible facilement.

Toute la plateforme sera conçue de manière à faciliter la navigation des utilisateurs :

- Le design est simple et intuitif

- L’application est disponible en deux langues différentes : Français et Anglais

### Performances

Les choix des technologies ont été surtout faits pour avoir les meilleures performances. Le chargement des pages devrait être court pour l’expérience utilisateur. Pour cela, les requêtes SQL seront optimisées pour augmenter la rapidité des chargements de page.

### Robustesse

Les données restent cohérentes avec les contrôles de valeurs.

### Sécurité

La sécurité est un critère important pour ce projet. Les mots de passes doivent être cryptés et les API doivent être protégés par des jetons d’authentification, les serveurs de déploiement seront configurés contre des attaques externes.

### Supportabilité (maintenabilité, évolutivité)

L’application est développée avec des langages populaires et offre ainsi une prise en main facile. De plus, on a respecté les règles de nommages standards pour chaque langage pour plus de lisibilité.

## Interfaces détaillées

### IHM

#### Cas d’utilisation 1 : Recherche d’emploi et d’offre

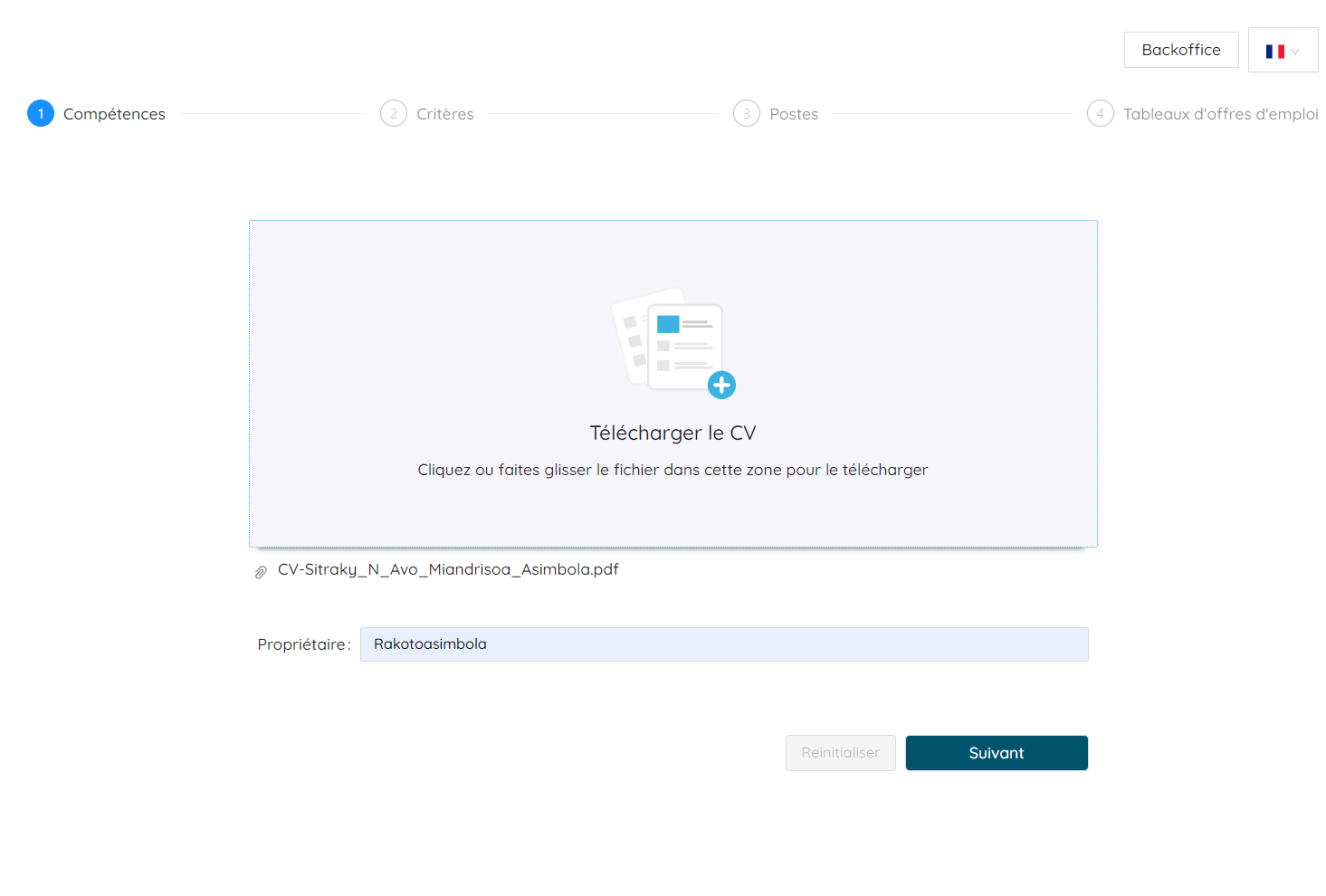


Figure 5 : Import de CV

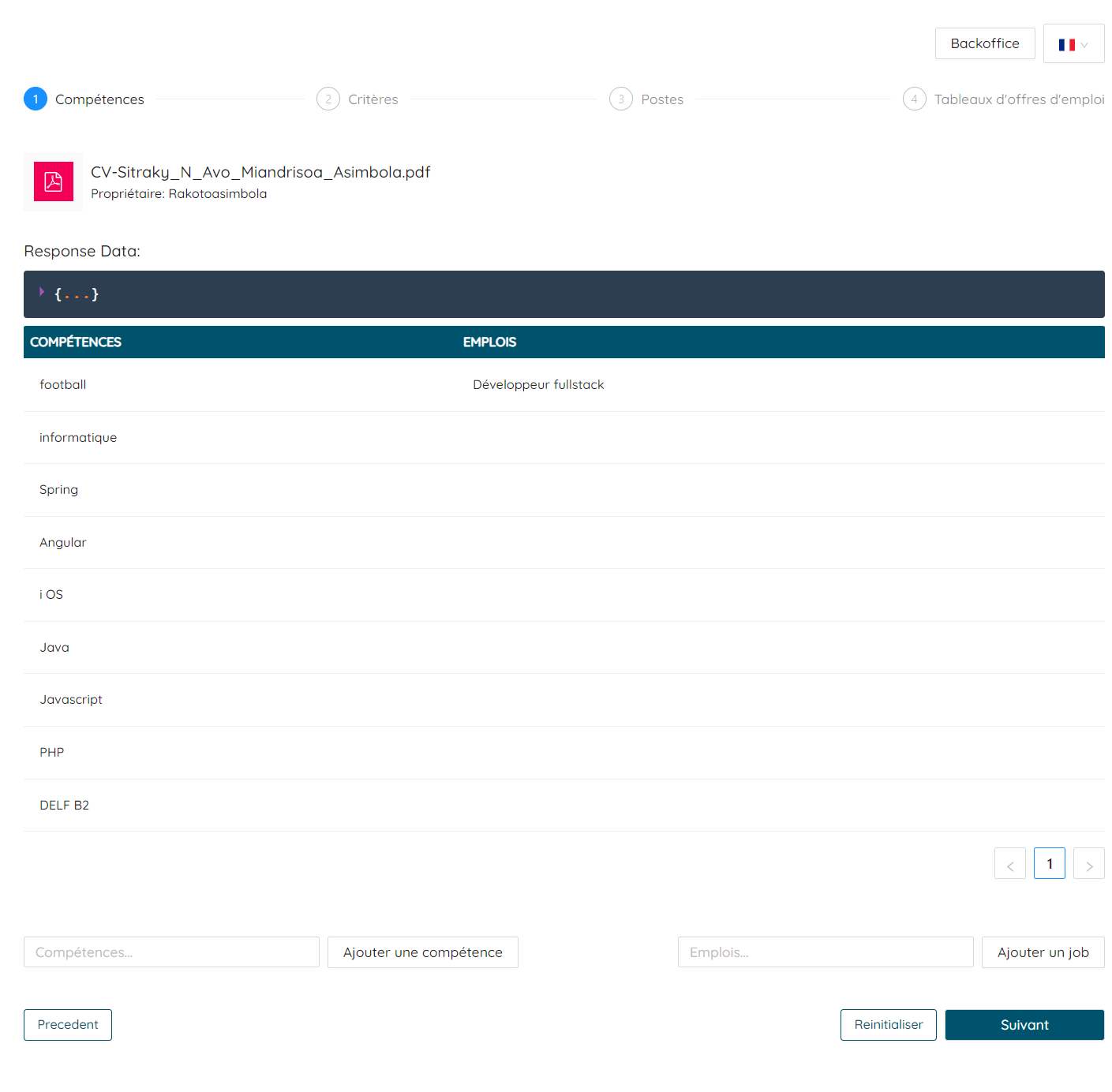


Figure 6 : Liste des compétences et professions

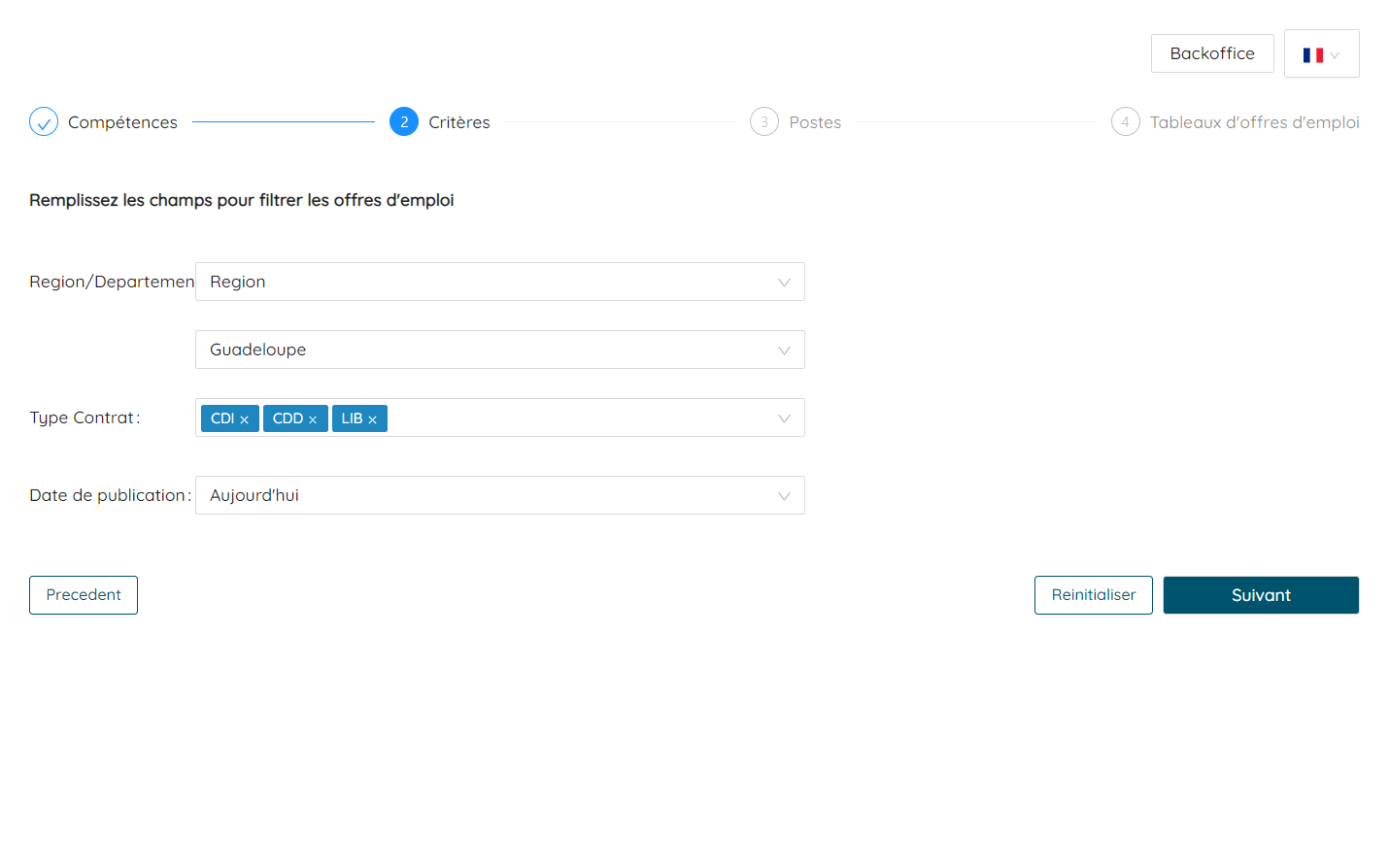


Figure 7 : Choix des critères

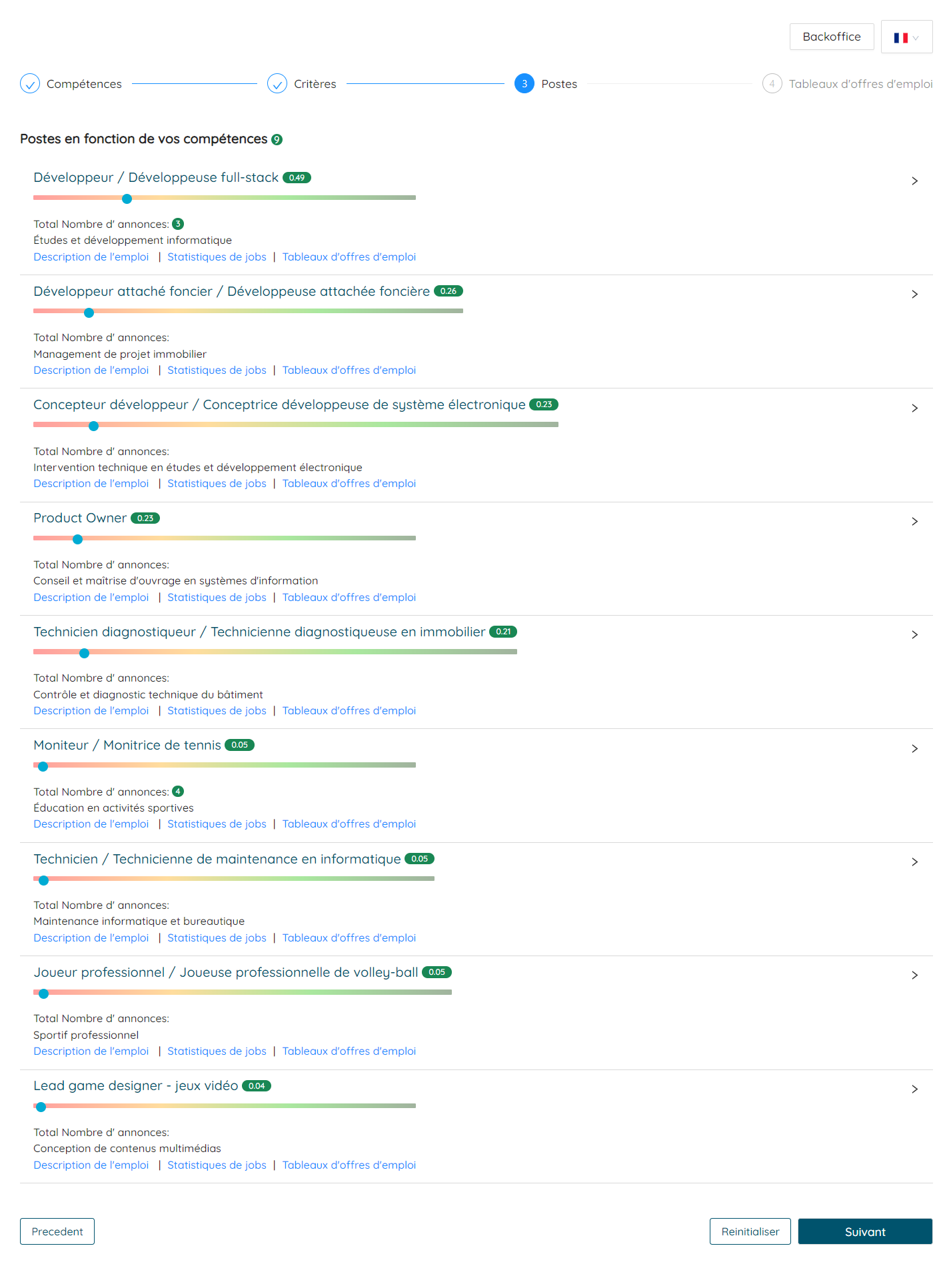


Figure 8 : Liste des emplois recherchés

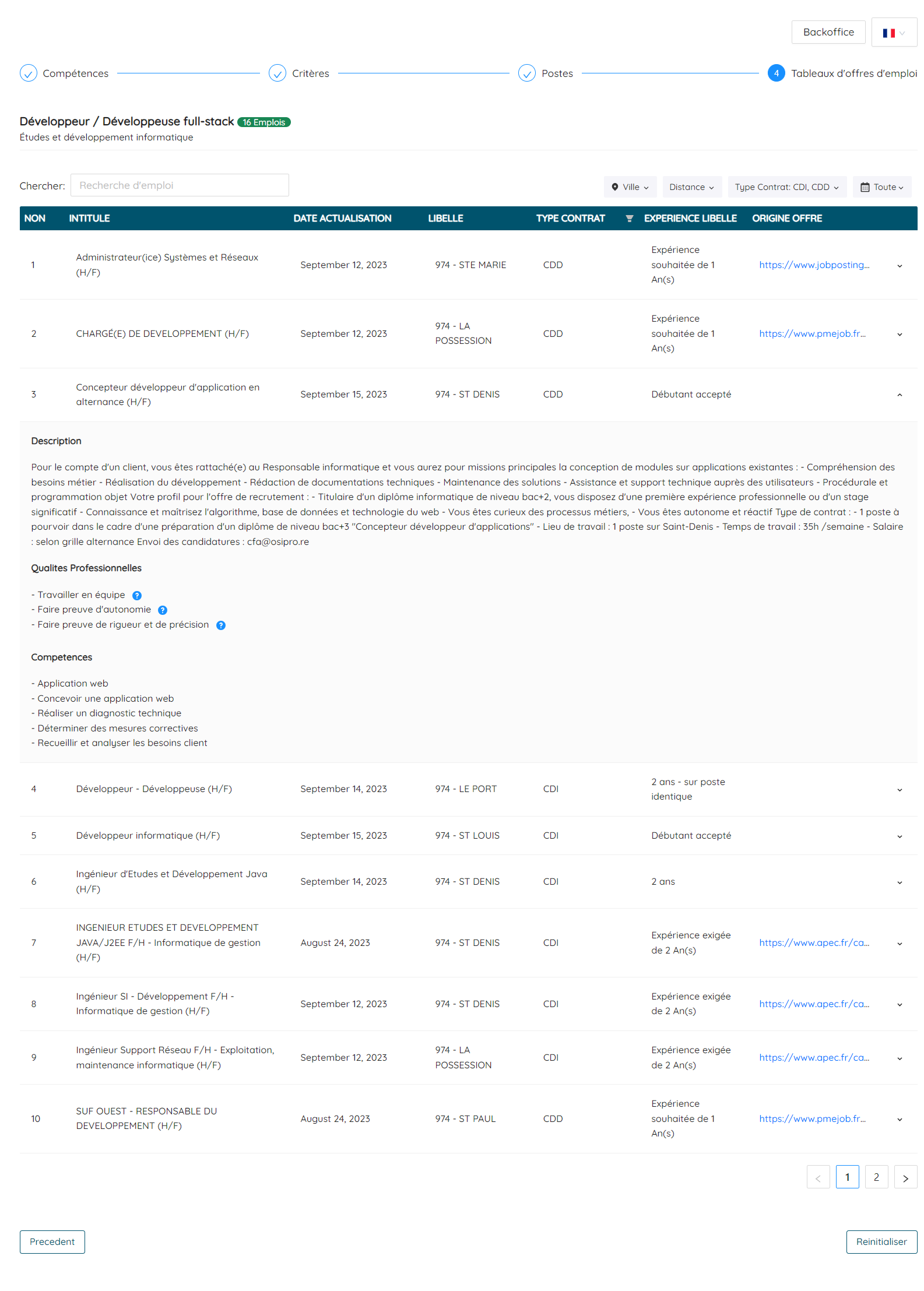


Figure 9 : Offres d'emplois

