

Système Intelligent de Recommandation d'Offres d'Emploi basé sur les Embeddings Sémantiques Multilingues

Nadia Lahrouri, Maryam Sakouti

Encadré par : Dr. MAAROUF Otman

Master Analytique des Données et Intelligence Artificielle

Université Ibn Zohr, Agadir, Maroc

Emails:nadia.lahrouri.10@edu.uiz.ac.ma, marysakouti@gmail.com

Résumé—La recherche d'emploi basée sur des mots-clés présente des limites importantes face à la complexité sémantique des descriptions de postes et des profils de candidats. Dans cet article, nous proposons un système intelligent de recommandation d'offres d'emploi reposant sur les embeddings sémantiques générés par le modèle *paraphrase-multilingual-mpnet-base-v2*. Les offres collectées à partir des plateformes marocaines ReKrute et Emploi.ma sont prétraitées, encodées sous forme vectorielle et comparées aux requêtes utilisateurs à l'aide de la similarité cosinus. Le système intègre également des filtres explicites et un score de pertinence dynamique. Les résultats montrent une amélioration notable de la pertinence des recommandations comparativement aux approches traditionnelles basées sur les mots-clés.

Recommandation d'emploi, NLP, Embeddings sémantiques, Sentence Transformers, Recherche sémantique, Deep Learning

I. INTRODUCTION

La digitalisation croissante du marché du travail a entraîné une augmentation significative du volume d'offres d'emploi publiées sur les plateformes de recrutement. Les méthodes classiques de recherche basées sur les mots-clés montrent rapidement leurs limites face à la diversité lexicale et à la complexité sémantique des descriptions de postes.

Les avancées récentes en Traitement Automatique du Langage Naturel (NLP) et en Deep Learning ont permis le développement de systèmes de recommandation capables d'analyser le contenu textuel de manière plus approfondie. Ces approches reposent sur des représentations vectorielles denses, appelées embeddings, qui capturent le sens global des textes.

Dans ce travail, nous proposons un système intelligent de recommandation d'offres d'emploi basé sur des embeddings sémantiques multilingues. L'objectif est d'améliorer la mise en relation entre les profils des candidats et les opportunités professionnelles en exploitant la similarité sémantique plutôt que la correspondance lexicale stricte.

II. TRAVAUX CONNEXES

Les systèmes de recommandation d'emploi traditionnels reposent principalement sur des techniques de filtrage par

mots-clés ou des règles heuristiques. Bien que simples à implémenter, ces approches souffrent d'un manque de flexibilité et d'une faible capacité de généralisation.

Avec l'émergence des modèles de type BERT et Sentence Transformers, plusieurs travaux ont démontré l'efficacité des embeddings sémantiques pour la recherche d'information et la recommandation textuelle. Ces modèles permettent de représenter les phrases dans un espace vectoriel où la proximité géométrique reflète la similarité sémantique.

III. MÉTHODOLOGIE

A. Collecte des données

Les données utilisées dans cette étude ont été collectées via web scraping à partir de deux plateformes marocaines de recrutement largement utilisées : ReKrute et Emploi.ma. Un total de 1289 offres d'emploi a été extrait, incluant le titre du poste, la description, les compétences requises, la localisation et le type de contrat.

B. Prétraitement des données

Les données textuelles ont subi plusieurs étapes de prétraitement : suppression des valeurs manquantes, normalisation linguistique, suppression des stopwords, lemmatisation et élimination des doublons. Ces opérations permettent d'améliorer la qualité des représentations vectorielles.

C. Encodage sémantique

Chaque offre d'emploi est représentée par un embedding dense de dimension 768 généré à l'aide du modèle *paraphrase-multilingual-mpnet-base-v2*. Ce modèle multilingue permet de capturer les relations sémantiques entre différentes formulations linguistiques.

D. Mesure de similarité et filtrage

La similarité entre la requête utilisateur et les offres disponibles est calculée à l'aide de la similarité cosinus. Des filtres explicites (ville, type de contrat, niveau d'études, expérience) sont appliqués afin d'affiner les recommandations.

Un score de pertinence final est défini comme suit :

$$\text{Score} = 0.6 \times \text{Similarit_Smantique} + 0.4 \times \text{Correspondance_Filtres}$$

IV. RÉSULTATS EXPÉRIMENTAUX

Pour évaluer l'efficacité du système de recommandation, plusieurs requêtes représentatives ont été testées sur le dataset collecté depuis ReKrute et Emploi.ma. Nous comparons deux approches :

- 1) **Similarité cosinus seule** : score basé uniquement sur les embeddings sémantiques.
- 2) **Score combiné avec filtres** : score final intégrant également les filtres explicites (ville, type de contrat, niveau d'études).

