SkiF Maker : un processus de classement du meilleur talent par rapport à une offre de mission

Jihen Karoui, Kamilia Hafid, Stéphane Pezeril

▶ **To cite this version:**

Jihen Karoui, Kamilia Hafid, Stéphane Pezeril. SkiF Maker : un processus de classement du meilleur talent par rapport à une offre de mission. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4 : Démonstrations et résumés d'articles internationaux, Jun 2020, Nancy, France. pp.38-41. hal-02768507v3

HAL Id: hal-02768507
https://hal.science/hal-02768507v3

Submitted on 23 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

SkiF Maker : un processus de classement du meilleur talent par rapport à une offre de mission

Jihen Karoui[1] Kamilia Hafid[1] Stéphane Pezeril[1]
(1) AUSY R&D, 6 rue Troyon, 92310 Sèvres Cedex, France
jkaroui@ausy.fr, kamilia.hafid@ausy.fr, spezeril@ausy.fr

RÉSUMÉ

SkiF Maker présente une solution innovatrice pour toute entreprise de services numériques (ESN) visant à améliorer son processus de recrutement de talents, la gestion des compétences de ses ressources humaines et la satisfaction des besoins de ses clients. Cet outil offre aux ESN un gain considérable en terme de temps de travail et leur évite un travail manuel souvent laborieux et non plaisant.

ABSTRACT

SkiF Maker : a process for ranking the best talent in relation to an assignment offer.

SkiF Maker presents an innovative solution for any digital services company aiming to improve its talent recruitment process, the management of its human resources skills and the satisfaction of the customer's requests. This tool provides digital service companies a considerable gain in terms of working time and avoids manual labor that is often laborious and unpleasant.

MOTS-CLÉS : Fouille de texte, annotation sémantique, apprentissage non supervisé, entité nommée.

KEYWORDS: Text mining, semantic annotation, unsupervised learning, named entity.

1 Contexte

Dans le métier du conseil en Hautes Technologies, le Curriculum-Vitae (CV) devient un outil du quotidien parce qu'il sert à mettre en avant les compétences d'un-e consultant-e. Le CV est un document commun, généralisé et incontournable dans le cadre d'un exposé de parcours professionnel mais il est loin d'avoir un contenu standardisé. Il est alors primordial d'adopter un format commun à tou-te-s, appelé "Dossier de Compétences (DC)". Le problème majeur est de trouver et d'ordonner les DC pour associer à l'offre de mission celui qui est le plus approprié. Appréhender un CV est donc une tâche complexe puisqu'il s'agit tant d'une analyse de structure de l'information sur une page que d'une langue.

L'évolution du marché du travail a prouvé que les méthodes classiques de recrutement ainsi que l'attribution des missions aux consultants sont devenues moins performant. Ceci est justifié par le développement considérable du recrutement en ligne ou l'e-recrutement ces dernières années (Faliagka *et al.*, 2014). Cette expansion a conduit à une croissance continue du nombre de portails de l'emploi et d'agences de recrutement sur Internet (Chen *et al.*, 2015), (Schmitt, 2016). Cela a également entraîné une augmentation constante du nombre de demandeurs d'emploi à la recherche de nouvelles opportunités de carrière (Hauff & Gousios, 2015). En conséquence, les portails d'emploi

Volume 4 : Démonstrations et résumés d'articles internationaux, pages 38–41.
HAL : hal-02768507.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