#### Cas d’utilisation 2 : Gestion du modèle de données

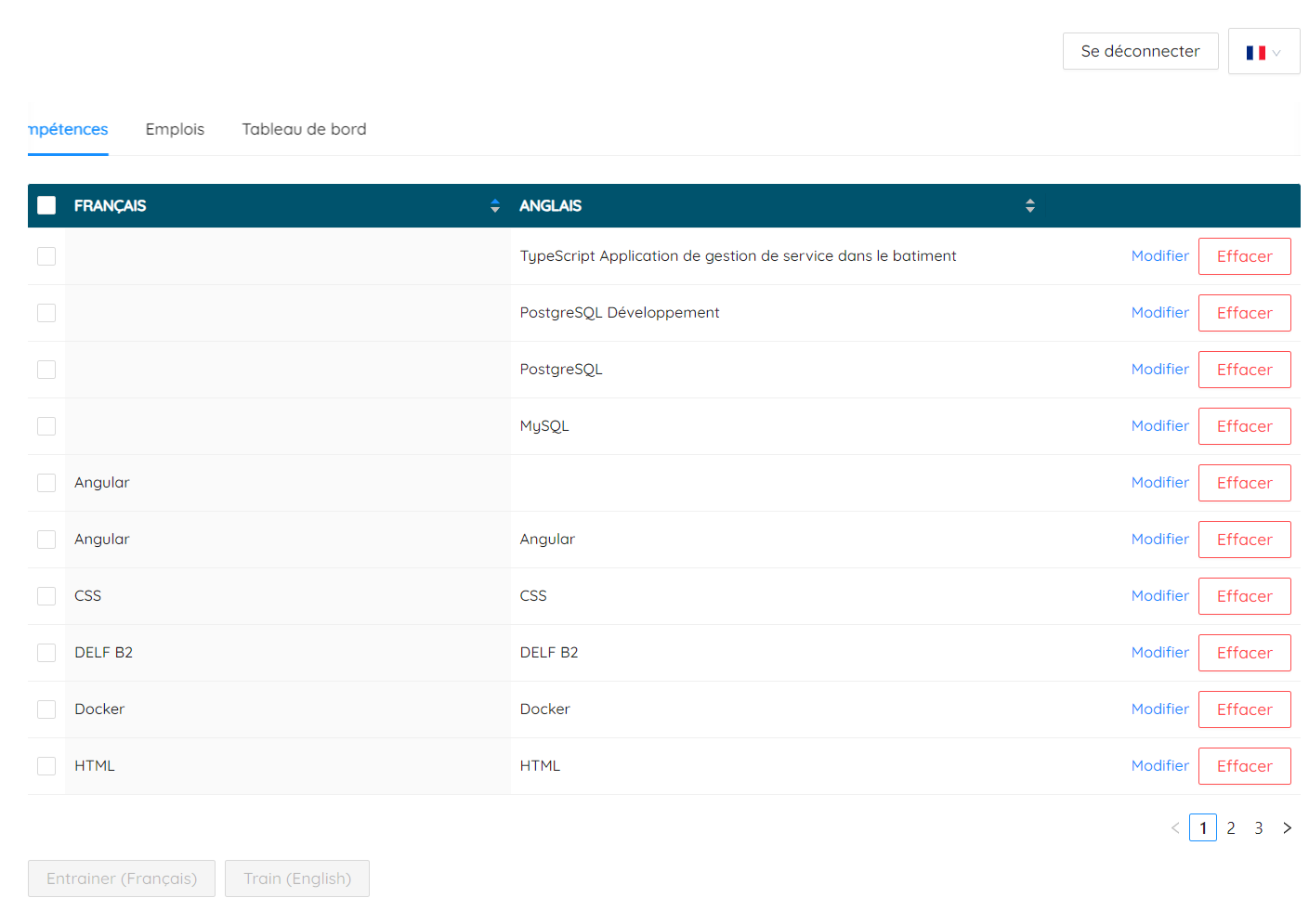


Figure 10 : Gestion des compétences

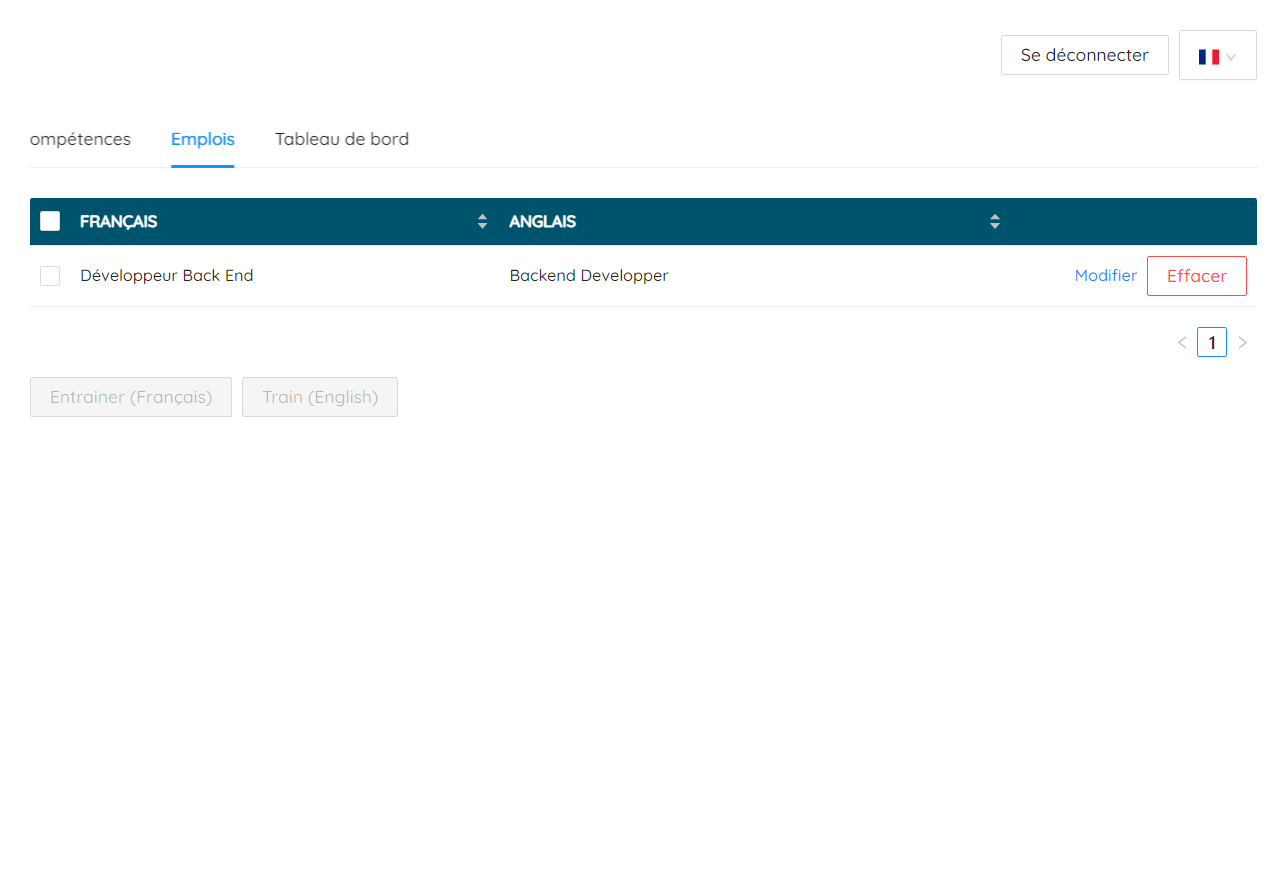


Figure 11 : Gestion des professions

#### Cas d’utilisation 3 : Tableau de bord



Figure 12 : Tableau de bord

### Interfaces avec d’autres systèmes

Concernant l’interface avec d’autres systèmes, plusieurs API ont été créées. Le type de données à échanger ici est le JSON. Pour cause de confidentialité, les APIs ne seront pas décrits dans cette section, la réponse donnée par le système comporte un code.