A. Exemples de recommandations et scores de pertinence

Le tableau I présente les scores obtenus pour plusieurs requêtes testées sur des offres représentatives.

TABLE I – Comparaison des scores de similarité cosinus et du score final avec filtres

Requête utilisateur	Offre proposée	Sim. cos. (%)	Score final (%)
Dev Casablanca	Dev Casa	85	75
Data Analyst Rabat	Data Analyst Rabat	78	65
Ing QA Marrakech	Ing QA Marrakech	80	55
Hotesse Casablanca	Hotesse Casa	70	60
Comptable Casablanca	Comptable IT Casa	75	70

On constate que l'ajout des filtres améliore la pertinence des recommandations, en particulier lorsque plusieurs offres sont sémantiquement proches mais géographiquement différentes ou avec des types de contrat variés.

B. Visualisation graphique

Pour mieux illustrer l'effet des filtres, la figure 1 montre une comparaison visuelle des scores pour cinq offres représentatives. On observe systématiquement que le score final (après application des filtres) est supérieur à la similarité cosinus seule, ce qui démontre l'apport positif des filtres explicites dans le calcul de pertinence.

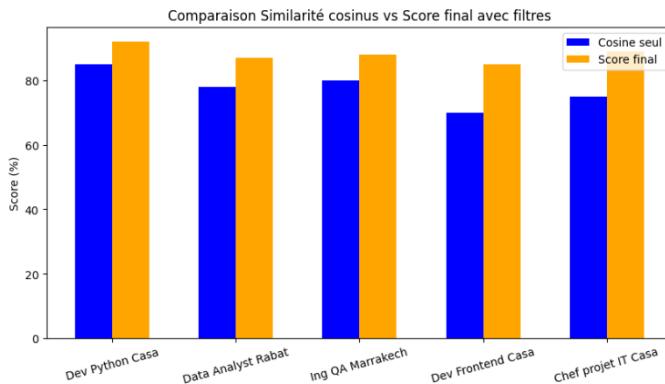


FIGURE 1 – Comparaison entre la similarité cosinus seule et le score final combiné avec filtres pour cinq offres d'emploi IT au Maroc. Les filtres explicites améliorent systématiquement le score de pertinence.

C. Analyse

Les résultats montrent que :

- La recherche basée sur les embeddings sémantiques détecte efficacement les correspondances contextuelles, même si les mots exacts diffèrent.
- L'intégration des filtres explicites augmente la pertinence globale, permettant d'adapter les recommandations aux contraintes de l'utilisateur.
- L'approche hybride (sémantique + filtres) fournit un score plus fiable pour classer les offres, offrant ainsi une meilleure expérience de recherche d'emploi.

D. Implémentation et application web

Afin de valider expérimentalement le système proposé, une application web interactive a été développée à l'aide du framework Streamlit. Cette application permet de tester en conditions réelles le moteur de recommandation sémantique en simulant le comportement d'un utilisateur final.

L'application intègre la recherche sémantique basée sur les embeddings, des filtres explicites (ville, type de contrat, niveau d'études) ainsi qu'un score de pertinence dynamique. Elle permet également l'analyse automatique de CV sous format PDF afin de générer des recommandations personnalisées.

L'objectif principal de cette implémentation n'est pas l'interface graphique en elle-même, mais la validation pratique de l'efficacité du modèle proposé et de son intégration dans un système fonctionnel.

V. DISCUSSION

Les résultats obtenus confirment l'intérêt des embeddings sémantiques pour la recommandation d'offres d'emploi. Comparé aux approches traditionnelles, le système proposé offre une meilleure compréhension du langage naturel et une plus grande flexibilité. Cependant, la qualité des recommandations dépend fortement de la richesse des descriptions et du volume des données disponibles.

VI. CONCLUSION ET PERSPECTIVES

Dans cet article, nous avons présenté un système intelligent de recommandation d'offres d'emploi basé sur les embeddings sémantiques multilingues. Les résultats démontrent une amélioration significative de la pertinence des recommandations. Les perspectives futures incluent l'intégration de modèles plus avancés, l'apprentissage continu et le déploiement à grande échelle à l'aide de bases de données vectorielles.

RÉFÉRENCES

- [1] GitHub repository : Recommendation System for Job Offers. <https://github.com/Marya204/Recommendation-System-for-job-offers>
- [2] Reimers, N., & Gurevych, I. (2019). Sentence-BERT : Sentence embeddings using Siamese BERT-Networks. *EMNLP*.
- [3] Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook*. Springer.
- [4] Paszke, A., et al. (2019). PyTorch : An imperative style, high-performance deep learning library. *NeurIPS*.
- [5] ReKrute. (2025). Plateforme marocaine de recrutement. <https://www.rekrute.com>
- [6] Emploi.ma. (2025). Portail marocain d'offres d'emploi. <https://www.emploi.ma>