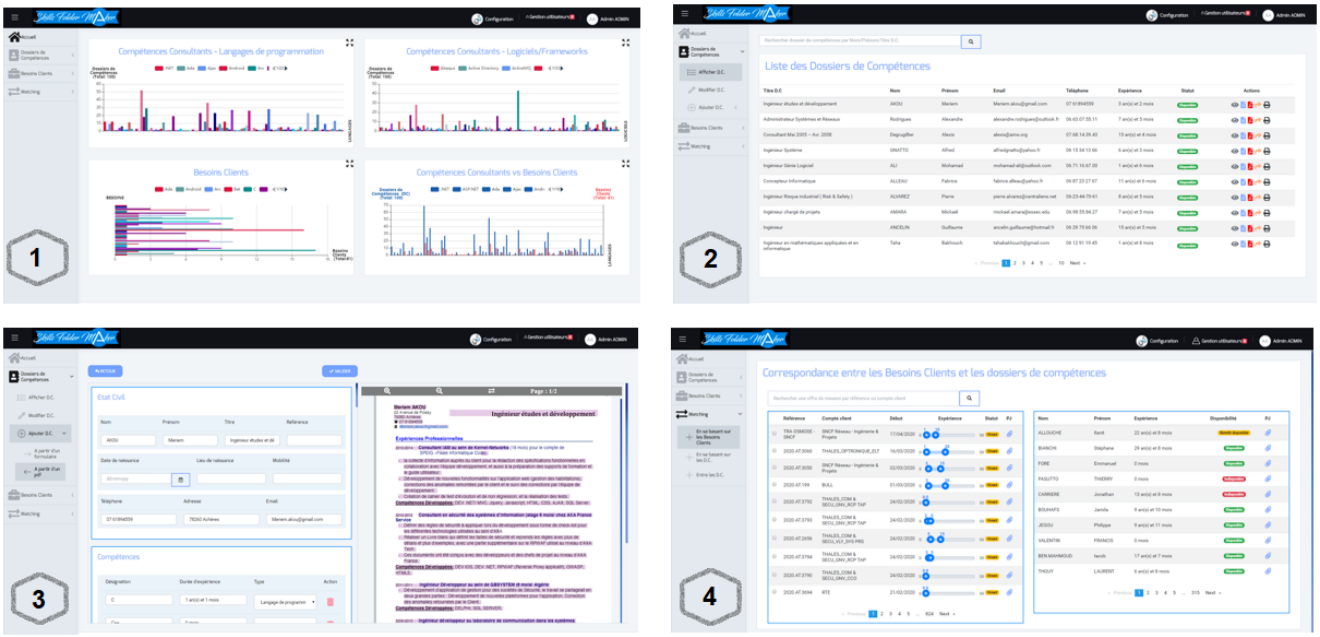

FIGURE 1 – Quelques interfaces graphiques de SkiF Maker

en ligne commencent à recevoir des milliers de CV (de styles et de formats divers) de demandeurs d'emploi possédant différents domaines d'expertise et spécialisés dans différents domaines (Kmail *et al.*, 2015). L'analyse du contenu des CV ainsi que l'extraction des informations utiles pour la correspondance automatique entre les CV et les besoins clients n'a pas été résolu dans les différentes solutions présentes sur le marché aujourd'hui. Devant cette problématique, le développement d'une solution devient crucial et représente un challenge pour le TAL.

2 Le projet SkiF Maker

SkiF Maker est une solution innovatrice pour la gestion des compétences des salariés et des nouvelles recrues vs. les besoins clients. C'est un outil permettant aux conseillers en Hautes Technologies (voir figure 1) :

1. avoir une vision sur le marché à travers des dashboards qui présentent des statistiques sur les compétences des salariés et les compétences demandées par les clients (voir image N°1 de la figure 1).
2. la visualisation des données de l'entreprise relatives aux consultants (voir image N°2 de la figure 1).
3. l'extraction d'information et la génération automatique des DC à partir des CV des candidats (voir image N°3 de la figure 1).
4. la correspondance automatique entre les DC et les fiches de missions (voir image N°4 de la figure 1).

Dans le cadre de ce projet, nous avons dû faire face à plusieurs problématiques. La première problématique consiste à l'identification des différentes sections d'un CV ainsi que la **reconnaissance des entités nommés (EN)** présentant l'*état civil* (nom, prénom), les *compétences* (langages de programmation, logiciels, framework), les *formations* (intitulé des diplômes, écoles, universités), etc.

Cela nécessite l'exploitation des vocabulaires de domaine d'étude que nous avons construits. Pour la construction de ces vocabulaires, nous avons exploité une ressource existante à savoir **wiki-français**. Vu l'absence d'une ressource complète de vocabulaire spécifique au domaine de l'ingénierie informatique, nous avons opté pour un enrichissement manuel du vocabulaire issue de *wiki-français* dans un premier temps ensuite nous avons automatisé la veille technologique en interrogeant *Google* via son API afin de construire un vocabulaire spécifique au domaine de l'ingénierie informatique.