# Architecture(s) système

## Architecture technique

L’architecture technique s’appuie sur la vision des flux d’information entre les composants d’une part, et le diagramme de déploiement applicatif d’autre part. Elle peut être schématisée de la manière suivante :

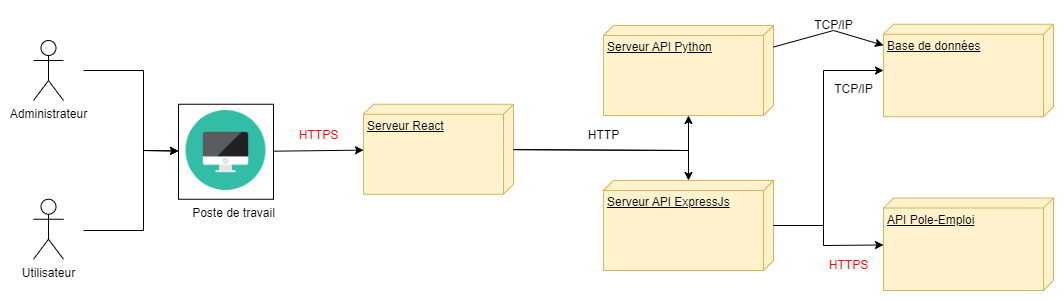


Figure 13: Architecture technique

## Architecture logicielle

L’architecture logicielle est une vue tournée sur l’organisation interne des logiciels du système informatique. Nous pouvons la représenter de la manière suivante :

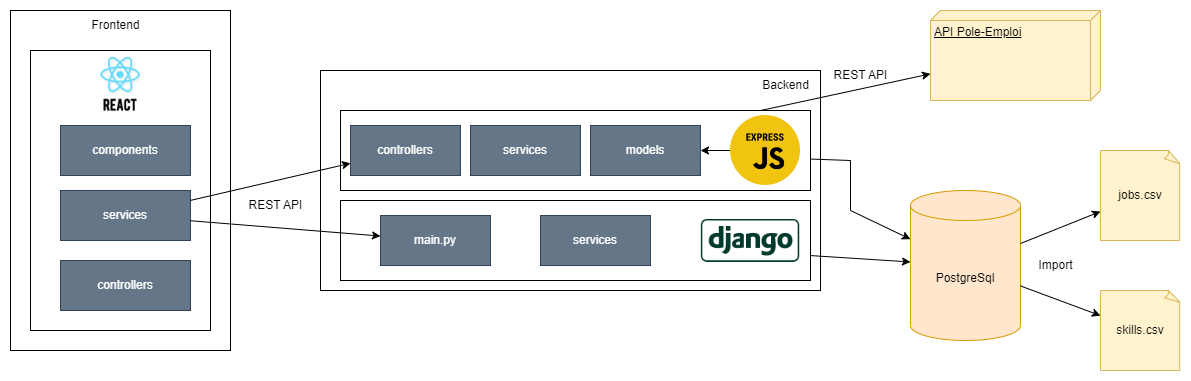


Figure 14 : Architecture logicielle

# Conception du système logiciel réalisée dans le projet (vision interne/développeur)

## Plate-forme technique

### Architecture matérielle

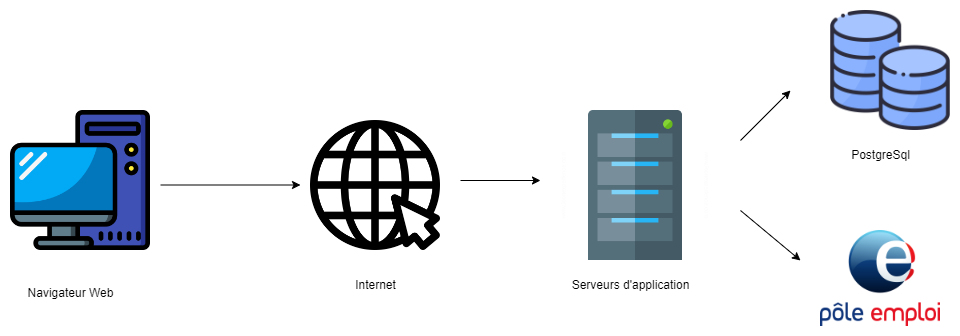


Figure 15 : Architecture matérielle

### Framework logiciel

Nous avons plusieurs nœuds de l’application :

* Frontend : React
* Backend : RESTful API
  + ExpressJs : pour l’authentification et le tableau de bord
  + Python et Django : pour le machine learning
* Base de données : PostgreSql
* Autres : API Pole-Emploi

## Conception du logiciel développé

### Conception du code source

* Pour les APIs codés en javascript ExpressJs, on a adopté l’architecture orientée services en séparant le code source en trois modules :

• Les contrôleurs : contenant la logique des interactions de l’utilisateur

• Les modèles : contenant les modèles de données

• Les services : contenant les fonctionnalités de l’application et les appels aux APIs.

* Pour les APIs de prédiction, nous avons suivi la structure de Django pour faciliter la prise en main de l’application par les membres de l’équipe. Ensuite, pour l’API REST, j’ai choisi cette structure :
* Un dossier « services » contenant les scripts d’analyses.
* Un fichier « main.py » se trouve sur le dossier principal, celui-ci contient les « endpoints » de l’API et lie les requêtes Https aux opérations correspondantes.
* La partie web qui est développée avec React suit une architecture orientée composants, dont le code source est structuré dans les packages suivants :

• components : contenant les composants correspondant aux pages de l’application.

• services : contenant les classes faisant appels aux APIs et les classes utilitaires.

• auth : contenant les classes responsables de l’authentification et des permissions à l’intérieur de l’application tels que les intercepteurs.