L'automatisation de la veille technologique s'effectue selon l'algorithme suivant. Considérons $WordsVocab$ l'ensemble de mots excluant les mots vides qui sont présents dans la section compétences d'un CV. Soit LP (langage de programmation), F (framework) et L (logiciel) les catégories recherchées pour le classement du vocabulaire. L'algorithme proposé est défini comme suit :

Étapes de l'algorithme

1. Segmenter $WordsVocab$ en un ensemble de tokens T.
2. Générer trois requêtes pour chaque token :
   (a) $Q1 = T \cup LP$
   (b) $Q2 = T \cup F$
   (c) $Q3 = T \cup L$
   soumettre à Google ces trois requêtes (une après l'autre) qui renverra pour chaque requête 20 résultats au maximum formés d'un titre et d'un extrait (snippet).
3. A partir des résultats renvoyés par Google, ne garder que les deux premiers. Ensuite, pour chaque résultat, si les mots-clés de la requête Google ont été trouvés dans le titre ou dans l'extrait, alors T appartient à la catégorie trouvée (LP ou F ou L).
4. Vérifier la présence de T dans le vocabulaire, si T n'existe pas alors il sera ajouter.

Cet algorithme nous a permis de construire un vocabulaire formé de 1200 termes. Pour évaluer l'exactitude de la catégorisation fourni par la veille technologique automatique, trois annotateurs ont également été invités à étiqueter l'ensemble des termes de $WordsVocab$. Les annotateurs ont confirmé l'exactitude des résultats fourni par l'algorithme.

Après avoir générer les DC et afin de résoudre la deuxième problématique, nous avons exploité un processus d'attribution de scores pour les compétences de chaque salarié/candidat. Ce processus consiste à calculer des scores pour l'ensemble des compétences présentes dans un DC en prenant en considération plusieurs critères. Parmi ces critères, nous citons le nombre d'année d'expérience global, le nombre d'année d'expérience pour chaque technos cité dans le DC, les compétences fonctionnelles effectués par le salarié durant ses expériences professionnelles précédentes (conception, développement, rédaction de la documentation, etc.). Cela a nécessité l'**analyse et l'annotation sémantique** du descriptif des expériences précédentes. Ces différents scores ont été exploités dans la phase d'**apprentissage non supervisé** et plus précisément le **Clustering** [1] afin que le système arrive à identifier et trier les DC qui correspondent à un besoin client donné et de retourner aux décideurs des ESN (manager, directeur, etc.) les meilleurs profils présents dans l'entreprise et les meilleurs profils à recruter.

1. Pour l'apprentissage, nous avons testé *K-means* avec deux mesures de similarité à savoir : la *distance Euclidienne* et la *distance de Manhatta*n. Les premières expériences ont prouvé que la *distance de Manhattan* permet d'obtenir les meilleurs résultats.

Références

CHEN J., NIU Z. & FU H. (2015). A novel knowledge extraction framework for resumes based on text classifier. In *International Conference on Web-Age Information Management*, p. 540–543 : Springer.

FALIAGKA E., ILIADIS L., KARYDIS I., RIGOU M., SIOUTAS S., TSAKALIDIS A. & TZIMAS G. (2014). On-line consistent ranking on e-recruitment : seeking the truth behind a well-formed cv. *Artificial Intelligence Review*, **42**(3), 515–528.

HAUFF C. & GOUSIOS G. (2015). Matching github developer profiles to job advertisements. In *Proceedings of the 12th Working Conference on Mining Software Repositories*, p. 362–366 : IEEE Press.

KMAIL A. B., MAREE M., BELKHATIR M. & ALHASHMI S. M. (2015). An automatic online recruitment system based on exploiting multiple semantic resources and concept-relatedness measures. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, p. 620–627 : IEEE.

SCHMITT, T. P. C. M. S. (2016). Matching jobs and resumes : a deep collaborative filtering task. In *Proceedings of the 2nd Global Conference on Artificial Intelligence*, p. 1–14.