### Le code source – vue statique

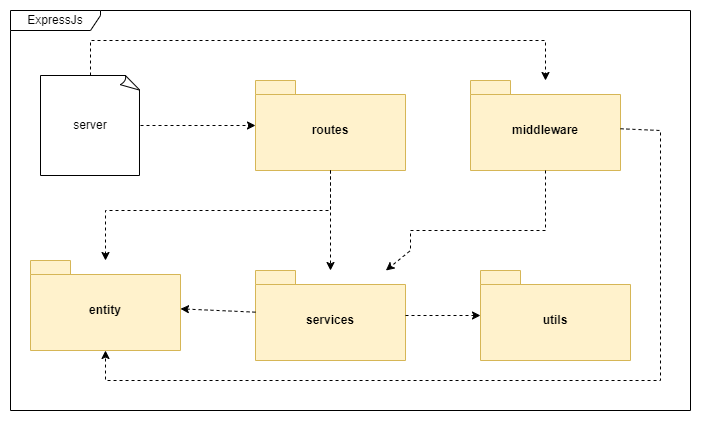


Figure 16 : Vue statique ExpressJs

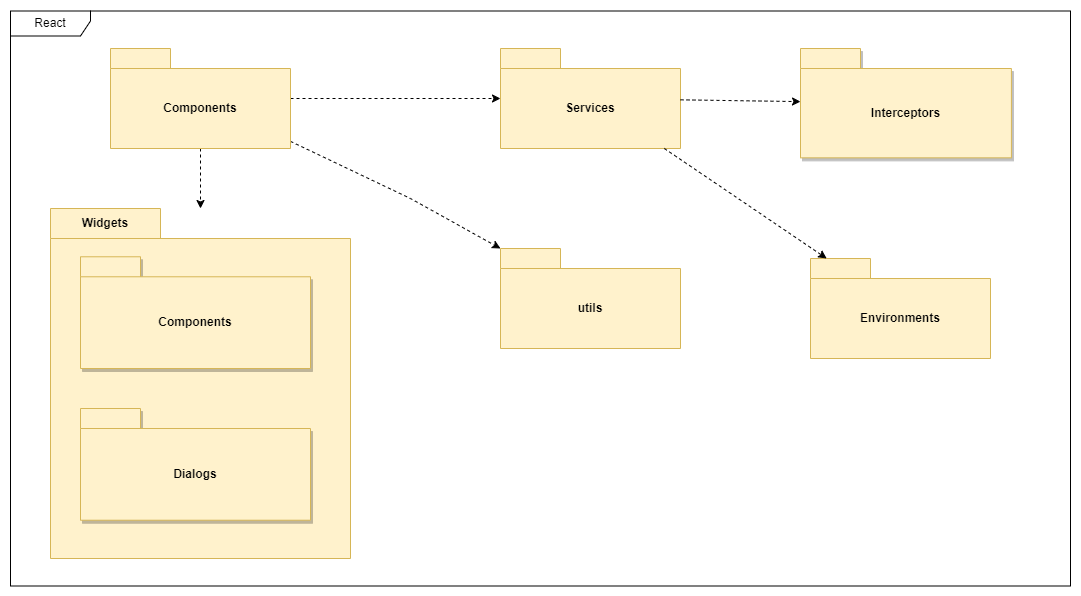


Figure 17 : Vue statique React

### Modélisation de données



Figure 18 : Modèle de données

### Réalisation des cas d’utilisation

La partie la plus importante de notre application est la recherche d’emplois, donc on va détailler ci-dessous la réalisation de ce cas d’utilisation.

Cas d’utilisation : Recherche d’emplois et d’offres

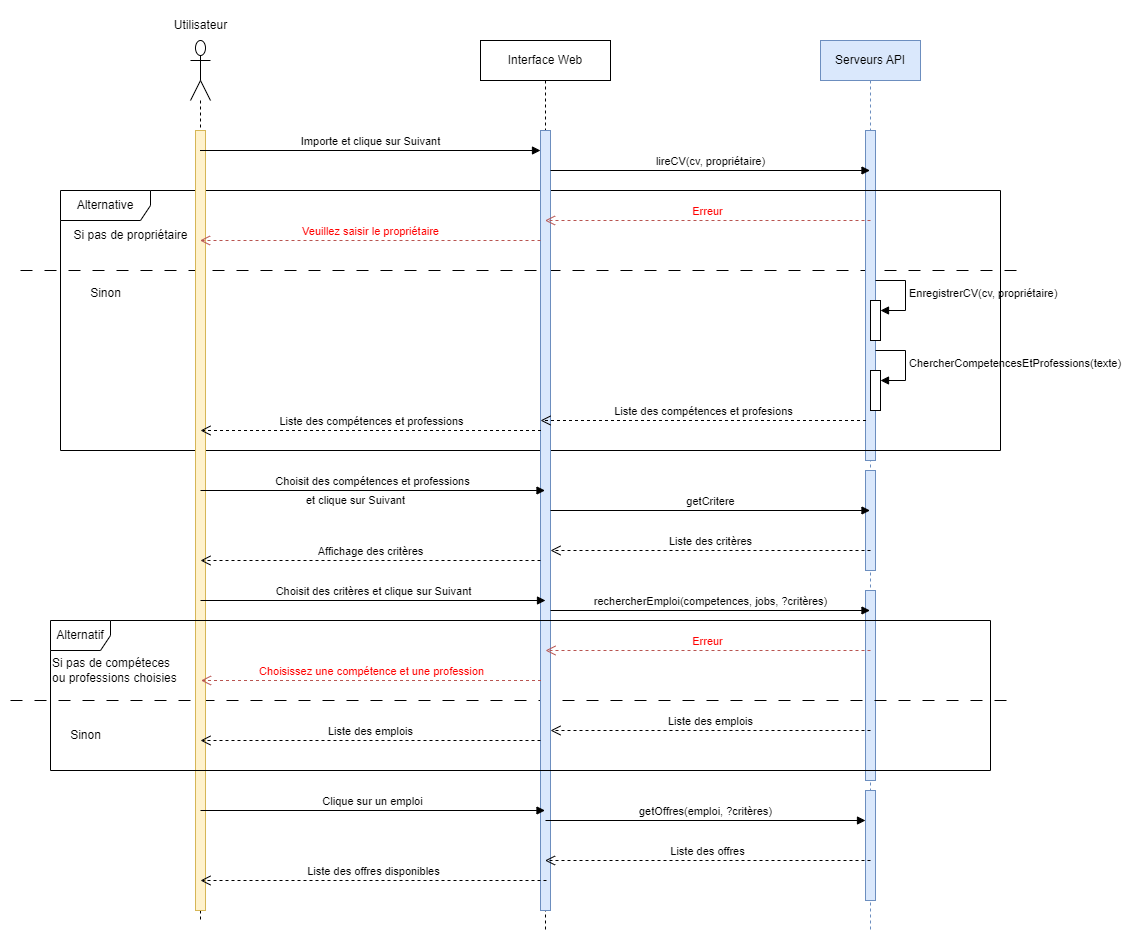


Figure 19 : Réalisation CU 1 - Recherche d'emplois et d'offres

### Les composants et leur déploiement

Toutes les applications sont déployées sur le serveur VPS (Virtual Private Server) de l’entreprise.

On a utilisé docker pour chacune des applications pour créer une image et un conteneur dans le serveur. La base de données est aussi déployée sur le même serveur.

# Tests du système logiciel

Après chaque livraison de jet, l’application est testée, d’abord par mes soins, ensuite par le chef de projet tout en leur expliquant le fonctionnement, en lançant l’application sur mon ordinateur.

Après les tests avec les métiers, je procède à la correction et aux rectifications s’il y en a.

# Conclusion générale

Ainsi, j’ai effectué mon stage de fin d’études en Master au sein de la société Wylog Madagascar. Ces travaux de quatre mois m’ont permis à appliquer les connaissances théoriques que j’ai acquises pendant les études à l’université. Sur le plan personnel, ce période de temps m’a permis de gagner beaucoup d’expériences et de professionnalisme, d’augmenter mes compétences, techniques ainsi qu’humaines.

Concernant AkeenJobs, les objectifs posés au début du stage ont été quasiment atteints:

* L’identification des mots à partir du CV est réalisée.
* L’interface Web est mise en place, sur le serveur de l’entreprise
* Les documents (Conception, DAT, MOP) sur l’outil ont été tous rédigés. Ils sont disponibles dans le dossier de partage de l’entreprise

Bien sûr, nous avons rencontré quelques difficultés durant mon stage, à citer le manque de données pour construire un bon modèle de données sur la catégorisation des mots. Nous avons pu résoudre ce problème en comparant plusieurs sources et en combinant les données. Personnellement, j’ai eu quelques soucis sur le choix du type d’algorithme utilisé pour cette catégorisation, à cause du manque de connaissance sur la catégorisation des mots. Mais au fur et à mesure des recherches faites, j’ai pu faire face à ce problème.

Le projet a encore de la place pour amélioration. En perspective, nous pouvons améliorer notre modèle de données actuel pour avoir plus de précision sur l’identification des mots, ceci à l’aide de nouvelles données accumulées pendant quelque mois, ou bien un meilleur algorithme que je n’ai pas pu appliquer.

# Références et Bibliographie

[1] Taux de chômage dans le monde, La Banque Mondiale, le 20 Août 2023  
<https://donnees.banquemondiale.org/indicateur/SL.UEM.TOTL.ZS>

[2] Nick Bostrom : « L'IA est comme une feuille blanche, elle a le potentiel d'être tout ce que nous pouvons imaginer »

[3] Summer 2021 Internship Report, par Jayawardena, Handapangoda Mudalige Gavindya Nuwandi, le 24 août 2021 - études sur PyPdf2

[4] Yann Lecun : « Les réseaux de neurones profonds sont des boîtes à outils extrêmement polyvalentes pour le traitement des données, avec un potentiel pratiquement illimité ».

[5] Pole-Emploi, le 02 juillet 2023   
<https://www.pole-emploi.fr/accueil/>

[6] CarreerBuilder, le 10 septembre 2023  
[https://www.careerbuilder.com/](https://www.careerbuilder.com/regional_sites)

[7] IEEE , 10 avril 2014 - A Review of Natural Language Processing Research, publié sur [IEEE Computational Intelligence Magazine](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=10207) ( Volume: 9, [Issue: 2](https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6786379&punumber=10207), May 2014), page 48-57

# Annexes

Annexe 1 : Interface Login administrateur

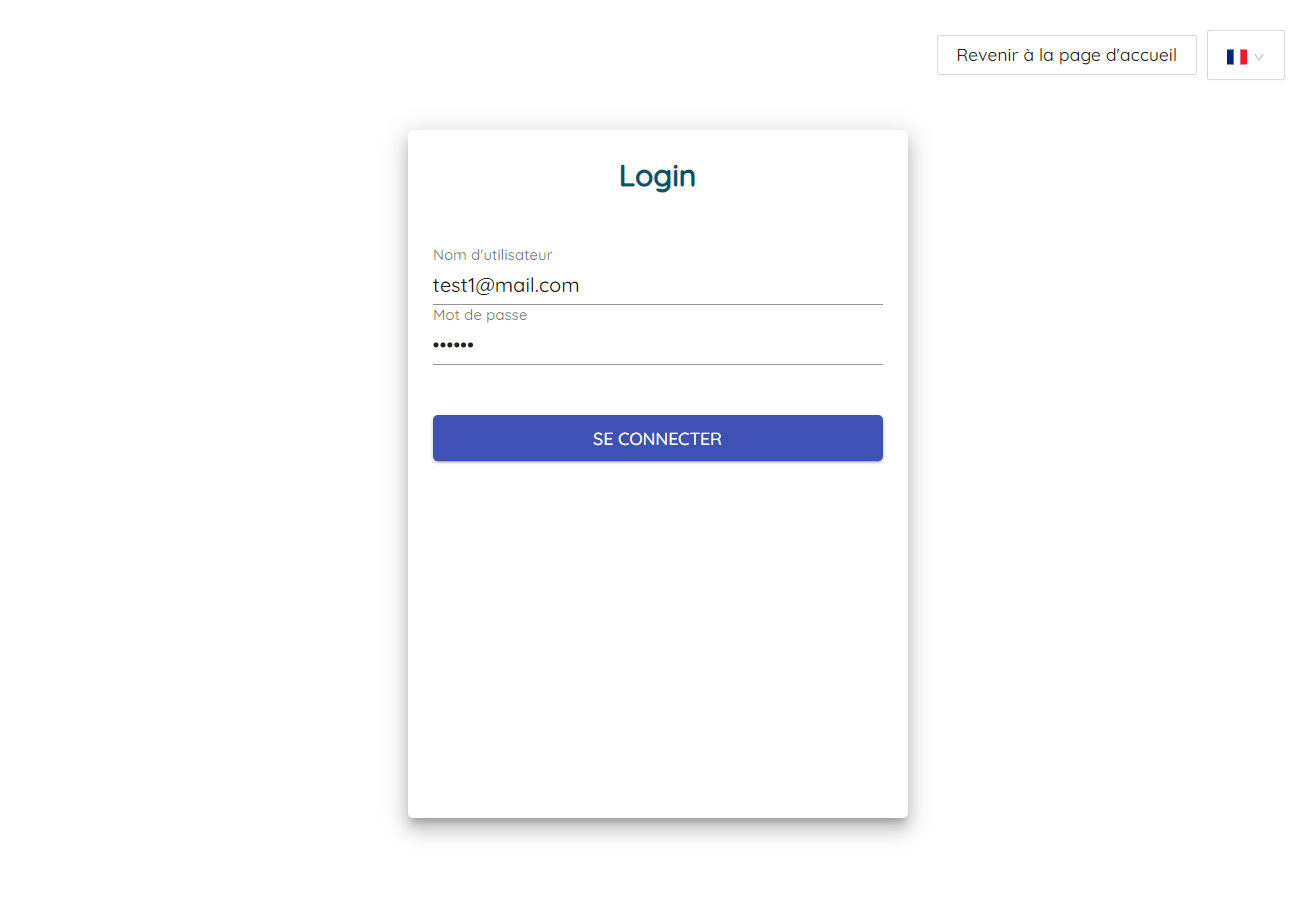


Figure 20 : Interface authentification

Annexe 2 : changement de langue

Le changement de langue se fait en cliquant le petit menu en haut de la page et en cliquant sur la langue.  
Quand le changement se fait, toutes pages de l’applications seront traduites en la langue voulue.

1. TIC : Technologies de l’Information et de la Communication [↑](#footnote-ref-1)